

Cuffless ambulatory blood pressure measurement using the photoplethysmogram and the electrocardiogram



Eoin Finnegan
St. Cross College
University of Oxford

Supervised by:
Prof. Lionel Tarassenko
Prof. Mauricio Villarroel

A thesis submitted for the degree of
Doctor of Philosophy
Hilary 2023

Declaration

I declare that this thesis is entirely my own work and, except where stated, describes my own research.

Eoin Finnegan
St. Cross College

Abstract

Blood pressure (BP), as with other vital signs such as heart rate and respiratory rate, exhibits endogenous oscillations over a period of approximately 24 hours, a phenomenon known as circadian rhythmicity. This rhythm typically reaches a nadir during sleep, however, different BP circadian rhythm phenotypes exist depending on the magnitude and direction of the nocturnal change. Analysis of these phenotypes has been shown to be an independent indicator for the onset of cardiovascular disease, the leading cause of non-communicable mortality and morbidity worldwide. However, currently the established technique for monitoring BP over 24 hours in the general population requires an inflatable cuff wrapped around the upper arm. This procedure is highly disruptive to sleep and daily life, and therefore rarely performed in primary care. Although commercial cuffless BP devices do exist, their accuracy has been questioned, and consequently, the clinical community do not recommend their use.

In this thesis, I investigated techniques to measure BP in an ambulatory environment without an inflatable cuff using two signals commonly acquired by wearable sensors: the photoplethysmogram (PPG) and the electrocardiogram (ECG). Given the diverse mechanisms by which the autonomic nervous system regulates BP, I developed methodologies using data from multiple individuals with BP perturbed by various, diverse, mechanisms.

To identify surrogate measures of BP derived from the PPG and ECG signals, I designed a clinical study in which significant BP changes were induced through a pharmacological intervention in thirty healthy volunteers. Using data from this study, I established that changes in the pulse arrival time (PAT, the time delay between fiducial points on the ECG and PPG waveforms) and morphological features of the PPG waveform could provide reliable cuffless indicators for changes in BP. Even at rest, however, these signals are confounded by factors such as the pre-ejection period (PEP) and signal measurement noise. Additionally, accurate absolute measurements of BP required calibration using a reference BP device.

Subsequently, I conducted a circadian analysis of these surrogate measures of BP using a large cohort of 1,508 patients during the 24-hour period prior to their discharge from an intensive care unit. Through this circadian analysis I suggest that PAT and a subset of features from the PPG waveform exhibit a phenotypically modified circadian rhythm in synchronicity with that of BP. Additionally, I designed a novel ordinal classification algorithm, which utilised circadian features of these signals, in order to identify BP circadian rhythm profiles in a calibration-free manner. This method may provide a cost-effective initial assessment of BP phenotypes in the general population. Notably, estimating absolute BP values using PPG and ECG signals in the ICU resulted in clinically significant mean absolute errors of 9.26 (5.01) mmHg.

Finally, I designed a clinical study to extend the work towards cuffless ambulatory BP estimation in a cohort of fifteen healthy volunteers. *Hybrid calibration* strategies

(where model personalisation was handled by user demographics, commonly utilised by commercial cuffless devices) led to clinically significant errors when estimating absolute values of BP, mean absolute error = 9.62 (19.73) mmHg. For the majority of individuals, a more appropriate estimation of BP values was achieved through an *individual calibration* strategy whereby idiosyncratic models were trained on personalised data, mean absolute error = 5.45 (6.40) mmHg. However, for a handful of individuals, notable estimation errors (> 10 mmHg) still persisted using this strategy largely as a result of motion artifacts, inherent intra- and inter-individual variability in PPG features, and inadequate training data.

Overall, I suggest that while beat-by-beat measurements of BP can be obtained using PPG and ECG signals, their accuracy is significantly limited in an ambulatory environment. This limitation, combined with the impracticality of *individual calibration* (due to the low tolerance for ABPM), suggest that cuffless ambulatory blood pressure measurement using the PPG and ECG signals may be infeasible. Nevertheless, macro assessments of cardiovascular health, such as an individual's BP phenotype, may be comparatively more accurately predicted using these signals with the potential to be recorded without calibration. Through further research on the relationship between the circadian rhythms of BP and the PPG and ECG waveforms, it is promising that these signals may be able to assist in detecting deterioration in cardiovascular health in the general population.

Acknowledgements

Four years (+ a little extra) of DPhil-ing has left me with an enormous debt to some truly incredible people. This A4 piece of paper is my attempt at paying it off – or at least making a dent.

Firstly, I would like to thank my supervisors. To Prof. Lionel Tarassenko (CBE, FEng, FMedSci, DPhil ...): I've never met a person with so many titles, and yet they're still too few! Thank you for your guidance over the past four years and for putting up with my incessant use of the word "*like*" when explaining anything technical. And to Prof. Mauricio Villarroel, thank you for your hard work, high expectations, and for always raising the bar.

To the BSP/Camera research group, thank you for creating a collaborative and stimulating environment. Dr. Shaun Davidson, your unrelenting help and guidance has been invaluable through all aspects of this DPhil. I will always be appreciative of the time and energy you generously gave me. Dr. (Dr.) Mirae Harford, thanks for bringing such positive vibes to the group – but seriously, starting your DPhil at the same time as me, taking a year of maternity leave, and still finishing before me? You're a superstar. And to my fellow DPhil students, Dineo, Alexis, Abdullah, and Marshia, all I can say is: *good luck*.

I would like to express my gratitude to the participants of the MOLLIE and nABPM clinical studies, I hope you spent your Amazon vouchers wisely.

I am truly lucky to have spent the last 4 years with some amazing people. Whether you're an Oxpal, a Basoluteoffballer, an MCK Cajoney, a Hobby hob hobsonite, a Live laugh Londer, an AJ33W/AS33W-ite, or one of the rabbits I lived with at the vicarage when I was homeless, thank you for being amazing friends for life and for not escaping from your hutch at night. Special thanks to Andy, Dave, Emmanuel, Hamish, Jorgen, Julius, Kai, Maddy, Navin, Ollie, Rachel, Sam, Sagar, Stanley, Stefan, Val and many more besides. And of course, a huge thank you to the Harcourt Arms.

To my brothers Ciaran, Niall, and Liam, and their (incredibly out of their league) partners, fiancés, and baby-mamas Alice, Sian, and Lucie. You all contributed absolutely nothing towards the content of this thesis, if anything you were a hindrance, but I still love you. To my wonderful God-daughter Aoife, thank you for coming into this world - I am so excited to see the person you grow up to be! If ever you're having problems sleeping, I'll get your dad to read you this thesis.

Finally, thank you from the bottom of my heart to my parents Margaret and John. The latter stages of this DPhil were more difficult than I could have imagined, and I would be nowhere without your unwavering support, (unfounded) belief, and unconditional love. Here's to you.

Related publications

Journal papers

- Eoin Finnegan, Shaun Davidson, Mirae Harford, Peter Watkinson, Lionel Tarassenko, and Mauricio Villarroel. "Features from the photoplethysmogram and the electrocardiogram for estimating changes in blood pressure." *Scientific Reports* 13, no. 1 (2023): 986.
- Eoin Finnegan, Shaun Davidson, Mirae Harford, Joao Jorge, Peter Watkinson, Duncan Young, Lionel Tarassenko, and Mauricio Villarroel. "Pulse arrival time as a surrogate of blood pressure." *Scientific Reports* 11, no. 1 (2021): 22767.
- Shaun Davidson, Mauricio Villarroel, Mirae Harford, Eoin Finnegan, Joao Jorge, Duncan Young, Peter Watkinson, and Lionel Tarassenko. "Day-to-day progression of vital-sign circadian rhythms in the intensive care unit." *Critical Care* 25, no. 1 (2021): 1-13.
- Shaun Davidson, Mauricio Villarroel, Mirae Harford, Eoin Finnegan, Joao Jorge, Duncan Young, Peter Watkinson, and Lionel Tarassenko. "Vital-sign circadian rhythms in patients prior to discharge from an ICU: a retrospective observational analysis of routinely recorded physiological data." *Critical Care* 24, no. 1 (2020): 1-13.

Conference presentations

- Eoin Finnegan, Shaun Davidson, Mirae Harford, Joao Jorge, Mauricio Villarroel, and Lionel Tarassenko. "Classifying nocturnal blood pressure patterns using photoplethysmogram features." In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 3401-3404. IEEE, 2022.
- Eoin Finnegan, Shaun Davidson, Mirae Harford, Joao Jorge, and Mauricio Villarroel. "The presence of a circadian rhythm in pulse arrival time and its application for classifying blood pressure night-time dip." In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 488-491. IEEE, 2021.
- Eoin Finnegan, Mauricio Villarroel, Shaun Davidson, Mirae Harford, Joao Jorge, and Lionel Tarassenko. "Comparing Circadian Rhythms in Blood Pressure Computed From the Arterial Line and Sphygmomanometer in the ICU." In 2020 Computing in Cardiology, pp. 1-4. IEEE, 2020.

Contents

	Page
LIST OF FIGURES	xix
LIST OF TABLES	xxiii
LIST OF ABBREVIATIONS	xxv
1 INTRODUCTION	1
1.1 Cuffless blood pressure measurement using the photoplethysmogram and the electrocardiogram	1
1.2 Clinical motivation	2
1.2.1 Blood pressure	2
1.2.2 Blood pressure measurement	2
1.2.3 Cuffless blood pressure measurement	3
1.3 Objectives	4
1.4 Contributions	4
1.4.1 Contributions from other members of the research group	5
1.5 Thesis structure	5
2 BACKGROUND	9
2.1 Introduction	10
2.2 Elements of the cardiovascular system	10
2.2.1 Structure of the heart	10
2.2.2 Cardiac cycle	12
2.2.3 Structure of blood vessels	14
2.3 Blood pressure	15
2.3.1 Short-term blood pressure regulation	16
2.3.2 Mid-term blood pressure regulation	18
2.3.3 Long-term blood pressure regulation	21
2.4 Measuring blood pressure	23
2.4.1 Arterial line	23
2.4.2 Sphygmomanometer cuff	25
2.4.3 Applanation tonometry	26
2.4.4 Volume clamping	27
2.4.5 Clinical assessment of hypertension	28
2.4.6 Home and ambulatory monitoring of blood pressure	28
2.5 Photoplethysmography	30
2.5.1 Modes of operation	31
2.5.2 Limitations and sources of error	32

2.6	Electrocardiography	33
2.6.1	Limitations and sources of error	35
2.7	Impedance cardiography	36
2.7.1	Limitations and sources of error	36
2.8	Estimating blood pressure using transit times	37
2.8.1	Background theory	37
2.8.2	Review of literature	43
2.8.3	Limitations	48
2.9	Estimating blood pressure from the photoplethysmogram	49
2.9.1	Background theory	49
2.9.2	Review of literature	52
2.9.3	Limitations	56
2.10	Estimating blood pressure from the electrocardiogram	57
2.10.1	Background theory	57
2.10.2	Review of literature	57
2.10.3	Limitations	59
2.11	Commercially-available devices	59
2.12	Conclusion	60
3	CLINICAL DATASETS	63
3.1	Introduction	63
3.2	Dataset requirements	64
3.2.1	Instrumentation requirements	64
3.2.2	Population size requirements	65
3.3	Mapping Of Lower Limb skIn pErfusion	65
3.3.1	Overview	65
3.3.2	Motivation	66
3.3.3	Study design and protocol	66
3.3.4	Research ethics and informed consent	66
3.3.5	Participants and eligibility criteria	67
3.3.6	Data acquisition	68
3.3.7	Cohort description	71
3.4	Medical Information Mart for Intensive Care - III	72
3.4.1	Overview	72
3.4.2	Patient selection criteria	74
3.5	Non-invasive ambulatory blood pressure monitoring study	75
3.5.1	Overview	75
3.5.2	Motivation	75
3.5.3	Study design and protocol	76
3.5.4	Research ethics and informed consent	77
3.5.5	Participants and eligibility criteria	77
3.5.6	Data acquisition	78
3.5.7	Cohort description	79

4	ESSENTIAL METHODOLOGY	83
4.1	Introduction	83
4.2	Segmentation of cardiac beats	83
4.2.1	Segmentation of the electrocardiogram	84
4.2.2	Segmentation of the photoplethysmogram	85
4.3	Signal quality indices	86
4.3.1	Signal quality index for the electrocardiogram	86
4.3.2	Signal quality index for the photoplethysmogram	88
4.3.3	Combined signal quality index for the electrocardiogram and photo- plethysmogram	91
4.4	Non-linear machine learning methods	92
4.4.1	Decision trees	93
4.4.2	Random forest	94
4.4.3	Mixed effects random forest	94
4.5	Shapley additive explanation values	95
5	PULSE ARRIVAL TIME AS A SURROGATE MEASURE OF BLOOD PRESSURE	97
5.1	Introduction	97
5.2	Methods overview	98
5.3	Computing reference blood pressure values	98
5.4	Estimating pulse arrival time	100
5.4.1	Pulse arrival time signal quality index	101
5.4.2	Kalman filter	102
5.5	Estimating pre-ejection period	104
5.5.1	Impedance cardiogram signal quality	104
5.5.2	Estimating beat-by-beat pre-ejection period	106
5.5.3	Pre-ejection period signal quality	107
5.6	Estimating blood pressure using transit times	109
5.6.1	Time alignment	109
5.6.2	Averaging measurements	110
5.6.3	Models for the estimation of blood pressure	111
5.6.4	<i>Baseline reference</i>	114
5.6.5	Model evaluation	114
5.7	Results	115
5.7.1	Overall response	115
5.7.2	Estimating blood pressure using transit times	117
5.8	Discussion	120
5.8.1	The relationship between pulse arrival time and blood pressure	120
5.8.2	Impact of pre-ejection period	122
5.8.3	Estimating blood pressure using transit times	123
5.8.4	Limitations	124
5.9	Conclusion	125

6	FEATURES FROM THE PHOTOPLETHYSMOGRAM AND ELECTROCARDIOGRAM FOR ESTI-	
	MATING CHANGES IN BLOOD PRESSURE	127
6.1	Introduction	127
6.2	Methods overview	128
6.3	Computing features from the photoplethysmogram	128
6.3.1	Fiducial point detection	128
6.3.2	Feature extraction	131
6.4	Computing features from the electrocardiogram	139
6.4.1	Feature extraction	139
6.5	Reference blood pressure values	142
6.6	Estimating changes in blood pressure	142
6.6.1	Feature set: X	143
6.6.2	Calibration	143
6.6.3	<i>Baseline reference</i>	144
6.6.4	Removing collinear features	144
6.6.5	Data augmentation	145
6.6.6	Regression models	145
6.6.7	Shapley additive explanation values feature importance ranking coefficient	146
6.7	Results	147
6.7.1	Removing collinear features	147
6.7.2	Model performance	149
6.7.3	Feature importance	152
6.8	Discussion	154
6.8.1	Observed changes in the photoplethysmogram waveform morphology . .	154
6.8.2	Model performance	155
6.8.3	Feature importance	158
6.8.4	Limitations	159
6.9	Conclusion	160
7	ANALYSIS OF THE CIRCADIAN RHYTHM OF PULSE ARRIVAL TIME AND PHOTOPLETHYS-	
	MOGRAM FEATURES	161
7.1	Introduction	161
7.2	Methods overview	162
7.3	Pulse arrival time estimation in the MIMIC database	163
7.4	Computing ensemble averaged photoplethysmogram features	165
7.4.1	Computing ensemble-averaged photoplethysmogram	165
7.4.2	Fiducial point detection	166
7.4.3	Feature extraction	167
7.5	Cohort selection	169
7.5.1	Blood pressure phenotype labelling	170
7.6	Estimating circadian rhythm profiles	171
7.6.1	Double-cosinor model	172

7.6.2	Goodness-of-fit F-test	173
7.6.3	Population-level double-cosinor model	174
7.6.4	Distributions of double-cosinor parameters	175
7.7	Results	176
7.7.1	Dataset	176
7.7.2	Population-level double-cosinor model	176
7.7.3	Distributions of double-cosinor parameters	178
7.8	Discussion	183
7.8.1	Limitations	185
7.9	Conclusion	187
8	CLASSIFICATION OF BLOOD PRESSURE PHENOTYPES	189
8.1	Introduction	189
8.2	Methods overview	190
8.3	Cohort selection	190
8.3.1	Blood pressure phenotype labelling	191
8.4	Method 1: Regularised <i>individual calibration</i>	191
8.4.1	Feature extraction	192
8.4.2	Model construction	192
8.4.3	Model generalisability	194
8.5	Method 2: Calibration-free	194
8.5.1	Feature extraction	194
8.5.2	Model construction	196
8.5.3	Model generalisability	197
8.6	Baseline reference method	198
8.7	Evaluation and generalisability of methods	198
8.7.1	Evaluation metrics	198
8.7.2	Generalisability of methods	199
8.8	Results	200
8.8.1	Dataset	200
8.8.2	Overall performance	201
8.8.3	Method 1: Regularised <i>individual calibration</i>	202
8.8.4	Method 2: Calibration-free	204
8.9	Discussion	204
8.9.1	Method 1: Regularised <i>individual calibration</i>	207
8.9.2	Method 2: Calibration free	209
8.9.3	Limitations	210
8.10	Conclusion	211
9	TOWARDS CUFFLESS ESTIMATION OF AMBULATORY BLOOD PRESSURE	213
9.1	Introduction	213
9.2	Methods overview	214

9.3	Activity detection	215
9.4	Ensemble averaged pulse arrival time and photoplethysmogram features	216
9.4.1	Ensemble averaged pulse arrival time	217
9.4.2	Ensemble averaged photoplethysmogram features	218
9.5	Establishing a reference blood pressure signal	220
9.5.1	Computing a reference heart rate	220
9.5.2	Heart rate-based removal of poor-quality blood pressure values	221
9.5.3	Double-cosinor model	223
9.6	Estimating Day ₇ blood pressure	224
9.6.1	<i>Individual calibration</i>	225
9.6.2	<i>Hybrid calibration</i>	226
9.6.3	Regression models	227
9.6.4	<i>Baseline references</i>	228
9.7	Evaluation and generalisability of models	228
9.7.1	Shapley additive explanation values feature importance and feature stability	228
9.7.2	Model complexity	229
9.8	Results	229
9.8.1	Data recording in a free-living context	229
9.8.2	Model performance	230
9.8.3	Participant-level results	233
9.8.4	Feature importance	234
9.9	Discussion	236
9.9.1	Model performance	239
9.9.2	Feature importance	242
9.9.3	Limitations	243
9.10	Conclusion	244
10	CONCLUSIONS AND FUTURE WORK	245
10.1	Introduction	245
10.2	Thesis review	245
10.2.1	Clinical datasets	245
10.2.2	Pulse arrival time as a surrogate measure of blood pressure	246
10.2.3	Features from the photoplethysmogram and electrocardiogram for estimating changes in blood pressure	246
10.2.4	Analysis of the circadian rhythm of pulse arrival time and photoplethysmogram features	247
10.2.5	Classification of blood pressure phenotypes	248
10.2.6	Towards cuffless estimation of ambulatory blood pressure	249
10.3	Future work	251
10.3.1	Study design and considerations	251
10.3.2	Predictive modelling	252
10.3.3	Developing a deeper understanding of photoplethysmogram features	253
10.3.4	Cuffless blood pressure device considerations	254
10.4	Concluding remarks	255

APPENDICES

A	COMPARISON BETWEEN AKTHIA AND AMBULATORY BLOOD PRESSURE MONITOR	259
A.1	Introduction	259
A.2	Methods	260
A.2.1	Dataset	260
A.2.2	Data processing	260
A.3	Results	261
A.3.1	Dataset	261
A.3.2	Overview	261
A.3.3	Vital sign estimation	261
A.3.4	DR _{SBP} estimation	262
A.4	Discussion	262
A.5	Conclusions	264
B	NABPM SLEEP QUESTIONNAIRE	265
C	STATISTICAL ANALYSIS	267
C.1	Hypothesis testing	267
C.1.1	Selecting a statistical test	267
C.1.2	Correcting for multiple hypothesis testing	269
C.2	Circular statistics	270
C.2.1	Circular measures of central tendency	270
C.2.2	Multi-sample test for equal circular median directions	271
D	LINEAR MACHINE LEARNING MODELS	273
D.1	Linear regression and classification methods	273
D.1.1	Ordinary least squares regression	273
D.1.2	Linear mixed effects modelling	274
D.1.3	Logistic regression	276
D.1.4	Least absolute shrinkage and selection operator (LASSO)	277
D.1.5	Elastic net regularisation	277
D.2	Evaluation approaches and metrics	277
D.2.1	Cross-validation	277
D.2.2	Regression metrics	278
D.2.3	Classification metrics	278
E	SUPPLEMENTARY RESULTS: PULSE ARRIVAL TIME AS A SURROGATE MEASURE OF BLOOD PRESSURE	281
E.1	Introduction	281
E.2	Detecting the B-point in the impedance cardiogram	281
E.2.1	Introduction	281
E.2.2	Methods	282

E.2.3	Results	284
E.2.4	Discussion and conclusion	284
E.3	Estimating respiratory rate from the nasal pressure airflow signal	285
E.4	Relationship between transit times and blood pressure	286
E.5	Resampling the impedance cardiogram to 1KHz	286
F	SUPPLEMENTARY RESULTS: FEATURES FROM THE PHOTOPLETHYSMOGRAM AND ELEC-	
	TROCARDIOGRAM FOR ESTIMATING CHANGES IN BLOOD PRESSURE	295
F.1	Introduction	295
F.2	Model performance for mean arterial pressure and diastolic blood pressure estimation	295
F.3	Individual results	297
F.4	Individual changes in cardiac output	300
G	SUPPLEMENTARY RESULTS: CLASSIFICATION OF BLOOD PRESSURE PHENOTYPES	301
G.1	Introduction	301
G.2	Methods	301
G.3	Results	302
G.4	Discussion and conclusion	303
H	SUPPLEMENTARY RESULTS: TOWARDS CUFFLESS ESTIMATION OF AMBULATORY BLOOD	
	PRESSURE	305
H.1	Introduction	305
H.2	Reference blood pressure double cosinor model fits	305
H.3	Model performance for mean arterial pressure and diastolic blood pressure estimation	305
	REFERENCES	309

List of Figures

	Page
2.1 Overview of the cardiovascular system and the heart.	11
2.2 Pressure and volume changes during a cardiac cycle.	13
2.3 Structure of blood vessel walls across the circulatory system.	15
2.4 Depiction of the different SBP circadian rhythm phenotypes: 'dipper' $\mathcal{P}_D$, 'non-dipper' $\mathcal{P}_{ND}$, and 'reverse-dipper' $\mathcal{P}_{RD}$	19
2.5 Averaged lifetime BP changes in 4,993 participants of the Framingham Heart Study.	22
2.6 Arterial blood pressure (ABP) waveform.	24
2.7 Example devices using the applanation tonometry and volume clamp methods. .	27
2.8 Origin of the photoplethysmogram (PPG) waveform.	32
2.9 Origin of the electrocardiogram (ECG) waveform.	34
2.10 Annotated impedance cardiogram (ICG) waveform.	36
2.11 Computing pre-ejection period (PEP) and pulse arrival time (PAT).	41
2.12 Interactions of the incident and reflected pressure waves on the morphology of the PPG waveform.	51
2.13 Example of PPG, VPG, and APG waveforms and their characteristic fiducial points.	51
2.14 Illustration of mechano-electric coupling governing the relationship between changes in BP and changes in the ECG.	58
3.1 MOLLIE study experimental timeline.	67
3.2 MOLLIE study equipment setup.	69
3.3 Distributions of the recorded reference vital signs in the MOLLIE dataset.	72
3.4 MIMIC database patient selection criteria.	75
3.5 nABPM study equipment setup.	76
3.6 Monitoring durations and periods of sleep for the 13 participants in the nABPM dataset.	79
3.7 Distributions of the recorded reference vital signs in the nABPM dataset.	81
4.1 Example of SQI_{ECG} computation for two 60-second ECG signals.	89
4.2 Example of SQI_{PPG} computation for two 60-second PPG signals.	91
4.3 Overview of steps for computing the combined SQI for the ECG and PPG waveforms.	92
4.4 Depiction of a random forest model constructed from a large ensemble of 'weak' decision trees.	93
5.1 Schematic overview of the methodology to estimate blood pressure using transit times.	98
5.2 Computing reference BP measurements for one participant using cubic smoothing splines.	100
5.3 Overview of steps for computing the beat-by-beat PAT time series.	103

5.4	Overview of steps for computing the beat-by-beat PEP time series.	108
5.5	ECG phase delay between Stowood and Cardioscreen recording devices.	109
5.6	Overview of PAT and PTT time alignment.	111
5.7	Histograms of τ_{PAT} and τ_{PTT} across the dataset.	112
5.8	Sensitivity analysis of the impact of γ_{SQI} on the proportion of good-quality windows.	112
5.9	Average deviations of BP and TT metrics from calibration across the cohort.	116
5.10	Distributions of PAT, PTT, and PEP computed in the MOLLIE dataset.	116
5.11	Frequency content of the PAT signal for a typical participant in the dataset.	117
5.12	Relationship between SBP and PAT for all participants in the MOLLIE study.	119
5.13	Relationship between SBP and PTT for all participants in the MOLLIE study.	120
5.14	Distributions of $\Delta BP/\Delta TT$ across the dataset.	121
5.15	The hysteresis effect in both PAT and PTT.	121
6.1	Overview of the fiducial points detected from the PPG waveform.	130
6.2	Gaussian decomposition of the PPG waveform.	136
6.3	Visualisation of the three principal components of the PPG, VPG, and APG waveforms.	139
6.4	Schematic of the ΔBP estimation pipeline for each of the proposed models.	144
6.5	Results of removing collinear features in PPG + ECG feature set.	149
6.6	Individual results of ΔSBP estimation using the $RF_{PPG+ECG}$ model across all participants in the study.	151
6.7	Agreement between the reference ΔSBP values and the estimated ΔSBP across the cohort using $RF_{PPG+ECG}$	152
6.8	Median SBP, MAP, and DBP SHAP values feature ranking coefficients for RF and LASSO+OLS both with PPG + ECG feature set.	153
6.9	Relationship between the top 9 ranking, non-demographic, features from the $RF_{PPG+ECG}$ model and ΔSBP	154
7.1	Schematic of the methodology implemented to estimate the circadian rhythm profiles of all signals of interest.	163
7.2	Sensor delays between the ECG and PPG waveforms recorded in the MIMIC dataset.	164
7.3	Overview of the four stages used to compute PPG_{EA} and SQI_{EA}	166
7.4	Fiducial point detection of PPG_{EA} in the MIMIC database.	168
7.5	Visual comparison of PPG features (PPG_{fts}) extracted on a beat-by-beat basis to those extracted from PPG_{EA}	168
7.6	Typical patient profile showing HR, SBP, RR, PAT, and three PPG_{fts} recorded during the final 24-hour period of a patient's stay in the ICU.	170
7.7	Sensitivity analysis of the impact of γ_{SQI} on the number of patients for whom 5 valid measurements of PAT or PPG_{fts} were available with at least one measurement during the day and one at night.	171
7.8	Visual comparison of the population-level double-cosinor circadian deviation models.	181
7.9	Visual comparison of the population-level double-cosinor circadian deviation models for all signals grouped by BP phenotypes.	182

8.1	Schematic of the 'regularised <i>individual calibration</i> ' phenotype classification pipeline.	191
8.2	Schematic of the 'calibration-free' phenotype classification pipeline.	195
8.3	Overview of the seven features extracted from each signal to parameterise its 24-hour profile.	196
8.4	Median BP phenotype classification results using the two methods and four models.	202
8.5	Median, upper, and lower quartile phenotype classification <i>recall</i> confusion matrices.	203
8.6	Regularised <i>individual calibration</i> , regression analysis for estimating SBP_{Night} and subsequent DR_{SBP} using Method 1.	205
8.7	Visualising the fixed and random effects of the $LMEM_{PPG+PAT}$ model.	206
8.8	Median $LASSO+LR_{PPG+PAT}$ and $RF_{PPG+PAT}$ calibration free phenotype classification ranking coefficients for all signals implemented.	206
9.1	Schematic of the methodology implemented to estimate ambulatory BP.	214
9.2	Example of detecting periods of activity in the dataset.	216
9.3	Determining γ_{acc} as an activity threshold.	217
9.4	Overview of the three stages used to compute PAT_{EA} and SQI_{PAT}^{EA}	218
9.5	Examples of two PPG_{R-R} where the number of previously detected PPG onsets was not one.	219
9.6	Visual comparison of pulse arrival time (PAT) computed on a beat-by-beat basis to PAT_{EA}	219
9.7	Agreement between HR_{ECG} and HR_{PPG} computed across all Day ₀ recordings. . .	221
9.8	Sensitivity analysis of w_{HR} and γ_{SQI}^{HR} to determine their optimum values.	222
9.9	HR-based detection of poor-quality BP values for a typical participant in the nABPM study.	223
9.10	Double-cosinor model smoothing for Day ₀ ABPM data for all participants to compute the reference BP time series.	224
9.11	24-hour availability of data performed as a sensitivity analysis of γ_{SQI}	226
9.12	Participant-wise median validation performance curves as PPG+PAT features were sequentially added to each model.	231
9.13	Comparison of reference and estimated Day ₇ SBP values for each participant. . .	233
9.14	Model RMSE as a function of the range (max-min) of BP values in Day ₀ and Day ₇ .	234
9.15	Top 15 median (across participants) SHAP (ϕ_j) values feature ranking coefficient for <i>Individual calibration</i> RF_{PPG} and <i>Hybrid calibration</i> $ELAS+OLS_{PAT+PPG}$	235
9.16	SHAP values feature rankings stability from the Day ₀ training set to the 'optimum' features determined when training equivalent models on Day ₇ data.	236
9.17	Participant ID: 1010, <i>individual calibration</i> , RF_{PPG} model. Analysis of ϕ_j in both training and test sets.	237
9.18	Participant ID: 1007, <i>individual calibration</i> , RF_{PPG} model. Analysis of ϕ_j in both training and test sets.	237
9.19	Participant ID: 1003, <i>hybrid calibration</i> , $ELAS+OLS_{PPG+PAT}$ model. Analysis of ϕ_j in both training and test sets.	238

9.20	Participant ID: 1002, <i>hybrid calibration</i> , ELAS+OLS _{PPG+PAT} model. Analysis of ϕ_j in both training and test sets.	238
A.1	Aktiia bracelet and ABPM device.	260
A.2	24-hour profile of vital signs recorded by the ABPM and Aktiia device.	262
A.3	Results of estimating SBP using the Aktiia bracelet.	262
A.4	Results of estimating DBP using the Aktiia bracelet.	263
A.5	Results of estimating HR using the Aktiia bracelet.	263
A.6	DR _{SBP} estimated values using the Aktiia bracelet and the reference ABPM.	264
B.1	nABPM sleep questionnaire.	265
C.1	Example of circular data α existing on the unit circle.	270
D.1	Example binary classification confusion matrix.	279
E.1	The PEP time series computed by the three methods across three participants.	285
E.2	The nasal pressure waveform.	286
E.3	Relationship between MAP and PAT for all participants in the MOLLIE study.	289
E.4	Relationship between MAP and PTT for all participants in the MOLLIE study.	290
E.5	Relationship between DBP and PAT for all participants in the MOLLIE study.	291
E.6	Relationship between DBP and PTT for all participants in the MOLLIE study.	292
E.7	PEP estimates from the original 200 Hz ICG signal compared to PEP estimates from the 1 kHz upsampled ICG signal.	293
E.8	Comparison of PEP estimates from the two ICG signals (original 200 Hz ICG and upsampled 1 kHz ICG).	293
F.1	Agreement between the reference Δ MAP values and the estimated Δ MAP across the cohort using RF _{PPG+ECG}	296
F.2	Agreement between the reference Δ DBP values and the estimated Δ DBP across the cohort using RF _{PPG+ECG}	296
F.3	Individual results of Δ MAP estimation using the RF _{PPG+ECG} model across all participants in the study.	298
F.4	Individual results of Δ DBP estimation using the RF _{PPG+ECG} model across all participants in the study.	299
F.5	Changes in CO observed in all participants.	300
G.1	Comparison of different linear models produced by the three approaches for a subset of 11 patients in the overall cohort.	303
G.2	Distribution of PAT-BP calibration slopes (Δ SBP/ Δ PAT) modelled by the three approaches across all patients in the cohort.	304
G.3	RMSE nocturnal SBP estimation as a function of the number of observations a patient had between 8am and midnight.	304
H.1	Double-cosinor model smoothing for all participant's Day ₇ ABPM data to compute the reference BP time series.	306

List of Tables

	Page
2.1 Sympathetic and parasympathetic mechanisms controlled by the ANS.	17
2.2 BP circadian rhythm phenotypes classification used in this thesis based on an individual's SBP day-to-night dip ratio (DR_{SBP}).	20
2.3 Summary of main BP measuring techniques and their associated advantages and disadvantages.	23
2.4 Standard clinical thresholds for assigning an individual into a BP category based on their SBP or DBP values recorded in the clinic.	28
2.5 Summary of key studies of BP estimation using PTT or PAT.	47
2.6 Summary of key studies of BP estimation using the PPG.	55
2.7 Overview of the three main available devices that make commercial claims of cuffless blood pressure monitoring.	61
3.1 Physiological parameters and sampling rates recorded by the three devices used in the MOLLIE study.	69
3.2 Demographics of the participants in the MOLLIE study.	71
3.3 Physiological parameters and sampling rates recorded by the two devices used in the nABPM study.	78
3.4 Demographics of the participants in the nABPM study.	79
5.1 Performance statistics of SBP estimation using the models proposed.	118
6.1 Criteria for identifying fiducial points on PPG waveform.	131
6.2 Summary of PPG_{fts} extracted, separated into PPG morphology, VPG morphology, APG morphology, Gaussian decomposition, and PCA.	131
6.3 First set of initial conditions for Gaussian decomposition optimisation.	136
6.4 Summary of ECG_{fts} extracted.	140
6.5 Resulting PPG_{fts} and ECG_{fts} included in the feature set post removing collinearity.	148
6.6 Performance statistics of ΔSBP estimation using the models proposed.	149
6.7 Spearman's correlation coefficient agreement between the feature ranking coefficients for pairs of ΔSBP , ΔMAP , and ΔDBP	153
7.1 Summary of the 33 PPG_{fts} extracted from each ensemble averaged PPG beat.	169
7.2 BP phenotypes classification based on an individual's SBP day-to-night dip ratio (DR_{SBP}).	171
7.3 Demographics of the subset of patients selected from the MIMIC database.	177
7.4 Results of double-cosinor analysis for all signals.	179
8.1 Demographics of the subset of patients selected from the MIMIC database.	201
8.2 Regularised <i>individual calibration</i> SBP regression performance statistics of SBP_{Night} and subsequent DR_{SBP} estimation using the models proposed in method 2.	204

9.1	Recording durations and data availability for the nABPM study.	230
9.2	Performance statistics of Day ₇ SBP estimation using the models proposed.	232
A.1	Demographics of the participants in the study.	261
A.2	Performance statistics of vital sign (SBP, DBP, and HR) estimation using the Aktiia device compared to an AND 2441 ABPM device.	261
E.1	Comparison of the three B-point detection algorithms.	284
E.2	Performance statistics of MAP estimation using the models proposed.	287
E.3	Performance statistics of DBP estimation using the models proposed.	288
F.1	Performance statistics of Δ MAP estimation using the models proposed.	295
F.2	Performance statistics of Δ DBP estimation using the models proposed.	297
G.1	Performance statistics of nocturnal SBP estimation using PAT and the three approaches.	303
H.1	Performance statistics of Day ₇ MAP estimation using the models proposed.	307
H.2	Performance statistics of Day ₇ DBP estimation using the models proposed.	307

List of Abbreviations

ABPM	Ambulatory blood pressure monitoring.
Acc	Classification accuracy.
ANS	Autonomic nervous system.
APG	Acceleration plethysmogram, second derivative of the PPG waveform.
BMI	Body mass index.
BP	Blood pressure.
CO	Cardiac output.
CV	Cross-validation.
CVD	Cardiovascular disease.
DBP	Diastolic blood pressure.
DR	Day-to-night dip ratio.
ECG	Electrocardiogram, measures the electrical activity of the heart.
ECG_{fts}	Morphological features derived from the ECG.
FFT	Fast Fourier transform.
HBPM	Home blood pressure monitoring.
HR	Heart rate.
HRV	Heart rate variability.
ICG	Impedance cardiogram.
ICU	Intensive care unit.
IQR	Inter-quartile range.
κ	Cohen's kappa.
LMEM	Linear mixed-effects model.
MAE	Mean absolute error.
MAP	Mean arterial blood pressure.
MEC	Mechano-electric coupling.
MERF	Mixed-effects random forest.
MIMIC	Medical Information Mart for Intensive Care.
ML	Machine learning.
MOLLIE	Mapping Of Lower Limb skIn pErfusion healthy volunteer study.
nABPM	Non-invasive ambulatory blood pressure monitoring healthy volunteer study.
$\mathcal{P}_D$	'dipper' blood pressure phenotype.

$\mathcal{P}_{ND}$	'non-dipper' blood pressure phenotype.
$\mathcal{P}_{RD}$	'reverse dipper' blood pressure phenotype.
PAT	Pulse arrival time.
PEP	Pre-ejection period.
ρ_p	Pearson's correlation coefficient.
PPG	Photoplethysmogram, monitors changing blood volume in an artery.
PPG_{fts}	Morphological features derived from the PPG.
ρ_s	Spearman's correlation coefficient.
PTT	Pulse transit time.
PWV	Pulse wave velocity.
RF	Random forest.
RMSE	Root mean squared error.
RR	Respiratory rate.
SBP	Systolic blood pressure.
SHAP	Shapley additive explanation values.
SMR	<i>Severe-misclassification rate</i> , the proportion of patients in the 'dipper' ($\mathcal{P}_D$) cohort that were incorrectly classified as 'reverse dipper' ($\mathcal{P}_{RD}$), and vice versa.
SpO_2	Peripheral oxygen saturation as computed by a pulse oximeter.
SQI	Signal quality index.
SV	Stroke volume.
TPR	Total peripheral resistance.
TT	Transit time, a collective term for pulse transit time and pulse arrival time.
VPG	Velocity plethysmogram, first derivative of the PPG waveform.
wF1	Weighted F1 score.

1.1 Cuffless blood pressure measurement using the photoplethysmogram and the electrocardiogram

Blood pressure (BP), as one of the five essential physiological vital signs, serves as a crucial indicator of cardiovascular health. Excessively elevated BP (hypertension) is the leading cause of non-communicable disease and mortality worldwide [1]. Similar to other vital signs, such as heart rate and respiratory rate, BP exhibits endogenous oscillations with a period of approximately 24 hours, a phenomenon known as circadian rhythmicity [2, 3, 4]. Analysis of this rhythm independently predicts all-cause mortality and cardiovascular events, including coronary heart disease, atrial fibrillation, and stroke [5, 6, 7].

Currently, the established technique for monitoring 24-hour variations of BP is known as ambulatory blood pressure monitoring (ABPM). However, as this technique requires an inflatable cuff to be wrapped around the arm, it is highly disruptive to sleep and daily life and is therefore rarely performed in primary care settings [5, 8].

Advances in computational capabilities have enabled research into the estimation of BP using other physiological signals [9]. For instance, variations in the cardiovascular system and autonomic nervous system (ANS) induces changes in the photoplethysmogram (PPG) and electrocardiogram (ECG), which are ubiquitous in a clinical setting and are increasingly available in an out-of-clinic setting where they can be unobtrusively recorded using wearables such as smartwatches. These signals, therefore, offer great promise for cuffless BP monitoring.

The ability to continuously monitor BP without the need for an inflatable cuff would offer substantial benefits over, or in addition to, traditional BP measurements. In primary care settings, more frequent monitoring of BP could raise awareness and facilitate earlier detection of hypertension. Additionally, by eliminating the requirement for cuff inflation, cuffless devices have the potential to monitor BP during sleep, enabling the assessment of BP circadian rhythm patterns in the general population. In pursuit of these benefits, this thesis investigates, in-depth, cuffless ambulatory blood pressure measurement using the PPG and ECG signals.

1.2 Clinical motivation

1.2.1 Blood pressure

As the heart pumps blood around the body, it transports oxygen and vital nutrients to cells and tissues. BP reflects the force the heart generates to maintain this essential blood flow [10]. The ANS closely regulates BP to remain within 'normal' limits in response to varying physiological demands. If BP drops too low (hypotension), blood flow to vital organs may become inadequate, potentially leading to cellular and organ dysfunction, and even shock due to reduced tissue perfusion [11]. Conversely, if BP rises too high (hypertension), blood vessels can be damaged, increasing the risk of clots and ruptures. Sustained hypertension has been shown to be one of the highest risk factors for the majority of cardiovascular diseases (CVDs) [12]. The global prevalence of sustained hypertension is currently estimated to be 31.1% with sharp increases in low-income countries where access to anti-hypertensive medication is limited [13].

BP additionally exhibits an endogenous rhythm, regulated by neurohormonal variations. This rhythm typically includes a 10%-20% decrease [14] during sleep, a phenomenon known as 'dipping' [5]. However, different BP circadian rhythm phenotypes exist depending on the magnitude and direction of the nocturnal change in BP. For example, there exists a cohort of individuals (thought to be approximately 10.8% of the general population [5]) for whom their highest BP occurs during the night, a pattern known as 'reverse-dipping'. 'Reverse-dipping' patterns are predictors of both fatal and nonfatal cardiovascular events that are independent of daytime BP values [15, 5].

Poor regulation of BP is often referred to as the 'silent killer' as it is largely asymptomatic [16]; however, it is highly treatable [17]. Accurate measurement of BP is, therefore, of significant importance.

1.2.2 Blood pressure measurement

Continuous measurement of BP can be obtained using an arterial line, a small thin catheter inserted into a large artery (typically the radial artery). Cannulation of an artery, however, is invasive, time-consuming, and has associated risks (such as infection [18]). As a result, inserting an arterial line requires an appropriately skilled practitioner with the procedure primarily performed in clinic or hospital settings for individuals requiring intensive care [19, 20]. As such, it is not a feasible option for widespread, continuous BP monitoring in primary care.

The current "gold standard" in non-invasive BP measurement uses a sphygmomanometer, an inflatable cuff wrapped around an individual's upper arm and inflated until the blood flow is temporarily occluded [20]. This provides a single reading of systolic blood pressure (highest blood

pressure when the heart contracts) and diastolic blood pressure (lowest blood pressure when the heart relaxes). To diagnose hypotension or hypertension, two BP measurements are typically taken by a trained clinician within a 1-2 minute interval, with the individual in a sitting position.

These clinical spot measurements, however, do not take into account the circadian nature of BP. For this assessment, ABPM is typically performed over a period of 24 or 48 hours with automatic cuff inflations every thirty minutes during the day and every hour during the night [21]. Studies have indicated that the cuff inflation process can result in disturbances to sleep, reduce daytime activity, and may even induce 'reverse-dipping' [22]. As a result, tolerance for ABPM in the general population is very low [23] and it is rarely performed in primary care settings [5, 8].

There is a need, therefore, for the development of an unobtrusive, cuffless ambulatory BP monitoring device that can accurately assess an individual's 24-hour BP profile without inducing sleep arousal or significant user discomfort.

1.2.3 Cuffless blood pressure measurement

Within the last five years, a number of companies have launched cuffless BP measurement devices [24, 25]. However, the accuracy of these devices has been debated and called into question in several independent studies [26, 27]. As accurate measurements are crucial for addressing the global burden of CVD, none of these devices have been accepted by the clinical community [25]. Therefore, additional research is required to investigate cuffless BP measurement options.

In this thesis, we investigate the utility of the PPG and ECG signals for deriving surrogate measures of BP. The PPG signal measures changes in the light absorption characteristics of blood [28] and is typically recorded by a pulse oximeter attached to the ear, finger, or toe. The ECG signal measures the electrical activity of the heart during a cardiac cycle through electrodes attached to the skin surface of the chest [29], or arms [30].

Multiple, cardiovascular-related, metrics may be derived from analysis of the PPG or ECG waveforms. One such example is the pulse arrival time (PAT) which indicates the velocity of the pulse wave produced as the heart pumps blood around the body [31]. More recent research has increasingly focused on relating morphological features of the PPG and ECG waveforms to BP using data-driven modelling [9]. However, the use of these metrics for continuous monitoring of ambulatory BP has not yet been fully established.

1.3 Objectives

The main objective of this thesis is **"To investigate PPG- and ECG-derived surrogate measures of ambulatory BP"**. The more specific goals are:

- To design clinical studies and collect novel datasets for the development of ambulatory BP estimation algorithms in healthy individuals.
- To identify relevant surrogate BP measures derived from the PPG and ECG waveforms and examine their relationship with BP.
- To develop robust methods for analysing PPG and ECG derived BP surrogates in an ambulatory setting, handling periods of poor signal quality caused by movement.
- To study 24-hour variations of these surrogate measures and develop cuffless solutions to estimate an individual's BP phenotype, as an alternative to ABPM.
- To compare different calibration strategies on the basis of BP estimation accuracy. Additionally, to investigate the stability of these strategies, 7 days after initial calibration.
- To compare BP estimation using only surrogate measures derived from the PPG waveform (as a single-source estimator), to then including the ECG waveform.

1.4 Contributions

Following the objectives outlined above, there are three major contributions of this thesis:

Surrogate blood pressure measures derived from the PPG and ECG waveforms The first major contribution of this work is a comprehensive assessment of surrogate BP measures derived from the PPG and ECG waveforms and their relationship with BP. We introduce robust signal-processing techniques to compute estimates of PAT as well as morphological features of the PPG and ECG waveforms. Deriving reliable features from continuous waveforms in an ambulatory environment is a non-trivial challenge. We, therefore further developed our methodology to compute these surrogate measures when in the presence of motion artefacts. The relationships between these surrogate measures and BP is assessed using data from multiple individuals in three distinct settings with BP perturbed by a variety of mechanisms. Our research provides a thorough evaluation of the significance of each surrogate measure in estimating BP within each specific setting.

Circadian rhythm analysis The second major contribution is a circadian rhythm analysis of PPG and ECG-derived surrogate BP measures in a large heterogeneous patient cohort. To our knowledge, no previous research has investigated whether any surrogate measure for BP exhibits an endogenous 24-hour rhythm in synchrony with reference BP measurements. This analysis additionally facilitated the development of a novel calibration-free ordinal BP phenotype classification model that may offer a cost-effective and unobtrusive means of assessing BP phenotypes in the general population.

Ambulatory blood pressure estimation seven days after initial calibration The third major contribution is the application of the algorithms developed throughout this thesis to the ambulatory domain. Due to the difficulties of recording data in an at-home environment, very few publications have explored cuffless ambulatory BP estimation. We conducted a unique clinical study and analysed appropriate calibration procedures for accurate 24-hour BP estimation, seven days after initial calibration.

1.4.1 Contributions from other members of the research group

This thesis incorporates collaborative contributions from members of our research group, specific occurrences of this are outlined below. While the thesis is presented in the first person plural, the author conducted all other work independently.

Design of the Mapping Of Lower Limb skin pErfusion The clinical study titled 'Mapping of Lower Limb Skin Perfusion' (introduced in section 3.3), was originally conceived by Dr. Mirae Harford as part of her own PhD research. Dr. Harford, who is medically qualified, oversaw the study's organisation and ensured compliance with medical research ethics as mandated by University regulations [32]. My involvement in the study included active participation in the planning process through regular meetings, as well as data collection.

Patient selection criteria of the Medical Information Mart for Intensive Care dataset The patient selection criteria of the Medical Information Mart for Intensive Care dataset (introduced in section 3.4) were adapted from previously published work by Dr. Shaun Davidson [2].

1.5 Thesis structure

This thesis is composed of 10 chapters. Below is an overview of the thesis structure:

Chapter 2 provides the relevant background regarding BP as a vital sign, circadian rhythms, and the PPG and ECG signals. This chapter explores the limitations of current in-clinic and at-home BP measurements. Existing methodologies, and their associated limitations, for cuffless estimation of BP are also discussed.

Chapter 3 introduces the three datasets employed in this thesis, with a description of the study populations, data collection processes, and an overview of the reference data.

Chapter 4 provides a technical background of the general methodologies implemented in this thesis. Chapter-specific methodology is introduced in each relevant chapter.

Chapter 5 investigates PAT as a surrogate measure of BP in a study involving thirty healthy volunteers, in which BP changes were induced by pharmacological intervention. In particular, we analyse the impact of the pre-ejection period (PEP) as an additional component of PAT. We establish a robust signal analysis pipeline for computing PAT and PEP estimates while handling periods of low signal quality, a step that has been largely overlooked in the literature. Finally, we highlight that population-based models are insufficient to model the PAT-BP relationship.

Chapter 6 extends the work of the previous chapter to investigate strategies for BP estimation using morphological features of the PPG and ECG waveforms. We extract a large and diverse set of features established from the relevant literature and develop linear (ordinary least squares) and non-linear (random forest) regression models to estimate changes in BP. Through a robust assessment of feature importance, we discuss the mechanisms behind the PPG-BP relationship and suggest that the ECG on its own is insufficient for a reliable estimation of changes in BP.

Chapter 7 provides a circadian rhythm analysis of vital signs (BP, heart rate, and respiratory rate), PAT, and the previously established PPG waveform features. This analysis is conducted on a large, heterogeneous, cohort of patient data available in an open-source intensive care database. Through statistical modelling, we establish that PAT and a subset of PPG waveform features exhibit a circadian rhythm in synchronicity with that of BP. We additionally demonstrate statistically significant differences in these circadian rhythm profiles between patients grouped according to BP circadian rhythm phenotypes.

Chapter 8 utilises the results of the previous chapters and introduces two new methods for classifying an individual's BP phenotype without a reliance on ABPM. Our first approach makes use of BP measurements taken while an individual is awake for a regularised individual calibration. Our second approach eliminates the need for BP calibration measurements altogether and utilises circadian features of PAT and PPG waveform features for calibration-free phenotype classification.

Chapter 9 further develops the methodology implemented throughout this thesis towards cuffless estimation of ambulatory BP, 7 days after initial calibration. We compare the performance of leave-one-subject-out regression, a technique commonly used by commercial cuffless devices, to personalised regression, which involves training individual regression models using only data from that specific individual. Our results suggest that leave-one-subject-out regression models may lead to clinically significant errors in BP measurements, highlighting the importance of individualised calibration.

Chapter 10 summarises the research carried out in this thesis, leading to a discussion of proposals for the future work needed to further advance the field of cuffless BP monitoring.

2

Background

Contents

2.1	Introduction	10
2.2	Elements of the cardiovascular system	10
2.2.1	Structure of the heart	10
2.2.2	Cardiac cycle	12
2.2.3	Structure of blood vessels	14
2.3	Blood pressure	15
2.3.1	Short-term blood pressure regulation	16
2.3.2	Mid-term blood pressure regulation	18
2.3.3	Long-term blood pressure regulation	21
2.4	Measuring blood pressure	23
2.4.1	Arterial line	23
2.4.2	Sphygmomanometer cuff	25
2.4.3	Applanation tonometry	26
2.4.4	Volume clamping	27
2.4.5	Clinical assessment of hypertension	28
2.4.6	Home and ambulatory monitoring of blood pressure	28
2.5	Photoplethysmography	30
2.5.1	Modes of operation	31
2.5.2	Limitations and sources of error	32
2.6	Electrocardiography	33
2.6.1	Limitations and sources of error	35
2.7	Impedance cardiography	36
2.7.1	Limitations and sources of error	36
2.8	Estimating blood pressure using transit times	37
2.8.1	Background theory	37
2.8.2	Review of literature	43
2.8.3	Limitations	48
2.9	Estimating blood pressure from the photoplethysmogram	49
2.9.1	Background theory	49
2.9.2	Review of literature	52
2.9.3	Limitations	56
2.10	Estimating blood pressure from the electrocardiogram	57
2.10.1	Background theory	57
2.10.2	Review of literature	57
2.10.3	Limitations	59
2.11	Commercially-available devices	59
2.12	Conclusion	60

2.1 Introduction

In this chapter, we provide an overview of the relevant background to cuffless blood pressure measurement. We first outline key elements of the cardiovascular system and discuss the need for accurate blood pressure monitoring. We detail why this thesis has a specific focus on 24-hour, circadian, changes in blood pressure. We then review the currently established techniques to measure blood pressure and outline their inadequacies for circadian rhythm analysis. We introduce key signals that are used throughout this thesis: the electrocardiogram and photoplethysmogram, as well as the impedance cardiogram. Finally, we review the state-of-the-art approaches and recent developments in the field of cuffless blood pressure measurement.

2.2 Elements of the cardiovascular system

The cardiovascular system is part of the circulatory system that pumps blood throughout the body. It consists of the heart, blood vessels, and blood (see figure 2.1(a)). Its primary function is to transport oxygen-rich blood and nutrients to all parts of the body and to carry deoxygenated blood back to the lungs. Secondary functions of the cardiovascular system include: removing carbon dioxide and metabolic waste from tissues and depositing them in the excretory organs (skin, lungs, and kidneys); distributing water, electrolytes, and hormones through the body; and regulating body temperature.

2.2.1 Structure of the heart

The heart pumps blood around the body in order to deliver oxygen-rich blood to the body's tissues and organs. Figure 2.1(b) depicts the heart's structure and its four compartments. Deoxygenated blood flows into the right atrium via the superior venae cavae (from the head) and inferior venae cavae (from the body). From the right atrium, blood is pumped into the right ventricle through the tricuspid valve. From the right ventricle, a small force is exerted on the blood to pump blood towards the lungs via the pulmonary valve and the pulmonary artery. Newly oxygenated blood is returned from the lungs to the heart's left atrium via the pulmonary veins. Blood is then pumped into the left ventricle via the mitral valve. Finally, a larger force is exerted by the left ventricle to pump blood to the rest of the body tissues via the aortic valve and the aorta. The pumping mechanism of the four chambers of the heart is highly coordinated and synchronised so that their actions are effective. On the exit of each chamber, blood passes through either an atrioventricular (between atrium and ventricles) or semilunar (between ventricles and arteries) valve. Each valve has flaps that open and close once per heartbeat to control the timing of blood flow and only allow blood to flow in one direction.

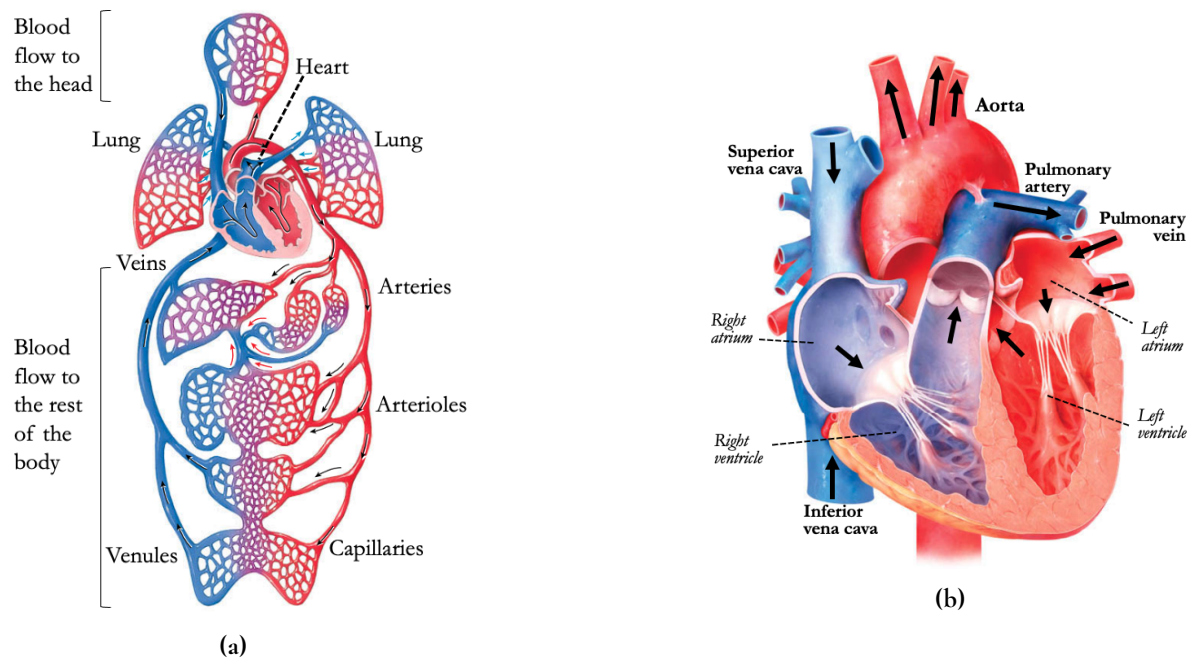


Figure 2.1: Overview of the cardiovascular system and the heart. Black arrows indicate the direction of blood flow. (a) Schematic representation of the cardiovascular system. The heart receives oxygenated blood (shown in red) from the lungs and distributes it through the arteries. Oxygenated blood travels through a series of capillaries to capillaries which are then supplied with oxygen and other nutrients. Deoxygenated blood (shown in blue) leaves the tissue cells and flows through the veins back to the heart and lungs. (b) Structure of the heart. The heart is divided into four chambers. The right atrium of the heart takes in deoxygenated blood and pumps it to the lungs through the right ventricle. The left atrium of the heart takes in oxygenated blood from the lungs and pumps it to the rest of the body through the left ventricle. Both figures were adapted from [33] with permission.

To summarise, the right side of the heart pumps oxygen-poor blood to the lungs, and the left side pumps oxygen-rich blood towards the body tissues. As the atria contract (atrial systole), the ventricles relax (ventricular diastole). Likewise, as the ventricles contract (ventricular systole), the atria relax (atrial diastole). This complete series of events make up a heartbeat, also called a cardiac cycle.

The wall of the heart is comprised of three layers: the outer pericardium, middle myocardium, and inner endocardium. The pericardium surrounds the outside of the heart and consists of connective tissue that protects the heart from infection. The myocardium is the thickest layer of the heart and consists of muscles responsible for the pumping force generated during a cardiac cycle. The right ventricle's muscular wall is thinner than that of the left ventricle, as it only pumps blood to the lungs with a comparatively lower resistance to blood flow. The endocardium is the inner layer and is made up of epithelium and connective tissue with many elastic and collagenous fibres. It also contains blood vessels and specialised cardiac muscle fibres known as Purkinje fibres that distribute electrical impulses to ventricular muscles.

2.2.2 Cardiac cycle

The cardiac cycle is the sequence of events that occurs during a heartbeat. A single cycle of cardiac activity can be divided into two phases: diastole and systole. Diastole, known as the filling phase, represents the period of time when the ventricles are relaxed. Throughout most of this time, blood is flowing passively from the left atrium and right atrium into the left ventricle and right ventricle respectively. Systole, known as the contraction phase, represents the time during which the left and right ventricles contract and eject blood into the aorta and pulmonary artery respectively.

Figure 2.2(a) shows the changes in pressure in key chambers of the heart during a single cardiac cycle, specifically in aortic pressure (AP), left ventricular pressure (LVP), and left atrial pressure (LAP). The onset of systole results in a sharp increase in LVP. Once LVP is greater than AP, blood is ejected from the left ventricle into the aorta. The pressure at which this occurs is known as afterload, the pressure required for blood to flow. Systole ends when LVP is less than AP and diastole begins. Figure 2.2(a) also shows changes in left ventricular volume (LV Vol) during a single cardiac cycle, clearly showing blood ejection during systole and blood filling during diastole.

Figure 2.2(b) shows the pressure-volume (P-V) relationship for the left ventricle during a single cardiac cycle. The shape of the loop is affected by the contractility and compliance of the left ventricle, and factors that alter refilling or ejection of blood. The cycle can be divided into four main phases, labelled a-d. Phase "a" consists of ventricular filling. When LVP falls below LAP during diastole, the mitral valve opens and the left ventricle begins to fill with oxygen-rich blood from the left atrium. Initially, LVP continues to fall as the ventricle fills because it is still relaxing (from phase "d" of the previous cardiac cycle). However, once the ventricle is fully relaxed, LVP gradually increases as the LV Vol increases. Left ventricular volume after filling, known as left ventricular end-diastolic volume (LVEDV), is typically around 120-140 mL [10]. The equivalent pressure, end-diastolic pressure (EDP), is typically < 10 mmHg [10].

Phase "b" consists of isovolumetric contraction. It is triggered by the mitral valve closing, resulting in a swift increase in LVP but no changes in volume. Phase "c" is the start of systole and consists of blood ejection. Once LVP exceeds AP, the aortic valve opens and rapid ejection of blood begins. During this phase, LV Vol decreases as LVP increases to a peak value (peak systolic pressure) and subsequently decreases as the ventricles begin to relax. LVP towards the end of the ejection phase is slightly less than AP in the aorta (see figure 2.2(a)), but blood continues to flow out due to momentum. Eventually, the flow briefly reverses, causing the closure of the aortic valve and a small increase in AP, giving rise to a characteristic feature in the pressure waveform known as the dicrotic notch. Left ventricular volume after ejection, typically around 50-70 mL, is known

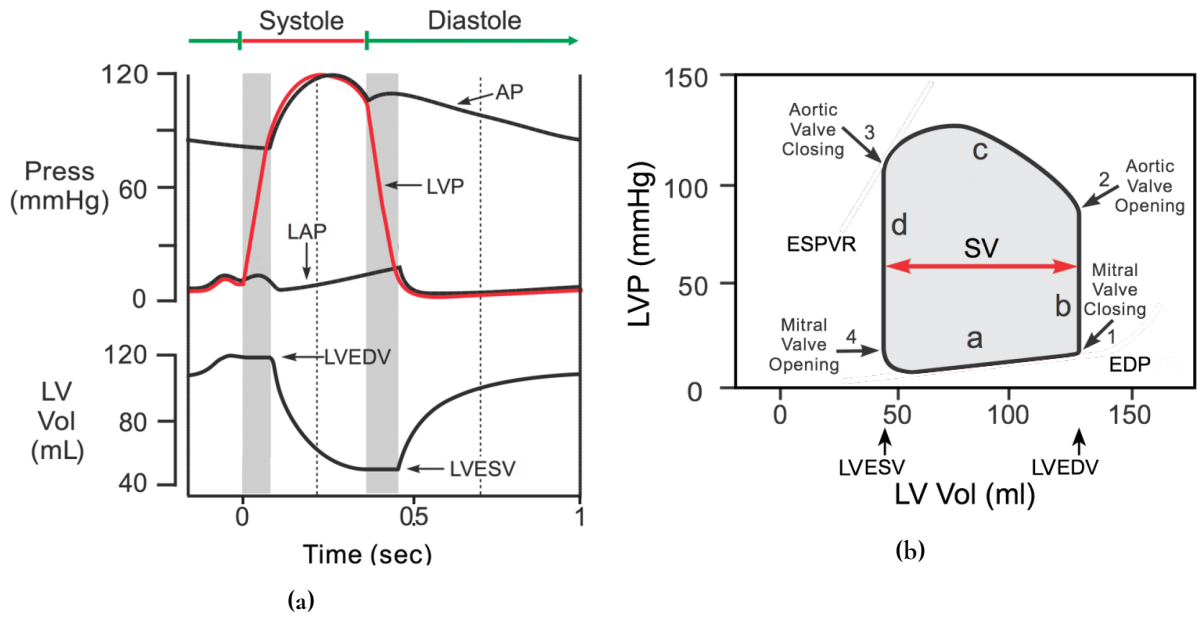


Figure 2.2: Pressure and volume changes during a cardiac cycle. (a) Changes in left ventricular pressure (LVP) and left ventricular volume (LV Vol) during cardiac contraction (systole) and relaxation (diastole). (b) Pressure-Volume relationship for the left ventricle. Acronyms: AP - aortic pressure; LVP - left ventricular pressure; LAP - left atrial pressure; LV Vol - left ventricular volume; LVESV - left ventricle end-systolic volume; LVEDV - left ventricle end-diastolic volume; EDP - end-diastolic pressure; ESPVR - end-systolic pressure-volume relationship; SV - stroke volume. Both figures were adapted from [10] with permission.

as left ventricular end-systolic volume (LVESV) [10]. Phase "d" marks the start of diastole and consists of isovolumetric relaxation, triggered by the closing of the aortic valve.

The width of the P-V loop represents the difference between LVEDV and LVESV. This is known as the stroke volume (SV) and equates to the volume of blood pumped out of the left ventricle during each cardiac cycle. Typically, SV is around 70mL in healthy adults [34]. The proportion of LVEDV that is ejected ($SV/LVEDV$) is known as the ejection fraction. The area within the P-V loop is the ventricular stroke work and is a measure of the work done during a heartbeat, an indicator of cardiac function. Cardiac output (CO) is the volume of blood ejected per minute and is equal to the product of heart rate (HR) (the rate at which the heart beats) and SV.

2.2.2.1 Frank-Starling law

The Frank-Starling law of the heart describes the relationship between SV and LVEDV (also known as preload, initial stretching of cardiac muscles). The law states that, when all other factors remain constant, an increase in LVEDV leads to a corresponding increase in SV due to stronger ventricular contractions caused by an increased initial length of the myocardial fibres. The most important consequence of the Frank-Starling law is that output is matched between the right and left

ventricles. Other factors that increase the force and rate of myocardial contraction include increased sympathetic nerve activity and circulating hormones such as adrenaline and noradrenaline.

2.2.3 Structure of blood vessels

The circulatory system (see figure 2.1(a)) is a closed loop consisting of two major circuits: the pulmonary circuit, which transports blood to and from the lungs; and the systemic circuit, which transports blood to and from the remaining tissues in the body. The systemic circuit is much longer than the pulmonary circuit and therefore requires more energy to complete. Within the circulatory system, there are five general classes of blood vessels: arteries, arterioles, capillaries, venules, and veins.

Oxygenated blood flows out of the aorta at high pressure and high velocity into the major arteries, each of which supplies oxygen to a major organ or body region. These arteries subdivide into arterioles (arteries with a diameter $< 100\mu\text{m}$). The arterioles lead to the smallest vessels, the capillaries, which form a dense network of blood supply within all body tissues. In the capillaries, the pressure and velocity of blood decrease as the total surface area of blood vessels increases. Cells receive oxygen and nutrient supply through the systemic capillaries in exchange for carbon dioxide and other waste products [35]. Deoxygenated blood is then transported from capillaries to the veins via microscopic vessels known as venules. Blood is returned to the heart using veins via the superior and inferior venae cavae.

The structure of each blood vessel varies due to their distinct functions and the varying demands placed upon them. The walls of these vessels typically consist of three distinct layers, as depicted in figure 2.3(a) [36]. The innermost layer, *tunica interna*, forms the inner lining of a blood vessel and is in direct contact with blood. Its smooth luminal surface is lined by endothelial cells which facilitate efficient blood flow by reducing surface friction. The basement membrane of the *tunica interna* is made up of elastic, collagenous fibres which provide tensile strength to the vessel. The middle layer, *tunica media*, is primarily made up of smooth muscle and is usually the thickest layer. It provides both structural support for the vessel and actively changes vessel diameter to regulate blood flow and BP. An increase in sympathetic stimulation typically stimulates the smooth muscle to contract, squeezing the vessel wall and narrowing the vessel diameter, this is known as vasoconstriction. Conversely, the widening of a blood vessel is known as vasodilation. The outermost layer, *tunica externa*, is composed of connective tissue that attaches the vessel to the surrounding tissue. This layer is made up of elastic and collagen fibres and also contains numerous sympathetic nerves in order to regulate vasoconstriction and dilation.

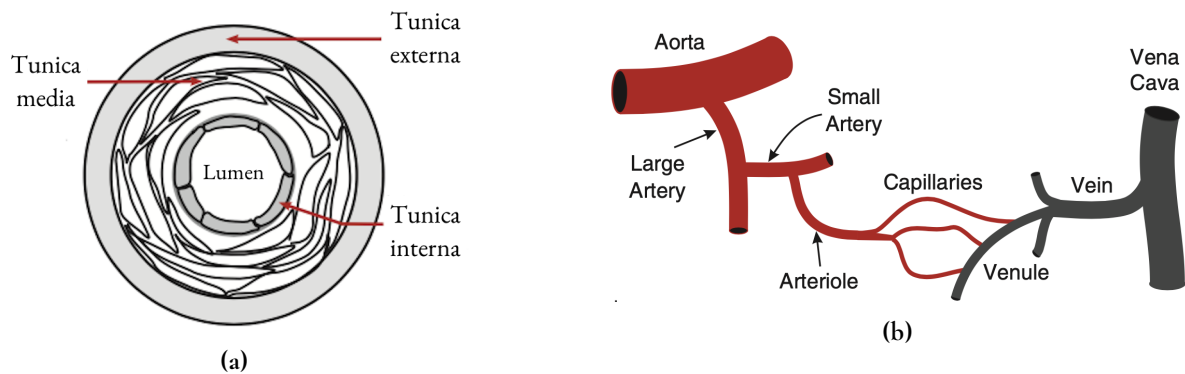


Figure 2.3: Structure of blood vessels. (a) Blood vessels are composed of three distinct layers: *tunica interna* (inner layer for smooth blood flow), *tunica media* (middle layer, provides strength to the vessel), and *tunica externa* (outer layer to attach the vessel to surrounding tissue). (b) Varying blood vessel structure across the circulatory system from main elastic arteries (with a large smooth muscle component), to capillaries (which have a large surface area), to veins (which have a small smooth muscle component and valves). Both figures were adapted from [10] with permission.

Because of the energy required to transport blood around the circulatory system, the heart exerts a strong force during left ventricular contraction. As a result, the pressure that blood exerts on the arterial walls is strongest when blood leaves the heart through the arteries, weaker in the arterioles, and lowest in the capillaries. Figure 2.3(b) depicts changes in the overall structure of the blood vessels across the circulatory system. In order to withstand high pressure, the arteries are the largest vessels in the body with the thickest smooth muscle layer. This gives the arteries large compliance, allowing them to swell without tearing at times of high pressure. Vessel swelling acts as a pressure reservoir, momentarily storing the mechanical energy which is then converted into kinetic energy of the blood as the elastic fibres recoil. As a result, elastic arteries are able to propel blood forwards while the ventricles are relaxing. Capillaries are designed for efficient gas exchange between blood and cells. They are, therefore, microscopically small (5-10 μm in diameter [36]) in order to increase the surface area to volume ratio. Capillaries contain a one-cell thick endothelium layer to allow for easy exchange of substances across the vessel walls.

The main function of veins is to transport deoxygenated blood back to the heart. Veins have a larger diameter than arteries and thus offer relatively little resistance to flow. Relative to their size, veins have thin walls that do not readily maintain their shape as the pressure of the blood is low and with little pulsation. Many veins, especially those in the lower limbs, contain valves to prevent the backflow of blood.

2.3 Blood pressure

BP is a measure of the force the blood exerts per unit area of blood vessel wall [10]. BP values are typically given in units of millimetres of mercury (mmHg) since the first gauges to measure BP

used mercury [37]. Regulating BP to be within 'normal' limits is vital to maintaining the essential flow of blood to the body's organs and tissues. If BP drops too low (hypotension), blood flow to vital organs may be inadequate, potentially resulting in shock (reduction in tissue perfusion) leading to cellular and organ dysfunction [11]. Conversely, if BP rises too high (hypertension), blood vessels can be damaged, increasing the risk of clots and possible ruptures. Additionally, sustained hypertension has been shown to be one of the highest risk factors for the majority of CVDs such as coronary disease, left ventricular hypertrophy, atrial fibrillation, cerebral stroke, and renal failure [12]. BP is, therefore, closely regulated by the ANS which consists of the sympathetic and parasympathetic nervous systems [10].

BP varies over a cardiac cycle. When the left ventricle contracts and pushes blood into the aorta, the pressure produced is called the systolic blood pressure (SBP), the maximum pressure over a cardiac cycle. When the heart is resting following the ejection of blood, the pressure within the arteries is much lower; it is called the diastolic blood pressure (DBP), the minimum pressure over a cardiac cycle. The difference between the two is known as the pulse pressure (PP). Mean arterial blood pressure (MAP) is the mean BP over a cardiac cycle. Based on the relative times spent in systole and diastole, MAP can be approximated as:

$$\text{MAP} = \text{DBP} + \frac{(\text{SBP} - \text{DBP})}{3} \quad (2.1)$$

where $1/3$ is the systole ratio of the cardiac cycle [10]. MAP can also be defined in terms of the properties of the cardiac cycle and vasculature of the blood vessels in the body [38]:

$$\text{MAP} = \text{CO} \cdot \text{TPR} \quad (2.2)$$

where TPR is the total peripheral resistance.

BP is thus maintained by constantly adjusting CO, the total amount of blood in the vasculature, and the degree of vasodilation and vasoconstriction in blood vessels. Vasodilation and vasoconstriction occur mainly in the arterioles, which have a *tunica media* composed almost entirely of smooth muscle. Different regulatory mechanisms are employed to control BP in the short (within seconds), medium (within days), and long-term (within years) intervals. We provide a brief overview of these in the following section.

2.3.1 Short-term blood pressure regulation

Short-term BP variation refers to beat-by-beat fluctuations in BP due to the interplay of multiple cardiovascular control mechanisms, such as baroreceptors, chemoreceptors, and Mayer waves.

2.3.1.1 Baroreceptor reflex

Baroreceptors are pressure receptors located in the walls of the aortic arch and carotid sinuses (at the bifurcation of the internal carotid artery at the level of the neck). When BP increases in these arteries, the baroreceptors are activated, increasing their input to the ANS. In response, the ANS increases parasympathetic nerve activity transmitted to the heart, slowing HR down. At the same time, sympathetic stimulation to the blood vessels is inhibited, inducing vasodilation and a decrease in TPR. The net result is a fall in BP. Conversely, if the pressure falls, the rate of baroreceptor discharge also falls. The ANS then increases the sympathetic drive to the heart to speed it up. Sympathetic activity in blood vessels is also increased leading to vasoconstriction, and consequentially an increase in TPR. The physiological effect of sympathetic and parasympathetic stimulation is summarised in table 2.1. This mechanism is called the baroreceptor reflex and acts as a negative feedback control system to reduce fluctuations in arterial pressure. However, the baroreflex cannot regulate BP over the medium and long-term because the mechanism that triggers baroreceptors resets itself once a more adequate BP is restored [39].

Table 2.1: Sympathetic and parasympathetic mechanisms controlled by the ANS. These mechanisms have opposing actions where one system activates a physiological response and the other inhibits it.

	Sympathetic stimulation (fight or flight)	Parasympathetic stimulation (rest and digest)
Heart	HR ↑ Force of contraction ↑	HR ↓ Force of contraction ↓
Blood vessels	Smooth arteries and arterioles constrict	Blood vessels do not have a parasympathetic supply
Net effect on BP	BP ↑	BP ↓

2.3.1.2 Chemoreceptors

Chemoreceptors are nerve endings spread over the arterial system that are sensitive to changes in carbon dioxide, oxygen and blood pH. Activation of these peripheral chemoreceptors is triggered by factors indicating failing tissue perfusion, such as rising CO₂, falling O₂, or falling blood pH. This increases the sympathetic nervous stimulation in the heart and causes a corresponding increase in HR and BP.

2.3.1.3 Mayer waves

Mayer waves are cyclical waves in BP thought to be driven by natural oscillations of the baroreceptor and chemoreceptor reflex control mechanisms [40]. These oscillations induce changes in the sympathetic vasomotor tone of arterial blood vessels. They occur at a relatively stable characteristic

frequency of around 0.1 Hz that is independent of sex, age and posture. The amplitude of the Mayer waves has been shown to be an indicator of sympathetic nervous activity.

2.3.2 Mid-term blood pressure regulation

Mid-term BP variation refers to changes in BP that occur over the course of 24 hours, otherwise known as circadian rhythms. It is important to distinguish a circadian rhythm from a diurnal rhythm. Circadian rhythms are endogenous oscillations in physiological functions driven by a combination of extrinsic factors (such as cycles in daily activity, stress and feeding), and intrinsic internal circadian clocks with a period of approximately 24 hours [3, 4]. Diurnal rhythms are processes that recur every 24 hours and are often linked to the natural light-dark cycles [4]. BP follows an endogenous circadian rhythm in order to appropriately meet the metabolic demands of tissues and cells over 24-hours [3]. The internal circadian clocks evolved in order for organisms to better adapt to the 24-hour day-night cycles on our planet. Circadian clocks are present in most major tissues such as the heart, vasculature, kidneys, and the neuroendocrine system. The pacemaker for these peripheral clocks resides in the suprachiasmatic nucleus (SCN) of the hypothalamus in the brain [7]. The SCN synchronises with the light/dark cycle via direct retinal input and communicates with peripheral clocks through hormonal and neuronal signals [41].

Long-term disruptions of the circadian system have been shown to be risk factors for metabolic disorders, obesity, diabetes, cardiovascular disease, and cancer [42, 43, 44, 45]. A large body of evidence also now shows a strong link between long-term disruption of circadian rhythms and the development of neurodegenerative diseases, such as Alzheimer's disease [41].

Circadian variations in BP have been extensively investigated in the literature. The underlying mechanisms of these variations are complex and multi-factorial, involving the circadian timing system, autonomic nervous system, and water and sodium regulation [46]. The typical profile for a healthy individual consists of a dip in BP when falling asleep (a phenomenon known as dipping), reaching its lowest value during slow-wave sleep and the highest variability during rapid eye movement (REM) sleep. The typical pattern while awake includes two peaks, a large peak upon waking (known as the 'morning BP surge') and a smaller peak at around 7pm [47]. The dip in BP during sleep is a phenomenon known as the 'day-to-night dip' and the degree to which an individual dips may be characterised by the day-to-night dip ratio in SBP (DR_{SBP}) [48]. DR_{SBP} is defined as the percentage difference between the mean night-time SBP and the mean daytime SBP, normalised by the latter:

$$DR_{SBP} = \frac{\overline{SBP}_{Night} - \overline{SBP}_{Day}}{\overline{SBP}_{Day}} \cdot 100\% \quad (2.3)$$

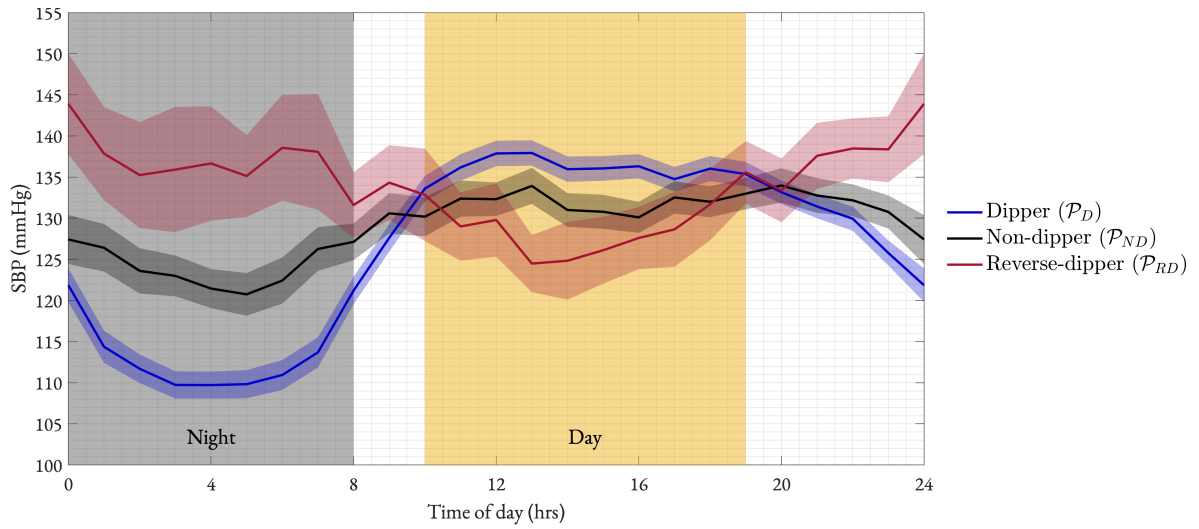


Figure 2.4: Depiction of the different SBP circadian rhythm phenotypes: 'dipper' $\mathcal{P}_D$, 'non-dipper' $\mathcal{P}_{ND}$, and 'reverse-dipper' $\mathcal{P}_{RD}$. Data was gathered from the 585 patients in the BP-Eth study, Armitage *et al.* [5]. Patients were grouped into $\mathcal{P}_D$, $\mathcal{P}_{ND}$, and $\mathcal{P}_{RD}$ phenotypes, and their hourly average and standard deviation of recorded BP values are shown. The DR_{SBP} criteria defining these phenotypes are outlined in table 2.2.

Healthy individuals typically experience a DR_{SBP} between -10% and -20%, driven by a reduction in sympathetic tone and an increase in vagal activity, both regulated by the SCN [49]. However, multiple BP phenotypes exist depending on the magnitude and direction of the change in nocturnal BP. Previous works have demonstrated the presence of 'non-dipping' ($\mathcal{P}_{ND}$, no nocturnal change in BP) and 'reverse dipping' ($\mathcal{P}_{RD}$, increase in BP during sleep) patterns [50]. Additionally, some of the literature further describes subcategories of 'dippers' ('extreme dippers', $\mathcal{P}_{ED}$ [15]) and 'reverse dippers' ('extreme reverse dippers', $\mathcal{P}_{ERD}$ [5]). The criteria for defining different BP circadian rhythm phenotypes are outlined in table 2.2.

To depict these phenotypes, we utilised data gathered from the 'BP in different ethnic groups' (BP-Eth) study from Armitage *et al.* [5]. BP-Eth was an observational study conducted between June 2010 and December 2012, with 585 patients registered at one of 28 general practices in the UK. All patients underwent 24-hour ABPM with BP measured every 30 minutes during the day and hourly during the night. Figure 2.4 depicts the averaged BP values for patients grouped into 'dipper' ($\mathcal{P}_D$), 'non-dipper' ($\mathcal{P}_{ND}$), and 'reverse dipper' ($\mathcal{P}_{RD}$) phenotypes. The width of each profile is proportional to the variance of the data for that data point, the wider width of the $\mathcal{P}_{RD}$ profile results from its relatively lower prevalence in the community.

In a systematic literature review, Cuspidi *et al.* [51] reported that the prevalence of the $\mathcal{P}_{RD}$ phenotype ranges from 3% to 39% depending on the study and the population in question. Armitage *et al.* [5] performed a retrospective analysis of $\mathcal{P}_{RD}$ prevalence in two patient populations. The first population consisted of 21,716 patients admitted to four acute hospitals in Oxford University

Table 2.2: BP circadian rhythm phenotypes classification used in this thesis based on an individual's SBP day-to-night dip ratio (DR_{SBP}). Note that $\mathcal{P}_{ED}$ and $\mathcal{P}_{ERD}$ are subcategories of $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ respectively and are not commonly analysed in the literature.

Phenotype	SBP Dip ratio (DR_{SBP}) range
'Extreme dipper' ($\mathcal{P}_{ED}$)	$DR_{SBP} < -20\%$
'Dipper' ($\mathcal{P}_D$)	$DR_{SBP} \leq -10\%$
'Non-dipper' ($\mathcal{P}_{ND}$)	$-10\% < DR_{SBP} < 0\%$
'Reverse dipper' ($\mathcal{P}_{RD}$)	$0\% \leq DR_{SBP}$
'Extreme reverse dipper' ($\mathcal{P}_{ERD}$)	$10\% < DR_{SBP}$

Hospitals NHS Foundation Trust, UK. Among this cohort, the prevalence of the $\mathcal{P}_{RD}$ phenotype was found to be 48.9%. The second population consisted of 770 primary care patients undergoing ambulatory BP monitoring. Among this cohort, the prevalence of the $\mathcal{P}_{RD}$ phenotype was found to be 10.8%. Overall, individuals in the $\mathcal{P}_{RD}$ cohort tended to be older, more frequently male, and with a higher prevalence of history of CVD than individuals in the $\mathcal{P}_D$ and $\mathcal{P}_{ND}$ cohorts. Indeed, substantial evidence over the past couple of decades has pointed to a significant link between disrupted dipping patterns in BP, such as $\mathcal{P}_{ND}$ or $\mathcal{P}_{RD}$ phenotypes, and development of CVD, such as myocardial infarction, stroke, angina, and heart failure [7]. In addition, an increased nocturnal BP has been associated with endothelial dysfunction [52], sleeping disorders [53], and reduced renal sodium excretion [54].

Obrien *et al.* [50] were the first to report a relationship between dipping status and cardiovascular events in their cohort of 123 patients. Those classified into the $\mathcal{P}_{ND}$ cohort had a significantly ($p < 0.01$) higher frequency of stroke than those in the $\mathcal{P}_D$ cohort. Iqbal *et al.* [6] examined 24-hour BP profiles in 297 patients from Chesterfield hospital, UK. Over half (54.7%) of the patients monitored were found to have a diminished nocturnal dip. Cardiovascular outcomes were then examined in all patients over a follow-up period of 2,015 days $\pm$ 116 days. A significant ($p < 0.05$) relationship was found between patients with a diminished nocturnal dip and future fatal and non-fatal cerebrovascular events. Ohkubo *et al.* [55] examined DR_{SBP} as a continuous variable as opposed to classifying individuals into phenotypes. Their study of 1,542 individuals from Japan showed that on average each 5% increase in DR was associated with a 20% greater mortality risk from cardiovascular disease.

Aside from the risk of cardiovascular disease, disruption to the BP circadian profile has been shown to be associated with neurodegenerative diseases [41]. For example, a study of 111 patients with Parkinson's disease (mean age 67.8 years) from Spain reported that 71% of patients did not experience a dip in BP [56].

The association between the absence of a nocturnal dip and adverse cardiovascular and neurodegenerative outcomes has prompted suggestions that BP-lowering medication should be administered at night-time in order to induce BP dipping. Hermida *et al.* [57] conducted a study in which they randomly assigned 2,156 patients with untreated hypertension into two groups, one taking medication in the morning and the other taking medication at bedtime. After a median follow-up period of 5.6 years, the patients in the bedtime cohort exhibited a significantly lower prevalence of total CVD events (cardiovascular death, myocardial infarction, ischaemic stroke, and hemorrhagic stroke) than the morning cohort ($p < 0.001$). However, it should be noted that several concerns surrounding these results have been raised [58, 59]. The Treatment in Morning versus Evening (TIME) study attempted to reproduce Hermida's results in a cohort of 21,104 adults with hypertension [60]. Following a similar protocol, after a median follow-up period of 5.2 years, Mackenzie found no differences in cardiovascular outcomes between either group. Neither of these studies, however, reported the prevalence of the different BP phenotypes, and thus cannot support or refute the hypothesis that individuals with different circadian BP phenotypes may respond differently to morning or evening antihypertensive dosing. In particular, those in the $\mathcal{P}_{RD}$ cohort may benefit from evening dosing.

Importantly, in multiple studies, the association between night-time dipping status and cardiovascular / neuro-degenerative diseases appears to be independent of baseline daytime BP values [15, 55, 61], highlighting the importance of appropriate tracking of an individual's nocturnal BP.

2.3.3 Long-term blood pressure regulation

Long-term BP variation refers to changes in BP that occur over multiple days, stretching onto the lifetime of an individual. These variations are typically a result of intrinsic, physiological changes as opposed to reactions from external excitation.

Seasonal fluctuations in temperature impact BP levels. In the colder winter months, daytime BP is typically higher than in the summer. This relationship is induced by vasoconstriction as a result of exposure to colder temperatures. In the summer, night-time BP levels tend to be elevated. This relationship is thought to be related to physical discomfort, dehydration, and poor sleep quality due to warm weather [62]. Kollias *et al.* [63] performed a meta-analysis of seasonal variation in BP across multiple studies. On average, daytime SBP was 6.1 (CI_{95} : 5.1 - 7.0) mmHg higher and night-time SBP was 1.3 (CI_{95} : 0.2-2.3) mmHg lower in the winter than in the summer.

Figure 2.5 depicts averaged lifetime variations in SBP and DBP observed in the Framingham Heart Study [64, 65]. The study followed a large cohort of 4,993 individuals over 4 years in Framingham, Massachusetts, USA. As ageing occurs, there are notable changes in the vasculature

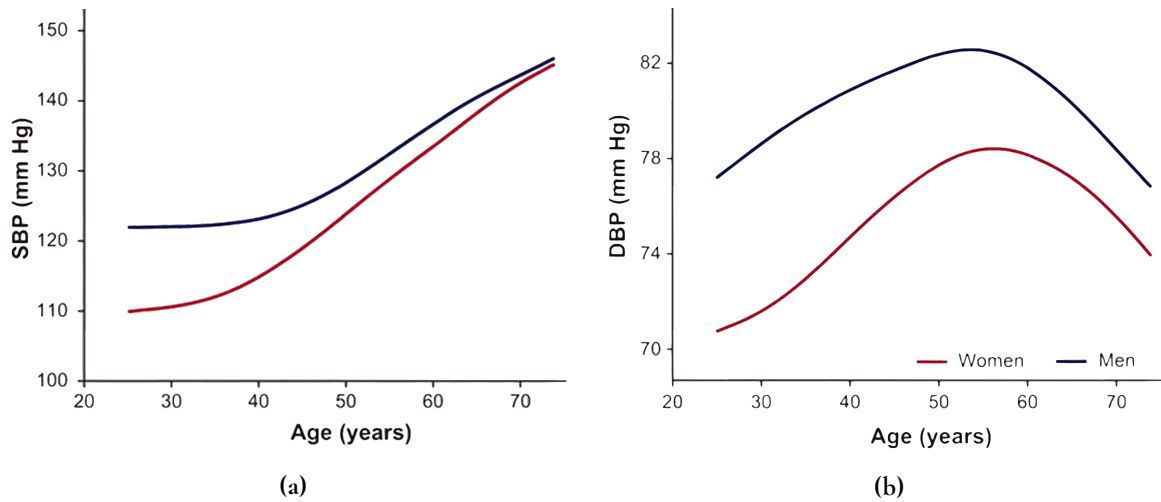


Figure 2.5: Averaged lifetime BP changes in 4,993 participants of the Framingham Heart Study. (a) SBP, (b) DBP. Results are split by sex. Figures from Chen *et al.* [64] with permission.

that cumulatively result in a consistent increase in SBP. Over time, the elastic structure of central arteries changes, resulting in arterial walls thickening, hardening, and losing their elasticity (a phenomenon known as arteriosclerosis). Additionally, the buildup of fatty plaques on the inner surface of the wall (known as both atherosclerosis and calcification) narrows the diameter of the artery. Arteriosclerosis and atherosclerosis both increase large artery stiffness (LAS) and afterload (resistance to flow), resulting in the heart needing to reach a higher LVP in order to eject blood. Driven by an increase in peripheral vascular resistance (PVR) DBP initially follows a pattern similar to that of SBP but reaches an apex around the fifth decade whereafter it decreases with the increase in LAS [66]. Men and women follow similar patterns with women generally having a lower BP than males at a young age. Females additionally experience hormonal changes linked to menopause that result in a steeper increase in SBP after their fourth decade [67]. The age and sex-related patterns in BP shown in figure 2.5 have been observed in virtually all developed and developing countries [68, 69]. Although genetics play a significant role in this long-term variation in BP [70], evidence suggests that lifestyle factors such as diet and physical exercise levels are the major causes of the changes in global BP trajectories [71].

Persistently high BP is known as hypertension. Based on an analysis of data from 135 population-based studies that included 968,419 adults from 90 countries, Mills *et al.* [13] concluded that the prevalence of hypertension across the globe was 31.1% (CI_{95} : 30.0%-32.2%). This prevalence is increasing, especially in low- and middle-income countries [13, 72] where access to anti-hypertensive medication is more limited than in high-income countries. Elevated BP increases the load on the heart and stress on the arteries, which can lead to a weakening of the heart muscles (cardiomyopathy) and a thickening of the left ventricle walls (left ventricular hypertrophy). Hypertension is the

Table 2.3: Summary of main BP measuring techniques and their associated advantages and disadvantages. Abbreviations: Sphyg. cuff - Sphygmomanometer cuff

Name	Use	Advantages	Disadvantages
Arterial line	Intensive care unit	Gold standard for continuous BP measurement.	Invasive. Risk of infection. Requires a skilled practitioner to appropriately place the catheter. Limits movement of the individual.
Auscultation Sphyg. cuff	Routinely used in clinical practice.	Non-invasive.	Requires training to recognise Korotkoff sounds. Irregularly sampled and manual.
Oscillometric Sphyg. cuff	Routinely used in both clinical practice and in a home setting	Non-invasive, easy to operate and automatic.	Irregularly sampled and uncomfortable.
Applanation tonometry	Limited practical use	Non-invasive, continuous measurement of BP.	Requires calibration in practice. Limits free movement. Reportedly low accuracy.
Volume-clamp method	Mostly restricted to research	Non-invasive, continuous measurement of BP.	Large. Expensive. Reported user discomfort and disruptive (finger numbness).

strongest risk factor for the majority of CVD, including coronary heart disease, atrial fibrillation and stroke. CVD is the largest cause of non-infectious mortality in the world [12, 72, 73]. In addition, chronic hypertension has been shown to be an important modifiable risk factor for neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease and related dementias [41, 74]. In a cross-sectional study of 153,996 adults spanning urban and rural areas across 17 countries, Chow *et al.* [16] found that despite the high risks of CVD, hypertension had a diagnosis rate of only 46%. Additionally, only a third of the individuals diagnosed with hypertension were able to keep their BP levels under control.

In summary, accurate diagnosis of hypertension is vitally important to minimise all-cause mortality. However, current diagnosis rates are especially low, leading to hypertension often being referred to as the "silent killer" [20].

2.4 Measuring blood pressure

Many different techniques and devices have been developed to measure BP. A summary of these, as well as their advantages and disadvantages, is provided in table 2.3.

2.4.1 Arterial line

The gold standard for BP measurement is an invasive technique using an intra-arterial catheter ("arterial line") [19, 20]. This measures instantaneous BP by placing a pressure transducer in contact

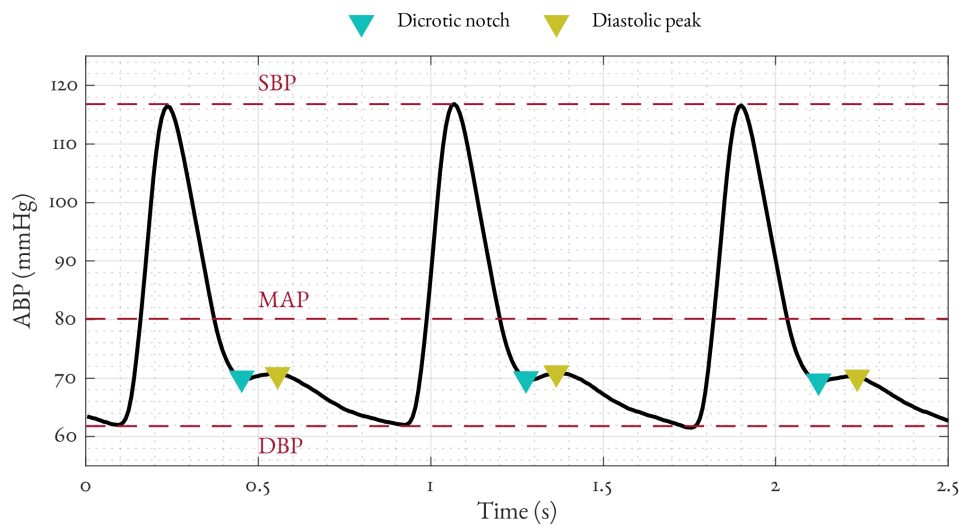


Figure 2.6: Example of an arterial blood pressure (ABP) waveform recorded in the brachial artery. SBP and DBP correspond to the peak and nadir of each beat respectively. MAP is the mean blood pressure over a cardiac cycle. The dicrotic notch and diastolic peaks are inflections in the descending limb of the ABP waveform as a result of the closure of the aortic valve.

with blood flow in an artery. Common sites for the insertion of the catheter are peripheral arteries such as the radial, brachial or femoral arteries [20]. The arterial pressure waveform is the result of a shock wave emitted from the left ventricle during left ventricular contraction. This shock wave travels much faster than the velocity of the blood being ejected. An example of the arterial BP (ABP) waveform recorded by an arterial line is shown in figure 2.6. The peaks and troughs indicate SBP and DBP respectively. Another distinctive feature of the ABP waveform is the inflection point (dicrotic notch) followed by a local maximum (diastolic peak) usually seen around one-third of the way down the descending limb. These points are the result of an increase in pressure due to the closure of the aortic valve [75] (see figure 2.2) and therefore marks the end of systole and the beginning of diastole [76]. The dicrotic notch is not present in all individuals, factors such as the stiffness of the peripheral arteries have been shown to have an important role in its appearance [77]. Typically, it is present in subjects with healthy, compliant arteries [78], whereas it has been reported to be diminished in elderly individuals due to the progression of atherosclerosis [77, 79].

Measuring BP using an arterial line is an invasive technique that has a number of disadvantages. The process of inserting a catheter is invasive, can be painful, cause discomfort, and carry the risk of complications such as infection and bleeding [19]. In severe cases, cannulation of an artery may lead to embolism, nerve damage, or ischaemia [80]. Given these risks, cannulation is typically only performed in intensive care settings using an appropriately trained clinician.

2.4.2 Sphygmomanometer cuff

Non-invasive BP measurements are typically performed using a sphygmomanometer cuff, which consists of an inflatable arm cuff, a pressure gauge, and a pump [80]. The cuff is typically wrapped around the upper left arm, as the left brachial artery carries blood directly from the aorta and thus provides the closest assessment of aortic pressure. If the upper arm is inaccessible, for example during surgery, there are cuffs that wrap around the forearm or the leg [81]. The cuff is inflated using the pump to a pressure above the physiological limit of SBP (known as suprasystolic pressure). At this point blood flow in the arteries within the upper arm is completely occluded. The cuff is then gradually deflated to below the physiological limit of DBP (subdiastolic pressure) allowing blood flow to gradually resume. The pressure gauge measures the air pressure inside the cuff during this process. There are two main methods for estimating SBP, MAP and DBP during cuff deflation: auscultation and oscillometry.

2.4.2.1 Auscultation

In the auscultatory method, a trained clinician uses a stethoscope to listen for characteristic Korotkoff sounds during cuff deflation. There are five characteristic phases of the Korotkoff sounds (snapping, murmurs, thumping, muffled, silent) that occur sequentially as blood flow in the artery is gradually resumed. SBP is typically associated with the cuff pressure at the first Korotkoff sound, whereas DBP is typically associated with the fifth Korotkoff sound [81, 36]. This technique, however, requires an appropriately trained professional to record an accurate measurement. Operator variability due to differing interpretations of DBP can also be significant [81].

2.4.2.2 Oscillometry

In the oscillometric method, the process of computing BP is automated, thus removing the need for a trained professional. The sphygmomanometer includes two, overlapping cuffs. One cuff inflates to occlude the artery (occluding cuff) and the other cuff (sensing cuff) senses the oscillations in the arterial wall as the occluding cuff is deflating. When the occluding cuff pressure is above the suprasystolic pressure, no oscillations are present as blood flow is completely occluded. During gradual deflation, oscillations in pressure are detected by the sensing cuff, reflecting the pulsatile arterial blood volume. The amplitude of these oscillations reaches a maximum at MAP. Values of SBP and DBP are then estimated from analysis of the relationship between the occluding cuff pressure and the amplitude of the sensing cuff oscillations. The algorithms used for estimating SBP and DBP are proprietary and specific to the device being used. They are typically not disclosed by the manufacturers [82, 83].

Many oscillometric devices have been validated against auscultation using the ANSI/AAMI/ISO (American National Standards Institute, Association of the Advancement of Medical Instrumentation, and International Organisation for Standardisation) protocol, which is accepted by the US Federal Drug Administration (FDA). This protocol involves a comparison of three pairs of BP measurements using both an auscultatory device as the gold standard and an oscillometric device as the test device. More than 85 subjects of diverse BP levels are required. The difference between the test and reference devices must yield mean and standard deviation of errors < 5 mmHg and < 8 mmHg respectively [84].

The oscillometric method has replaced the auscultatory method as the most widely used BP measuring technique in clinical practice [83, 25, 85]. Despite being considered the standard method, sphygmomanometers come with significant limitations [25]. Most significantly, they only provide intermittent BP values, removing the ability to assess high-frequency BP variations [85]. The size of the cuff is also a critical factor in obtaining accurate readings, with cuffs that are too small or too large leading to overestimations or underestimations of BP, respectively [86, 87]. The measurements can be easily disrupted by movement, and so the user is required to remain still for the duration of the recording, which can take up to a minute [37]. Finally, users often report significant discomfort from the pressure generated during cuff inflation. This can lead to cuff-inflation hypertension where users experience a short increase in BP due to the stress associated with the measurement [81].

2.4.3 Applanation tonometry

Applanation tonometry is a non-invasive method for recording a BP waveform from large, superficial arteries. The method involves measurement of the physical motion of major arteries as pulsatile blood flows [88]. A miniature transducer is attached to the skin above a pulsating artery with strong bone support (typically the radial artery is used). The pressure of the sensor flattens ('applanates') the artery. Under optimal conditions (complete applanation of the artery, and the transducer completely encompassing the flattened artery), the pressure gradient between the arterial wall and the sensor is zero and, therefore, the arterial pressure waveform can be recorded by the sensor [89]. However, satisfying the optimum conditions mentioned above has proved difficult, requiring the non-linear relationship between vessel volume and pressure in the elastic vessel walls to be accounted for. This is typically performed by routinely calibrating with cuff BP values [90, 89].

Cheng *et al.* [90] reported the accuracy of applanation tonometry devices, such as the T-line device shown in figure 2.7(a), using a meta-analysis across 22 studies (857 subjects and 1167 measurements). Errors in BP measurements were reported as high as -8.2 ± 10.3 mmHg for SBP and 7.6 ± 8.7 mmHg for DBP. The poor performance of BP measurement and the



Figure 2.7: Example devices using the applanation tonometry and volume clamp methods. (a) T-line using the applanation tonometry method (T-Line, Tensys Medical). (b) Finapres using the volume clamp method (Finapres, Finapres).

cumbersome equipment set-up (see figure 2.7(a)) result in applanation tonometry not being used in clinical practice.

2.4.4 Volume clamping

Volume clamping is a non-invasive method for recording a continuous BP waveform at the finger. As the pressure wave propagates through the arterial tree, compliant arteries swell and shrink in order to carry the load. The volume clamp method estimates a surrogate of BP as the pressure that needs to be applied to an artery to keep the diameter of the artery constant (to 'clamp' it) [91]. The system uses an inflatable finger cuff combined with a photodiode. The photodiode continuously measures the diameter of the artery. A feedback control system adjusts the finger cuff pressure in response to the arterial diameter measurement [92].

However, the volume clamp method is highly susceptible to movement. In cases of severe vasoconstriction, peripheral vascular disease, or distorted fingers due to arthritis, clinical experience has shown that it may be difficult to obtain a valid BP measurement using finger cuffs. This limits its use in a hospital setting [93]. Additionally, some users report discomfort from the congestion in venous return from the fingertip where the cuff is placed [80]. For this reason, manufacturers recommend changing the cuff to another finger after a certain period of monitoring. Furthermore, compared to conventional intermittent BP measurement devices, the volume clamp method is relatively expensive and large (see figure 2.7(b)).

Finapres is a commercial device that uses the volume clamp method [94] (see figure 2.7(b)). Silke *et al.* [95] reported the accuracy of the Finapres, against the gold-standard arterial line, using a meta-analysis across 20 publications (449 patients, number of measurements not reported). Errors in BP measurements were 2.2 ± 12.4 mmHg for SBP and -0.3 ± 7.9 mmHg for DBP, suggesting that the Finapres has a good average performance but with large variance. Chin *et al.* [96] additionally

Table 2.4: Standard clinical thresholds for assigning an individual into a BP category based on their SBP or DBP values recorded in the clinic. Values are given in units of mmHg.

Blood pressure category	SBP	DBP
Hypotensive	≤ 90	OR < 60
Normotensive	90 - 120	OR 60 - 80
Pre-hypertensive	120 - 140	OR 80 - 90
Hypertensive	> 140	OR > 90

examined the reproducibility of the Finapres by simultaneous measurements of BP with two independent devices. Significant ($p < 0.05$) differences were observed between the two devices for both SBP and DBP estimation, demonstrating the unreliability of the device.

2.4.5 Clinical assessment of hypertension

To assess if an individual has hypertension, the 2018 guidelines from the American Heart Association Task Force on Clinical Practice Guidelines recommend that sphygmomanometer BP measurements are recorded in the clinic [97]. Typically, two BP readings are taken by a trained clinician 1-2 minutes apart while the individual is in the sitting position. Using the average of these measurements, the clinician will then decide on a treatment plan using the categories outlined in table 2.4. These thresholds are general guidelines on the state of the cardiovascular system for an individual and clinicians may make adjustments based on the individual (for example their sex and age) when determining treatment plans. For individuals that are classified as hypertensive, a clinician will first recommend lifestyle changes (such as exercise and diet). An additional clinical measurement of BP is taken three months later. If the patient is persistently hypertensive home monitoring is recommended.

2.4.6 Home and ambulatory monitoring of blood pressure

Persistent hypertension when in an ambulatory, or at-home, setting has been shown to be a better predictor of CVD morbidity and mortality than clinical BP measurements [98, 99, 100, 19]. Home monitoring also increases an individual's awareness of how their BP responds to medication or dietary changes. There is evidence to suggest that individuals whose care incorporates self-monitoring of BP are much more likely to achieve adequate BP control compared to those using the current systems of care [101]. Two types of devices have been developed for out-of-clinic monitoring of BP: Home and ambulatory BP monitors.

Home blood pressure monitoring (HBPM) is typically performed over a time span of 4 to 7 days [102, 103]. The user is provided with an oscillometric cuff monitor and instructed to take a number of BP measurements at certain times of the day, including one reading in the

morning just before breakfast, and one reading at night one hour before sleep [102]. HBPM is typically recommended when there is a suspicion of the white coat effect (higher BP values in the presence of a clinician) or masked hypertension (lower BP values in the clinic), or to study a patient's response to anti-hypertensive treatment.

ABPM is typically performed over 24 or 48 hours with readings taken using an automated oscillometric cuff monitor. The monitor is typically programmed to automatically inflate every 30 minutes during the day and every hour during the night [21]. The user is required to wear the cuff at all times. ABPM is typically recommended when a patient has an apparent drug resistance, suspected autonomic dysfunction or for the assessment of their dipping status [103]. At the end of the recording procedure, the BP values are downloaded onto a computer for a clinician to analyse and decide on an adequate treatment plan.

The comparison between clinic-based and out-of-clinic BP measurements has revealed several discrepancies. The first is the "white coat effect" whereby a patient's BP increases in the presence of a clinician due to stress. This can lead to normotensive patients being categorised as hypertensive in a clinical setting. This effect is prevalent in up to 43% of patients, with even higher rates in the elderly population [104, 21]. The second form is "masked hypertension", whereby a hypertensive patient appears normotensive in a clinical setting. Causes of masked hypertension have been linked with a number of factors such as regular alcohol consumption, smoking, and obesity [105]. The prevalence of masked hypertension is between 8% and 20% among adults [105]. The third form is "nocturnal hypertension" whereby a patient's BP is in the hypertensive range at night-time but not during the day. These individuals can be further categorised into the $\mathcal{P}_{RD}$ phenotype. For these individuals, their cumulative BP load (defined as the percentage of abnormally elevated BP readings over a period of time [85]) cannot be determined without the use of ABPM.

While ABPM provides valuable insights into cardiovascular health, it also has several significant limitations. ABPM relies on a sphygmomanometer cuff, which assumes that BP does not change significantly during cuff inflation and deflation (which can often take up to a minute [37]). This assumption holds when considering clinical measurements, as the clinician instructs the individual to sit completely still with their arms by their side. However, when used for ambulatory monitoring, the individual can be moving, eating or changing posture during this time, all of which affect the recorded BP measurement. In addition, frequent cuff inflations can cause significant discomfort to users, particularly during sleep. Viera *et al.* [23] assessed the tolerability of ABPM among 60 participants and found that 70.2% reported the device woke them from their sleep at least once, 33% reported pain from the device, and 19.6% reported that it stopped them from sleeping altogether. Elliot *et al.* [106] assessed the tolerability of ABPM across 650 individuals, with 75% reporting

that they found the procedure unpleasant. A survey comparing the attitudes of 200 individuals with hypertension towards different BP measurement techniques found that ABPM was the least accepted method [107]. Furthermore, it has been suggested that cuff inflations cause sleep arousal, which transiently increases BP so that the levels detected by the ABPM are higher than during undisturbed sleep [14]. Davies *et al.* [108] studied sleep arousal via electroencephalography (EEG) in six subjects undergoing ABPM. Beat-by-beat BP was also obtained via the volume clamp method. It was found that sleep arousal from cuff inflation lasted on average 16 seconds. During this time, the SBP was shown to rise by a mean value of 6.3mmHg. No significant rise in DBP was found. A study by Sherwood *et al.* [109] showed that 'non-dippers' had worse sleep quality compared to 'dippers'. This indicates a link between sleep arousal and an increase in BP. There is evidence to suggest, therefore, that ABPM can overestimate true SBP during sleep.

As measurements are taken manually and not at night-time, HBPM is often the preferred method of measuring BP in an out-of-clinic setting [110]. Nasothimiou *et al.* [111] reported that a significantly ($p < 0.05$) higher preference for 7-day HBPM over 24-hour ABPM existed among 105 patients from a hypertension clinic in Athens, Greece. Indeed, ABPM is largely unavailable to primary care practitioners in the United States [112]. In the United Kingdom, NHS England and NHS Improvement have initiated the distribution of HBPM monitors to individuals with suspected hypertension as part of the Blood Pressure @home program [113].

However, the drawback of HBPM is that it is unable to provide data on nocturnal BP values, thus rendering an assessment of the individual's circadian rhythm infeasible. There is a need, therefore, for an unobtrusive, cuffless ambulatory BP monitoring device which will accurately assess an individual's 24-hour BP profile without causing sleep arousal or significant user discomfort.

2.5 Photoplethysmography

Photoplethysmography is a non-invasive optical technique that monitors pulsations in blood volume within a peripheral vascular bed [28]. By measuring the attenuation of the light after passing through tissue, changes in the tissue's composition can be recorded over time [114]. The resulting PPG signal is a valuable source of physiological information since it is influenced by the cardiac, vascular, and autonomic nervous systems [28]. As a result, the PPG can be used to estimate important physiological parameters such as peripheral oxygen saturation (SpO_2), HR, and respiratory rate (RR).

A PPG signal is typically obtained using a pulse oximeter probe, which shines light into the skin using a light-emitting diode (LED). A photodiode then measures the amount of light transmitted

through the tissue (transmission PPG) or, the amount reflected from it (reflectance PPG). This measurement is converted into a voltage which is then sampled to produce a PPG time-series. The units of the PPG signal are therefore analogue-to-digital units (a.d.u).

Light travelling through biological tissue is absorbed by different substances that give rise to a quasi-static (DC) component as well as a pulsatile (AC) component (see figure 2.8). The AC component accounts for approximately 5% of the total absorption [114]. The static component consists of light absorbed by (for example) skin, bone, melanin, and venous blood. This component varies slowly due to respiration, vasomotor activity and the Traube Hering Mayer (THM) waves (see section 2.3.1.3) [28]. The AC component is driven by the pulsatile nature of blood flow and consists of changes in blood volume, blood vessel wall movement, and the orientation of the red blood cells [28]. The fundamental frequency of this AC component is dependent on HR.

The Bouguer-Beer-Lambert law characterises the attenuation of light travelling through a light-absorbing medium [115]. The intensity, I of light of wavelength λ passing through the medium, decreases exponentially with respect to the path length, L , and the medium's absorption characteristics:

$$I(\lambda) = I_0(\lambda)e^{-\alpha(\lambda)L} \quad (2.4)$$

where I_0 is the transmitted intensity, and $\alpha(\lambda)$ is the absorption coefficient of the medium at wavelength λ .

2.5.1 Modes of operation

The PPG signal can be recorded using devices operating in either transmission or reflectance mode. Figure 2.8 shows the origin of the PPG waveform, including the impact of the mode of operation on the observed PPG waveform.

In transmission mode, the light emitted from the LED is transmitted through the medium and detected by a photodiode positioned on the opposite side of the tissue. During systole, there is an increase in the absorption coefficient, α , and path length, L , leading to a decrease in the intensity of light recorded by the photodiode. For effective operation, the device must be located at a site where transmitted light can be easily detected, such as the fingertip or earlobe [114]. In order to restrict any environmental light interfering with the signal, these sensors will typically have a casing fitted around the LED and photodiode.

In reflectance mode, the photodiode is positioned adjacent to the LED and detects light that is reflected from the tissue. There are a number of possible locations for reflectance mode PPG devices, including the wrist (which is commonly used in smartwatches), forehead, and ring finger. These devices are generally more convenient for daily use than transmission mode devices

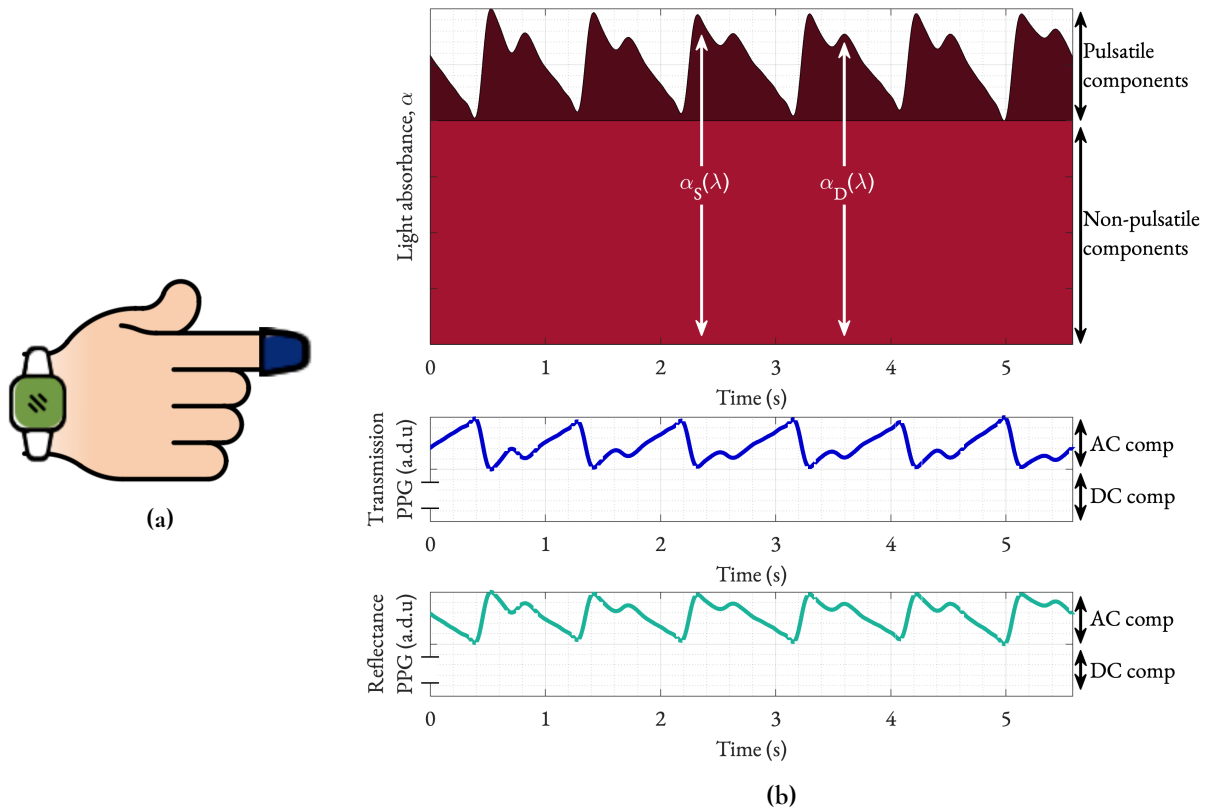


Figure 2.8: Origin of the PPG waveform. (a) Transmission mode PPG (blue) recorded by a finger pulse oximeter; reflectance mode PPG (green) recorded by a wearable smartwatch. (b) The PPG signal is composed of a quasi-static DC component, and an AC pulsatile component. During systole, the path length and α increase ($\alpha_S(\lambda)$), leading to a nadir in the transmission mode PPG and a peak in the reflectance mode PPG. In practice, the transmission mode PPG is often inverted when displayed.

(which have to enclose the whole finger). However, the quality of the PPG signal obtained with a device in reflectance mode is typically inferior to that obtained with a device using transmission mode [116]. This is due to the smaller fraction of the emitted light reaching the photodiode. Furthermore, a reflectance mode PPG is often more susceptible to movement artefacts. For instance, a smartwatch can move in relation to the wrist, which can distort the path length of the reflected light. Additionally, controlling the amount of ambient light detected by the photodiode becomes more difficult, and sudden increases in the amount of ambient light can saturate the signal.

2.5.2 Limitations and sources of error

In the past decade, PPG waveforms have become one of the most useful signals for monitoring the physiology of an individual [117]. Nevertheless, there are several factors that can affect the accuracy of the readings provided by a pulse oximeter. Some of the potential sources of errors are summarised below.

Low arterial perfusion The PPG waveform morphology is highly dependent on the presence of sufficient arterial perfusion in the measurement site (typically the fingertip or earlobe). In cases of low arterial perfusion, such as in individuals with compromised circulation due to cardiovascular diseases, hypotension, or peripheral vascular disorders, PPG signals may exhibit reduced amplitude and distorted waveforms [28].

Motion artefacts PPG signal recordings are susceptible to motion artefacts caused by patient movement, muscle contractions, or even minor sensor displacement in wearable devices. These motion artefacts can disrupt the PPG waveform, introducing noise and making it difficult to extract meaningful information from the signal. An accurate assessment of signal morphology and time-based parameters becomes compromised in the presence of motion artefacts.

Skin colour Skin pigmentation can influence the accuracy of PPG measurements, as different skin tones absorb and reflect light differently [28, 118]. Darker skin tones tend to absorb more light, making it more challenging for the PPG sensor to capture a reliable signal. Calibration and signal processing techniques that account for variations in skin colour can help improve accuracy for diverse patient populations.

2.6 Electrocardiography

The cardiac cycle is governed by electrical impulses transmitted throughout the heart [29]. A diagram of the electrical system of the heart is shown in figure 2.9(a). The sinoatrial (SA) node is the heart's natural pacemaker which governs the timings of the cardiac cycle. An impulse from the SA node causes atrial depolarisation forcing arterial contraction, and pushing blood into the ventricles. The impulse then moves to the atrioventricular (AV) node where it is delayed for a few milliseconds to allow for the ventricles to fill with blood. The signal travels from the AV node and divides into the left and right bundle branches, one for each ventricle. These two main branches divide further into a system of conducting Purkinje fibres that spread the signal through the ventricles, causing ventricular depolarisation and contraction. After atrial and ventricular contraction, the atrium and ventricles relax by repolarization to accept an inflow of blood.

As the heart goes through depolarization and repolarization, the electrical currents that are generated spread not only within the heart but also throughout the body tissues. This electrical activity can be measured by an array of electrodes placed on the body surface. The pattern of the electrical impulses in the heart is known as the ECG.

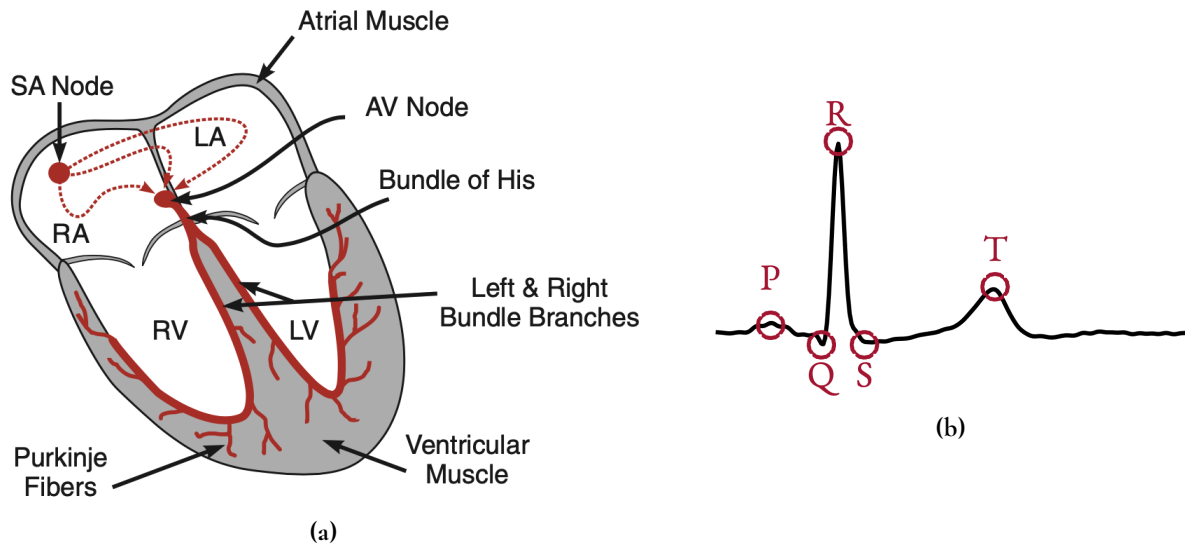


Figure 2.9: Origin of the ECG waveform. (a) The electrical system of the heart. The SA node is the heart's natural pacemaker. Electrical impulses from this node spread eventually to the Purkinje fibres and control the cardiac cycle. Figure from [10] with permission. (b) Annotated ECG waveform over a cardiac cycle. The P-wave indicates depolarisation of the atria. The QRS complex indicates ventricular depolarisation. The T-wave indicates ventricular repolarisation.

An ECG monitoring system typically uses electrodes in good electrical contact with the skin through a conductive gel. These electrodes detect the electrical activity of the heart as it beats, producing a periodic signal which can be displayed to clinicians for analysis. The potential difference across two electrodes is known as a lead. Common-mode interference (for example from electrical power lines) is commonly eliminated through a driven right leg circuit [114]. For basic monitoring, a single-lead ECG is recorded using three electrodes placed in Eithoven's triangle [29]. A diagnostic-quality ECG uses 10 electrodes and provides a more complete picture of heart rhythm recording 12 leads. There are four electrodes placed on each limb forming six leads (I, II, III, AVR, AVL, AVF) and six electrodes placed on the chest forming six leads (V1-V6).

The ECG waveform has a characteristic morphology, which varies depending on the exact placement of the electrodes. The waveform has three main features per heartbeat (see figure 2.9(b)): the P-wave, which is produced by the depolarisation of the atria; the QRS complex, which is due to ventricular depolarisation, indicating the start of ventricular contraction; and the T-wave, which is caused by the repolarisation of the ventricles. Atrial repolarisation occurs during the QRS complex, however, it is negligible when compared to ventricular depolarisation. Ventricular depolarisation is sudden and strong, therefore the QRS complex is the most prominent feature in the ECG. The QRS complex is used as a salient point when computing HR from the ECG. QRS detection algorithms such as the Pan-Tompkins [119] detector may be used to identify QRS complexes in an ECG waveform.

Abnormalities in the morphology of the ECG waveform can be interpreted by trained clinicians and are used in the diagnosis of multiple cardiovascular conditions. By analysing the shape of the ECG, a clinician may be able to determine the presence of arrhythmias, cardiac ischaemia, atrial fibrillation, myocardial infarction and other cardiovascular conditions [29].

2.6.1 Limitations and sources of error

ECG signals are widely used for diagnosing cardiac abnormalities and monitoring heart health. However, several factors can introduce errors in ECG signal recordings. Some of the potential sources of errors are summarised below.

Electrode placement The accuracy of ECG signal acquisition depends on precise electrode placement and consistent skin-electrode contact. Variations in electrode positioning and skin impedance, influenced by factors such as perspiration, hair, or skin conditions, can lead to distortions in ECG waveforms. These variations introduce considerable variability, potentially complicating ECG interpretation. Minimising this limitation necessitates adherence to standardised electrode placement protocols and skin preparation techniques to reduce electrode-skin interface-related errors.

Muscle and motion artefacts ECG recordings can be susceptible to artefacts originating from muscle electrical activity, posing a challenge in accurate signal interpretation. Muscle contractions and movement generate electrical signals that can overlap with cardiac signals, potentially leading to confusion and misinterpretation [120]. These artefacts can be especially prominent during patient movement or activities that engage skeletal muscles. To mitigate this issue, techniques such as patient relaxation and the use of specialised adhesive electrodes designed to reduce motion-related noise are employed. Additionally, advanced signal processing methods, such as adaptive filtering, may be applied to disentangle muscle-related electrical activity from genuine cardiac signals.

Electromagnetic interference ECG signals are sensitive to electromagnetic interference (EMI) from neighbouring electronic devices, power sources, and electrical equipment within clinical environments [121]. EMI can introduce noise, making accurate ECG interpretation difficult. Additionally, ambient noise within clinical settings exacerbates this issue. Effective strategies, including robust shielding, meticulous grounding, and signal filtering, are essential to mitigate EMI and noise effects.

2.7 Impedance cardiography

As blood is an electrical conductor, each heartbeat produces a change in the body's electrical impedance [122]. Impedance cardiogram is a non-invasive method of monitoring these changes in electrical impedance in order to derive estimates of haemodynamic parameters related to blood flow such as SV and PEP [123]. Typically, devices recording the impedance cardiogram (ICG) signal induce a constant low amplitude ($\approx 5\text{mA}$ [124]), alternating current ($\approx 20\text{-}100\text{ kHz}$ [124]) along the thorax and measure bioimpedance (Z) changes with each heartbeat. The ICG signal is then computed as the first derivative of the bioimpedance signal ($\frac{dZ}{dt}$). A typical ICG pulse and its characteristic points are shown in figure 2.10. The A-point presents the beginning of electromechanical systole and is associated with volume changes occurring due to atrial contraction. The B-point indicates the opening of the aortic valve and therefore denotes the beginning of left ventricular ejection. The C-point is the maximum of the ICG pulse and indicates the point of maximum mechanical contraction. The X-point indicates the closure of the aortic valve and marks the end of systole. The O-point indicates the opening of the mitral valve.

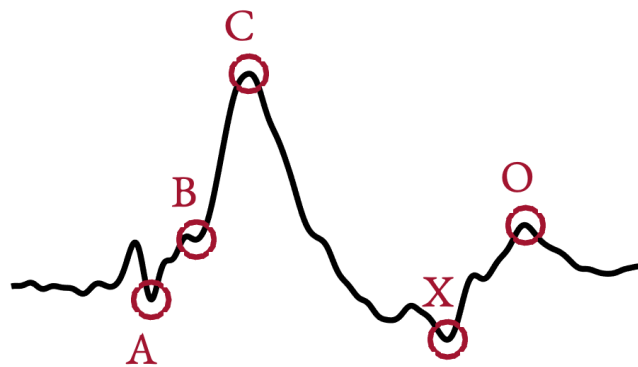


Figure 2.10: Annotated ICG waveform over a cardiac cycle. The A-point indicates the beginning of electromechanical systole. The B-point indicates the opening of the aortic valve. The C-point indicates the point of maximum contraction. The X-point indicates the closure of the aortic valve. The O-point indicates the opening of the mitral valve.

2.7.1 Limitations and sources of error

ICG is a valuable technique for assessing cardiac function. However, ICG signal recordings are subject to specific limitations and potential sources of error. Some of the potential sources of errors are summarised below.

Susceptibility to motion artefacts Patient movement during ICG measurements, including respiration-induced changes in thoracic impedance and variations in electrode-skin contact impedance, can compromise signal quality. These motion artefacts pose a significant challenge in clinical applications of ICG, necessitating the development and implementation of robust motion artefact detection and correction methods to ensure the accuracy of cardiac parameter assessments [125, 126].

Limitations in ambulatory monitoring Accurate ICG signal recordings typically require patients to remain at rest due to the limited portability of the equipment setup and the signal's susceptibility to motion artefacts. Accurate ICG measurement in an ambulatory environment remains a significant research challenge [125].

Frequency and blood velocity dependency The impedance of biological tissue exhibits a dependence on the frequency of the injected AC current. This is because the reactive component of the cell membrane behaves as a capacitor and reacts differently to variations in electrical current frequency [124, 127]. Moreover, the resistivity of blood has also been shown to be velocity dependent due to changes in the orientation of red blood cells induced by changing shear forces during blood flow [127]. Understanding the intricate relationship between frequency, blood velocity, and impedance signals is crucial for accurate interpretations of hemodynamic changes and underscores the complexity of ICG data analysis in various physiological contexts.

2.8 Estimating blood pressure using transit times

2.8.1 Background theory

2.8.1.1 Arterial stiffness

As blood flows through an artery, the arterial walls swell and stretch to accommodate the increase in pressure (known as the Windkessel effect [128]). This deformation transiently increases arterial wall stiffness (also called arterial tone) in accordance with the Young-Laplace law [129]. Stiffer arteries also increase cardiac afterload (the pressure against which the heart must work to eject blood) and thereby elevate BP. Hence, monitoring arterial stiffness can provide an indication of BP. Moreover, arterial stiffness has been shown to be an independent predictor of mortality in hypertensive patients [130]. The guidelines on the management of hypertension from the European Society of Hypertension acknowledge the potential role of arterial stiffness measurement in clinical management [131].

The main load-bearing components in large conduit arteries are elastin and collagen [132]. These components are located in the *tunica media* and *tunica externa*, respectively, of the arterial wall (see figure 2.3(a)). Collagen is about 400 times stiffer than elastin and only exerts tension when the arterial wall is stretched [133]. As a result, elastin predominantly determines arterial elasticity at low BP, whereas collagen significantly contributes to the elasticity at high BP [134]. Consequently, there theoretically exists a non-linear, monotonic, relationship between arterial stiffness and BP. In 1991, Geddes [135] derived the following relationship between the elastic modulus (stiffness) and BP in large conduit arteries:

$$E = E_0 \cdot \exp(a \cdot \text{BP}) \quad (2.5)$$

where E is the elastic modulus, E_0 is the elasticity at zero pressure, and a is a coefficient depending on the vessel, ranging from 0.016 to 0.018 mmHg⁻¹ [136].

It is well known that age-related arteriosclerosis progressively increases arterial wall stiffness, largely owing to the depletion of elastin, which is gradually replaced by stiffer collagen [134]. This increase in arterial stiffness is the main driving factor for the long-term BP relationship seen in figure 2.5.

In addition to the passive components of elastin and collagen, smooth muscle (located in the *tunica media* of the arterial wall) represents an active component of arterial stiffness [133]. Smooth muscle activation is controlled by the ANS to vary arterial stiffness in order to regulate the flow of blood. The smooth muscle component is strongest in the peripheral arteries and considered a negligible component in central arteries [31].

Aside from invasive pressure catheter recordings that are only occasionally used in technical validation studies, directly measuring arterial stiffness in humans is difficult [137]. To combat this, arterial stiffness is often indirectly approximated using pulse wave velocity.

2.8.1.2 Pulse wave velocity

Pulse wave velocity (PWV) is the group velocity of the pulse wave (also known as a pressure wave) produced by left ventricular contraction. It is important to make the distinction that the pulse wave is a transverse wave, whose oscillations are perpendicular to the flow of blood. Therefore, PWV should not be confused with the phase velocity of blood itself which is often 10 times slower [138]. In 1878, Moens and Korteweg [133] modelled the theoretical relationship between the pulse wave velocity of an incompressible fluid (such as blood) in an elastic, isotropic, tube (artery) and the elastic modulus (arterial stiffness) of the tube's wall. The resulting Moens-Korteweg equation is:

$$\text{PWV} = \sqrt{\frac{Eh}{\rho D}} \quad (2.6)$$

where E is the elastic modulus, ρ is the density of blood, h is the wall thickness, and D is the tube diameter. The Moens-Korteweg equation relies upon two assumptions: the arterial wall is thin such that it can be modelled as a thin shell; the thickness and radius of the artery remain fixed both elastically (in relation to changes in BP) and geometrically (along the arterial tree). As such, the Moens-Korteweg equation holds only for large conduit arteries [132].

Thus, across equations (2.5) and (2.6) changes in PWV, theoretically relate with changes in BP and could therefore be used to monitor changes in cardiovascular health [139]. There are multiple different methods to estimate PWV such as echocardiography, ultra-fast imaging, and ultrasonic shear waves [140]. In this thesis, we will focus on transit times such as the pulse transit time (PTT) and pulse arrival time (PAT).

2.8.1.3 Pulse transit time

Pulse transit time (PTT) is the time delay for the pressure wave to travel between proximal (close to the heart) and distal (away from the heart) arterial sites. PTT is related to PWV by:

$$\text{PWV} = \frac{L_a}{\text{PTT}} \quad (2.7)$$

where L_a is the distance along the arterial path between the proximal and distal points. Combining equations (2.5) to (2.7) results in the following relationship between PTT and BP:

$$\text{BP} = \frac{1}{a}(-2 \ln(\text{PTT}) + \ln(\frac{\rho L_a^2 D}{h E_0})) \quad (2.8)$$

Assuming that the values of a , ρ , and the ratio of h to D remain constant, and considering small variations in PTT within the physiological range of BP, it has been demonstrated in previous studies [141, 31] that a Taylor series expansion of equation (2.8), which considers only the linear component, is suitable for estimating BP. This results in two alternative models for estimating BP ($\hat{\text{BP}}$) that have been extensively employed in the literature, both of which are reported to have similar levels of accuracy [31]:

$$\hat{\text{BP}} = K_1^- \cdot \text{PTT} + K_2 \quad (2.9)$$

$$\hat{\text{BP}} = K_1^+ \cdot \frac{1}{\text{PTT}} + K_2 \quad (2.10)$$

where K_i $i = \{1, 2\}$ are subject-specific parameters with K_1^- being negative and K_1^+ being positive. Equation (2.9) is largely utilised throughout this thesis. In general, when BP rises the arteries become distended and their stiffness increases, causing the pulse wave to travel faster and consequently reducing the PTT. Similarly, a fall in BP should lead to an increase in PTT.

2.8.1.4 Proximal and distal waveforms

The measurement of PTT requires proximal and distal waveforms to indicate the timings of the pulse wave. Various signals have been explored in the literature, for example: arterial blood pressure (ABP) [142], ballistocardiogram (BCG) [143], ultrasound [31], impedance cardiogram (ICG) [144] and video plethysmography (iPPG) [145].

Typically, the PPG is employed as the distal waveform. The timing of the arrival of the PPG pulse is often recorded using the waveform foot as the characteristic fiducial point. This can be detected using the intersecting tangents method [31]. The PPG waveform is often recorded using a pulse oximeter located on the index finger, or more recently, using a PPG sensor in a smartwatch located on the wrist [146]. However, PTT measured with this distal waveform can be confounded by smooth muscle contraction.

The most commonly used proximal waveform for PTT estimation is the ECG [31]. The ECG waveform is easy to measure and relatively robust to movement artefacts. The ECG also contains several high-frequency fiducial points, notably the R-wave [147] (see figure 2.9(b)), which is easily detected.

Most researchers, when they measure the time difference between the peak of the R-wave in the ECG signal and a fiducial point in the photoplethysmogram (PPG) waveform (see figure 2.11(b)), describe this erroneously as the PTT. That is in fact the PAT.

2.8.1.5 Pulse arrival time and pre-ejection period

PAT includes not only the PTT, but also the time delay between electrical depolarisation of the heart's left ventricle and the subsequent ejection of blood through the aorta [148], a time period known as the PEP:

$$\text{PAT} = \text{PTT} + \text{PEP} \quad (2.11)$$

PEP comprises two key components: the temporal delay between electrical depolarisation and the onset of ventricular compression (a period known as ventricular electro-mechanical delay); and the duration of ventricular isovolumic contraction [149]. Notably, PEP can exhibit significant variations, which depend on multiple factors, including myocardial contractility, sympathetic nervous system activity, preload, and afterload [150]. Moreover, several external factors, such as exercise, medication, and alterations in posture, have been shown to play a significant role in altering PEP [151, 149, 152, 153].

To measure PEP, the time of aortic valve opening can be determined via an invasive arterial line [152], the echocardiogram [154], ballistocardiogram [143], or impedance cardiogram [151] (see

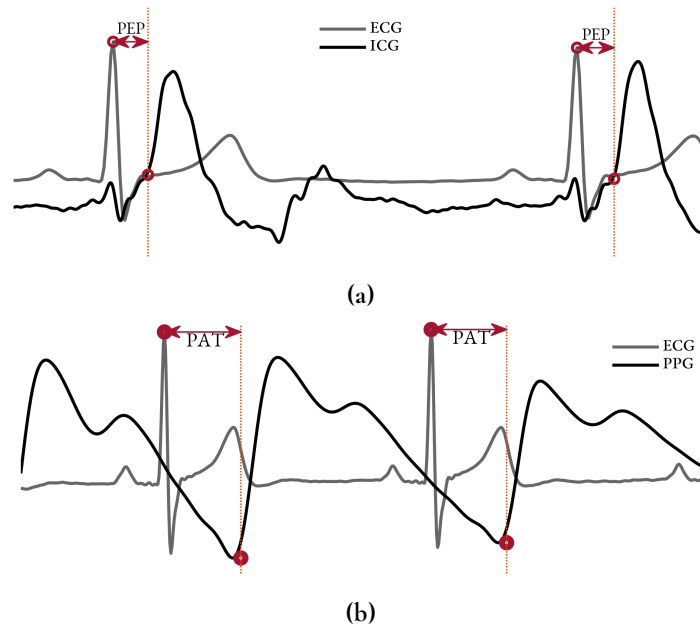


Figure 2.11: Computing PEP and PAT. (a) Computing PEP as the time delay between the R-peak of the ECG and the B-point on the ICG. (b) Computing PAT as the time delay between the R-peak of the ECG and the foot of the PPG.

section 2.7). An example of PEP estimation from the ECG and ICG waveforms and an example of PAT estimation from the ECG and PPG waveforms are shown in figure 2.11.

In this thesis, we will collectively refer to PTT and PAT as transit time (TT).

2.8.1.6 Calibrating TT measurements to BP

Estimating BP (in units of mmHg) using TT (in units of seconds) requires parameterising an estimation model through calibration. In the literature, calibration strategies are categorised into three groups: *individual*, *hybrid*, and *population* [9]. In *individual calibration*, all model parameters are determined by multiple paired recordings of BP and TT from a single individual. Techniques such as cold pressor test, medication, or exercise are used to perturb BP and model TT through a range of BP values. In *hybrid calibration* (also called single-point calibration), only one calibration BP datapoint is required for a single individual. The remaining model parameters are estimated using the individual's demographics and a training set comprised of multiple BP-TT pairs from a cohort of other individuals. Previous studies have reported demographic-related dependencies on the calibration parameters [155, 31]. The gradient of the TT to BP relationship reflects arterial properties such as compliance and resistance [31], where shallow gradients are associated with young, healthy individuals and steeper gradients with more elderly individuals [156]. In *population calibration*, no calibration recordings are required, and BP is estimated using the TT estimate combined with the individual's demographics and a training set acquired from a large population cohort. *Individual*, *hybrid*, and *population* calibration strategies offer increasing convenience with decreasing accuracy.

Population-based models Various *hybrid calibration* models have been developed based on population averages that do not require an additional dataset of BP and TT pairs. These models use a single-point calibration value, typically obtained at rest, for both BP and TT (BP^{calib} and TT^{calib} respectively). Three highly cited models that follow this approach are outlined below.

Poon *et al.* [157] derived the following model from a combination of the Moens-Korteweg equation [158] and a model relating the elastic modulus of a vessel to the mean pressure of the fluid inside it [159].

$$\hat{DBP} = \frac{SBP^{\text{calib}}}{3} + 2\frac{DBP^{\text{calib}}}{3} + \frac{2}{\gamma} \ln\left(\frac{TT^{\text{calib}}}{TT}\right) - \frac{SBP^{\text{calib}} - DBP^{\text{calib}}}{3} \left(\frac{TT^{\text{calib}^2}}{TT^2}\right) \quad (2.12)$$

$$\hat{SBP} = \hat{DBP} + (SBP^{\text{calib}} - DBP^{\text{calib}}) \left(\frac{TT^{\text{calib}^2}}{TT^2}\right) \quad (2.13)$$

$$\hat{MAP} = \frac{2}{3}\hat{DBP} + \frac{1}{3}\hat{SBP} \quad (2.14)$$

where $\gamma = 0.031 \text{ mmHg}^{-1}$ and represents a vascular parameter relating the distending pressure in an artery to the elasticity of the arterial wall [159].

The Gesche *et al.* [160] model parameters were estimated by least squares regression on a group of 13 volunteers who underwent a bicycle test. The model was then evaluated on a group of 50 volunteers and reported a mean Pearson's correlation coefficient (ρ_p) between reference and estimated BP of 0.83.

$$PWV = \frac{0.5 \cdot h}{TT}, \quad PWV^{\text{calib}} = \frac{0.5 \cdot h}{TT^{\text{calib}}} \quad (2.15)$$

$$BP_{\text{ref}} = 700 \cdot PWV^{\text{calib}} \cdot \exp(-PWV^{\text{calib}}) + 766,000 \cdot PWV^{\text{calib}^9} \quad (2.16)$$

$$\hat{BP} = 700 \cdot PWV \cdot \exp(-PWV) + 766,000 \cdot PWV^9 - (BP_{\text{ref}} - BP^{\text{calib}}) \quad (2.17)$$

where h is the volunteer's height in meters.

The Fung *et al.* [161] model was derived from a model of laminar blood flow through a rigid pipe.

$$\hat{BP} = \frac{(0.6 \cdot h)^2 \cdot \rho}{1.4 \cdot TT^2} + (BP^{\text{calib}} - \frac{(0.6 \cdot h)^2 \cdot \rho}{1.4 \cdot TT^{\text{calib}^2}}) \quad (2.18)$$

where $\rho = 1035 \text{ kg/m}^3$ and represents the average density of blood.

Recalibration Over time, various changes in the arterial geometric properties such as the progression of arteriosclerosis, atherosclerosis, vessel thickness, arterial diameter, and blood density, can affect the arterial wall structure. Consequently, an individual's calibration parameters become increasingly more obsolete, leading to the accuracy of BP estimation through TT estimates degrading over time [136]. The severity of this degradation remains to be established, however,

it is mitigated by frequent recalibrations. Theoretical predictions suggest that the maximum recalibration period to keep the BP error within 1 mmHg is no more than one year for a 30-year-old. This period theoretically declines linearly to 6 months for a 70-year-old [156], due to the faster recruitment of collagen fibres in the elderly.

Several studies have investigated the need for recalibration *in vivo*. For instance, Wong *et al.* [162] and Yavarimanesh *et al.* [163] examined the requirement for recalibration after six months and one year, respectively. Wong *et al.* recorded PAT and BP measurements during an exercise test in 14 normotensive subjects and repeated the study six months later. The differences in calibration parameters (slope and intercept) computed six months apart were found to be non-significant. However, notable increases in estimation errors were recorded, highlighting the sensitivity to slight changes in calibration parameters. Yavarimanesh *et al.* on the other hand recorded PAT and BP measurements during a multiple perturbation study (slow breathing, mental arithmetic, cold pressor test, and nitroglycerin) in 10 subjects, three of whom were hypertensive. The study was repeated one year later, and using the same calibration coefficients as the initial test increased the root mean squared error (RMSE) value by more than 5 mmHg.

2.8.2 Review of literature

In 1974, Jernstedt and Newcomer [164] developed the first device for measuring PAT. They recorded the time between the R-wave of the ECG and the foot of the PPG recorded using a pulse oximeter on the finger. Together with a measured distance from the heart to the finger, PAT was then used to estimate PWV (equation (2.7)). BP changes, measured using a sphygmomanometer cuff with the auscultatory method, were induced in two healthy volunteers by exercise (running). The two subjects recorded 11 and 8 data points of paired SBP and PWV values with corresponding ρ_p of 0.97 and 0.92 respectively.

Following from this work, many experimental studies have been conducted to assess the accuracy of BP monitoring via TT. A summary of key studies in this field is provided in table 2.5. These studies are diverse in terms of the methods and signals used for estimating TT, calibration strategies, and procedures for evaluation. Therefore, it is difficult to aggregate the studies to compare the overall accuracy of the approaches as metrics such as RMSE and ρ_p depend on the investigated BP range. For example, DBP has a lower natural variability than SBP [165], which might contribute to DBP having both a lower correlation and error. Additionally, different researchers employed different calibration strategies such as *individual* or *hybrid* calibration in order to convert TT to BP. Appropriate *individual calibration* should always yield better performance metrics than *hybrid calibration*. Mukkamala *et al.* [31] reviewed a large number of studies using TT to estimate BP. They

report that the unweighted average (across all reviewed studies) ρ_p between BP estimates using TT and reference BP was 0.79 ± 0.05 for SBP and 0.58 ± 0.09 for DBP.

There are only a handful of papers that have compared BP estimates via both PTT and PAT. Payne *et al.* [151] studied changes in PAT and PEP across 12 healthy volunteers under an infusion of various medications, including: glyceryl trinitrate (GTN), angiotensin, norepinephrine, and salbutamol. Full details of PAT and PEP computation were not provided. However, the authors concluded that changes in PEP contributed a significant factor (12% - 35%) to changes in PAT. These changes were independent of changes in BP suggesting that PAT may not be an appropriate surrogate for PTT. In 2011, Zhang *et al.* [150] measured both PAT and PTT in 6 anaesthetised dogs who were given an infusion of different vasoactive medications including (but not limited to) phenylephrine, nitroglycerin, and norepinephrine. PTT was measured using two invasive arterial lines located in the aorta and femoral artery respectively. PAT was measured using the ECG and the BP waveform from a femoral arterial line. PEP was estimated as the difference between PAT and PTT. Vasoconstriction caused PTT to decrease due to BP increase and PEP to increase due to the afterload increase. These variations occurred in opposite directions and blunted the change in PAT. Thus the RMSE value for PAT was approximately twice as large as for PTT. However, the extent to which this effect can translate to humans was unclear. Kim *et al.* [143] recorded PAT and PTT estimates in 22 healthy individuals during three haemodynamic interventions (Mental arithmetic, cold pressor test, and stair climbing) lasting 8 mins with 6 BP measurements recorded using an oscillometric sphygmomanometer. The time of pulse wave ejection from the aorta was estimated via the ballistocardiogram (BCG), which records the ballistic forces on the heart caused by the sudden ejection of blood. Similar to Zhang *et al.* Kim *et al.* found that BP estimation using PAT resulted in RMSE values approximately twice those obtained when using PTT. Conversely, Block *et al.* [166] reported stronger performances using PAT estimated with a pulse oximeter on the toe as opposed to PTT estimates. However, in this work PTT was computed as the time delay between two PPG sensors located on the ear and the toe respectively. This should not be considered a reliable measurement of PTT as blood does not flow directly from ear to toe.

A large proportion of the papers listed in table 2.5 used data recorded in the hospital, including the ICU. These databases benefit from the high-frequency and high-quality vital-sign monitoring recorded from patients in those settings. The hospital, additionally, is one of the few environments in which a patient will have their BP measured using an arterial line (the gold standard in BP monitoring). This also allows for the assessment of TT as a beat-by-beat predictor of BP. However, the data is often analysed in a *post-hoc* manner, with a lack of understanding of the patient's physiological status at the time of measurement. Hence, studies based on the analysis of these

datasets often report comparatively poor performance. Zhang *et al.* [167] extracted 2 hours of data from 20 patients in the MIMIC-II intensive care unit (ICU) dataset [168]. The authors developed a signal quality index (SQI) for PAT estimation in order to automatically detect periods of poor signal quality caused by, for example, motion artefacts. A Kalman filter [169] was then implemented to provide reliable estimates of PAT values during periods of poor signal quality. *Individual calibration* models were estimated using the first 20 beats. BP was estimated for the remaining 2 hours of data. Including the signal quality and Kalman filter steps significantly improved performance compared to previous work. The mean absolute error (MAE) was 7.45 ± 9.79 mmHg for SBP and 4.41 ± 5.85 mmHg for DBP. Lee *et al.* [170] analysed the PAT-BP relationship across a large number ($N = 2,309$) of patients who underwent surgery. ECG, PPG, and arterial line BP were recorded for over 30 minutes for each patient. Using *hybrid calibration* to predict BP resulted in poor performance metrics for SBP prediction (mean error = 0.00 ± 11.04 mmHg, $\rho_p = 0.37 \pm 0.37$). The authors used the mean of the BP and PAT values across the entire recording as the single-point calibration value for each individual, hence a mean error of zero. Performance metrics (specifically ρ_p) improved for individuals who experienced larger changes in BP, showing that accurate calibration required large changes in BP. Studies for which BP was not sufficiently perturbed invariably reported poor performance [171]. Lee [170] was also able to demonstrate significant ($p < 0.001$) differences in calibration gradients as an individual ages, with no reduction in performance. These results suggest that frequent re-calibration, over time, may be sufficient for long-term tracking of BP. Yoon *et al.* used 23 patients in an ICU database to show that including HR as an additional term in the BP estimation model can significantly improve results. This is most likely driven by the baroreflex response and the relationship between short-term changes in HR and BP.

Circadian variations in TT have not been explored in great detail in the literature, mostly due to the complications involved in simultaneous recordings of 24-hour BP, ECG, and PPG. Zheng *et al.* [172] developed an armband wearable capable of recording simultaneous ECG and PPG on the same internal clock for 24 hours. ECG was measured by two e-textile electrodes sewn into the armband and another electrode placed on the right side of the subject's thorax. A PPG signal was recorded in reflectance mode by a near-infrared LED emitter and photodiode on the inside of the armband. While this set-up prioritises convenience for the user, it was reported to result in reduced performance metrics over conventional PPG measured by a pulse oximeter on the finger [171]. The armband was wrapped around the upper left arm and was set to record ECG and PPG signals for 1-minute intervals every 30 minutes. BP was recorded by an ABPM wrapped around the upper right arm and was set to measure BP every 30 minutes during both the day and night. BP was estimated from PAT using a *hybrid calibration* strategy [157] with the single point calibration value

being the first BP value recorded. The study involved 10 individuals who were asked to go about their normal life for 24 hours and thus natural, 24-hour variations in BP were recorded. However, the authors did not compute circadian metrics such as night-time dip. It was reported that PAT correlated significantly more strongly with SBP during the night (median $\rho_p = -0.80$) as opposed to during the day (median $\rho_p = -0.24$). This is most likely a combined effect of the largest natural change of BP occurring during the night as well as it being a period of limited motion artefacts.

Ganti *et al.* [146] developed a wearable device, similar to HBPM, in which the user manually records PTT measured between the BCG signal and the reflectance-mode PPG waveform recorded on the wrist. However, the BCG in particular was very susceptible to motion artefacts [173], making the device's ability to estimate ambulatory BP limited.

More comprehensive literature reviews of BP estimation using TT can be found in [31, 174, 175].

Table 2.5: Summary of key studies of BP estimation using PTT or PAT. 'BP ref' indicates the reference BP device used. '# calib points' indicates the number of calibration datapoints used to estimate the model parameters for each subject. In cases where multiple results are reported, the best-performing one is listed.

Reference	# of subjects and type	BP intervention	BP ref	Waveforms	Model	# calib points	Results*			
							SBP		DBP	
							Error	ρ_p	Error	ρ_p
Sharwood-Smith <i>et al.</i> (2006) [176]	58 NTs	Spinal anesthesia (variable length)	Cuff	ECG $\rightarrow$ toe PPG	NA	NA	-	-0.74	-	-
	15 HTs			-	-0.67	-	-			
Zhang <i>et al.</i> (2011) [150]	6 dogs under anesthetic	Vasoactive medication	ABP	Aortic ABP $\rightarrow$ femoral ABP	$BP = K_1/PTT + K_2$	All	$7.5 \pm 2.2^\dagger$	-	$5.3 \pm 1.2^\dagger$	-
				ECG $\rightarrow$ ABP	$BP = K_1/PAT + K_2$	All	$11.9 \pm 6.1^\dagger$	-	$9.8 \pm 5.2^\dagger$	-
Zheng <i>et al.</i> (2014) [172]	10 HVs	Natural / environmental changes (24 hours)	ABPM	ECG $\rightarrow$ PPG (armband wearable)	Poon <i>et al.</i> [157]	1	2.4 ± 5.7	-0.8 ± 0.1	-	-
Zhang <i>et al.</i> (2017) [167]	20 ICU patients	Clinical (2 hours)	ABP	ECG $\rightarrow$ finger PPG	$BP = K_1 \ln(PAT) + K_2$	20	0.092 ± 5.05	-	-0.15 ± 3.5	-
Yoon <i>et al.</i> (2018) [177]	23 ICU patients	Clinical (> 24 hours)	ABP	ECG $\rightarrow$ finger PPG	$BP = K_1 PAT + K_2 HR + K_3$	12 hours of data	9.8 ± 2.8	0.64 ± 0.18	5.2 ± 2.1	-
Kim <i>et al.</i> (2018) [143]	22 HVs	Mental arithmetic, cold pressor, stair climbing (8 mins)	Volume clamp	BCG $\rightarrow$ toe PPG	$BP = K_1/PTT + K_2$	All	-	-	$5.7 \pm 0.4^\dagger$	-
				ECG $\rightarrow$ finger PPG	$BP = K_1/PAT + K_2$	All	-	-	$10 \pm 0.4^\dagger$	-
Rajala <i>et al.</i> (2018) [171]	30 HVs	No perturbation. 3 BP taken over 10 mins	Cuff	ECG $\rightarrow$ finger PPG	$BP = K_1/PAT + K_2$	All	-	0.44	-	-
				ECG $\rightarrow$ wrist PPG	$BP = K_1/PAT + K_2$	All	-	0.37	-	-
Liang <i>et al.</i> (2019) [178]	121 ICU patients	Clinical (variable length)	ABP	ECG $\rightarrow$ finger PPG	NA	NA	-	-0.30 ± 0.25	-	-
Lee <i>et al.</i> (2019) [170]	2,309 surgery patients	Surgery (variable length, > 30 mins)	ABP	ECG $\rightarrow$ finger PPG	$BP = K_1 PAT + K_2$	1	0 ± 11.04	0.37 ± 0.37	0 ± 6.25	0.30 ± 0.38
Block <i>et al.</i> (2020) [166]	32 HVs	Slow breathing, mental arithmetic, cold pressor, vasoactive medication (28 mins)	Cuff	Ear PPG $\rightarrow$ finger PPG	NA	NA	-	-0.29	-	-0.22
				ECG $\rightarrow$ finger PPG	-	-0.42	-	-0.28		
				ECG $\rightarrow$ toe PPG	-	-0.63	-	-0.31		
Ganti <i>et al.</i> (2020) [146]	21 HVs	Natural / environmental changes (24 hours)	HBPM	BCG $\rightarrow$ wrist PPG	$BP = K_1/PTT + K_2$	All	$4.75 \pm 2.29^\dagger$	0.33	$2.72 \pm 0.75^\dagger$	0.69
						2	-	-	$3.86 \pm 1.53^\dagger$	-

Acronyms: NTs - Normotensives; HTs - Hypertensives; ICU - Intensive care unit; HVs - Healthy volunteers. Cuff - BP recorded by an oscillometric sphygmomanometer cuff. Poon refers to the hybrid calibration model detailed in [157] * Error values were reported differently in each study, $\dagger$ indicates errors given as RMSE, otherwise, error values are given as mean $\pm$ standard deviation, all in units of mmHg.

2.8.3 Limitations

Continuous estimation of BP via transit times (both PTT and PAT) has several limitations. Since TT measurements are derived from two independently recorded signals, they are especially vulnerable to various forms of noise, such as motion artefacts. For instance, Foo *et al.* [179] showed that motion artefacts caused by hand movements can impact TT measurements by as much as 25 ms. While appropriate signal quality indices (SQIs) can help mitigate the impact of motion artefacts, these steps are often overlooked in the literature. Moreover, changes in the contact pressure of the PPG sensor, which refers to the external pressure applied by the PPG sensor to the skin surface [180], can also affect the accuracy of proximal and distal timing computation. Indeed, Chandrasekhar *et al.* [181] reported that variations in the contact pressure of the sensor that records the PPG signal may result in TT measurement errors of up to 40 ms.

Even small distortions in TT values can cause significant errors in BP estimation. For example, a typical linear calibration gradient relating TT to BP, as reported in the literature, is about 500 mmHg s⁻¹ in magnitude [182, 162, 181, 151]. Based on this calibration gradient, a 25 ms error in TT estimation would lead to a BP estimation error of 12.5 mmHg. This error would be greater in the elderly population, for whom steeper calibration slopes are often observed [31]. Such an error could result in normotensive individuals being misclassified as hypertensive (see table 2.4).

It is important to also note that changes in TT reflect changes in arterial stiffness as opposed to BP. While there is a strong relationship between arterial stiffness and BP [132], TT can also reflect changes in arterial structure, sympathetic activity and disease state [136] which can act independently of BP. Furthermore, PPG signals are typically recorded on peripheral arteries and therefore are significantly impacted by smooth muscle contraction. Smooth muscle contraction significantly alters the pressure-volume relationship by introducing a viscous delay element [31]. This can cause variations in arterial stiffness which are not reflected in BP [183]. PAT, additionally, includes the PEP term which has been previously suggested as a significant limitation to cuffless BP estimation.

Another challenge is the independent estimation of both SBP and DBP at the same time. This challenge is important, as isolated systolic hypertension often occurs in the elderly (see figure 2.5). One potential solution is to include additional covariates, such as HR, heart rate variability (HRV), or a marker of neural control, in the calibration model. Another approach is to estimate multiple PTT values per beat via arterial modelling [184].

Finally, the requirement to record two signals, from different locations on the body, makes for cumbersome wearable equipment and can bring time synchronicity issues. As a result, the research landscape has shifted in the last few years towards estimating BP using either the PPG or the ECG on their own.

2.9 Estimating blood pressure from the photoplethysmogram

2.9.1 Background theory

As discussed previously, changes in the cardiovascular and autonomic nervous systems are reflected in changes in PPG waveform morphology. Therefore, it may be possible to predict BP using only the PPG signal. This is additionally validated by the noticeable similarity between the ABP and PPG waveforms (see figures 2.6 and 2.13). Martinez *et al.* [185] highlighted the strong morphological similarity between ABP and PPG, suggesting that the PPG holds key information related to BP. More recently, Tang *et al.* [186] developed a deep learning model to reconstruct ABP signals from simultaneously recorded PPG signals.

The PPG offers significant benefits for BP monitoring over both conventional cuff-based measurements and TT estimates. Most notably, the PPG can be recorded by a single, unobtrusive optical sensor on a wearable device such as a smartwatch [24, 187].

The PPG-BP relationship is driven, in part, by the theoretical relationship between changes in pressure and volume of blood in a localised region of an artery [9], as well as the impact of reflected waves [188]. The Kelvin-Voigt model of visco-elastic vessels suggests that the AC components of the PPG (ΔV) can be considered a low-pass filtered version of the AC components of the BP (ΔP) waveform [9, 189] with the following transfer function:

$$\Delta V(\omega) = \frac{1}{j\omega\eta + E} \Delta P(\omega) \quad (2.19)$$

where E and η are the elastic modulus and coefficient of viscosity of the arterial wall respectively [190, 9] and ω is the angular frequency in radians. The viscous element is introduced by the presence of smooth muscle in arteries, particularly in peripheral arteries where a pulse oximeter is often placed. The model transfer function is a low-pass filter with $1/E$ gain and E/η cut-off frequency. This low-pass filter impacts the phase delay and amplitude attenuation of the volume of blood in response to the pressure wave produced by left ventricular contraction. The magnitudes of the amplitude attenuation and phase delay are determined by the mechanical properties of the arterial wall, as well as the transmural pressure on the walls.

An additional explanation relating changes in PPG to changes in BP is the impact of reflected wave interference. The PPG waveform is the summation of a forward pressure wave and reflected pressure wave(s) [191]. The forward pressure wave is the result of left ventricular contraction and travels from the aorta to the distal point of PPG measurement in peripheral sites such as the fingertip. Reflected pressure wave(s) result at points of impedance mismatch (such as reduction in vessel diameter or bifurcations) further down the arterial tree and travel back to the point of

PPG measurement. Changes in arterial stiffness have a significant impact on the timing, amplitude and morphology of the reflected pressure waves [188].

There currently is no consensus in the literature on the number of reflected pressure waves that affect the PPG waveform. Millaseu *et al.* [192, 77] concluded that the shape of the PPG waveform results from the interference of one backwards travelling reflected wave, occurring during diastole and is caused by impedance mismatch at the bifurcation of the iliac arteries (located in the pelvic region providing blood supply to the legs, see figure 2.12). Increases in arterial stiffness results in an earlier return of the reflected wave (and vice versa) which significantly impacts the morphology of the PPG. Rubins *et al.* [193] used Gaussian decomposition of the PPG to propose that there are three reflected waves, all occurring during diastole, although the authors provided no physiological explanation for these.

Baruch *et al.* [188] proposed that there are two main reflected waves from two distinct points of significant impedance mismatch, with the addition of minor reflected waves that do not have a notable impact on the PPG waveform. These two reflected waves make up a late systole component (also known as the tidal wave [194]) caused by reflections at the renal arteries, and a diastole component (also known as the dicrotic wave) caused by reflection at the bifurcation of the iliac arteries. A visualisation of the interactions between the incident and reflected pressure waves, and their impact on the PPG waveform, is provided in figure 2.12. The tidal wave and dicrotic wave follow the incident wave at a delay of between 70-140 ms and 180-400 ms respectively [188]. Specific arterial properties of these two sites of reflection result in a strong relationship between changes in arterial stiffness with the amplitude and timing of the reflected pressure waves. Therefore, analysis of the morphology of the PPG waveform, with a specific focus on the effect of reflected pressure waves, might provide insight into changes in arterial stiffness.

Typically, BP estimation using the PPG is performed by mapping characteristic morphological features from the PPG to BP values using data-driven machine-learning models. We refer to these features as PPG features (PPG_{fts}) and they are often derived from the PPG morphology using key fiducial points. In the literature, the nomenclature used for these fiducial points varies considerably. In this thesis, we follow the nomenclature of Elgendi *et al.* [196]. The first and second derivatives of the PPG are known as the velocity plethysmogram (VPG) and acceleration plethysmogram (APG) respectively. Figure 2.13 provides an example of the fiducial points located on the PPG (onset O , systolic peak S , dicrotic notch N and diastolic peak D), VPG (maximum W), and APG (a , b , c , d , e , and f waves). The projection of a point onto a n^{th} higher or lower derivative is given a subscript $+n$ or $-n$. For example, the projection of W onto the PPG is labelled W_{-1} , and the projection of S onto the APG is labelled S_{+2} .

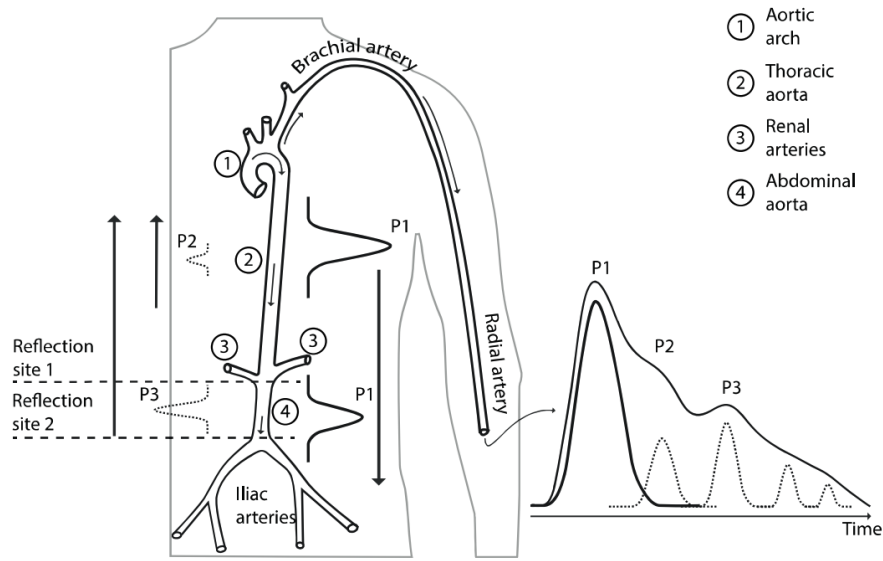


Figure 2.12: The morphology of the PPG waveform, measured at the wrist or fingertip along the brachial artery, is influenced by the interactions of the incident and reflected pressure waves. The incident wave is generated by left ventricular contraction and travels along the brachial and thoracic arteries. As the wave reaches the junction of the renal or iliac arteries, a portion of it is reflected back up the thoracic artery, eventually travelling along the brachial artery to interact with the incident wave at the measurement site. Although there are additional minor reflected waves, their impact on the waveform shape is negligible. Figure from [195] with permission.

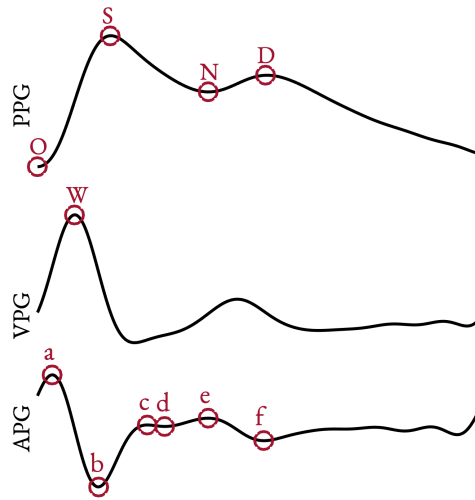


Figure 2.13: Example of PPG, VPG, and APG fiducial points. The following fiducial points are present on the PPG: onset *O*, systolic peak *S*, dicrotic notch *N* and diastolic peak *D*. The maximum of the VPG is labelled *W*. The APG is characterised by the *a*, *b*, *c*, *d*, *e*, and *f* waves.

A wide range of PPG_{fts} have been reported in the literature. Examples include: amplitude variations [197, 198, 199], time delays [200, 201, 202], area under the PPG contour [203, 204], Gaussian decomposition [205, 206], derivatives [201, 79], and features in the frequency domain [207, 208].

PPG_{fts} values are mapped to BP in the same manner as TT. A data-driven model is first defined to relate the feature values to BP and then an appropriate calibration strategy is implemented to fit

these models to the individual. Much like TT, features of the PPG are thought to be subject-specific with different relationships to BP for different individuals. As a result, these models require calibration. As the parametric models used to relate PPG_{fts} to BP are often multidimensional, *individual calibration* is often not feasible without long-term, longitudinal, individual tracking of both BP and PPG. *Population calibration* uses an individual's demographics to calibrate the model. Demographic-related dependencies on the morphology of the PPG waveform have been previously reported for age [192], sex [155] and body mass index (BMI) [209]. However, these are often not strong enough to allow for good accuracy of *population calibration* strategies. As a result, the majority of studies proposing PPG-based BP estimation opt for a *hybrid calibration* strategy.

Miao [210] used support vector regression on 14 PPG_{fts} from 73 subjects to track BP changes induced by physical exercise (rope skipping for 5 minutes). The stability of the developed models was evaluated in a follow-up study for 1 day, 10 days and 6 months after the initial test. The results suggested that, similar to that found for TT [162, 156], the estimation models lose their accuracy over time and may, therefore, require frequent recalibration.

2.9.2 Review of literature

A summary of key studies in this field is provided in table 2.6. As was the case with transit times, the heterogeneous nature of publications in this field makes the comparison of results between studies difficult. A significant proportion of studies use large open-source datasets that are typically obtained from an ICU [211, 212, 213, 214]. These large databases allow for improved accuracy of machine learning models via increased dataset sizes, which may further assist in the development of new hybrid calibration strategies. However, these models have poor interpretability, driven by a lack of understanding of the patient's physiological status at the time of measurement. Additionally, hospital patients are often under a high level of stress, which significantly alters their physiology compared to healthy individuals. Clinical studies which are able to offer high interpretability despite the expense of poor model performance are still sparse in this field [215, 187, 190].

Ruiz-Rodriguez *et al.* [211] in 2013 was the first to estimate continuous changes in BP using the PPG. The authors conducted a study using data from 572 patients in an intensive care setting, with an invasive arterial line as the reference signal. A deep neural network was trained on the data from 525 patients and tested on the data from the remaining 47 patients. Personalisation of the model was handled only by demographics as opposed to calibration. The PPG_{fts} used to train the model were undisclosed. Estimated SBP values achieved an error of -2.98 ± 19.35 mmHg. The large variance of error highlights the importance of appropriate model calibration to the individual.

Sun *et al.* [215] evaluated the use of a linear regression model for estimating SBP measured by a commercial Finapres (see section 2.4.4), in 19 subjects undergoing an exercise test. 19 morphological PPG_{fts} were extracted from each PPG pulse. PPG signal quality was assessed by a combination of visual inspection and an algorithm that highlighted sharp changes in peaks and the width of the PPG pulses. An RMSE of 10.44 mmHg was reported when estimating SBP. Normalised weights of the linear model were used to highlight features that had the most predictive power. However, the study was only conducted on 19 volunteers and BP changes were induced by exercise, which would introduce a significant amount of motion artefacts.

Hasanzadeh *et al.* [212] extracted PPG_{fts} in a subset of 1,000 individuals from the MIMIC-III dataset. The arterial line blood pressure waveform was used as a reference. The authors implemented a linear regression model as well as a non-linear tree-based model, AdaBoost. AdaBoost significantly outperformed the linear model for their dataset. Additionally, the authors highlighted the sensitivity of PPG_{fts} extraction in the presence of random noise.

Xing *et al.* performed two studies [216]. The first study estimated continuous BP for patients in an ICU with ABP and PPG signals. 85 patients were used in total. The amplitudes and phases from the FFT of each PPG beat were fed into a multi-layer perceptron with 36 hidden units to estimate SBP and DBP. The second study evaluated the trained network on an additional set of 23 volunteers outside of a hospital setting. In this study, the PPG sensor was located on the finger and a BP reading was taken using a sphygmomanometer. An 8-second segment of the PPG signal, centred on the BP reading, was extracted. The FFT features for the average beat were used as inputs to the neural network to be compared with the cuff measurements. The first study achieved an error of 0.06 ± 5.57 mmHg for SBP and 0.01 ± 3.69 mmHg for DBP. The second study achieved an error of -1.67 ± 2.46 mmHg for SBP and -1.29 ± 1.71 mmHg for DBP.

Radha *et al.* [187] investigated PPG-based ambulatory BP estimation. The authors ran a study with 106 healthy individuals with the PPG signal being recorded using a smartwatch on their wrist and an ABPM cuff programmed to inflate every 30 minutes both day and night. An accelerometer was used to determine periods of rest. Each participant took part in the study up to 3 times. The final cohort consisted of 226 days of simultaneous PPG and ABPM. PPG_{fts} based on heart rate variability, pulse morphology, Gaussian decomposition analysis, and frequency domain analysis were used. All data from 82 randomly selected participants was used for model training and validation. Data from the remaining participants were reserved for testing. Using a long short-term memory (LSTM) neural network, the authors reported an SBP estimation error of 8.22 ± 1.49 mmHg and a ρ_p of 0.69 between reference and estimated DR_{SBP}. However, Radha *et al.* z-normalised all signals across each 24-hour recording prior to estimation, thus circumventing any calibration requirements.

The Aurora project, conducted by a team at Microsoft [217], involved a similar study with a larger cohort of participants ($N = 483$). They used standard linear regression with a *hybrid calibration* approach to estimate 24-hour BP using PPG_{fts} obtained from a wrist-worn sensor. The performance of this method was compared with a *baseline reference* model that only used the initial calibration value and the time of day to predict BP. Surprisingly, the *baseline reference* model outperformed the waveform-based model, although better performance was observed in younger participants.

In general, there has been some potential demonstrated for predicting BP using PPG_{fts} . However, there is a lack of consensus regarding the best features and the best models to use, as well as its efficacy in ambulatory settings. This lack of understanding contributes to the challenge of generalisability within the field. Recent developments in the field have explored the application of deep learning algorithms [213, 214]. However, while these approaches may improve the estimation accuracy, their black box nature mean that they lose model interpretability.

More comprehensive literature reviews of BP estimation using the PPG can be found in [218, 219, 9].

Table 2.6: Summary of key studies of BP estimation using the PPG. 'BP ref' indicates the reference BP device used. In cases where multiple results are reported, the best-performing one is listed.

Reference	# of subjects and type	BP intervention	BP ref	Waveforms	PPG _{fts} used	Model & calibration	Results*			
							SBP Error	ρ_p	DBP Error	ρ_p
Ruiz-Rodriguez <i>et al.</i> (2013) [211]	572 / Critically ill (47 for testing)	Clinical (30 mins)	ABP	Finger PPG	Various waveform & demographic features	Deep neural network / <i>population calibration</i>	-9.01 ± 14.55	-	5.22 ± 9.55	-
Sun <i>et al.</i> (2016) [215]	19 / HVs (LOSOCV)	Exercise and posture changes (40 mins)	Finapres	Finger PPG	Various waveform features	Linear regression / Single cuff <i>hybrid calibration</i>	$8.99^\dagger$	0.85	-	-
Hasanzadeh <i>et al.</i> (2019) [212]	1000 / Clinically ill (10 fold CV)	Clinical (variable time frame)	ABP	Finger PPG	Various waveform features	Adaboost / <i>population calibration</i>	0.09 ± 10.38	0.78	0.23 ± 4.22	0.72
Slapnicar <i>et al.</i> (2019) [213]	510 / Critically ill (LOSOCV)	Clinical (1.3 h)	ABP	Finger PPG	Whole PPG; VPG and APG signals	ResNet / <i>Hybrid calibration</i>	$9.43^{\dagger\dagger}$	-	$6.88^{\dagger\dagger}$	-
Radha <i>et al.</i> (2019) [187]	103 / HVs (21 for testing)	Natural (nocturnal dip)	ABPM	Wrist PPG + accelerometer	Various waveform features + HRV + activity	Random forest / individual z-normalised features	$7.86^\dagger$	0.69	$6.49^\dagger$	0.77
Luo <i>et al.</i> (2019) [220]	1328 / HVs (200 for testing)	No perturbation in BP	CNAP	iPPG on face	Various waveform + demographic features	Multilayer perceptron / <i>population calibration</i>	0.39 ± 7.30	0.67	-0.20 ± 6.00	0.47
Liu <i>et al.</i> (2020) [221]	35 / Clinically stable patients (60:20:20 tr:v:te per patient)	Clinical; during arrhythmias (10 mins)	ABP	Finger PPG	Small number of features + PAT	Random forest / <i>Individual calibration</i>	-0.04 ± 6.11	-	0.11 ± 3.62	-
Schlesinger <i>et al.</i> (2020) [214]	329 / Critically ill (60:20:20 tr:v:te)	Clinical (30 s)	ABP	Finger PPG	Spectrogram of PPG segments	CNN / Siamese network <i>individual calibration</i>	MAE: 0.49; STD: 11.27	-	MAE: 0.47; STD: 5.94	-
Natarajan <i>et al.</i> (2021) [190]	32 / HVs (LOSOCV)	Slow breathing; mental arithmetic; cold pressor and medication (28 mins)	Cuff	Finger PPG	PAT + b-time	Linear regression / Single cuff <i>individual calibration</i>	$9.50^\dagger$	0.45	$7.00^\dagger$	-

Acronyms: HVs - Healthy volunteers; CV - cross-validation; LOSOCV - Leave-One-Subject-Out cross-validation; tr:v:te - training:validation:test split ratio; ABP - Arterial blood pressure; CNAP - CNAP Monitor 500 (a finger cuff-PPG) [191]; iPPG - imaging PPG from camera; HRV - Heart rate variability; CNN - convolutional neural network; STE - Standard deviation of error. * Error values were reported differently in each study, $\dagger$ indicates errors given as RMSE, $\dagger\dagger$ indicates errors given as MAE, otherwise, error values are given as mean $\pm$ standard deviation, all in units of mmHg.

2.9.3 Limitations

Estimation of BP from the PPG signal has a number of limitations. The location of the sensor (finger, wrist, ear etc), contact pressure, and environmental factors (ambient light absorption, motion artefacts) have all been shown to significantly alter the shape of the PPG waveform independently of BP [114]. Additionally, as outlined previously the PPG waveform may be impacted by the user demographics such as ethnicity.

When considering a PPG sensor located on a finger, something as simple as a change from the right to the left hand can have a major impact on the accuracy of BP estimation. Allen *et al.* [28] outlined the difference in the shape of the PPG measured at different sites (ears, fingers, toes) and on different sides (left and right), showing significant changes in the morphology at each site. The contact pressure of the PPG sensor markedly impacts the amplitude and shape of the PPG waveform [114, 222]. This is due to the non-linear arterial blood volume-transmural pressure relationship (see equation (2.19)). Insufficient or excessive pressure can both result in low AC signal amplitude due to inadequate contact or occlusion of the artery respectively. For example, features from the APG have been shown to be inconsistent when recorded at non-optimal probe contact pressures [114]. Block *et al.* showed that the PPG amplitude should never be used as an indicator of BP for this reason. Additionally, there are many external factors unregulated by BP that can impact the PPG signal. Conditions such as local hypothermia can result in peripheral vasoconstriction which can influence the PPG trace without a change in central BP [118].

Many PPG_{fts} rely on the accurate detection of salient fiducial points on the PPG, VPG or APG waveforms. There are a number of algorithms proposing accurate detection of these points [223, 224, 225]. However, their performance is dependent on the level of noise in the signal. With the PPG waveform, in particular, being susceptible to motion artefacts [118], the potential for BP estimation using the PPG in a free-living ambulatory setting remains unclear. In addition, many features use the dicrotic notch to define the separation between systole and diastole. However, as discussed in section 2.4.1, age-related atherosclerosis results in a diminished, and sometimes absent, dicrotic notch. Often in the ICU for example, the dicrotic notch is unobservable [226]. This is especially concerning given that ICU patients are a cohort analysed extensively in this field (see table 2.6).

Finally, many of the algorithms proposed for estimating BP via the PPG estimating rely heavily on machine learning algorithms without consideration of the underlying physiology. Such systems are often able to generate good results on a single test set, however, may be unable to generalise to a broader population.

2.10 Estimating blood pressure from the electrocardiogram

2.10.1 Background theory

The general theory governing the relationship between changes in BP and changes in the ECG waveform is based on a cyclical process known as mechano-electric coupling (MEC), depicted in figure 2.14. Under normal conditions, MEC plays an important role in the electrical regulation of cardiac performance in response to mechanical changes and demands [227]. Changes in the electrical properties of the heart directly affect its contractility, resulting in changes in the ECG waveform. This process is known as excitation-contraction coupling (ECC). The tissues surrounding the heart are also sensitive to mechanical changes, and these changes can be detected by mechanosensitive calcium ion channels. Acute mechanical stimuli can lead to increased stretching of pacemaker cells (Bainbridge effect) and atrial stretch receptors, which impact the electrophysiology of the heart through stretch-activated channels and mechano-sensitive channels [228, 229]. Consequently, mechanical changes in the surrounding tissues (caused by, for example, changes in aortic stiffness) can be reflected in changes in the ECG waveform. This process is known as mechano-electrical feedback (MEF).

MEC, therefore, describes the cyclical process whereby changes in the ECG waveform can reflect changes in the mechanical properties surrounding the heart. These mechanical properties will undoubtedly reflect changes in BP through changes in afterload and preload.

While MEC defines how changes in BP may be reflected in changes in the ECG, it does not characterise how these changes may impact the ECG morphology. Consequently, unlike PPG_{fts}, researchers are unable to define appropriate features to extract from the ECG based on a theoretical understanding. Typically, whole-based features that characterise the overall shape of a short ECG segment are extracted, we refer to these as ECG features (ECG_{fts}).

2.10.2 Review of literature

In comparison to the PPG, fewer studies have investigated the potential use of the ECG for BP estimation. Villecco *et al.* [231] first reported a significant correlation between high BP and ischaemic ECG patterns in 1984. It was not until 2014 that Monroy Estrada *et al.* [232] attempted to predict absolute values of BP from the ECG. The authors used a neural network with the raw ECG beat as the input vector. However, the data used to validate this approach involved only 12 volunteers (6 male, 6 female) who each only contributed one BP value during rest. Hence the reported error values should not be considered as valid as they do not account for calibration and there was no effort to perturb BP.

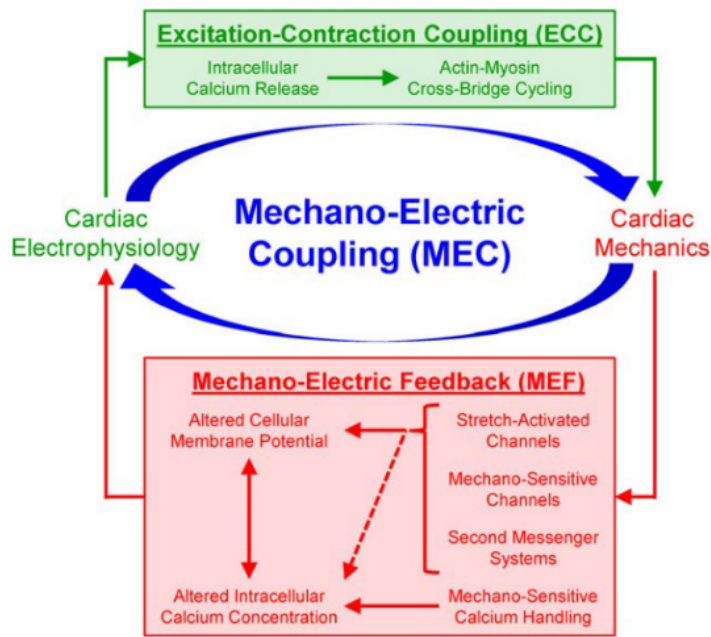


Figure 2.14: Illustration of the general theory governing the relationship between changes in BP and changes in the ECG: MEC. MEC is a cyclical process involving excitation-contraction coupling and mechano-electric feedback. Figure from [230] with permission.

Mousavi *et al.* [233] used frequency domain analysis of 15-second segments of the ECG signal from 660 patients in the MIMIC-II ICU database. The authors used a random forest to predict BP, without accounting for calibration for each individual, and the model produced large error values. MAE and standard deviation of error for SBP and DBP were 6.04 ± 6.42 mmHg and 12.75 ± 12.15 mmHg respectively.

Complexity features are the most commonly extracted ECG_{fts} for BP estimation [234, 235, 236]. These features quantify the level of regularity and unpredictability of fluctuations over a time series. Generally, a higher-level complexity indicates a more irregular dynamic system. A lower-level complexity indicates the presence of central trends or cyclical patterns. Changes in entropy of the ECG time series have also been shown to track changes in heart rate variability (HRV) caused by myocardial ischaemia [237] as well as to indicate periods of cardiac arrhythmia [238].

Simjanoska *et al.* [234] used a dataset containing data from 51 individuals from a mixture of 4 open-source datasets, three of them included data from healthy volunteers recorded using commercial ECG sensors with BP measured by a cuff, and one was from an ICU dataset with clinical ECG sensors and BP measured by an arterial line. Complexity features, such as mobility and entropy, were extracted from 30-second segments of ECG waveform data and used to train a random forest regression model using a train-validation-test split. After calibration, the model still reported poor performance with MAE of 7.72 mmHg for SBP, 9.45 mmHg for DBP, and 8.13 mmHg for MAP. Additionally, the authors only derived a small number of complexity features,

and the mixture of data sources (especially with regard to mixing healthy and ICU populations) makes interpreting the results difficult.

Yang *et al.* [236] extracted complexity features in addition to features derived from multifractal analysis from both the ECG and the PPG. BP was recorded by a sphygmomanometer cuff from 14 subjects while they conducted an exercise test to induce changes in BP. Using support vector regression, the authors suggested that complexity features were better predictors of DBP using the PPG, and SBP using the ECG. However, the sample size of the dataset was far too small for evidence-based conclusions.

To the author's knowledge, only one study has analysed the relationship between 24-hour ECG and ABPM. Yang *et al.* [121] recorded simultaneous ECG and ABPM in 177 men. A significant relationship ($p \leq 0.05$) was observed between an ECG index (the Cornell index [239]) and 24-hour BP values. For every 1 standard deviation increment in SBP, the Cornell index increased by 9.6 mVms.

2.10.3 Limitations

MEC is influenced not only by cardiac control mechanisms, such as ANS and hormonal changes, but also by environmental mechanisms, such as ion concentrations and temperature [229]. These mechanisms can occur independently of changes in BP and therefore any pressure-related information in the ECG may be overshadowed. This is evidenced by the relatively poor performance of ECG-based analysis techniques for BP estimation reported in the literature.

ECG waveform morphology can also vary notably depending on the lead being recorded. The effect of model performance when using different ECG leads has been completely overlooked in the literature. Additionally, the presence of arrhythmias, ectopic beats, and other abnormal heart rhythms could have a significant impact on model performance.

2.11 Commercially-available devices

Within the last five years, a number of companies have introduced cuffless BP measurement devices to the commercial market [24, 25]. An overview of the claims and published validation studies for three of the most prominent devices (Sotera Wireless, Samsung Galaxy Watch Active 2, and Aktiia) is provided in table 2.7. While these devices may offer promise, their validation and reliability for clinical applications are yet to be established. At the time of writing this thesis, there are currently no clinically-approved wearable cuffless BP devices. Apple, for example, has repeatedly delayed announcing that the latest version of the Apple Watch will be able to measure BP [240].

In appendix A we provide results from a small healthy volunteer study comparing 24-hour BP from a reference ABPM and estimated BP from an Aktiia device [241]. In this work, we suggest that the Aktiia SBP estimates are primarily scaled HR estimates and that the device significantly underestimates any deviations from the initial calibration value. This leads to large error values and an underestimation of the DR_{SBP} . These conclusions were recently corroborated by Tan *et al.* [26] using a larger cohort of 42 individuals all of whom wore the Aktiia device continuously for 6-12 days. At the beginning and the end of this period, 24-hour ABPM was performed. It was found that the Aktiia device significantly ($p < 0.001$) underestimated nocturnal changes in BP resulting in a significantly ($p < 0.01$) higher proportion of participants being classified as non-dippers compared to a standard ABPM. Furthermore, an additional three participants with hypertension, wore the Aktiia device and performed HBPM continuously one week before and 2 weeks after antihypertensive medication uptitration. The average medication-induced SBP/DBP decline after 2 weeks of uptitration was $-1.0/-0.8$ mmHg with Aktiia vs. $-19.7/-11.5$ mmHg with HBPM ($p = 0.03$ for difference) demonstrating that the Aktiia is additionally unable to track medication-induced BP changes.

The European Society of Hypertension Working Group on Blood Pressure Monitoring and Cardiovascular Variability recently released a statement that they currently do not recommend the use of any commercially available cuffless devices for the diagnosis and management of hypertension [27]. The authors also highlight the current inadequacies of the validation protocols that are often reported for cuffless BP devices. In particular, the ANSI/AAMI/ISO [242] validation criterion of mean $\pm$ standard deviation of error $< 5 \pm 8$ mmHg was specifically designed for validating oscillometric devices against the auscultatory method. Cuffless devices will produce biasedly favourable results due to their calibration procedure which effectively sets the mean error to 0 at the start of the recording. If the study protocol does not significantly vary BP, the standard deviation of the error will remain low and thus it can be comparatively easy for a cuffless device to meet the current ANSI/AAMI/ISO validation criteria. To combat this, the ISO is currently developing a new protocol specific to continuous recordings of BP [242]. Until this protocol is released, validation thresholds should not be reported on for cuffless devices.

2.12 Conclusion

In this chapter, we provided a comprehensive review of the relevant background on cuffless BP measurement, with a particular focus on 24-hour circadian variations in BP. Inadequate BP control is a leading global cause of non-communicable mortality, and evidence strongly suggests that

Table 2.7: Overview of the three main available devices that make commercial claims of cuffless blood pressure monitoring.

Device	Mode of operation	Claims	Validation
Sotera Wireless	PAT-based with periodic recalibrations.	Continuous multi-parameter vital-sign monitoring including BP.	FDA clearance only when used by a trained clinician such that abnormal deviations in BP can be verified by appropriate means. Independent studies have highlighted significant errors in BP estimation. [243]
Samsung Galaxy Watch Active 2	BP estimation through wrist reflectance mode PPG.	Continuous BP after initial calibration and recalibration every month.	To the author's knowledge, no technical validation report has been released by Samsung.
Aktiia	BP estimation through wrist reflectance mode PPG.	Continuous BP after initial calibration and recalibration every month.	Only internal validation reports state that the device meets ANSI/AAMI/ISO validation criteria [24]. Work by this group and Tan <i>et al.</i> [26] has shown notable errors in estimating the magnitude of changes.

analysing 24-hour ambulatory BP variations provides a superior indicator of cardiovascular health compared to current clinical practices. However, current ambulatory monitoring methods, which necessitate cuff inflations, can be inconvenient, costly, and disruptive, leading to infrequent use.

Accurate cuffless BP measurement holds great promise, particularly for individuals with suspected hypertension or those requiring discreet night-time monitoring. Its comfort, convenience, real-time data, and long-term monitoring capabilities could significantly enhance diagnostic accuracy and treatment effectiveness.

Over recent decades, various methods have emerged that utilise ECG or PPG waveforms for cuffless BP estimation. However, these approaches are by no means validated, and there remain considerable challenges regarding the optimal signals, appropriate modelling, and translation to the ambulatory domain. Additionally, very few studies have assessed the accuracy of these techniques for monitoring the variations in BP that occur over 24 hours.

3

Clinical datasets

Contents

3.1	Introduction	63
3.2	Dataset requirements	64
3.2.1	Instrumentation requirements	64
3.2.2	Population size requirements	65
3.3	Mapping Of Lower Limb skIn pErfusion	65
3.3.1	Overview	65
3.3.2	Motivation	66
3.3.3	Study design and protocol	66
3.3.4	Research ethics and informed consent	66
3.3.5	Participants and eligibility criteria	67
3.3.6	Data acquisition	68
3.3.7	Cohort description	71
3.4	Medical Information Mart for Intensive Care - III	72
3.4.1	Overview	72
3.4.2	Patient selection criteria	74
3.5	Non-invasive ambulatory blood pressure monitoring study	75
3.5.1	Overview	75
3.5.2	Motivation	75
3.5.3	Study design and protocol	76
3.5.4	Research ethics and informed consent	77
3.5.5	Participants and eligibility criteria	77
3.5.6	Data acquisition	78
3.5.7	Cohort description	79

3.1 Introduction

Open-source datasets specific to the research on cuffless BP monitoring are limited in number, primarily due to data protection laws. At the time of writing this thesis, to our knowledge, the only relevant example is Liang *et al.* [244] who published a public dataset in which BP and PPG data were recorded from 657 individuals. However, this dataset only includes one BP measurement per individual under resting conditions, making the investigation of *individual calibration* strategies, as well as the study of circadian rhythms, impossible.

This chapter describes the three datasets used in this thesis: Mapping Of Lower Limb skIn pErfusion (MOLLIE), Medical Information Mart for Intensive Care (MIMIC), and non-invasive ambulatory blood pressure monitoring (nABPM). We conducted two clinical studies (MOLLIE and

nABPM) as well as analysed a large, well-known, public dataset recorded in the Intensive Care Unit (ICU). Each dataset provided valuable insights to answer research questions specific to this thesis.

3.2 Dataset requirements

This section outlines the core dataset requirements employed for the clinical studies designed for this thesis. These requirements detail the minimum criteria for instrumentation and population sizes when collecting data. Specific requirements tailored to individual studies are introduced in their respective sections.

3.2.1 Instrumentation requirements

To guarantee the precision of all signal recordings, we opted to collect all physiological signals using clinical-grade equipment, in contrast to consumer-grade alternatives that often prioritise user comfort over data accuracy.

3.2.1.1 Electrocardiogram signal

- **Electrode setup:** The ECG to be recorded using a minimum 3-electrode ECG setup in Einthoven's triangle configuration to optimise signal quality.
- **Sampling rate:** A minimum sampling rate of 100 Hz was defined to minimise predicted BP resolution. This rate was determined by calculating the necessary time resolution to achieve the ANSI/AAMI/ISO ± 5 mmHg BP validation standard, [242] assuming a typical PTT calibration slope of 500 mmHg s^{-1} , as documented in previous studies [182, 162, 181, 151].
- **Amplitude resolution:** The ECG signal to be recorded with a 16-bit analogue-to-digital converter (ADC) amplitude resolution, providing an adequate dynamic range for ECG signals [245].

3.2.1.2 Photoplethysmogram signal

- **Sensor type:** Pulse oximeter, operating in the transmission mode, to be placed on the fingertip.
- **Sampling rate:** Similar to the ECG, the PPG signal to be sampled at a frequency greater than 100 Hz.
- **Amplitude resolution:** Similar to the ECG, the PPG signal to be recorded with a 16-bit ADC amplitude resolution, providing an adequate dynamic range for PPG signals [246].

3.2.1.3 Blood pressure measurement

BP to be recorded using an oscillometric sphygmomanometer cuff that meets the ANSI/AAMI/ISO 81060-2 validation criteria [242]. The choice of an appropriate cuff size will be based on each participant's arm circumference.

3.2.1.4 Additional requirements

- **Clock synchronisation:** ECG and PPG signals to be recorded on a synchronised internal clock to ensure precise signal alignment.
- **Data transmission and storage:** All signal data to be recorded and stored for subsequent analysis.
- **Compliance with Regulatory Standards:** All aspects of data collection and equipment usage to adhere to relevant regulatory standards and ethical guidelines, prioritising participant safety and ethical treatment.

3.2.2 Population size requirements

In light of the exploratory nature of this research, the studies conducted as part of this thesis primarily serve as feasibility studies aimed at assessing whether this area warrants further investigation. Therefore, the determination of study population sizes for this thesis was largely guided by pragmatic considerations concerning participant recruitment and practical considerations for running the studies, without specific power calculations for sample sizes. If the results of these feasibility studies demonstrate the viability of accurate BP estimation through ECG and PPG signals, any future work will necessitate careful consideration of statistical power and sample size calculations.

3.3 Mapping Of Lower Limb skin perfusion

3.3.1 Overview

The MOLLIE study is a healthy volunteer study in which changes in blood perfusion were induced via an infusion of phenylephrine under a strict protocol [32]. Phenylephrine is a synthetic α -1-adrenoceptor agonist that acts selectively on the receptors present on peripheral blood vessel walls, both arterial and venous, without significant direct effect on the cardiac myocytes [247]. It has the clinical impact of increasing BP via peripheral vasoconstriction and as well as inducing reflex bradycardia. Phenylephrine is frequently used to treat hypotension associated with spinal or epidural anaesthesia.

3.3.2 Motivation

The MOLLIE study protocol induced significant changes in BP safely and quickly while participants remained in a supine and still position. Data from this study, therefore, facilitated the initial development of algorithms for cuffless estimation of BP in a controlled setting. Using this dataset, we developed methods to address the following two research questions:

- Does PAT provide a reliable surrogate measure for BP?
- Are morphological features extracted from the ECG and PPG waveforms reliable and robust enough for the accurate estimation of changes in BP?

3.3.3 Study design and protocol

The study took place at the Cardiovascular Clinical Research Facility (CCRF) within the John Radcliffe Hospital, Oxford, United Kingdom. The CCRF is a dedicated research environment with expertise in echocardiography, exercise physiology, and vascular assessments. It was chosen for the MOLLIE study to maximise the cardiovascular safety of the study participants. The facility was well equipped for potential haemodynamic resuscitation to safely support cardiovascular clinical studies. The haemodynamic resuscitation equipment was checked daily and was available on standby for immediate use as necessary.

3.3.4 Research ethics and informed consent

The study was reviewed and approved by the Oxford University Research and Ethics Committee and Clinical Trials and Research Governance teams (R63796/RE001). The study was registered with International Standard Randomised Controlled Trial Number (ISRTN) registry (ISRCTN10417167). Written informed consent was obtained from all participants. Participants were identified only by a unique participant identification number on study documents and on any electronic database. Documents that contained identifying data or information allowing the association between participant identification number and personal data were stored separately under strict access controls. The participants retained the right to have their data deleted at any time upon request. All participants underwent the same study protocol without randomisation.

Each participant attended for one in-person visit. Participants were asked to refrain from drinking caffeinated drinks for four hours prior to the study visit as caffeine is a vasoconstrictor. The study visit began with informed consent taken by the lead clinician, a final eligibility check, measurement of height, weight, and baseline vital-sign observations including HR, BP, RR, and peripheral SpO_2 . A urinary β HCG test was performed on all female participants to rule out

pregnancy before proceeding with the study. A cardiovascular examination was performed to detect any underlying abnormalities that would make the infusion of phenylephrine unsafe.

Participants were positioned on their backs with the head and trunk raised to between 15 and 45 degrees (semi-Fowler's position). A Becton Dickson Venflon (Eysins, Switzerland) intravenous cannula with injection port (size 22G blue 25mm) was placed in the participant's right antecubital fossa using a sterile non-touch technique. The intravenous cannula was kept in place using a Smith & Nephew IV3000 transparent dressing. Phenylephrine hydrochloride, provided as 10 mg.mL⁻¹ clear colourless solution, was used in the study. One mL of phenylephrine hydrochloride was diluted in 100 mL of 0.9% saline to create concentrations of 100 µg.mL⁻¹. A labelled 50 mL BD Plastipak Luerlock syringe connected to a double-lumen extension (with the second lumen connected to the 0.9% saline infusion) was used.

A schematic of the experimental timeline is shown in figure 3.1. Each session began with a 5-minute resting period followed by a 9-minute infusion rate escalation phase during which the dosage of phenylephrine was steadily increased. The infusion started at 0.2 µg.kg⁻¹.min⁻¹ (adjusted by body weight) and increased by 0.2 µg.kg⁻¹.min⁻¹ increments every minute until a maximum infusion rate of 2 µg.kg⁻¹.min⁻¹. The maximum absolute peak infusion rate was set at 140 µg.min⁻¹ (regardless of participant body weight). This maximum rate was chosen with guidance from the clinical advisory team as part of the Central University Research Ethics Review for the study. After the escalation phase, the infusion rate was held at its maximum value for 6 minutes. Each session then ended with an 8-minute washout period.

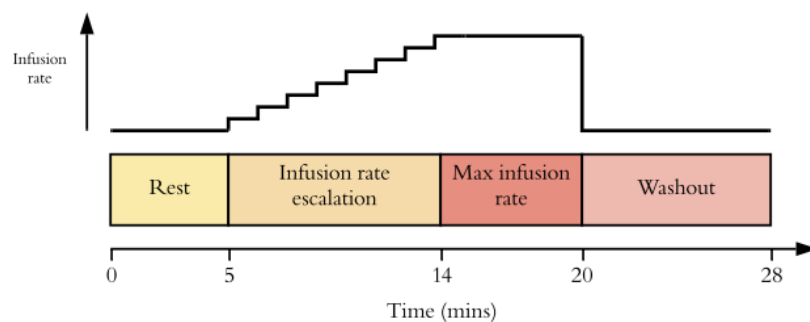


Figure 3.1: MOLLIE study experimental timeline. During the study, participants were administered phenylephrine, a vasoconstrictor.

3.3.5 Participants and eligibility criteria

We recruited thirty participants between the ages of 18 and 65. We used a combination of in-person recruitment, poster advertisements, and snowball recruitment to reach eligible individuals.

All individuals who expressed interest were enrolled serially as long as they met the following inclusion and exclusion criteria:

Inclusion criteria:

- Healthy adults aged 18 to 65 years.
- Willing and able to give informed consent for participation in the study.

Exclusion criteria:

- Participants with baseline SBP above 140 mmHg or baseline DBP above 90 mmHg on repeated measurements.
- Participants who, upon a cardiovascular examination from the lead clinician, showed signs of cardiovascular disease, palpation of central pulse, and auscultation of heart sounds.
- Allergy to silver chloride ECG sensors.
- Hyperthyroidism (intravenous phenylephrine contraindicated).
- Any regular medication except oral contraception.
- Pregnant or breastfeeding.
- History or current neurological condition affecting peripheral circulation.
- History of cardiovascular disease where phenylephrine or glyceryl trinitrate are contraindicated.
- History of severe headaches

3.3.6 Data acquisition

Throughout the study, for each participant we simultaneously recorded ECG, PPG, nasal airflow, and ICG physiological signals, alongside reference BP measurements. It's worth noting that, to the best of our knowledge, there isn't a single commercially available device capable of recording all these essential signals while meeting the specified instrumentation requirements outlined above. Consequently, we employed three separate recording devices to accomplish this task. Figure 3.2 shows the typical data collection set-up used in the MOLLIE study. Table 3.1 provides an overview of the physiological parameters recorded by the three devices: the Philips patient monitor, the Stowood device, and the Cardioscreen bioimpedance monitor.

Since data were recorded from three different devices, time synchronisation was required at the beginning of each recording session in order to ensure the accuracy of timestamps across the devices. At the start of each session, all devices were time-synchronised using the network time protocol (NTP).

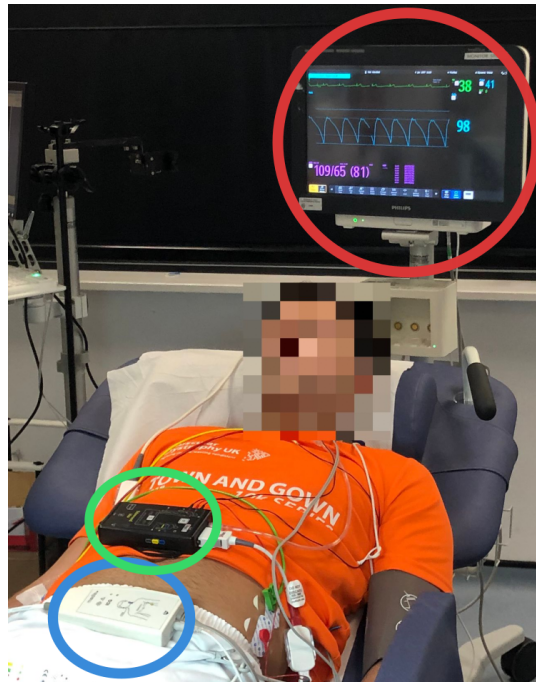


Figure 3.2: MOLLIE study equipment setup. The Philips patient monitor is circled in red, the Stowood device is circled in green and the Cardioscreen monitor is circled in blue. Consent was obtained from the participant to use the image for publication.

Table 3.1: Physiological parameters and sampling rates recorded by the three devices used in the MOLLIE study (Philips, Stowood, CardioScreen).

Device	Signal name	Description	Sampling rate (Hz)
Philips	ECG	Single-lead ECG signal	500
	BP	BP signal from the sphygmomanometer cuff	1/60
	HR	Heart rate derived from the ECG signal	1
Stowood	ECG	Single-lead ECG signal	256
	PPG	PPG signal recorded using a finger pulse oximeter	512
	Airflow	Airflow from nasal cannula signal	32
	HR	Heart rate derived from Masimo pulse oximeter	5
	SpO_2	Oxygen saturation from Masimo pulse oximeter	5
Cardioscreen	ECG	Single-lead ECG signal	200
	Z	Thoracic impedance signal	200
	HR	Heart rate derived from the ECG signal	1

All data were stored on a secure University server. All study documents were stored securely and only accessible by study staff (Kadoorie Centre for Critical Care Research and Education, and the Institute of Biomedical Engineering) and authorised personnel from the University of Oxford for monitoring or audit of the study to ensure compliance with regulations. The data will be stored securely for 5 years after the release date of the last publication arising from the study.

3.3.6.1 Philips patient monitor

A Philips IntelliVue MX800 patient monitor (Philips, Netherlands) [248], alongside a Philips IntelliVue MMX Multi-Measurement Module, was used to record a single-lead ECG signal as well as reference vital signs (BP and HR). ECG Cable 3 Leadset Grammer IEC ICU (M1672A) connected to an ECG Cable 3 Lead (M1669A) was used to record the ECG signal at a sampling frequency of 500 Hz. HR was computed by the Philips monitor at a frequency of 1 Hz. A Comfort Care Reusable Adult NIBP cuff (M1574A) blood pressure cuff was wrapped around the upper left arm of the participant directly above the brachial artery as recommended [102]. BP was recorded at a rate of one measurement per minute. Data from the Philips patient monitor was transmitted to a dedicated workstation using the ixTrend software (Ixellence GmbH, Germany) onto the dedicated workstation hard drive.

3.3.6.2 Stowood device

The Stowood Visi Black Series Polygraphy device (Stowood, Oxford, UK) [249] is a portable device designed to record physiological signals for overnight studies. In this study, we used the Stowood device to record ECG, PPG, and airflow signals; and reference vital signs such as HR and SpO_2 . A Masimo M-LNCS DBI (Masimo, Irvine, CA, USA) pulse oximeter probe was placed on the middle finger of the participant's right hand. The ECG signal was recorded at a rate of 256 Hz, and the PPG signal at a rate of 512 Hz. SpO_2 was computed by the Masimo pulse oximeter at a rate of 5 Hz. Using Stowood proprietary software, HR was derived from the PPG signal at 5 Hz. Participants breathed through a tight-fitting nasal cannula (EMBLA, Embla, Ontario, Canada) fitted with a nasal pressure transducer. The airflow signal was recorded at a sampling frequency of 32 Hz. The Visi-download software was used to download the data recorded by the Stowood device.

3.3.6.3 Cardioscreen monitor

The Cardioscreen monitor (Cardioscreen 1000, Medis, Germany) is a commercially available thoracic bioimpedance monitor. The device was used to record the bioimpedance (Z) signal as well as a single-lead ECG waveform, both at a sampling rate of 200 Hz. HR was computed by the Cardioscreen monitor at a rate of 1 Hz. To record Z, four dual sensors were placed on the left lateral thorax and left side of the neck, and the Arterial Compliance Modulation earlobe sensor was placed onto the left earlobe. The monitor was connected to a portable computer using the manufacturer's own software, Cardio-Vascular-Lab.

Table 3.2: Demographics of the participants in the MOLLIE study.

Descriptor	Value
Total number of participants	26
Male	12
Female	14
Average length of session (mins) ¹	28.0 (0.1)
Age (years) ¹	28.0 (9.0)
Height (cm) ¹	170.0 (18.0)
Weight (kg) ¹	69.5 (23.0)
Body Mass Index (kg/m ²) ¹	22.5 (5.2)

¹ Participant-wise median (inter-quartile range (IQR)).

3.3.7 Cohort description

3.3.7.1 Participant demographics

We recruited thirty participants for the study. We discarded the data from four participants from subsequent analyses. The reasons for their exclusion are as follows:

- For one participant, the ECG electrodes from the Stowood device were connected incorrectly, resulting in a poor-quality ECG signal recording.
- For two participants, their ECG morphologies indicated that they were not in normal sinus rhythm, with QRS complexes in bigeminy pattern [250], likely due to the phenylephrine-induced bradycardia [247].
- For one participant, there were errors in recording the BP signal.

Therefore, we included 26 participants in all data analyses. The demographics of the included participants are provided in table 3.2. All participants were healthy with a median BMI of 22.5 kg/m² and no history of cardiovascular disease. The median age of participants was 28 years and there was an even split of sexes (53.8% female).

3.3.7.2 Recorded vital signs

The distributions of vital signs recorded in the study are shown in figure 3.3. Figure 3.3(a) shows a strong agreement between the HR values computed by the three devices used in the study. Figure 3.3(b) shows SpO_2 computed by the Stowood finger pulse oximeter which exist within typical healthy limits [28]. Figure 3.3(c)-(e) show the distributions of SBP, MAP and DBP respectively, all recorded by the BP cuff and Philips patient monitor. Given that phenylephrine is a vasodilator, the recorded BP measurements exhibit higher values than what would typically be observed in a healthy population. The means of these distributions fall within the pre-hypertensive range, as detailed in table 2.4.

Figure 3.3(f)-(h) show the distributions of changes (Δ) in SBP, MAP and DBP respectively from their baseline value recorded during the 5-minute rest period at the start of the study. The precise clinical effect of the weight-based dosing of phenylephrine in an individual depends on the balance between their sensitivity to the increase in afterload, the effect of bradycardia on cardiac filling (therefore contractility, the Frank-Starling law [251]), and the proportion of venous/arterial action of phenylephrine causing the increase in preload and afterload. Therefore, variations in Δ BP across the cohort were expected. The participant-wise median maximum Δ SBP (from rest to peak infusion) was 20 mmHg, with a maximum increase of 40 mmHg in a subset of four participants.

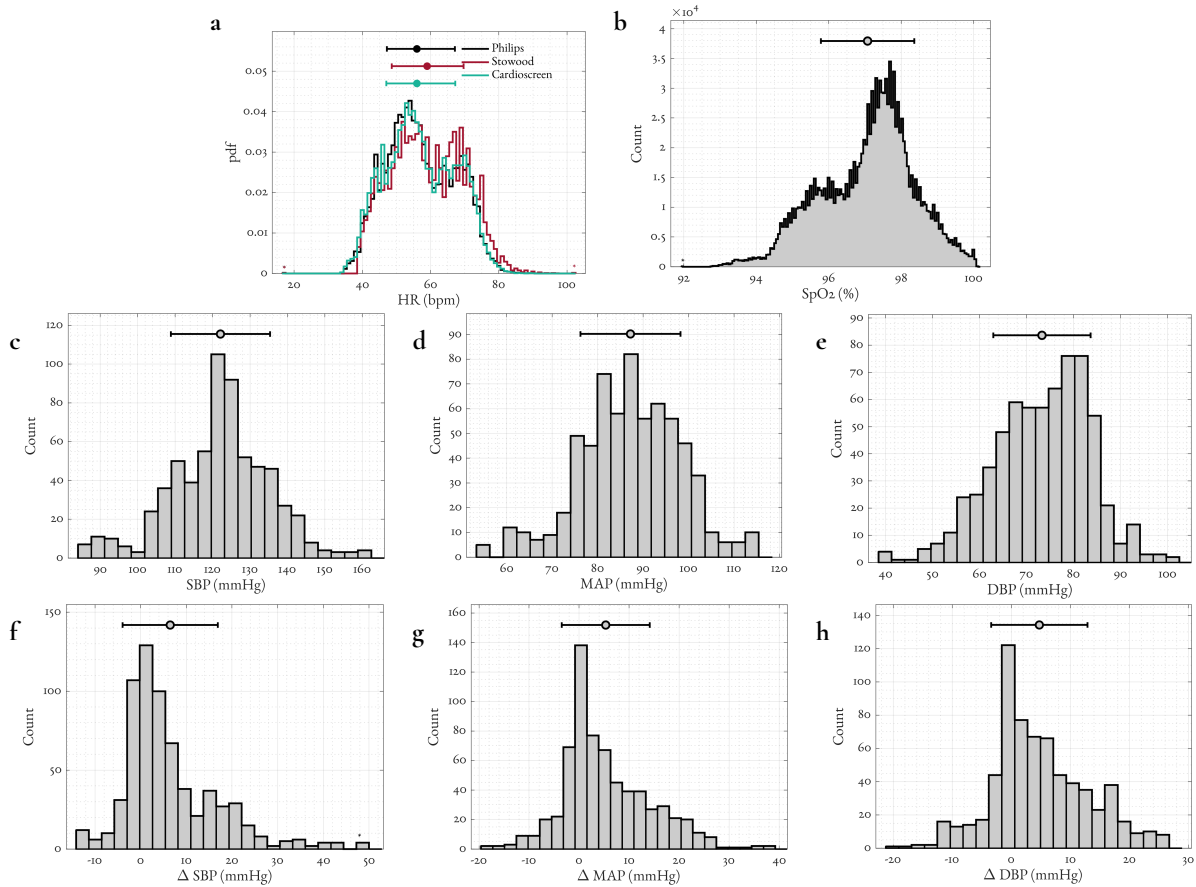


Figure 3.3: Distributions of recorded vital signs in the MOLLIE dataset. (a) HR values computed from the three devices. (b) SpO_2 computed by the Stowood finger pulse oximeter. (c)-(e) SBP, MAP and DBP respectively all recorded by the BP cuff and Philips patient monitor. (f)-(h) changes (Δ) in SBP, MAP and DBP respectively from their baseline value recorded during the 5-minute rest period at the start of the study. The boxplots show the mean and standard deviation.

3.4 Medical Information Mart for Intensive Care - III

3.4.1 Overview

Due to the COVID-19 pandemic and the resulting lockdowns, we were unable to conduct any clinical studies or collect data from human participants. Therefore, during this period, our research

primarily involved analysing an existing public domain database.

The MIMIC - III database [168] is a publicly available database comprising de-identified clinical data of patients admitted to the Beth Israel Deaconess Medical Centre (BIDMC) in Boston, Massachusetts, USA between 2001 and 2012. The database covers 46,520 patients during their stay in the ICU. All data were gathered during routine hospital care and were de-identified in accordance with the Health Insurance Portability and Accountability Act (HIPAA). Use of this database was approved by the Institutional Review Boards of the BIDMC and Massachusetts Institute of Technology (Cambridge, MA).

The MIMIC database is comprised of two parts: the waveform database and the clinical database. The waveform database contains physiological signals such as the ECG and PPG recorded by a Philips patient monitor at a sampling rate of 125 Hz across thousands of patients. PPG signals were recorded using finger pulse oximeters. The clinical database is a set of clinical notes for each patient recorded in the hospital information system. These notes were recorded by the doctors and included admission reports, procedures, medications taken, discharge summaries, and more. Additionally, vital signs (such as HR, RR, and BP measured by a sphygmomanometer cuff) were regularly measured and recorded by the clinical team. This typically occurred once per hour; however, the frequency was also driven by the level of care and attention each patient required. The records were collected using one of two critical care information systems: Philips CareVue Clinical Information System or iMDsoft MetaVision ICU.

The clinical database was saved independently of the waveform database and with different patient identifiers. Many of the records in the waveform database have been matched to the patients in the clinical database by considering the time overlap between waveform records and clinical records and if derived parameters (such as HR) agree. We refer to these records as the matched subset.

3.4.1.1 Motivation

The assessment of circadian rhythms of BP using cuffless signals is limited by a requirement for a large and heterogeneous cohort of individuals being monitored. Assembling a large and diverse database of individuals wearing an ABPM, ECG, and PPG sensors is not practically feasible given the low user acceptability and tolerance of ABPM [252]. To combat these limitations, an ICU database represents a setting with a large set of physiological parameters recorded across a diverse range of individuals.

Patients in the ICU may often experience circadian disruption caused by tight controls of vital signs through the use of medication, in addition to sedation, patient examinations, and unnatural

environmental light. Despite this, it has been previously observed that in a select group of patients in the ICU who were discharged home healthy, circadian rhythms in vital signs can be observed in the last 24-hour period prior to discharge [2, 253]. Using the MIMIC dataset we developed methods to address the following two research questions:

- Do the cuffless surrogate measures for BP, developed using the MOLLIE database, also exhibit circadian-like patterns in synchrony with the reference clinical BP values?
- Can we use this information to improve cuffless BP phenotype classification?

3.4.2 Patient selection criteria

We restricted our analysis to a subset of patients who, whilst in the ICU, were closest to recovery and thus most likely to exhibit physiological profiles comparable to out-of-hospital individuals, such as individuals in an ambulatory setting or at home. Davidson *et al.* [2, 253], have previously shown that the following selection criteria are sufficient to obtain patients, within the MIMIC-III database, whose 24-hour variation in vital signs (SBP, HR, RR, and core body temperature) conform to those reported for non-ICU cohorts.

The patient selection criteria are outlined in figure 3.4. From the initial cohort of 10,282 patients in the matched subset, we considered only adult patients (aged between 15 and 90 years old) that were discharged home healthy and did not return to the hospital within 6 months. The patients must not have been given a "Do Not Resuscitate (DNR)", "Do Not Intubate (DNI)" or "Comfort Measures Only (CMO)" code during their hospitalisation. In these cases, palliative care was initiated which indicates a deviation from typical ICU care.

We restricted our analysis to the last 24 hours during which a patient's vital signs were being recorded by the bedside monitor. If the gap from the end of this day to the time of discharge from the ICU was more than two days, the patient was not included in the database as they were deemed to be sufficiently far from recovery. Patients that had a pacemaker were removed as this acts as an external excitation source for the cardiovascular system (this was checked by manually reviewing the patient's discharge summary).

Measurements taken while the patient was under the effect of medications that were likely to significantly affect the vital signs being measured (such as vasopressors and β -blockers) were excluded as these treatments may have had a significant impact on their circadian profiles. The medications considered and the period of time for which they were considered active were:

- Up to one hour after a patient was on catecholamine vasopressors, which have a half-life of 2 - 3 minutes, or Phenylephrine, which has a duration of action of 20 minutes [254, 255].

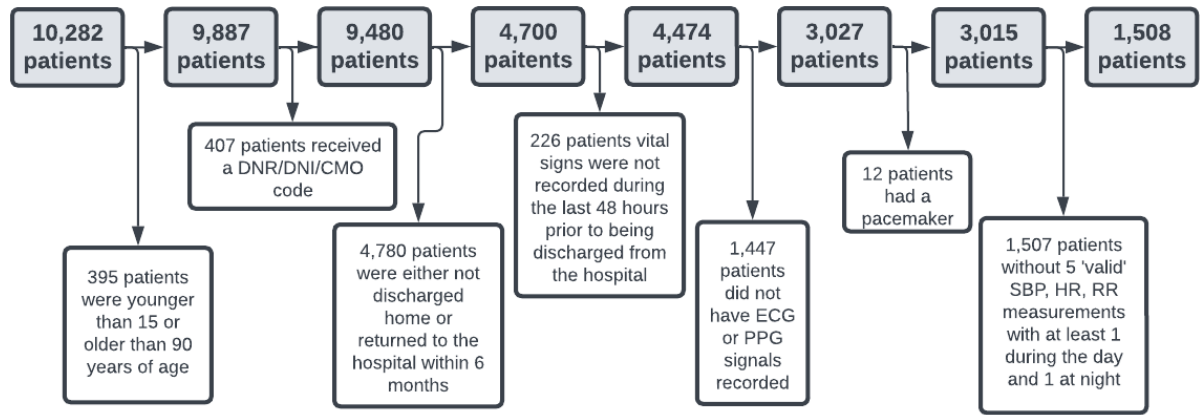


Figure 3.4: MIMIC patient selection criteria. From the 10,282 patients in the matched subset, we selected 1,508 patients who were most likely to exhibit physiological profiles comparable to out-of-hospital individuals.

- Up to two hours after a patient was on Vasopressin (half-life 16 - 24 minutes), Propofol (duration of action 10 minutes, but complex discharge pathways), or Magnesium Sulfate [256, 257, 258].
- Up to one day after a patient was on PDE3 inhibitor Milrinone (half-life of 2 hours), β -blocker Metoprolol (half-life of 3 hours), or Hydralazine [254, 259, 260].

Patients must have had at least five separate hours during which valid measurements of SBP, HR, and RR were recorded. Measurements outside of broad physiological bounds ($60 \text{ mmHg} < \text{SBP} < 280 \text{ mmHg}$, $30 \text{ bpm} < \text{HR} < 240 \text{ bpm}$, $4 \text{ breaths/min} < \text{RR} < 60 \text{ breaths/min}$) were excluded [2]. At least one of these measurements must have been recorded during the night (defined as midnight to 8am) and at least one during the day (defined as 10am to 8pm). These patient selection criteria resulted in 1,508 patients that we analysed as our patient dataset.

3.5 Non-invasive ambulatory blood pressure monitoring study

3.5.1 Overview

The nABPM study is a healthy volunteer study designed to further develop the methodology introduced in the MOLLIE and MIMIC datasets towards estimation of ambulatory BP using the PPG and ECG signals. This study consisted of two 24-hour sessions, spaced one week apart, during which participants carried out their normal daily activities while physiological data was recorded.

3.5.2 Motivation

We extended our methods developed using the previous two datasets (MOLLIE and MIMIC) to this ambulatory domain to address the following three research questions:

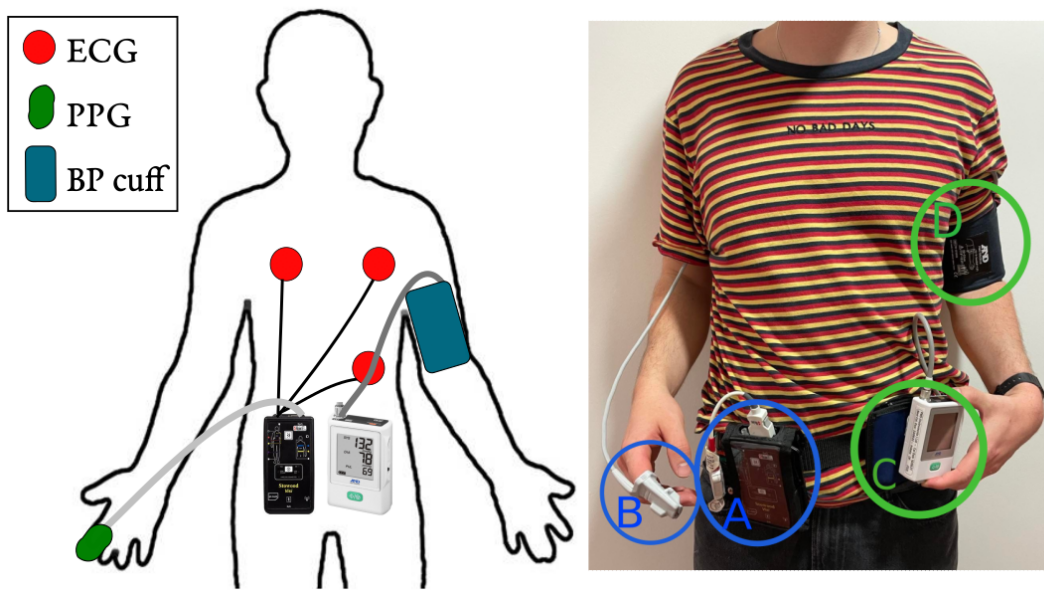


Figure 3.5: nABPM equipment setup. Data were recorded from two devices. The Stowood device (circled in blue) recorded ECG, PPG, and accelerometer signals. A: the Stowood device; B: the Masimo pulse oximeter placed on the index finger of the participant's right hand. The A&D ABPM (circled in green) recorded BP. C: the A&D ABPM device, D: the cuff wrapped around the participant's left arm. All equipment was housed in a belt strapped to the participant's waist which allowed for free movement. Consent was obtained from the participant to use the image for publication.

- Is it feasible to evaluate reliable and robust BP surrogate measures in an ambulatory setting?
- Can the need for *individual calibration* be circumvented by implementing *hybrid calibration*?
- Does this calibration hold over one week?

3.5.3 Study design and protocol

The study consisted of two 24-hour sessions spaced one week apart. The initial study setup occurred at the Institute of Biomedical Engineering (IBME), Oxford. The setup began with informed consent taken by the lead clinician, followed by verification that the participant met the eligibility criteria. Measurements of height and weight were subsequently recorded. All equipment was then attached to the participant and programmed to record data for 24 hours. Figure 3.5 shows the equipment setup used, including an ECG, PPG, accelerometer sensor, and ABPM. The ABPM was programmed to inflate once every 30 minutes during the day and once per hour during the night. The participant specified their planned night-time schedule to configure the ABPM. All equipment was housed in a belt strapped around the participant's waist which allowed for free movement.

Finally, a calibration reading was taken by the lead clinician. This consisted of the participant remaining still for two minutes while all signals, including a reference BP measurement, were being recorded.

Following the study set-up, the participant was free to live their normal life with all of the equipment on for 24 hours, taking the equipment off only for bathing and travelling. At the end of the 24 hours, the equipment was returned and the participant was asked to complete a short sleep questionnaire detailing their sleep quality. This sleep questionnaire is provided in appendix B. The whole process was then repeated 7 days later.

3.5.4 Research ethics and informed consent

The study was reviewed and approved by the Oxford University Research and Ethics Committee and Clinical Trials and Research Governance teams (R80677/RE001). Written informed consent was obtained from all participants. Participants were identified only by a participant identification number on study documents and on any electronic database. Documents that contained identifying data or information allowing the association between participant identification number and personal data were stored separately under strict access controls. The participants retained the right to have their data deleted at any time upon request. All participants underwent the same study protocol without randomisation.

3.5.5 Participants and eligibility criteria

We recruited fifteen participants between the ages of 18 and 69. We used a combination of in-person recruitment, poster advertisements, and snowball recruitment to reach eligible individuals. All individuals who expressed interest were enrolled serially as long as they met the following inclusion and exclusion criteria:

Inclusion criteria:

- Healthy adults aged 18 to 65 years.
- Willing and able to give informed consent for participation in the study.

Exclusion criteria:

- Individuals suffering from tachycardia or atrial fibrillation.
- Individuals suffering from diabetes or renal dysfunction.
- Individuals who are pregnant.
- Individuals with significant sleeping difficulties or who use sleeping medications.
- Active smokers.
- History of, or current psychiatric illness.
- History of, or current neurological condition (e.g. epilepsy).
- Individuals on blood pressure medication.
- Individuals with injuries (such as broken bones or large amounts of bruising) in their arms.
- Individuals that have had surgery involving replacing an artery.
- Individuals that are allergic to adhesive stickers.
- Individuals that perform active or manual jobs (e.g. electricians).

Table 3.3: Physiological parameters and sampling rates recorded by the two devices used in the nABPM study (Stowood and ABPM).

Device	Signal name	Description	Sampling rate (Hz)
Stowood	ECG	Single-lead ECG signal	256
	PPG	PPG signal recorded using a finger pulse ox.	128
	HR	Heart rate derived from Masimo pulse ox.	5
	SpO_2	Oxygen saturation from Masimo pulse ox.	5
	Accel	Accelerometer located in the Stowood device	32
A&D	BP	BP from the ABPM cuff	<i>All signals recorded once every 30 minutes during the day & once an hour during the night.</i>
ABPM	HR _{Cuff}	HR from the ABPM cuff	

3.5.6 Data acquisition

In addition to the requirements specified in section 3.2, a key aspect considered in instrumentation was device portability. This ensured that participants could maintain their regular daily routines without the burden of cumbersome equipment. With these factors in mind, data collected in this study came from two sources: the Stowood device and the A&D ABPM. Table 3.3 provides an overview of the physiological parameters and sampling rates recorded by the two devices.

3.5.6.1 Stowood device

Physiological signals (ECG, PPG, and acceleration) and reference vital-signs (HR and SpO_2) were recorded using the Stowood device. A Masimo M-LNCS DBI (Masimo, Irvine, CA, USA) pulse oximeter was placed on the index finger of the participant's right hand. The ECG signal was recorded at a sampling rate of 256 Hz, and the PPG signal at a rate of 128 Hz. The Stowood device records triaxial acceleration at a rate of 32 Hz. The three axes were combined using a Stowood proprietary algorithm to give a single time series indicating participant movement. The Visi-download software was used to download the data from the Stowood device onto a dedicated workstation.

3.5.6.2 A&D ABPM

Reference BP measurements were recorded using an A&D TM-2441 ABPM (A&D Instruments, Tokyo, Japan). The device also computed HR measurements at the times of cuff inflation (HR_{Cuff}). The correct cuff size for each participant was determined by measuring their left arm circumference. The cuff was wrapped around the upper left arm of the participant directly above the brachial artery, as recommended by the manufacturer [102]. The manufacturer provided proprietary software to download data from the ABPM to a dedicated workstation.

Table 3.4: Demographics of the participants in the nABPM study.

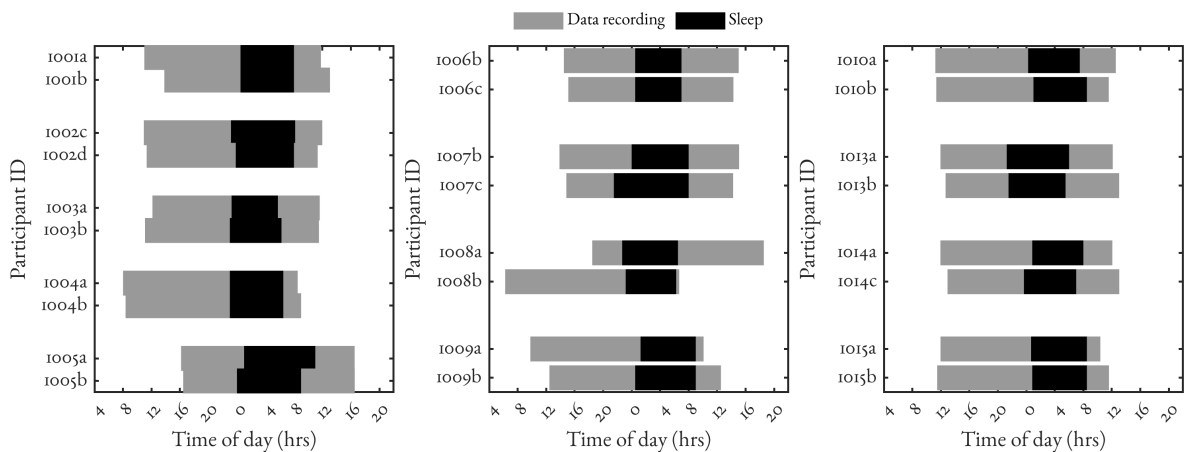
Descriptor	Value
Total number of participants	13
Male	6
Female	7
Age (years) ¹	31.0 (11.5)
Height (cm) ¹	168.0 (17.3)
Weight (kg) ¹	59.0 (22.4)
Body Mass Index (kg/m ²) ¹	21.7 (2.8)
Total number of sessions	26
Average length of session (hours) ¹	24.2 (0.4)
Average length of sleep (hours) ¹	7.6 (0.9)

¹ Participant-wise median (IQR).

3.5.7 Cohort description

3.5.7.1 Participant demographics

We recruited 15 participants for our clinical study. We discarded the data from two participants. For one participant, the pulse oximeter recording the PPG signal malfunctioned approximately an hour into a session and, therefore, no PPG data was recorded for the remaining 23 hours. For the second participant, the Stowood device was turned off while the participant was asleep. As a result, we included data from 13 participants in our dataset. The demographics of these remaining participants are shown in table 3.4. The recording time and reported periods of sleep for each session are shown in figure 3.6. Aside from one notable exception, 1008, participants largely followed a similar protocol of equipment on and off times and sleep times.

**Figure 3.6:** Monitoring durations and periods of sleep for the 13 participants in the nABPM dataset.

3.5.7.2 Recorded vital signs

The distributions of vital signs recorded in the nABPM study are provided in figure 3.7. Figure 3.7(a) shows the HR values from the two devices. As the AND device recorded vital sign measurements at a much lower frequency (once every 30 minutes during the day and hourly during the night), the bin widths used for the histogram are much wider than those used for the Stowood device (where HR measurements were reported five times a second). The boxplots, however, demonstrate that similar distributions of HR values are observed between the two devices. Figure 3.7(b) shows SpO_2 computed by the Stowood finger pulse oximeter. Figure 3.7(c)-(e) show the distributions of SBP, MAP and DBP respectively, all recorded by the ABPM. Figure 3.7(f)-(h) show the distributions of changes (Δ) in SBP, MAP and DBP respectively from their baseline values recorded during the 2-minute calibration period at the start of the study. Collectively, figure 3.7(c)-(f), show that typically healthy ranges of BP were recorded, with a significant variation in BP for each participant.

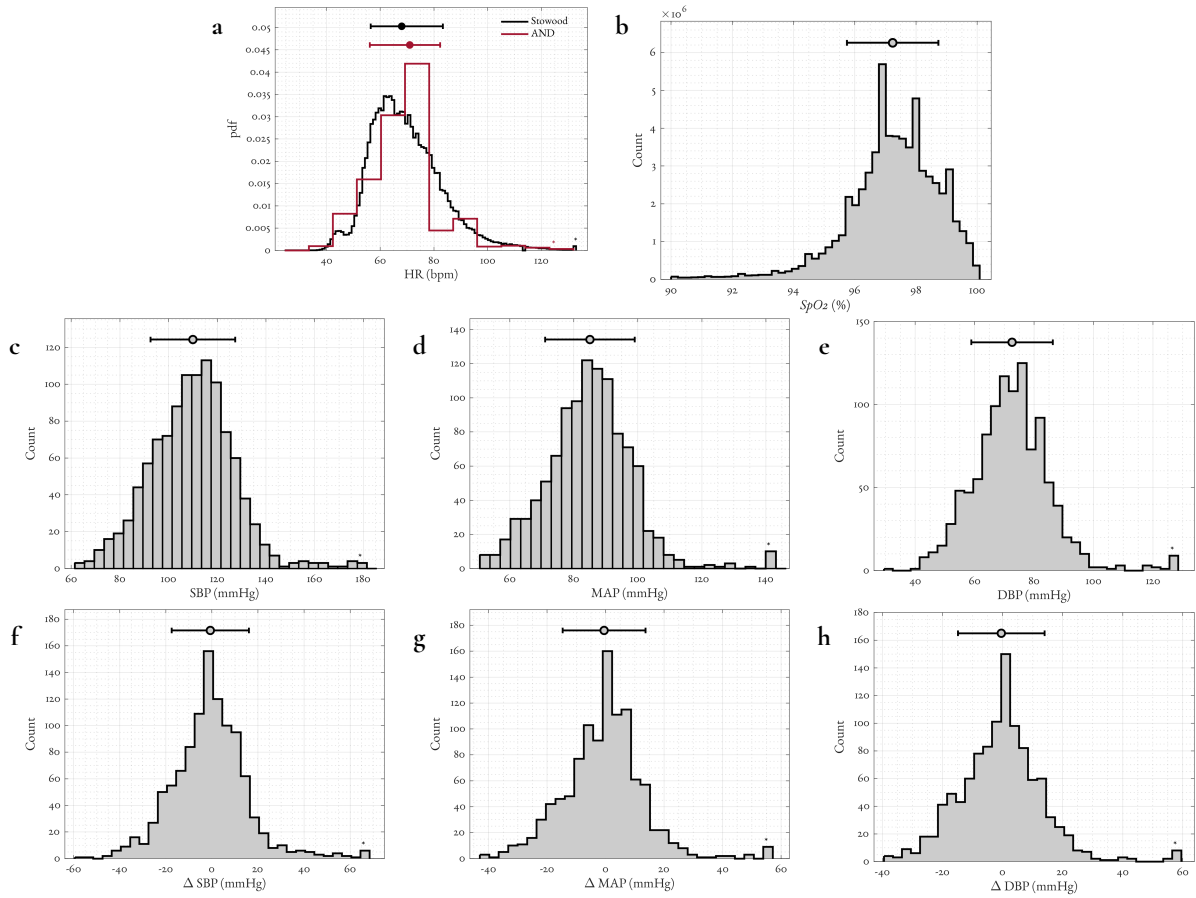


Figure 3.7: Distributions of recorded vital signs in the nABPM dataset. (a) HR values computed from the Stowood and the ABPM devices. Note the differences in bin widths as a result of different sampling frequencies. (b) SpO_2 computed by the Stowood finger pulse oximeter. (c)-(e) SBP, MAP and DBP respectively, all recorded by the ABPM. (f)-(h) changes (Δ) in SBP, MAP and DBP respectively from their baseline value recorded during the 2-minute calibration period at the start of the study. The boxplots show the mean and standard deviation.

4

Essential methodology

Contents

4.1	Introduction	83
4.2	Segmentation of cardiac beats	83
4.2.1	Segmentation of the electrocardiogram	84
4.2.2	Segmentation of the photoplethysmogram	85
4.3	Signal quality indices	86
4.3.1	Signal quality index for the electrocardiogram	86
4.3.2	Signal quality index for the photoplethysmogram	88
4.3.3	Combined signal quality index for the electrocardiogram and photoplethysmogram	91
4.4	Non-linear machine learning methods	92
4.4.1	Decision trees	93
4.4.2	Random forest	94
4.4.3	Mixed effects random forest	94
4.5	Shapley additive explanation values	95

4.1 Introduction

To develop our algorithms for cuffless ambulatory blood pressure monitoring using the ECG and PPG signals, a combination of different signal processing, statistical analysis, and classical machine learning (ML) methods were employed. In this chapter, we outline the essential methodology used throughout this thesis. In sections 4.2 and 4.3, methods specific to segmenting the cardiac beats and establishing the signal quality indices respectively for the ECG and PPG are outlined. Section 4.4 provide a background to regression and classification tasks and the non-linear methods implemented to address them. Section 4.5 introduces Shapley additive explanation (SHAP) values as an approach towards model interpretability through assessing feature importance.

In appendices C and D we outline the additional methodology used in this thesis. All statistical methods used to highlight statistical significance are introduced in appendix C. Linear ML models relevant to this thesis are outlined in appendix D.

4.2 Segmentation of cardiac beats

The segmentation of cardiac beats in ECG or PPG signals relies on detecting fiducial points within each waveform. Typically, high-frequency features (the QRS complex in the ECG and the

onset of the PPG pulse) are easier to detect as they are less corrupted by low-frequency noise such as movement related to breathing.

4.2.1 Segmentation of the electrocardiogram

The ECG waveform morphology has been outlined in section 2.6. Detection of the QRS complex is performed using the well-known Pan-Tompkins algorithm [119]. We additionally outline details of the Zong *et al.* [261] QRS detector given its role (alongside the Pan-Tompkins algorithm) in the derivation of ECG signal quality indices. Both algorithms detect the QRS complex using adaptive thresholding on a preprocessed ECG signal, ECG^{filt} . In both cases, ECG^{filt} was derived in order to enhance the features of the QRS complex¹. The location of the QRS complex was then determined as the point where ECG^{filt} exceeded a given threshold. For QRS detection to be performed accurately even when the ECG amplitude has changed, an adaptive threshold was implemented.

In this thesis, learning periods distributed across each recording were used to determine the initial threshold value and the orientation of the ECG waveform. The ECG waveform may be inverted if the electrodes are connected to the wrong ports on the ECG measuring device. This was relatively common in the MIMIC dataset as data collection occurred during routine care in the ICU. For long recordings (24-hour recordings in MIMIC and nABPM), ten learning periods of 5 minutes in length, evenly distributed over the entire recording, were used. For short recordings, (28-minute recordings in MOLLIE dataset), the entire recording was used as the learning period.

4.2.1.1 Pan-Tompkins

The Pan-Tompkins QRS detection algorithm is based on digital filtering and integration techniques [119]. The preprocessing steps to compute ECG^{filt} consist of: filtering, differentiation, squaring, moving window integration, and finally median filter smoothing [119]. For a clean ECG time series, the resulting ECG^{filt} signal reaches its maximum around the time of the QRS complex. Adaptive filtering was then performed to detect the QRS complexes. First, the time series of mean ECG^{filt} values over all learning periods was obtained. The threshold was then set as twice the median value over this time series. The orientation of the ECG waveform was determined by finding the crossings of ECG^{filt} over all of the learning periods with the threshold. The sign (positive or negative) of the point at which the absolute value of the ECG between the thresholds was at its maximum was used to determine the orientation of the ECG.

¹In order to not introduce unnecessary phase delays between waveforms (which would significantly impact any transit time computations) all filtering in this thesis was performed using zero-phase filters by processing the data in both the forwards and reverse directions.

The R-peak was identified as the maximum value (or minimum value if the ECG waveform was inverted) in the ECG when ECG^{filt} exceeded the threshold. A local search strategy was then applied to determine the location of the Q and S waves. To avoid missing beats, the threshold was reduced by half if the inter-beat (RR) interval between successive R-peaks was larger than twice the median of the RR intervals of R-peaks already detected. To reduce the possibility of multiple contiguous beat detections, a refractory period of 250 ms was introduced after a beat was detected.

4.2.1.2 Zong *et al.*

The Zong *et al.* QRS detection algorithm is based on the curve-length transform (LT) [261]. Initially, ECG^{filt} was formed by a sequence of: filtering, differentiation and squaring. The curve length transform (LT) of ECG^{filt} over an interval $[i - w, i]$ was defined as:

$$LT(w, i) = \sum_{k=i-w}^i \sqrt{1 + \left(\frac{\text{ECG}^{\text{filt}}_k - \text{ECG}^{\text{filt}}_{(k-1)}}{\Delta t} \right)^2 \Delta t} \quad (4.1)$$

where Δt is the sampling interval and w was chosen to be the maximum duration of the QRS complex, taken to be 150ms, [262].

The initial threshold was set to be the mean of the LT signal over the learning periods. The threshold was dynamically adjusted to 1/3 of the median of the maximum LT values for the detected QRS complexes in a 15-second window. For each threshold-crossing point, the algorithm searched backwards to find a Q point and forward to find an S point. The R-peak was defined as the maximum (or minimum if the ECG waveform was inverted) of the ECG signal between each pair of Q and S points. To reduce the possibility of multiple contiguous beat detections, a refractory period of 250 ms was introduced after a beat was detected.

4.2.2 Segmentation of the photoplethysmogram

Villarroel *et al.* [263] designed a robust beat onset detection algorithm for pulsatile signals known as the Boxed Slope Sum Function (BSSF). The algorithm builds on the Slope Sum Function (SSF) work from Zong *et al.* [264] by adaptively tracking the upslopes of the signal. SSF enhances the slope of the pulse and suppresses the remainder of the waveform. Given x_k as the PPG signal at time k , the SSF was defined as:

$$\text{SSF}(i) = \sum_{k=i-w}^i u_k, \quad u_k = \begin{cases} \Delta x_k & \text{if } \Delta x_k > 0 \\ 0 & \text{if } \Delta x_k \leq 0 \end{cases} \quad (4.2)$$

where w is the analysis window length and $\Delta x_k = x_k - x_{k-1}$. The window length w was chosen to be the typical duration of the upslope of the pulse, $w = 166.7$ ms.

The boxed slope sum function (BSSF) was computed from the SSF signal by finding the zero-crossing points at the beginning of the upslope and at the end of the downslope for each pulse. Adaptive thresholding and a decision rule were applied to the upslope points in the BSSF signal (uBSSF) to identify the onsets of each PPG beat [263].

4.3 Signal quality indices

A SQI of the ECG and PPG waveforms (SQI_{ECG} and SQI_{PPG} respectively) is a beat-by-beat assessment indicating a degree of confidence that each signal is recording its respective physiological process, i.e. that the signal is not corrupted with some degree of noise. The SQI is a number that varies between 0 and 1 (inclusive), representing a poor-quality or good-quality beat respectively.

4.3.1 Signal quality index for the electrocardiogram

The derivation of SQI_{ECG} followed the work of Li *et al.* [265]. The overall SQI_{ECG} was computed for each ECG beat by combining the following metrics: the agreement of multiple beat detection algorithms ($SQI_{\text{ECG}}^{\text{b}}$); the kurtosis of the waveform ($SQI_{\text{ECG}}^{\text{k}}$); the instantaneous heart rate (IHR) computed using the interval between adjacent R-peaks ($SQI_{\text{ECG}}^{\text{f}}$); and the analysis of the signal-to-noise ratio of the ECG signal ($SQI_{\text{ECG}}^{\text{SNR}}$).

4.3.1.1 Beat detection agreement ($SQI_{\text{ECG}}^{\text{b}}$)

Different QRS detection algorithms are susceptible to different forms of noise. Variations in the detected R-peaks of multiple algorithms can, therefore, indicate the presence of signal artefacts in the ECG signal. $SQI_{\text{ECG}}^{\text{b}}$ employs two different beat detectors: the Pan-Tompkins [119] and Zong *et al.* [261] QRS detectors (described in section 4.2.1). Both algorithms have shown high accuracy in previous studies using the single-channel ECG records in the MIT-BIH Arrhythmia database [266]. $SQI_{\text{ECG}}^{\text{b}}$ was defined as:

$$SQI_{\text{ECG}}^{\text{b}}(k) = \frac{N_{\text{matched}}(w)}{N_{\text{all}}(w)} \quad (4.3)$$

where w is a 30-second window centred on the k^{th} beat. $N_{\text{matched}}(w)$ is the number of beats for which both methods agree, within 150 ms as recommended by ANSI/AAMI [262], and $N_{\text{all}}(w)$ is the number of all beats that both algorithms detected minus $N_{\text{matched}}(w)$.

4.3.1.2 Kurtosis ($\text{SQI}_{\text{ECG}}^k$)

From the central limit theorem, random uncorrelated processes, such as thermal noise, tend to have Gaussian distributions. Kurtosis (which measures the relative "tailedness" of a distribution) provides a metric of how Gaussian-like the ECG trace is and thus the presence of common forms of noise such as muscle artefact or mains interference. The kurtosis ($Kurt$) of an ECG signal segment, x , within the window w of length 30-seconds centered on the k^{th} beat was defined as:

$$Kurt(x) = \frac{1}{N_x} \frac{\sum_{i=1}^{N_x} (x_i - \mu_x)^4}{\sigma_x^4} \quad (4.4)$$

where N_x , μ_x and σ_x are the length, mean and standard deviation of x respectively. $Kurt$ of a Gaussian-like distribution is approximately 3. $Kurt$ of a clean ECG trace in normal sinus rhythm has been shown to exceed 5 [265]. Therefore, $\text{SQI}_{\text{ECG}}^k$ was defined as:

$$\text{SQI}_{\text{ECG}}^k(k) = \begin{cases} 1 & \text{if } Kurt(x) > 5 \\ 0 & \text{otherwise} \end{cases} \quad (4.5)$$

4.3.1.3 Frequency bounding ($\text{SQI}_{\text{ECG}}^f$)

Errors in beat detection resulting from noise can be assessed based on whether the computed IHR falls within a physiological range. The physiological range for HR was taken to be between 30 and 200 beats/min. This range covered the HRs computed by the reference monitors in all datasets outlined in chapter 3. IHR was computed as the time difference between consecutive R-peaks detected using the Pan-Tompkins QRS detector (see section 4.2.1.1). Given $\text{IHR}(k)$ as the instantaneous heart rate of the k^{th} beat, $\text{SQI}_{\text{ECG}}^f$ was defined as:

$$\text{SQI}_{\text{ECG}}^f(k) = \begin{cases} 1 & \text{if } 30 \leq \text{IHR}(k) \leq 200 \\ 0 & \text{otherwise} \end{cases} \quad (4.6)$$

4.3.1.4 Signal to noise ratio ($\text{SQI}_{\text{ECG}}^{\text{SNR}}$)

$\text{SQI}_{\text{ECG}}^{\text{SNR}}$ measured the strength of the signal carrying cardiac information. Since the spectrum of the ECG is concentrated below 14 Hz [267], the ratio of the information in this band to the overall signal provides a measure of the overall signal-to-noise ratio (SNR). Given an ECG signal segment, x , within window w centred on the k^{th} beat, x^{flt} was computed using a Butterworth infinite impulse response (IIR) 4th -order band-pass filter with cut-off frequencies of 0.5 and 14 Hz. The SNR was defined as:

$$\text{SNR}(k) = 10 \log_{10} \frac{\text{Var}(x^{\text{flt}})}{\text{Var}(x - x^{\text{flt}})} \quad (4.7)$$

where $\text{Var}(\cdot)$ computes the variance. A high SNR implies that the signal contains a strong amount of cardiac information. Therefore, $\text{SQI}_{\text{ECG}}^{\text{SNR}}$ was defined as:

$$\text{SQI}_{\text{ECG}}^{\text{SNR}}(k) = \begin{cases} 1 & \text{if } \text{SNR}(k) \geq 10 \\ \frac{\text{SNR}(k)}{10} & \text{if } 0 \leq \text{SNR}(k) \leq 10 \\ 0 & \text{otherwise} \end{cases} \quad (4.8)$$

4.3.1.5 Combined electrocardiogram signal quality index (SQI_{ECG})

The combined ECG SQI, SQI_{ECG} , of the k^{th} beat was calculated as:

$$\text{SQI}_{\text{ECG}}(k) = \begin{cases} \text{SQI}_{\text{ECG}}^{\text{b}}(k) \cdot \text{SQI}_{\text{ECG}}^{\text{f}}(k) \cdot \text{SQI}_{\text{ECG}}^{\text{SNR}}(k) & \text{if } \text{SQI}_{\text{ECG}}^{\text{k}}(k) = 1 \\ \text{SQI}_{\text{ECG}}^{\text{b}}(k) \cdot \text{SQI}_{\text{ECG}}^{\text{f}}(k) \cdot \text{SQI}_{\text{ECG}}^{\text{SNR}}(k) \cdot \eta & \text{if } \text{SQI}_{\text{ECG}}^{\text{k}}(k) = 0 \end{cases} \quad (4.9)$$

where η is a coefficient denoting the presence of statistical noise which was determined to be present whenever the kurtosis of the signal was lower than 5 [265]. The value of η was chosen to be 0.9. Figure 4.1 shows two examples of the computation of SQI_{ECG} from data collected from the same participant in the nABPM study. Figure 4.1(b) was gathered during a transient period of movement from the participant, resulting in a poor agreement between the two beat detectors and as a result, a low overall SQI_{ECG} .

4.3.2 Signal quality index for the photoplethysmogram

The derivation of SQI_{PPG} follows the work of Villarroel *et al.* [263] and Li *et al.* [268]. The overall SQI_{PPG} was computed for each PPG beat by combining the following metrics: an assessment of clipping in the signal ($\text{SQI}_{\text{PPG}}^{\text{c}}$); sharp changes in amplitude ($\text{SQI}_{\text{PPG}}^{\text{a}}$); IHR outside of the physiological bounds ($\text{SQI}_{\text{PPG}}^{\text{f}}$); and the deviation of beats from running-average window template ($\text{SQI}_{\text{PPG}}^{\text{DTW}}$).

4.3.2.1 Clipping detection ($\text{SQI}_{\text{PPG}}^{\text{c}}$)

Clipping generally occurs as a result of motion artefacts. Signal clipping can be detected when the absolute value of the first derivative falls below a given threshold. Given that $N_{\text{length}}(k)$ is the length of the k^{th} PPG beat and $N_{\text{clipped}}(k)$ is the amount of the absolute first derivative of the k^{th} beat that falls below a clipping threshold of 0.1, the $\text{SQI}_{\text{PPG}}^{\text{c}}$ of the k^{th} beat was set to 0 when more than one-third of the first derivative was clipped:

$$\text{SQI}_{\text{PPG}}^{\text{c}}(k) = \begin{cases} 1 & \text{if } N_{\text{clipped}}(k)/N_{\text{length}}(k) \leq 1/3 \\ 0 & \text{otherwise} \end{cases} \quad (4.10)$$

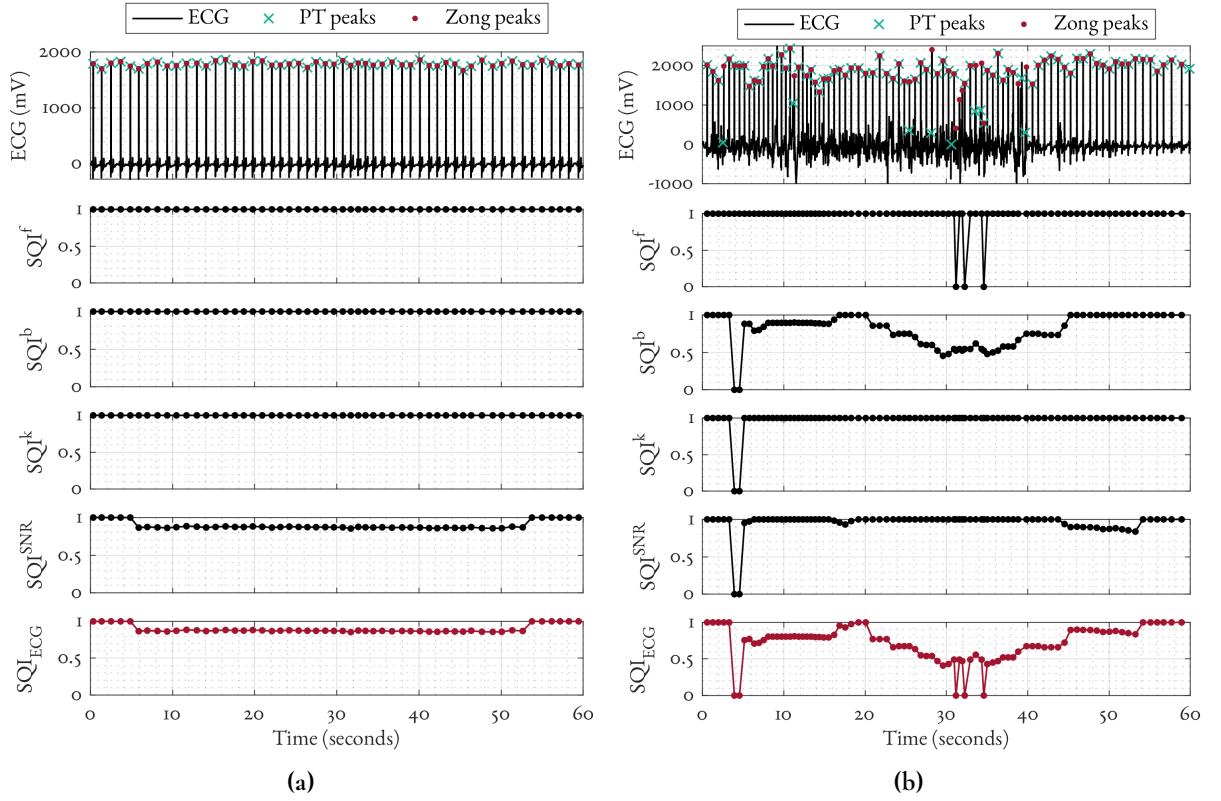


Figure 4.1: Example SQI_{ECG} computation for two 60-second ECG signals. Both examples are from the same participant in the nABPM study. (a) Period of good-quality data where all SQI metrics are high. (b) ECG signal recorded during a transient period of movement, resulting in a poor agreement between the two beat detectors and as a result, a low overall SQI_{ECG} . PT peaks and Zong peaks are ECG R-waves detected using the Pan-Tompkins and Zong *et al.* algorithms respectively.

4.3.2.2 Amplitude thresholding (SQI_{PPG}^a)

Amplitude thresholding was performed to indicate sudden changes in the amplitude of the PPG signal. Given a PPG signal segment, x , within a window, w , of length 30-seconds centred on the k^{th} beat with a corresponding amplitude z_k , $SQI_{PPG}^a(k)$ was defined as:

$$SQI_{PPG}^a(k) = \begin{cases} 1 & \text{if } \mu_x - (3 \cdot \sigma_x) < z_k < \mu_x + (3 \cdot \sigma_x) \\ 0 & \text{otherwise} \end{cases} \quad (4.11)$$

where μ_x and σ_w are the mean and standard deviation of x respectively.

4.3.2.3 Frequency bounding (SQI_{PPG}^f)

Similar to SQI_{ECG}^f , frequency bounding determined whether the IHR fell within a valid physiological range, taken to be between 30 and 200 beats/min as before. For the PPG, IHR was estimated as the time difference between successive beat onsets detected using the BSSF algorithm (see section 4.2.1.2). Given $IHR(k)$ as the instantaneous heart rate of the k^{th} beat, SQI_{PPG}^f was defined as:

$$\text{SQI}_{\text{PPG}}^f(k) = \begin{cases} 1 & \text{if } 30 \leq \text{IHR}(k) \leq 200 \\ 0 & \text{otherwise} \end{cases} \quad (4.12)$$

4.3.2.4 Dynamic Time Warping ($\text{SQI}_{\text{PPG}}^{\text{DTW}}$)

Dynamic time warping (DTW) is a time series technique used to determine a distance (or a degree of similarity) between two signals based on the best possible alignment between the two [269]. Importantly, DTW is able to compute a metric of similarity for a time series whose period changes over time. As a result, DTW was ideal for computing how similar PPG beats were even when the HR changed [268].

The algorithm starts by constructing a beat template, y , over a 30-second window by averaging all beats that fall within the window. Once a template has been computed for a given window, the DTW algorithm was used to compute the distance between the k^{th} PPG beat, x , of length N_x within the current window and the corresponding window template, y , of length M_y . DTW finds the optimum warping path, $p = \{p_1, p_2, \dots, p_L\}$ of length L , between the beat and the template. To achieve this, a $N_x \cdot M_y$ cost function, $c(x, y)$ was constructed whereby each element (i, j) defines the corresponding local distance between $x(i)$ and $y(j)$. The cost has a small value if $x(i)$ and $y(j)$ are similar to each other, and a large value if they are different. DTW was then computed by determining the path, p that minimises this cost as:

$$\text{DTW}(x, y) = \arg \min_p \sum_{l=1}^L c(x_{nl}, y_{ml}) \quad (4.13)$$

where nl and ml define the warping path for x and y respectively.

The classical DTW algorithm is computationally expensive as it needs to evaluate every possible warping path in order to obtain an optimal alignment. Therefore, multi-scale dynamic time warping for pulsatile signals was performed to refine the search space for the optimum alignment between the two time-series. Multi-scale DTW is described in more detail together with its mathematical derivations in [263].

A good-quality beat will have a low DTW value such that it has a similar shape to the average beat in the time window. $\text{SQI}_{\text{PPG}}^{\text{DTW}}$ was defined as:

$$\text{SQI}_{\text{PPG}}^{\text{DTW}}(k) = 1 - \frac{\text{DTW}(x, y)}{100} \quad (4.14)$$

4.3.2.5 Combined photoplethysmogram Signal quality index (SQI_{PPG})

The combined PPG SQI, SQI_{PPG} , of the k^{th} beat was derived by multiplying all the SQI metrics together:

$$SQI_{PPG}(k) = SQI_{PPG}^c(k) \cdot SQI_{PPG}^a(k) \cdot SQI_{PPG}^f(k) \cdot SQI_{PPG}^{DTW}(k) \quad (4.15)$$

Figure 4.2 shows two examples of the computation of SQI_{PPG} for the same participant in the nABPM study.

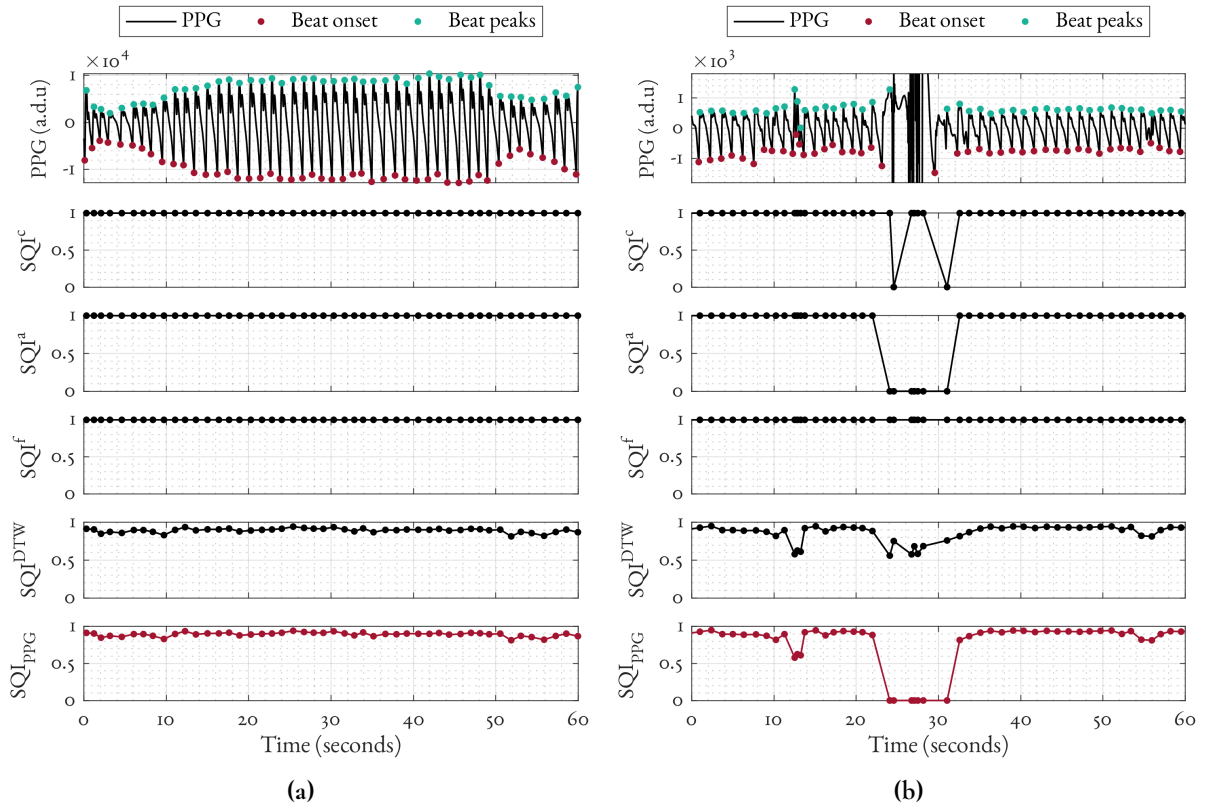


Figure 4.2: Example SQI_{PPG} computation for two 60-second PPG signals. Both examples are from the same participant in the nABPM study. (a) Period of good-quality data where all SQI metrics are high. (b) PPG signal recorded while the participant was on the computer, the sudden spikes are the result of the index finger being used to click on a mouse. This is shown by the sudden change in amplitude which is detected by SQI_{PPG}^a and as a result, a low overall SQI_{PPG} .

4.3.3 Combined signal quality index for the electrocardiogram and photoplethysmogram

During recording periods with synchronous ECG and PPG signals, differences in the measured cardiac frequency can indicate the presence of noise in at least one of the signals. HR values were computed for each signal separately (labelled HR_{ECG} and HR_{PPG} respectively) using the average of the IHR within a sliding window, w , of length 30 seconds and step size of 10 seconds. Only beats of good quality (defined as having an $SQI > 0$ for both signals) were included in the

analysis. If $|\text{HR}_{\text{ECG}}(w) - \text{HR}_{\text{PPG}}(w)|$ was greater than a threshold of 5 beats/min (the ANSI/AAMI maximum HR measurement error for cardiac monitors [270]) then the SQI for all ECG and PPG beats that fall within w was set to 0. Figure 4.3 outlines this process when computed using data from the nABPM study.

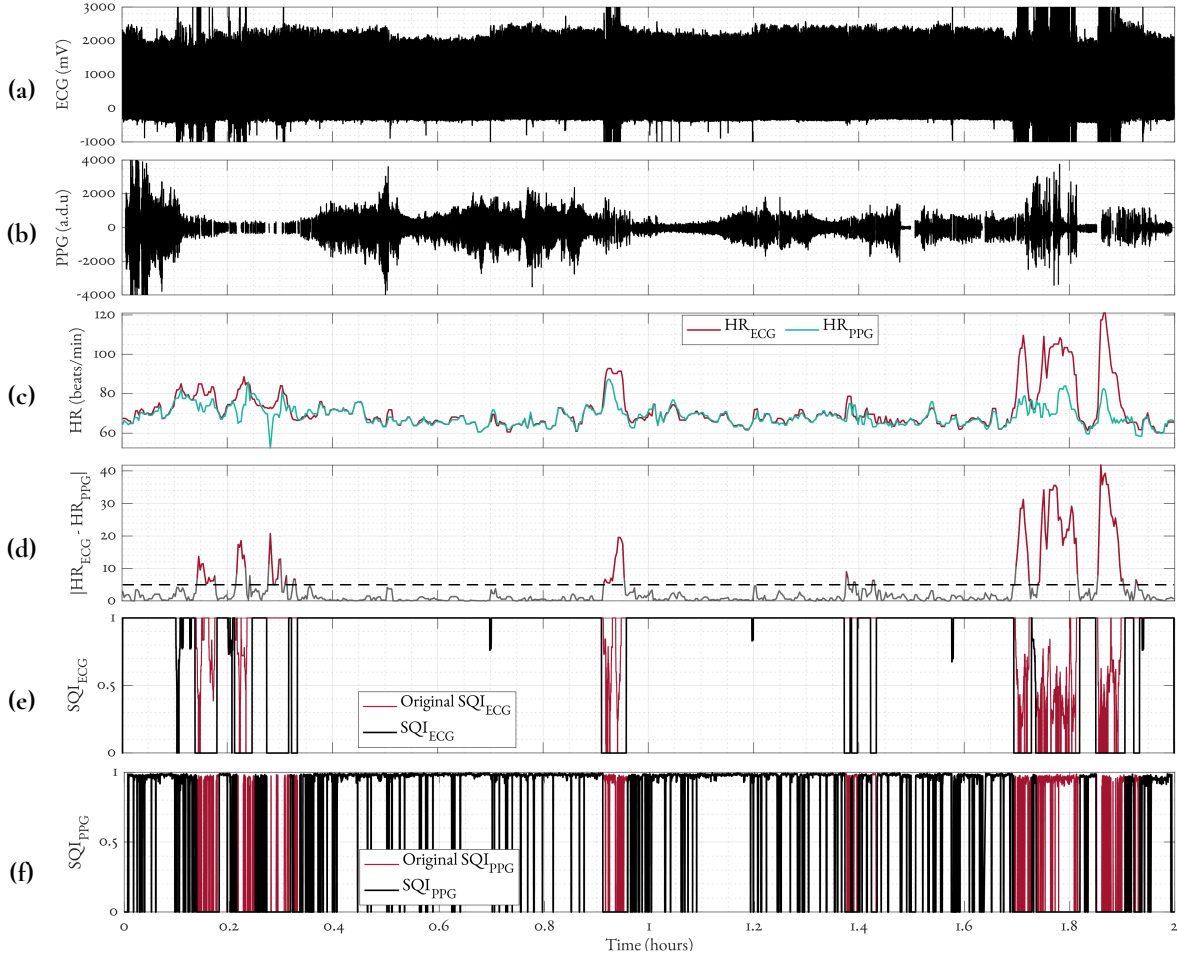


Figure 4.3: Overview of steps for computing the combined SQI for the ECG and PPG. Data comes from a participant in the nABPM study over 2 hours. (b) Recorded ECG signal, periods of poor quality ($\text{SQI}_{\text{ECG}} = 0$) are removed. (a) Recorded PPG signal, periods of poor quality ($\text{SQI}_{\text{PPG}} = 0$) are removed. (c) HR estimates from the two signals are computed using the mean IHR within the window w . (d) The absolute difference between HR_{ECG} and HR_{PPG} . Periods where the absolute difference exceeds the 5 beats/min threshold are shown in red. (e) Updated SQI_{ECG} using the absolute difference between the two HRs. (f) Updated SQI_{PPG} using the absolute difference between the two HRs.

4.4 Non-linear machine learning methods

ML describes the process of estimating a response variable, y , (with N observations), from some measured data, X , (N observations each of dimension M). X can also be known as a feature set with M features. The estimated response variable, $\hat{y}$, is a function of the feature set: $\hat{y} = f(X)$.

The parameters defining $f(\cdot)$ are determined by optimising some loss function that minimises an appropriate error between y and $\hat{y}$ using training data. There are two main machine-learning tasks: regression and classification. In regression, y represents a continuous variable, whereas in classification, y represents a discrete variable. We use x_i to represent the i^{th} observation of all features and x_{ij} to represent the datapoint of the i^{th} observation of the j^{th} feature.

Linear ML models, in which the response is estimated using a weighted linear combination of the predictor variables, are outlined in appendix D. In non-linear modelling, the response is estimated using a non-linear combination of the predictor variables. Many different non-linear algorithms exist, however in this thesis, we focus on random forests (RFs), which are formed of a large ensemble of decision trees (see figure 4.4).

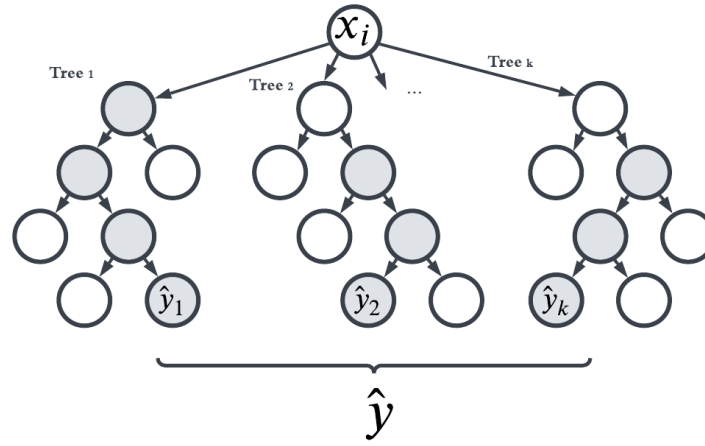


Figure 4.4: Depiction of a random forest model constructed from a large ensemble of 'weak' decision trees. Each decision tree is trained on a bootstrap sample of the total features in the dataset. The predictions from all trees $\{\hat{y}_1, \hat{y}_2, \dots\}$ are then aggregated to form the final prediction $\hat{y}$.

4.4.1 Decision trees

A single decision tree consists of a series of nodes and branches arranged in a hierarchical structure. From the first 'parent' node a binary decision is made regarding a particular feature input. For example, the splitting criterion may be whether a particular feature is greater than 10. From each node, two branches are formed which represent alternative responses to the splitting criterion. At the end of each branch is a child node with its own splitting criterion. These child nodes in turn become parent nodes and so on until reaching a final set of terminating leaf nodes.

Training a decision tree defines the different splitting criteria at each node as well as the tree structure (how many nodes). For the n^{th} node, the splitting criterion is defined as the one that maximises the decrease of some impurity measure $i(n)$. This decrease is computed as:

$$\Delta i(n) = i(n) - p(n_L)i(n_L) - p(n_R)i(n_R) \quad (4.16)$$

where n_L and n_R indicate the children nodes of n on the left and right branches. $p(n) = N(n)/N$ is the proportion of the total number of data points (N) at node n . For classification tasks, Gini impurity [271] is used as $i(n)$. For regression tasks, the summed squared error is used [271]. A tree is typically expanded until all leaves are pure (i.e each node can only have one possible class in it) or until all leaves contain fewer than the minimum number of samples in a partition required to split a node.

4.4.2 Random forest

Despite techniques such as pruning, which reduce the complexity of a tree, decision trees are susceptible to overfitting [272, 273]. RFs, therefore, build a large ensemble of 'weak' decision trees and aggregate the results of all of the trees to make the final prediction. Each individual tree is trained on a bootstrap sample (sampling with replacement) of features. For each node, the optimum splitting criterion is determined from a random subset of the features available. Once all trees have been grown, the results of the 'weaker' trees are aggregated to make the final prediction, $\hat{y}$. Classification predictions are formed as the majority predicted class label. Regression predictions are formed as the mean of the predicted response across all trees. Typically, RF models are not very sensitive to the choice of the number of trees (labelled N_{trees}), provided it is sufficiently high [271]. In this thesis, N_{trees} was kept at 300. However, the number of features randomly selected for each node (labelled $mtry$) was optimised for each problem addressed in the thesis.

4.4.3 Mixed effects random forest

In cases involving repeated measures (such as multiple tests taken from the same group/subject), there is likely to be an inter-dependency of the observations on the particular groups. An extension of the RF model that accounts for dependence between groups of data is the mixed effects random forest (MERF) model introduced by Hajjem *et al.* in 2014 [274]. Mixed effect random forest models are so called as they separate the terms into fixed-effects (which are equivalent to a non-linear RF, $f(X)$, outlined above), and random-effects (which denote the additional group-specific variation). This approach allows each group, g , to have its own intercept (u_0^g) and slope (u_1^g) for features that are defined to be random effects:

$$y_i^g = \underbrace{f(x_i^g)}_{\text{fixed-effect}} + \underbrace{u_0^g + u_1^g x_{i1}^g + u_1^g x_{i3}^g + \dots + \epsilon_i^g}_{\text{random-effect}} \quad (4.17)$$

where u_i defines the random effects. The superscript g denotes all observations in the g^{th} group cluster. Note in equation (4.17) that not all features in X have random effects. Indeed, features

that are not included in X may also be included in the random effect structure. In matrix notation, this can be written as:

$$y^g = f(X^g) + Z^g u^g + \epsilon^g \quad (4.18)$$

where u is the vector of q random effects, and Z is the design matrix defining the random effects. u^g and ϵ^g are assumed to follow a Normal distribution:

$$\begin{bmatrix} u^g \\ \epsilon^g \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & \sigma_\epsilon^2 I \end{bmatrix}\right) \quad (4.19)$$

where G is the covariance matrix of the random effect parameters.

An expectation-maximisation (EM) technique is used to fit MERF with the aim of determining estimates for $f(\cdot)$, u_i , G , and σ_ϵ^2 . EM is an iterative alternative optimisation procedure where for each iteration one parameter is optimised in turn while the others remain fixed. Full details of the EM algorithm are provided in [274]; here we briefly overview the main steps.

The initial conditions are set as $u^g = 0$, $G = I$, and $\sigma_\epsilon^2 = 1$. For each iteration, the EM steps are:

- Fix u^g . Compute $y_*^g = y^g - Z^g \hat{u}^g$. Fit a RF, $f(X)$ with y_*^g as the corresponding training set responses.
- Fix $f(\cdot)$, G , and σ_ϵ . Optimise to find u^g . There is a closed-form solution.
- Fix $f(\cdot)$, and u^g . Optimise to find G and σ_ϵ . Again, there is a closed-form solution.

EM is iterated until the maximum number of iterations is met. In this thesis, we set the maximum number of iterations to be 300 as recommended [274].

To predict the response for a new observation i that belongs to a cluster g among those used to fit the MERF model, we use both its corresponding population-averaged RF prediction, $f(x_i^g)$, and the predicted random part corresponding to its cluster, $Z^g u_i^g$.

4.5 Shapley additive explanation values

ML model interpretation is typically performed by establishing the importance of each feature towards model predictions. We assessed the importance of each feature in our models through SHAP values [275]. SHAP values, ϕ_j , represent the marginal contribution of each feature to individual model predictions. For a given ML model $f(\cdot)$, the marginal contribution of the j^{th} feature is computed through the difference in model outputs when trained with that feature compared to when trained without it, $f(z_i \cup \{x_{ij}\}) - f(z_i)$, where z_i is a subset of features in x_i not including x_{ij} ($z_i \subseteq x_i \setminus \{x_{ij}\}$). Since the impact of adding a feature can depend on the

other features in the model, ϕ_j is computed as the weighted sum of marginal contributions of the j^{th} feature over all possible subsets $z_i \subseteq \{x_{i1}, \dots, x_{iM}\} \setminus \{x_{ij}\}$:

$$\phi_j(f, x_i) = \sum_{z_i \subseteq x_i \setminus \{x_{ij}\}} \frac{|z_i|!(M - |z_i| - 1)!}{M!} (f(z_i \cup \{x_{ij}\}) - f(z_i)) \quad (4.20)$$

where $|z_i|$ is the number of features in z_i . As most models cannot handle missing data, SHAP values are computed by replacing $f(\cdot)$ in equation (4.20) with $f_x(\cdot)$, a conditional expectation function of the original model. This step equates to replacing each missing feature with a random value representative of the data over which the original model was trained. Repeating this process, and averaging the results integrates out the missing value. It is demonstrated in [275] that SHAP values computed in this manner correspond to the only method of assessing feature importance that satisfies three desirable properties known as local accuracy, missingness, and consistency.

A result of these properties is that for a complex, non-linear, model such as RF, a simpler explanation model $f_e(\cdot)$ can be approximated through the linear combination of SHAP values across all features:

$$f_e(x_i) = \phi_0 + \sum_{j=1}^M \phi_j \quad (4.21)$$

where ϕ_0 is the expected model value over the training set. Equation (4.21) allows for local interpretability of model outputs by examining each feature's SHAP value for a given estimation. As SHAP values are computationally expensive to estimate (there are 2^M distinct permutations of feature values), typically approximations of SHAP values are computed through KernelSHAP or TreeSHAP (for tree-based models) implementations. These techniques were implemented using the SHAP Python library [275, 276].

5

Pulse arrival time as a surrogate measure of blood pressure

Contents

5.1	Introduction	97
5.2	Methods overview	98
5.3	Computing reference blood pressure values	98
5.4	Estimating pulse arrival time	100
5.4.1	Pulse arrival time signal quality index	101
5.4.2	Kalman filter	102
5.5	Estimating pre-ejection period	104
5.5.1	Impedance cardiogram signal quality	104
5.5.2	Estimating beat-by-beat pre-ejection period	106
5.5.3	Pre-ejection period signal quality	107
5.6	Estimating blood pressure using transit times	109
5.6.1	Time alignment	109
5.6.2	Averaging measurements	110
5.6.3	Models for the estimation of blood pressure	111
5.6.4	<i>Baseline reference</i>	114
5.6.5	Model evaluation	114
5.7	Results	115
5.7.1	Overall response	115
5.7.2	Estimating blood pressure using transit times	117
5.8	Discussion	120
5.8.1	The relationship between pulse arrival time and blood pressure	120
5.8.2	Impact of pre-ejection period	122
5.8.3	Estimating blood pressure using transit times	123
5.8.4	Limitations	124
5.9	Conclusion	125

5.1 Introduction

One of the most promising cuffless surrogate measures of BP is the PAT, computed as the time delay between characteristic fiducial points of the ECG and the PPG signals. However, there are several potential limitations to its use in BP monitoring. PAT includes the PEP component, which could make PAT an unreliable surrogate of BP. Additionally, it remains unclear whether the requirement for *individual calibration* can be circumvented using population-based, *hybrid calibration* strategies [9]. These issues are explored in depth in this chapter.

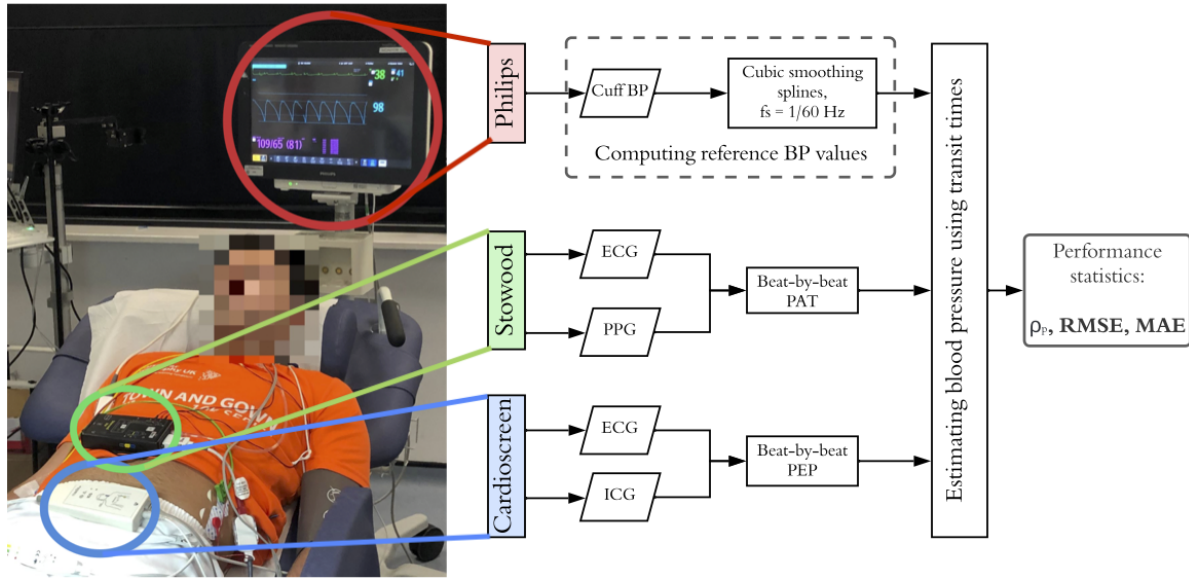


Figure 5.1: Schematic overview of the methodology to estimate blood pressure using transit times. Data was gathered from the 26 participants of the MOLLIE study in which three different devices (Philips, Stowood, and Cardioscreen) were used to estimate the reference BP signal, PAT, and PEP respectively. Finally, we estimated BP values using the TTs.

5.2 Methods overview

Data was gathered from the 26 participants in the MOLLIE clinical study (see section 3.3) for whom BP changes were induced by the infusion of phenylephrine. Figure 5.1 provides a schematic overview of the methodology employed in this chapter. BP values were recorded by a Philips patient monitor using a sphygmomanometer cuff wrapped around the upper left arm of the participant. The BP time series were subsequently smoothed and interpolated, using a cubic smoothing splines algorithm, to mitigate the effects of noise and to provide reference BP values at a rate of once per minute. Beat-by-beat PAT values were computed using the ECG and PPG waveforms recorded by the Stowood device. The algorithm developed as part of this thesis research focuses on detecting and handling periods of low signal quality which are often overlooked in the literature. Beat-by-beat PEP estimates were computed using the ECG and thoracic bioimpedance (Z) signals recorded by the Cardioscreen monitor. PTT was computed as the difference between PAT and PEP. Finally, estimates of BP were computed using the TT values.

5.3 Computing reference blood pressure values

Measurements of SBP, MAP, and DBP using sphygmomanometer cuffs have known limitations depending on posture, cuff-inflation hypertension, and cuff size [105, 277]. The sphygmomanometer cuff used in our study was programmed to inflate once per minute. However, errors in cuff inflation

prevented the Philips device from registering an accurate estimate on every occasion, resulting in missed data points in the recorded BP time series. Data recorded from the cuff was, therefore, both noisy and sparse. In order to reduce the impact of these sources of error, we processed the cuff data using a cubic smoothing splines algorithm [278]. This allowed for both filtering and interpolation of the noisy blood pressure readings to generate a smoothed time series at a sampling frequency (f_{BP}) of 1/60 Hz (once per minute).

For each participant, let the i^{th} BP observation, y_i , at time t_i be modelled by the relation:

$$y_i = f(t_i) + \epsilon_i, \quad i = \{1, \dots, N_{\text{meas}}\} \quad (5.1)$$

where N_{meas} is the number of BP measurements recorded for a participant. ϵ_i forms a sequence of independently distributed random variables with zero mean. Cubic smoothing splines define an estimate, $\hat{f}$, of f that equates to a cubic spline with knots (transition points) at $f(t_i)$. At these transition points, the values of $\hat{f}$, $\hat{f}'$, and $\hat{f}''$ (where f' and f'' denote the first and second derivative of f respectively) all match. The exact form of $\hat{f}$ is determined by minimising a loss L^{BP} :

$$L^{\text{BP}} = \lambda \sum_{i=1}^{N_{\text{meas}}} (y_i - \hat{f}(t_i))^2 + \int_{t_1}^{t_{N_{\text{meas}}}} \hat{f}''(t)^2 dt \quad (5.2)$$

The first term of equation (5.2) imposes a penalty for the squared distance between observed BP values, y_i , and smoothed values $\hat{f}(t_i)$. The second term imposes a penalty for the complexity of $\hat{f}$, modelled as the accumulated second derivative of $\hat{f}$. λ is a constant that defines the relative weight placed on minimising the residual sum of squares against the complexity of $\hat{f}$. A very low value of λ will result in the regressed function converging to a linear least squares estimate. A very high value of λ will result in the smoothing spline converging to a cubic spline that passes through all data points.

As all participants were under the same protocol, we implemented a λ value for SBP, MAP, and DBP (λ_{SBP} , λ_{MAP} , and λ_{DBP} respectively) that was common among participants. Each respective λ value was determined by extending the ordinary cross-validation strategy proposed in [279] by a grid search across the log-scaled range $[10^{-3}, \dots, 10^8]$. For each participant, the leave-one-out (LOO) RMSE was computed across the entire λ range. The λ value that minimised the participant-wise average LOO error was implemented. Figure 5.2 shows the original observations of the BP cuff data and the processed BP data sampled at 1/60 Hz (once per minute), for a typical participant in the study.

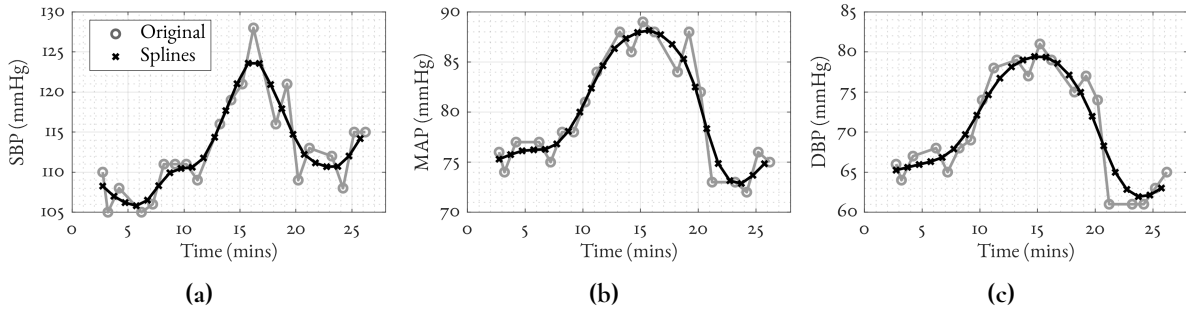


Figure 5.2: Computing reference BP measurements using cubic smoothing splines, (a)-(c) for SBP, MAP and DBP respectively. The original data was both noisy and sparse. Our proposed algorithm was able to provide a smoother profile and interpolate the missing data to provide a reference BP measurement once per minute.

5.4 Estimating pulse arrival time

PAT was estimated using the ECG and PPG signals recorded by the Stowood device. The R-peak (point of left ventricular depolarisation) of the ECG waveform and the foot of the PPG waveform (recorded using the pulse oximeter attached to the participant's finger) were used as the proximal and distal sites respectively. For both signals, accurate segmentation and labelling of the cardiac beats was required. Additionally, assessing the quality of the ECG and PPG signals (SQI_{ECG} and SQI_{PPG} respectively) was important as data corrupted by noise or motion artefacts is often difficult to avoid during continuous recordings.

To suppress the impact of baseline wander, the ECG was filtered using an 8th-order Butterworth infinite impulse response (IIR) high-pass filter with a cut-off frequency of 0.5 Hz. To suppress power-line interference, a 2nd-order IIR notch filter with a centre frequency of 50 Hz (the frequency of mains power in the UK) was used. The QRS complex of the ECG was detected using the Pan-Tompkins algorithm [119] (see section 4.2.1.1). The quality of the ECG, SQI_{ECG} , was computed using the methods outlined in section 4.3.1.

The PPG signal was filtered using an 8th-order Butterworth IIR band-pass filter with cut-off frequencies of 0.5 Hz and 10 Hz as recommended in [31]. Cardiac beat onsets were detected from the PPG using the Boxed Slope Sum Function outlined in section 4.2.2 [263]. The intersecting tangents method was used to define the location of the foot of the PPG, as is done commonly in the literature [151, 182, 280]. This point is defined by the intersection of the tangent to the maximum gradient of the upslope of the PPG and the horizontal tangent from the onset value of the waveform. The quality of the PPG signal, SQI_{PPG} , was computed using the methods outlined in section 4.3.2. As both the ECG and the PPG signals were recorded simultaneously, SQI_{ECG} and SQI_{PPG} were further adjusted with the combined SQI outlined in section 4.3.2.5.

PAT was estimated as the time delay between a proximal ECG R-peak and a distal PPG foot prior to the following ECG R-peak. If there were either none or more than 1 PPG beat located between two successive R-peaks, this indicated an error in the beat detection algorithm (for either the ECG or the PPG) and no PAT value for that beat was computed.

5.4.1 Pulse arrival time signal quality index

The quality of the PAT signal will naturally depend on the degree of confidence of the detected proximal ECG R-peak and the distal PPG foot. The ECG waveform morphology may be corrupted by electrical activity of the muscles or power line interference [121]. The PPG waveform may be corrupted by motion artefacts caused by movement of the fingers. Additionally, as PAT was computed from two independent waveforms located separately on the body, the susceptibility to noise due to (but not limited to) motion artefacts, was high.

Despite this, a significant proportion of work described in the literature does not document any attempt to reduce the impact of low signal quality on PAT estimates. In this work, the SQI for each PAT beat (SQI_{PAT}) was calculated as a combination of outlier detection from the PAT time series and the SQI metrics of the proximal ECG beat and distal PPG pulse.

5.4.1.1 Outlier detection (SQI_{PAT}^o)

Outliers in the PAT time series can result from noise, errors in the beat detection algorithms used, motion artefacts or other errors arising from issues relating to sensor attachment. Outliers were detected by constructing a moving-average window, w , of size 30 seconds and step 25 seconds, in which each PAT beat was given a score based on its deviation from the average of the window. To mitigate the influence of outliers, the average of the window was computed as the median, med_w , and the deviation was computed as the median absolute deviation, MAD_w [281, 282]. The outlier SQI (SQI_{PAT}^o) was calculated for the k^{th} PAT beat in the sliding window as:

$$SQI_{PAT}^o(k) = \begin{cases} 1 & \text{if } med_w - (3 \cdot MAD_w) < PAT_k < med_w + (3 \cdot MAD_w) \\ 0 & \text{otherwise} \end{cases} \quad (5.3)$$

5.4.1.2 Combined pulse arrival time signal quality index (SQI_{PAT})

The SQI of each PAT beat was subsequently computed as:

$$SQI_{PAT}(k) = SQI_{PAT}^o(k) \cdot \min(SQI_{ECG}(k), SQI_{PPG}(k)) \quad (5.4)$$

5.4.2 Kalman filter

Kalman filtering was applied to adjust beat-by-beat PAT values based on their signal quality, as previously described by Zhang *et al.* [167]. The Kalman filter is an algorithm to compute an estimate of the current state, x_k (representing the true, underlying, PAT signal), at time point k of a linear dynamic system as a weighted average of the system's previous state estimate and the current measurement, z_k (representing the estimated PAT values, which may contain a significant degree of noise). This reduces the impact of transient changes from noise and motion artefacts. The Kalman filter is a recursive estimator consisting of two stages: the prediction stage and the update stage. Full details of the Kalman filter algorithm can be found in [169]. We implemented a Kalman filter for a one-dimensional signal (the PAT time series), following the work of Li *et al.* [265] and Tarassenko *et al.* [283]. The prediction and update equations were:

Prediction stage:

$$\hat{x}_k^- = \hat{x}_{k-1} \quad (5.5)$$

$$P_k^- = P_{k-1} + Q \quad (5.6)$$

Update stage:

$$\mathbb{K}_k = P_k^- (P_k^- + R_k)^{-1} \quad (5.7)$$

$$\hat{x}_k = \hat{x}_k^- + \mathbb{K}_k (z_k - \hat{x}_k^-) \quad (5.8)$$

$$P_k = (1 - K_k) P_k^- \quad (5.9)$$

where $\hat{x}_k^-$ and $\hat{x}_{k-1}$ are the prior and posterior state estimates for the k^{th} and $(k^{\text{th}} - 1)$ PAT beat respectively. P_k^- and P_k are the prior and posterior error variances respectively. $\mathbb{K}_k$ is the Kalman gain that minimises the posterior error variance. Q is the state noise variance, R_k is the measurement noise variance.

Equation (5.5) states that we assume the PAT estimate at each time step is equal to the estimate at the previous time step. Variation in the PAT time series was then modelled by assigning a value of 1 to Q . Similar to [263], in order to assign a higher weight to PAT estimates of good-quality, R_k was modified using the equation:

$$R_k = \exp\left(\frac{1}{\text{SQI}_{\text{PAT}}(k)^2} - 1\right) \quad (5.10)$$

In this instance, each new measurement of beat-by-beat PAT, z_k , was updated based on its respective $\text{SQI}_{\text{PAT}}(k)$. Figure 5.3 provides an overview of PAT processing, including the effect of Kalman filtering on two different time scales. As the participants remained supine and still for the duration of the study, the signal quality indices remained high with occasional periods of 0 often caused by a participant's finger twitching.

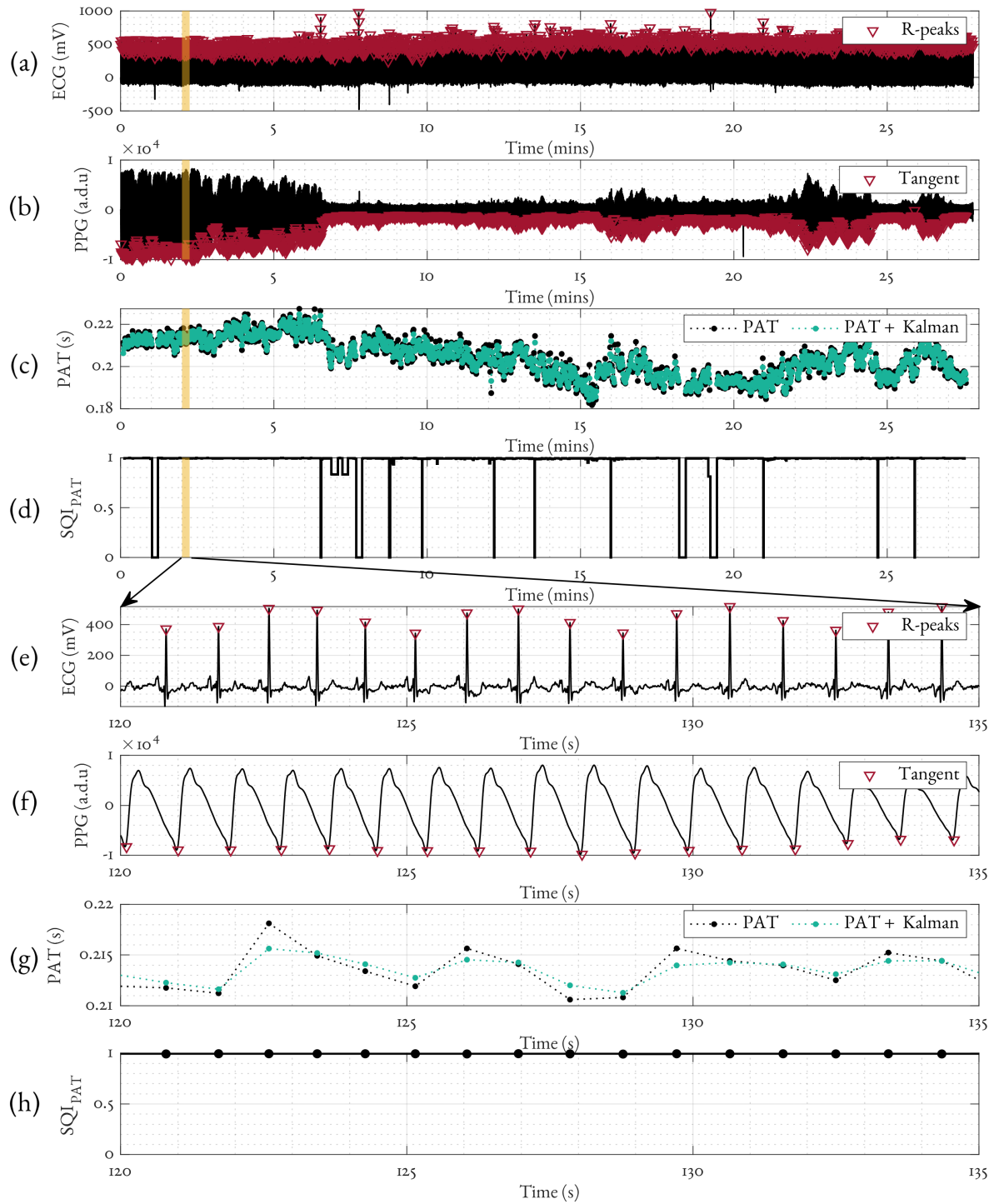


Figure 5.3: Overview of the steps for computing beat-by-beat PAT time series across a typical 28 minutes of data recorded in the MOLLIE study. (a) ECG time series and R-peaks. (b) PPG time series and tangent foot locations. (c) Computed beat-by-beat PAT and the impact of including Kalman filtering. (d) SQI_{PAT} . (e)-(h) repeats these plots for a zoomed-in 15-second portion of data (highlighted in yellow in (a)-(d)).

5.5 Estimating pre-ejection period

PEP was estimated using the ECG and bioimpedance (Z) signals recorded by the Cardioscreen monitor. The ICG signal was computed from Z using a 6th-order Savitzky-Golay derivative filter.

Savitzky-Golay filters [284] are digital smoothing filters computed by fitting a q^{th} degree polynomial to sliding windows of data. The Savitzky-Golay filter can additionally provide a smooth approximation for the derivative of the signal, obtained by differentiation of the fitted polynomial rather than the original data [285]. Their smoothing properties make Savitzky-Golay filter derivatives preferable to point differences when there is a high level of noise present in the signal [286]. When the data points are evenly spaced, an analytical solution can be found such that the derivative of the smoothed signal, $\tilde{y}'$, can be determined through convolution of the original signal, y , with the filter coefficients, C .

$$\tilde{y}'_j = \sum_{i=-(m-1)/2}^{(m+1)/2} C_i \cdot y_{j+i}, \quad j = \left\{ \frac{m+1}{2}, \dots, N_{\text{sig}} - \frac{m+1}{2} \right\} \quad (5.11)$$

where N_{sig} is the number of datapoints in the signal, m is the filter window size equal to $q + 1$. In practice, q is chosen as a trade-off between noise reduction and resolution loss. High values of q are more effective at removing noise, but at the expense of distorting small details too much [287].

For a 6th-order first derivative, the filter coefficients are [284]:

$$C = \frac{1}{21} [-2, 3, 6, 7, 6, 3, -2] \quad (5.12)$$

As is commonly done in the literature [288, 289], to remove the influence of high-frequency noise (such as mains interference) and low-frequency drift (such as respiration), the ICG signal was band-pass filtered using a 4th-order Butterworth IIR filter with cut-off frequencies set at 0.5 Hz and 40 Hz. The location of the R-peaks and beat-by-beat SQI of the ECG signal were computed using the steps outlined in the previous sections describing PAT estimation. The location of the R-peaks of the ECG defined the onsets of the beats in the ICG signal.

5.5.1 Impedance cardiogram signal quality

As the ICG signal is particularly vulnerable to noise, such as motion [288], an adequate signal quality metric was vital for PEP estimation. The overall SQI for the ICG (SQI_{ICG}) was computed for each beat by combining two metrics: activity index ($\text{SQI}_{\text{ICG}}^{\text{Act}}$), and the deviation from an ensemble average ($\text{SQI}_{\text{ICG}}^{\text{EA}}$).

5.5.1.1 Activity index ($\text{SQI}_{\text{ICG}}^{\text{Act}}$)

The activity index (Act), initially proposed by Forouzanfar *et al.* [290], indicates variations in the ICG waveform over a cardiac cycle. For the k^{th} ICG beat, ψ , of length N_ψ , Act was defined as:

$$Act_k = \sqrt{\frac{1}{N_\psi} \sum_{j=1}^{N_\psi} \left(\frac{\psi_j}{C_a} - \mu_k \right)^2} \quad (5.13)$$

where C_a was defined as the median amplitude of all ICG beats in the recording. μ_k was computed as:

$$\mu_k = \frac{1}{N_\psi} \sum_{j=1}^{N_\psi} \left(\frac{\psi_j}{C_a} \right) \quad (5.14)$$

Very low values of Act indicate ICG beats with very little variation, likely to be resulting from sensor disconnect. Very high values of Act indicate ICG beats with significant variation, likely to be resulting from sudden motion artefacts. Forouzanfar *et al.* [290] defined an acceptable range for Act . $\text{SQI}_{\text{ICG}}^{\text{Act}}$ was computed as:

$$\text{SQI}_{\text{ICG}}^{\text{Act}}(k) = \begin{cases} 1 & \text{if } 0.1 < Act_k < 0.4 \\ 0 & \text{otherwise} \end{cases} \quad (5.15)$$

5.5.1.2 Ensemble average ($\text{SQI}_{\text{ICG}}^{\text{EA}}$)

$\text{SQI}_{\text{ICG}}^{\text{EA}}$ was initially proposed by Sheikh *et al.* [289] to remove ICG beats that significantly deviated in morphology from an ensemble-averaged ICG beat (ICG_{EA}) computed across a window. A four-stage strategy was used to construct the ICG_{EA} within a window of length 30 seconds, centred on each ICG beat. At each stage, ICG beats located within the window that passed the stage criteria were first resampled to be 100 samples long using cubic-splines interpolation, and then subsequently averaged to compute the ICG_{EA} .

For the first stage, all ICG beats for which $\text{SQI}_{\text{ICG}}^{\text{Act}} = 0$ were removed from the analysis. In the second stage, the ICG_{EA} was computed using valid beats (from the previous stage) whose amplitude was within $\pm$ three standard deviations for that window. In the third stage, the cross-correlation (R_{xy}) was used to measure the similarity between each beat and the ICG_{EA} .

R_{xy} measures the similarity between two time series, x and y , as a function of the delay, t , of one signal relative to the other.

$$R_{xy}(t) = \frac{\sum_{i=1}^{N_o} N_o (x_i - \mu_x)(y_{i-t} - \mu_y)}{(N_o - 1)\sigma_x\sigma_y} \quad (5.16)$$

where μ_x and σ_x are the mean and standard deviation of x respectively (and similarly for μ_y and σ_y). N_o is the number of overlapping data points. The cross-correlation may be used to

estimate the time delay, τ_{lag} , between two signals as the value of t that maximises (or minimises in cases where a negative correlation is expected) $R_{xy}(t)$:

$$\tau_{\text{lag}} = \underset{t}{\operatorname{argmax}}(R_{xy}(t)) \quad (5.17)$$

Any ICG beat with $|\tau_{\text{lag}}| > 50\text{ms}$, or $R_{xy}(0) < 0.5$ was discarded. At the final stage, the τ_{lag} of each beat with ICG_{EA} was computed as before. Any beat with $R_{xy}(\tau_{\text{lag}}) < 0.8$ with the ICG_{EA} was removed from the analysis.

The SQI_{ICG}^{EA} was then computed for the k^{th} ICG beat as:

$$\text{SQI}_{\text{ICG}}^{\text{EA}}(k) = \frac{|\Omega|}{N_w} \quad (5.18)$$

where N_w is the total number of ICG beats initially in w (i.e. prior to stage 1), Ω is the subset of ICG beats that were ultimately not discarded from analysis post stage 4.

5.5.1.3 Combined ICG signal quality index (SQI_{ICG})

The overall SQI for the ICG was subsequently computed as:

$$\text{SQI}_{\text{ICG}}(k) = \text{SQI}_{\text{ICG}}^{\text{Act}}(k) \cdot \text{SQI}_{\text{ICG}}^{\text{EA}}(k) \quad (5.19)$$

5.5.2 Estimating beat-by-beat pre-ejection period

PEP was estimated as the time difference between the R-peak of the ECG and the B-point of the ICG. As discussed in section 2.7, the exact location of the B-point has proven in prior literature to be extremely difficult to automatically detect from ICG beats. A number of different algorithms have been proposed for automatic B-point detection [290, 291, 292]. In appendix E.2, we provide an overview of three of the most popular algorithms and conclude that the Lozano method [292] provided the most robust estimates of PEP in our dataset. This method is commonly applied in the literature [291, 293, 294] as it avoids the complications associated with detecting the B-point of the ICG through a data-driven polynomial relationship which estimates PEP using the time delay from the R-peak to the C-point (the maximum of the ICG beat). Using the Lozano method, PEP was defined using the following quadratic relationship:

$$\text{PEP} = 1.233 \cdot t_{RC} - 0.0032 \cdot t_{RC}^2 - 31.59 \quad (5.20)$$

where t_{RC} is the time between ECG R-peak and ICG C-point.

5.5.3 Pre-ejection period signal quality

The signal quality index of the PEP signal, SQI_{PEP} , was computed by a combination of outlier detection (SQI_{PEP}^o), physiological thresholding (SQI_{PEP}^{phys}), and the SQI metrics of the proximal ECG beat and distal ICG beat.

5.5.3.1 Outlier detection (SQI_{PEP}^o)

The outlier SQI (SQI_{PEP}^o) was calculated for the k^{th} PEP beat in a moving-average window, w , of size 30 seconds and step 25 seconds, as:

$$SQI_{PEP}^o(k) = \begin{cases} 1 & \text{if } \text{med}_w - (3 \cdot \text{MAD}_w) < \text{PEP}_k < \text{med}_w + (3 \cdot \text{MAD}_w) \\ 0 & \text{otherwise} \end{cases} \quad (5.21)$$

where med_w and MAD_w are the median and median absolute deviation of w respectively.

5.5.3.2 Physiological range (SQI_{PEP}^{phys})

The physiological range for PEP was taken to be between 50 and 160 ms following the work of Cybulski *et al.* [295]. SQI_{PEP}^{phys} was, therefore, defined for the k^{th} beat as:

$$SQI_{PEP}^{phys}(k) = \begin{cases} 1 & \text{if } 50\text{ms} < \text{PEP}_k < 160\text{ms} \\ 0 & \text{otherwise} \end{cases} \quad (5.22)$$

5.5.3.3 Combined pre-ejection period signal quality index (SQI_{PEP})

The overall SQI for PEP was computed as:

$$SQI_{PEP}(k) = SQI_{PEP}^o(k) \cdot SQI_{PEP}^{phys}(k) \cdot \min(SQI_{ECG}(k), SQI_{ICG}(k)) \quad (5.23)$$

Similar to the PAT time series, the PEP time series was processed by a Kalman filter using the SQI_{PEP} of the beats to weight the measurement noise, R_k :

$$R_k = \exp\left(\frac{1}{SQI_{PEP}(k)^2} - 1\right) \quad (5.24)$$

Figure 5.4 provides an overview of PEP processing steps, on two different time scales. Compared to the PAT time series (see figure 5.3), the PEP time series was often of lower quality due to the ICG's relatively higher susceptibility to noise (compared to the PPG and ECG). During sustained periods of poor signal quality (for example minute 16 in figure 5.4(e)), the Kalman filter places a higher emphasis on previous state estimates, reducing the effect of transient changes.

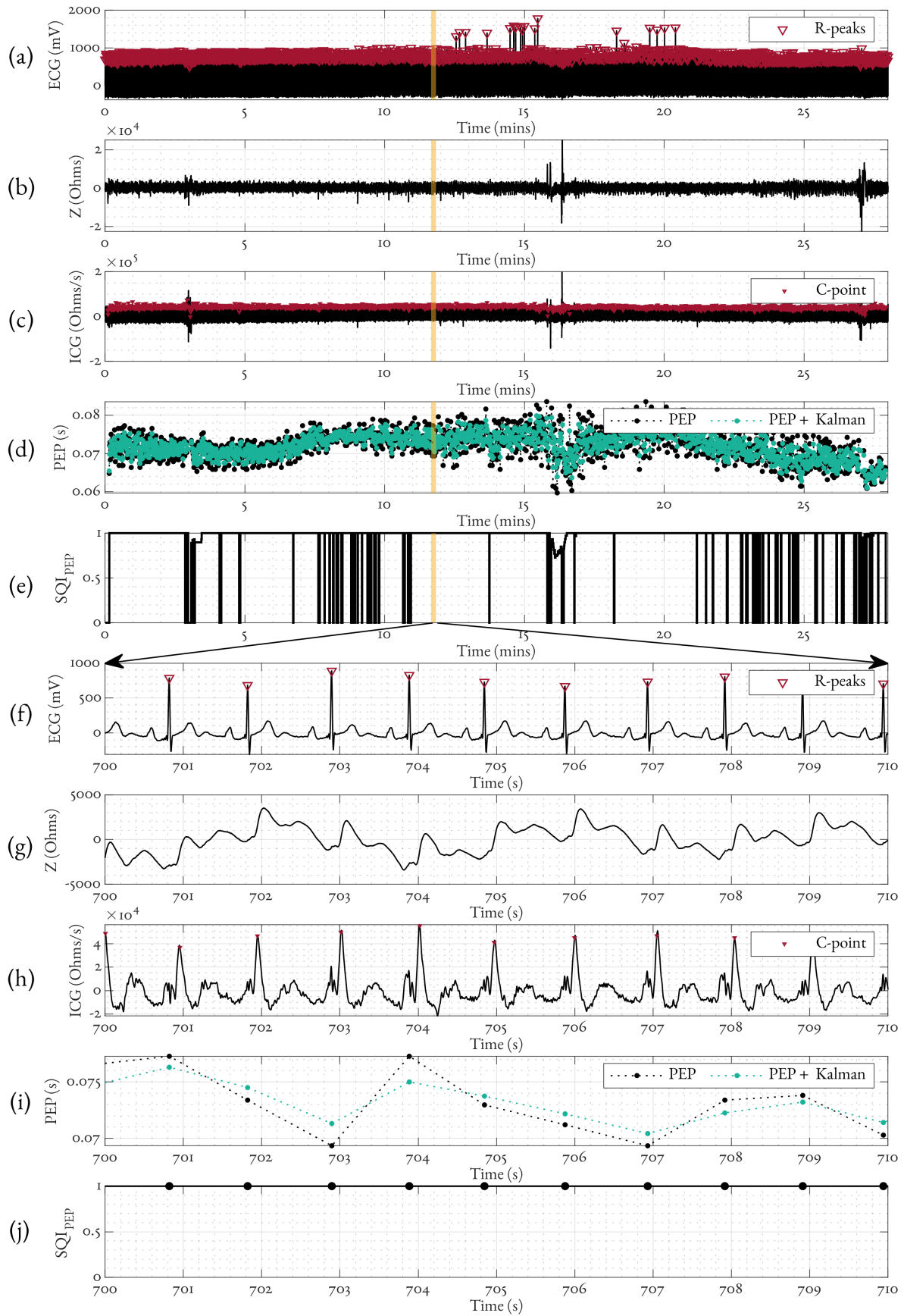


Figure 5.4: Overview of the steps for computing beat-by-beat PEP time series across a typical 28 minutes of data recorded in the MOLLIE study. (a) ECG time series and R-peaks. (b) Bio-impedance signal. (c) ICG time series and C-points. (d) Computed beat-by-beat PEP and the impact of including Kalman filtering. (e) SQI_{PEP} . (f)-(j) repeats these plots for a zoomed-in 15-second portion of data (highlighted in yellow in (a)-(e)).

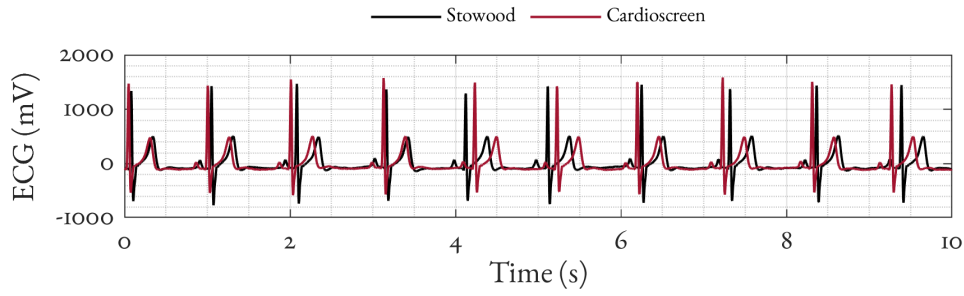


Figure 5.5: ECG phase delay between Stowood and Cardioscreen recording devices. Different internal device filtering, as well as potential differences in the drift of the two internal clocks, has resulted in a phase delay between the two. As a result, it was not possible to compute beat-by-beat PTT estimates in this dataset.

5.6 Estimating blood pressure using transit times

BP estimation using TTs was a three-step process involving time alignment, window averaging, and finally model fitting.

PAT and PEP were estimated from signals recorded by the Stowood and Cardioscreen devices respectively. While the devices were time synchronised using the network time protocol (NTP) at the start of each session, differences in internal signal filtering and potential differences in the drift of the two internal clocks, meant that it was not possible to perfectly align the timing of the signals recorded from the two devices for the entire duration of the recording session. Figure 5.5 shows the Stowood and Cardioscreen ECG time series which highlights the differences in phase lags between the two signals. A result of these phase lags is that a beat-by-beat assessment of PTT, as the difference between PAT and PEP (see section 2.8.1.5), was not possible using the MOLLIE clinical study. PTT was therefore computed using both a filtered PAT and PEP time series (PAT_{filt} and PEP_{filt}), and using PAT and PEP values averaged within a time window (see below).

We evaluated five different models proposed in the literature for the estimation of BP from TT. The first two models (LMEM random-intercepts and random-slopes) use all data points available for that individual; we refer to them as *a posteriori* models. The final three models (Poon, Gesche, and Fung) allow for real-time estimation of BP using *hybrid calibration* with model parameters determined from population averages and physiological constants. We refer to these three models as population-based models.

5.6.1 Time alignment

To align the PAT and PEP signals with BP, high-frequency variations in these signals were attenuated through filtering to ignore any frequencies above the Nyquist limit¹ of the BP cuff ($1/(2 \times 60)$

¹For a signal sampled at a rate f_s , the Nyquist limit, $f_s/2$, indicates the maximum frequency which can be captured [296].

Hz). To compute the filtered PAT signal, PAT_{filt} , any PAT beat with a corresponding SQI_{PAT} equal to zero was first removed from analysis. The irregularly sampled beat-by-beat PAT time series was then resampled at 1 Hz using cubic-splines interpolation. PAT_{filt} was then computed using a 6th-order Butterworth low-pass filter with a cut-off frequency of $1/(3 \times 60)$ Hz. This process was repeated to compute PEP_{filt} . Using these filtered signals, PTT_{filt} was then computed as:

$$PTT_{\text{filt}} = PAT_{\text{filt}} - PEP_{\text{filt}} \quad (5.25)$$

PAT_{filt} and PTT_{filt} were aligned with respect to the BP time series by subtracting the time delay (referenced τ_{PAT} and τ_{PTT} respectively) computed as the lag which minimised the cross-correlation (see section 5.5.1.2). This process was done independently for PAT_{filt} and PTT_{filt} and for SBP, MAP, and DBP, for each participant. In order to compute the cross-correlation, the two time series are required to be sampled at the same sampling frequency. Therefore, for the purposes of time alignment only, the reference BP cuff data were interpolated using the cubic smoothing splines algorithm (see section 5.3), at the same frequency as the TT_{filt} time series (1 Hz). This process is shown in figure 5.6 for the data recorded from a participant in the dataset. The computed τ_{PAT} and τ_{PTT} values across the whole dataset are shown in figure 5.7.

5.6.2 Averaging measurements

In order to compare the beat-by-beat PAT and PEP time series with the reference BP measurements (computed once per minute), we averaged the, time-aligned, PAT and PEP signals within a window, w , of size 40 seconds centred around each BP measurement (20s to the left and 20s to the right). The size of w was chosen as a trade-off between localisation of data points to the time of the BP reference value and minimising noise. Small values of w provide a better approximation of the local values of PAT and PEP at the time of cuff inflation, whereas large values of w allow for a better approximation of the mean within the window.

For PAT, we averaged the beat-by-beat PAT time series, time delayed by τ_{PAT} . A window based $SQI_{\text{PAT}}(w)$ was formed as its mean value within the window w . If $SQI_{\text{PAT}}(w)$ fell below a given threshold γ_{SQI} , then a valid measurement of PAT could not be made for that window. We computed PTT as the difference between the averaged PAT and PEP time series, both time delayed by τ_{PTT} . We formed a window-based $SQI_{\text{PAT}}(w)$ and $SQI_{\text{PEP}}(w)$ as the mean value within w . If either of these SQIs fell below the γ_{SQI} threshold, then a valid measurement of PTT could not be made for that window.

To determine the optimal value for γ_{SQI} , we conducted a sensitivity analysis to assess its impact on the proportion of windows for which a valid measurement was possible throughout the dataset.

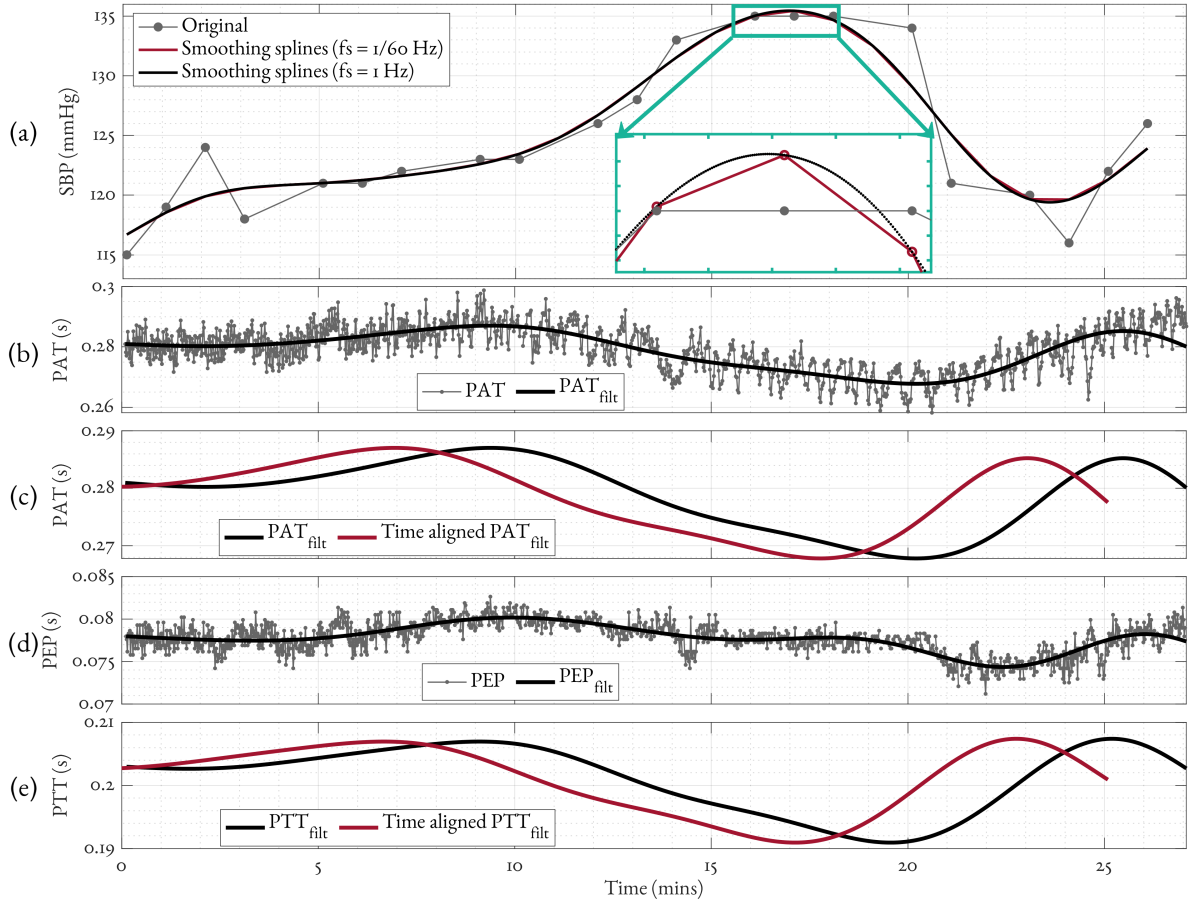


Figure 5.6: Overview of PAT and PTT time alignment to SBP for a typical participant in the dataset. (a) Original and interpolated SBP time series. The zoomed-in region demonstrates the interpolation of the SBP data to 1Hz for the purposes of time alignment. (b) Beat-by-beat PAT and PAT_{filt} . (c) PAT_{filt} and time aligned PAT_{filt} . In this instance τ_{PAT} was 146 seconds. (d) Beat-by-beat PEP and PEP_{filt} . (e) PTT_{filt} , and time aligned PTT_{filt} . In this instance τ_{PTT} was 116 seconds. Note that this process was repeated independently for MAP and DBP

As shown in figure 5.8, due to the ICG's relatively higher susceptibility to noise (compared to the PPG and ECG), there were consistently fewer good-quality windows available for PTT estimates than for PAT estimates. We selected γ_{SQI} as the maximum value that ensured at least 95% of windows remained for analysis for both PAT and PTT. Based on this analysis, we set γ_{SQI} to 0.7, resulting in 100% of PAT windows and 95.6% of PTT windows being available for analysis.

5.6.3 Models for the estimation of blood pressure

Below, we outline the models that we employed to estimate BP using TT for each participant in our dataset. These models can be categorised as: *a posteriori* models (linear mixed effects random-intercepts and random-slopes models), and population-based models (Poon [157], Gesche [160], and Fung [161]).

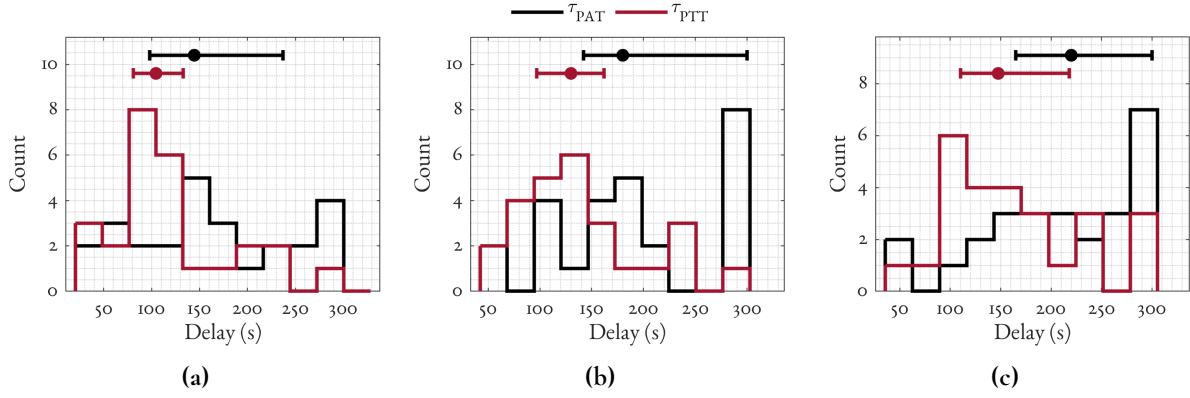


Figure 5.7: Histograms of τ_{PAT} and τ_{PTT} across the dataset. (a)-(c) for SBP, MAP, and DBP respectively. All PAT delays were significantly greater than PTT delays (Wilcoxon signed-rank, $p < 0.001$). Box plot shows the median and upper and lower quartiles.

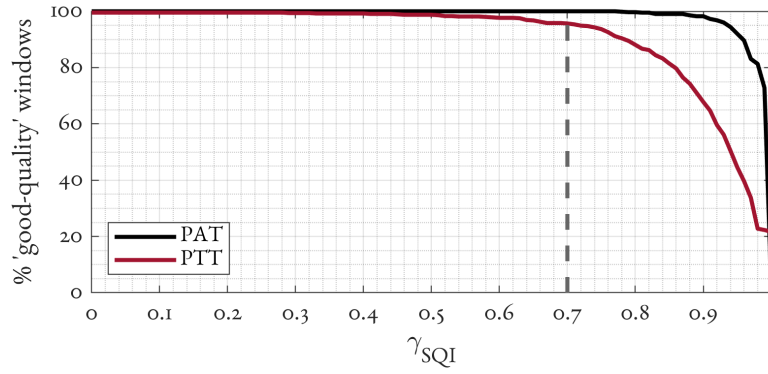


Figure 5.8: Sensitivity analysis of the impact of γ_{SQI} on the proportion of good-quality 40-second windows for which a valid measurement was possible. From this analysis, γ_{SQI} was set as 0.7. This resulted in 100% of PAT windows and 95.6% of PTT windows remaining for analysis.

5.6.3.1 *a posteriori* models (random-intercepts and random-slopes)

A posteriori models utilise *individual calibration* [9] in order to estimate BP using TT. These models require all available data for an individual in order to determine the model parameters, representing the best-case scenario for BP estimation where model parameters are known *a posteriori*.

We implemented *a posteriori* models using linear mixed-effects modelling (LMEM, see appendix D.1.2) to account for inter-participant variability in linear calibration parameters. We compared the estimation performance of two models: random-intercepts, and random-slopes. A random-intercepts model accounted for individual calibration intercepts for each participant. This model can be summarised as:

$$\begin{aligned} BP^p &= K_1 \cdot TT^p + K_0 + u_0^p + \epsilon^p \\ u_0 &\sim \mathcal{N}(0, \sigma_0^2), \quad \epsilon^p \sim \mathcal{N}(0, \sigma_\epsilon^2) \end{aligned} \quad (5.26)$$

where the superscript p denotes data from the p^{th} participant. K_1 and K_0 are the slope and intercept for the fixed effect component. u_0^p is the random-effect intercept for the p^{th} participant assumed to be drawn from a Normal distribution with variance σ_0^2 . ϵ^p is the model residual with variance σ_ϵ^2 .

A random-slopes model further accounted for individual linear calibration gradients ($\Delta\text{BP}/\Delta\text{TT}$). This model can be summarised as:

$$\begin{aligned} \text{BP}^p &= K_1 \cdot \text{TT}^p + u_1^p \cdot \text{TT}^p + K_0 + u_0^p + \epsilon^p \\ \begin{bmatrix} u_0 \\ u_1 \end{bmatrix} &\sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}\right), \quad \epsilon^p \sim \mathcal{N}(0, \sigma_\epsilon^2) \end{aligned} \quad (5.27)$$

where u_1^p is the random-effect slope for the p^{th} participant assumed to be drawn from a Normal distribution with variance σ_1^2 . u_{01} is the covariance between the intercept and slope residual for the random effect.

The requirement for personalised calibration gradients was assessed using a likelihood ratio test (LRT). For a random-intercepts model with parameter vector θ_0 and a random-slopes model with additional random effects and a parameter vector θ , LRT tests the null hypothesis (H_0) that the variance of the additional random effects are 0 (i.e. there is no requirement for participant specific calibration slopes). The likelihood ratio test statistic, LR , is given as:

$$LR = -2 \log\left(\frac{\mathcal{L}(\theta_0; y, X)}{\mathcal{L}(\theta; y, X)}\right) = -2(\ell(\theta_0; y, X) - \ell(\theta; y, X)) \quad (5.28)$$

where $\mathcal{L}$ and ℓ denote the likelihood and log-likelihood functions that the observed data was produced by the model, respectively (see appendix D.1.2). LR follows a chi-squared, χ^2 , distribution with the degrees of freedom equal to the difference in the number of degrees of freedom from θ and θ_0 , which in this work was 2 (σ_1^2 , the variance of the calibration slopes, and σ_{01} , the covariance of the slopes and intercepts). Thus, H_0 is rejected at a confidence level α when $LR > \chi_{1-\alpha}^2(2)$.

Hysteresis effect A hysteresis effect has been previously reported for the TT-BP relationship [297, 298]. In this work, we quantified the presence of a hysteresis effect as differences in $\Delta\text{BP}/\Delta\text{TT}$ computed during the dose increase and washout periods of the study protocol. We tested this effect for significance by a one-sample Wilcoxon signed-rank test. We compared the hysteresis effect of PAT to that of PTT by a paired two-sample Wilcoxon signed-rank test. All p -values were corrected using the Benjamini-Hochberg method [299].

5.6.3.2 Population-based models (Poon, Gesche, and Fung)

Population-based models utilise *hybrid calibration* [9] in order to estimate BP. These models require only a single-point calibration and derive the remaining parameters from population averages and physiological constants. In this work, we implemented three of the most highly cited population-based models for BP estimation using TTs: Poon [157], Gesche [160], and Fung [161]. These models were introduced in section 2.8.1.6.

We defined the calibration values for BP and TT (BP^{calib} and TT^{calib} respectively) as their respective mean values during the 5-minute resting period at the start of the study. In the specific example for one participant where no BP readings were recorded during the rest period, we used the mean values in a one-minute window centred on their first cuff inflation as the calibration readings.

5.6.4 Baseline reference

For the purpose of benchmarking performance metrics, we implemented a simple *baseline reference* that assumed no changes in BP from BP^{calib} for each individual. The *baseline reference* provides two key insights. Firstly, it indicates the degree to which BP was perturbed in the study with large values highlighting that there was a significant change. Secondly, it provides reference metrics against which all other models' performance must be evaluated. If, for example, one of the models reports worse performance metrics than the *baseline reference*, then clearly that model is worse than a naive assumption of no changes in BP in this dataset.

5.6.5 Model evaluation

We used the following performance metrics to evaluate model performance: Pearson's correlation coefficient (ρ_p), root mean squared error (RMSE), and mean absolute error (MAE).

ρ_p was used to measure the linear correlation between reference measurements, y , and model estimates, $\hat{y}$. Correlation coefficients provide a metric of the association or dependence between two variables. The values of correlation are in the range -1 to +1, with -1 signifying perfect negative correlation, +1 signifying perfect positive correlation, and values close to 0 signifying weak or no correlation between the two variables. ρ_p is defined as:

$$\rho_p = \frac{\text{cov}(y, \hat{y})}{\sigma_y \sigma_{\hat{y}}} = \frac{\sum_{i=1}^N (y_i - \mu_y)(\hat{y}_i - \mu_{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \mu_y)^2 \sum_{i=1}^N (\hat{y}_i - \mu_{\hat{y}})^2}} \quad (5.29)$$

where μ_y and $\mu_{\hat{y}}$ are the mean of y and $\hat{y}$ respectively, and σ_y and $\sigma_{\hat{y}}$ are the standard deviations of y and $\hat{y}$ respectively.

RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5.30)$$

where N is the number of data points. MAE is defined as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (5.31)$$

Both MAE and RMSE characterise the average error of the prediction. As MAE reports the average magnitude of errors, its values are easily interpreted across multiple datasets. In contrast, RMSE gives a higher weight to larger error values, and so it is an important metric when large errors are undesirable [300].

Differences in performance metrics across all participants were evaluated for statistical significance by a two-tailed Wilcoxon signed-rank test (see appendix C.1.1). We adjusted the p-values for multiple comparisons using the Benjamini-Hochberg method (see appendix C.1.2).

5.7 Results

5.7.1 Overall response

Figure 5.9 shows the participant-wise average deviations (Δ) in SBP, DBP, MAP, PAT, PTT and PEP from each participant's calibration values. The error bars show the 95% confidence intervals, (CI_{95}). On average SBP, MAP, and DBP increased by 20.5 (CI_{95} : 15.1-25.9) mmHg, 17.2 (CI_{95} : 13.7-20.6) mmHg, and 13.8 (CI_{95} : 12.0-17.8) mmHg respectively from their calibration values. The peaks of the recorded BP occurred during minute 17 (halfway through the max infusion phase of the study). PAT and PTT decreased by 13.3 (CI_{95} : 10.5-16.0) ms and 18.7 (15.6-21.7) ms respectively. The nadirs of the transit times occurred during minute 20. PEP increased by 5.82 (CI_{95} : 3.8-7.9) ms. The peak of PEP occurred during minute 19. Figure 5.9 (g) shows the ratio between PEP and PAT as a percentage, with a mean value of 28.6 % at the start of the session, rising to a peak of 32.5 % during minute 19.

Figure 5.10 shows the distributions of PAT, PTT, and PEP values computed across all participants. Mean (standard deviation) values for PAT, PTT, and PEP were: 252.6 (27.5) ms, 177.34 (26.74) ms, and 75.4 (5.99) ms respectively.

Figure 5.11 illustrates the frequency content of the PAT time series for a typical participant in the dataset. Figure 5.11(a) shows the PAT time series, which for analysis of the frequency content only, has been resampled to a 1 Hz signal using cubic-splines interpolation. Figure 5.11(c) shows the frequency spectrum, computed by the fast Fourier transform (FFT), of the PAT signal. The FFT shows three frequency ranges of interest. The first is a dominant component close to DC and below 0.0083 Hz, the Nyquist limit of the reference cuff measurements. This component is clearly driven by the large changes in BP observed using the cuff for all individuals in the dataset. The second is a component between 0.0083 Hz to 0.1 Hz. We suggest this component is driven by BP oscillations that cannot be recorded using the low sampling frequency cuff measurements. An example of these may be the Traube-Hering-Mayer (THM) waves, often known as Mayer waves

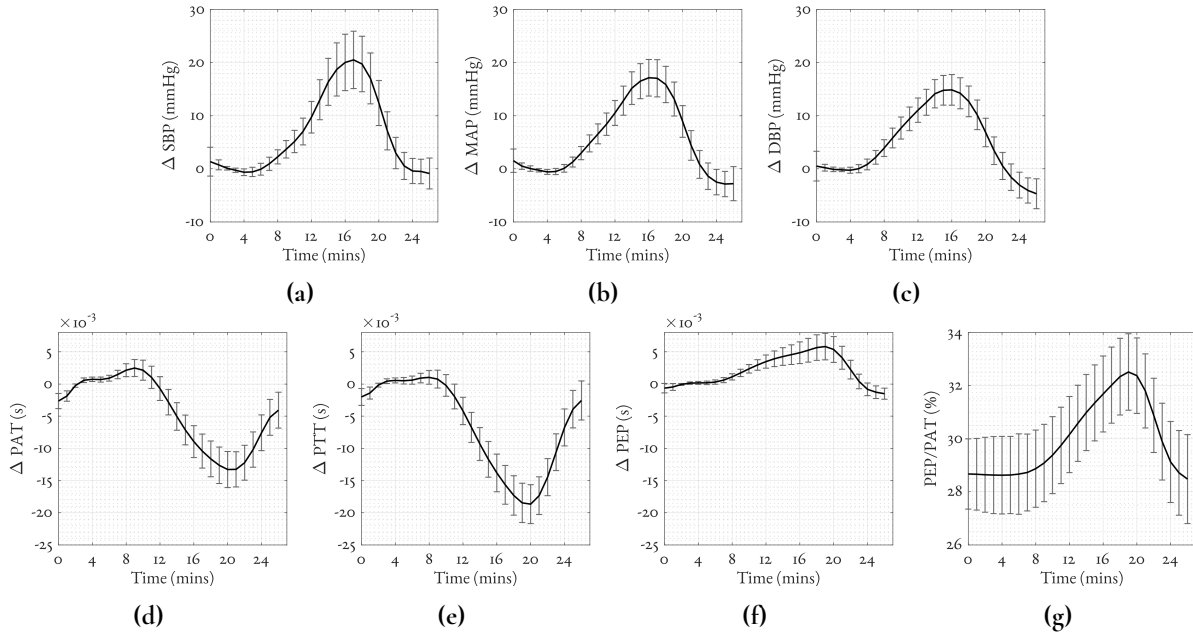


Figure 5.9: Average deviations of BP and TT metrics from calibration across the cohort. (a) SBP, (b) MAP, (c) DBP, (d) PAT, (e) PTT and (f) PEP. (g) variations in PEP as a percentage of PAT. The error bars show the CI_{95} .

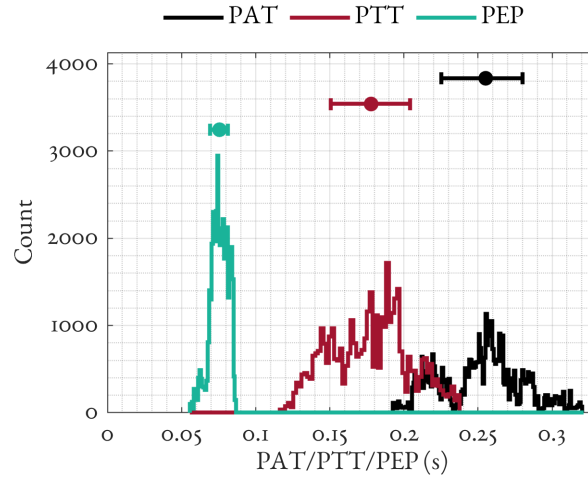


Figure 5.10: Distributions of PAT, PTT, and PEP computed in the MOLLIE dataset. Boxplots show mean and one standard deviation.

[40], see section 2.3.1.3. The third component is between 0.2 Hz and 0.4 Hz. This range is likely to be driven by respiration. Figure 5.11(b) shows the spectrogram of the PAT signal that has been high-pass filtered with a 0.1 Hz cut-off frequency. The estimated RR (the steps for estimating RR are provided in appendix E.3) from the nasal cannula is superimposed in red, showing that there is a clear respiratory component in the PAT signal.

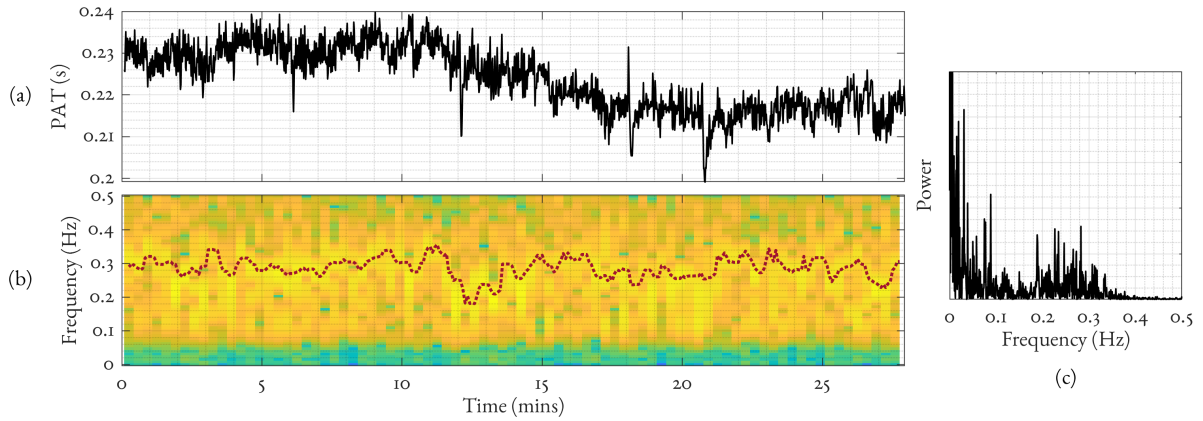


Figure 5.11: Frequency content of PAT signal for a typical participant (24 years old, female) in the dataset. (a) PAT time series. (b) Spectrogram of a high-pass filtered PAT signal showing a large respiratory component. The reference RR from the nasal cannula is superimposed in red. (c) fast Fourier transform (FFT) frequency spectrum of PAT signal.

5.7.2 Estimating blood pressure using transit times

Table 5.1 shows the median and IQR performance statistics for all models over all participants in the dataset for SBP estimation. Results for MAP and DBP estimation are provided in appendix E.4. The *baseline reference* resulted in large RMSE and MAE values, indicating that participants experienced a significant change in their SBP values in response to the weight-based dosing of phenylephrine. On average, we achieved an increase of 20 mmHg in SBP, with a maximum increase of 40 mmHg in a subset of four participants.

PAT, PTT, and PEP had a median (IQR) ρ_p with SBP of $-0.85(0.10)$, $-0.90(0.07)$, and $0.81(0.30)$. The corresponding values of ρ_p with MAP were $-0.82(0.19)$, $-0.88(0.15)$, and $0.80(0.23)$. The corresponding values of ρ_p with DBP were $-0.83(0.33)$, $-0.85(0.13)$, and $0.77(0.20)$. PTT had a consistently stronger ρ_p with BP than PAT (Wilcoxon signed-rank, $p < 0.001$ for SBP, MAP, and DBP).

Most likely to be driven by the linear relationship between TT and BP (see section 2.8.1), there was little variation observed in ρ_p across all models. Therefore, the error values (RMSE and MAE) are the primary parameters of interest in this work when comparing model performance.

5.7.2.1 *A posteriori* models

All *a posteriori* models significantly outperformed the *baseline reference* ($p < 0.001$ for all). The addition of personalised individual slopes was tested for significance for PAT and PTT (and for all SBP, MAP, and DBP) via a LRT. For all tests the null hypothesis was rejected with a p value close to or at 0. This highlights the requirement for individual calibration curves for appropriate BP estimation. In figures 5.12 and 5.13 we show the relationship between SBP and TT for all

Table 5.1: Performance statistics of SBP estimation using the models proposed. Results are given as median (IQR) computed across all participants. Entries in bold indicate the best performance for that metric for either *a posteriori* or population-based models.

Model		ρ_p	RMSE *		MAE *	
Baseline reference [†]		- (-)	10.46	(6.18)	7.40	(4.87)
<i>A posteriori</i> models						
PAT	Random-intercepts	0.85 (0.10)	4.59	(2.46)	3.54	(1.84)
	Random-slopes	0.85 (0.10)	4.18	(2.50)	3.26	(1.89)
PTT	Random-intercepts	0.90 (0.07)	4.12	(1.61)	3.23	(1.46)
	Random-slopes	0.90 (0.07)	3.06 (2.22)	2.45 (1.64)		
Population-based models						
PAT	Poon [157]	0.85 (0.10)	8.61	(5.01)	6.27	(3.91)
	Gesche [160]	0.85 (0.10)	11.67	(18.99)	8.54	(15.83)
	Fung [161]	0.85 (0.10)	5.68 (2.82)	4.48 (2.31)		
PTT	Poon [157]	0.87 (0.07)	6.13	(3.04)	4.95	(2.68)
	Gesche [160]	0.87 (0.11)	503.50	(1806.88)	383.11	(1349.34)
	Fung [161]	0.87 (0.07)	21.43	(23.28)	15.96	(17.49)

* results given in units of mmHg. [†] Baseline reference assumed no changes in SBP for each participant from SBP^{calib}.

participants in the dataset. We plot both the random-intercepts regression line and the, more appropriate, random-slopes regression line in order to highlight the difference between the two. The relationship for MAP and DBP is provided in appendix E.4.

For all *a posteriori* models, PTT outperformed PAT. The random-slopes model using PAT achieved similar performance statistics as the random-intercepts model using PTT and non-significant differences were observed between the two.

Figure 5.14 shows the distributions of calibration gradients ($\Delta\text{BP}/\Delta\text{TT}$) computed across the dataset. It was found that for SBP and MAP, the calibration gradient of PAT had a significantly larger amplitude than for PTT (Wilcoxon signed-rank, $p < 0.01$ for both). Non-significant differences were observed for DBP. The gradients of the SBP-PAT relationship varied from $-2,946$ to -470 mmHg s⁻¹ in our dataset. The gradients of the SBP-PTT relationship varied from $-3,185$ to -324 mmHg s⁻¹.

5.7.2.2 Hysteresis effect

Figure 5.15 shows the results of the hysteresis analysis. It is clear that for both PAT and PTT, the calibration gradient was larger during the dose increase than during the washout. Note that as the calibration gradients were negative (see figure 5.14), this means that the gradient during washout was steeper than the gradient during dose increase. This result is consistent for SBP, DBP, and

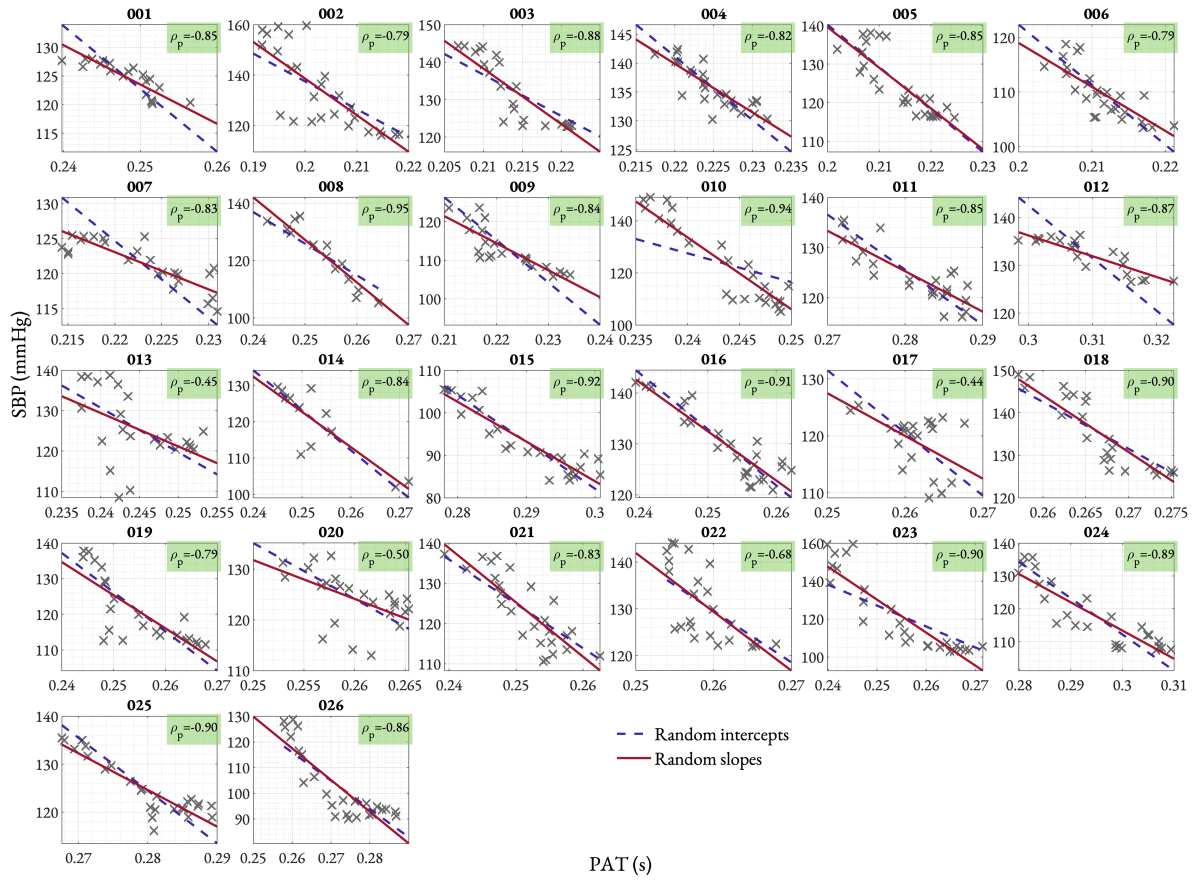


Figure 5.12: Individual relationships between SBP (mmHg) and PAT (s) for all 26 participants in the MOLLIE study. The random intercepts regression line is shown in dashed blue. The random slopes regression line is shown in red.

MAP. These effects were all found to be statistically significant ($p < 0.05$ for all). The median difference in dose increase and washout $\Delta\text{SBP}/\Delta\text{PAT}$ was 224 mmHg s^{-1} ($p = 0.033$) and the difference for $\Delta\text{SBP}/\Delta\text{PTT}$ was 110 mmHg s^{-1} ($p = 0.048$). The hysteresis effect appears to be reduced for PTT when compared to PAT; however, it should be noted that non-significant differences were found between the two. This overall effect can also be observed in figure 5.9 in which the averaged PAT and PTT profiles do not return to their baseline value, despite BP returning (and in the case of DBP, overshooting). The difference with respect to baseline at the end of the study was smaller for PTT than for PAT.

5.7.2.3 Population-based models

All of the population-based models recorded worse performance metrics than all of the *a posteriori* models. For all performance metrics, the Gesche model performed worse than the *baseline reference*. This was especially the case when using PTT as an input where the median RMSE value exceeded 500 mmHg. The Poon model significantly outperformed the *baseline reference* ($p < 0.001$ for

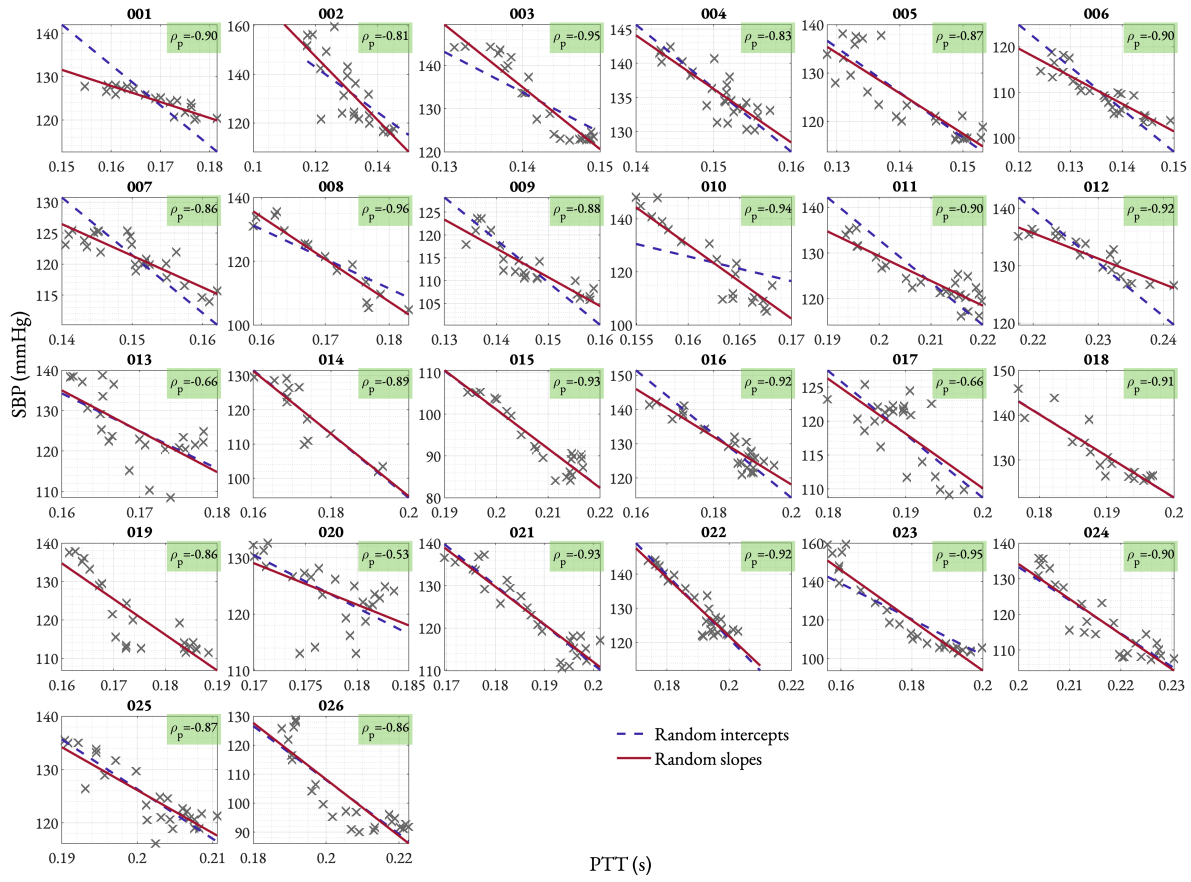


Figure 5.13: Individual relationships between SBP (mmHg) and PTT (s) for all 26 participants in the MOLLIE study. The random intercepts regression line is shown in dashed blue. The random slopes regression line is shown in red.

both PAT and PTT). A significant improvement in results was observed when implementing the Poon model with PTT as opposed to PAT ($\Delta\text{RMSE} = 2.48 \text{ mmHg}$, $p < 0.001$; $\Delta\text{MAE} = 1.32 \text{ mmHg}$, $p < 0.001$). The Fung model with PAT reported the best performance statistics across the population-based models ($\text{RMSE} = 5.68 \text{ mmHg}$, $\text{MAE} = 4.48 \text{ mmHg}$). The performance of the Fung model was significantly ($p < 0.001$) worse when using PTT as an input and this model did not outperform the *baseline reference*.

5.8 Discussion

5.8.1 The relationship between pulse arrival time and blood pressure

The PAT frequency response exhibited a dominant component driven by the dosing regime ($f < 0.0083 \text{ Hz}$). In addition, we hypothesised that there were extra cardiovascular fluctuations exhibited in the 0.0083 Hz to 0.1 Hz range such as Mayer waves. Mayer waves are oscillations in BP, with a frequency of around 0.1 Hz , thought to be due to oscillations of the sympathetic

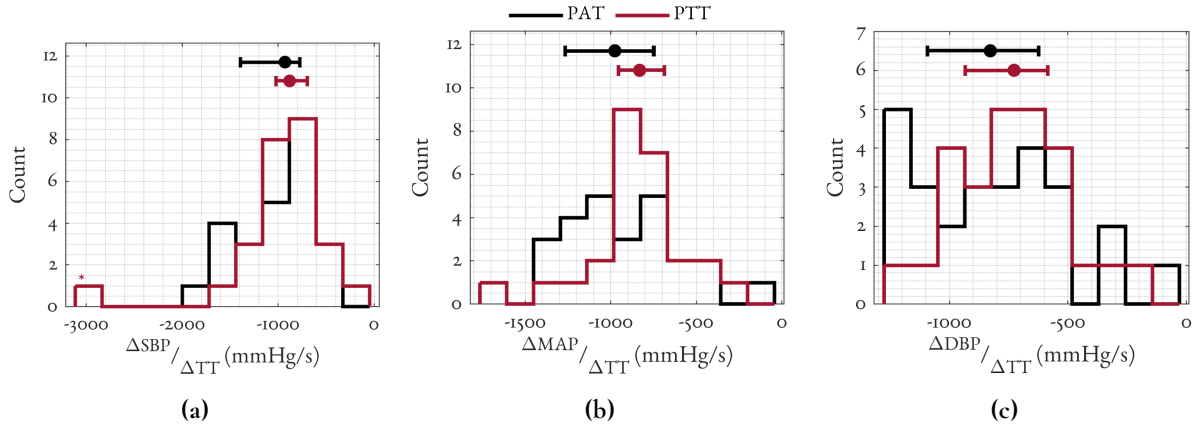


Figure 5.14: Distributions of $\Delta BP / \Delta PAT$ and $\Delta BP / \Delta PTT$ across the dataset for (a) SBP, (b) MAP, and (c) DBP. $\Delta BP / \Delta PAT$ had significantly (Wilcoxon signed-rank, $p < 0.01$) larger amplitude than $\Delta BP / \Delta PTT$ for SBP and MAP, but not for DBP. Box plot shows the median and upper and lower quartiles.

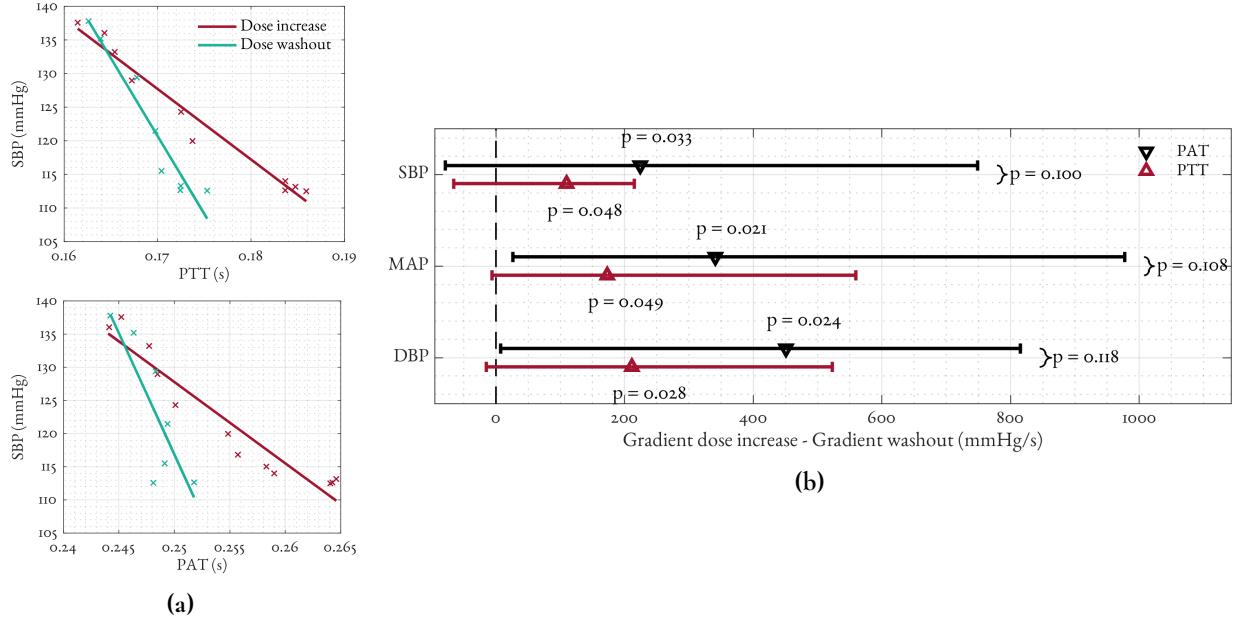


Figure 5.15: The hysteresis effect in both PAT and PTT. (a) Example dose increase and washout calibration gradients for PAT and PTT observed in one participant. (b) Median and inter-quartile range differences in dose increase and washout calibration gradients across all participants. A significant (Wilcoxon signed-rank, $p < 0.05$ for all) hysteresis effect was present for both PAT and PTT.

vasomotor tone of arterial blood vessels. The amplitude of the Mayer waves has been shown to be an indicator of sympathetic nervous activity [40]. Finally, PAT showed a strong respiratory component. This finding has been previously demonstrated by Ding *et al.* [301] and is driven by respiratory-induced phase delays of the PPG signal.

PAT was negatively correlated with BP and the majority of participants had their own unique and strong relationship between changes in PAT and SBP (see figure 5.12). This response was also consistently shown for MAP and DBP.

5.8.2 Impact of pre-ejection period

It has been suggested that the contribution of PEP makes PAT an unreliable surrogate for BP [151, 152]. Payne *et al.* [151] studied the influence of PEP under different vasoactive medications in 12 healthy individuals. It was reported that PEP accounted for between 12-35 % of PAT. For drugs where only a small change in BP was observed (salbutamol), the PEP component was dominant and resulted in PAT and PTT changes in opposing directions. However, for drugs where a significant change in BP was observed (angiotensin and norepinephrine), PAT and PTT had comparable behaviour. Indeed, in studies where PEP is reported to be a non-negligible variable component of PAT, it is often the case that BP was not significantly perturbed, for example using exercise or posture changes [153, 280]. These effects are likely to lead to small changes in PTT and it can be expected that under these circumstances, the PEP contribution is non-negligible.

In our study, the maximum change from the initial value at rest of PEP had a mean value (standard deviation) of 5.8 (5.0) ms (see figure 5.9). However, the changes in PTT were considerably larger in magnitude -13.3 (7.1) ms. As a result, we found that PEP contributed between 28.6% and 32.5% of PAT in our study (see figure 5.9 (g)). Therefore, while changes in PEP cannot be neglected, their relative contribution to PAT was nearly constant. Changes in PTT accounted for most of the changes in PAT during the 28 minutes of the study. Collectively, these results suggest that PAT may be an unreliable marker for small (for example, resting beat-to-beat) changes in BP. However, for significant changes (for example, 24-hour circadian rhythms), the variable influence of PEP might be considered to be negligible, making PAT a reliable surrogate for PTT.

There were three significant differences observed between the PAT and PTT estimates which highlight the effect of PEP in this dataset. Firstly regarding the hysteresis effect presented in section 5.7.2.2. The phenomenon of PAT hysteresis has been previously investigated by Liu *et al.* in the context of an exercise test [297]. In their study, hysteresis was defined as the area between the PAT-BP relationship curves observed during both the exercise and recovery phases. Notably, Liu *et al.* found a positive correlation between the extent of hysteresis and increased ANS activity and smooth muscle tone. Consequently, it is suggested that distinct baroreflex responses are activated when BP is rising or falling when induced by smooth muscle activation, potentially necessitating the use of separate calibration coefficients [298]. Our dataset confirms that in situations where smooth muscle activation dominates, such as under the influence of phenylephrine, which significantly affects arterial stiffness, especially at peripheral measurement sites, a hysteresis effect is present. Interestingly, our data also suggests a less pronounced hysteresis effect in PTT compared to PAT. This is counter-intuitive as the origin of the hysteresis effect

is thought to be due to changing arterial properties as opposed to changes in cardiac properties. In a broader context, the presence of the hysteresis effect holds significant implications for BP estimation using TT. It indicates the potential need for fine-tuned calibration coefficients based on ANS activity levels and the anticipated direction of BP fluctuations. This effect should be explored in more detail in multiple follow-up studies.

Secondly, the delay values computed between BP and TT (see figure 5.7) suggest that τ_{PAT} values were significantly larger than the τ_{PTT} values. In our study, the origin of the delay between BP and TT is likely to have been largely dominated by instrumentation error as opposed to a physiological delay. The Philips patient monitor recording the BP measurements reported the timestamps as if they occurred at the start of each cuff inflation. However, the process of cuff inflation and deflation may take up to 40 seconds, and the reference value is only localised to the time of maximum amplitude variations. Therefore, there is immediately a delay present between the reported cuff value times and the actual BP at that time. This instrumentation delay, however, fails to account for the differences observed between PAT and PTT. This difference is likely to be the result of the differences in hysteresis discussed above, resulting in PTT reaching its nadir earlier than PAT, and hence a shorter delay.

The third and final observed difference was that PTT demonstrated stronger ρ_p with BP than PAT. This resulted in all *a posteriori* models implementing PTT outperforming all equivalent models implementing PAT. These differences were consistent as highlighted by a p -value < 0.001 when comparing the random-slopes models using PAT and PTT. However, the absolute values of the differences were minor, with only a 1.12 mmHg difference in median RMSE observed between the two. In fact, it was found that the differences between the PAT random-slopes model and the PTT random-intercepts model were insignificant. This suggests that cuffless BP estimation may be more easily addressed by ensuring accurate PAT calibration values as opposed to developing new devices that mitigate for changes in PEP.

5.8.3 Estimating blood pressure using transit times

We implemented both population-based models and *a posteriori* models in order to estimate BP using TT. Three different population-based models were tested: Poon [157], Gesche [160] and Fung [161], three of the most highly cited population-based models for BP estimation using PAT or PTT. These models require only a single calibration reading and approximate the remaining model parameters using averaged physiological constants (such as the density of blood) and global averages. These population average models may well provide an effective initial assessment of changes in BP, with some of them performing better than the *baseline reference*. However, they all

failed to outperform either of the *a posteriori* random-intercept models which assumed a constant TT-BP slope for all participants without any adjustments for demographics. While we analysed a homogeneous cohort in this work, the calibration gradients varied significantly and indeed the LRT results highlighted that *individual calibration* curves require both a unique intercept and slope. It is clear, therefore, that population averages are not appropriate for estimating personalised model parameters. We suggest, that if cuffless devices recording TTs are to deliver clinical-grade BP monitoring (as some claim they already do, see section 2.11), it is unlikely that the current population average models will provide sufficient accuracy.

5.8.4 Limitations

While the results of this chapter demonstrate the utility of the PAT as a surrogate for BP, there are several limitations that should be noted. The MOLLIE study analysed consisted of a small number of, relatively homogeneous, healthy participants ($N_p=26$) The size of the dataset was limited by practical considerations related to conducting the MOLLIE study, as opposed to being determined by power calculations on the minimum sample size required to confidently conclude on the method's validity. Consequently, the extent to which the findings presented in this chapter can be extrapolated to the broader population remains uncertain.

Moreover, BP was perturbed via an infusion of phenylephrine, an α_1 -adrenergic receptor agonist that induces arterial and venous smooth muscle contraction [247]. In daily life, BP changes result from a diverse set of physiological mechanisms governed by the autonomic nervous system. α_1 receptors are typically activated in response to shock or low blood pressure [302]. Whereas, the resulting smooth muscle contraction is often activated during daily life activities, such as exercise [303], in order to ensure adequate blood flow. Overall, it must be considered that this was a feasibility study with relatively few subjects. Further studies, especially with a more heterogeneous and diverse set of subjects, is required to understand how the relationship observed in this work compares to that present in daily life.

PEP was estimated using the Lozano method [292]. This method avoids detecting the B-point (point of aortic valve opening) by detecting the C-point (maximum of ICG pulse) of the ICG and a derived relationship between the ECG R-peak and the ICG B and C points. However, van Lien et al [304] reported considerable errors between PEP estimation using the Lozano method and hand annotated PEP values, especially for extreme values. Other techniques for detecting the B-point exist [291, 290], however, in our study, the Lozano method produced considerably less noisy estimates than these techniques. Additionally, the ICG signal was sampled at 200 Hz, resulting in a time resolution of 5ms for the detection of the C-point. This could potentially lead

to errors in the estimation of PEP. Appendix E.5 provides a comparison of PEP estimates from the original 200 Hz signal against PEP estimates from an upsampled ICG waveform at 1 kHz using cubic splines. Although we found that this limitation did not significantly impact the results and conclusions of this chapter, further studies should aim to record the ICG waveform at a high sampling rate, such as 1 kHz, especially to estimate a wide range of PEP values.

Finally, measurements of BP using a sphygmomanometer cuff are susceptible to various forms of noise that can distort the readings. The oscillometric device used as a BP reference in this study was compliant with the IEC 60601-2-30/EN60601-2-30 and with the American National Standard for Electronic or Automated Sphygmomanometers (ANSI/AAMI SP 10/92) [242] with a maximum mean error of ± 5 mmHg (± 0.7 kPa) and a maximum standard deviation of 8 mmHg (1.1 kPa). Despite this, it was observed that the recorded BP values were both noisy and sparse (see figure 5.2). To address these challenges, we implemented a cubic smoothing splines algorithm to generate a smoother BP profile and interpolate missing data, ultimately providing a reference BP measurement at one-minute intervals. However, it's important to note that this data processing step introduced its own set of limitations, including the potential for smoothing artefacts and interpolation-related uncertainties, which should be taken into consideration when interpreting the BP-related findings of this study.

5.9 Conclusion

In this chapter, we investigated the potential use of PAT as a surrogate for BP. We introduced new processing methods for both the PAT and PEP time series in order to provide robust estimates of each signal. We highlighted the presence of a hysteresis effect in both the PAT and PTT signals and recommend that this effect be studied in more detail in a follow-up study. Additionally, we propose that the impact of PEP may be too significant for a beat-by-beat assessment of small changes of BP using PAT. However, for scenarios in which large changes in BP are observed (such as was in the case in our study) changes in PTT are likely to be significantly larger than changes in PEP and so the use of PAT instead of PTT is justifiable. Finally, we suggest that population-based *hybrid calibration* strategies may not adequately reflect the unique and individualised relationship between changes in BP and changes in TT. Therefore, they should be used with caution and only as a potential indicator of directional changes over time as opposed to a clinical assessment of BP, for example for the diagnosis of hypertension.

Computing estimates of PAT requires two signals (ECG and PPG) recorded synchronously. In the following chapter, we assess whether the ECG or PPG waveforms on their own may be used for BP estimation.

6

Features from the photoplethysmogram and electrocardiogram for estimating changes in blood pressure

Contents

6.1	Introduction	127
6.2	Methods overview	128
6.3	Computing features from the photoplethysmogram	128
6.3.1	Fiducial point detection	128
6.3.2	Feature extraction	131
6.4	Computing features from the electrocardiogram	139
6.4.1	Feature extraction	139
6.5	Reference blood pressure values	142
6.6	Estimating changes in blood pressure	142
6.6.1	Feature set: X	143
6.6.2	Calibration	143
6.6.3	<i>Baseline reference</i>	144
6.6.4	Removing collinear features	144
6.6.5	Data augmentation	145
6.6.6	Regression models	145
6.6.7	Shapley additive explanation values feature importance ranking coefficient . . .	146
6.7	Results	147
6.7.1	Removing collinear features	147
6.7.2	Model performance	149
6.7.3	Feature importance	152
6.8	Discussion	154
6.8.1	Observed changes in the photoplethysmogram waveform morphology	154
6.8.2	Model performance	155
6.8.3	Feature importance	158
6.8.4	Limitations	159
6.9	Conclusion	160

6.1 Introduction

In chapter 5, BP estimation using PAT was introduced and evaluated. PAT, however, is limited by the requirement for two synchronous signals to be recorded from different locations on the body. Another category of BP estimation techniques focuses on relating morphological features of the PPG or ECG waveforms to BP (or changes in BP, Δ BP) using data-driven models. This approach addresses many of the major limitations of PAT. However, the appropriate features

and models required for accurate BP estimation are not well established. Additionally, whether the requirement for *individual calibration* is diminished using this approach compared to PAT remains unclear in the literature. In this chapter, we investigate the utility of features from the PPG or ECG waveforms to estimate ΔBP .

6.2 Methods overview

As in the previous chapter, data was gathered from the 26 participants in the MOLLIE clinical study (see section 3.3). The PPG and ECG signals relevant to this chapter were recorded using the Stowood device. We extracted a large set of features from both the PPG signal (PPG_{fts}) and from the ECG signal (ECG_{fts}), based on the existing literature. We additionally propose new features from both signals. Using *hybrid calibration*, we trained, tuned, and evaluated linear (ordinary least squares, OLS) and non-linear (random forest, RF) machine learning models to estimate ΔBP within a nested leave-one-subject-out cross-validation framework. We assessed the individual importance of each feature for estimating ΔBP through SHAP values and a ranking coefficient. We compared the performance of ΔBP estimation using PPG_{fts} or ECG_{fts} to the performance metrics computed using PAT and equivalent machine learning models.

6.3 Computing features from the photoplethysmogram

To preprocess the PPG waveform, detect the cardiac beats, and assess the signal quality of the PPG (SQI_{PPG}), we followed the steps previously outlined in section 5.4. Amplitude modulation of the PPG can result from respiration [305] as well as changes in contact pressure of the pulse oximeter sensor on the surface of the skin [222]. In order to reduce the dependence on this amplitude modulation, as well as changes in HR (a phenomenon known as heart rate variability), we normalised each PPG beat to have unit amplitude and time duration. The first and second derivatives of the PPG beats, known as the VPG and APG respectively, were computed using a 6th-order Savitzky-Golay derivative filter (see section 5.5).

6.3.1 Fiducial point detection

Figure 6.1 shows the typical changes in the morphology of the PPG, VPG, and APG beat observed for a participant during the four main stages of the study protocol (rest, dose increase, maximum infusion, and washout). Each of these waveforms contains salient fiducial points utilised for feature extraction. As we normalised the time duration of each beat, we define a normalised time vector t' such that t' starts at 0 and is equal to 1 at the end of the beat.

From the PPG the following fiducial points were detected: systolic peak (S), dicrotic notch (N), and diastolic peak (D). S corresponds to the maximum compressive pressure during left ventricular ejection and was defined as the first turning point of the PPG beat waveform above its midpoint. The closing of the aortic valve produces a small inflection in pressure that gives rise to N (the local minimum) and D (the local maximum). N was detected through analysis of the APG and was characterised by the e wave (early diastolic positive wave, defined below) [78, 189]. D was detected as the local maximum of the PPG in the window $N_{t'} < t' < 0.8$, where $N_{t'}$ denotes the time of the dicrotic notch on the normalised time vector. The limit of 0.8 in the search window accounts for the upper bound of the expected location of D [189].

The maximum of the VPG, W , defines the point of the steepest rise in the ascending limb of the PPG. From the APG waveform, four systolic waves and one diastolic wave were detected: a (early systolic positive wave), b (early systolic negative wave), c (late systolic re-increasing wave), d (late systolic re-decreasing wave), and e (early diastolic positive wave) [78]. a , the initial maximum curvature of the PPG upslope, was detected as the local maximum of the APG prior to $W_{t'}$. b , the minimum curvature prior to maximum pressure, was detected as the local minimum of the APG between a and S .

Prior work by Gemignani *et al.* [306] has suggested that the maximum duration of systole in an adult population is 0.5 seconds. Thus, in the window $b_{t'} < t' < 0.5/T$, where T is the original time duration of the beat, two peaks in the APG should appear: the c wave and e wave. e , as noted above, corresponds to the increase in pressure resulting from the closing of the aortic valve. c is the maximum curvature of the descending limb of the PPG pulse from peak to dicrotic notch. If two (or more) peaks in the APG appear in this window, we then defined the second peak to be the location of the e wave and the first peak to be the location of the c wave. However, it is not uncommon for c to reflect an inflection point, in which case only one peak in the APG appears in this window corresponding to the e wave [78, 189]. In this instance, the location of the c wave was defined as the minimum of the fourth derivative of the PPG (also known as the jerk plethysmogram, JPG) between the detected b and e waves. Finally, the d wave, the minimum curvature of the descending limb of the PPG pulse from its peak to the dicrotic notch, was defined as the minimum of the APG between c and e waves. However, again d may be reflected by an inflection point, in which case no minimum would exist in this region and d was set to be coincident with e . As a summary, the criteria used for identifying the PPG fiducial points are outlined in table 6.1.

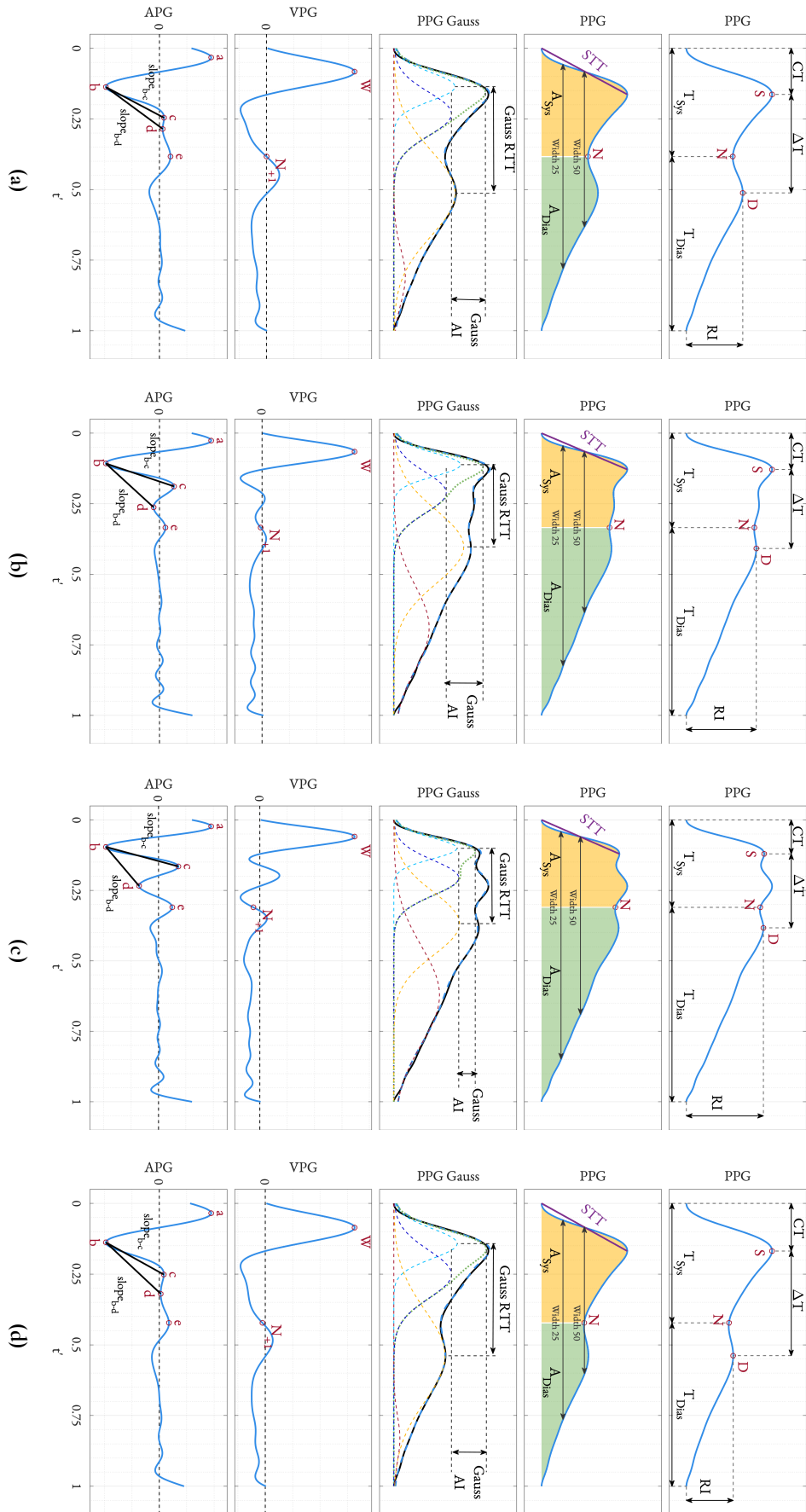


Table 6.1: Criteria for identifying fiducial points on PPG waveform [189]. T is the time duration of the pulse.

Signal	Fid. point	Criterion
PPG	S	The first turning point of the PPG pulse above its midpoint.
	N	Coincident with e .
	D	First local maximum of the PPG pulse after N and before $0.8T$. If no maxima is in this window then the first local maximum of the APG pulse in the window.
VPG	W	The maximum of the VPG pulse.
APG	a	The maximum of the APG pulse before W .
	b	The local minimum of the APG pulse after a and before S .
	c	The local maximum of the APG pulse after b and before e . If no maxima is in this window then the first local minima of the first derivative of the APG signal (also known as the jerk plethysmogram, JPG) in the window.
	d	The local minimum of the APG pulse after c and before e . If there is no minimum then the point is an inflection and is coincident with c .
	e	The second maximum of the APG pulse after W and before $0.6T$.

6.3.2 Feature extraction

The PPG_{fts} we extracted are grouped into: PPG morphology, VPG morphology, APG morphology, Gaussian decomposition, and principal component analysis (PCA). These features were identified from previous publications describing methods for estimating ΔBP from the PPG. All PPG_{fts} were computed on a beat-by-beat basis. Table 6.2 provides a summary of these PPG_{fts} which are then outlined in more detail in the subsequent sections.

We processed PPG_{fts} using a method similar to that proposed in section 5.4. Firstly outlier values were detected using a moving window average. A sliding window of size 30 seconds with a step size of 25 seconds was used. Subsequently, Kalman filtering was applied to reduce the impact of transient changes from noise and motion artefacts. The SQI_{PPG} of the k^{th} beat was used to weight the measurement noise as (see section 5.4.2):

$$R_k = \exp\left(\frac{1}{\text{SQI}_{\text{PPG}}(k)^2} - 1\right) \quad (6.1)$$

Table 6.2: Summary of PPG_{fts} extracted, separated into PPG morphology, VPG morphology, APG morphology, Gaussian decomposition, and PCA. References to previous applications of the features are also provided.

Category	Feature notation and publication	Description or formula
PPG morphology	Notch amplitude (N_{amp})	Amplitude of the dicrotic notch (see figure 6.1)
	Reflective index (RI) [307]	Amplitude of the diastolic peak (see figure 6.1)
	ΔT [78]	Time from systolic peak to diastolic peak (see figure 6.1)
	Crest Time (CT) [192]	Time from onset to systolic peak (see figure 6.1)
	T_{Sys} [308]	Time in systolic phase (see figure 6.1)

Continued on next page

Table 6.2 – Continued from previous page

Category	Feature notation and publication	Description or formula
	T_{Dia} [308]	Time in diastolic phase (see figure 6.1)
	T_{Ratio} [308]	T_{Sys}/T_{Dia}
	Slope transit time (STT) [309]	Slope of straight line from onset to peak (see figure 6.1)
	Stress-Induced Vascular Response Index (sVRI) [310]	μ_{Dias}/μ_{Sys}
	A_{Sys} [311]	Area under PPG in systolic phase (see figure 6.1)
	A_{Dias} [311]	Area under PPG in diastolic phase (see figure 6.1)
	Inflection point area (IPA)[311]	A_{Dias}/A_{Sys}
	Width ₂₅ [78]	Width of the PPG at 25% of its amplitude (see figure 6.1)
	Width ₅₀ [78]	Width of the PPG at 50% of its amplitude (see figure 6.1)
	Pressure index (PI) [202]	$\frac{N_{t'} - S_{t'}}{N_{t'} - W_{t'}} \cdot h$
	Normalised harmonic area (NHA) [204]	$\sum_{n=2}^{N_h} FFT^2(n \cdot f_1) / \sum_{n=1}^{N_h} FFT^2(n \cdot f_1)$
	Inflection and harmonic area ratio (IHAR) [204]	(1-NHA)/ IPA
	Skewness [213]	Equation (6.4)
	Kurtosis [213]	Equation (6.4)
VPG morphology	Sys_{μ} [215]	Mean of VPG in the systolic phase
	Sys_{σ} [215]	Variance of VPG in the systolic phase
	Dia_{μ} [215]	Mean of VPG in the diastolic phase
	Dia_{σ} [215]	Variance of VPG in the diastolic phase
APG morphology	$\frac{b}{a}, \frac{c}{a}, \frac{d}{a}$ & $\frac{e}{a}$ [79]	
	Ageing index (AGI) [79]	$(b - c - d - e)/a$
	slope _{b-c} [308]	Slope of a straight line between b and c , normalised by a (see figure 6.1)
	slope _{b-d} [308]	Slope of a straight line between b and d , normalised by a (see figure 6.1)
	PPG AI [312]	PPG augmentation index d_{-2}/b_{-2} (see section 6.3.2.3 for an explanation of the suffixes)
Gaussian decomposition	$A_{g1-4}, \sigma_{g1-4}, \mu_{g1-4}$	Amplitude, variance, and mean of the four decomposed Gaussians
	Gaussian augmentation index (Gauss AI) [187]	$\max(g_s) - A_{g3}$ (see figure 6.1)
	Gaussian reflection index (Gauss RI) [187]	$\sum(g_s) - \sum(g_3)$
	Gaussian reflected wave transit time (Gauss RTT) [193]	$\mu_{g3} - \mu_{g1}$ (see figure 6.1)
	Gaussian augmentation index _R (Gauss AI _R) [193] [†]	$(A_{g1} - A_{g2})/(A_{g1})$
	Gaussian reflection index _R (Gauss RI _R) [193] [†]	A_{g3}/A_{g1}
	Gaussian approximation of left ventricular ejection time (Gauss LVET) [206]	See section 6.3.2.4 for definition.
	Gauss _{Sys/Dias} *	$\sum(g_s) / \sum(g_d)$
	Gauss _{A4/A1} *	A_{g4}/A_{g1}
	Gauss _{$\sigma 4/A1$} *	σ_{g4}/A_{g1}
PCA	PPG PCA ₁₋₃ [313]	First 3 principal components of PPG beat
	VPG PCA ₁₋₃ [313]	First 3 principal components of VPG beat
	APG PCA ₁₋₃ [313]	First 3 principal components of APG beat

Continued on next page

Table 6.2 – Continued from previous page

Category	Feature notation and publication	Description or formula
* Indicates features that, to our knowledge, have not been previously implemented for BP estimation. † Note that there are two Gaussian indices referenced reflection index and augmentation index. We refer to the second set using the subscript R to reflect the authors: Rubins <i>et al.</i> [205] μ_{Sys} and μ_{Dias} represents the mean of the PPG during systole and diastole respectively. h is the height of the participant in cm. FFT is the Fast Fourier Transform. g_i refers to the i^{th} decomposed Gaussian; $g_s = g_1 + g_2$ representing the systolic wave; $g_d = g_3 + g_4$ representing the diastolic wave. $\sum(g_i)$ is the area under the i^{th} Gaussian.		

6.3.2.1 Photoplethysmogram morphology

BP is determined by the product of total peripheral resistance (TPR) and CO. Therefore, features reflecting variations in these two metrics are likely to provide a strong indication of ΔBP .

Changes in TPR can significantly impact the morphology of the PPG through interactions between the incident forward travelling wave produced by left ventricular contraction and the reflected pressure waves [188]. A review of reflected pressure waves is provided in section 2.9.1. As TPR increases, the impedance mismatch at the renal and iliac arteries (the two sites responsible for reflected pressure waves) increases, and the resulting reflected pressure waves (known as the tidal and dicrotic waves respectively) increase in both velocity and amplitude. To model the varying amplitude of the reflected pressure waves, we included the dicrotic notch amplitude (N_{amp}), and the diastolic peak amplitude (labelled reflection index, RI, for consistent feature notation with respect to the literature) as features [307]. As the tidal and dicrotic reflected waves typically arrive during systole and diastole respectively, Lyu *et al.* [310] proposed the stress-induced vascular response index (sVRI) as the ratio of the mean of the PPG in the diastolic phase, μ_{Dias} , to the mean of the PPG in the systolic phase, μ_{Sys} . This effect was additionally modelled using the areas under PPG during the systolic (A_{Sys}) and diastolic (A_{Dias}) phases [311]. The ratio of these two areas ($A_{\text{Dias}}/A_{\text{Sys}}$) was labelled the inflection point area ratio (IPA) by Wang *et al.* [311].

The influence of the reflected pressure waves was additionally modelled through analysis of the frequency domain by assessing the level of distortion in the PPG waveform as the relative power of its harmonic frequencies [314]. This feature is referred to as the normalised harmonic area (NHA) [204] and was computed as:

$$\text{NHA} = \frac{\sum_{n=2}^{N_h} \text{FFT}^2(n \cdot f_1)}{\sum_{n=1}^{N_h} \text{FFT}^2(n \cdot f_1)} \quad (6.2)$$

where FFT is the Fast Fourier Transform of the PPG beat. f_1 is the fundamental frequency of the spectrum, the frequency of the highest power in the FFT. Harmonic frequencies occur at

integer factors of the fundamental frequency. N_h equates to the number of harmonics in the frequency spectrum. As we normalised the time duration of each PPG beat, the Nyquist limit of the frequency spectrum for the normalised beat is $(f_s \cdot T)/2$, where f_s is the original sampling frequency of the PPG (512 Hz) and T is the original time duration of the beat. N_h is therefore the number of integer factors of f_1 below this Nyquist limit.

The increased velocity of the reflected waves was modelled by relative time delays such as ΔT (time between S and D) [78], crest time (CT, time from onset to S) [192], T_{Sys} (time in systolic phase), T_{Dia} (time in diastolic phase), and T_{Ratio} ($T_{\text{Sys}}/T_{\text{Dia}}$) [308]. Additionally, the time delay between the systolic peak and dicrotic notch is related to the transit time of the dicrotic wave to the iliac artery junction and back. Shin *et al.* [202] suggest that this time delay may be calibrated to the individual by normalising it by the time delay between the maximum derivative of the PPG and the dicrotic notch, and multiplying the resultant by the individual's height (h). This feature is labelled pressure index (PI). Lin *et al.* [315] recently suggested that these time delay features may additionally reflect changes in CO. Wang *et al.* [204] put forward the inflection and harmonic area ratio (IHAR) as a parameter related to CO, computed as:

$$\text{IHAR} = \frac{1 - \text{NHA}}{\text{IPA}} \quad (6.3)$$

To model variations in the velocity of the incident forward travelling pulse wave produced by left ventricular contraction, we included the slope transit time (STT) as the gradient of the straight line from onset to the peak of the PPG waveform [309].

A small number of PPG_{fts} parameterise changes in the overall PPG morphology. For example, Width_{25} and Width_{50} track variations in the width of the PPG at 25% and 50% of the overall amplitude respectively [78, 201]. Skewness and kurtosis both characterise the distribution of the PPG beat values [213]. Skewness is a measure of the asymmetry of the data around the mean. Kurtosis is a measure of the relative "tailedness" of the distribution. For a PPG beat, ψ , they are both computed as:

$$\text{Skewness} = \frac{1}{N_\psi} \frac{\sum_{i=1}^{N_\psi} (\psi_i - \mu_\psi)^3}{\sigma_\psi^3}, \quad \text{Kurtosis} = \frac{1}{N_\psi} \frac{\sum_{i=1}^{N_\psi} (\psi_i - \mu_\psi)^4}{\sigma_\psi^4} \quad (6.4)$$

where N_ψ , μ_ψ and σ_ψ are the length, mean, and standard deviation of ψ respectively.

6.3.2.2 Velocity-plethysmogram morphology

We included four features derived from analysis of the VPG morphology. In order to model the changing velocity of the pulse wave during the two main phases of the cardiac cycle, we included the mean and standard deviation of the VPG during systole (Sys_μ and Sys_σ respectively) and diastole (Dia_μ and Dia_σ respectively) [215].

6.3.2.3 Acceleration-plethysmogram morphology

We included eight features derived from analysis of the APG morphology. The amplitude of each APG wave, normalised by the amplitude of a , exhibits a moderate linear relationship with age [79]. $\frac{b}{a}$ increase with age, whereas $\frac{c}{a}$, $\frac{d}{a}$, and $\frac{e}{a}$ decrease with age. These relationships were combined into a single feature using the ageing index (AGI) $\frac{(b-c-d-e)}{a}$ [79]. In addition, Mok Ahn *et al.* [308] proposed the gradient of the straight lines joining the b wave to the c and d wave (slope_{b-c} and slope_{b-d} respectively) as additional indicators of age-related arterial stiffness. Finally, Pilt *et al.* [312] suggested that the PPG augmentation index (PPG AI), computed as the ratio of the PPG waveform coincident with the d wave of the PPG (labelled d_{-2}) to the ratio of the PPG waveform coincident with the b wave of the PPG (labelled b_{-2}), reliably tracks changes in arterial stiffness.

6.3.2.4 Gaussian decomposition

Each PPG beat (of unit amplitude and duration) was decomposed into the summation of four Gaussians [193, 316]. This approach has the advantage of providing a representation of the PPG waveform without reliance on fiducial point detection. Additionally, it allows for the modelling of reflected wave interactions which are thought to have a Gaussian profile [317]. For a given PPG beat, ψ , we computed the modelled pulse, ψ^{Gauss} , as:

$$\psi^{\text{Gauss}}(t', \Theta) = \sum_{i=1}^4 g_i(t', \theta_i) \quad (6.5)$$

where g_i represents the i^{th} Gaussian component:

$$g_i(t', \theta_i) = A_{gi} \cdot \exp\left(-\frac{(t' - \mu_{gi})^2}{2\sigma_{gi}^2}\right) \quad (6.6)$$

θ_i is a vector, $[A_{gi}, \mu_{gi}, \sigma_{gi}]$, containing the amplitude, mean, and variance of each Gaussian respectively. $\Theta = [\theta_1, \theta_2, \theta_3, \theta_4]$ and thus Gaussian decomposition parameterises each PPG beat into 12 components. To determine the optimum value for Θ , $\tilde{\Theta}$, we implemented a bounded Levenberg-Marquart optimisation algorithm to minimise the root mean squared error loss, L^{Gauss} , between ψ and ψ^{Gauss} given as:

$$L^{\text{Gauss}}(\Theta) = \sqrt{\frac{1}{N} \sum (\psi - \psi^{\text{Gauss}}(t', \Theta))^2} \quad , \quad \tilde{\Theta} = \arg \min_{\Theta} (L^{\text{Gauss}}(\Theta)) \quad (6.7)$$

The optimisation procedure was bounded such that all parameters were positive and the amplitudes were all less than 1.0. Additionally, we constrained the sequence $[\mu_1, \mu_2, \mu_3, \mu_4]$ to be monotonically increasing. Given that L^{Gauss} is non-convex, the optimised values depended on the choice of initial conditions. For each beat, we ran two optimisations with different initial

conditions, and we kept the optimisation results that produced the lowest value of L^{Gauss} . The first set of initial conditions, shown in figure 6.2 (a) and outlined in table 6.3, resulted in a set of Gaussians with equal widths and amplitudes that matched the contour of a typical PPG beat. To encourage the continuity of parameters from beat-to-beat, we used the optimised parameters for the previous beat as the second set of initial conditions for the current beat. However, for the first beat in a PPG signal, we ignored the second set of initial conditions and used only the first set. The resulting Gaussian decomposition for a typical beat is illustrated in figure 6.2. Following the work of [318], we set the SQI_{PPG} of each beat to 0 if the value of $L^{\text{Gauss}}(\tilde{\Theta})$ for that beat was greater than 0.03.

Table 6.3: First set of initial conditions for Gaussian decomposition optimisation. These parameters resulted in an evenly distributed set of four Gaussians with equal widths. The amplitudes were determined in order to match the contour of a typical PPG beat, see figure 6.2.

	g_1	g_2	g_3	g_4
A_{gi}	0.9	$2/3$	$1/2$	$1/3$
μ_{gi}	0.2	0.4	0.6	0.8
σ_{gi}	0.015	0.015	0.015	0.015

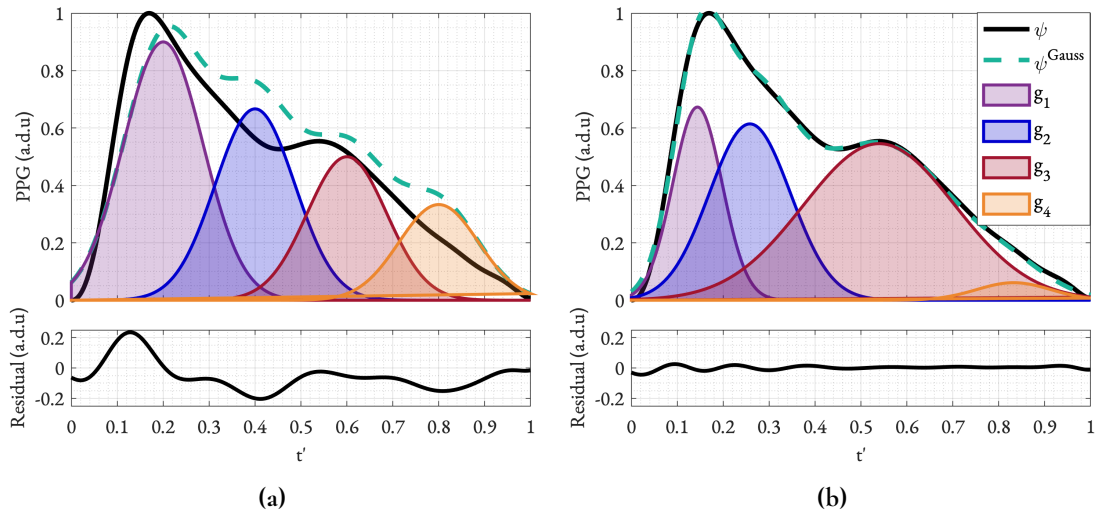


Figure 6.2: Gaussian decomposition of the PPG waveform. (a) The first set of initial conditions outlined in table 6.3 resulting in an evenly distributed set of Gaussians with equal width. (b) Optimised Gaussian decomposition. The optimisation process involved two sets of initial conditions, those shown in (a), and the optimised parameters from the previous beat. The results of whichever optimisation resulted in the lowest value of L^{Gauss} was kept.

Previous literature on pulse decomposition analysis provides physiological interpretations behind each of the four decomposed Gaussians [188]. g_1 reflects the incident forward travelling wave produced by left ventricular contraction. g_2 represents the tidal wave [194]. g_1 and g_2 are typically observed during systole, therefore $g_s = g_1 + g_2$ reflects the systolic component. g_3 represents the dicrotic wave [319]. g_4 represents the additional minor reflections and re-reflections in the systemic

vascular structure with lower amplitude than the two main reflected waves [316]. g_3 and g_4 are both typically dominant in the diastole phase, therefore $g_d = g_3 + g_4$ reflects the diastolic component.

To parameterise the characteristics of each wave, the values of $\tilde{\Theta}$ were included as features. A number of additional features modelled interactions between the Gaussians:

$$\text{Gauss AI} = \max(g_s) - A_{g3} \quad (6.8)$$

$$\text{Gauss RI} = \sum(g_s) - \sum(g_3) \quad (6.9)$$

$$\text{Gauss AI}_R = \frac{A_{g1} - A_{g2}}{A_{g1}} \quad (6.10)$$

$$\text{Gauss RI}_R = \frac{A_{g3}}{A_{g1}} \quad (6.11)$$

$$\text{Gauss RTT} = \mu_{g3} - \mu_{g1} \quad (6.12)$$

where $\sum g_i$ computes the area under the curve. The Gaussian augmentation index (Gauss AI) reflects the relative amplitude difference between the systolic component and the dicrotic wave [187]. The Gaussian reflection index (Gauss RI) is a similar parameter to Gauss AI, however, it reflects the contribution of the systolic component through the area under the curve [187]. Rubins *et al.* [193] additionally proposed separate parameters labelled the Gaussian augmentation index and the Gaussian reflection index. We use the subscript R to note this distinction. Gauss AI_R and Gauss RI_R parameterise the contribution of the tidal and dicrotic wave respectively. Gaussian reflection transit time (Gauss RTT) reflects the changing transit time of the dicrotic wave to the iliac arteries and back to the point of measurement [193].

Left ventricular ejection time (LVET), the time interval from the opening of the aortic valve to its subsequent closure, has been shown to be associated with CO [320]. We followed the method proposed by Couceiro *et al.* [206] to compute an approximation of LVET through the Gaussian decomposition of the PPG. The algorithm first computes the second derivative of g_s , g_s'' , using point differences. The first and third maxima of g_s'' reflect the onset and offset of systolic ejection and therefore the time delay between these two points reflects LVET. We label this feature Gauss LVET.

Finally, through observations of Gaussian decomposition in our dataset, we propose three new features: the ratio of the area under the systolic component to the area under the diastolic component (Gauss_{Sys/Dias}); the amplitude of the fourth Gaussian scaled by the amplitude of the first Gaussian (Gauss_{A4/A1}); and the variance of the fourth Gaussian scaled by the amplitude of the first Gaussian (Gauss _{$\sigma_4/A1$})¹.

¹Scaling by variance rather than the amplitude of the first Gaussian gave a less informative parameter.

6.3.2.5 Principal components

Principal component analysis (PCA) is a mathematical technique used to reduce the dimensionality of high-dimensional data while retaining its key features [321]. By mapping the data to a lower dimension along orthogonal principal axes, PCA identifies the principal components that account for the majority of the variation in the original data. This allows us to focus on the regions of the data that are most significant, and identify patterns and relationships that might not be apparent in the original high-dimensional space. In the context of PPG, VPG, and APG waveforms, PCA was used to highlight regions of significant change across the cohort. We computed PCA features using the following steps:

1. Resample all good-quality beats (defined as an $SQI_{PPG} > 0$), across all participants, from the PPG, VPG, and APG signals to be 100 samples in length using cubic spline interpolation.
2. Pool all resampled PPG, VPG, and APG beats from all participants to form 3 matrices: Ψ_{PPG} , Ψ_{VPG} , and Ψ_{APG} respectively.
3. Mean normalise each Ψ matrix.
4. Compute the covariance, Σ_{Ψ} , of the mean normalised Ψ matrix as:

$$\Sigma_{\Psi} = \frac{1}{N-1} \sum_{i=1}^N \Psi_i \Psi_i^T \quad (6.13)$$

where N is the number of beats in Ψ , and Ψ_i corresponds to the i^{th} column of Ψ .

5. Derive the eigenvectors and eigenvalues from Σ_{Ψ} using eigen-decomposition:

$$\Sigma_{\Psi} = V \Lambda V^{-1} \quad (6.14)$$

Each column of V is an eigenvector representing the direction of a principal component. Λ is a diagonal matrix with corresponding eigenvalues.

6. Extract the first 3 principal components that correspond to the largest eigenvalues.

We selected the first 3 principal components as this was found empirically to explain more than 85% of the variation in the Ψ_{PPG} , Ψ_{VPG} and Ψ_{APG} datasets. Figure 6.3 shows a visualisation of computed PCA eigenvectors (e.g. the visualisation of PPG PCA 1 is shown as the eigenvector corresponding to the largest eigenvalue in the covariance of Ψ_{PPG}).

Principal components analysis (see figure 6.3) highlights the regions of most significant change in the PPG, VPG and APG. The first principal component of the PPG shows a strong weighting on

the region corresponding to the location of the tidal wave (the first reflected wave). As discussed above, the amplitude of the tidal wave is significantly impacted by the changes in arterial tone and therefore large variations in this region were expected in our dataset. Whereas, the second component of the PPG was almost entirely focused on changes in the upslope, which has been previously suggested as a useful region for cuff-less measurement of BP [190]. All 3 components of the VPG reflect changes in the upslope. Less consistency was found in the APG. As we included PPG, VPG and APG beats from all participants in the study, some of the variation in the dataset may result from inter-participant differences as opposed to the intra-participant changes caused by the arterial stiffness modulation. However, we still chose to implement PCA across the whole cohort (as opposed to unique PCA per participant) in order to maintain consistency of feature definitions (i.e. so that PPG PCA 1 from one participant represents similar changes in another participant).

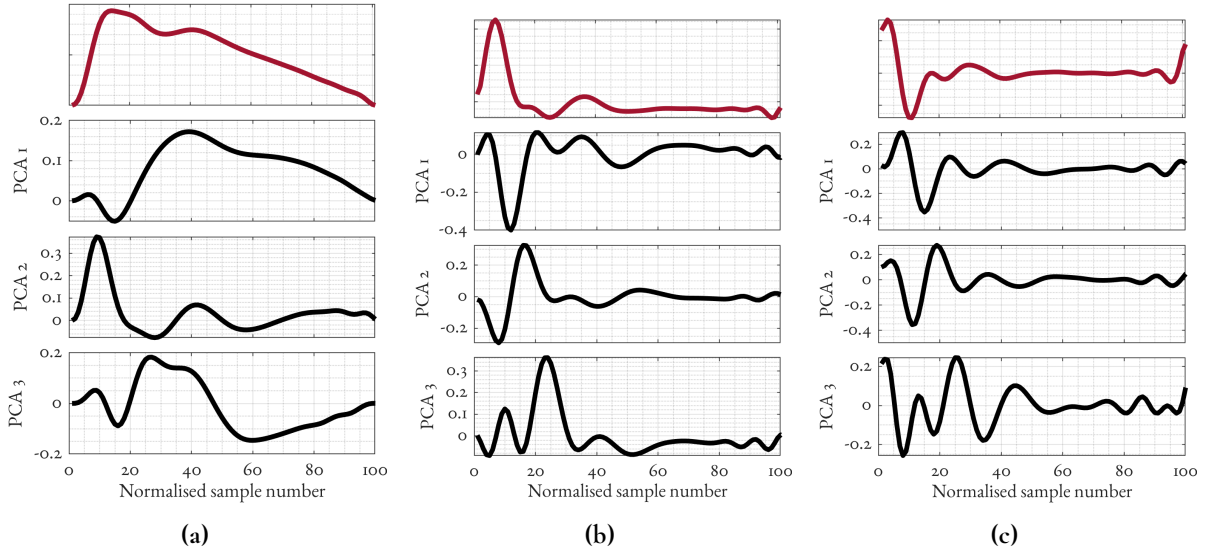


Figure 6.3: Visualisation of the 3 principal components of the (a) PPG, (b) VPG, and (c) APG waveforms using the corresponding eigenvectors. The top row, in red, shows a typical PPG, VPG and APG beat from the dataset.

6.4 Computing features from the electrocardiogram

To preprocess the ECG waveform, detect the cardiac beats, and assess the signal quality of the ECG (SQI_{ECG}), we followed the steps previously outlined in section 5.4.

6.4.1 Feature extraction

Features relating to complexity and entropy of the ECG signal have been previously proposed to track changes in BP [234, 236]. These features quantify the level of regularity and unpredictability of fluctuations over a time series. Generally, a higher-level complexity indicates a more irregular

dynamic system. A lower-level complexity indicates the presence of central trends or cyclical patterns. Changes in the entropy of the ECG time series have also been shown to track changes in HRV caused by myocardial ischaemia [237] and also denote periods of cardiac arrhythmia [238].

Table 6.4 provides a summary of the features we extracted from the ECG (ECG_{fts}) which are outlined in more detail in the subsequent sections. All features were computed within a short segmented window of the ECG time series. We outline our implementation of these windows in section 6.6.

Table 6.4: Summary of ECG_{fts} extracted. References to previous applications of the features are also provided.

Feature notation and publication	Description or formula
Hjorth mobility [234]	Estimate of the signal's mean frequency (equation (6.15))
Hjorth complexity [234]	Estimate of the signal's bandwidth (equation (6.15))
Fractal dimension [234]	Computed using Higuchi's algorithm [322] with $k_{\text{max}} = 17$
Shannon entropy (SE) [234]	Uncertainty of information content based on probability distribution (equation (6.16))
Approximate entropy (approxEnt) *	Quantifies regularities of signal
Sample entropy (sampEnt) *	approxEnt computed without self-comparisons
Multi-level sample entropy (MSE) *	Coarse approximations of sample entropy, computed at scales 2, 4, 6, 8

* Indicates features that, to our knowledge, have not been previously utilised for BP estimation.

6.4.1.1 Hjorth parameters

The Hjorth mobility and complexity parameters indicate activity variations in a signal [323]. The mobility parameter represents the signal's mean frequency. The signal complexity is an estimate of the signal bandwidth. For an ECG signal segment, ψ , Hjorth mobility and complexity were computed as:

$$\text{Mobility} = \sqrt{\frac{\text{var}(\psi')}{\text{var}(\psi)}} \quad \text{Complexity} = \frac{\text{Mobility}(\psi')}{\text{Mobility}(\psi)}. \quad (6.15)$$

where $\text{var}(\cdot)$ denotes the variance.

6.4.1.2 Fractal dimension

The fractal dimension of an ECG signal segment provides a measure of self-similarity, quantifying how the detail in the signal varies with respect to the length scale with which it is measured. The fractal dimension of a time series ranges between 1 (straight line) and 2 (defining a surface), where a higher number represents a signal with higher fluctuations and complexity. To calculate the fractal dimension, we used the Higuchi algorithm [322], whereby increasing length scales (up to k_{max}) are used to estimate the signal length. The fractal dimension was then defined as the proportionality

constant, representing the relationship between length scale and signal length. Brezinski [324] showed that increasing values of k_{max} beyond a value of 17 resulted in only a marginal increase in the fractal dimension for regular ECG beats in sinus rhythm. Therefore, we set k_{max} as 17 in this work.

6.4.1.3 Shannon's entropy

Shannon's entropy (SE) was used to measure the uncertainty of the information content in a time series based on its probability distribution. The probability distribution of an ECG signal segment, ψ , was estimated by a normalised histogram. SE was then computed as:

$$SE = - \sum_{k=1}^{N_\psi} p_k \cdot \log\left(\frac{p_k}{w_k}\right) \quad (6.16)$$

where p_k and w_k are the probability and width of the k^{th} bin of the histogram. N_ψ is the length of ψ .

6.4.1.4 Approximate entropy

Approximate entropy (approxEnt) quantifies the regularity of a time series and the likelihood that similar patterns of observations will not be followed by additional similar observations. For example, a time series containing many repetitive patterns has a small approxEnt. For an ECG signal segment, ψ , approxEnt was computed using the following steps:

1. Divide the signal, ψ , into consecutive segments of length m . Following the work of Li *et al.* [238], we set $m = 2$.
2. For each segment, i , compute the Chebyshev distance to all other segments and calculate $C_i(m, r)$ as the number of segments with a Chebyshev distance less than r to the i^{th} segment. Following the work of Li *et al.* [238], we set $r = 0.2$.
3. Define $\zeta(m, r)$ as the average number of segments of length m that are suitably similar to each other within a tolerance of r :

$$\zeta(m, r) = \frac{1}{N_\psi - m + 1} \sum_{i=1}^{N_\psi - m + 1} \ln(C_i(m, r)) \quad (6.17)$$

4. To compare $\zeta(m, r)$ to the subsequent data point, increase the dimension to $m + 1$ and compute $\zeta(m + 1, r)$
5. The approximate entropy was computed as:

$$\text{approxEnt} = \zeta^m(r) - \zeta^{m+1}(r) \quad (6.18)$$

6.4.1.5 Sample entropy

Sample entropy (sampEnt) is a modification of approxEnt that excludes self-comparison. In approxEnt, the comparison between each segment and the rest of the segments also includes a comparison of the segment with itself. As a result, signals are interpreted to be more regular than they actually are. These self-matches are not included in sampEnt, resulting in a more stable estimate of entropy, reducing bias. We computed sampEnt using the same hyperparameters ($m = 2$, and $r = 0.2$) as approxEnt.

6.4.1.6 Multi-scale entropy

Multi-scale entropy (MSE) is an extension of sample entropy. MSE is the application of sampEnt to the signal at increasingly coarser scales. For each s^{th} scale, the original signal samples are grouped into non-overlapping windows, of length s , and the windowed samples are averaged. sampEnt is then applied to this averaged signal. We computed MSE at scales 2, 4, 6 and 8.

6.5 Reference blood pressure values

As discussed in section 5.3, the BP values were smoothed and interpolated using a cubic smoothing splines algorithm to mitigate the effects of noise and to provide reference BP values sampled at once per minute.

6.6 Estimating changes in blood pressure

A schematic outlining our proposed steps for estimating ΔBP is shown in figure 6.4. We adopted a *hybrid calibration* [25] approach in order to estimate ΔBP using regression models (ordinary least squares, OLS, and random forest, RF) as a function of an input feature set, X , defined in the subsequent section. We calibrated all feature and BP values to the individual participant using data recorded during the rest period of the study (first 5 minutes of the recording). We then removed all collinear features and implemented data augmentation to increase the size of the training set, X_{Aug} .

We used nested leave-one-subject-out cross-validation (LOSOCV) to train, tune, and evaluate the models. For each fold of the LOSOCV, one participant in turn was set to be the test participant. Data from the X_{Aug} set for all participants apart from the test participant were used to train and tune the models. Nested LOSOCV was used for model validation to optimise model hyperparameters with the aim of minimising RMSE. Data from X for the test participant was then used to evaluate the performance of the model.

We used the following metrics to evaluate model performance: ρ_p , RMSE, and MAE. Differences in performance metrics across all folds from the models were evaluated for statistical significance by a two-tailed Wilcoxon signed-rank test (see appendix C.1). We adjusted the p-values for multiple comparisons using the Benjamini-Hochberg method (see appendix C.1.2).

6.6.1 Feature set: X

We implemented four different feature sets, $X \in \{\text{PPG}_{\text{fts}}, \text{ECG}_{\text{fts}}, \text{PPG}_{\text{fts}}+\text{ECG}_{\text{fts}}, \text{PAT}\}$. To explore ΔBP estimation using a single source of data, the first feature set was constructed using only PPG_{fts} and the second feature set was constructed using only ECG_{fts} . The power of combining these features was established by constructing the third feature set using both PPG_{fts} and ECG_{fts} . Finally, to compare the estimation of ΔBP using the features introduced in this chapter with PAT estimates, introduced in the previous chapter, the final dataset was constructed using only PAT.

Each observation in X , x_i , was computed using relevant feature values within a feature window w_f of length 40s centred on each reference BP measurement (20s to the left, 20s to the right). PPG_{fts} and PAT were estimated on a beat-by-beat basis thus, x_i consisted of the average values of these signals within w_f . Any beat with an SQI_{PPG} or SQI_{PAT} equal to zero was excluded from w_f . As all ECG_{fts} were window-based features, the segmented ECG signal within w_f was used to compute all observations of ECG_{fts} .

Window-based $\text{SQI}_{\text{PPG}}(w_f)$, $\text{SQI}_{\text{ECG}}(w_f)$, and $\text{SQI}_{\text{PAT}}(w_f)$ were then computed as their respective mean values within w_f . For each of PPG_{fts} , ECG_{fts} , and PAT if their respective window-based SQI fell below the $\gamma_{\text{SQI}} = 0.7$ threshold defined in section 5.6.2, then a valid measurement for that signal could not be made for that window. We handled missing data, caused by poor signal quality, by nearest neighbour imputation for each participant.

6.6.2 Calibration

We adopted a *hybrid calibration* [25] approach to personalise the estimation models for each participant in our dataset. We use the superscript p to reflect an observation from the p^{th} participant. In this work, for each participant, we defined the baseline calibration value of BP ($\text{BP}^{p, \text{calib}}$) and the j^{th} waveform feature ($f_j^{p, \text{calib}}$) as their respective mean values during the 5-minute resting period at the start of the study. In the specific example of one participant (ID = 008) for whom no BP readings were taken during the rest period, we used the mean values in a one-minute window centred on the first cuff inflation as the calibration readings.

For the p^{th} participant, BP values were subtracted from their calibration value to compute ΔBP_i^p , as observations of the regression target vector, y_i . Additionally, each observation of a

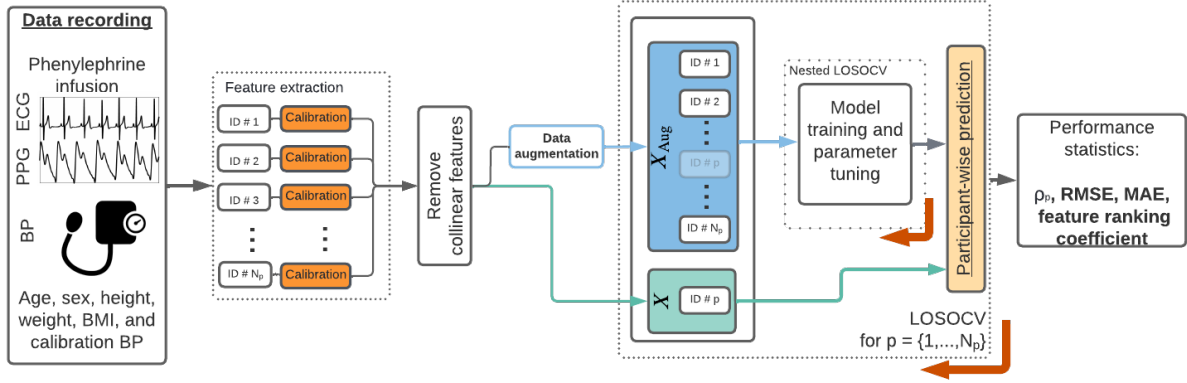


Figure 6.4: Schematic of the ΔBP estimation pipeline for each of the proposed models. We extracted features from the PPG and ECG and averaged their values at times of cuff inflations. We then implemented a *hybrid calibration* approach such that the proposed models estimate ΔBP from a baseline calibration value determined during the rest period. Data augmentation was implemented to increase the training and validation set sizes by interpolating between cuff inflations. Models were trained and evaluated in a nested leave-one-subject-out cross-validation (LOSOCV) framework shown here by the iterator p which indicates the test participant for that iteration. Participant p is then removed from the training/validation set (X_{Aug}) for that iteration.

feature in X , x_{ij} , consisted of relative changes of that feature from its calibration value [190]. Thus ΔBP and x_{ij}^p took the form:

$$y_i = \Delta\text{BP}_i^p = \text{BP}_i^p - \text{BP}^{p, \text{calib}}, \quad x_{ij}^p = \frac{f_{ij}^p - f_j^p}{f_j^{p, \text{calib}}}, \quad \begin{aligned} i &= \{1, \dots, N\}, \\ j &= \{1, \dots, M\}, \\ p &= \{1, \dots, N_p\} \end{aligned} \quad (6.19)$$

where N and M are the number of observations and features in X respectively. N_p is the number of participants.

Hybrid calibration strategies use participant demographics to personalise the model outputs. Thus, for all feature sets, we added participant age, sex, height, weight, BMI, and BP calibration ($\text{BP}^{p, \text{calib}}$) as static categorical features.

6.6.3 Baseline reference

For reference performance metrics, we implemented a simple *baseline reference* that assumed no BP changes for each participant from their calibration value $\text{BP}^{p, \text{calib}}$ (i.e. $\Delta\text{BP}_i^p = 0$ for all i and p). We refer to this as *baseline reference* and it indicates the minimum performance that must be achieved by the regression models.

6.6.4 Removing collinear features

Collinearity occurs when there is intercorrelation between multiple features [325, 326], thus violating the independent identically distributed (i.i.d.) assumption that is common in regression

models. Additionally, the presence of collinearity inflates the variance of the regression parameters and makes it difficult to assess the importance of features. Collinearity in a feature set, X , can be highlighted by the condition number κ representing the ratio of the largest singular value of X to the smallest singular value. It can be computed as:

$$\kappa(X) = \|X^+\| \cdot \|X\| \quad (6.20)$$

where $\|\cdot\|$ is the 2-norm of a matrix and X^+ is the pseudo-inverse of the matrix X . Typically, a condition number greater than 30 is thought to indicate the presence of strong multicollinearity in the dataset [326].

It is likely that collinearity exists in the feature sets presented in this chapter as there are multiple features describing similar characteristics; for example, entropies of the ECG or time durations of the PPG. In order to mitigate the effect of collinearity and to allow for parsimonious models, we removed collinear features by investigating the variance inflation factor (VIF), defined as:

$$\text{VIF}_j = \frac{1}{1 - R_j^2} \quad (6.21)$$

where R_j^2 is the unadjusted coefficient of determination for regressing the j^{th} feature on the remaining ones. If a feature can be accurately predicted using one, or more, other features then VIF_j would be large for that feature. Removing collinear features is an iterative process whereby on each iteration, the feature with the largest corresponding VIF_j is removed from the feature set until no features had a VIF greater than 10 (corresponding to $R_i = 0.9$) [326].

6.6.5 Data augmentation

We implemented data augmentation in order to increase the feature set size whilst incorporating information on feature values between the reference BP measurements. As shown in figure 6.4, we performed model training and validation on the augmented feature set, hereafter referred to as X_{Aug} , and performance metrics were reported using the original dataset X . We constructed X_{Aug} by interpolating between the reference BP values for each participant using the cubic smoothing splines algorithm (see section 5.3) at a new frequency, $f_{\text{BP}} = 1/15$ Hz (four measurements a minute, as opposed to one a minute) with a smaller feature window size, $w_f = 15$ s to prevent overlapping windows violating the i.i.d assumption common in regression models (see appendix D.1.1).

6.6.6 Regression models

The models we implemented to estimate ΔBP are outlined below. Different models were independently built for estimating ΔSBP , ΔMAP , and ΔDBP .

6.6.6.1 LASSO+OLS

To explore the linear relationship between each of the feature sets X and ΔBP , ordinary least squares (OLS) linear regression was implemented. This model is introduced in appendix D.1.1. To prevent over-fitting, reduce the dimensionality, and improve model interpretability, we employed the Least Absolute Shrinkage and Selection Operator (LASSO) method to remove redundant features prior to linear regression (see appendix D.1.4). We refer to this model as LASSO+OLS. We optimised the λ hyper-parameter by a nested LOSOCV loop to minimise RMSE. For each loop of the LOSOCV, LASSO feature selection was implemented and features with non-zero coefficients were used by OLS to compute ΔBP estimates.

We refer to the different LASSO+OLS and input feature set, $X \in \{\text{PPG}_{\text{fts}}, \text{ECG}_{\text{fts}}, \text{PPG}_{\text{fts}}+\text{ECG}_{\text{fts}}, \text{PAT}\}$, combinations as $\text{LASSO+OLS}_{\text{PPG}}$, $\text{LASSO+OLS}_{\text{ECG}}$, $\text{LASSO+OLS}_{\text{PPG+ECG}}$, and $\text{LASSO+OLS}_{\text{PAT}}$ respectively.

6.6.6.2 Random forest

To explore a potential non-linear relationship between X and ΔBP , we additionally trained a Random Forest (RF) regression model. This model is introduced in section 4.4.2. We optimised the number of features randomly selected for each node (labelled *mtry*) by a nested LOSOCV loop to minimise RMSE.

We refer to the different RF and input feature set, $X \in \{\text{PPG}_{\text{fts}}, \text{ECG}_{\text{fts}}, \text{PPG}_{\text{fts}}+\text{ECG}_{\text{fts}}, \text{PAT}\}$, combinations as RF_{PPG} , RF_{ECG} , $\text{RF}_{\text{PPG+ECG}}$, and RF_{PAT} respectively.

6.6.7 Shapley additive explanation values feature importance ranking coefficient

A key objective of this chapter was to highlight features that have strong predictive power for estimating ΔBP . We assessed the importance of each feature through SHAP values [275]. This method is introduced in section 4.5. SHAP values, ϕ_j^i , represent the marginal contribution of the j^{th} feature to the i^{th} model prediction. An explanation model, $f_e(\cdot)$, is developed which represents the model predictions as the summation of ϕ_j^i over all j :

$$f_e(x_i) = \phi_0 + \sum_{j=1}^M \phi_j^i \quad (6.22)$$

where ϕ_0 is the expected model value over the training set.

For each loop of the LOSOCV, we assessed the overall importance of each feature as the mean absolute SHAP value across the training data:

$$\text{imp}_j = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} |\phi_j^i| \quad (6.23)$$

where N_{train} is the number of observations in the training set. imp_j indicates the average impact of the j^{th} feature on model output magnitude.

To report the variability of the SHAP values feature importance across the LOSOCV, we computed a ranking coefficient. The ranks of all features were determined at each fold and normalised by the total number of features (1 being the highest rank, 0 being the lowest). For each feature, the distributions of the ranking coefficients across all folds were analysed. This step allowed for each fold of the LOSOCV to contribute equally towards assessing the overall importance, while also providing a fair comparison across ΔSBP , ΔMAP , and ΔDBP estimation.

The comparison of feature ranking coefficients for estimating ΔSBP , ΔMAP , and ΔDBP was performed using Spearman's correlation coefficient (ρ_s) [327]. ρ_s measures monotonic relationships between two variables x and y , and is defined as:

$$\rho_s = 1 - \frac{6 \sum_{i=1}^M d_i^2}{M(M^2 - 1)} \quad (6.24)$$

where d_i is the difference between the two ranks of each x_i and y_i observation. ρ_s values can be interpreted in the same way as ρ_p .

6.7 Results

6.7.1 Removing collinear features

Figure 6.5 (a) shows the correlation matrix of the total feature set ($\text{PPG}_{\text{fts}} + \text{ECG}_{\text{fts}} + \text{demographics}$), including 77 features. There were a large number of features (58.5%) which were significantly correlated ($|\rho_p| > 0.8$, $p < 0.05$) with at least one other feature. This indicates a high level of collinearity, highlighted by a condition number κ of 315. Figure 6.5 (b) shows the correlation matrix after removing all collinear features. The remaining dataset contained 45 features with a condition number κ of 11, suggesting independence of features and encouraging parsimonious models.

Table 6.5 outlines the remaining PPG_{fts} and ECG_{fts} and the subset of the total feature set with which they had a strong correlation (defined as $|\rho_p| > 0.8$, $p < 0.05$). Additionally, the linear correlation with ΔSBP for each feature across the whole cohort and on a participant-wise basis is provided. The overall correlations of these features to ΔSBP across the cohort was in general quite low with only one feature ($\text{Gauss}_{\sigma 4/A1}$) having $|\rho_{p\Delta\text{SBP}}| > 0.5$. There were a number of features that had significant participant-wise correlations (PWC) to ΔSBP , 21 features with an absolute median $\text{PWC} > 0.5$.

Table 6.5: Resulting PPG_{fts} and ECG_{fts} included in the feature set post removing collinearity. ‘Correlated features’ indicates the subset of the total feature set with which each included feature demonstrated a strong linear correlation with ($|\rho_p| > 0.8$, $p < 0.05$). Features are ordered by their absolute median participant-wise correlation with Δ SBP. ECG_{fts} are highlighted in red for clarity.

Feature name	Correlated features	$\rho_p \Delta$ SBP	p	PWC [q1, q3]
Gauss σ_4/A_1	PPG AI; Gauss A_4/A_1 ; A_{g4} ; μ_{g4} ; σ_{g4}	0.51	< 0.001	[0.57, 0.92]
σ_{g1}	Gauss _{sys/dias} ; Gauss RTT; μ_{g1} ; μ_{g2} ; σ_{g2} ; μ_{g3}	-0.25	< 0.001	[-0.90, -0.49]
APG PCA ₁		-0.11	< 0.01	[-0.91, -0.12]
VPG PCA ₁		-0.18	< 0.001	[-0.88, 0.04]
d / a		-0.18	< 0.001	[0.50, 0.88]
IHAR	T _{Sys} ; T _{Dia} ; T _{Ratio} ; A _{Sys}	-0.30	< 0.001	[-0.80, -0.38]
Hjorth mobility		-0.42	< 0.001	[-0.85, -0.47]
c / a		0.02	0.576	[0.16, 0.84]
Gauss RI _R	A_{g1}	0.41	< 0.001	[0.29, 0.72]
approxEnt		-0.17	< 0.001	[-0.78, -0.21]
Gauss AI		0.15	< 0.001	[-0.81, -0.20]
NHA	N _{amp} ; Sys_μ	0.12	< 0.01	[0.11, 0.73]
Width ₅₀	Width ₂₅	0.17	< 0.001	[-0.32, 0.75]
MSE scale 8	MSE scale 4; MSE scale 6	-0.28	< 0.001	[-0.77, -0.20]
IPA	T _{Dia} ; N _{amp} ; RI; A _{Dias} ; sVRI; Sys_μ	0.30	< 0.001	[0.26, 0.85]
A_{g2}		-0.25	< 0.001	[-0.74, -0.28]
Kurtosis		0.42	< 0.001	[-0.11, 0.92]
Fractal dimension		-0.41	< 0.001	[-0.83, -0.13]
Gauss RI		0.23	< 0.001	[-0.12, 0.71]
PPG PCA ₁		-0.16	< 0.001	[-0.72, 0.01]
σ_{g3}	μ_{g4}	-0.31	< 0.001	[-0.79, -0.29]
slope _{bd}	AGI	-0.11	< 0.01	[-0.75, -0.06]
VPG PCA ₃		-0.32	< 0.001	[-0.79, 0.20]
e / a		0.21	< 0.001	[-0.03, 0.76]
Gauss LVET		0.32	< 0.001	[-0.19, 0.55]
PPG PCA ₃		0.06	0.121	[-0.68, -0.07]
APG PCA ₃		0.09	< 0.05	[-0.69, 0.56]
APG PCA ₂		0.00	0.933	[-0.67, 0.43]
b / a	A_{g1}	-0.25	< 0.001	[-0.72, 0.19]
sampEnt	MSE scale 2; MSE scale 4	-0.01	0.789	[-0.55, 0.10]
Hjorth complexity		0.21	< 0.001	[-0.07, 0.48]
VPG PCA ₂		0.11	< 0.01	[-0.65, 0.30]
PPG PCA ₂		-0.02	0.695	[-0.62, 0.25]
A_{g3}		-0.18	< 0.001	[-0.63, -0.02]
Gauss AI _R		-0.08	0.052	[-0.56, 0.47]
Sys σ		-0.03	0.469	[-0.29, 0.52]
Dia σ		0.13	< 0.01	[-0.65, 0.30]
SE		0.02	0.527	[-0.42, 0.33]
STT	CT	0.09	< 0.05	[-0.55, 0.36]
Skewness		-0.05	0.212	[-0.57, 0.58]

$\rho_p \Delta$ SBP: correlation with Δ SBP across all participants, with resulting p -value. PWC: participant-wise correlation, q1 and q3 are the first and third quartile PWC values across the cohort.

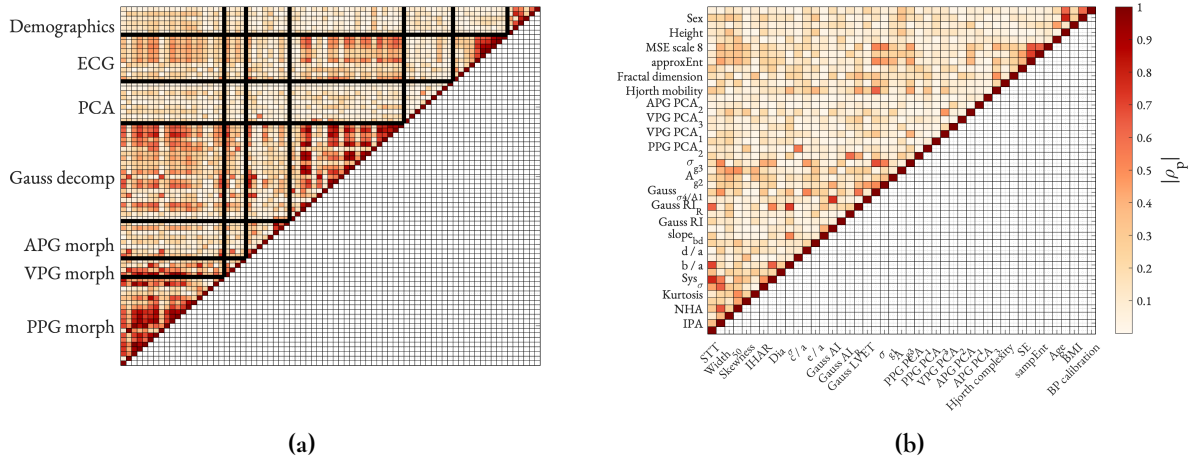


Figure 6.5: Results of removing collinear features in PPG + ECG feature set. Pairwise linear absolute Pearson's $|\rho_p|$ correlation matrix. (a) pre and (b) post removal of collinear features. Blank spaces represent non-significant ($p > 0.05$) or weak ($|\rho_p| < 0.2$) correlations. For clarity of labelling on the post-collinear feature set (b), every other label is presented on the y-axis with the remaining labels on the x-axis. Abbreviations: Gauss decomp - Gaussian decomposition, APG morph - features derived from the morphology of the APG, VPG morph - features derived from the morphology of the VPG, PPG morph - features derived from the morphology of the PPG.

6.7.2 Model performance

Table 6.6 shows the median and IQR performance statistics for all models computed over all 26 folds of the LOSOCV for Δ SBP estimation. The results for Δ MAP and Δ DBP estimation are provided in the appendix in table F.1 and table F.2 respectively.

Table 6.6: Performance statistics of Δ SBP estimation using the models proposed. Results are given as median (IQR) computed across all folds of the LOSOCV. Entries in bold indicate the best performance for that metric.

Model name	ρ_p	RMSE *	MAE *
Baseline reference	- (-)	10.46 (6.18)	7.40 (4.87)
LASSO+OLS _{PPG}	0.85 (0.44)	6.19 (4.05)	5.41 (2.41)
RF _{PPG}	0.88 (0.26)	5.75 (4.73)	4.75 (3.66)
LASSO+OLS _{ECG}	0.64 (0.70)	7.48 (6.37)	6.54 (4.63)
RF _{ECG}	0.69 (0.45)	6.79 (5.66)	5.28 (4.57)
LASSO+OLS _{PPG+ECG}	0.82 (0.40)	7.99 (5.74)	6.30 (4.62)
RF _{PPG+ECG}	0.86 (0.23)	5.66 (4.76)	4.86 (4.29)
LASSO+OLS _{PAT}	0.83 (0.10)	6.09 (4.25)	4.91 (3.33)
RF _{PAT}	0.80 (0.15)	6.39 (4.23)	5.21 (3.69)

* results given in units of mmHg. [†] Baseline reference assumed no changes in SBP for each participant from SBP_{calib}.

All models outperformed the results obtained with the *baseline reference*, and all apart from LASSO+OLS_{ECG} reported statistically significant p -values indicating that consistent improvements in performance statistics were observed ($p < 0.05$ for all). For the PPG_{fts}, ECG_{fts}, and

PPG_{fts}+ECG_{fts} feature sets, the RF model consistently achieved stronger performance metrics than LASSO+OLS. When comparing these models, statistically significant improvements were recorded only in RMSE and MAE ($p < 0.05$ for all), whereas the Wilcoxon signed-rank test failed to reject the null hypothesis of equal median ρ_p between LASSO+OLS and RF.

The PPG_{fts} feature set significantly outperformed the ECG_{fts} feature set for all performance metrics and regression models. The absolute difference in the median ρ_p , RMSE, and MAE between RF_{PPG} and RF_{ECG} was 0.19 ($p = 0.00007$), 1.04 mmHg ($p = 0.005$) and 0.53 mmHg ($p = 0.004$) respectively. RF_{PPG} and RF_{PPG + ECG} reported similar performance statistics with non-significant p -values, indicating that adding ECG_{fts} to a feature set of PPG_{fts} offers little or no performance gain. For PAT, LASSO+OLS significantly outperformed RF across all performance metrics ($p < 0.05$ for all). LASSO+OLS_{PAT} achieved similar performance metrics to the RF_{PPG + ECG}. LASSO+OLS_{PAT} consistently resulted in the smallest IQR for all performance metrics.

Figure 6.6 shows the individual reference Δ SBP and estimated Δ SBP values using the RF_{PPG + ECG} model across all participants. These results are presented for Δ MAP and Δ DBP estimation in appendix F (figures F.3 and F.4 respectively). Under the results for each participant, we show the accumulated SHAP values for each observation grouped into PPG_{fts}, ECG_{fts}, and demographics. PPG_{fts} consistently demonstrated significant positive SHAP values during peak infusion. At peak infusion, the median (IQR) value of Δ SBP across the cohort was 20 (8) mmHg. We found large errors in the data for the four participants (IDs: 002, 010, 023, 026) whose Δ SBP at peak infusion exceeded 30 mmHg.

Figure 6.7 summarises these results using scatter and Bland-Altman plots for Δ SBP estimation using the RF_{PPG + ECG} model. Figures F.1 and F.2 in the appendix shows the scatter and Bland-Altman plots for Δ MAP and Δ DBP estimation respectively. The data from individual participants are colour and marker-coded. The ρ_p value of the overall estimation was 0.64. The median participant-wise correlation coefficient was 0.86, with a range of 0.34 to 0.95. Figure 6.7 (a) shows the distributions of the reference and estimated values. The reference Δ SBP values ranged from -16.4 to 53.8 mmHg, but the estimated Δ SBP values had a much tighter range of -3.37 to 22.2 mmHg. The bias of the overall error was 0.30 mmHg with a standard deviation of 8.05 mmHg (see figure 6.7 (b)). We note also an additional bias whereby large values of Δ SBP were underestimated and small values were overestimated.

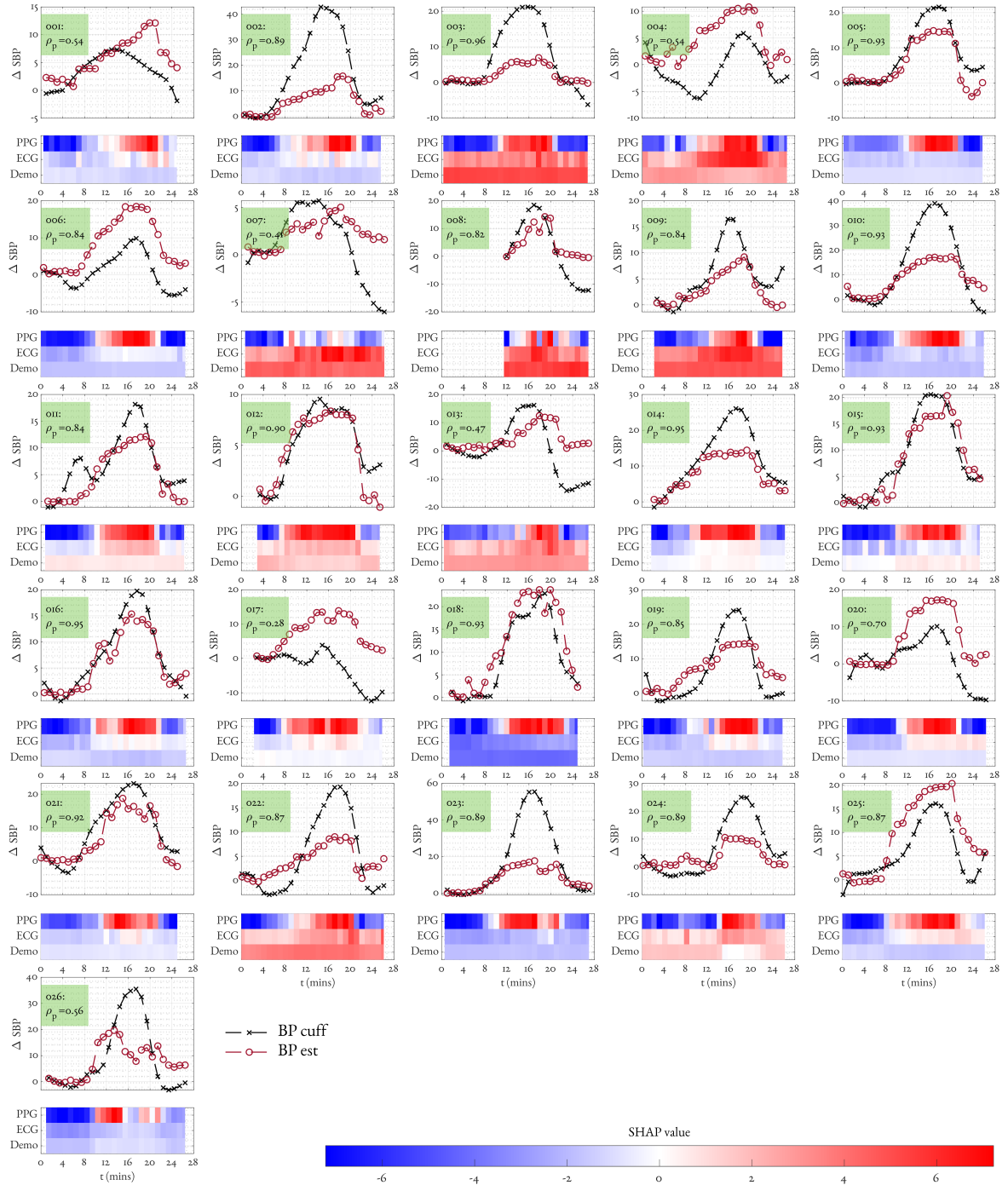


Figure 6.6: Individual results of Δ SBP estimation using the $RF_{PPG + ECG}$ model across all participants in the study. Note the differences in the y-axis scale. The reference Δ SBP values from the cuff are shown in black and the estimated values are shown in red. Under each participant's data, we show the summed SHAP values for each observation grouped into PPG_{fts}, ECG_{fts}, and demographics. Red hues indicate a positive marginal contribution to the model prediction for each observation. Blue hues indicate a negative marginal contribution to the model prediction for each observation.

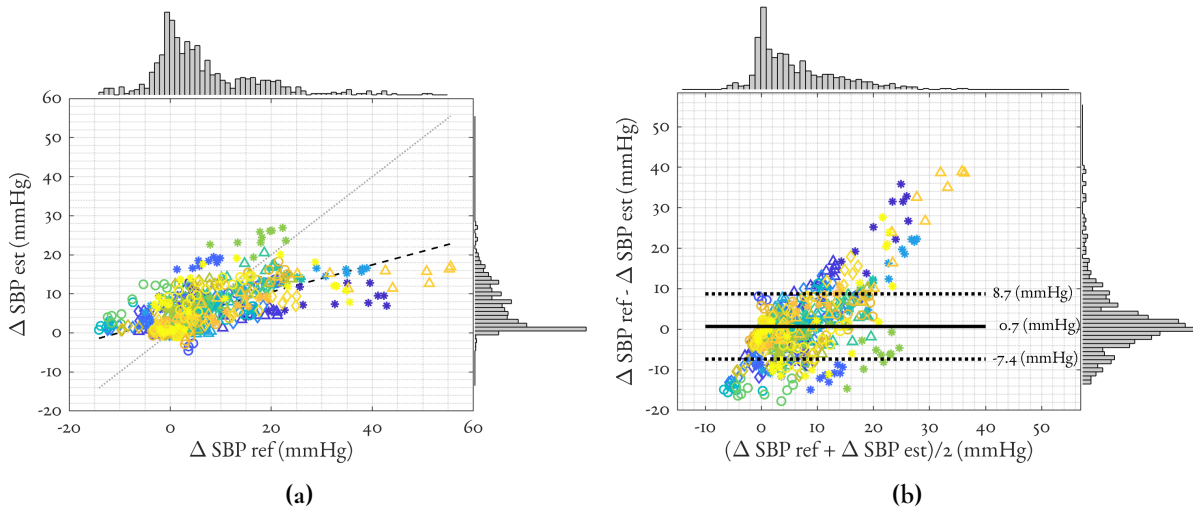


Figure 6.7: Agreement between the reference ΔSBP values and the estimated ΔSBP across the cohort using $\text{RF}_{\text{PPG} + \text{ECG}}$. Data from individual participants are colour and marker-coded. (a) Scatter plot, with an overall correlation of 0.64, a median participant-wise correlation coefficient of 0.86 with a range of 0.34 to 0.95. The black striped line shows the regression line. The Grey dotted line shows the line of unity. (b) Bland-Altman analysis, with the bias of the overall error at 0.3 mmHg, with a standard deviation of 8.05 mmHg

6.7.3 Feature importance

To mitigate variations in the training data between folds of the LOSOCV, the overall SHAP value feature importance was assessed through a ranking coefficient. Figure 6.8 shows the median (across folds) RF and LASSO ranking coefficients for the PPG+ECG feature set for predicting ΔSBP , ΔMAP , and ΔDBP .

Table 6.7 quantifies the agreement between the feature ranks for pairs of ΔSBP , ΔMAP , and ΔDBP using Spearman's correlation coefficient (ρ_s). It was found that the feature importance showed good agreement with each other ($\rho_s > 0.55$ for all) and as a result, we investigate feature importances only for SBP estimation in this chapter. The top five features as determined by the median RF ranking coefficient were: kurtosis, σ_{g1} , $\text{Gauss}_{\sigma 4/A1}$, Hjorth mobility, and VPG PCA_1 . The top five features as determined by the median LASSO+OLS ranking coefficients were: kurtosis, $\text{Gauss}_{\sigma 4/A1}$, Gauss LVET, BP calibration, and IPA. Kurtosis had the highest feature importance in 13 out of 26 folds for both RF and LASSO+OLS (2nd highest in the remaining folds). In general, the ranking of features by LASSO+OLS followed the order of feature correlation values shown in table 6.5. The only ECG feature that demonstrated significant importance in ΔBP estimation was Hjorth mobility which had a median ranking coefficient of 0.93 for RF.

Figure 6.9 shows the relationship between ΔSBP and the top 9 ranking, non-demographic, features from the $\text{RF}_{\text{PPG} + \text{ECG}}$ model. The data from different participants are coded by colour and marker to highlight clusters of features suggesting individual-specific feature changes. Both

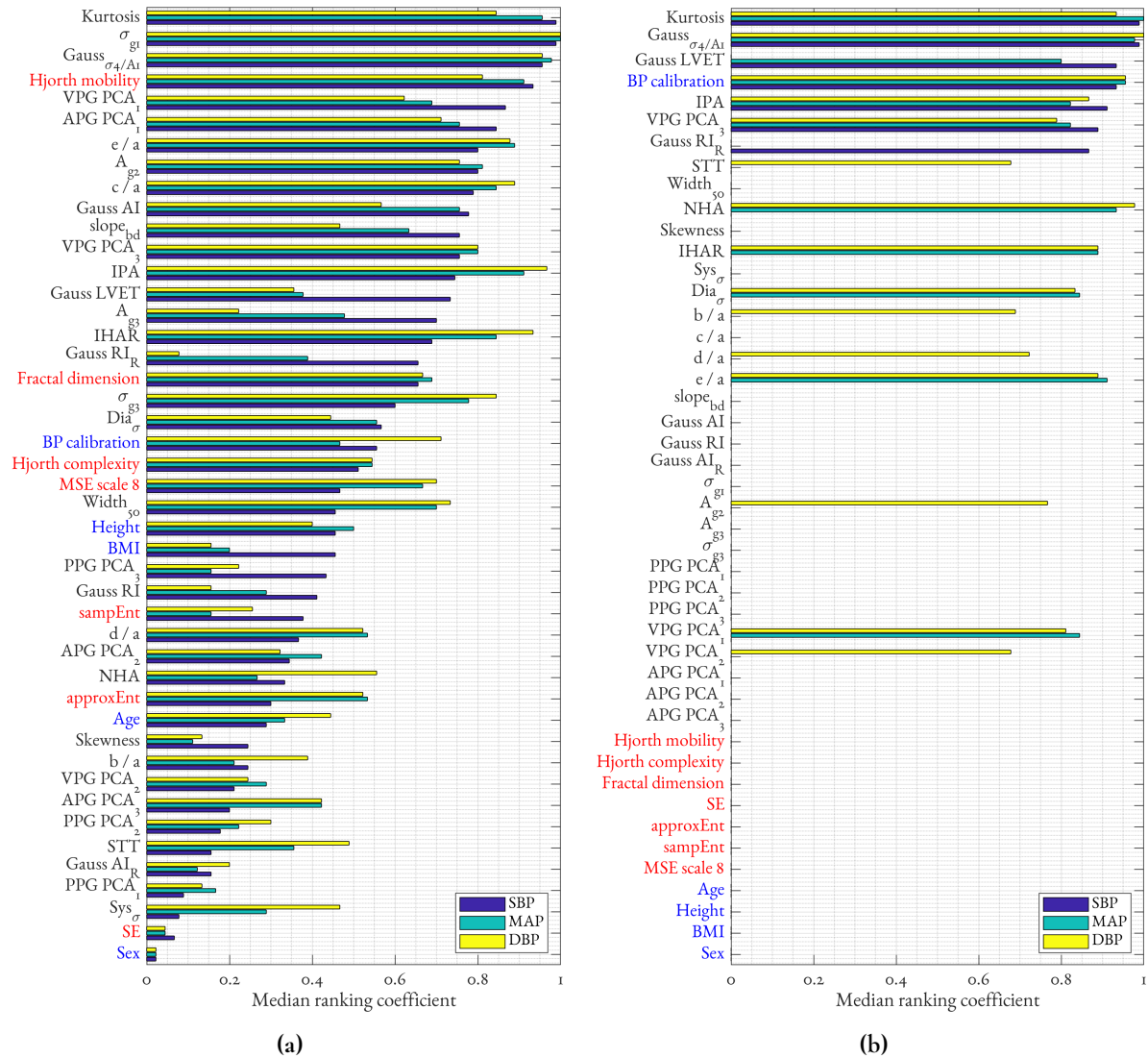


Figure 6.8: Median SBP, MAP, and DBP SHAP values feature ranking coefficients for (a) RF and (b) LASSO+OLS both with PPG + ECG feature set. As LASSO allows feature coefficients to equal zero, many features had a median ranking coefficient of zero. Features are ordered by their respective median ranking coefficient for SBP. In instances where the median ranking coefficient of SBP is equal (such as all features in LASSO+OLS with a zero ranking coefficient), features are ordered by their order of appearance in this chapter. Demographic features are highlighted in blue and ECG features in red for clarity.

Table 6.7: Spearman's correlation coefficient (ρ_s) quantifying the agreement between the feature ranking coefficients for pairs of Δ SBP, Δ MAP, and Δ DBP.

	(Δ SBP, Δ MAP)	(Δ MAP, Δ DBP)	(Δ DBP, Δ SBP)
LASSO+OLSPPG+ECG	0.63	0.83	0.55
RFPPG+ECG	0.85	0.90	0.70

Pearson's ρ_p and Spearman's ρ_s correlation coefficients are provided. ρ_s indicates monotonic (but not necessarily linear) relationships and so may provide further insight into the RF models (see for example the scatter plot of σ_{g1}).

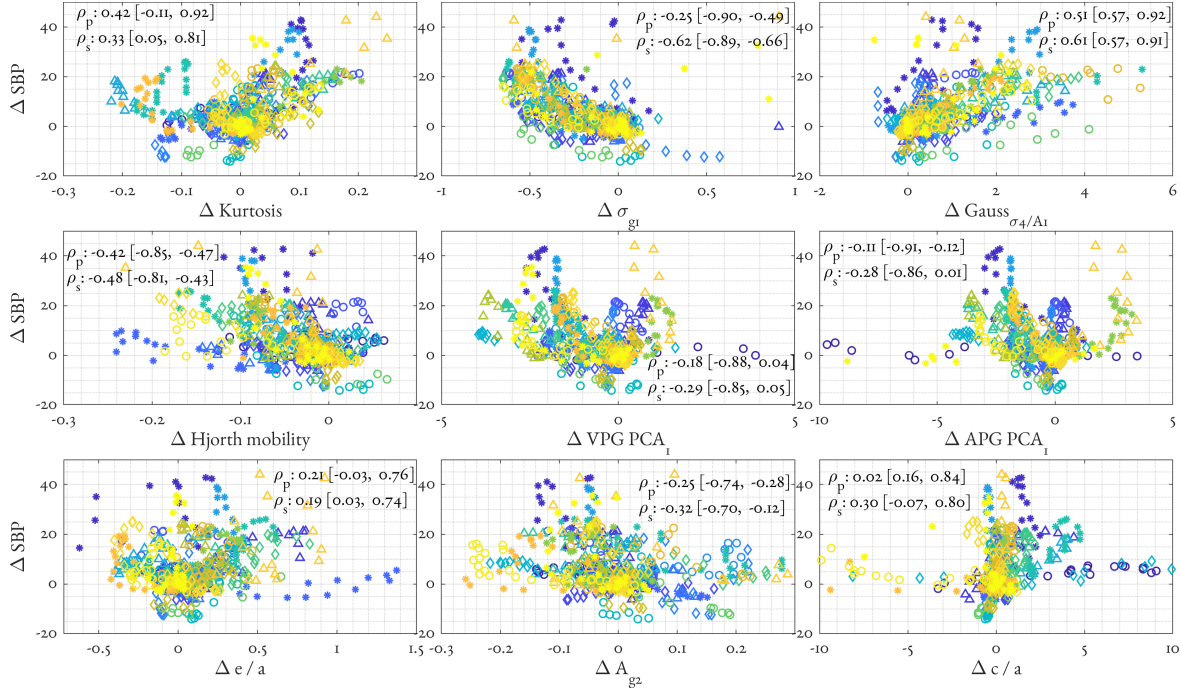


Figure 6.9: Relationship between the top 9 ranking, non-demographic, features from the RF_{PPG+ECG} model and ΔSBP . The data from individual participants are colour and marker-coded. Inter Pearson's ρ_p and Spearman's ρ_s correlation coefficients are shown as median and upper and lower quartiles. Additionally, the 1st and 3rd quartiles of the participant-wise correlation coefficients are shown. Note that feature values have been both subtracted and divided by their value at baseline, see section 6.6.2.

6.8 Discussion

6.8.1 Observed changes in the photoplethysmogram waveform morphology

Figure 6.1 shows the changes in PPG beat for a typical participant during the four main stages of our study: rest, dose increase, max infusion, and washout. These changes include a loss of a clear diastolic notch and a rising middle peak (referred to as the tidal wave [194] and thought to be caused by reflected waves at the renal arteries [188]) that can overshadow the initial peak seen at rest. These changes have been reported previously [215, 317, 78] and are thought to be due to increased amplitude and speed of reflected waves due to arterial stiffening. Additionally, they have also been shown to occur for age-related arterial stiffening [194].

During the phenylephrine infusion, we also observed variations in the amplitude and time duration of the PPG waveform morphology. Many factors can impact the amplitude of the PPG beat, such as respiration [305], changes in the contact pressure of the pulse oximeter [181], and changes in

peripheral blood volume [28]. As these factors can occur independently of changes in blood pressure, we normalised the PPG beat to have unit amplitude, as recommended in [181]. Additionally, the phenylephrine infusion activates the baroreflex response, which increases the parasympathetic nerve activity to the heart, slowing down HR [328]. Therefore, we additionally normalised the time duration of the PPG pulse, allowing features such as ΔT to be independent of HR.

Principal components analysis (see figure 6.3) highlights the regions of most significant change in the PPG, VPG, and APG. The first principal component of the PPG shows a strong weighting on the region corresponding to the location of the tidal wave (the first reflected wave). Whereas, the second component of the PPG was almost entirely focused on changes in the upslope, which has been previously suggested as a useful region for cuff-less measurement of BP [190]. All 3 components of the VPG reflect changes in the upslope. Less consistency was found in the APG. As we included PPG, VPG, and APG beats from all participants in the study, some of the variation in the dataset may result from inter-participant differences as opposed to the intra-participant changes caused by the arterial stiffness modulation. However, we still chose to implement PCA across the whole cohort (as opposed to unique PCA per participant) in order to maintain consistency of feature definitions (i.e. so that PPG PCA 1 from one participant represents similar changes in another participant).

6.8.2 Model performance

In order to increase the training set size, the models were trained, tuned, and evaluated within a LOSOCV framework. We implemented data augmentation by interpolating between the cuff measurements using cubic smoothing splines with an additional 3 augmented measurements every minute. Across all models, data augmentation significantly improved the performance statistics at the $p < 0.001$ level, computed using a Wilcoxon signed-rank test. Figure 6.5 (a) splits all features into groups (PPG morphology, VPG morphology, APG morphology, Gaussian decomposition, PCA, ECG, and demographics) and shows strong intra-group correlations indicating multi-collinearity in the dataset. We reduced feature collinearity by removing features with a VIF > 10 . This increased the confidence of the model parameters, allowed for an appropriate examination of feature importance, and encouraged parsimonious models.

For the PPG_{fts} , ECG_{fts} , and $\text{PPG}_{\text{fts}} + \text{ECG}_{\text{fts}}$ feature sets, the RF model significantly outperformed the LASSO+OLS model at the $p < 0.05$ level in terms of RMSE and MAE. A non-significant difference was found for ρ_p . This indicates that LASSO+OLS was able to detect directional changes in ΔSBP , although it was unable to determine the magnitudes of these changes. This points to a calibration issue whereby the LASSO+OLS model was unable to approximate the feature calibration gradients for each participant. RF models may have been better at estimating

the feature calibration gradients as these models allow for interactions between features and demographics. Indeed, larger overall importance was placed on the static demographic features by $\text{RF}_{\text{PPG}+\text{ECG}}$ than $\text{LASSO}+\text{OLS}_{\text{PPG}+\text{ECG}}$ (see figure 6.8). In linear models, these static features can only influence intercepts and not gradients. Thus, since the feature set was calibrated in order to predict ΔBP from a baseline calibration value, low overall importance was placed on demographics by $\text{LASSO}+\text{OLS}_{\text{PPG}+\text{ECG}}$, with the exception of BP calibration. Figure 6.6 demonstrates the marginal influence of demographics on the $\text{RF}_{\text{PPG}+\text{ECG}}$ model estimates through the use of SHAP values. Demographics were found to have a varying influence on the ΔSBP estimation, with a positive impact on some participants (IDs: 003, 009 for example) and a negative impact on others (IDs: 018, 025 for example). Demographics are known to influence the PPG contour; in particular, c/a has been previously explored as a marker for age-related arterial stiffening [78].

Despite possible interactions with demographics, a significant calibration issue still persisted as shown in figure 6.7. This issue was particularly evident in the four individuals with the largest ΔSBP at peak infusion (IDs: 002, 010, 023, and 026 shown in figure 6.6). These four individuals experienced $\Delta\text{SBP} > 33$ mmHg at peak infusion, whereas the median value across the cohort was 20 mmHg. The precise clinical effect of the weight-based dosing of phenylephrine in an individual would depend on the balance between their sensitivity to the increase in afterload, the effect of bradycardia on cardiac filling (therefore contractility, the Frank-Starling law [251]), and the proportion of venous/arterial action of phenylephrine causing the increase in preload and afterload. As a result, variations in ΔBP at peak infusion were expected across the cohort. The median weight of the four outlier individuals (73 kg) was marginally larger than the cohort average (69.5 kg) and so, on average, a larger dose of phenylephrine would have been given. Other *post-hoc* assessments of demographics were not sufficient to distinguish these four individuals. Although in such a small cohort it is difficult to draw conclusions for these individuals, we suggest two possible explanations for the large errors found. Firstly, these individuals clearly experienced a significant change to their cardiovascular system in response to the dosing of phenylephrine and so the changes in their cardiovascular system may not be adequately represented by the features available. Secondly, the *hybrid calibration* strategy may be impacted by the small sample size. As a result, there will be difficulty in calibrating individuals that may be classed as outliers.

A secondary explanation for the improved performance of the RF model relative to the $\text{LASSO}+\text{OLS}$ model may be that the features presented in this study, particularly the high-performing features presented in figure 6.8 (a), have a non-linear relationship with ΔSBP . This is further corroborated by the relationship between the top 9 ranking (non-demographic) features and ΔSBP shown in figure 6.9. Only σ_{g1} shows a discernible global relationship to ΔSBP . Individual,

participant-specific clusters are apparent, demonstrating low intra-participant variability but high inter-participant variability. This finding is consistent with previous studies by Radha *et al.* [187] and Hasanzadeh *et al.* [212], who reported that using an RF instead of a linear model improved the estimation of BP.

This non-linearity significantly impacts the ability to develop *individual calibration* models using these features. For PAT, typically only 2 model parameters (a slope and an intercept) are required for accurate *individual calibration*. Therefore, theoretically, a dataset containing only two measurements of PAT and BP for an individual is required for accurate *individual calibration*, although in practice this number is much higher for accurate estimation of the model parameters [329]. Whereas, for non-linear modelling of PPG_{fts} or ECG_{fts} , many more parameters are required to be estimated, thus forcing a much larger dataset requirement for accurate *individual calibration*.

LASSO+OLS_{PAT} outperformed the RF_{PAT} due to the strong negative linear correlation of PAT with BP. Comparing the LASSO+OLS_{PAT} results to the PAT random intercepts model (which assumed a constant calibration gradient) evaluated in the previous chapter highlights that in this dataset, the inclusion of demographics had a negative impact on model performance.

The results presented suggest that PPG_{fts} exhibits a stronger relationship with changes in BP than the complexity features extracted from the ECG. Although Hjorth mobility demonstrated notable feature importance, with a median ranking coefficient of 0.93, the addition of ECG_{fts} to a PPG_{fts} feature set did not result in any performance improvements. As shown in figure 6.6, the marginal contribution of ECG_{fts} to the overall model prediction was low, and the estimations were largely dominated by PPG_{fts} . For a small number of participants (IDs: 004, 007, 009, and 011), ECG_{fts} demonstrated a significant contribution to model estimates, but in general, the poor performance of the ECG feature set suggests that external, non-cardiac, control mechanisms may have a significant impact on the mechano-electric coupling BP-ECG relationship, affecting the latter's ability to estimate ΔBP . An additional explanation for the poor performance of the ECG feature set may be the choice of features used to explain the changes in the ECG morphology. However, to our knowledge, there is no other work suggesting alternatives to ECG complexity features for BP estimation. While RF_{ECG} showed improvements over a naive *baseline reference* assuming constant BP values, we suggest that ECG_{fts} on their own may not provide a viable solution for cuffless BP monitoring.

6.8.3 Feature importance

We explored a large and comprehensive set of features from both the PPG and the ECG gathered from a wide range of existing literature. The overall correlations of these features to ΔSBP across the cohort were in general low with only one feature ($\text{Gauss}_{\sigma 4/A1}$) having $|\rho_{p\Delta\text{SBP}}| > 0.5$. There were a number of features that had significant participant-wise correlations (PWC) to ΔSBP , 21 features with a median absolute PWC > 0.5 . The disparity between low correlations across the cohort and high correlations on a participant-wise basis underpins the important need for *individual calibration* due to low intra-participant variability and high inter-participant variability.

BP is determined by CO and TPR, and changes in either of these may be represented by different features [315]. Phenylephrine causes a direct increase in TPR [330] via an increase in both arterial and venous vasoconstriction. Therefore, a large impact from the reflected waves was not unexpected. This was corroborated by the observation that the majority of features which in this study had either a strong correlation or importance in estimating ΔSBP , characterise the impact of the reflected pressure waves ($\text{Gauss}_{\sigma 4/A1}$, c/a , IPA). On the other hand, phenylephrine causes a mixed response in CO with the relationship governed by preload dependency [330]. The majority of participants in the study experienced a decrease in CO (see figure F.5 in the appendix), driven largely by a decrease in HR. Therefore, it is not surprising that at least one of the most important features (σ_{g1}) represents changes in the upslope of the PPG which are driven by changes in CO [315]. However, it is not always clear how to link one feature to a specific BP control mechanism. Kurtosis (the feature with the highest importance for both RF and LASSO+OLS) for example, represents changes in the overall shape of the PPG.

We found some agreement between the features of the highest importance in our models to those reported in the literature. For example, features derived from the APG have been previously demonstrated to reflect age-related arterial stiffening [79, 308]. In agreement with this, in our work, e/a , c/a , and slope_{bd} were found to have median ranking coefficients of 0.89, 0.84, and 0.75 respectively. In general, however, we suggest that there is little consensus on the appropriate features for ΔBP estimation, both in our work and across the literature. Additionally features such as STT, proposed by Addison *et al.* [309] and supported by a recent publication from Natarajan *et al.* [190], had a very low median ranking coefficient of 0.16 in our dataset. Similarly, the best-performing features from Sun *et al.* [215] (Sys_{μ} , Sys_{σ} , Dia_{μ} , and Dia_{σ}) and Miao *et al.* [210] (b/a and RI) all demonstrated poor performance in this work. Furthermore, to our knowledge, no previous works have suggested that Kurtosis may be a feature of significant importance for BP estimation.

It should be noted that many of the features used in this study were derived from fiducial points such as the dicrotic notch. Relying on fiducial point detection has a number of limitations for BP

estimation. The detection algorithms often set arbitrary decisions or thresholds for fiducial point locations. For example, the typical definition for the systolic peak was not appropriate in this study due to the increasing influence of the tidal wave. Additionally, fiducial point detection algorithms will be valid up to a precision; small changes in BP observed in, for example, an ambulatory setting may result in very small perturbations in feature values that are indistinguishable from errors in fiducial point detection [212]. Finally, the fiducial points are not always detectable. The dicrotic notch has been reported to diminish in elderly individuals due to atherosclerosis (hardening of vessel walls and recruitment of collagen fibres to support walls) [77, 79]. We found that the majority of the features with high importance (for example: kurtosis, PCA features and Gaussian decomposition features) did not require fiducial point detection, and so these may offer a more desirable representation of changes of the PPG.

6.8.4 Limitations

Further to the dataset constraints presented in the previous chapter, there are a few limitations to the work presented in this chapter. Firstly, machine learning algorithms require large amounts of data to make robust estimations. In this work, we were limited by the number of observations per participant (typically once per minute for the 28 minutes of the recording per participant) and as a result, we were unable to explore *individual calibration* strategies, limiting our approach to predicting ΔBP using *hybrid calibration* in a LOSOCV framework. With only 26 participants, however, LOSOCV *hybrid calibration*, with personalisation performed by participant demographics, struggled to calibrate appropriately to the individual. This was particularly highlighted in the calibration issue for the four individuals with the largest errors (see figure 6.6).

Furthermore, the MOLLIE study design featured a single perturbation of BP for each participant, involving an increase from baseline followed by a return to the baseline level. As highlighted by Mukkamala *et al.* [9], single-perturbation studies may yield biasedly favourable results since models tend to overfit to the specific mechanisms activated by the intervention. Consequently, the extent to which the conclusions drawn in this chapter can be extrapolated to an ambulatory setting, which is the primary focus of this thesis, remains to be evaluated.

In the literature, the systolic peak is conventionally defined as the maximum of the PPG beat [189]. However, in our study, it was observed that under the influence of phenylephrine at peak infusion, the tidal wave may have an amplitude greater than that of the original systolic peak (see figure 6.1). In such instances, the standard definition of the systolic peak may incorrectly classify it as part of the tidal wave. Consequently, features like CT, ΔT , and STT, which rely on precise systolic peak detection, may be significantly affected. Although this PPG morphology has

been documented previously [215, 317, 78, 194], the precise definition of the systolic peak in such cases remains ambiguous. To address this issue, we identified the systolic peak as the first turning point of the PPG pulse above its midpoint. While this definition proved suitable for our brief, single-perturbation study with minimal motion artefacts, it may pose a significant limitation to BP estimation using PPG in real-world scenarios where motion artefacts are a substantial source of noise and the PPG may contain multiple turning points.

Finally, the accuracy of the BP cuff is a significant limitation to using single-point or *hybrid calibration* for BP estimation. Minor errors in a single cuff reading, caused by instrumentation or user factors (such as movement or incorrect cuff sizing), can result in a consistent offset in BP estimation. For instance, consider participant 006 figure 6.6, where the initial calibration during the rest period indicated $\Delta\text{SBP}_{\text{cuff}}$ and $\Delta\text{SBP}_{\text{est}}$ values of 0 mmHg. Over the subsequent five cuff inflations, the $\Delta\text{SBP}_{\text{cuff}}$ readings decreased to just under -5 mmHg, within the resolution of the ANSI/AAMI protocol. It remains unclear whether this SBP change reflects an actual physiological shift (potentially caused by the participant's relaxation during the study) or is an outcome of instrumental errors in the blood pressure cuff. In any case, a consistent DC offset of 5-10 mmHg was observed in the remaining BP estimates for this participant.

6.9 Conclusion

In this chapter, we explored the potential for using features derived from the morphology of the PPG and ECG to estimate ΔBP . Our results indicate that the currently proposed ECG_{fts} cannot accurately track ΔBP . Therefore, we do not pursue further investigation of ECG_{fts} as a marker for changes in BP in the rest of this thesis. However, we did observe clear changes in the PPG waveform in response to the dose increase of phenylephrine and hypothesised that these changes were caused by smooth muscle activation leading to an increase in the amplitude of the reflected tidal wave. Several PPG_{fts} , including kurtosis, σ_{g1} , and $\text{Gauss}_{\sigma_4/A1}$, were able to reflect these changes, and non-linear modelling of these features achieved similar performance to linear modelling of PAT. It is evident that estimating BP using PPG_{fts} is limited by the same calibration constraints as PAT. However, the apparent non-linear relationship between BP and PPG_{fts} is likely to exacerbate the calibration problem, requiring a larger database of reference measurements for adequate *individual calibration*. Ultimately, we conclude from this work that features from the PPG may be further explored as an indicator for changes in BP within a given individual.

This analysis has all been performed in a controlled setting with ΔBP induced by a pharmacological agent. In the next chapter, we examine 24-hour variations in these features and question whether they exhibit an endogenous circadian rhythm in line with that of BP.

Analysis of the circadian rhythm of pulse arrival time and photoplethysmogram features

Contents

7.1	Introduction	161
7.2	Methods overview	162
7.3	Pulse arrival time estimation in the MIMIC database	163
7.4	Computing ensemble averaged photoplethysmogram features	165
7.4.1	Computing ensemble-averaged photoplethysmogram	165
7.4.2	Fiducial point detection	166
7.4.3	Feature extraction	167
7.5	Cohort selection	169
7.5.1	Blood pressure phenotype labelling	170
7.6	Estimating circadian rhythm profiles	171
7.6.1	Double-cosinor model	172
7.6.2	Goodness-of-fit F-test	173
7.6.3	Population-level double-cosinor model	174
7.6.4	Distributions of double-cosinor parameters	175
7.7	Results	176
7.7.1	Dataset	176
7.7.2	Population-level double-cosinor model	176
7.7.3	Distributions of double-cosinor parameters	178
7.8	Discussion	183
7.8.1	Limitations	185
7.9	Conclusion	187

7.1 Introduction

Circadian patterns of BP have been extensively documented in the literature (see section 2.3.2). This endogenous rhythm, driven by neurohormonal variations, typically includes a decrease in BP during sleep, a phenomenon known as 'dipping' [5]. However, different BP circadian rhythm phenotypes exist depending on the magnitude and direction of the nocturnal change in BP. Recent research on this rhythm has strengthened the evidence that 'non-dipping' (no nocturnal change in BP), and 'reverse dipping' (increase in BP during sleep) patterns are predictors of both fatal and nonfatal cardiovascular events that are independent of daytime BP values [15, 5]. Currently, ABPM is the only technique available to monitor the circadian rhythm of BP in a home environment. However, ABPM is extremely disruptive to sleep and daily life and, as a result, is rarely performed

in primary care [5, 8]. The development of portable cuffless devices for monitoring 24-hour BP patterns at home would, therefore, be of significant importance.

The previous two chapters presented the fundamental methods to estimate BP using PAT and features from the PPG signal (PPG_{fts}). These methods were explored over a short-term clinical study that induced variations in arterial tone through pharmacological intervention. In this chapter, we extend this analysis to investigate whether PAT and PPG_{fts} have their own endogenous circadian rhythm in synchronicity with that of SBP. We additionally investigate whether the circadian rhythms of PAT and PPG_{fts} may be able to differentiate between BP phenotype patterns.

7.2 Methods overview

Large longitudinal datasets from which it might be possible to evaluate circadian rhythms of PAT and PPG_{fts} in a healthy population are not readily available. As a result, in this chapter, data was gathered from the MIMIC-III intensive care unit (ICU) database (see section 3.4). As is it not the goal of this thesis to propose cuffless BP circadian rhythm estimation within the ICU, we selected a cohort of ICU patients who were most likely to exhibit circadian variation in vital signs comparable to those observed in an out-of-hospital setting.

An overview of the signals extracted from each patient is provided in figure 7.1(a). During each patient's stay in the ICU, vital signs such as SBP, HR, and RR were routinely measured and recorded by the clinical team often at approximately hourly increments. Unlike the previous two chapters, we only focused on SBP in this work, and not MAP or DBP, as the literature regarding the circadian rhythm of BP and the corresponding phenotypes is largely dominated by analysis of SBP [5]. Additionally, ECG and PPG waveforms were recorded from each patient. PPG signals were recorded using a finger probe pulse oximeter. From these signals, we derived estimates of PAT and PPG_{fts} . To mitigate against motion artefacts, we extracted PPG_{fts} using an ensemble-averaged PPG (PPG_{EA}).

An overview of the steps implemented to investigate the circadian rhythms of the relevant signals is shown in figure 7.1(b). For each patient and each signal, we performed statistical modelling of circadian rhythms using the rhythmometric double-cosinor model. The model goodness-of-fit was assessed using an F-test and for each signal, any patient that did not pass the F-test, at the $p < 0.05$ level, was removed from subsequent analysis. We performed a population-level analysis by comparing representative circadian rhythm profiles across the cohort. Finally, we performed a statistical comparison of the distributions of model parameters across the cohort. We grouped these analyses by patient's BP phenotype in order to investigate phenotype estimation using PAT and PPG_{fts} .

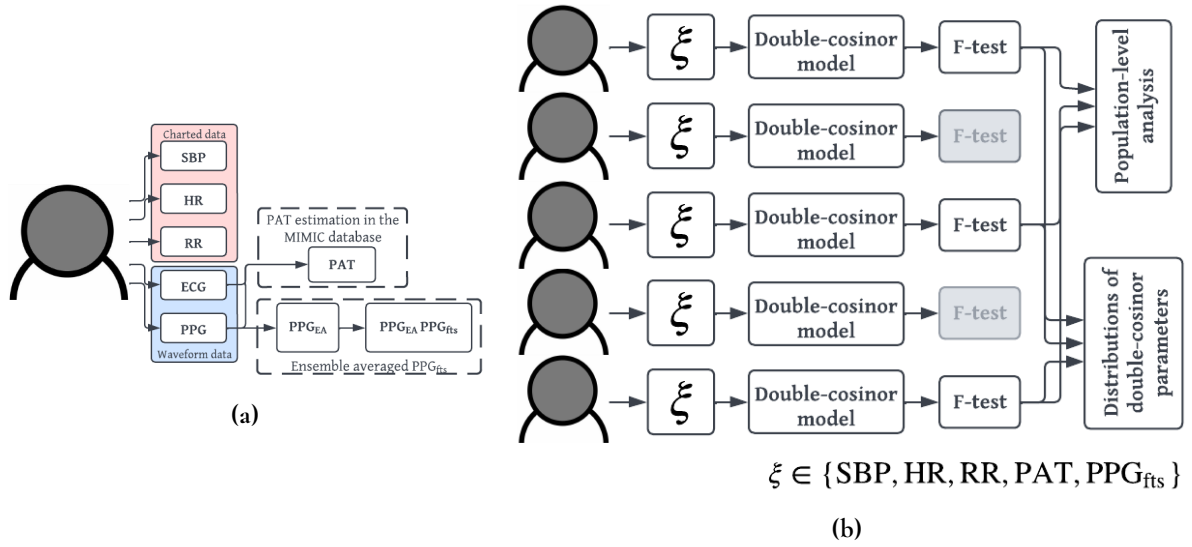


Figure 7.1: Schematic of the methodology implemented to estimate the circadian rhythm profiles of all signals of interest in this chapter. (a) Signal extraction from each patient in the cohort. Vital signs: SBP, HR, and RR were all extracted from the data charted by the clinical team in the hospital, typically once per hour. PAT was estimated from the ECG and PPG signals. Ensemble averaged PPG_{fts} were estimated from the PPG. (b) Estimating circadian rhythm profiles. For each patient and each signal of interest (denoted here by ξ), we modelled the circadian rhythm using the rhythmometric double-cosinor model. Model goodness-of-fit was assessed using an F-test; any patient that did not pass the F-test at the $p < 0.05$ level was removed from the subsequent analysis (demonstrated here by the greyed-out blocks). Finally, we analysed the circadian rhythms of all signals on a population level as well as through investigating the distributions of the double-cosinor parameters.

7.3 Pulse arrival time estimation in the MIMIC database

We followed the steps outlined in section 5.4 to preprocess the PPG and ECG waveforms, detect their respective cardiac beats and assess their signal quality (SQI_{PPG} and SQI_{ECG} respectively). The ECG mains frequency notch filter was redesigned with a centre frequency of 60Hz, the frequency of mains power in the US.

It has been previously observed that PAT estimates computed using data recorded in the MIMIC database contain an artefact resulting from internal clock differences in the equipment used for data collection [182, 331, 332, 333]. The ECG and PPG waveforms were recorded by separate Phillips modules each with its own internal clock and individual sampling rates. The signals were then later resampled for the MIMIC database to have the same sampling frequency (125 Hz) to allow for easy comparison by the research community. The sampling rate conversion resulted in a relative drift of the PPG waveform with respect to the ECG waveform. The resulting artefact can be characterised by a consistent DC time offset as well as an AC sawtooth pattern with a period of around 100s. The sawtooth artefact can be seen in figure 7.2(a).

This sensor artefact can make the comparison of beat-by-beat PAT estimates with BP difficult. For example, the time offset can push PAT values to be just over the R-R interval in the ECG signal,

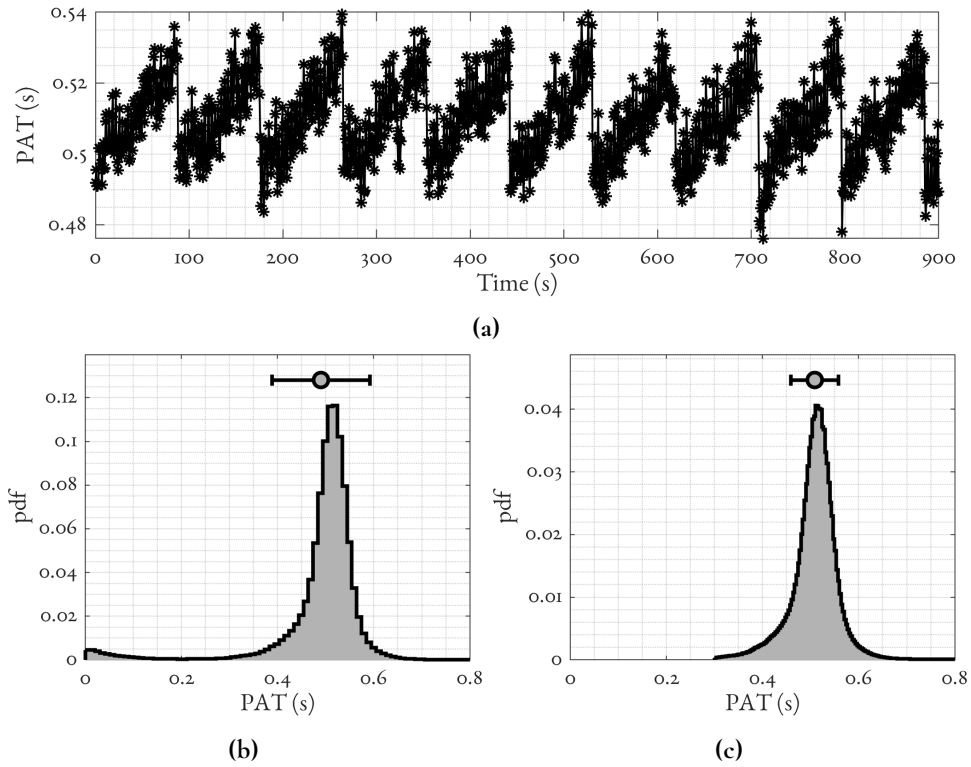


Figure 7.2: Sensor delays between the ECG and PPG waveforms recorded in the MIMIC dataset. (a) Sawtooth pattern seen in PAT due to clock differences between the ECG and PPG devices. (b) Distribution of PAT values computed using the typical definition identifying PPG beats between successive ECG R-peaks. (c) Distribution of PAT values used in this chapter by identifying PPG beats within a 0.30 s to 0.80 s window from each ECG R-peak. Note that the mean PAT value computed in the MIMC database is significantly larger than typically observed in healthy individuals (see figure 5.10 and [166]) as well in the ICU [182], highlighting the impact of the consistent DC offset.

which would be physiologically impossible given the typical speed of the pulse wave [138]. In these instances, the standard algorithm for computing PAT values (the time from the R-peak in the ECG signal to the first PPG beat) may fail to track the correct PPG pulse. Figure 7.2(b) shows the distribution of PAT values computed using the same algorithm as in the previous chapters. A small peak occurs just above zero seconds, driven by PPG beats drifting over an R-peak as a result of the sensor delays. In order to mitigate the effect of this, we computed PAT values as the time from each R-peak in the ECG signal to the first PPG beat within a window of 0.30 s to 0.80 s. The distribution of the resulting PAT values is shown in figure 7.2(c). Note that the mean of the distribution of PAT values computed in the MIMC database is significantly larger than typically observed in both healthy individuals (see figure 5.10 and [166]) as well in the ICU [182], highlighting the sensor artefact consistent DC offset on top of which the sawtooth AC waveform is superimposed.

The quality of the computed PAT signal, SQI_{PAT} , was estimated using the methods outlined in section 5.4.1 with an outlier window length and stride length of 5 minutes and 4 minutes respectively. Subsequently, Kalman filtering was employed to reduce the effects of transient artefacts.

7.4 Computing ensemble averaged photoplethysmogram features

In an uncontrolled setting such as the ICU, PPG signals recorded using a pulse oximeter with a finger probe can often be corrupted by sudden motion artefacts caused by, for example, hand movements. In order to mitigate this, we computed PPG_{fts} from an ensemble-averaged PPG beat (PPG_{EA}) representative of the average PPG signal within a 60-second sliding window, with a step size of 30 seconds.

7.4.1 Computing ensemble-averaged photoplethysmogram

To compute the overall PPG_{EA} for each 60-second window, we adopted a four-stage strategy. At the end of each stage, a PPG_{EA} was constructed using PPG beats located within the window that met the stage criteria. The PPG_{EA} was constructed by first normalising all valid PPG beats to have unit amplitude. Subsequently, all beats were then resampled to be 100 samples long using cubic-splines interpolation. Finally, all normalised and resampled PPG beats were averaged to construct the PPG_{EA} for the relevant stage and window.

The four stages involved in computing the PPG_{EA} are outlined in figure 7.3. For the first stage, all beats of poor signal quality (defined as an SQI_{PPG} equal to 0) were removed from analysis. In the second stage, the PPG_{EA} was computed using only the valid beats (from the previous stage) that did not exceed the mean $\pm$ three standard deviations for that window. In the third stage, the ρ_p and relative lag (τ_{lag}) of each PPG beat with the PPG_{EA} were computed. τ_{lag} was computed as the time delay that maximised the cross-correlation, R_{xy} , between the two waveforms (see section 5.5.1.2). Any beat with $\rho_p < 0.5$ or $|\tau_{\text{lag}}| > 50\text{ms}$ was discarded. Finally, in the fourth stage, the τ_{lag} of each beat with the PPG_{EA} was recomputed as before and any beat with $R_{xy}(\tau_{\text{lag}}) < 0.8$ with the PPG_{EA} was discarded.

The resulting SQI of the PPG_{EA} for each 60-second window, w , ($\text{SQI}_{\text{PPG}}^{\text{EA}}(w)$) was computed as:

$$\text{SQI}_{\text{PPG}}^{\text{EA}}(w) = \frac{1}{N_w} \sum_{j \in \Omega} \text{SQI}_{\text{PPG}}(j) \quad (7.1)$$

where N_w is the total number of PPG beats initially in w (i.e. prior to stage 1), Ω is the subset of PPG beats that were ultimately not discarded from analysis post stage 4. Note that $\text{SQI}_{\text{PPG}}^{\text{EA}}(w)$ is a particularly strict signal quality index (i.e. values are much more likely to be closer to zero than beat-by-beat SQI_{PPG} values) as it is impacted not only by the number of beats discarded from analysis but also by the SQI_{PPG} of the remaining valid beats.

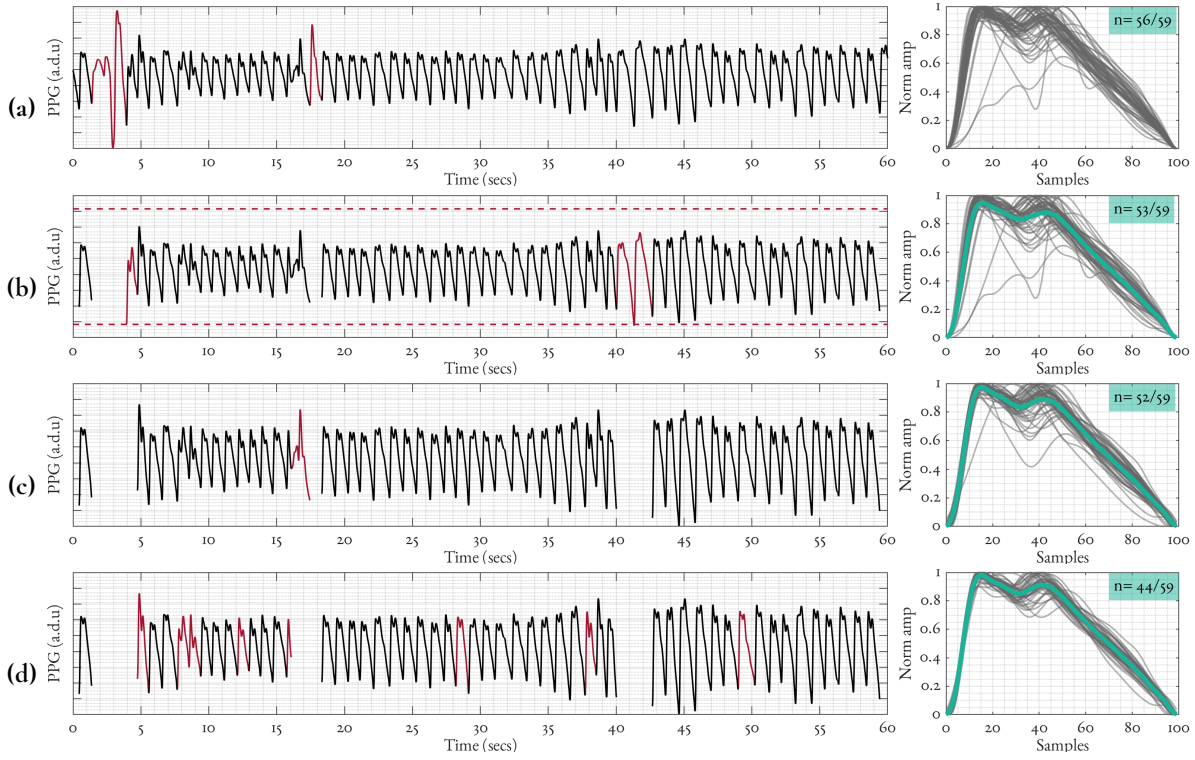


Figure 7.3: Overview of the four stages used to compute PPG_{EA} and SQI_{EA} . At each stage, beats that did not meet the stage criteria are shown in red and the resulting PPG_{EA} for that stage is shown in the right-hand column. n indicates the fraction of PPG beats in the window that met the stage criteria up to that stage. (a) Stage 1: PPG_{EA} is constructed from PPG beats with an SQI_{PPG} greater than 0. (b) Stage 2: PPG_{EA} is constructed from beats that do not exceed the mean $\pm$ three standard deviations for that window. (c) Stage 3: PPG_{EA} is constructed from beats whose correlation with PPG_{EA} was greater than 0.5 and lag was less than 50 ms. (d) Stage 4: PPG_{EA} is constructed from beats whose circularly shifted correlation was greater than 0.8.

7.4.2 Fiducial point detection

With the exception of S (peak of the PPG pulse), N (corresponding to the dicrotic notch), and D (corresponding the diastolic peak), we used the same criteria as were outlined in section 6.3.1 to detect the fiducial points of the PPG_{EA} . For each PPG_{EA} we define a normalised time vector t' such that t' starts at 0 and is equal to 1 at the end of the beat.

S was defined as the maximum of the PPG_{EA} , at time $S_{t'}$. The dicrotic notch has been reported to diminish in elderly individuals due to the development of atherosclerosis [77, 79], N and D can, therefore, often be missing in PPG signals recorded from patients in the ICU. We followed the method proposed by Balmer *et al.* [334] for end-of-systole detection in dicrotic notch-less pressure waveforms. The algorithm places a prior weighting function, $w_N(t')$, on the expected dicrotic notch location in the APG waveform (referred to as APG_{EA} here as it is computed as the second

derivative of PPG_{EA}) using a beta distribution. $w_N(t')$ was defined as:

$$w_N(t') = \begin{cases} 0 & : t' \leq S_{t'} \\ \left(\frac{t' - S_{t'}}{1 - S_{t'}}\right)^{\alpha-1} \left(\frac{t' - 1}{S_{t'} - 1}\right)^4 & : t' > S_{t'} \end{cases} \quad (7.2)$$

for $t' > S_{t'}$, $w_N(t')$ is a beta-distribution whose position is characterised by α , computed as:

$$\alpha = \frac{3t'_{Nexp} - 4S_{t'} + 1}{1 - t'_{Nexp}} \quad (7.3)$$

where t'_{Nexp} is the expected location of the dicrotic notch for each PPG_{EA} . α was bounded between 1.5 and 4.5 in order to ensure that $w_N(t')$ was not too prominent in late systole or early diastole. For the n^{th} PPG_{EA} , t'_{Nexp} was defined as:

$$t'_{Nexp} = \begin{cases} 0.45/T_{EA} - 0.1/T_{EA}^2 & \text{if } n \leq 10 \\ \frac{1}{10} \sum_{i=1}^{10} T'_{sys}(n-i) & \text{otherwise} \end{cases} \quad (7.4)$$

where $T'_{sys}(j)$ is the detected systolic duration from the j^{th} PPG_{EA} and T_{EA} is equal to the median time duration of the beats in Ω (i.e. effectively a scaling factor of the normalised time vector t'). The " $0.45/T_{EA} - 0.1/T_{EA}^2$ " estimate stems from the observed empirical relationship between the duration of systole and HR in a study involving 103 adults previously outlined in section 6.3.1 and projected on our normalised time vector t' [306].

With $w_N(t')$ computed, the dicrotic notch, N , was located as the most prominent peak of $\text{APG}_{\text{EA}} \cdot w_N(t')$. Following the work of Millasseau *et al.* [192], the location of the diastolic peak D was detected as the earlier of the two following points: the first turning point of APG_{EA} following N , or as the zero-crossing point of the VPG_{EA} following N . An overview of these steps is shown in figure 7.4 for a typical PPG_{EA} waveform in the database.

7.4.3 Feature extraction

We extracted the same 33 PPG_{fts} analysed in the previous chapter from each PPG_{EA} beat. These features are outlined in table 7.1. Note that, similar to the previous chapter, as the time duration of PPG_{EA} was normalised, all PPG_{fts} were independent of HR. To allow for consistency of feature definitions, features derived from principal component analysis (PCA) were computed using the same principal orthogonal axes as in section 6.3.2.5. We processed the PPG_{fts} using outlier detection to remove notable deviations in values and a Kalman filter to reduce the effect of transient artefacts caused by noise. As the sampling frequency of the PPG_{fts} was 1/30 Hz (a

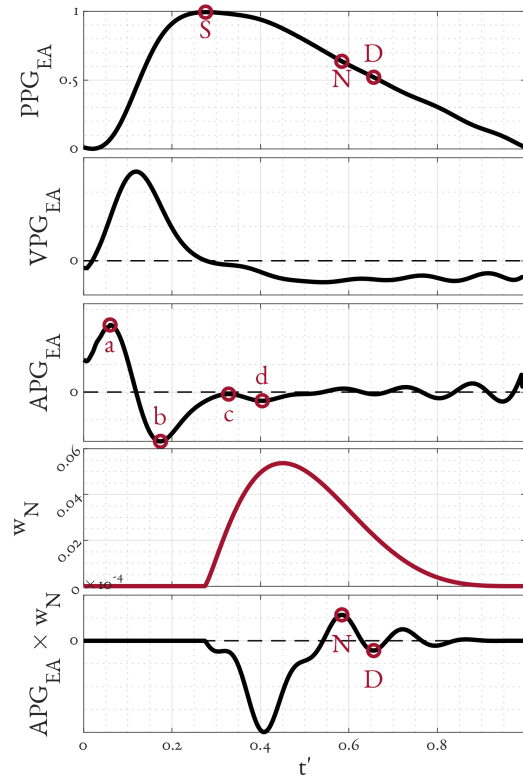


Figure 7.4: Fiducial point detection of PPG_{EA} in the MIMIC database. S was detected as the maximum of the PPG_{EA} . The a , b , c , and d waves of the APG_{EA} were detected in the same manner as previously outlined in section 6.3.1. N and D were detected through analysis of the APG_{EA} multiplied by a prior weighting function w_N computed as in equation (7.2). t' is the normalised time vector.

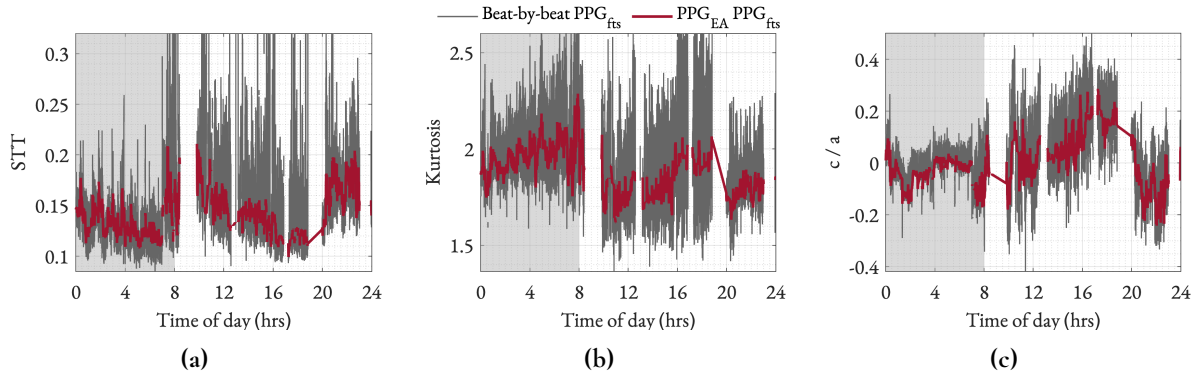


Figure 7.5: Visual comparison of three PPG_{fts} (STT, Kurtosis, c/a) extracted on a beat-by-beat basis to those extracted from PPG_{EA} for a typical patient over a 24-hour monitoring period. PPG_{EA} PPG_{fts} present a smooth approximation of the overall trends. The shaded area indicates night-time (defined as midnight to 8am).

result of the step size of PPG_{EA} computation) the outlier detection was performed over a larger 30-minute window with a 25-minute step.

Figure 7.5, visually compares PPG_{fts} extracted on a beat-by-beat basis to those extracted from PPG_{EA} for a typical patient over a 24-hour monitoring period. We note that ensemble-averaged PPG_{fts} represent a smooth approximation of the overall trends of each individual feature.

Table 7.1: Summary of the 33 PPG_{fts} extracted from each ensemble averaged PPG beat. All nomenclature as introduced in table 6.2.

Feature notation	Description or formula
Slope transit time (STT) [309]	Slope of straight line from onset to systolic peak
Inflection point area (IPA)[204]	A_{Dias}/A_{Sys}
Width ₅₀ [78]	Width of the PPG at 50 % of its amplitude
Normalised Harmonic Area (NHA) [204]	$\sum_{n=2}^{N_h} FFT^2(n \cdot f_1) / \sum_{n=1}^{N_h} FFT^2(n \cdot f_1)$
Inflection and Harmonic area ratio (IHAR) [204]	(1-NHA)/ IPA
Skewness [213]	Asymmetry of the data around the mean
Kurtosis [213]	Relative "tailedness" of the distribution
Sys $_{\sigma}$ [215]	Variance of VPG in the systolic phase
Dia $_{\sigma}$ [215]	Variance of VPG in the diastolic phase
$\frac{b}{a}, \frac{c}{a}, \frac{d}{a}$ & $\frac{e}{a}$ [79]	
slope _{b-d} [308]	Slope of a straight line between b and d , normalised by a
$\sigma_{g1}, A_{g1}, \sigma_{g3}, A_{g3}$	Variance and amplitude of the first and third decomposed Gaussians
Gaussian augmentation index (Gauss AI) [187]	$\max(g_s) - A_{g3}$
Gaussian reflection index (Gauss RI) [187]	$\sum(g_s) - \sum(g_3)$
Gaussian augmentation index _R (Gauss AI _R) [193]	$(A_{g1} - A_{g2})/A_{g1}$
Gaussian reflection index _R (Gauss RI _R) [193]	A_{g3}/A_{g1}
Gaussian approximation of left ventricular ejection time (Gauss LVET) [206]	See section 6.3.2.4 for definition.
Gauss $_{\sigma 4/A1}$	σ_{g4}/A_{g1}
PPG PCA ₁₋₃ [313]	First 3 principal components of PPG beat
VPG PCA ₁₋₃ [313]	First 3 principal components of VPG beat
APG PCA ₁₋₃ [313]	First 3 principal components of APG beat

7.5 Cohort selection

In section 3.4.2 we introduced our MIMIC patient selection criteria which restricted our analysis to a subset of 1,508 patients most likely to exhibit physiological vital-sign profiles comparable to non-hospitalised individuals. For each patient, we averaged all signals (SBP, HR, RR, PAT, and all PPG_{fts}) within each hour of the 24-hour period we are analysing. This process avoids weighting continuous data (PAT and PPG_{fts}) above the irregularly sampled (typically once per hour) charted vital-sign measurements. These mean hourly values were recorded using centre alignment (e.g. the mean of measurements between 6:00 pm and 7:00 pm was recorded as occurring at 6:30 pm). We show this process in figure 7.6.

In an ICU, measurement equipment and sensors may be frequently removed as patients get up out of bed, eat, or when examinations were performed. As a result, all signals were not present for the full 24 hours for each patient. In order for there to be enough data to appropriately model circadian variations for each patient, we required there to be at least 5 distinct hourly bins with valid measurements for all signals with at least one measurement during the day (10am to 8pm) and at least one during the night (midnight to 8am). For both PAT and PPG_{EA} , we calculated an hourly

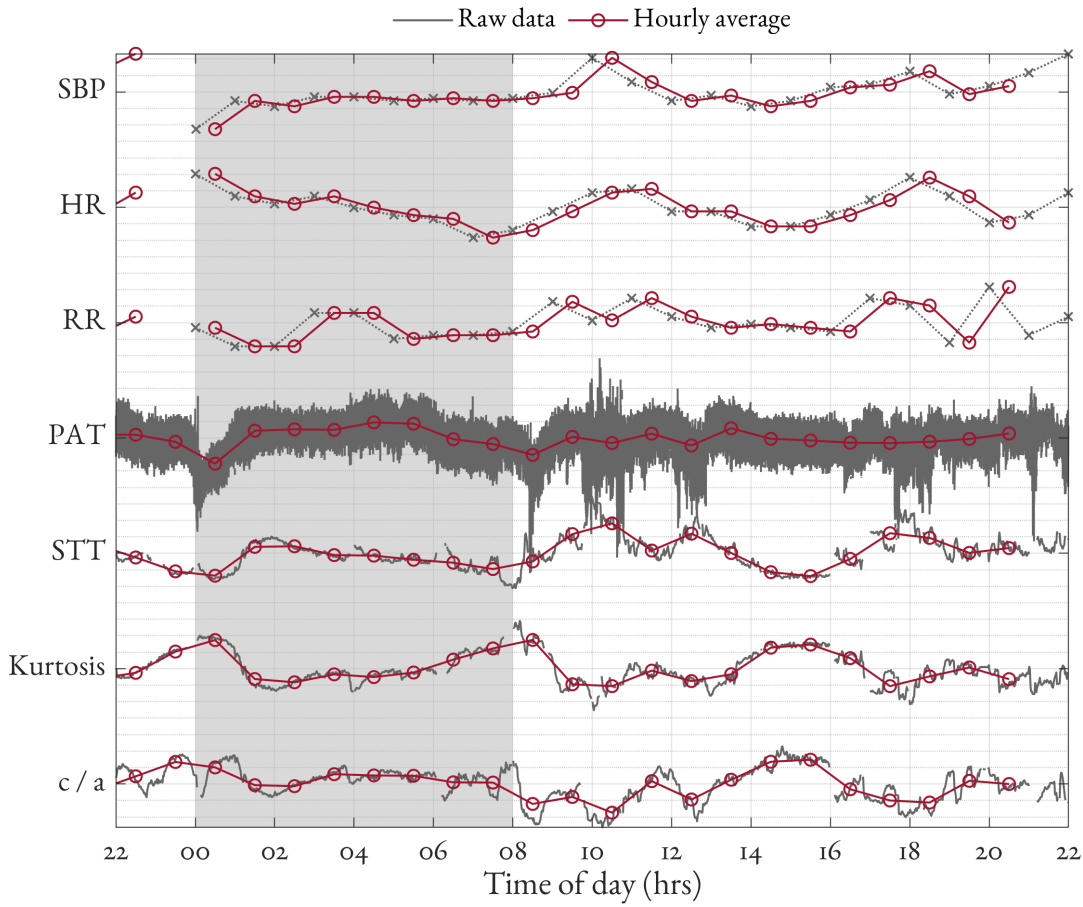


Figure 7.6: Typical patient profile showing SBP, HR, RR, PAT, and PPG_{fts} recorded during the final 24-hour period of a patient's stay in the ICU. Values of SBP, HR, and RR were recorded typically once an hour. High-frequency variation in PAT demonstrates the sawtooth artefact (see figure 7.2(a)). Only three PPG_{fts} (STT, Kurtosis, and c/a) are shown here as a representation. The shaded area indicated night-time (defined as midnight to 8am).

SQI as the mean value of SQI_{PAT} and SQI_{PPG}^{EA} respectively. If either of these SQIs fell below a given threshold γ_{SQI} , then a valid measurement of PAT or PPG_{fts} could not be made for that window. To determine the optimum value of γ_{SQI} , we performed a sensitivity analysis outlining the threshold's impact on the number of patients for whom the criteria of at least 5 valid measurements, with at least one daytime and one night-time measurement, was met for PAT or PPG_{fts} . This analysis is shown in figure 7.7. As a result of the strict SQI developed for PPG_{EA} (see equation (7.1)), PPG_{fts} exhibited a significantly higher sensitivity to γ_{SQI} than PAT. From this analysis, γ_{SQI} was set as 0.6. This resulted in 1,051 patients from which we computed an estimate of their circadian rhythm.

7.5.1 Blood pressure phenotype labelling

Section 2.3.2 identified three BP phenotypes according to the percentage difference between the mean night-time (midnight to 8am) SBP and the mean daytime (10am to 8pm) SBP, normalised

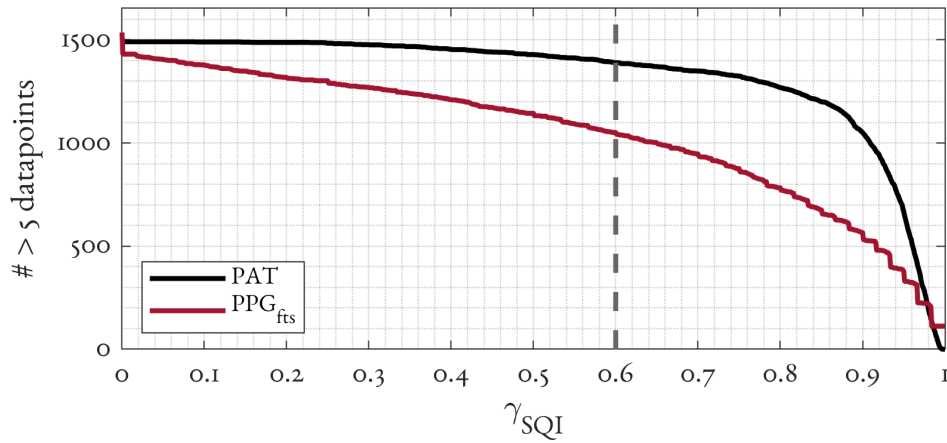


Figure 7.7: Sensitivity analysis of the impact of γ_{SQI} on the number of patients for whom 5 valid measurements of PAT or PPG_{fts} were available with at least one measurement during the day and one at night. From this analysis, γ_{SQI} was set as 0.6.

by the latter (DR_{SBP}):

$$DR_{SBP} = \frac{\overline{SBP}_{Night} - \overline{SBP}_{Day}}{\overline{SBP}_{Day}} \cdot 100\% \quad (7.5)$$

Patients were placed in the $\mathcal{P}_D$, $\mathcal{P}_{ND}$, or $\mathcal{P}_{RD}$ categories using the DR_{SBP} criteria outlined in table 7.2.

Table 7.2: BP phenotype labelling according to an individual's SBP day-to-night dip ratio (DR_{SBP}). Daytime was defined as 10am to 8pm, night-time was defined as midnight to 8am.

Phenotype	SBP Dip ratio (DR_{SBP}) range
Dipper ($\mathcal{P}_D$)	$-10\% \geq DR_{SBP}$
Non-dipper ($\mathcal{P}_{ND}$)	$0\% > DR_{SBP} \geq -10\%$
Reverse dipper ($\mathcal{P}_{RD}$)	$DR_{SBP} \geq 0\%$

7.6 Estimating circadian rhythm profiles

There is no gold standard method for analysing 24-hour variations in vital signs such as SBP. The approaches typically employed in the literature include the rhythmometric cosinor method [335, 336, 337], cubic smoothing splines [338, 339], and a double logistic model [340]. Of these methods, cosinor analysis has traditionally been the most common approach for statistical modelling of circadian variations. For example, ABPM data is traditionally analysed through a double-cosinor model (the summation of two cosinors with 24-hour and 12-hour periods respectively) [335, 341]. This approach has been shown to be able to model the early morning BP peak [337] as well as demonstrate significant differences in the $\mathcal{P}_D$ and $\mathcal{P}_{ND}$ BP phenotypes [342].

Figure 7.1(b) provides an outline of the steps implemented in our analysis of the circadian rhythm of the following signals: SBP, HR, RR, PAT, and all PPG_{fts}.

7.6.1 Double-cosinor model

The rhythmometric cosinor method was initially developed in the 1960s by Halberg in order to analyse short sparse data series, such as the breakdown of births and deaths according to the hours of the day [343, 344]. At the time, methods typically used to analyse periodic data (such as periodograms) required data to be equidistant and cover multiple cycles. The cosinor method is a curve fitting method that attempts to fit a sinusoid to data through least squares procedures, thus removing the limitations of data length and uniform sampling [206]. However, limitations in this method still exist, primarily through the sensitivity to outliers.

Fitting a single-cosine model to observations of signal Y_i taken at time t_i for a patient can be written as:

$$Y_i = M + A \cos(2\pi t_i/\tau + \phi) + \epsilon_i, \quad i = \{1, \dots, N_{\text{meas}}\} \quad (7.6)$$

where M represents the midline estimating statistic of rhythm (MESOR) or the level around which the oscillations cycle, A is the amplitude (half the extent of predictable variation within a cycle), τ is the period (duration of one cycle) and ϕ the acrophase (the time of overall high values recurring in each cycle). N_{meas} is the number of signal observations for a patient. The parameter ϵ_i represents the error at time t_i , with the error terms assumed to be normally distributed with zero mean. For the case of assessing circadian rhythms, τ was fixed to 24 hours. This allowed equation (7.6) to be written as:

$$Y_i = M + \beta x_i + \gamma z_i + \epsilon_i, \quad i = \{1, \dots, N_{\text{meas}}\} \quad (7.7)$$

where:

$$\beta = A \cos \phi; \quad \gamma = -A \sin \phi; \quad x_i = \cos(2\pi t_i/\tau); \quad z_i = \sin(2\pi t_i/\tau) \quad (7.8)$$

Equation (7.7) is therefore linear in terms of M , β , and γ and can be solved using least squares. The principle underlying the least squares method is the minimisation of the residual sum of squares (RSS), the sum of the squared differences between reference measurements Y_i and modelled values:

$$RSS = \sum_i^{N_{\text{meas}}} (Y_i - (M + \beta x_i + \gamma z_i))^2 \quad (7.9)$$

RSS is minimised by differentiating equation (7.9) with respect to M , β , and γ and setting the resultant to 0. This results in the following set of equations:

$$\begin{bmatrix} \sum_i Y_i \\ \sum_i Y_i x_i \\ \sum_i Y_i z_i \end{bmatrix} = \begin{bmatrix} N_{\text{meas}} & \sum_i x_i & \sum_i z_i \\ \sum_i x_i & \sum_i x_i^2 & \sum_i x_i z_i \\ \sum_i x_i & \sum_i x_i z_i & \sum_i z_i^2 \end{bmatrix} \begin{bmatrix} M \\ \beta \\ \gamma \end{bmatrix} \quad \text{or} \quad d = Su \quad (7.10)$$

where the values of M , β , and γ can be computed by solving:

$$u = S^{-1}d \quad (7.11)$$

A and ϕ can then be computed using trigonometric relations from equation (7.8):

$$A = \sqrt{\beta^2 + \gamma^2} \quad \phi = \tan^{-1}\left(-\frac{\gamma}{\beta}\right) \quad (7.12)$$

The single-cosinor model is easily extended to include the summation of multiple cosinor components at harmonic frequencies of the fundamental 24-hour component (i.e. components with 12-hour, 8-hour, 6-hour etc periods) [335]. Multiple components are often more appropriate for modelling data that contain multiple peaks and nadirs and so deviate from perfect sinusoidality. For example, the 24-hour profile of SBP typically includes two peaks, one soon after awakening and one in late afternoon (see section 2.3.2), and as a result, is commonly analysed using a double-cosinor model consisting of cosines with both 24-hour (also known as a ultradian pattern) and 12-hour periods [337, 342, 335, 341]. The double-cosinor model is defined as:

$$Y_i = M + A_1 \cos(2\pi t_i/\tau_1 + \phi_1) + A_2 \cos(2\pi t_i/\tau_2 + \phi_2) + \epsilon_i, \quad i = \{1, \dots, N_{\text{meas}}\} \quad (7.13)$$

where τ_1 is 24 hours, and τ_2 is 12 hours. The system increases from 3 equations with 3 unknowns to 5 equations with 5 unknowns (M and 2 pairs of $(A_j$ and $\phi_j)$). It is for this reason that we required all patients to have a minimum of 5 valid measurements of each signal. The parameters can be computed using the same method as the single component cosinor, through solving equation (7.11).

The multiple-component cosinor model also introduces two new parameters: orthophase and bathyphase. Orthophase (θ_{Orth}) is the time to the nadir of the model. Bathyphase (θ_{Bathy}) is the time to the peak of the model. For signals such as SBP, these parameters have been shown to be especially important to distinguish 'dippers' and 'reverse-dippers' [342].

7.6.2 Goodness-of-fit F-test

A goodness-of-fit test assesses model adequacy based on how well the model fits the observed data. In this work, we assessed the goodness-of-fit of the double-cosinor model for each signal and each patient based on an F-test. An F-test provides a statistical assessment of the presence of a rhythm (i.e. the amplitudes of the two cosinor terms are non-zero) in the observed data [335]. The test statistic, F_{rhythm} , is computed as the ratio of the RSS (defined in equation (7.9) as the variance of the observed data that is unexplained by the model), and the model sum of squares (MSS, the variance of the observed data that is explained by the model). MSS is computed as:

$$\text{MSS} = \sum_{i=1}^{N_{\text{meas}}} (M + A_1 \cos(2\pi t_i/\tau_1 + \phi_1) + A_2 \cos(2\pi t_i/\tau_2 + \phi_2) - \frac{1}{N_{\text{meas}}} \sum_{j=1}^{N_{\text{meas}}} (Y_j)) \quad (7.14)$$

If there is a rhythm present in the data, then the majority of the variance in the data would be captured by the double-cosinor model and so MSS would be notably larger than RSS. Thus, F_{rhythm} is computed as:

$$F_{rhythm} = \frac{\left(\frac{MSS}{4}\right)}{\left(\frac{RSS}{N_{meas} - 5}\right)} \quad (7.15)$$

where 4 and $N_{meas} - 5$ are the degrees of freedom attributed to the model (5 model parameters - 1) and to the error term ($N_{meas} - 5$ model parameters). F_{rhythm} follows the F -distribution and thus, the null hypothesis (H_0) that there is no rhythm present is rejected at a 5% confidence level when $F_{rhythm} > F_{95\%}(4, N_{meas} - 5)$. For each signal, the proportion of patients that pass the F-test is referred to as % *rhythm*. Signals that have their own endogenous circadian rhythm are likely to exhibit high % *rhythm* values.

7.6.3 Population-level double-cosinor model

In order to investigate the circadian rhythm profile present across the whole cohort, we computed a representative population-level double-cosinor model for each signal.

As outlined in the previous chapters, when estimating BP using PAT or PPG_{fts}, individual-specific calibration coefficients are often handled by subtracting a calibration data point and as such changes in SBP (ΔSBP) are often predicted using changes in PAT (ΔPAT) or PPG_{fts} (ΔPPG_{fts}). To account for this, for each patient and each signal we subtracted the corresponding MESOR value around which the double-cosinor models oscillate. As a result, in this chapter, we analysed circadian rhythm deviations for all signals.

To compute the population-level model for each signal, each patient in the cohort was first modelled by the double-cosinor model to yield estimates of $u_j = [\beta_{1j}, \gamma_{1j}, \beta_{2j}, \gamma_{2j}]$, $j = \{1, \dots, N_{Circ}\}$ where N_{Circ} is the number of patients in the cohort. Patients that did not pass the F-test for the presence of a rhythm were removed from the population-level analysis. The resulting population size for each signal is labelled N'_{Circ} .

Individual vectors u_j are assumed to be a random sample from a 4-dimensional, multi-variate Gaussian distribution with mean: $u^* = [\beta_1^*, \gamma_1^*, \beta_2^*, \gamma_2^*]$. The population estimates are unweighted averages of the individual parameters [335]:

$$u^* = \frac{1}{N'_{Circ}} \sum_{i=1}^{N'_{Circ}} [\beta_{1i}, \gamma_{1i}, \beta_{2i}, \gamma_{2i}] = [\beta_1^*, \gamma_1^*, \beta_2^*, \gamma_2^*] \quad (7.16)$$

and the population A_1^* , A_2^* , ϕ_1^* and ϕ_2^* parameters can be estimated using the relations:

$$A_1^* = \sqrt{\beta_1^{*2} + \gamma_1^{*2}}, \quad \phi_1^* = \tan^{-1}\left(-\frac{\gamma_1^*}{\beta_1^*}\right), \quad A_2^* = \sqrt{\beta_2^{*2} + \gamma_2^{*2}}, \quad \phi_2^* = \tan^{-1}\left(-\frac{\gamma_2^*}{\beta_2^*}\right) \quad (7.17)$$

For each signal, the population-level double cosinor model was subsequently computed using equation (7.13) (with M set to 0). Agreement between each signal's population-level profile was characterised by visual inspection of the profiles as well as linear correlation, ρ_p . To examine their consistency across BP phenotypes, these results were grouped into $\mathcal{P}_D$, $\mathcal{P}_{ND}$, and $\mathcal{P}_{RD}$ cohorts.

7.6.4 Distributions of double-cosinor parameters

Analysing the distributions of the double-cosinor parameters provides an insight into the response at a patient level (which is then explored in more detail in chapter 8). Here we outline our steps to produce confidence intervals for θ_{Orth} and θ_{Bathy} as well as a comparison of model parameters across BP phenotypes.

7.6.4.1 Confidence intervals for θ_{Orth} and θ_{Bathy}

The orthophase (θ_{Orth}) and bathyphase (θ_{Bathy}) of double-cosinor models indicate times of signal nadirs and peaks respectively. This is especially important for non-invasive tracking of circadian rhythms of BP in order to indicate timings of night-time dips and whether these timings are consistent across individuals. As θ_{Orth} and θ_{Bathy} are not thought to take on a normal distribution [335], 95% confidence intervals for the population mean θ_{Orth} ($CI_{95}^{\theta_{\text{Orth}}}$) and θ_{Bathy} ($CI_{95}^{\theta_{\text{Bathy}}}$) were approximated by a bootstrap estimate using 1,000 repeats [345, 346]. Wide $CI_{95}^{\theta_{\text{Orth}}}$ or $CI_{95}^{\theta_{\text{Bathy}}}$ for a given signal, indicate notable inconsistencies of nadir or peak timings respectively across all patients.

Bootstrapping is a sampling technique used to estimate the distribution of a parameter of interest [346]. By randomly resampling (with replacement) θ_{Orth} and θ_{Bathy} from the cohort, a bootstrap estimate of the population mean can be calculated for these parameters. Furthermore, the distribution of each parameter around its population mean can be used to calculate CI_{95} . For each signal, the procedure starts by selecting a random sample (also known as a replicate), with replacement, of θ_{Orth} and θ_{Bathy} from all patients that passed the F-test for the presence of a rhythm. From this replicant the sample mean $\bar{\theta}_{\text{Orth}}$ and $\bar{\theta}_{\text{Bathy}}$ were computed. As phases represent a circular parameter ($\theta_{\text{Orth}}, \theta_{\text{Bathy}} \in [0, 2\pi]$), the standard linear mean would not account for the circular properties. As a result, the circular mean equivalent was implemented (see appendix C.2). Repeating this process 1,000 times (as recommended [346]) a large number of estimations of $\bar{\theta}_{\text{Orth}}$ and $\bar{\theta}_{\text{Bathy}}$ are derived. For each parameter, taking the 2.5 and 97.5 percentiles of this set of values returns the CI_{95} .

7.6.4.2 Comparison of double-cosinor parameters across phenotypes

To test the null hypothesis that the circadian features derived using the double-cosinor models ($[A_1, \phi_1, A_2, \phi_2, \theta_{\text{Orth}}, \theta_{\text{Bathy}}]$) did not differ between patients grouped in the $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ cohorts (see table 7.2), a non-parametric comparison of these features was performed. For A_1 and A_2 this assessment was performed using a Mann-Whitney U Test. Details of this statistical test are provided in appendix C.1. For the phase parameters ($\phi_1, \phi_2, \theta_{\text{Orth}}$, and θ_{Bathy}) this was performed using a non-parametric test for equal circular medians [347]. Details of this statistical test are provided in appendix C.2. We adjusted the p-values for multiple comparisons using the Benjamini-Hochberg method (see appendix C.1.2).

7.7 Results

7.7.1 Dataset

Table 7.3 shows the overall demographics of the cohort of patients evaluated. A total of 1,051 patients made up the cohort. 54.9% of patients were male, and on average stayed for 2.4 days in the ICU. The mean Oxford Acute Severity of Illness Score (OASIS), a severity of illness score used for predicting patient outcomes, was 26.5, suggesting that, on average, a patient in our cohort had a 2.60% chance of dying in the ICU [348]. On average, around 20 measurements of SBP, HR, RR, and PAT were recorded for each patient during the 24-hour period of analysis. In comparison, there were, on average, 14.1 hourly estimates of PPG_{fts} .

7.7.2 Population-level double-cosinor model

Table 7.4 provides an overview of the evaluation metrics computed by estimating the population-level circadian rhythm deviations for all signals, grouped into vital signs (ΔSBP , ΔHR , ΔRR), ΔPAT , and $\Delta\text{PPG}_{\text{fts}}$. Intra-group signals are ordered by their absolute correlation with the population-level circadian rhythm profile of ΔSBP across all patients.

Of the patients studied, 68.9%, 82.2%, and 55.5% exhibited a rhythm in ΔSBP , ΔHR , and ΔRR , respectively. Both ΔHR and ΔRR were highly correlated with ΔSBP across the whole cohort ($\rho_p = 0.91$ for both). However, this relationship was not consistent across phenotypes, as a weak negative correlation ($\rho_p = -0.42$ for ΔHR and -0.79 for ΔRR) was reported for 'reverse dippers'.

Most of the patients (88.2%) exhibited a rhythm in ΔPAT , which had a strong negative correlation ($\rho_p = -0.99$) with ΔSBP across the whole cohort. This result was consistent across BP phenotypes ($\mathcal{P}_D$: $\rho_p = -0.99$; $\mathcal{P}_{ND}$: $\rho_p = -0.98$; $\mathcal{P}_{RD}$: $\rho_p = -0.83$).

Table 7.3: Demographics of the subset of patients selected from the MIMIC database. Data was recorded during the last 24 hours of a patient's stay in the ICU. Parameters such as height and weight were not recorded for all patients and so are not presented here. The admission location provides information about the previous location of the patient prior to arriving in the intensive care unit (ICU). The reference BP phenotype was computed using DR_{SBP} .

Descriptor	Value
Number of patients	1051
Male	577
Female	474
Age (years) ¹	56.0 (15.8)
Length of stay in the ICU (days) ¹	2.4 (2.4)
OASIS ^{1,2}	26.5 (7.2)
Ethnicity ³	
White	706 (67%)
Black	127 (12%)
Unknown	117 (11%)
Hispanic	61 (6%)
Asian	40 (4%)
Admission location ³	
Referral	740 (70%)
Emergency room admission	177 (17%)
Transfer from hospital	134 (13%)
BP phenotype ³	
Dipper ($\mathcal{P}_D$)	121 (12%)
Non-dipper ($\mathcal{P}_{ND}$)	515 (49%)
Reverse dipper ($\mathcal{P}_{RD}$)	415 (39%)
# measurements ¹	
SBP	19.5 (4.0)
HR	21.3 (2.6)
RR	21.4 (2.5)
PAT	20.5 (4.2)
PPG _{fts}	14.1 (5.2)

¹ Patient-wise mean (standard deviation). ² Oxford Acute Severity of Illness Score (OASIS), a severity of illness score used for predicting patient outcomes [348]. ³ Values given as count (% of total number of patients)

Approximately 70% of patients exhibited a rhythm across all ΔPPG_{fts} . $\Delta Gauss_{\sigma 4/A1}$ was reported to be the lowest % *rhythm* metric, 49.48%. For 14 (42.42%) ΔPPG_{fts} there was a strong ($|\rho_p| > 0.7$) correlation with ΔSBP across the whole cohort. 18 (54.55%) ΔPPG_{fts} had consistent directional relationships across all BP phenotypes. 5 ΔPPG_{fts} had strong correlations across all phenotypes with consistent directions: ΔVPG_{PCA1} , ΔSys_{σ} , ΔPPG_{PCA1} , ΔSTT , and $\Delta Width_{50}$.

These results are visualised in figure 7.8 showing the population-level double-cosinor deviation profiles for all signals. Figure 7.9 shows these profiles for all signals grouped by BP phenotype. As expected, a clear distinction between the phenotypes was observed for ΔSBP . ΔHR and ΔRR showed differences in their profiles only during the night, with daytime values appearing to

be nearly identical across all phenotypes. ΔPAT and certain $\Delta\text{PPG}_{\text{fts}}$ show clear differences in their profiles across the three phenotypes.

7.7.3 Distributions of double-cosinor parameters

Table 7.4 reports $CI_{95}^{\theta_{\text{Orth}}}$ and $CI_{95}^{\theta_{\text{Bathy}}}$ as hours relative to midnight, so for example, $CI_{95}^{\theta_{\text{Bathy}}} = [-1.69, 3.06]$ indicates that there was 95% confidence that the mean peak timing existed between 10.19pm to 3.04am for that signal. Signals with tight phase confidence intervals indicate consistent peak and nadir timings. The intersection of these confidence intervals with that of SBP suggests the signals were synchronous. The nadir of ΔSBP typically occurred between 1.52am and 4.52am. The peak of ΔSBP typically occurred between 4.32pm and 6.19pm. The nadirs and peaks of ΔHR were delayed compared to ΔSBP and occurred between 4.53am and 6.13am, and between 6.20pm and 7.21pm respectively. Note that for signals such as ΔPAT where an inverse relationship to SBP was expected, the peaks occurred at similar times to the nadirs of SBP, and vice versa. Typically, $\Delta\text{PPG}_{\text{fts}}$ reported much wider confidence intervals than any of the vital signs or PAT.

Figure 7.8 additionally highlights the $CI_{95}^{\theta_{\text{Orth}}}$ and $CI_{95}^{\theta_{\text{Bathy}}}$ for all signals. Note that there is no guarantee that $CI_{95}^{\theta_{\text{Orth}}}$ and $CI_{95}^{\theta_{\text{Bathy}}}$ align with the respective nadir and peak of the population-level profile. This is a result of the different methods through which the two were estimated. The population level profile was produced by computing the population mean of $[\beta_1, \gamma_1, \beta_2, \gamma_2]$ (see section 7.6.3) across the cohort and using the trigonometric relations to produce the phases of the population double-cosinor model, whereas the confidence intervals were produced by bootstrap circular mean resampling. An example of these two methods not producing the exact same results can be viewed for the peak of ΔSBP in figure 7.8 where the bathphase of the population-level circadian rhythm proceeds $CI_{95}^{\theta_{\text{Bathy}}}$.

Table 7.4 provides the results of the non-parametric statistical analysis of circadian parameters between patients in the $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ cohorts. As previously reported [342, 349] and demonstrated in figure 7.9, SBP contains a number of circadian parameters that exhibit notable differences between the two phenotypes. Significant ($p < 0.001$) differences were observed between the peak and nadir timings of the SBP double-cosinor model. Additional significant differences were found for A_1 ($p < 0.001$), ϕ_1 ($p = 0.0016$), and ϕ_2 ($p = 0.017$). A number of other signals also exhibited statistically significant differences between the $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ cohorts. 7 (non-SBP) signals exhibited more than 2 significant differences at the $p < 0.01$ level: ΔHR , ΔPAT , ΔVPG , PCA_1 , ΔSys_σ , $\Delta\text{PPG PCA}_1$, $\Delta\text{Kurtosis}$, and $\Delta d/a$.

Table 7.4: Results of double-cosinor analysis for all signals evaluated divided into three groups: vital signs (containing ΔSBP , ΔHR and ΔRR), ΔPAT , and $\Delta\text{PPG}_{\text{fts}}$. 'Population-level correlation' indicates the strength of agreement of the signal to the population-level double-cosinor profile of ΔSBP for All patients as well as those in the $\mathcal{P}_D$, $\mathcal{P}_{ND}$, and $\mathcal{P}_{RD}$ cohorts. Intra-group signals are ordered by their $|\rho_p|$ with ΔSBP on all patients. $CI_{95}^{\theta_{\text{Orth}}}$ and $CI_{95}^{\theta_{\text{Bathy}}}$ values are reported as hours relative to midnight. 'Double-cosinor model parameters across phenotypes' provides a statistical analysis of circadian parameters between patients in the $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ cohorts. For A_1 and A_2 , the comparison was performed using a Mann-Whitney U Test. For phases (ϕ_1 , ϕ_2 , θ_{Orth} , θ_{Bathy}), the comparison was performed using a non-parametric test for equal circular medians.

Signal name	% <i>rhythm</i>	<i>Population-level correlation</i>				θ_{Orth} & θ_{Bathy} 95% CI		<i>Double-cosinor params across phenotypes</i>					
		All	$\mathcal{P}_D$	$\mathcal{P}_{ND}$	$\mathcal{P}_{RD}$	θ_{Orth}	θ_{Bathy}	A_1	A_2	ϕ_1	ϕ_1	θ_{Orth}	θ_{Bathy}
<i>Vital signs</i>													
ΔSBP	68.89%	---	---	---	---	[1.88, 4.86]	[16.54, 18.32]	***	n.s.	*	n.s.	***	***
ΔHR	82.21%	0.91	0.87	0.92	−0.42	[4.89, 6.23]	[18.34, 19.35]	***	n.s.	n.s.	n.s.	*	n.s.
ΔRR	55.47%	0.91	0.98	0.98	−0.79	[4.19, 5.84]	[16.69, 18.15]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
ΔPAT													
ΔPAT	88.20%	−0.99	−0.99	−0.98	−0.83	[16.61, 18.58]	[2.32, 6.54]	**	n.s.	***	n.s.	***	**
$\Delta\text{PPG}_{\text{fts}}$													
$\Delta\text{VPG PCA}_1$	73.36%	−1.00	−0.98	−1.00	−0.76	[15.09, 17.34]	[9.59, 21.59]	**	*	n.s.	*	n.s.	n.s.
$\Delta\text{Sys}_{\sigma}$	71.08%	−0.97	−0.91	−0.97	−0.70	[15.13, 17.48]	[−3.28, 8.72]	**	**	n.s.	n.s.	n.s.	n.s.
$\Delta\text{PPG PCA}_1$	71.46%	−0.94	−0.94	−0.91	−0.86	[14.33, 16.61]	[−1.69, 3.06]	**	*	n.s.	n.s.	n.s.	***
$\Delta\text{APG PCA}_1$	74.98%	−0.94	−0.92	−0.97	−0.11	[16.34, 18.12]	[2.00, 14.00]	**	*	n.s.	n.s.	n.s.	n.s.
ΔSTT	72.69%	0.91	0.90	0.99	0.95	[9.81, 21.81]	[14.02, 17.29]	**	*	n.s.	**	*	n.s.
$\Delta\text{APG PCA}_2$	73.17%	0.90	0.85	0.54	−0.80	[2.30, 14.30]	[16.02, 18.94]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
$\Delta\text{e} / \text{a}$	60.50%	0.84	−0.14	0.73	0.03	[11.74, 20.62]	[13.89, 19.54]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
ΔWidth_{50}	72.41%	0.83	0.84	0.85	0.78	[21.16, 25.20]	[13.07, 15.38]	**	*	**	n.s.	***	n.s.
$\Delta\sigma_{g3}$	66.51%	−0.77	−0.29	−0.81	0.70	[13.66, 23.24]	[7.75, 19.75]	*	n.s.	n.s.	n.s.	n.s.	n.s.
$\Delta\text{b} / \text{a}$	68.41%	−0.76	−0.55	−0.36	0.78	[17.33, 19.99]	[5.67, 10.30]	*	n.s.	n.s.	n.s.	n.s.	**
$\Delta\text{APG PCA}_3$	75.07%	0.74	0.82	0.63	0.61	[15.43, 20.02]	[10.19, 13.75]	n.s.	n.s.	***	n.s.	n.s.	**
$\Delta\sigma_{g1}$	66.79%	0.72	0.77	0.86	0.68	[10.46, 14.43]	[15.96, 19.02]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

Continued on next page...

...Continued from the previous page

Signal name	% rhythm	Population-level correlation				θ_{Orth} & θ_{Bathy} 95% CI		Double-cosinor params across phenotypes					
		All	$\mathcal{P}_D$	$\mathcal{P}_{ND}$	$\mathcal{P}_{RD}$	θ_{Orth}	θ_{Bathy}	A_1	A_2	ϕ_1	ϕ_1	θ_{Orth}	θ_{Bathy}
$\Delta\text{APG PCA}_3$	75.07%	0.74	0.82	0.63	0.61	[15.43, 20.02]	[10.19, 13.75]	n.s.	n.s.	***	n.s.	n.s.	**
$\Delta\sigma_{g1}$	66.79%	0.72	0.77	0.86	0.68	[10.46, 14.43]	[15.96, 19.02]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
$\Delta\text{Gauss}_{\sigma 4/A1}$	49.48%	0.71	0.84	0.75	-0.42	[-2.12, 9.88]	[14.84, 18.10]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
$\Delta\text{Dia}_{\sigma}$	61.46%	-0.71	-0.84	-0.83	-0.67	[16.95, 19.76]	[5.25, 17.25]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
$\Delta\text{Skewness}$	71.93%	-0.70	-0.75	-0.82	-0.88	[12.99, 14.93]	[21.83, 24.80]	**	*	**	n.s.	n.s.	***
$\Delta\text{Gauss AI}_R$	66.03%	-0.68	-0.72	-0.66	-0.19	[17.13, 19.68]	[7.33, 19.33]	*	n.s.	n.s.	n.s.	n.s.	n.s.
$\Delta\text{VPG PCA}_3$	76.50%	0.67	0.55	0.78	-0.86	[5.23, 8.18]	[16.62, 18.29]	*	n.s.	n.s.	n.s.	n.s.	n.s.
$\Delta\text{Gauss LVET}$	69.17%	0.62	0.86	0.58	0.72	[9.83, 18.43]	[16.76, 20.01]	n.s.	n.s.	n.s.	n.s.	**	*
$\Delta\text{PPG PCA}_2$	75.26%	-0.54	-0.18	-0.17	0.93	[15.78, 17.76]	[1.22, 9.03]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
$\Delta\text{slope}_{\text{bd}}$	71.93%	0.50	-0.41	-0.63	-0.89	[8.70, 14.56]	[17.37, 19.94]	n.s.	n.s.	***	n.s.	*	*
$\Delta\text{Gauss RI}_R$	71.08%	0.49	0.83	0.68	0.48	[20.74, 24.24]	[12.32, 14.63]	n.s.	n.s.	n.s.	n.s.	**	n.s.
ΔA_{g2}	63.56%	0.41	0.66	0.29	0.35	[12.01, 17.26]	[17.88, 21.46]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
$\Delta\text{Kurtosis}$	72.22%	-0.37	-0.81	-0.75	-0.89	[12.31, 14.92]	[18.95, 22.49]	**	*	***	n.s.	*	***
$\Delta\text{VPG PCA}_2$	73.74%	-0.34	0.20	-0.03	0.79	[18.59, 21.44]	[10.74, 13.55]	**	n.s.	n.s.	n.s.	**	n.s.
ΔIPA	75.55%	0.31	0.51	0.44	-0.95	[22.16, 25.83]	[13.21, 15.04]	n.s.	n.s.	n.s.	n.s.	n.s.	*
$\Delta c / a$	69.55%	0.24	0.40	-0.30	-0.53	[18.39, 23.04]	[12.45, 15.73]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
$\Delta\text{PPG PCA}_3$	75.64%	-0.23	-0.43	-0.44	-0.07	[11.84, 14.33]	[19.51, 22.94]	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
ΔIHAR	72.88%	-0.22	-0.42	-0.41	0.84	[12.90, 14.78]	[21.44, 25.08]	n.s.	n.s.	n.s.	n.s.	*	n.s.
ΔA_{g3}	73.36%	0.20	0.33	0.54	0.62	[20.77, 23.37]	[11.94, 14.01]	*	*	*	n.s.	**	n.s.
$\Delta\text{Gauss AI}$	73.17%	-0.20	-0.42	-0.59	-0.63	[12.07, 14.21]	[20.48, 23.28]	n.s.	n.s.	***	n.s.	n.s.	**
$\Delta\text{Gauss RI}$	73.26%	0.18	0.08	-0.24	-0.46	[11.64, 13.71]	[18.61, 22.01]	n.s.	n.s.	n.s.	n.s.	n.s.	*
ΔNHA	60.50%	0.17	0.04	-0.10	-0.73	[10.36, 15.38]	[16.17, 19.59]	**	*	n.s.	n.s.	n.s.	n.s.
$\Delta d / a$	71.93%	0.15	-0.75	-0.52	-0.80	[11.66, 15.53]	[18.56, 21.51]	n.s.	n.s.	***	n.s.	n.s.	**

n.s. Indicates non-significant differences. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

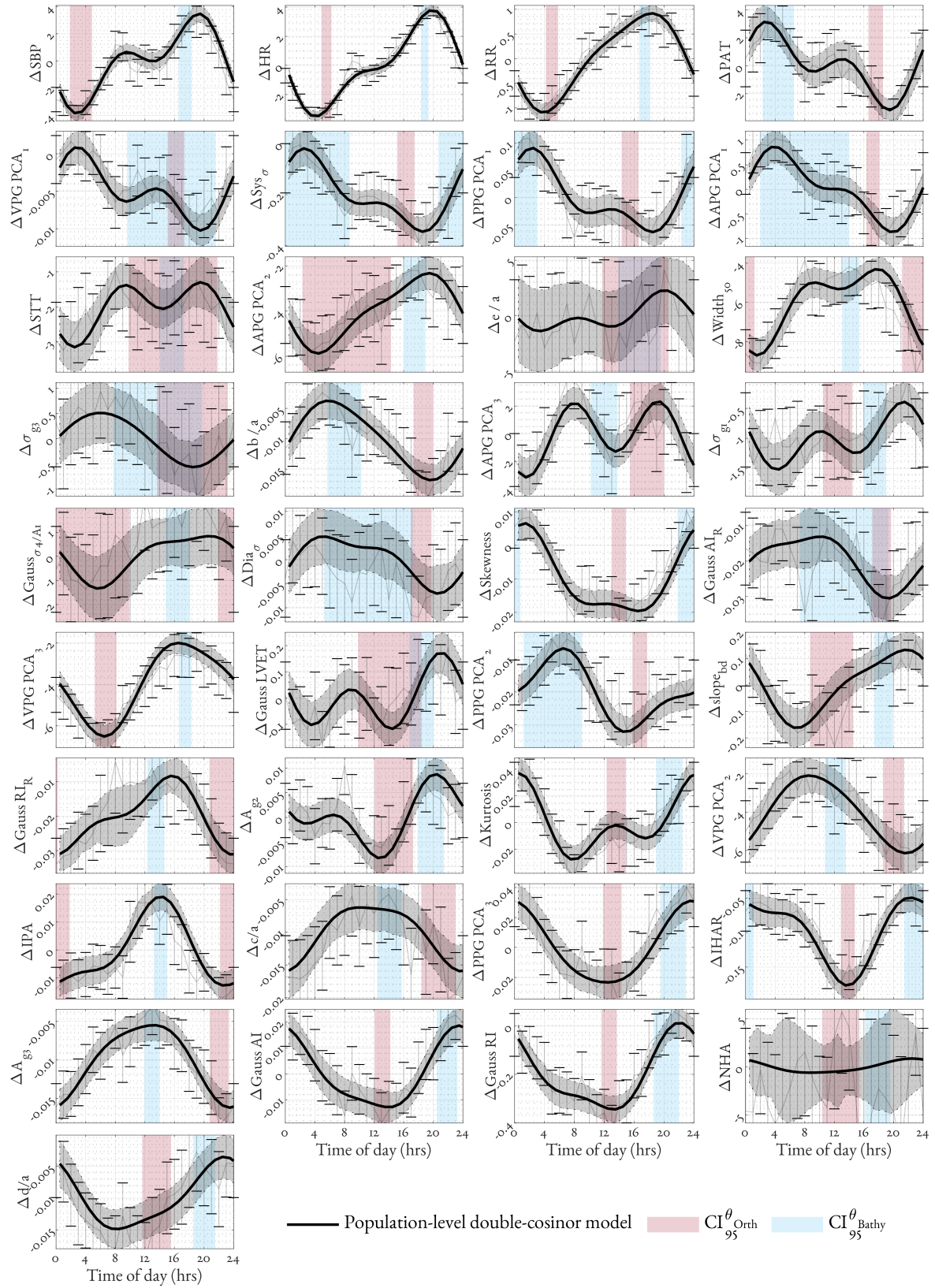


Figure 7.8: Visual comparison of the population-level double-cosinor circadian deviation models. The results for all signals are shown: vital signs (ΔSBP , ΔHR , and ΔRR), ΔPAT , and ΔPPG_{fts} . The error bars denote the distributions of the signal recorded for each hourly bin. The grey area indicates the 95% confidence interval for the resulting double-cosinor model. The red and blue areas show the $CI_{95}^{\theta_{Orth}}$ and $CI_{95}^{\theta_{Bathy}}$ respectively for all signals.

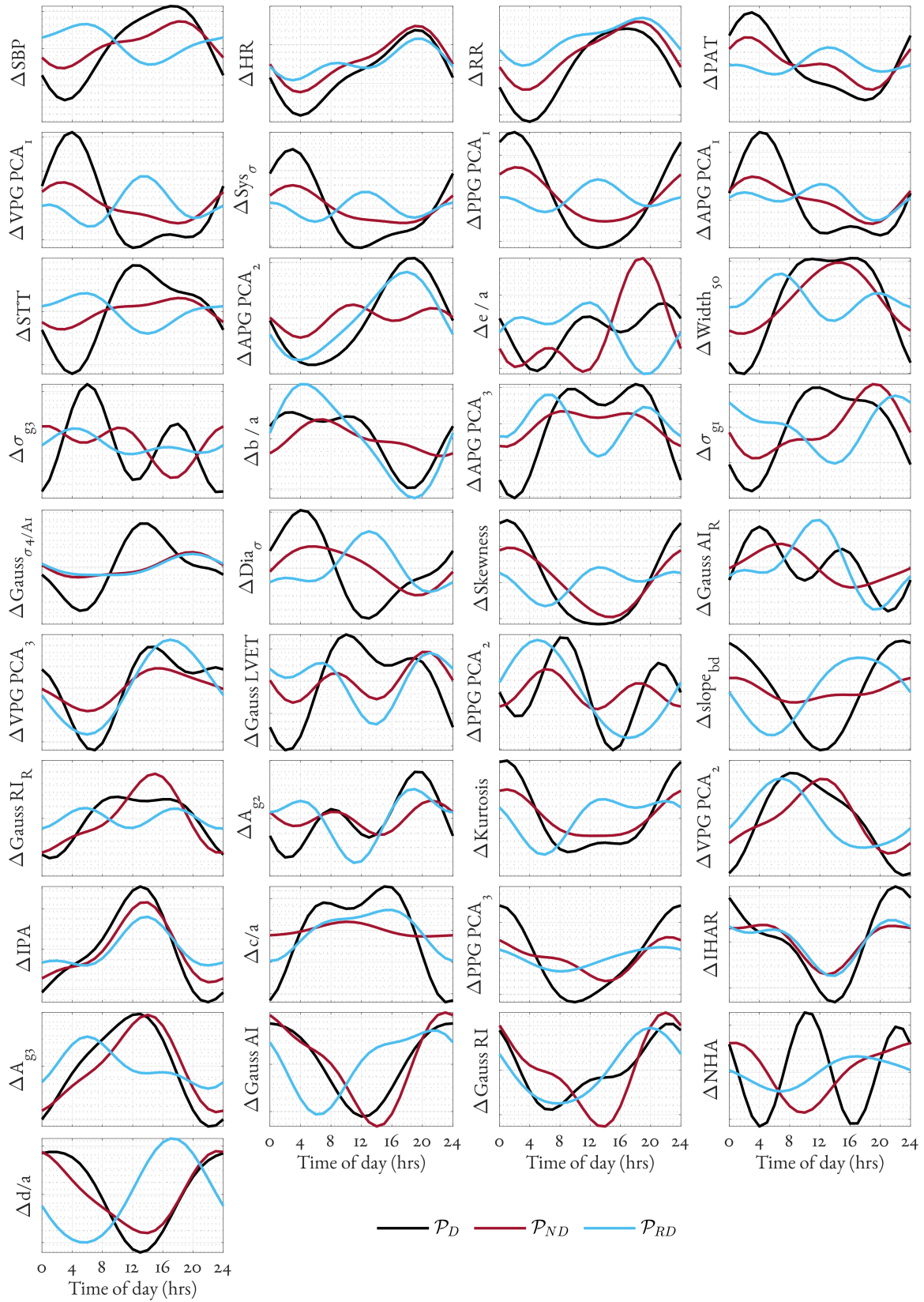


Figure 7.9: Visual comparison of the population-level double-cosinor circadian deviation models for all signals grouped by BP phenotypes. $\mathcal{P}_D$ - Dipper, $\mathcal{P}_{ND}$ - Non-dipper, $\mathcal{P}_{RD}$ - Reverse dipper.

7.8 Discussion

Patient care in the ICU typically involves maintaining homeostasis or 'normalisation' of vital signs [350]. Routine clinical interventions, combined with drug treatments such as sedatives; and external factors such as ventilation, environmental light, and noise levels can disrupt a patient's natural circadian rhythm [351]. Despite this, previous work by Davidson *et al.* [2] has highlighted that circadian patterns of vital signs can be observed during the final 24-hours of stay in patients who recovered from the ICU, as opposed to patients who did not survive or were discharged with palliative care.

In this work, the periodic nature of vital signs (SBP, HR, RR), PAT, and PPG_{fts} was characterised in a simple manner using the rhythmometric double-cosinor model. The vital-sign population-level circadian rhythm deviations conform to those documented in a non-hospital setting [21, 352, 353], see figure 7.8. This suggests that the rhythms observed in Δ PAT and Δ PPG_{fts} also represent endogenous natural variations in the signals rather than being governed by environmental influences. However, proving this to be the case is not possible for the ICU dataset we analysed.

Despite the cohort including patients across all SBP phenotypes, the population-level pattern of Δ SBP contained the salient characteristics of a typical healthy individual in a non-hospital setting. This includes a dip at night-time (with a nadir typically around 3am), a morning surge (typically around 9am), and an afternoon peak (typically around 5pm). The population-level rhythm of HR reported a strong correlation with that of SBP; however, this relationship was not consistent across phenotypes. In numerous clinical studies, a strong link between nocturnal BP reduction and increased baroreflex sensitivity has been established [354, 355, 356]. This relationship subsequently plays a significant role in the weakening association between HR and SBP in the $\mathcal{P}_{RD}$ cohort. As a result, HR on its own is not able to track SBP circadian rhythms reliably across all BP phenotypes.

With 88.2% of patients exhibiting a rhythm, double-cosinor modelling of Δ PAT across 24 hours appears to be appropriate. A clear and observable population-level circadian rhythm of Δ PAT was present in the cohort and was in synchronicity with the circadian profile of Δ SBP. The peak timings of Δ PAT were closely aligned with the nadirs of Δ SBP, and vice versa. When subdividing the data according to BP phenotype this relationship persisted. We propose, therefore, that these results suggest that PAT has its own endogenous, phenotypically modified, circadian pattern that aligns closely with that of SBP. Furthermore, as demonstrated in chapter 5, calibrating PAT measurements to an individual requires estimation of both a calibration intercept and slope. In this work, the calibration intercept was accounted for by computing circadian rhythm deviations around the computed MESOR (the midline around which the double-cosinor models

oscillate). However, the individualised calibration slopes were not appropriately handled. It is likely, therefore, that accounting for these individual differences in slopes would significantly improve the modelled population-level circadian rhythm profile of PAT. Failing this, it appears that longitudinal monitoring of PAT as its own signal may provide vital insights into long-term BP variability (BPV) in the general population without the need for cumbersome cuff measurements. Indeed, there is increasing evidence across multiple clinical studies that increased long-term BPV is independently associated with cardiovascular mortality [357, 358] and neurodegenerative cognitive impairment [359, 360]. PAT may provide a simple unobtrusive method through which BPV can be assessed.

Many $\Delta\text{PPG}_{\text{fts}}$ exhibited an observable, phenotypically modified, circadian pattern (see figure 7.9). However, only a handful of these signals reported consistently strong correlations with SBP across the three phenotypes analysed. $\Delta\text{VPG PCA}_1$, ΔSys_σ , $\Delta\text{PPG PCA}_1$, ΔSTT , and ΔWidth_{50} all reported strong ($|\rho_p| > 0.7$) correlations with SBP, with consistent directional relationships across all BP phenotypes. There were, however, some notable inconsistencies with the results of the previous chapter where Kurtosis, σ_{g1} , and $\text{Gauss}_{\sigma4/A1}$ were all found to have the highest feature importance for accurate BP estimation. This lack of consistency may be driven by the fact that non-linear relationships between PPG_{fts} and SBP were observed in the MOLLIE dataset (see figure 6.9), whereas the analysis performed in this chapter largely favours linear relationships. $\Delta\text{Kurtosis}$ for example, shows a clear rhythm with tight θ_{Orth} and θ_{Bathy} confidence intervals (see figure 7.8). However, a weak negative correlation ($\rho_p = -0.37$) with ΔSBP was reported, whereas in figure 6.9 a positive correlation was reported (median $\rho_p = 0.55$). This supports the recent conclusions from Lin *et al.* [315] that there are inconsistencies in the relationships between certain PPG_{fts} and SBP depending on the physiological mechanisms through which changes in SBP are induced. This may be a significant limitation to BP estimation using PPG_{fts} and requires further work. Nonetheless, the aforementioned PPG_{fts} that demonstrated strong population-level correlations across all phenotypes in this work may well offer a linear relationship to estimate BP circadian rhythms as opposed to the non-linear relationship reported in the previous chapter.

Strong performance at a population level, however, does not indicate similar performance on a patient level. Indeed, all $\Delta\text{PPG}_{\text{fts}}$ signals reported wide $CI_{95}^{\theta_{\text{Orth}}}$ and $CI_{95}^{\theta_{\text{Bathy}}}$ limits that were inconsistent with the peak and nadir timings of ΔSBP (see figure 7.8). This demonstrates that PPG_{fts} show significantly greater inter-patient variability than PAT. One possible cause of this behaviour may be small random errors in fiducial point detection (see section 7.4.2), which are then averaged out when computing the population-level circadian rhythms. Nevertheless, poor performance of individual PPG_{fts} at a patient level does not suggest that individual circadian rhythms cannot be

identified from the PPG as BP estimation typically involves combining a number of features using a machine learning approach. We explore this in more detail in the subsequent chapter.

Additionally, we have demonstrated that there are statistically significant differences in double-cosinor model parameters between patients in the $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ phenotype cohorts. 7 (non-SBP) signals exhibited more than two significant differences at the $p < 0.01$ level: ΔHR , ΔPAT , $\Delta\text{VPG PCA}_1$, ΔSys_σ , $\Delta\text{PPG PCA}_1$, $\Delta\text{Kurtosis}$, and $\Delta d/a$. In general, these differences stemmed from the phases of the double-cosinor models, however, ΔHR , $\Delta\text{VPG PCA}_1$, and ΔSys_σ all reported significant differences for A_1 , the amplitude of the 24-hour period cosine. More appropriate calibration (by, as discussed above, incorporating calibration slopes) should improve the ability of cosinor amplitudes to separate individuals into $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ cohorts. Regardless, these results propose a completely calibration-free procedure through which cosinor features of PAT and PPG_{fts} may be able to classify nocturnal blood pressure patterns into one of the three phenotypes. We additionally explore this in more detail in the subsequent chapter.

7.8.1 Limitations

There are several noteworthy limitations to the conclusions made in this chapter.

Firstly, our approach was based on prior research indicating the presence of vital sign circadian rhythmicity during the final 24 hours of stay in the ICU for patients who ultimately recovered [2, 253]. However, the ICU environment is far from ideal for circadian rhythm analysis, given the frequent medical interventions, disruptions in environmental conditions, and irregular sleep patterns experienced by patients. These complex factors can significantly confound the accurate identification and analysis of circadian patterns.

Moreover, ICU patients are under significant physiological stress. The median length of stay in our dataset was 2.4 days (see table 7.3), which is relatively short compared to ICU stays in other countries [361, 362]. This implies that the 24-hour period analysed covered a substantial portion of each patient's ICU stay. During their ICU stay, patients often receive heavy medication, and although we excluded measurements taken while patients were under the influence of medications likely to impact the vital signs being measured (e.g., vasopressors and β -blockers), accurate documentation of medication administration may not always have been feasible due to the demands of patient care. Consequently, some vital sign measurements in this dataset may reflect the effects of specific medications rather than endogenous variations. The high proportion of 'reverse-dippers' (39%) compared to a community dataset (11%) [5] further suggests that the observed patterns may not represent typical circadian variations in the wider population.

Furthermore, it is important to distinguish between circadian and diurnal patterns (see section 2.3.2). While our analysis primarily focused on circadian rhythms, some of the variations observed may actually be diurnal in nature, influenced by factors other than the internal circadian clock. The distinction between circadian and diurnal patterns is particularly complex in a clinical setting where patients are exposed to irregular schedules and medical interventions.

The circadian rhythms we studied were computed at the population level by averaging across the cohort. Circadian rhythms can significantly differ between individuals [7], and accurately characterising individual-level patterns would have required more extensive longitudinal data than available in the MIMIC dataset. The absence of individual-level analysis makes it challenging to determine whether the observed patterns are consistent across all patients, potentially overlooking critical individual differences. Additionally, our study did not stratify results based on factors like sex, age, or ethnicity, all of which have been shown to influence circadian rhythms. These factors can introduce substantial variations in circadian patterns, underscoring the need for caution when generalising our findings to diverse populations. Additionally, averaging signals, particularly in the context of cuffless surrogates, presents a significant challenge. While the calibration issue was partially addressed by computing circadian rhythm deviations around the MESOR, it did not fully account for individualised calibration slopes. Proper handling of these calibration slopes is essential to enhance the accuracy in capturing population-level circadian rhythms. The presence of inter-individual variability in calibration may further impact the reliability of cuffless surrogate signals for monitoring circadian rhythms.

Finally, there is no accepted gold standard technique for analysing circadian rhythms in a cohort, and the double-cosinor model employed in this study relies on specific assumptions. Two core assumptions of the double-cosinor modelling are the stationarity of the rhythms and the linearity of their representation as combinations of cosine functions [335]. Indeed, it is noteworthy that only 55% of patients exhibited a rhythm in RR. Given the typical circadian pattern of RR, which often comprises a single peak during late afternoon and a trough during sleep [352], it may have been appropriate to employ a single-cosinor model, as previously utilised [363], for a more accurate representation of the RR circadian pattern. This phenomenon (that more appropriate modelling may be achieved through fewer or more than two cosinor terms) may also exist for PAT and PPG_{fts} . However, in order to maintain a consistent comparison with SBP (where double-cosinor modelling has been previously reported to be appropriate [337, 342, 335, 341]), we decided to implement double-cosinor modelling for all signals.

7.9 Conclusion

In this chapter, we explored 24-hour variations in PAT and PPG_{fts} across a large, heterogenous, cohort of patients in the ICU. To our knowledge, the work presented in this chapter presents the largest cohort of individuals from which 24-hour variations of PAT and PPG_{fts} have been examined. We have demonstrated that many of these signals have their own endogenous circadian rhythm that can be modelled using the rhythmometric double-cosinor model at the population level. Strong correlations between the population-level profile of ΔSBP and ΔPAT were observed and this relationship persisted across all BP phenotypes. A number of $\Delta\text{PPG}_{\text{fts}}$ also exhibited a strong population-level relationship with ΔSBP . However, we noted inconsistencies in the peak and nadir timings for these signals, suggesting weaker relationships at the patient level. Finally, we showed statistically significant differences in the double-cosinor parameters of HR, PAT, and a subset of PPG_{fts} between patients in the $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ phenotype cohorts.

The results of this chapter offer a strong indication that it may be possible to compute a cuffless estimate of the circadian rhythm of BP at the patient level. In the next chapter, we extend this analysis to propose two novel methods to classify an individual's BP phenotype using the signals evaluated in this chapter.

8

Classification of blood pressure phenotypes

Contents

8.1	Introduction	189
8.2	Methods overview	190
8.3	Cohort selection	190
8.3.1	Blood pressure phenotype labelling	191
8.4	Method 1: Regularised <i>individual calibration</i>	191
8.4.1	Feature extraction	192
8.4.2	Model construction	192
8.4.3	Model generalisability	194
8.5	Method 2: Calibration-free	194
8.5.1	Feature extraction	194
8.5.2	Model construction	196
8.5.3	Model generalisability	197
8.6	Baseline reference method	198
8.7	Evaluation and generalisability of methods	198
8.7.1	Evaluation metrics	198
8.7.2	Generalisability of methods	199
8.8	Results	200
8.8.1	Dataset	200
8.8.2	Overall performance	201
8.8.3	Method 1: Regularised <i>individual calibration</i>	202
8.8.4	Method 2: Calibration-free	204
8.9	Discussion	204
8.9.1	Method 1: Regularised <i>individual calibration</i>	207
8.9.2	Method 2: Calibration free	209
8.9.3	Limitations	210
8.10	Conclusion	211

8.1 Introduction

The results of chapter 7 suggest that PAT, and a subset of features from the PPG waveform, have an endogenous circadian rhythm at the population level. A number of these signals were also found to be synchronous with BP across all BP phenotypes and we highlighted statistically significant differences in the circadian parameters of these signals between patients in the $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ cohorts. In this chapter, we present two novel methods to classify BP phenotypes at the individual level. Our first approach requires calibration measurements taken while the individual is awake, while the second approach is completely calibration-free. Both methods can

be easily integrated into primary care practice without the need for ABPM, the current gold standard for evaluating 24-hour BP patterns.

8.2 Methods overview

Data was gathered from the MIMIC-III ICU dataset constructed in the previous chapter. In this chapter, we propose two new signal-based methods for BP phenotype classification through analysis of the PAT and PPG_{EA} PPG_{fts} signals. As a reference comparison, we implemented a *baseline reference* method that examined the extent to which an individual's phenotype may be estimated from their age and sex alone.

The first proposed method ('regularised *individual calibration*') utilised measurements of SBP, PAT, and PPG_{fts}, taken between 8am and midnight, as training data for mixed-effects regression models (linear mixed-effects model and mixed-effects random forest) to estimate nocturnal (midnight to 8am) SBP values. A patient's BP phenotype was subsequently classified using the estimated nocturnal SBP, and the reference daytime measurements. This approach addresses a major limitation of ABPM where nocturnal cuff inflations have been shown to cause sleep disruption as well as sleep arousal, wherein BP increases as the measurement is taken [23, 111].

The second method ('calibration-free') did not require any SBP calibration measurements altogether. Instead, characteristic circadian features, including double-cosinor parameters and dip ratios, of the 24-hour profile of PAT and PPG_{fts} were extracted. We subsequently trained linear (logistic regression) and non-linear (random forest) ordinal classification models to classify a patient's BP phenotype.

8.3 Cohort selection

We made use of the MIMIC dataset constructed in the previous chapter (see section 7.5). From this dataset, we selected the subset of the patients whose 24-hour SBP time series passed the double-cosinor F-test for the presence of a rhythm at the $p < 0.05$ level. This step reduced the likelihood of predicting phenotypes for patients whose 24-hour SBP profile was significantly corrupted by noise such that their true phenotype will have been different from that determined by the day-to-night SBP dip ratio (DR_{SBP}). Additionally, these individuals exhibited patterns in SBP that were most similar to those circadian patterns observed in a community setting. Hence they were likely to be individuals well enough to have a valid assessment of circadian rhythms [253].

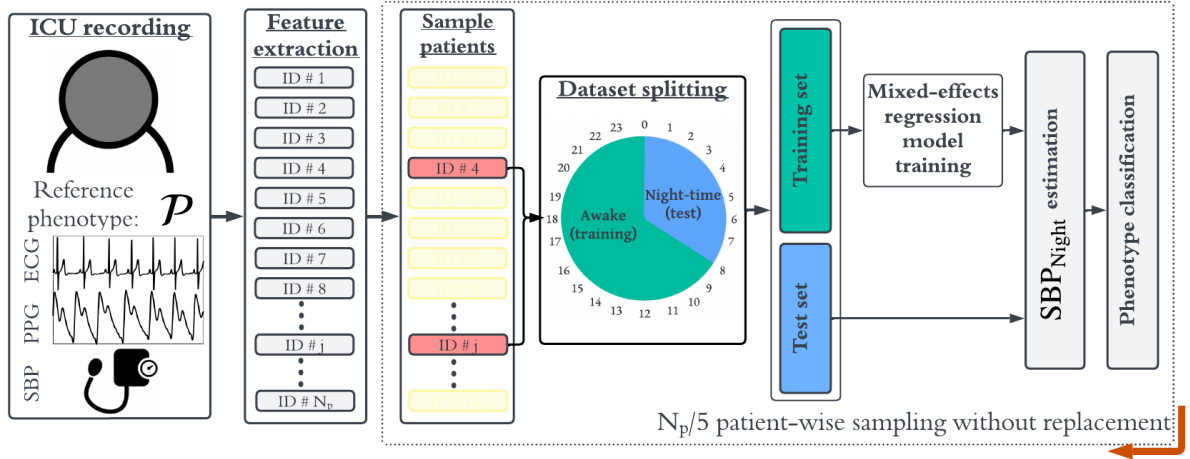


Figure 8.1: Schematic of the 'regularised *individual calibration*' phenotype classification pipeline. SBP, ECG, and PPG time series data was recorded during each patient's final 24 hours in the ICU. Hourly averaged measurements of SBP, PAT, VPG PCA₁, Sys_σ, PPG PCA₁, STT, and Width₅₀ were extracted. We used patient-wise sampling without replacement with a sample size of $N_p/5$ (where N_p is the number of patients in the cohort). Mixed effects regression models (LMEM and MERF) were then trained on data collected while the patient was likely to be awake (8am to midnight) to produce cuffless estimates of $\hat{SBP}_{\text{Night}}$ and correspondingly $\hat{DR}_{\text{SBP}}$ from which an estimate of the individual's BP phenotype could be determined.

8.3.1 Blood pressure phenotype labelling

Each patient's reference phenotype label (i.e. the ground truth): $\mathcal{P}_D$, $\mathcal{P}_{ND}$, or $\mathcal{P}_{RD}$, was determined using the DR_{SBP} criterion outlined in table 7.2.

8.4 Method 1: Regularised *individual calibration*

Our proposed first method ('regularised *individual calibration*') utilised data collected while the patient was awake (defined as 8am to midnight), to enable the estimation of nocturnal (midnight to 8am) SBP without the use of a cuff. A schematic representation of the steps we implemented for this method is shown in figure 8.1. We extracted SBP, PAT, and a relevant subset of the PPG_{EA} PPG_{fts} signals from all patients. Data gathered between 8am to midnight was used to produce cuffless estimates of the SBP time series during the night-time. To evaluate this method and its generalisability, we used patient-wise sampling without replacement with a sample size of $N_p/5$ (where N_p is the number of patients in the cohort).

An estimate of DR_{SBP} ($\hat{DR}_{\text{SBP}}$) was then computed as the percentage difference between the mean estimated night-time SBP ($\hat{SBP}_{\text{Night}}$) and the mean reference daytime (10am to 8pm)

SBP (SBP_{Day}), normalised by the latter:

$$\hat{\text{DR}}_{\text{SBP}} = \frac{\frac{1}{N_N} \sum_i^{N_N} \hat{\text{SBP}}_{i, \text{Night}} - \frac{1}{N_D} \sum_j^{N_D} \text{SBP}_{j, \text{Day}}}{\frac{1}{N_D} \sum_j^{N_D} \text{SBP}_{j, \text{Day}}} \cdot 100\% \quad (8.1)$$

where N_D and N_N are the number of observations during the day and night, respectively. A patient's BP phenotype was subsequently classified based using $\hat{\text{DR}}_{\text{SBP}}$, and the standard SBP day-to-night dip ratio classification rules (see table 7.2).

8.4.1 Feature extraction

In the previous chapter, we demonstrated that the population-level circadian rhythms of PAT and a subset of five PPG_{EA} PPG_{fts} had strong and consistent linear correlations with the circadian rhythm of SBP, across all phenotypes. These five PPG_{fts} were: VPG PCA_1 , Sys_σ , PPG PCA_1 , STT, and Width_{50}^1 . We extracted hourly-averaged measurements of each of these 6 signals and SBP from all patients.

Any poor-quality values from the PAT or PPG_{fts} signals, defined as having a SQI_{PAT} or $\text{SQI}_{\text{PPG}}^{\text{EA}}$ equal to 0 respectively, were not included in the hourly window. Similarly, any hourly window with an averaged SQI_{PAT} or $\text{SQI}_{\text{PPG}}^{\text{EA}}$ less than the $\gamma_{\text{SQI}} = 0.6$ threshold defined in section 7.5 was removed from analysis.

As this method implements a regularised form of *individual calibration*, patient demographics were not included in the feature set. To explore BP phenotype classification using a single source of data (i.e. only the PPG signal), we also compared performance when using only the five PPG_{fts} as input signals.

8.4.2 Model construction

Two mixed-effects regression models: linear mixed-effects model (LMEM) and mixed-effects random forest (MERF), were developed for nocturnal SBP ($\text{SBP}_{\text{Night}}$) estimation. These mixed-effects regression models address the interdependence of multiple observations from the same patient by separating terms into fixed-effects (population-level trends) and random-effects (individual-level variations).

As the mixed-effects models perform a regression task prior to the patient's phenotype classification, $\text{SBP}_{\text{Night}}$ and DR_{SBP} estimation was assessed using the following metrics: ρ_p , RMSE, and MAE.

¹Note that HR was not extracted as a feature in this method.

8.4.2.1 Linear mixed-effects model

A random slopes LMEM was developed in order to explore the linear relationship between the input feature set and SBP while additionally incorporating individual linear calibration slopes for all signals (see appendix D.1.2). Starting from the foundations of an ordinary least squares model, for the j^{th} signal (e.g. PAT), from the p^{th} patient, x_j^p , patient-specific linear calibration slopes can be modelled by $\beta_j + u_j^p$ where β_j is the fixed-effect model coefficient and u_j^p is the random-effect deviation for that patient. The LMEM model can, therefore, be summarised as:

$$y^p = (\beta_0 + u_0^p) + (\beta_1 + u_1^p)x_1^p + \cdots + (\beta_M + u_M^p)x_M^p + \epsilon^p \quad (8.2)$$

$$= \underbrace{\beta_0 + \beta_1 x_1^p + \cdots + \beta_M x_M^p}_{\text{fixed-effect}} + \underbrace{u_0^p + u_1^p x_1^p + \cdots + u_M^p x_M^p}_{\text{random-effect}} + \epsilon^p \quad (8.3)$$

$$= X\beta + Zu^p + \epsilon \quad (8.4)$$

The target vector, y^p , represents the SBP values and the residuals are captured in the vector, ϵ . The number of signals in the feature set, M , is 5 when only PPG_{fts} are used, and 6 when PAT is added. The fixed effects component, $X\beta$, is similar to an ordinary least squares model. As *individual calibration* slopes were modelled for each signal, the random effects design matrix, Z , was equivalent to the feature set, X .

The random effect parameters, u^p , are assumed to be sampled from a multivariate Gaussian: $u \sim \mathcal{N}(0, G)$, where G is the covariance matrix. As a result, an individual's calibration slope for a given signal was regularised against the response across all patients, facilitating a regularised *individual calibration* approach. This effect helped to mitigate against potential overfitting for patients who had only a few calibration datapoints taken during the period between 8am and midnight or whose data was heavily corrupted by motion artefacts. A more in-depth analysis of the impact of mixed-effects regularisation is provided in appendix G. We refer to an LMEM model using only PPG_{fts} as LMEM_{PPG} and as LMEM_{PPG+PAT} when also including PAT as a sixth input signal.

For each signal, we additionally analysed the distributions of the modelled, patient-specific, linear calibration slopes $\beta_j + u_j^p$, $j = \{1, \dots, M\}$, $p = \{1, \dots, N_p\}$ across all patients in the cohort.

8.4.2.2 Mixed effects random forest model

To explore potential non-linear relationships between the feature set X and SBP, while still accounting for multiple observations from the same patient, a MERF model was developed [274]² (see section 4.4.3). MERF maintains the same random-effect structure as LMEM, which assumes

²Implemented using the *merf* Python package: <https://pypi.org/project/merf/>

linear patient-specific variations from the population fixed-effects model. However, the structure of the fixed effect component is defined as an RF regression model, $f(X)$, with the number of features randomly selected at each node ($mtry$) set to the default value of $M/3$ [271]. The MERF model can be summarised as:

$$y^p = f(X) + Zu^p + \epsilon \quad (8.5)$$

with the same nomenclature as equation (8.2). Again, individual calibration slopes were modelled for all signals and so the random effects design matrix, Z , was equal to the feature set X . We refer to a MERF model using only PPG_{fts} as input signals as MERF_{PPG} and as MERF_{PPG+PAT} when including PAT as an input signal.

8.4.3 Model generalisability

Model generalisability was established through patient-wise sampling without replacement with a sample size of $N_p/5$. For each iteration, approximately³ 20% of the patients in the overall cohort were selected and used to train and evaluate the performance of the mixed-effects regression models. This procedure was iterated five times, in order to evaluate all patients in the cohort.

8.5 Method 2: Calibration-free

Our proposed second method ('calibration-free') classified a patient's BP phenotype based on a 24-hour analysis of the PAT and PPG_{EA} PPG_{fts} signals, without any requirement for reference calibration BP values. A schematic representation of the steps we implemented for this method is shown in figure 8.2. We computed estimates of PAT and a relevant subset of the PPG_{EA} PPG_{fts} signals from each patient in our cohort and extracted a set of features to parameterise the 24-hour profile of each signal. Using these features, we implemented a 5-fold ordinal classification cross-validation (CV) procedure for individual patient BP phenotype classification.

8.5.1 Feature extraction

In the previous chapter, we demonstrated that parameters derived from double-cosinor modelling of certain signals exhibited significant differences between the $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ phenotypes. We selected the seven signals that reported significant differences between these phenotypes for at least two cosinor parameters at the $p < 0.01$ level (see section 7.7.3). These seven signals were HR, PAT, and five PPG_{fts}: VPG PCA₁, Sys _{σ} , PPG PCA₁, Kurtosis, and d/a . From each of these

³Approximately 20% because N_p may not divide perfectly by 5

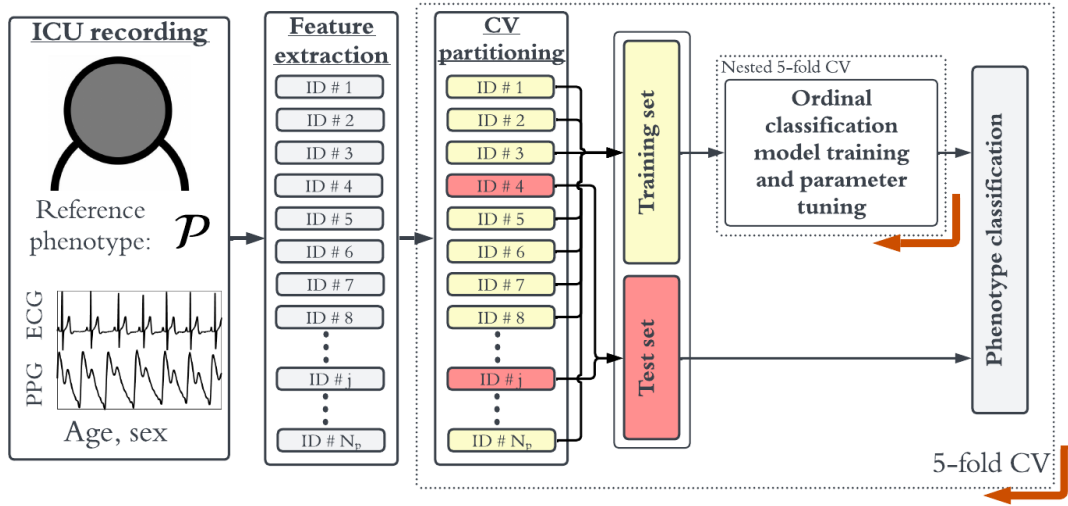


Figure 8.2: Schematic of the 'calibration-free' phenotype classification pipeline. SBP, ECG, and PPG time series data, as well as age and sex demographics, were recorded during each patient's final 24 hours in the ICU. The patient's reference phenotype was identified from their SBP dip ratio (DR_{SBP}). We estimated HR, PAT, VPG PCA_1 , Sys_σ , PPG PCA_1 , Kurtosis, and d/a signals from the ECG and PPG time series. From each patient and each signal, we extracted seven features that characterise the signal's 24-hour profile. Using 5-fold CV we trained, tuned and evaluated ordinal classification models to predict a patient's BP phenotype.

seven signals, we extracted seven features $[A_1, A_2, \phi_1, \phi_2, \theta_{Bathy}, \theta_{Orth}, \tilde{DR}]$ (see below) that parameterised different aspects of its 24-hour profile.

To explore phenotype classification using a single source of data (i.e. only the PPG signal), we compared performance when using only the PPG_{fts} as input signals, to then adding PAT to the feature set. In this analysis, HR was included in the definition of PPG_{fts} . Although HR was computed from the ECG in the MIMIC database, it has been demonstrated in prior research that HR computed from both the ECG and PPG signals has a strong agreement [364]. Finally, we added the patient's sex and age to the feature set as static features.

8.5.1.1 Double-cosinor modelling

The double cosinor modelling technique was applied to each signal. This technique decomposes the signal's 24-hour profile into the summation of a 24-hour cosine (C_1 , see figure 8.3 (a)) and a 12-hour cosine (C_2 , see figure 8.3 (b)). The amplitudes (A_1 and A_2) and phases (ϕ_1 and ϕ_2) of these cosinors were included as features. The overall model is given by $C_1 + C_2$ (see figure 8.3 (c)), with the phase of this model at its peak (θ_{Bathy}) and nadir (θ_{Orth}) also included as two of the seven features.

8.5.1.2 Modified dip ratio

To characterise the diurnal to nocturnal variations of each signal, we introduced a modified version of the day-to-night dip ratio (DR), labelled as $\tilde{DR}$. This modification addresses two limitations

of the standard definition of DR: the assumption that baseline values represent a comparable physiological state across the population (hence normalisation by the daytime value), and the assumption that values of the signal do not decrease below zero. For both PAT and PPG_{fts}, the first of these assumptions is not valid and for PPG_{fts} the second assumption is not valid (for example, PCA features may go below zero). To overcome these limitations, $\tilde{DR}$ was calculated by first z-normalising the signal over the full 24-hour period, and then computing the mean day to mean night difference. An example of computing $\tilde{DR}$ is shown in figure 8.3 (b).

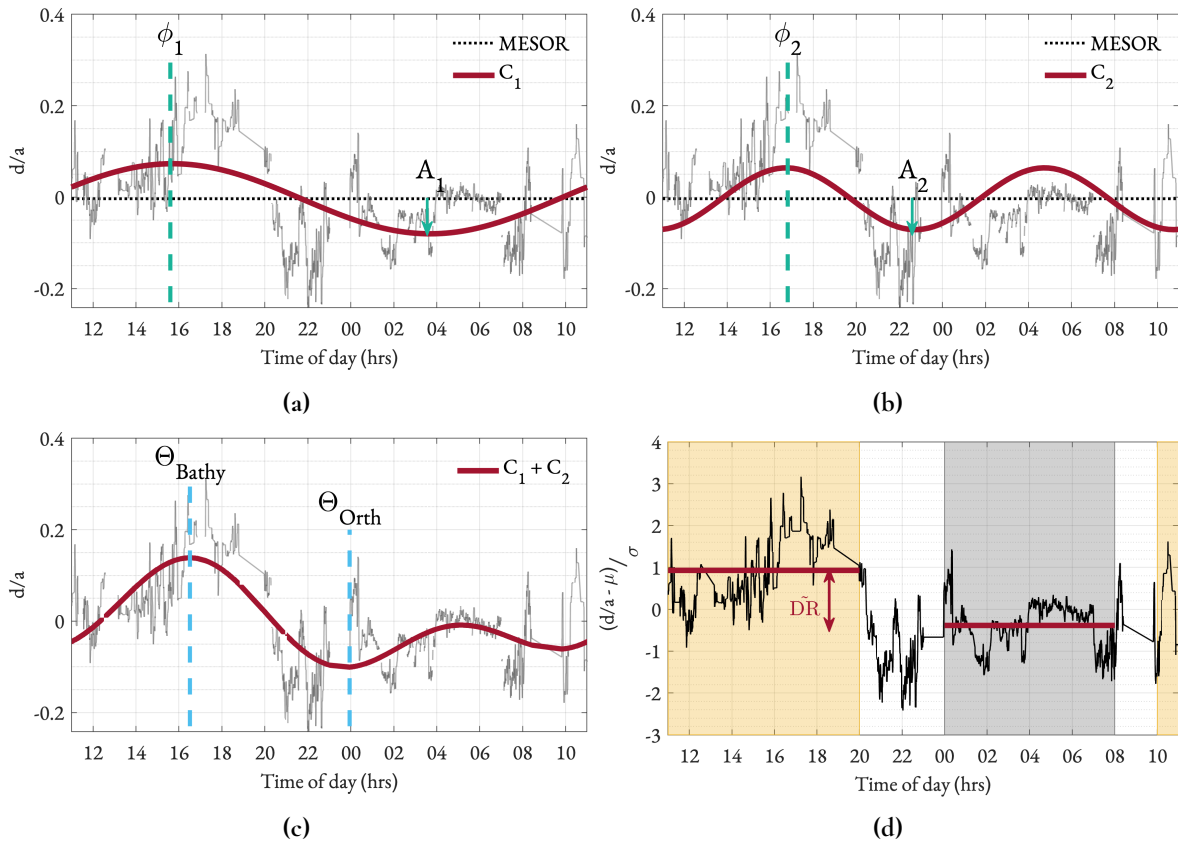


Figure 8.3: Overview of the seven features extracted from each signal (in this instance the d/a PPG_{EA} signal is used as a visual reference) to parameterise their 24-hour profile. Double cosinor modelling of the signal was performed as in the previous chapter. (a) C_1 , the cosinor with a 24-hour period, from which the amplitude A_1 and phase ϕ_1 were derived. (b) C_2 , the cosinor with a 12-hour period, from which the amplitude A_2 and phase ϕ_2 were derived. (c) $C_1 + C_2$, the double-cosinor model, the phase of this model at its peak (θ_{Bathy}) and nadir (θ_{Orth}) were also included as features. (d) Overview of computing the modified dip ratio, $\tilde{DR}$. First, d/a is z-normalised by subtracting the mean (μ) and dividing by the variance (σ). $\tilde{DR}$ is then computed as the mean day to mean night difference of the z-normalised signal.

8.5.2 Model construction

Classifying BP phenotypes presents an ordinal classification problem, where $\mathcal{P}_D \rightarrow \mathcal{P}_{ND} \rightarrow \mathcal{P}_{RD}$ indicates transitions of increasing DR_{SBP} . To address this, the class labels were treated as an ordered

set, and ordinal classification was performed as proposed by Frank *et al.* [365]. This method transforms the 3-class ordinal classification problem into two interlinked binary classification tasks. First, a classification model was trained to predict individuals not in the $\mathcal{P}_D$ cohort ($\mathcal{P}_D$ vs. rest). Then, a separate classification model of the same type was trained to estimate individuals not in either the $\mathcal{P}_D$ or $\mathcal{P}_{ND}$ cohorts ($\mathcal{P}_D + \mathcal{P}_{ND}$ vs. rest). Using the results of these two models, the probability of an individual, y_i , belonging to each of the three phenotypes was computed as follows:

$$\begin{aligned} \Pr(y_i = \mathcal{P}_D) &= 1 - \Pr(\mathcal{P}_D \text{ vs. rest}) \\ \Pr(y_i = \mathcal{P}_{ND}) &= \Pr(\mathcal{P}_D \text{ vs. rest}) - \Pr(\mathcal{P}_D + \mathcal{P}_{ND} \text{ vs. rest}) \\ \Pr(y_i = \mathcal{P}_{RD}) &= \Pr(\mathcal{P}_D + \mathcal{P}_{ND} \text{ vs. rest}) \end{aligned} \quad (8.6)$$

To classify the phenotype of a patient, the two binary classifiers were applied and the phenotype with the highest predicted probability was assigned to the patient. Ordinal classification using this approach requires models that return predicted class likelihoods.

Logistic regression (LR) was implemented to map the input feature set to the class probabilities through a linear combination of features. This model is introduced in appendix D.1.3. To prevent overfitting, remove redundant features, and improve model interpretability, we applied the Least Absolute Shrinkage and Selection Operator (LASSO) regularisation method (see appendix D.1.4). When using only PPG_{fts} as input signals, we refer to the model as $\text{LASSO} + \text{LR}_{\text{PPG}}$. When including PAT as an input signal, we refer to the model as $\text{LASSO} + \text{LR}_{\text{PPG} + \text{PAT}}$. We optimised the regularisation hyperparameter (λ) using a nested 5-fold CV to maximise the accuracy of each classification task.

To investigate non-linear combinations of the input feature set, we additionally implemented an RF classification model. This model is introduced in section 4.4. The optimal number of features randomly selected for each node (labelled m_{try}) was determined through a nested 5-fold CV to maximise the accuracy of each classification task. We refer to an RF model using only PPG_{fts} as input signals as RF_{PPG} and as $\text{RF}_{\text{PPG} + \text{PAT}}$ when including PAT as an input signal.

8.5.3 Model generalisability

The generalisability of this method was determined through 5-fold, patient-wise, CV. After CV partitioning patients into training and test sets, phenotype class imbalances (unequal proportions of patients in the $\mathcal{P}_D$, $\mathcal{P}_{ND}$, and $\mathcal{P}_{RD}$ cohorts) existed. Without accounting for this imbalance, the classification models would be biased towards predicting the phenotype with the highest occurrence. Solutions for handling class imbalances include cost-sensitive and resampling methods [366]. Cost-sensitive methods require learning the costs associated with misclassifying examples. This procedure can be subjective, as it depends on the clinical use case. Therefore, we handled phenotype class imbalances by random upsampling of minority classes in the training set until all classes were equally represented [366].

8.5.3.1 Shapley additive explanation values feature importance

We assessed the importance of each feature for phenotype classification through SHAP values. SHAP values, ϕ_j^i , reflect the marginal contribution of the j^{th} feature to the i^{th} prediction. For classification tasks, ϕ_j^i measures the change in predicted probability assigned to the positive class when the j^{th} feature is added to the feature set. In this work, we used an ordinal classification approach which involved two nested binary classification tasks ($\mathcal{P}_D$ vs. rest, and $\mathcal{P}_D + \mathcal{P}_{ND}$ vs. rest), so the SHAP value of each prediction was split into two parts: $\phi_j^i(\mathcal{P}_D)$ and $\phi_j^i(\mathcal{P}_D + \mathcal{P}_{ND})$ respectively.

For each fold of the CV, we assessed the overall importance of each feature over the training set as:

$$imp_j = \frac{1}{N_{train}} \sum_{i=1}^{N_{train}} (|\phi_j^i(\mathcal{P}_D)| + |\phi_j^i(\mathcal{P}_D + \mathcal{P}_{ND})|)/2 \quad (8.7)$$

where N_{train} is the number of observations in the training set. imp_j indicates the average impact of the j^{th} feature on model output magnitude. The distributions of imp_j across CV folds were assessed using a ranking coefficient, which ranks the imp_j values between 0 and 1 at each fold.

8.6 Baseline reference method

Demographic variables such as age and sex, are known to correlate with BP [64, 65]. It is important to establish the extent to which these demographic parameters alone are sufficient for accurate phenotype classification. For example, if an individual's BP phenotype can be mainly predicted by age and sex, then there would be no requirement for a cuffless device. Therefore, as proposed by Mukkamala *et al.* [25], reference performance metric values were computed using only patient age and sex. We refer to this model as *baseline reference*.

Equivalent to Method 2, the *baseline reference* was computed using the same 5-fold CV process and an ordinal classification framework. An RF classification model was implemented with $mtry = 2$.

8.7 Evaluation and generalisability of methods

8.7.1 Evaluation metrics

We used the following metrics to evaluate phenotype classification performance: accuracy (Acc), weighted F1 score (wF1), and Cohen's kappa (κ). Acc is the simplest metric, it corresponds to the proportion of correct predictions. However, in situations where there is a class imbalance in

the dataset, Acc can be an unreliable metric. Therefore, metrics that are more robust to these imbalances were additionally reported. The F1 metric for $\mathcal{P}_D$, $F1_{\mathcal{P}_D}$, was computed as:

$$F1_{\mathcal{P}_D} = 2 \cdot \frac{precision_{\mathcal{P}_D} \cdot recall_{\mathcal{P}_D}}{precision_{\mathcal{P}_D} + recall_{\mathcal{P}_D}} \quad (8.8)$$

where $precision_{\mathcal{P}_D}$ is the proportion of patients, classified as 'dippers', that were actually 'dippers'. $recall_{\mathcal{P}_D}$ (also known as true positive rate) is the proportion of 'dippers' that were correctly classified. To compute a single metric across all labels, wF1 was computed by weighting the mean of F1 by the number of true instances for each phenotype label.

$$wF1 = \frac{1}{N} \left(\sum_{\mathcal{P} \in \{\mathcal{P}_D, \mathcal{P}_{ND}, \mathcal{P}_{RD}\}} |\mathcal{P}| \cdot F1_{\mathcal{P}} \right) \quad (8.9)$$

where the $|\cdot|$ operator returns the number of true instances for that phenotype. N is the number of observations. We additionally used a *recall* confusion matrix to summarise the agreement between predicted labels and ground truth, or reference, labels.

κ is also a robust classification performance metric that accounts for imbalanced data, as well as chance agreement [367]. κ is defined as:

$$\kappa = \frac{Acc - p_e}{1 - p_e} \quad (8.10)$$

where p_e is the estimated probability of random agreement, computed as:

$$p_e = \frac{1}{N} \sum_k n_k^{GT} n_k^{est} \quad (8.11)$$

where n_k^{GT} and n_k^{est} denote the number of instances which are labelled as the k^{th} state in the ground-truth (GT) and estimated values respectively. Unlike wF1, κ is a more conservative metric. Values near 1.0 denote almost perfect classification and values close to 0.0 tend towards chance agreement.

We additionally recorded the proportion of patients in the $\mathcal{P}_D$ cohort that were incorrectly classified as $\mathcal{P}_{RD}$, and vice versa. We refer to this metric as the *severe-misclassification rate* (SMR) as it indicates the proportion of patients whose predicted phenotype was 2 steps away on the ordinal scale.

8.7.2 Generalisability of methods

In this chapter, we presented three methods for BP phenotype classification: two signal-based methods (Method 1: 'regularised *individual calibration*', and Method 2: 'calibration-free') and a *baseline reference* method. Method 2 and the *baseline reference* are population-based methods, meaning that the BP phenotypes for a set of patients were classified using a classification model trained on a set of other patients. Method 1, on the other hand, utilised a regularised *individual*

calibration approach in which nocturnal SBP (and therefore subsequently the BP phenotype) for a set of patients was estimated using a mixed-effects regression model trained on data collected only from that set of patients. As a result, when classifying the BP phenotypes for the same set of patients, method 1 has a different training set requirement to method 2 and the *baseline reference*. To ensure a direct comparison of the evaluation metrics, the methods were evaluated simultaneously in a modified CV procedure.

The overall modified CV process is as follows: the cohort of patients, denoted Ψ , was randomly divided into five disjointed subsets of patients: $\psi_1, \psi_2, \psi_3, \psi_4, \psi_5$ where $\Psi = \bigcup_{j \in [1,5]} \psi_j$. For method 2 and the *baseline reference*, for the i^{th} fold of the CV, the $\bigcup_{j \in [1,5] \setminus i} \psi_j$ sets of patients were used to train and tune ordinal classification models which were then evaluated on the ψ_i set of patients⁴. Whereas, for method 1, the data collected between 8am and midnight from the ψ_i set of patients was used to train a mixed-effects regression model to predict nocturnal SBP also in the ψ_i set of patients⁵. As a result, for each fold of the modified CV, phenotype classification was performed on the same ψ_i set of patients across all methods. This modified CV approach, therefore, facilitated a direct comparison of the classification performance metrics, and their distributions, across the cohort.

CV is often repeated with varying random splits in order to reduce bias in data splitting and to report the distributions of the performance metrics [368]. The number of CV repetitions is often determined based on computational limitations. In this work, the modified 5-fold CV was repeated 10 times. Differences in performance metrics across the 10 repetitions of the CV procedure from the models were evaluated for statistical significance by a two-tailed Wilcoxon signed-rank test (see appendix C.1). We adjusted the p-values for multiple comparisons using the Benjamini-Hochberg method (see appendix C.1.2).

8.8 Results

8.8.1 Dataset

Table 8.1 presents the demographic information of the patient cohort. The dataset consists of 715 patients, with a majority of them being male (55.2%). The average length of stay in the ICU was 2.3 days, with a mean Oxford acute severity of illness score (OASIS) score of 26.4, suggesting that, on average, a patient in our cohort had a 2.56% chance of dying in the ICU [348]. The distribution of BP phenotypes among the 715 patients was $\mathcal{P}_D$ (15.8%), $\mathcal{P}_{ND}$ (48.7%), and $\mathcal{P}_{RD}$ (35.5%).

⁴This process is equivalent to a patient-wise 5-fold CV

⁵This process is equivalent to $N_p/5$ patient-wise sampling without replacement

Table 8.1: Demographics of the subset of patients selected from the MIMIC database. Data was recorded during the last 24 hours of a patient’s stay in the ICU. Parameters such as height and weight were not recorded for all patients and so are not presented here. The admission location provides information about the previous location of the patient prior to arriving in the ICU. The reference BP phenotype was computed using DR_{SBP} .

Descriptor	Value
Total Number of patients ³	715
Male	395 (55.2%)
Female	320 (44.8%)
Age (years) ¹	56.0 (15.5)
Total length of stay in the ICU (days) ¹	2.3 (2.2)
OASIS ^{1, 2}	26.4 (7.2)
<i>Ethnicity</i> ³	
White	487 (68%)
Black	83 (12%)
Unknown	75 (10%)
Hispanic	38 (5%)
Asian	32 (4%)
<i>Admission location</i> ³	
Referral	499 (70%)
Emergency room admission	120 (17%)
Transfer from hospital	96 (13%)
<i>Reference BP phenotype</i> ³	
Dipper ($\mathcal{P}_D$)	113 (16%)
Non-dipper ($\mathcal{P}_{ND}$)	348 (49%)
Reverse dipper ($\mathcal{P}_{RD}$)	254 (36%)

¹ Patient-wise mean (standard deviation). ² Oxford Acute Severity of Illness Score (OASIS), a severity of illness score used for predicting patient outcomes [348]. ³ Values given as count (% of total number of patients)

8.8.2 Overall performance

The results of BP phenotype classification using the *baseline reference* and the two proposed methods are presented in Figure 8.4. The median performance statistics, along with their interquartile range (IQR), for each method are reported for the 10 repetitions of the CV procedure. The median *baseline reference* wF1, Acc, κ , and SMR values were: 0.27, 0.28, 0.03, and 21.3% respectively. The proposed methods showed improved performance with respect to the *baseline reference*, with statistically significant improvements across all performance metrics ($p < 0.001$).

Figure 8.5 shows the median and IQR of the *recall* confusion matrices for the best-performing models from Method 1 (LMEM_{PPG+PAT}), Method 2 (RF_{PPG+PAT}), as well as for the *baseline reference*, computed across the 10 repetitions of CV. In general, the two methods classified each phenotype with similar recall. Non-significant differences were observed between the two models for wF1, Acc, and κ . Notably, both models showed poor recall values for estimating the $\mathcal{P}_D$ phenotype

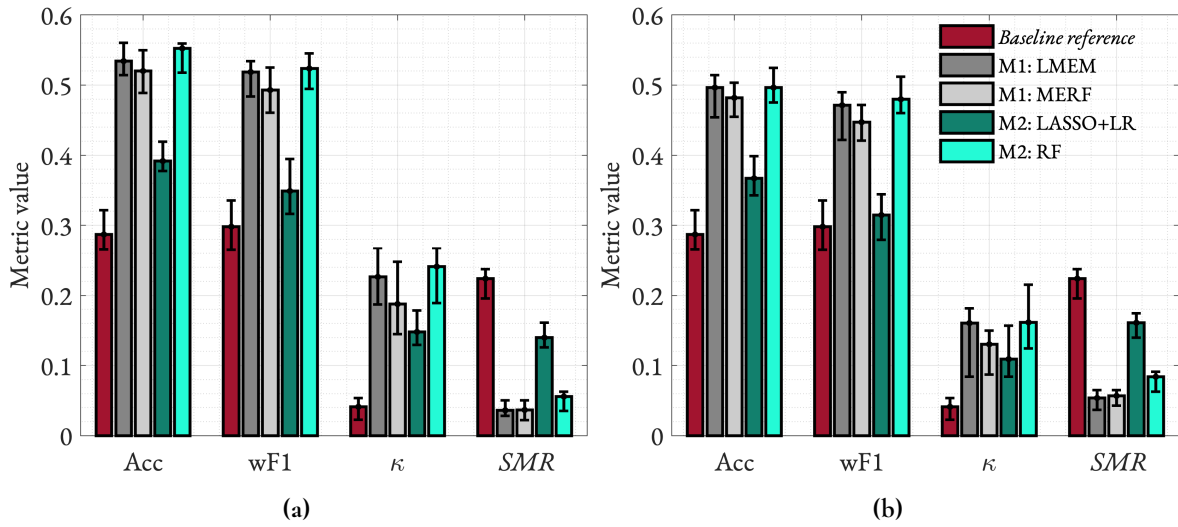


Figure 8.4: Median BP phenotype classification results using the methods proposed in this chapter across the 10 repetitions of the CV. The *baseline reference*, which classified BP phenotypes based on patient age and sex, results are shown in red. The two Method 1 models are shown in grey. The two Method 2 models are shown in green. (a) PAT + PPG_{fts}, (b) PPG_{fts}. Error bars indicate upper and lower quartiles. Note that for all metrics apart from SMR, a higher value indicates better performance.

(recall Method 1 = 15.8%, Method 2 = 12.4%) with the majority of 'dippers' being classified as 'non-dippers' (Method 1 = 62.5%, Method 2 = 61.1%). LMEM_{PPG+PAT} reported a statistically significantly ($p < 0.001$) lower SMR value than RF_{PPG+PAT} and this can be observed by the relatively lower upper-right and lower-left cells of the confusion matrices. RF_{PPG+PAT} reported a higher recall for $\mathcal{P}_{ND}$ than LMEM_{PPG+PAT}, whereas LMEM_{PPG+PAT} was comparatively better at estimating the $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ phenotypes.

8.8.3 Method 1: Regularised *individual calibration*

Method 1 estimated SBP_{Night} ($\hat{SBP}_{Night}$) leading to subsequent estimation of DR_{SBP} ($\hat{DR}_{SBP}$) and BP phenotype. The regression performance metrics for these values are presented in table 8.2 and visualised in figure 8.6. In order to account for the calibration procedure, figure 8.6 shows the deviation (Δ) of the estimated values from the mean reference SBP_{Day} for each patient, colour and marker-coded. For both the PPG_{fts}+PAT and PPG_{fts} feature sets, LMEM outperformed MERF across all regression performance statistics. The absolute difference in mean ρ_p , RMSE, and MAE between LMEM_{PPG+PAT} and MERF_{PPG+PAT} was 0.14 ($p < 0.001$), 0.39 mmHg ($p < 0.001$), and 0.29 mmHg ($p = 0.0076$) respectively. The absolute difference in mean ρ_p , RMSE, and MAE between LMEM_{PPG} and MERF_{PPG} was 0.11 ($p < 0.001$), 0.64 mmHg ($p < 0.001$), and 0.5 mmHg ($p < 0.001$) respectively. However, overall performance was still poor with mean RMSE values greater than 10 mmHg for both LMEM_{PPG+PAT} and LMEM_{PPG}. Despite implementing regularised *individual calibration* through mixed-effects modelling, figure 8.6 shows a large calibration issue whereby

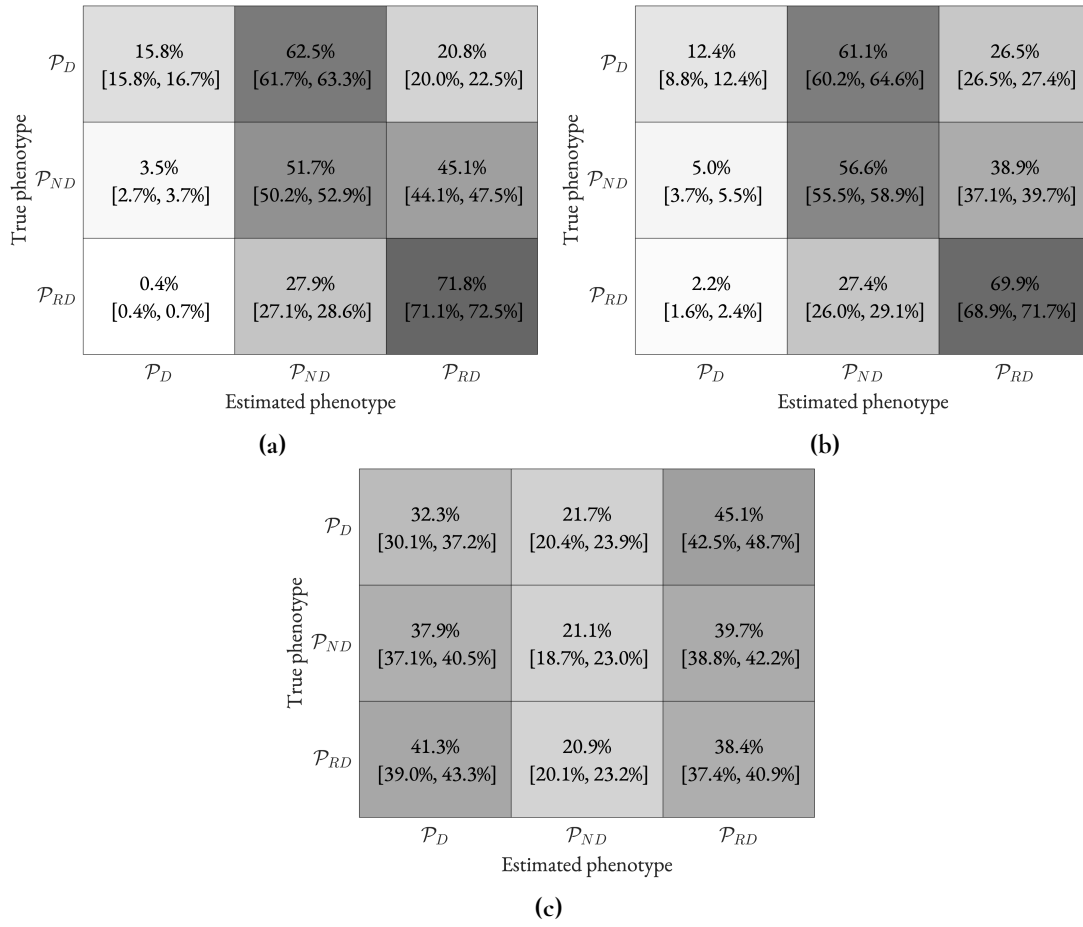


Figure 8.5: Median, upper, and lower quartile phenotype classification *recall* confusion matrices across the 10 repetitions of the CV procedure. (a) Method 1: LMEM_{PPG+PAT}, (b) Method 2: RF_{PPG+PAT}, and (c) baseline reference.

small values of $\Delta\text{SBP}_{\text{Night}}$ were overestimated and large values of $\Delta\text{SBP}_{\text{Night}}$ were underestimated. This calibration issue persists for DR_{SBP} estimation; however, its severity appears to be reduced with the least squares regression line being closer to unity.

Figure 8.7 shows the distributions of fixed + random effects for the LMEM_{PPG+PAT} model. Each subplot shows a distribution of the modelled calibration slopes for each signal, highlighting the regularisation effect of mixed-effects modelling.

For both the PPG_{fts}+PAT and PPG_{fts} feature sets, LMEM and MERF reported similar classification performance statistics, with LMEM reporting marginally better results. However, these differences were not found to be statistically significant. For both models, adding the PAT signal to the feature set was found to significantly improve performance. The absolute difference in median wF1, Acc, κ , and SMR between LMEM_{PPG+PAT} and LMEM_{PPG} was 0.058 ($p < 0.001$), 0.048 ($p < 0.001$), 0.085 ($p < 0.001$), and 1.4% ($p = 0.00013$) respectively.

Table 8.2: Regularised *individual calibration* SBP regression performance statistics of SBP_{Night} and subsequent DR_{SBP} estimation using the models proposed in method 1. Results are given as patient-wise mean (standard deviation). Entries in bold indicate the best performance for that metric. These results are depicted in figure 8.6.

Model name		ρ_p		RMSE		MAE	
SBP _{Night} ¹	LMEM _{PPG+PAT}	0.27	(0.54)	10.73	(5.47)	9.26	(5.01)
	MERF _{PPG+PAT}	0.17	(0.54)	11.11	(5.5)	9.56	(5.04)
	LMEM _{PPG}	0.21	(0.55)	11.28	(5.9)	9.82	(5.49)
	MERF _{PPG}	0.11	(0.54)	11.81	(5.93)	10.25	(5.5)
DR _{SBP} ²	LMEM _{PPG+PAT}	0.60	(0.01)	7.42	(0.07)	5.80	(0.07)
	MERF _{PPG+PAT}	0.51	(0.03)	7.58	(0.19)	5.91	(0.14)
	LMEM _{PPG}	0.49	(0.01)	8.06	(0.09)	6.3	(0.1)
	MERF _{PPG}	0.45	(0.03)	8.21	(0.24)	6.36	(0.17)

¹ Error values given in units of mmHg, ² Error values given in units of %.

8.8.4 Method 2: Calibration-free

RF models significantly outperformed LASSO+LR models for both PPG_{fts}+PAT and PPG_{fts} feature sets ($p < 0.001$ for all performance metrics). Adding PAT to the feature set significantly improved performance, with all corresponding p values less than 0.001. The absolute difference in median wF1, Acc, κ , and SMR between RF_{PPG+PAT} and RF_{PPG} (see figure 8.4) was 0.035, 0.034, 0.061, and 1.3% respectively.

To mitigate against variations in the training data between folds of the repeated CV, overall feature importance was assessed through mean absolute SHAP values and a ranking coefficient. Figure 8.8 illustrates the median (across all repeated CV folds) LASSO+LR_{PPG+PAT} and RF_{PPG+PAT} ranking coefficients for all signals using a spider plot, such that the relative contribution of each of the 7 features derived from each signal can be easily compared. Between the two models, there was a large disparity in the relative importance placed on the features. LASSO+LR_{PPG+PAT} placed a considerably higher emphasis on cosinor amplitudes (A_1 and A_2) than RF_{PPG+PAT}. Higher sex and age demographics ranking coefficients were observed for LASSO+LR_{PPG+PAT} (median ranking coefficients 0.46 and 0.71 respectively) than RF_{PPG+PAT} (median ranking coefficients 0.02 and 0.19 respectively). For RF_{PPG+PAT}, features derived from PAT consistently exhibited high importance values. Across all signals, $\tilde{D}R$ and θ_{Ortho} were found to offer a large contribution to classification performance. $\tilde{D}R_{PAT}$ had the highest feature importance for 74% of all folds across all CV repetitions.

8.9 Discussion

We have introduced two methods for estimating BP phenotypes without the requirement for night-time cuff measurements (as with ABPM), which could be integrated into everyday primary

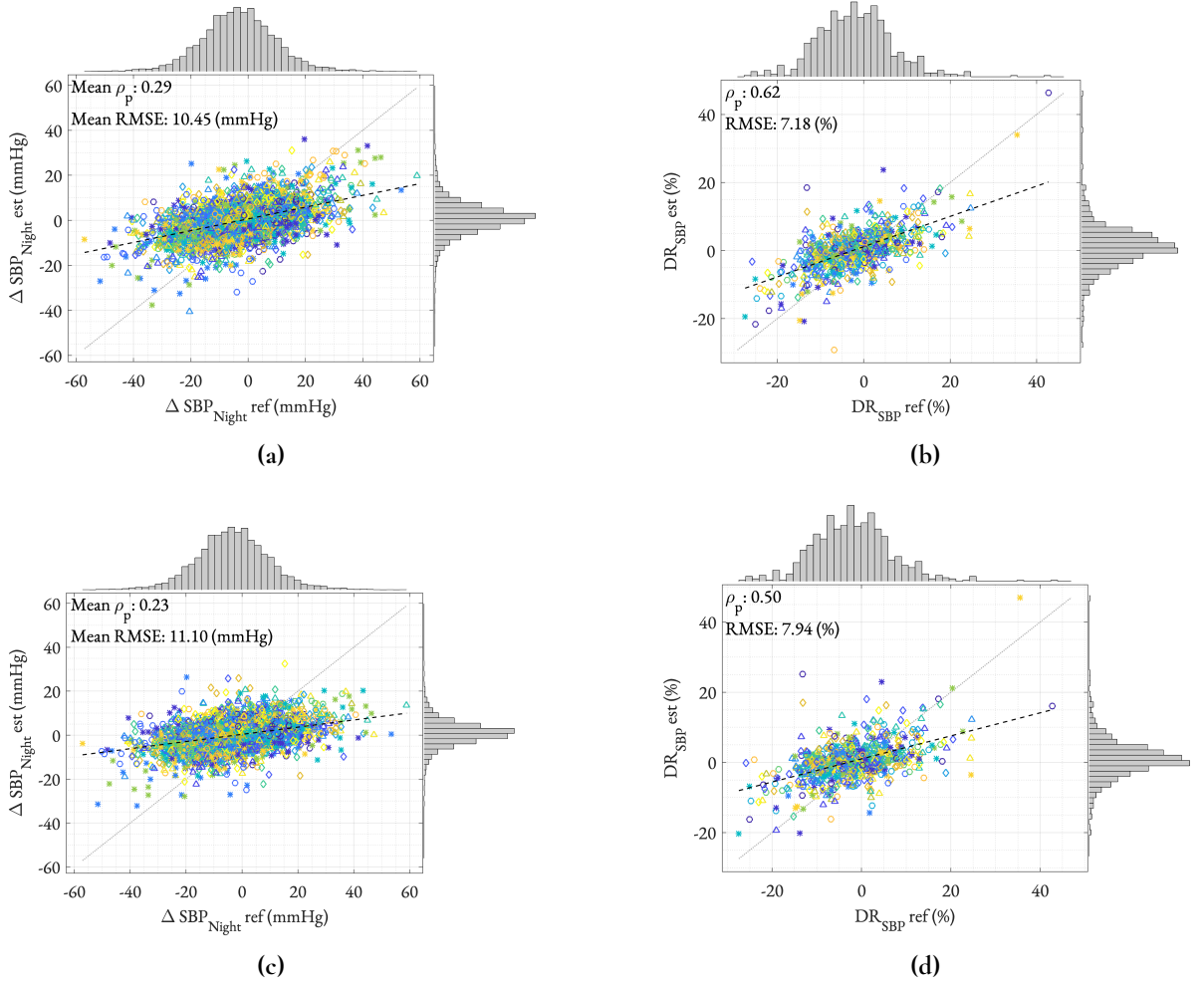


Figure 8.6: Regularised *individual calibration*, regression analysis for estimating SBP_{Night} and subsequent DR_{SBP} using Method 1. (a)-(b) shows $LMEM_{PPG+PAT}$ model performance. (a) SBP_{Night} and the estimated $\hat{SBP}_{Night}$ values. In order to normalise the effect of the calibration procedure, for each patient their mean SBP_{Day} was subtracted from the reference and estimates to report the deviations (Δ). (b) DR_{SBP} and estimated $\hat{DR}_{SBP}$ values. (c)-(d) show $LMEM_{PPG}$ model performance. (c) SBP_{Night} and the estimated $\hat{SBP}_{Night}$ values. (d) DR_{SBP} and estimated $\hat{DR}_{SBP}$ values. Data from individual patients are colour and marker-coded.

care practice. Our methods were compared to a *baseline reference* that relied solely on demographic information. Constructing a dataset of healthy volunteers with a large variation in DR_{SBP} is not feasible practically given the low user acceptability and tolerance of ABPM [23]. We, therefore, conducted a retrospective observational study on a large cohort of ICU patients during their final 24 hours in the unit.

Despite the implementation of class-balancing solutions (random upsampling of minority classes), the low prevalence of the $\mathcal{P}_D$ phenotype (15.8%) in the dataset still had a significant impact on the accuracy of its classification (recall Method 1 = 15.8%, Method 2 = 12.4%). This prevalence is comparable to other hospital datasets (10.5% in [5]), but much lower than the typical

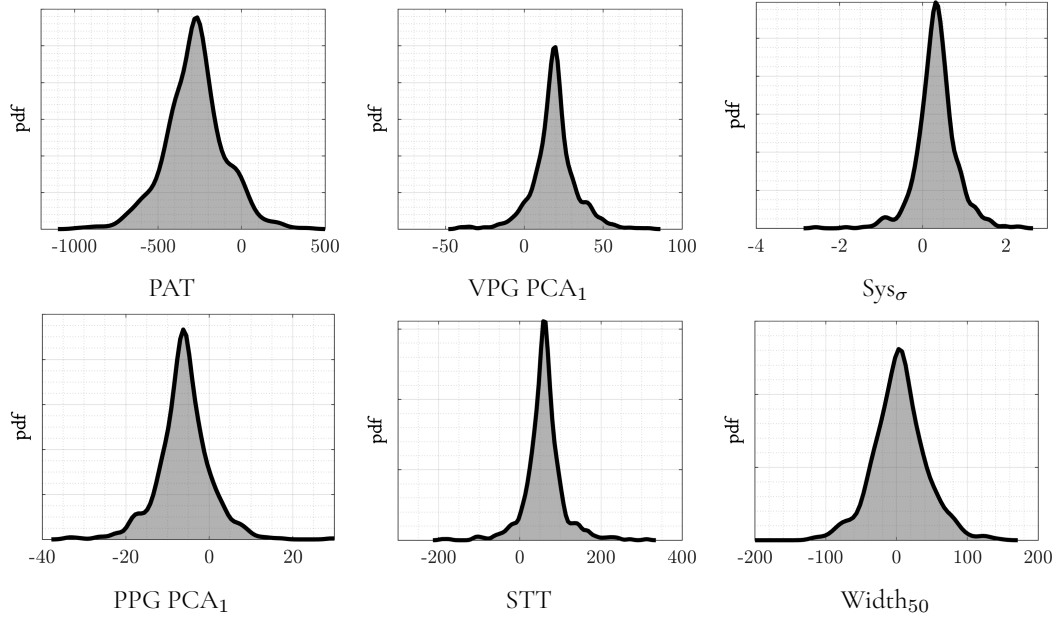


Figure 8.7: Regularised *individual calibration*, visualising the fixed and random effects of the $\text{LMEM}_{\text{PPG}+\text{PAT}}$ model. Each subplot shows the distribution of modelled calibration slopes for each signal. The mean of each distribution indicates the fixed effect, the spread indicates the added patient-wise random effect.

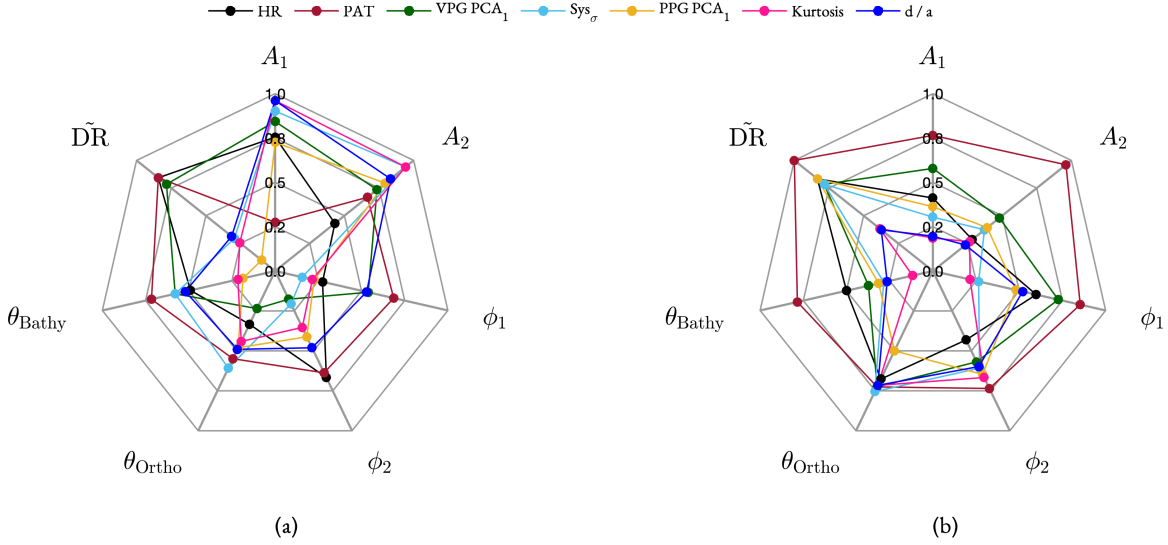


Figure 8.8: Calibration-free median feature ranking coefficient for all signals implemented for method 2. Each signal contributed 7 features to the calibration-free classification method: $[A_1, A_2, \phi_1, \phi_2, \theta_{\text{Bathy}}, \theta_{\text{Ortho}}, \tilde{D}R]$. Values at the edge of the spider plot indicate a high importance for phenotype classification. (a) $\text{LASSO}+\text{LR}_{\text{PPG}+\text{PAT}}$, the median ranking coefficients for sex and age were 0.46 and 0.71 respectively. (b) $\text{RF}_{\text{PPG}+\text{PAT}}$, the median ranking coefficients for sex and age were 0.02 and 0.19 respectively.

prevalence observed in the general population (around 50% [5, 51]). To address this issue, further improvement in the accuracy of $\mathcal{P}_D$ classification could be achieved by either increasing the size of the cohort through the inclusion of new individuals or by implementing cost-sensitive methods that increase the penalty for incorrect 'dipper' classification [366]. Accurately determining an appropriate penalty for misclassification necessitates a careful evaluation of the associated costs. For example, an individual classified as a 'reverse-dipper' may be required to perform ABPM with a validated cuff, or be asked to take anti-hypertensive medication before sleeping [57].

The prevalence of the $\mathcal{P}_{RD}$ phenotype (35.5%) was much higher than has been previously observed in the general population (around 10% in [51]). This high prevalence of $\mathcal{P}_{RD}$ phenotype facilitated the development of effective methods for its classification, which would otherwise have been a challenging task due to its rarity. This was reflected in the relatively high recall values reported for $\mathcal{P}_{RD}$ prediction (Method 1 = 71.8%, Method 2 = 69.9%). The $\mathcal{P}_{RD}$ phenotype has been linked with an increased risk for various cardiovascular and neurodegenerative diseases, regardless of daytime blood pressure values [15, 55, 61]. Therefore, early detection of this phenotype may be possible using the methods proposed in this chapter.

8.9.1 Method 1: Regularised *individual calibration*

Method 1 used data recorded between 8am to midnight to train mixed-effects regression models for regularised individual cuffless $\text{SBP}_{\text{Night}}$ estimation. This approach addresses a major limitation of ABPM, which requires cuff-based BP measurements to be taken during the night-time causing sleep disruption and low user compliance [23, 111]. A more favourable alternative to ABPM is HBPM, in which a user manually records multiple BP measurements during the daytime, including one reading in the morning and one reading before sleep, over a period of between 4 to 7 days [102, 103, 111]. Combining the regularised *individual calibration* method with HBPM in a primary care setting could facilitate an assessment of BP phenotypes in the general population, without the need for nocturnal cuff-based measurements.

The individualised relationship between BP and PAT or PPG_{fts} should be taken into account to estimate night-time BP accurately. One approach towards this would be to build individual regression models for each patient. However, this approach would require a large amount of longitudinal data over a clinically useful range of BP, which was not available in the current dataset. We, therefore, implemented mixed-effects regression (LMEM and MERF) modelling to predict BP values across the cohort while accounting for the inter-dependency of multiple observations of the same patient. Appendix G provides additional analysis of the regularisation

effect of mixed-effects modelling, and how this is advantageous when recording data in a noisy environment such as the ICU, over individual modelling.

The computational cost of the LMEM maximum likelihood algorithm scales to the quadratic of the number of random-effect covariance terms [369]. Therefore, we decided to limit the number of signals incorporated into the feature set based on the population-level analysis of the previous chapter. This signal selection process largely favoured signals that exhibited strong linear relationships with SBP. Although the relationship between PPG_{fts} and BP may be non-linear, as indicated in chapter 6, the decision to use LMEM and MERF, which both assume linear random-effect contributions, was still deemed appropriate. However, as reported by Rutten *et al.* [370] and Gottard *et al.* [371], when the relationship between predictors and target is largely linear, LMEM outperforms MERF. As a result, for both the PPG_{fts} and $PPG_{fts}+PAT$ feature sets, LMEM was found to have slightly better (but not statistically significant) classification performance in comparison to MERF (see figure 8.4).

Despite implementing an *individual calibration* strategy, the regression performance of SBP_{Night} estimation was clearly limited by inadequate calibration (see figure 8.6). The $LMEM_{PPG+PAT}$ model achieved weak correlations (median $\rho_p = 0.29$) and large errors (median RMSE = 10.45 mmHg) with respect to the reference SBP_{Night} values. Assessing the fixed and random effect distributions (see figure 8.7) provides an interpretation for this. All distributions contained both positive and negative values, highlighting uncertainty in the calibration slopes. This uncertainty limits the effectiveness of the random effect regularisation across the cohort (i.e. allowing a, non-physiological, positive PAT calibration slope [31]). As a result, *individual calibration* could not be appropriately handled and so any dependence between observations was maintained. As MERF handles random effects in the same manner as LMEM, this issue persisted for both mixed-effects models.

The issue of poor calibration is very likely to be the result of comparatively low BP variability during the day [48], coupled with errors in signal recording (such as motion artefacts) in the ICU that cannot be appropriately accounted for in a *post-hoc* assessment of the MIMIC database. Indeed, analysis of $CI_{95}^{\theta_{Orth}}$ and $CI_{95}^{\theta_{Bath}}$ in the previous chapter suggested that at the patient level, the selected PPG_{fts} showed significantly greater inter-patient variability than BP, possibly driven by a higher susceptibility to motion artefacts (see figure 7.8). A single day of measurements may, therefore, not provide sufficient data for appropriate *individual calibration*.

The use of devices that record user motion, such as accelerometers, in primary care home recordings, may help in the filtering of noisy data or outliers. Additionally, multi-day measurements (as is typically required when monitoring blood pressure at home [102]) can enable the creation of a larger, longitudinal dataset for each individual, which may enhance performance.

Unlike Method 2, regularised *individual calibration* facilitates assessing DR_{SBP} as a continuous variable, as opposed to discretisation into phenotypes based on arbitrary thresholds, as suggested by multiple studies [372, 55, 373]. The $LMEM_{PPG+PAT}$ model recorded comparatively better results in estimating DR_{SBP} (median $\rho_p = 0.62$, median RMSE = 7.18%), and the limitation of poor calibration may be reduced for this metric. When compared to other methods in the literature, Radha *et al.* [187] reported a comparable mean $\rho_p = 0.69$ value for dip estimation using a long short-term memory (LSTM) neural network with only PPG_{fts} as inputs. This further indicates that the ECG or PPG-derived signals may be better suited for an assessment of macro changes in BP, as opposed to beat-by-beat, or even hour-by-hour variations.

8.9.2 Method 2: Calibration free

Method 2 is a calibration-free approach that estimates a patient's BP phenotype based on features derived from the 24-hour profiles of multiple signals. The aim was to provide a convenient and efficient way of determining an individual's BP phenotype without the need for reference BP measurements. The double-cosinor phases (in particular peak timings) of all signals were a large contributing factor in RF model estimations (see figure 8.8). As these phase features are circular parameters (they exist in the range $[0, 2\pi]$), the relative performance gains of the non-linear RF model over the linear LASSO+OLS becomes clear. Relying on signal phases rather than amplitudes offers several benefits, including the avoidance of the need for individual-specific calibration and increased robustness against signal noise. The double-cosinor amplitudes of PPG_{fts} were found to offer little contribution to $RF_{PPG+PAT}$ model performance. The relative amplitude of diurnal to nocturnal differences ($\tilde{DR}$) was of high importance across all signals. PAT consistently reported higher feature importance than individual PPG_{fts} , and $RF_{PPG+PAT}$ reported significantly better performance metrics than RF_{PPG} .

Armitage *et al.* [5] previously suggested that 'reverse-dippers' tended to be older and more frequently male. Despite this, in our results, sex and age demographics had the lowest and third-lowest $RF_{PPG+PAT}$ feature ranking coefficients respectively. This suggests that the influence of age and sex on phenotypes is not a prominent factor in our ICU data, and helps to explain the poor performance of the *baseline reference* method. Factors such as ICU admission and medication may have proved to be more important than age and sex, which are more likely to be relevant in a community study.

With a median SMR of 6.3%, this calibration-free approach could offer a practical and cost-effective initial evaluation of dipping status in the wider population. The potential for further

improvement may be derived from a more extensive, longitudinal dataset for each individual. Portoluppi *et al.* [374] suggests that cosinor modelling is consistent across multiple days, indicating the reproducibility of daily routines. By modelling the variations in signals over multiple days, the signal-to-noise ratio of each signal can be improved, leading to more accurate cosinor feature estimates.

8.9.3 Limitations

The work presented in this chapter has some limitations that are worth discussing.

Firstly, as detailed in the previous chapter, the ICU is not an ideal dataset for an analysis of circadian phenotypes. Patients in the ICU exhibit distinct physiological responses to critical illness and medication, which may not align with typical primary care or ambulatory environments. The noisy nature of the data and the presence of medical staff may have influenced the results, raising questions about the generalisability of the findings to uncontrolled, real-world settings. Given the focus on patient-wise analysis conducted in this chapter, this limitation is likely more significant than in the previous chapter.

The estimation of patient-wise BP phenotypes presents a challenge due to the subjective nature of the ground-truth phenotype labelling. This process is based on the definition of arbitrary thresholds of DR_{SBP} , which inherently includes a degree of uncertainty. Measurements of BP using sphygmomanometer cuffs have known limitations depending on posture, cuff-inflation hypertension, and cuff size [105, 277]. Slight errors in the BP measurements can result in a patient being mislabelled into a different phenotype group. As a result, it was expected that the confusion matrices shown in figure 8.5 would reveal a significant number of misclassifications.

In order to mitigate against this occurrence, we selected patients for analysis whose 24-hour SBP time series passed the double-cosinor F-test for the presence of a rhythm at the $p < 0.05$ level. However, it is crucial to recognise that this selection approach introduces specific biases and assumptions into the study. The selection of patients with SBP data that fits a double-cosinor model assumes that these patients have relatively regular circadian rhythms in their BP. This assumption may not hold for all patients, as individuals can exhibit varying degrees of rhythm regularity or irregularity. Consequently, the study may inadvertently favour patients with more predictable BP patterns, potentially excluding those with atypical or irregular circadian rhythms.

Finally, a critical limitation is the challenge of translating the developed methods from the ICU to an at-home monitoring environment. The controlled and supervised nature of data collection in the ICU differs significantly from the independence and variability of home-based monitoring. Factors such as patient movement, variable adherence to measurement schedules, and different signal qualities in home settings could substantially impact the accuracy and applicability of the methods.

8.10 Conclusion

In this chapter, we developed two methods for classifying an individual's BP phenotype using PAT or PPG_{fts}. Both methods are suitable for integration into current primary care practices. We found that the two methods showed similar classification performance, with non-significant differences observed for Acc, wF1, and κ . The regularised *individual calibration* method was able to estimate DR_{SBP} as a continuous variable, allowing for a more nuanced interpretation of an individual's BP patterns. However, we encountered challenges in mixed-effects modelling due to the presence of noise in the dataset that couldn't be accounted for when performing a *post-hoc* assessment. The calibration-free method used circadian features of the PPG_{fts} or PAT signals. This calibration-free approach may provide an unobtrusive and cost-effective assessment of dipping status in the wider population. Adding PAT to the feature set could result in improved performance for phenotype identification, but requires the additional recording of an ECG signal.

This assessment was performed on a cohort of patients close to recovery in the ICU. In the following chapter, we extend this approach to a population of healthy individuals in an ambulatory setting.

9

Towards cuffless estimation of ambulatory blood pressure

Contents

9.1	Introduction	213
9.2	Methods overview	214
9.3	Activity detection	215
9.4	Ensemble averaged pulse arrival time and photoplethysmogram features	216
9.4.1	Ensemble averaged pulse arrival time	217
9.4.2	Ensemble averaged photoplethysmogram features	218
9.5	Establishing a reference blood pressure signal	220
9.5.1	Computing a reference heart rate	220
9.5.2	Heart rate-based removal of poor-quality blood pressure values	221
9.5.3	Double-cosinor model	223
9.6	Estimating Day7 blood pressure	224
9.6.1	<i>Individual calibration</i>	225
9.6.2	<i>Hybrid calibration</i>	226
9.6.3	Regression models	227
9.6.4	<i>Baseline references</i>	228
9.7	Evaluation and generalisability of models	228
9.7.1	Shapley additive explanation values feature importance and feature stability	228
9.7.2	Model complexity	229
9.8	Results	229
9.8.1	Data recording in a free-living context	229
9.8.2	Model performance	230
9.8.3	Participant-level results	233
9.8.4	Feature importance	234
9.9	Discussion	236
9.9.1	Model performance	239
9.9.2	Feature importance	242
9.9.3	Limitations	243
9.10	Conclusion	244

9.1 Introduction

To address the increasing global burden of CVD, there is a pressing need for BP monitoring in at-home environments [5, 1]. In this chapter, we expand upon the research conducted in this thesis by exploring the estimation of ambulatory BP in healthy individuals seven days after their initial calibration. Specifically, we assess the performance of two calibration strategies: *individual*

calibration and *hybrid calibration*. These strategies aim to provide accurate assessments of both absolute BP values and relative BP changes.

9.2 Methods overview

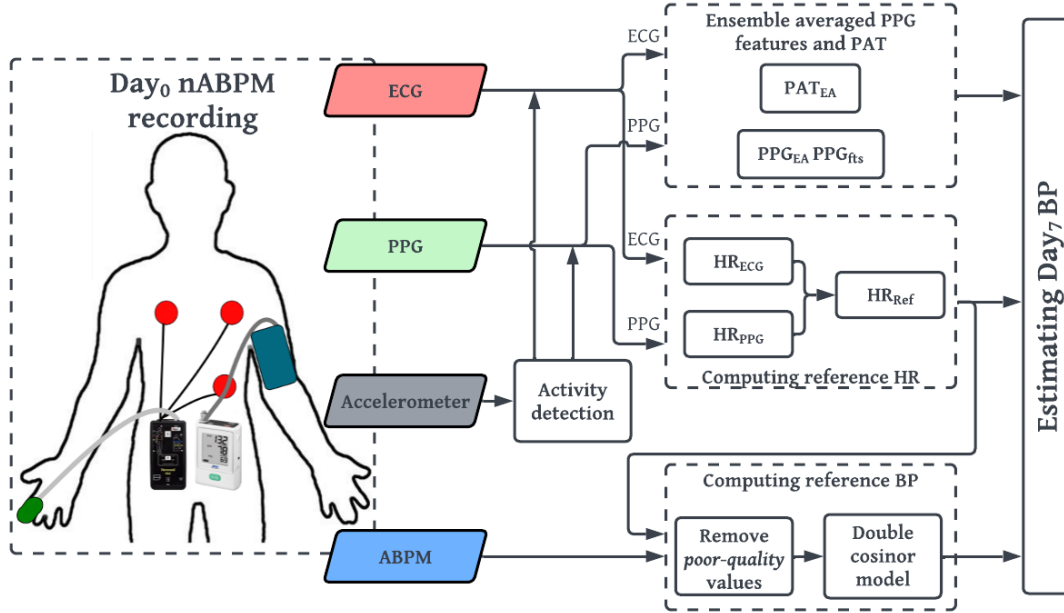


Figure 9.1: Schematic of the methodology used to estimate Day₇ BP using data collected during Day₀ in the nABPM study. The accelerometer data was employed to detect periods of activity and remove them from the analysis of the ECG and PPG signals. These signals were then used to compute reference HR (HR_{Ref}) values and ensemble-averaged PAT and PPG_{fts} . Two ML regression models, regularised ordinary least squares and random forest, were trained on the data collected during Day₀ to estimate the Day₇ BP from the PAT and PPG_{fts} signals.

Data was gathered from the 13 participants enrolled in the nABPM study, presented in section 3.5. This study consisted of two 24-hour sessions, spaced one week apart, during which participants carried out their normal daily activities while physiological data was recorded. During the two 24-hour sessions, PPG, ECG, and accelerometer signals were continuously recorded using a portable physiological monitoring device (Stowood, Oxford, UK). Reference BP measurements were recorded using an ABPM device (A&D Instruments, Tokyo, Japan). The first session, referenced Day₀, and the second session, referenced Day₇, each began with a 2-minute calibration period during which all data, including a reference BP measurement, was recorded with the participants at rest. During the day, BP was recorded every 30 minutes. During the night-time, BP was recorded hourly. Participants reported their sleep times and sleep quality at the end of each session using the sleep questionnaire provided in Appendix B.

Figure 9.1 provides an overview of the methodology used, which was developed using only data collected during Day₀. In order to mitigate against motion artefacts, the accelerometer signal

was used to detect periods of sustained activity. During these active periods measurements of the PPG and ECG signals were labelled as being of poor quality. Subsequently, we established ensemble-averaged estimates of PAT and PPG_{fts} during all inactive periods (see section 9.4 for further details). A reference HR (HR_{Ref}) signal was computed by averaging the heart rate values extracted from the ECG and PPG waveforms. To record the availability of good-quality HR_{Ref}, PAT, and PPG_{fts} data across all sessions.

Disagreements between HR_{Ref} and the HR value recorded by the ABPM device (HR_{Cuff}) enabled the removal of poor-quality BP values. The resulting good-quality ABPM data was smoothed using a double-cosinor model to attenuate high-frequency variations.

Finally, we trained linear (regularised ordinary least squares, OLS) and non-linear (random forest, RF) ML regression models implementing both *individual calibration* (personalised regression) and *hybrid calibration* (leave-one-out regression) strategies to estimate a participant's Day₇ BP ($\hat{BP}$) using only data recorded during Day₀. Additionally, we compared estimates of DR_{SBP} ($\hat{DR}_{SBP}$) to the reference values computed from the ABPM. We assessed the importance of each feature, and their stability, from Day₀ to Day₇, using SHAP values.

9.3 Activity detection

In uncontrolled settings, such as individuals performing their normal daily tasks, recordings of physiological signals may be heavily corrupted by motion artefacts [375]. To mitigate the impact of noise resulting from motion artefacts, we utilised the motion signal recorded by a three-axis accelerometer located in the Stowood device to detect periods of sustained activity. Stowood's proprietary software computed a single vector magnitude time series with arbitrary units (a.u) from the three-axis accelerometer signal. The original motion signal exhibited short sharp bursts of unknown origin. Consequently, we processed this signal using a 10-second running average mean to avoid flagging these bursts whilst still flagging periods of sustained activity. The result of this smoothing is shown in figure 9.2(a) for a typical 24-hour recording.

We located periods of significant participant activity through a threshold (γ_{acc}) of the smoothed motion signal. Figure 9.3 shows the cumulative distribution function (CDF) of the smoothed motion signal grouped into periods of self-reported sleep, and wakefulness during Day₀ recordings. Setting the value of γ_{acc} involved a trade-off; at high values, periods of motion would not be adequately detected and removed from analysis, and at low values, a significant proportion of each session would be discarded from analysis because of perceived activity. With these factors in mind, based on where the CDF of data collected during sleep reached a plateau in figure 9.3,

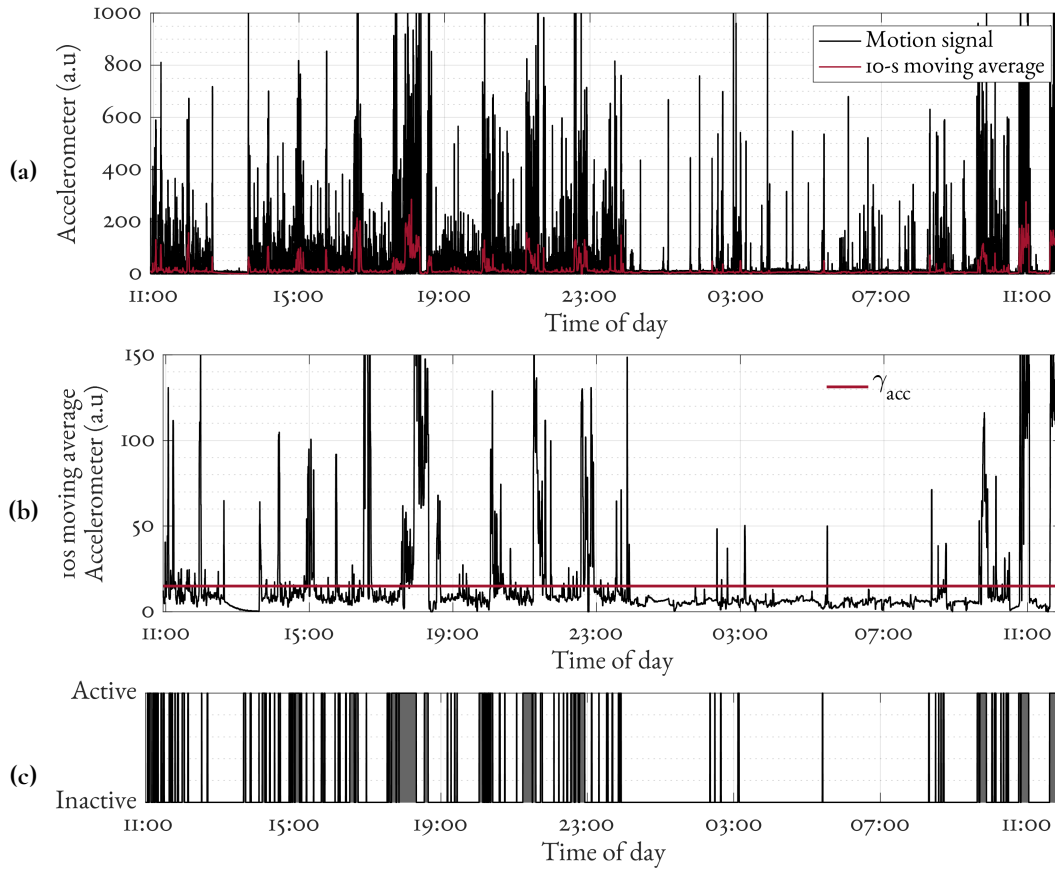


Figure 9.2: Example of detecting periods of activity in the dataset. (a) Original accelerometer signal recorded by the Stowood device. This signal was processed by a 10-second moving average filter to reduce the impact of short sharp bursts and to indicate periods of sustained activity. (b) The 10-second moving average accelerometer signal and threshold γ_{acc} . (c) Periods of activity detected by the accelerometer signal crossing γ_{acc} . Periods labelled as 'Active' were removed from analysis.

we set γ_{acc} to 15. Figure 9.2 provides an example of the activity detection results. Periods during which the 10-second moving average motion signal crossed the γ_{acc} threshold were labelled as 'Active'. The majority of these periods occurred during the day with short periods of activity resulting from movement during sleep.

9.4 Ensemble averaged pulse arrival time and photoplethysmogram features

We computed an approximation of the underlying PAT and PPG_{fts} signals through an ensemble-averaged technique. We followed the steps outlined in section 5.4 to preprocess the PPG and ECG waveforms, detect their respective cardiac beats and assess their signal quality (SQI_{PPG} and SQI_{ECG} respectively). Note that the signal quality assessment included the combined SQI which set both SQI_{PPG} and SQI_{ECG} to zero when the estimated heart rates in the two signals differed by more than 5 beats/min. Additionally, during any detected 'Active' periods, SQI_{ECG} and SQI_{PPG} were set to zero.

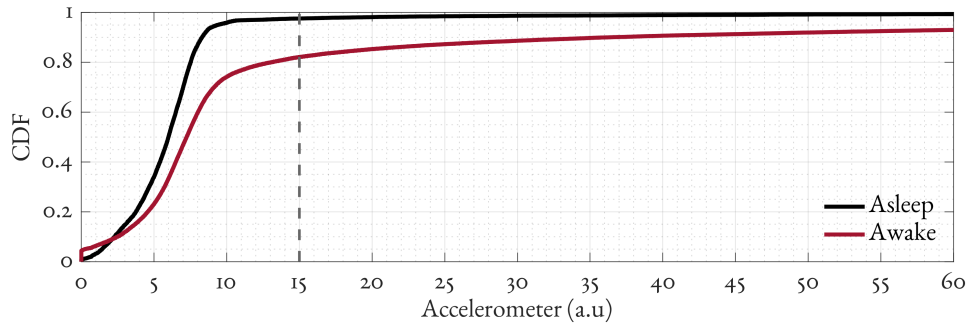


Figure 9.3: Determining γ_{acc} as a threshold that appropriately separates the accelerometer values recorded during self-reported periods of sleep to those recorded during the day. As a result for this analysis γ_{acc} was set to 15.

9.4.1 Ensemble averaged pulse arrival time

We implemented a three-stage strategy to compute an ensemble-averaged estimate of PAT (PAT_{EA}). We first computed the PPG_{R-R} waveform as the average of the PPG signal segmented between two consecutive R-peaks in the ECG signal within a 60-second sliding window with a step size of 30 seconds. At the end of each stage, each PPG_{R-R} waveform was included only if it passed that stage's criterion (see below). Subsequently, the valid PPG_{R-R} waveforms were normalised to have unit amplitude and were averaged, with the length of the averaged waveform set to be the median R-R interval within the window. PAT_{EA} was finally computed from the ensemble-averaged PPG_{R-R} waveforms using the intersecting tangents method [280].

The stages involved in computing PAT_{EA} are outlined in figure 9.4. For the first stage, a PPG_{R-R} waveform was removed from analysis if any of the following criteria were met:

- If either of the R-peaks defining PPG_{R-R} had a corresponding SQI_{ECG} value of 0.
- If within PPG_{R-R} , the number of previously detected PPG onsets was not one (see figure 9.5 for two examples where this criterion was not met in the cohort).
- If the single PPG beat within PPG_{R-R} had a corresponding SQI_{PPG} value of 0.

In the second stage, the ensemble average was computed using valid PPG_{R-R} waveforms (from the first stage) whose amplitude was within the mean $\pm$ three standard deviations for that window. Finally, in the third stage, the linear correlation (ρ_p) of each PPG_{R-R} waveform with the ensemble-averaged PPG_{R-R} waveform was computed, and any waveforms with $\rho_p < 0.8$ were discarded. The resulting SQI of the PAT_{EA} for each 60-second window, w , ($SQI_{PAT}^{EA}(w)$) was computed as:

$$SQI_{PAT}^{EA}(w) = \frac{1}{N_w} \sum_{j \in \Lambda} \min(SQI_{ECG}(j), SQI_{PPG}(k_j)) \quad (9.1)$$

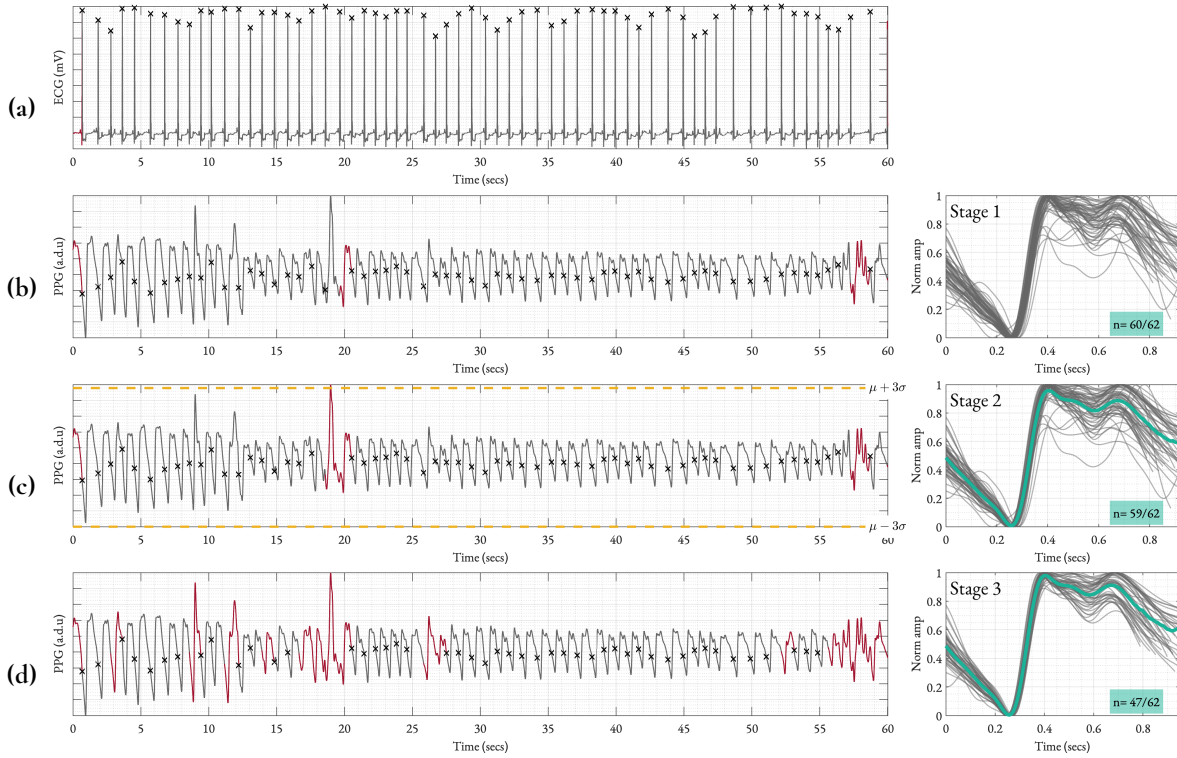


Figure 9.4: Overview of the three stages used to compute PAT_{EA} and SQI_{PAT}^{EA} . At each stage, beats that did not pass the stage criterion are shown in red and the resulting ensemble averaged PPG_{R-R} for that stage is shown in the right-hand column. n indicates the fraction of PPG_{R-R} in the window that met the stage criteria up to that stage. (a) ECG signal with detected R-peaks. (b) Stage 1: PPG_{R-R} waveforms with corresponding SQIs equal to 0 were removed. (c) Stage 2: PPG_{R-R} waveforms that exceed the mean $\pm$ three standard deviations for that window were removed. (d) Stage 3: PPG_{R-R} waveforms that have a $\rho_p < 0.8$ with the ensemble average were removed.

where N_w is the total number of ECG beats initially in w (i.e. prior to stage 1), Λ is the subset of ECG beats that were not discarded from analysis post stage 3 (i.e. defining the onsets of PPG_{R-R}), k_j is the index of the PPG beat succeeding j (i.e. the index of the PPG beat within PPG_{R-R}).

Outlier values of PAT_{EA} were subsequently detected within a 30-minute window with a 25-minute step (see section 5.4.1.1). The SQI_{PAT}^{EA} for any detected outliers was set to zero. Subsequently, Kalman filtering was applied to reduce the effect of transient artefacts caused by noise. Figure 9.6 visually compares PAT computed on a beat-by-beat basis (computed using the steps outlined in section 5.4) to PAT_{EA} for a typical participant in the study. PAT_{EA} represents a smooth approximation of the overall trend of PAT .

9.4.2 Ensemble averaged photoplethysmogram features

We computed the ensemble-averaged PPG waveform (PPG_{EA}) within a 60-second window with a step size of 30-seconds, following the methods described in section 7.4. With the exception of the systolic peak (S), which was detected as the maximum of the PPG_{EA} , we used the criteria

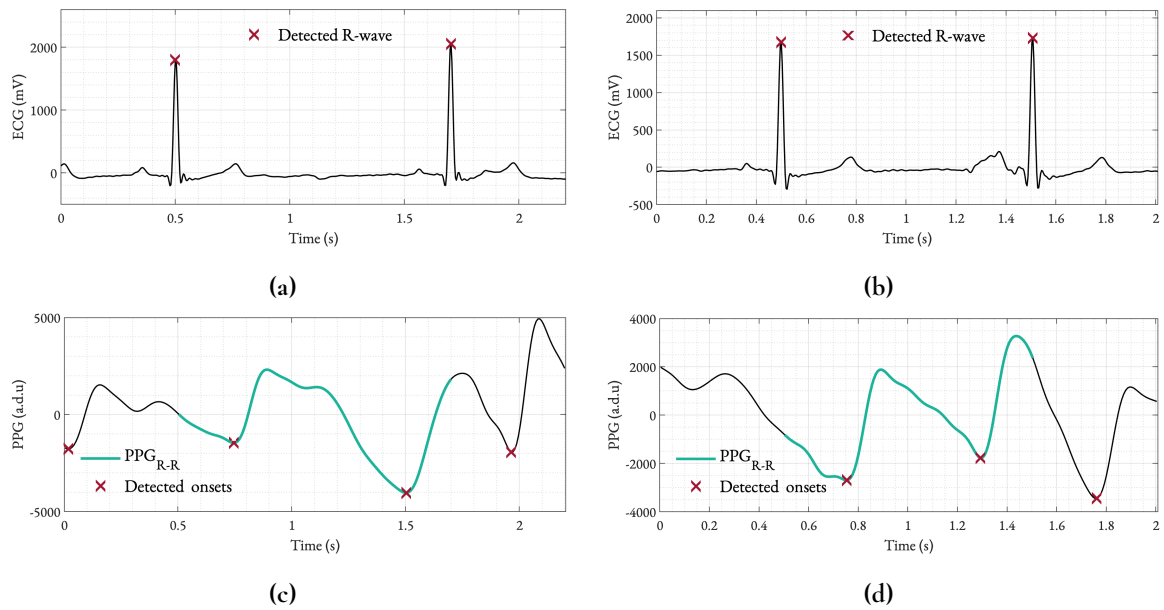


Figure 9.5: Examples of two PPG_{R-R} waveforms where the number of previously detected PPG onsets was not one. In both of these examples, movement of the index finger introduced an artefact into the PPG signal, leading to a misinterpretation as a pulse. (a) & (b): Example ECG waveforms whose R-waves define the start and end of the PPG_{R-R}. (c) & (d): Corresponding PPG_{R-R} and the onsets previously detected using the bSSF algorithm.

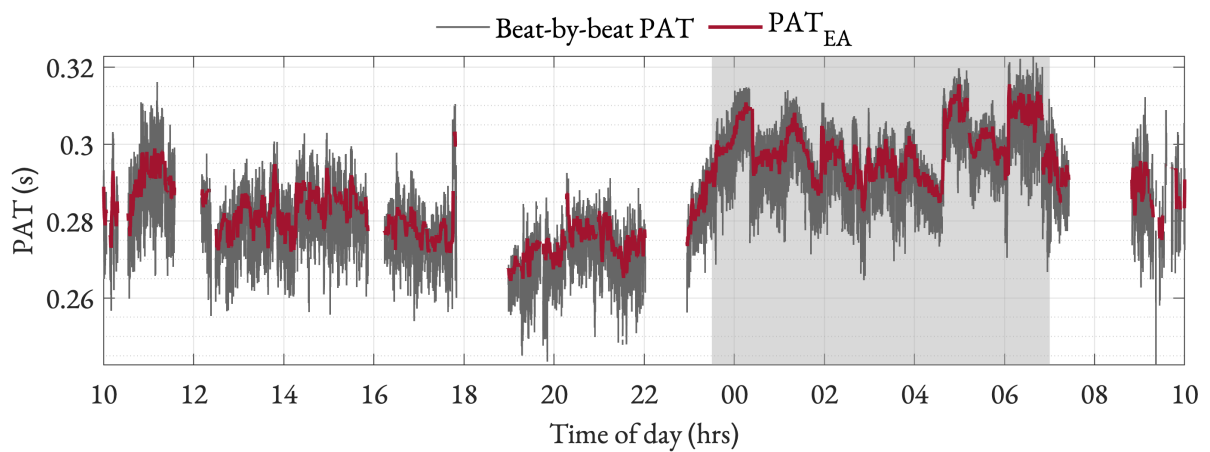


Figure 9.6: Visual comparison of PAT computed on a beat-by-beat basis (computed using the steps outlined in section 5.4) to PAT_{EA} for a typical participant in the study. PAT_{EA} represents a smooth approximation of the overall trend of the PAT signal. Only data corresponding to an SQI=0 is shown. The shaded area indicates night-time.

outlined in section 6.3.1 to detect the characteristic fiducial points of the PPG_{EA} . The 33 PPG_{fts} defined in table 7.1 were extracted from each PPG_{EA} . We similarly processed the PPG_{fts} using outlier detection to remove notable deviations in values and a Kalman filter to reduce the effect of transient artefacts caused by noise.

9.5 Establishing a reference blood pressure signal

Two forms of noise can corrupt BP data recorded by an ABPM device. Firstly, there is a dependence on the precision of the device itself. The TM-2441 ABPM used in the nABPM study has been validated against the ISO 81060-2: 2013 protocol. This protocol requires that the mean $\pm$ standard deviation of error must be $< 5 \pm 8\text{mmHg}$ when compared to a reference mercury sphygmomanometer operated by a trained professional [376, 242]. Secondly, data recorded by an ABPM device also exhibit non-stationary noise, typically caused by movement. For example, often ABPM data will exhibit fewer high-frequency fluctuations during the night than during the day [85]. Training ML regression models in the presence of significant ground truth noise can lead to overfitted models that assume this variation represents the structure of all data. Models trained in such circumstances will often not generalise well to new data [377]. To address this issue, we first detected and removed erroneous BP values, and subsequently smoothed the resulting time series in order to attenuate high-frequency components.

As the blood pressure cuff deflates, the oscillations measured in the sensing cuff are utilised by the ABPM device to report an estimate of HR (HR_{Cuff}). To indicate the presence of significant measurement uncertainty during each cuff deflation, we compared the HR_{Cuff} values to the reference HR signal (HR_{Ref}) computed from the ECG and PPG signals.

9.5.1 Computing a reference heart rate

For each session, HR_{Ref} was computed by calculating the heart rates in the ECG and PPG signals and averaging the results. HR values were computed for each signal separately (labelled HR_{ECG} and HR_{PPG} respectively) using the average of the instantaneous heart rates (IHR) within a sliding window, w , of length 30 seconds and step size of 10 seconds. To compute IHR, the time delay between consecutive R-peaks was used for the ECG signal, while the time delay between consecutive beat onsets was used for the PPG signal. Only beats of good quality (defined as an SQI > 0 for both signals) were included in the analysis. The agreement between HR_{ECG} and HR_{PPG} computed across all Day₀ recordings is shown in figure 9.7 via a Bland-Altman and scatter plot. The two

HR measurements showed very strong agreement¹ with lower dispersion at low HR values, most probably because of the good quality of both signals during sleep.

For each window, HR_{Ref} and its corresponding signal quality index (SQI_{HR}) were computed as the average of the respective values from HR_{ECG} and HR_{PPG} .

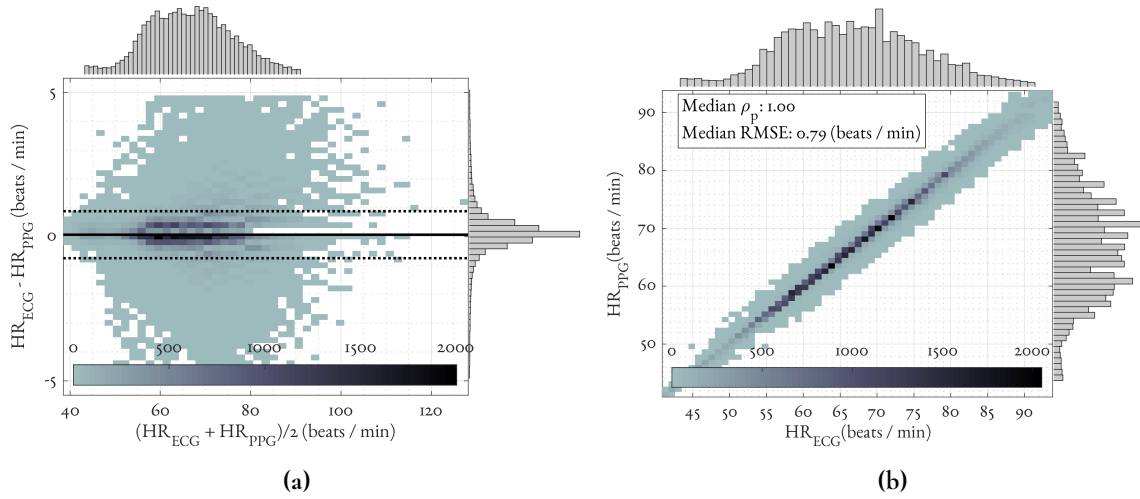


Figure 9.7: Agreement between HR_{ECG} and HR_{PPG} computed across all Day₀ recordings. (a) Bland-Altman plot (b) Scatter plot. The colour of the values indicates the density of the data. HR_{Ref} was computed as the average of these signals.

9.5.2 Heart rate-based removal of poor-quality blood pressure values

Poor-quality BP values were detected based on disagreements between HR_{Ref} and HR_{Cuff} . For each HR_{Cuff} observation, we averaged the HR_{Ref} time series within a window of size w_{HR} centred on the HR_{Cuff} measurement. The % error for each measurement was computed as:

$$\% \text{ error} = \frac{|HR_{Cuff} - HR_{Ref}|}{HR_{Ref}} \cdot 100\% \quad (9.2)$$

if % error exceeded 10% (the ANSI/AAMI maximum HR measurement % error for cardiac monitors [270]), then it was assumed that there was significant noise in the measurement and it was considered to be of poor quality. Additionally, if the averaged SQI_{HR} within the window w_{HR} was less than a threshold, γ_{SQI}^{HR} , then it was assumed that there was significant noise present during data recording and so the measurement was considered to be of poor quality.

To determine the optimum values of the hyperparameters w_{HR} and γ_{SQI}^{HR} , a sensitivity analysis was performed to evaluate their impact on the RMSE agreement between the good-quality HR_{Ref} and HR_{Cuff} values, as well as the number of windows labelled as poor-quality, across all Day₀ sessions.

¹Note that previously, periods with significant deviation in the heart rate between the two signals were removed from analysis, and so a strong agreement was expected.

The goal was to minimise the RMSE, suggesting a close agreement between the two measurements, while also minimising the number of detected poor-quality windows, to achieve a high specificity.

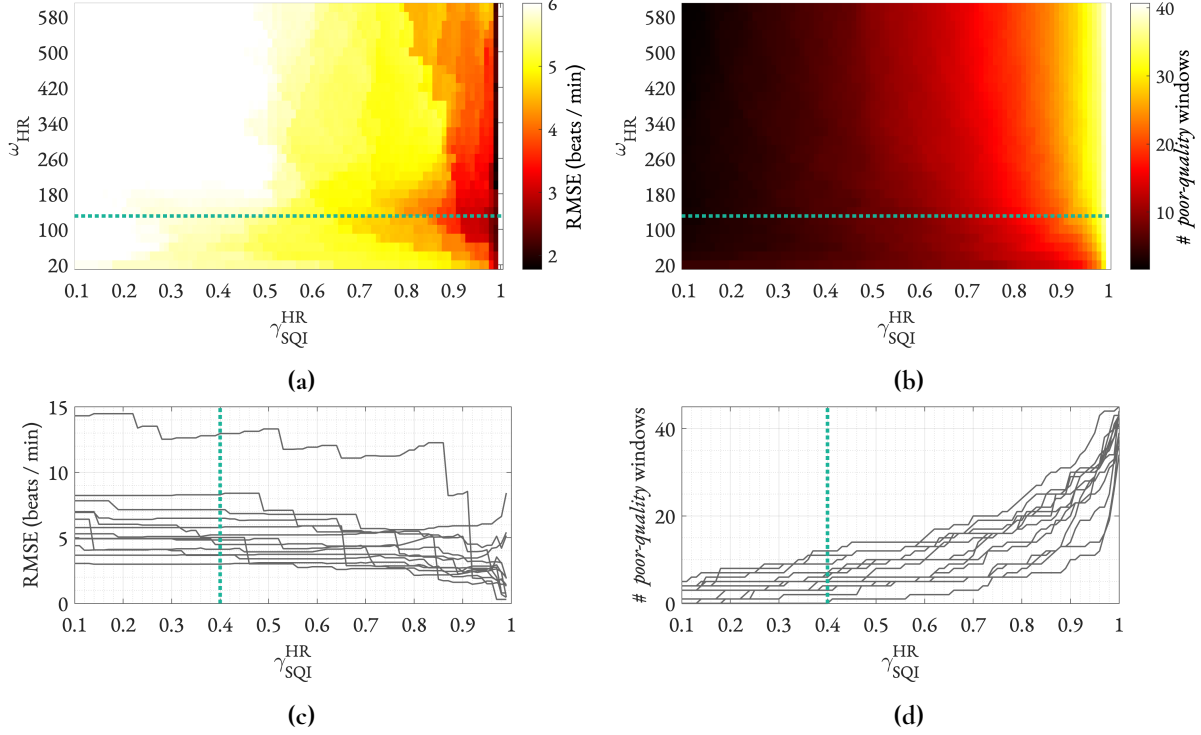


Figure 9.8: Sensitivity analysis of w_{HR} and γ_{SQI}^{HR} to determine their optimum values. (a) and (b): sensitivity of the variables on participant-wise median RMSE between HR_{Ref} and HR_{Cuff} and the number of poor-quality windows respectively. From this analysis w_{HR} was set to be 120s. (c) and (d): sensitivity of γ_{SQI}^{HR} on RMSE and the number of poor-quality windows respectively at a w_{HR} value of 120s. From this analysis γ_{SQI}^{HR} was set to be 0.4. These thresholds are shown in green in the relevant plots.

Figures 9.8(a) and 9.8(b) show the impact of w_{HR} and γ_{SQI}^{HR} on the participant-wise median RMSE and the number of poor-quality windows respectively. For w_{HR} values between 0s and 120s, RMSE values improved as w_{HR} increased, across all γ_{SQI}^{HR} . However, for window sizes larger than 120s, RMSE was largely independent of w_{HR} . Additionally, the number of poor-quality windows showed little sensitivity to w_{HR} . As a result, w_{HR} was set to 120s.

Figures 9.8(c) and 9.8(d) show the participant-wise sensitivity to γ_{SQI}^{HR} at $w_{HR} = 120$ s. Increasing values of γ_{SQI}^{HR} decreased the RMSE between the two HR measurements; however, there was a corresponding rapid increase in the number of windows rejected. γ_{SQI}^{HR} was therefore set at a value of 0.4.

Figure 9.9 shows an example from our proposed HR-based detection of poor-quality BP values for a typical participant in the nABPM study.

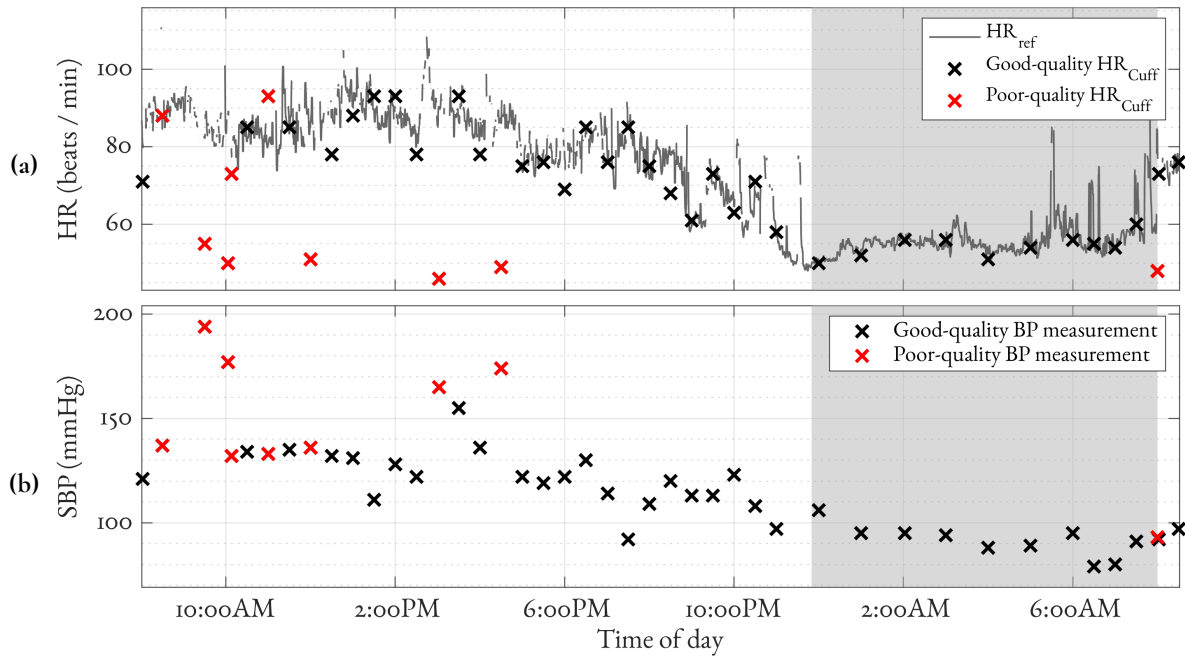


Figure 9.9: HR-based detection of poor-quality BP values for a typical participant in the nABPM study. BP values were removed from analysis if the % error between HR_{Ref} and HR_{Cuff} was greater than 10 %, or if SQI_{HR} exceeded γ_{SQI}^{HR} (a) HR_{Ref} and HR_{Cuff} values. (b) SBP time series recorded by the ABPM.

9.5.3 Double-cosinor model

For each participant, we processed the reference good-quality ABPM signal using a double-cosinor statistical model (see section 7.6.1). The double-cosinor model performs a filtering operation such that high-frequency components are attenuated, resulting in a smooth approximation of the overall ABPM time series [335, 341]. Figure 9.10 shows the results of the double-cosinor model fitted to the good-quality SBP measurements for the Day₀ data from all participants. Assessing model goodness-of-fit using an F-test (see section 7.6.2) indicated that the double-cosinor model was appropriate, at the $p < 0.05$ level, for all participants apart from ID: 1007. For this participant (who reported significant difficulty sleeping due to the measurement equipment) there was a significant dispersion in nocturnal BP values.

The differences between DR_{SBP} values computed from the double-cosinor smoothed ABPM time series and the original good-quality data were found to be statistically non-significant (Wilcoxon signed-rank test: $p = 0.123$). This suggests that the smoothing process and attenuation of high-frequency variations did not result in a loss of information related to an individual's dipping status. All participants reported a fall in BP during night-time hours. Using the DR_{SBP} values computed with the smoothed BP signal and the phenotype classification rules outlined in table 7.2, we classified three participants (ID: 1002, 1007, and 1015) as 'non-dippers', while the remaining participants were classified as 'dippers'.

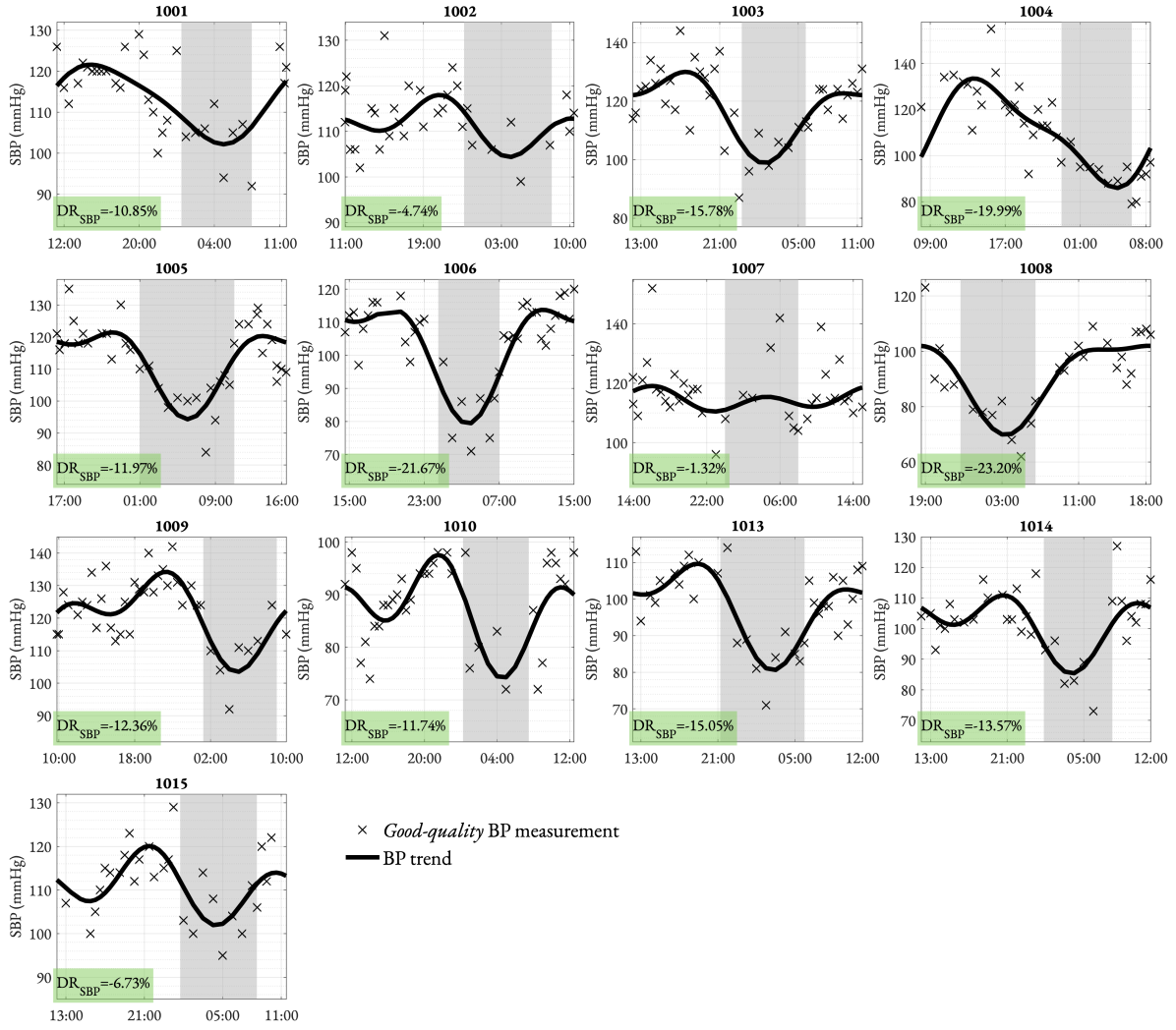


Figure 9.10: Double-cosinor model smoothing for Day₀ ABPM data for all participants to compute the reference BP time series. The timestamp at the left-hand edge of the plot indicates the time at which the BP monitoring started for that study participant, and it notably varies across different participants. For one participant (ID: 1007) the double-cosinor model did not pass the goodness-of-fit F-test at the $p < 0.05$ level. The shaded area indicates self-reported sleep time.

9.6 Estimating Day₇ blood pressure

The main aim of this chapter was to report the accuracy of the estimation of 24-hour BP values, 7 days after initial calibration. In particular, we compare the performance of *individual calibration* (personalised regression) to *hybrid calibration* (leave-one-subject-out regression) strategies. For each of these strategies, we derived different *baseline references* against which performances were evaluated. We developed linear (ordinary least squares, OLS) and non-linear (random forest, RF) ML regression models, using two different feature sets, $X \in \{\text{PPG}, \text{PPG}+\text{PAT}\}$. The PPG feature set included all PPG_{EA} , PPG_{fts} and HR_{PPG} . The PPG+PAT feature set included all PPG_{EA} , PPG_{fts} , PAT_{EA} , and HR_{Ref} .

For each 24-hour session, observations of X , x_i , and the regression target vector y , y_i , were

established in a feature window w_f of length 10 minutes and step 10 minutes. We computed x_i by taking the mean average of the relevant features for each w_f , including only data points with a corresponding $SQI > 0$ in the window.

Window-based $SQI_{PPG}^{EA}(w_f)$, $SQI_{PAT}^{EA}(w_f)$, and $SQI_{HR}(w_f)$ were then computed as their mean values within each window, w_f . For each of PPG_{EA} , PAT_{EA} , and HR, if their respective window-based SQI fell below a γ_{SQI} threshold, then we excluded that window from further analysis. To determine the optimum value of γ_{SQI} we conducted a sensitivity analysis to evaluate its impact on the proportion of w_f windows with valid measurements of PPG_{fts} , PAT_{EA} , or HR across all Day₀ sessions, for each 10-minute window of the day.

The proportion of good-quality windows varied throughout the day, with fewer windows meeting the SQI threshold during the day than at night. This is demonstrated in Figure 9.11, where participant-wise median values are shown for each signal over 24 hours. As a reference, figure 9.11 (d) shows the participant-wise median accelerometer values for each 10-minute window, w_f , demonstrating high activity levels during the day. Compared to HR, PPG_{EA} and PAT_{EA} showed higher sensitivity to changes in γ_{SQI} , indicating that their respective SQIs were more stringent than that of HR. Based on the sensitivity analysis, we set γ_{SQI} to 0.5 (shown by the green circle in figure 9.11), resulting in the exclusion of 40.5% of PPG_{EA} , 43.4% of PAT_{EA} , and 21.9% of HR measurements during the daytime, and 5.3%, 11.5%, and 7.6% during the night-time across all Day₀ sessions.

We handled missing data by participant-wise cubic splines imputation. Subsequently, for each participant we smoothed each feature with a moving median filter 3 samples long (corresponding to 30 minutes) in order to mitigate the impact of spurious data, which is common in a free-living context.

To establish the regression target vector y , for each session, double-cosinor modelled BP time series were sampled every 10 minutes (corresponding to the centre of w_f).

9.6.1 Individual calibration

Individual calibration accounts for the individualised nature of calibrating PAT or PPG_{fts} to BP. In this study, we implemented *individual calibration* through independent personalised regression models, trained on a participant's Day₀ data and tested on the same participant's Day₇ data.

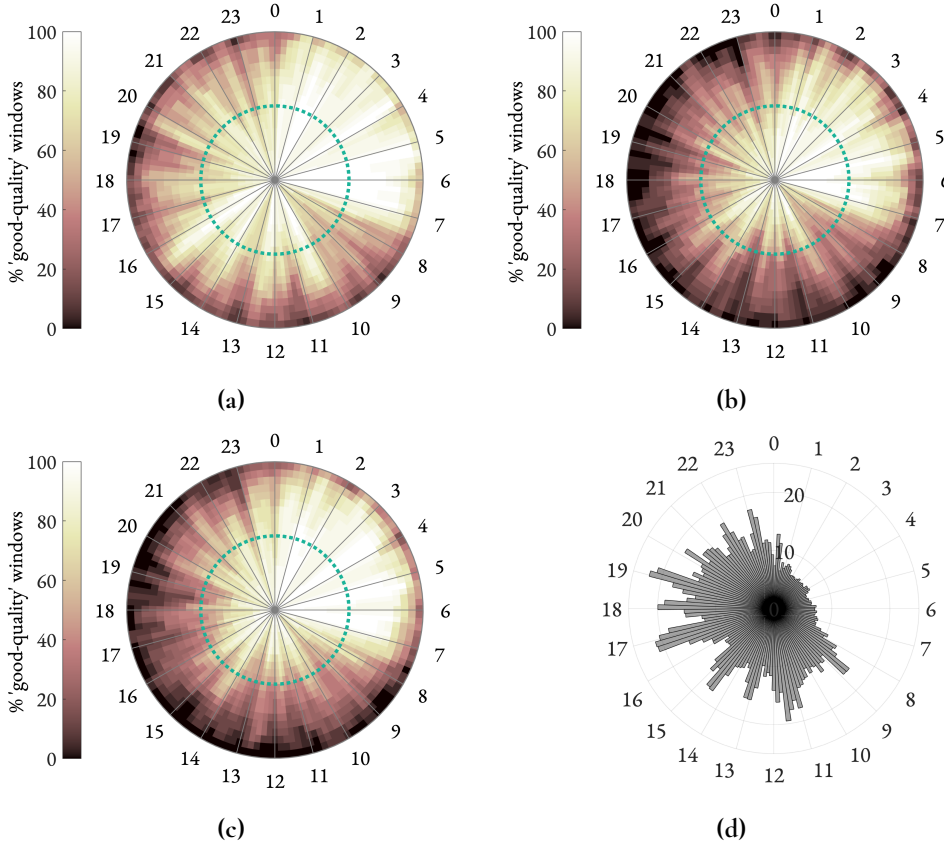


Figure 9.11: Sensitivity analysis of γ_{SQI} on the participant-wise median % good-quality windows across all Day₀ sessions for (a) HR_{Ref}, (b) PAT_{EA}, and (c) PPG_{EA}. For each 10-minute window of the day (shown by the perimeter of the plot as the hour of the day), γ_{SQI} (shown by the radius of the plot, starting at 0 at the centre and ending at 1 at the circumference) was varied and the heatmap shows the participant-wise median % good-quality w_f windows (i.e. at values close to the centre, the γ_{SQI} threshold is low and so the majority of windows were considered to be good-quality, as γ_{SQI} increases a higher proportion of windows are labelled as poor-quality). From this analysis, γ_{SQI} was set as 0.5, as shown by the green circle. For reference, (d) shows the participant-wise median accelerometer values (a.u.) for each 10-minute window.

9.6.2 Hybrid calibration

To minimise the number of calibration measurements required before measuring an individual's BP, commercial cuffless BP devices adopt *hybrid calibration* strategies [24]. To model this approach, we implemented *hybrid calibration* models trained in a nested leave-one-subject-out cross-validation (LOSOVCV) framework. Within this framework, Day₇ BP values were estimated for the p^{th} participant using data collected from Day₀ sessions for all participants apart from p .

As with ΔBP estimation in the MOLLIE dataset (see chapter 6), we implemented *hybrid calibration* by subtracting a reference calibration value from each BP measurement and waveform feature. The participant-wise calibration value of BP ($\text{BP}^{p, \text{calib}}$) and the j^{th} waveform feature ($f_j^{p, \text{calib}}$) were computed as their respective mean values during the 2-minute calibration period at the start of the Day₀ recording. The Day₀ calibration values were also used for Day₇ test

data in order to establish whether this calibration was still relevant 7 days later. Therefore, y_i^p and x_{ij}^p took the form:

$$\begin{aligned} i &= \{1, \dots, N\}, \\ y_i^p &= \text{BP}_i^p - \text{BP}^{p, \text{calib}}, \quad x_{ij}^p = f_{ij}^p - f_j^{p, \text{calib}}, \quad j = \{1, \dots, M\}, \\ p &= \{1, \dots, N_p\} \end{aligned} \quad (9.3)$$

where N and M are the number of observations and features in X respectively. N_p is the number of participants. Finally, we added participant age, sex, height, weight, and BMI as static features to the input feature set.

9.6.3 Regression models

To estimate the Day₇ BP time series, linear combinations of features using regularised ordinary least squares (OLS), and non-linear combinations using a RF regression model were both investigated. We have previously implemented LASSO regularisation to minimise the likelihood of linear models overfitting to the training set. LASSO imposes the L1-norm penalty on the feature coefficients. The regularisation value, λ , is typically optimised based on the minimisation of RMSE in a validation set. However, in situations where the validation data contains a significant degree of noise, it is not uncommon for LASSO regularisation to set all feature coefficients to 0, leading to predictions that use only the intercept term [378]. As a result, we employed elastic net regularisation (see appendix D.1.5). This includes an L2-norm penalty in addition to the L1-norm imposed by LASSO, overcoming the limitations of LASSO by significantly reducing the tendency of the regularisation to set a feature coefficient to zero [378]. We refer to this model as ELAS+OLS.

Model hyperparameters (λ for ELAS+OLS and $mtry$ for RF) were optimised using CV to minimise the median RMSE in the validation set. For *individual calibration*, a 5-fold CV approach was repeated 10 times with new random partitions in order to reduce the bias in resampling. For *Hybrid calibration*, a nested LOSOCV approach was implemented.

Finally, model outputs were smoothed using the double-cosinor model to compute the estimated BP time series, $\hat{\text{BP}}$. Separate models were trained for estimating SBP, MAP, and DBP.

9.6.3.1 DR_{SBP} estimation

Estimates of DR_{SBP} ($\hat{\text{DR}}_{\text{SBP}}$) were computed as the percentage difference between the mean estimated night-time SBP ($\hat{\text{SBP}}_{\text{Night}}$) and the mean estimated daytime SBP ($\hat{\text{SBP}}_{\text{Day}}$), normalised by the latter. In this study, night-time was defined as the self-reported sleeping period from each participant for each session. Daytime was defined as 10 am to 8 pm.

9.6.4 Baseline references

To establish reference performance statistics, separate *baseline references* were developed for *individual* and *hybrid* calibrations. For *individual calibration*, the *baseline reference* assumed no BP changes for each participant from their Day₀ calibration value $BP^{p, \text{calib}}$ (i.e. $\hat{BP}_i^p = BP_i^{p, \text{calib}}$ for all i and p).

For *hybrid calibration*, following the work of Mieloszyk *et al.* [217], we trained both ELAS+OLS and RF regression models using features not derived from the PPG or ECG signals that have previously demonstrated a significant relationship with BP over 24 hours. These models incorporated demographic static features such as age, sex, height, weight, and BMI, which have been linked to the circadian rhythm of BP [2, 379]. We additionally modelled the time of measurement by accounting for its circular nature using the sine & cosine transforms:

$$t_c = \cos(2\pi \frac{\text{hour}}{24}), \quad t_s = \sin(2\pi \frac{\text{hour}}{24}) \quad (9.4)$$

where 'hour' is the time of the day in hours. These *hybrid calibration baseline references* models were trained within a LOSOCV framework as outlined in section 9.6.2.

9.7 Evaluation and generalisability of models

We used the following metrics to evaluate BP estimation performance: ρ_p , RMSE, and MAE. DR_{SBP} estimation was assessed as the absolute difference with respect to the reference. Differences in performance metrics across all folds from the models were evaluated for statistical significance by a two-tailed Wilcoxon signed-rank test (see appendix C.1). We adjusted the p-values for multiple comparisons using the Benjamini-Hochberg method (see appendix C.1.2).

9.7.1 Shapley additive explanation values feature importance and feature stability

To determine the importance of each feature, we utilised the mean absolute SHAP value (ϕ_j) computed across the training data. SHAP values, which represent the contribution of each feature to the model predictions, were computed as detailed in section 4.5. To account for variability in the training data across folds, we employed a ranking coefficient that normalised the SHAP values between 0 and 1 at each fold, as described in section 6.6.7. We further evaluated the SHAP values for each model prediction to highlight variations in feature contributions from training and test observations.

It remains unclear whether the best-performing features for an individual on one day are stable, and therefore equally as important for the same individual seven days later. Moreover, whether the

'optimum' features (defined as the features most relevant at test time) for an individual may be ascertained using LOSO regression also remains to be explored. To investigate feature stability, we identified each participant's 'optimum' feature rankings by training ELAS+OLS and RF regression models on data collected during their Day₇ session. The agreement between the Day₀ feature rankings and the 'optimum' Day₇ feature rankings was assessed by ρ_s (see section 6.6.7).

9.7.2 Model complexity

Highly complex models require a large number of features in order to make accurate predictions. However, these models also tend to suffer from overfitting whereby they perform well in one context but are not generalisable [377]. We assessed the complexity requirements of the models by analysing the validation performance (ρ_p and RMSE) as more features were included in the feature set X . Starting with a null feature set, $X_\emptyset$, PPG+PAT features were ordered by their absolute ρ_s with the training target vector y : $\{f_1, f_2, \dots, f_M\}$. Ordered features were then sequentially added to the feature set and the models outlined above were trained, with their optimum validation performance then reported.

9.8 Results

9.8.1 Data recording in a free-living context

Table 9.1 summarises the recording durations and data availability for all sessions in the nABPM study. Accurate data recording in a free-living context can be challenging, and approximately 10-20% of each session was classified as 'Active' during which all ECG and PPG data were removed from analysis. A valid HR_{Ref} signal was available for approximately 70-80% of each session. There was a considerably lower and more varied availability of PAT_{EA} and PPG_{fts} across all sessions.

The reference Day₇ SBP time series was computed using double-cosinor modelling, as shown in figure H.1 in the appendix. The double-cosinor model was found to be appropriate, at the $p < 0.05$ level, for all individuals apart from one, ID: 1013. This participant reported significant difficulty sleeping during the session because of the high night-time temperature at the time of recording. Across all Day₇ recordings, the DR_{SBP} values between the good-quality SBP data and the double-cosinor smoothed SBP time series also showed non-significant differences (Wilcoxon signed-rank test: $p = 0.091$).

Table 9.1: Recording durations and data availability for the nABPM study. The summary row indicates the median statistic over the recordings.

ID	Total recording time (HH:mm:ss)	'Inactive'	HR _{Ref}	PAT _{EA}	PPG _{fts}	# BP meas.
<i>Day₀ recordings (training set)</i>						
1001	24:47:00	93.23%	73.30%	68.59%	66.58%	32 (84.21%)
1002	25:02:10	82.23%	61.24%	49.26%	57.92%	32 (82.05%)
1003	23:28:19	85.51%	77.40%	47.57%	66.04%	38 (90.48%)
1004	24:30:19	86.61%	77.53%	46.93%	45.57%	35 (79.55%)
1005	24:21:19	83.39%	73.22%	53.38%	62.27%	39 (90.70%)
1006	24:31:49	92.16%	84.93%	81.53%	86.97%	38 (97.44%)
1007	25:12:19	93.83%	82.65%	62.82%	70.75%	38 (88.37%)
1008	24:04:49	91.99%	83.05%	57.45%	57.45%	28 (82.35%)
1009	24:18:10	81.96%	76.12%	69.95%	78.87%	38 (88.37%)
1010	25:19:30	84.11%	75.02%	62.52%	56.60%	36 (83.72%)
1013	24:10:10	81.06%	81.37%	53.10%	46.20%	34 (82.93%)
1014	24:07:19	75.12%	75.31%	51.82%	54.58%	34 (82.93%)
1015	22:26:00	81.91%	73.55%	52.75%	58.69%	28 (71.79%)
Summary	24:21:19	84.11%	76.12%	53.38%	58.69%	35 (83.72%)
<i>Day₇ recordings (test set)</i>						
1001	23:14:49	94.85%	81.01%	70.26%	73.13%	39 (95.12%)
1002	24:00:19	80.29%	59.71%	40.27%	44.43%	36 (80.00%)
1003	24:24:19	83.43%	77.17%	47.80%	58.73%	39 (82.98%)
1004	24:38:00	85.46%	77.81%	49.39%	46.68%	32 (71.11%)
1005	24:04:40	85.07%	75.45%	60.22%	65.76%	37 (84.09%)
1006	23:10:00	89.99%	88.49%	79.86%	90.65%	39 (95.12%)
1007	23:24:10	90.84%	92.58%	72.64%	76.20%	38 (100.00%)
1008	24:26:10	92.83%	79.12%	54.56%	59.34%	31 (88.57%)
1009	24:04:10	84.80%	86.56%	78.25%	90.02%	39 (95.12%)
1010	24:12:10	89.33%	86.77%	77.13%	79.19%	41 (97.62%)
1013	24:21:19	81.42%	80.06%	53.38%	50.64%	32 (76.19%)
1014	24:06:00	79.53%	77.46%	56.71%	56.71%	34 (82.93%)
1015	24:06:30	85.80%	72.59%	49.78%	51.16%	39 (95.12%)
Summary	24:06:00	85.46%	79.12%	56.71%	59.34%	38 (88.57%)

'Inactive' indicates the proportion of the recording that the motion signal crossed the γ_{acc} threshold. HR, PAT_{EA}, and PAT_{EA} indicate the proportion of the recording for which a valid estimate of the respective signal was available. # BP meas. indicates the number (and proportion) of good-quality BP measurements.

9.8.2 Model performance

Figure 9.12 shows the ρ_p and RMSE validation performance curves as PPG+PAT features were cumulatively added to the feature set. Note that different validation sets were implemented for *individual* (5-fold CV) and *hybrid calibration* (LOSOCV) strategies and so a direct comparison of absolute performance values between the two is not possible. Overall, all models showed smooth improvements in validation performance as more features were added. For each calibration strategy, better validation metrics were observed in RF models than in ELAS+OLS models. These non-linear

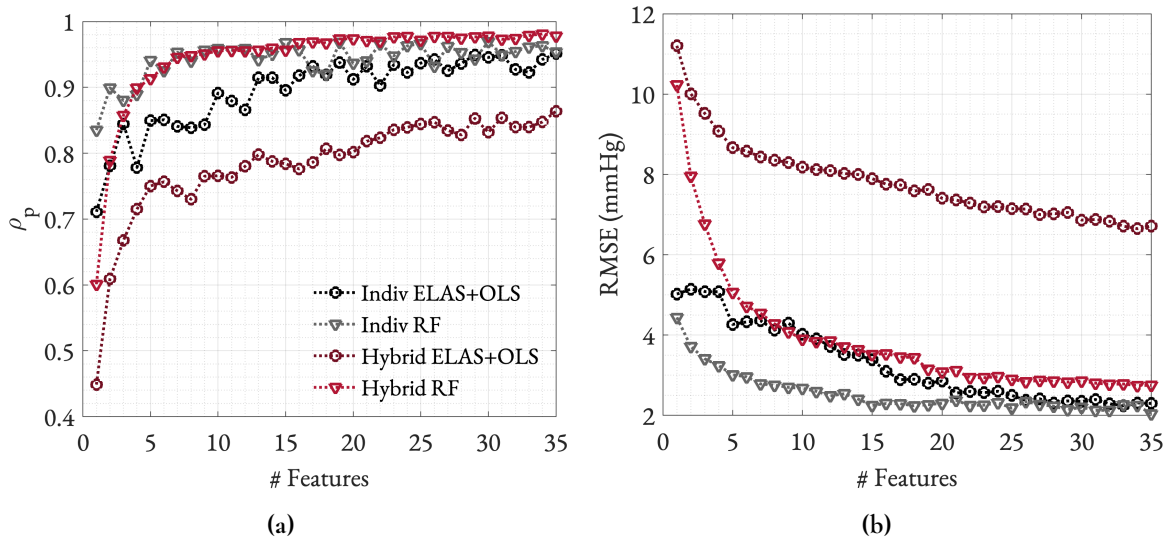


Figure 9.12: Assessing model complexity based on the participant-wise median validation performance curves as PPG+PAT features were sequentially added to each model. Features were ranked by their absolute Spearman's rank correlation ($|\rho_s|$) with the target vector. (a) Median ρ_p computed across the validation data. (b) Median RMSE across the validation data.

models also tended to reach their optimum performance for a fewer number of input features. ELAS+OLS displayed consistent improvements in performance as more features were added, demonstrating that the LASSO limitation of setting all feature weights to zero was addressed through the elastic net. *Hybrid calibration* models reported notably sharper improvements in validation performance across the first five features. The performance across all models reached a plateau after around 15 features, after which marginal improvements were recorded.

Table 9.2 shows the median and IQR performance metrics for SBP estimation by all models computed across the Day7 sessions for all 13 participants. Additionally, the results of DR_{SBP} estimation are provided. The corresponding results for MAP and DBP estimation are provided in the appendix tables H.1 and H.2 respectively.

Using *individual calibration*, all models outperformed the results obtained using the *baseline reference* method. With the exception of $ELAS+OLS_{PPG+PAT}$, all models reported statistically significant p -values indicating that consistent improvements in all performance statistics were observed ($p < 0.05$ for all). For both the PPG and the PPG+PAT feature sets, the RF model, compared to ELAS+OLS, resulted in decreased error values (RMSE and MAE), for estimating absolute values of BP, with weaker ρ_p and DR_{SBP} error values for tracking relative changes in BP.

The corrected Wilcoxon signed-rank test failed to reject the null hypothesis of equal medians between ELAS+OLS and RF across RMSE, MAE and ρ_p . A significant ($p < 0.001$) difference was noted for DR_{SBP} estimation. Adding PAT_{EA} to the feature set did not yield a consistent and significant improvement in SBP estimation. Overall, the RF_{PPG} model reported the lowest median

Table 9.2: Performance statistics of Day7 SBP estimation using the models proposed. DR_{SBP} errors are given as absolute differences with respect to the reference. Results are given as median (IQR) across all participants. Entries in bold indicate the best performance for that metric.

Model name	ρ_p	RMSE *	MAE *	$DR_{SBP}^\dagger$
<i>Individual calibration</i>				
Baseline reference	- (-)	10.02 (6.95)	8.41 (6.50)	16.75 (7.73)
ELAS+OLS _{PPG}	0.86 (0.22)	8.12 (5.50)	7.01 (4.04)	7.52 (11.12)
RF _{PPG}	0.83 (0.23)	6.47 (7.96)	5.45 (6.40)	10.75 (13.81)
ELAS+OLS _{PPG+PAT}	0.90 (0.16)	9.35 (5.43)	8.21 (5.64)	7.07 (10.48)
RF _{PPG+PAT}	0.81 (0.19)	6.90 (8.03)	5.63 (6.23)	10.87 (13.54)
<i>Hybrid calibration</i>				
Baseline reference ELAS+OLS	0.82 (0.18)	11.57 (9.90)	10.67 (9.99)	7.12 (4.76)
Baseline reference RF	0.80 (0.11)	11.11 (10.19)	10.25 (9.45)	7.58 (7.62)
ELAS+OLS _{PPG}	0.82 (0.29)	12.09 (20.13)	10.46 (19.46)	8.95 (16.02)
RF _{PPG}	0.47 (0.66)	11.21 (9.83)	10.50 (9.77)	14.15 (11.27)
ELAS+OLS _{PPG+PAT}	0.84 (0.29)	11.02 (20.41)	9.62 (19.73)	8.98 (15.59)
RF _{PPG+PAT}	0.46 (0.73)	10.53 (8.89)	9.88 (8.83)	14.19 (11.97)

* results given in units of mmHg. † results given in units of %.

RMSE (6.47 mmHg) and MAE (5.45 mmHg) values, whereas the ELAS+OLS_{PPG+PAT} had the highest median ρ_p (0.90) values and lowest median DR_{SBP} (7.07%) errors.

Using *hybrid calibration*, none of the models outperformed the *baseline reference* RF for RMSE, MAE, or DR_{SBP} . ELAS+OLS models resulted in improved ρ_p values, however, with the associated wide IQR resulting in this difference being non-significant. For both the PPG and PPG+PAT feature sets, the RF model produced significantly ($p < 0.05$ for all) weaker ρ_p values and DR_{SBP} errors than the two ELAS+OLS models. The RF models had lower median error values (RMSE and MAE) with considerably tighter IQR than the ELAS+OLS models, suggesting better consistency of results.

The corrected Wilcoxon signed-rank test suggests there were no notable differences between any of the *hybrid calibration* models for either RMSE or MAE. Of the waveform-based models, ELAS+OLS_{PPG} had the lowest DR_{SBP} errors, although this error was 1.9% higher than the *baseline reference* ELAS+OLS.

When comparing *individual calibration* and *hybrid calibration* strategies, we observed large and statistically significant ($p < 0.01$ for all) differences in the error values between the RF models, indicating that the *individual calibration* models were significantly more appropriate for estimating absolute BP. Whereas, non-significant differences were observed in ρ_p and DR_{SBP} between the two ELAS+OLS models, suggesting that both calibration strategies were able to equivalently track relative changes in BP.

All models reported large IQR values demonstrating inconsistencies in estimation performance between participants. Comparing the results reported in table 9.2 to the validation errors reported in figure 9.12 suggests that overfitting occurred.

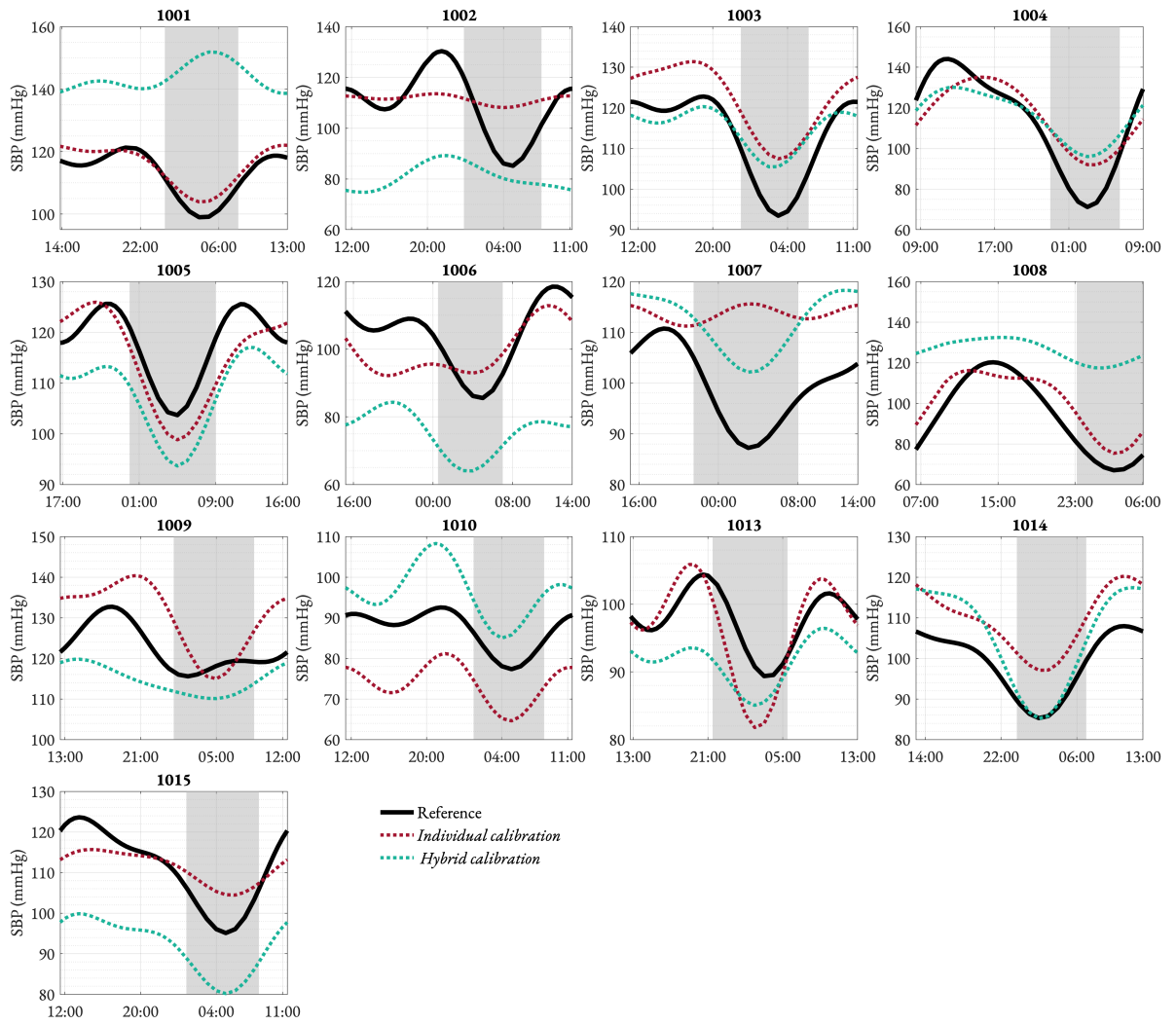


Figure 9.13: Comparison of reference and estimated Day7 SBP values for each participant. *Individual calibration* is from the ELAS+OLS_{PPG}+PAT model, whereas *hybrid calibration* is from the ELAS+OLS_{PPG} model. These two models reported the lowest DR_{SBP} error values for each calibration strategy. The timestamp at the left-hand edge of the plot indicates the time at which the BP monitoring started for that study participant, and it notably varies across different participants

9.8.3 Participant-level results

Figure 9.13 shows the reference and estimated Day7 SBP values, for each participant. For *individual calibration*, we show the results of the ELAS+OLS_{PPG}+PAT model. For *hybrid calibration*, we show the results of the ELAS+OLS_{PPG} model. These two models are associated with the lowest DR_{SBP} error values for each calibration strategy. In general, the former model provides a much closer fit to the reference values than the latter. The largest *individual calibration* error was reported for ID: 1007, 14.46 mmHg. As reported in figure 9.10, this individual's Day0 SBP time series did not pass the goodness-of-fit F-test at the $p < 0.05$ level. For the *hybrid calibration* model, a notable DC offset was observed for a number of participants, IDs: 1001, 1002, 1006, 1008,

1010, and 1015, contributing to the high RMSE and MAE values found in table 9.2. The largest error was reported for ID: 1002, 39.19 mmHg.

Both calibration strategies reported similar DR_{SBP} values and non-significant ($p = 0.68$) differences were observed between the two. The differences between the median reference and estimated DR_{SBP} was 1.08% ($p = 0.047$) and 5.13% ($p = 0.027$) for *individual* and *hybrid calibration* models respectively.

Figure 9.14 shows the *individual calibration* RMSE as a function of the range (max-min) of BP values in Day₀ and Day₇. For RF models, large errors only occurred when the Day₀ range was lower than the Day₇ range demonstrating their inability to extrapolate out of the range they were trained on. ELAS+OLS models did not exhibit the same dependence.

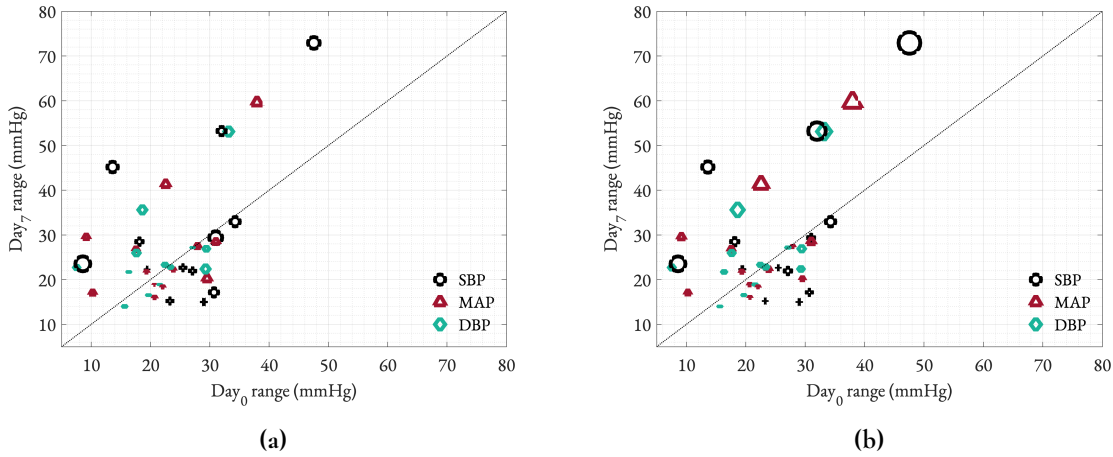


Figure 9.14: *Individual calibration* (a) ELAS+OLS_{PPG+PAT} and (b) RF_{PPG} model RMSE as a function of the range (max-min) of BP values in Day₀ and Day₇. RMSE is denoted by the marker size, with larger markers indicating higher estimation error. The dotted line shows the line of unity, any data above this line refers to participants whose training data had a lower range than their test data.

9.8.4 Feature importance

Figure 9.15 shows the median SHAP (ϕ_j) values feature ranking coefficients for the top 15 (chosen as the number for which the performance reaches a plateau in figure 9.12) features computed across all participants. Figure 9.15(a) shows the results for *individual calibration* RF_{PPG} and Figure 9.15(b) shows the results for ELAS+OLS_{PPG+PAT}. These two models achieved the lowest MAE for each calibration strategy. The top five features for *individual calibration* as determined by their median ranking coefficient were: HR, NHA, Skewness, APG PCA₂, and Sys σ . The top five features for *hybrid calibration* were Sex, PPG PCA₂, IPA, c/a , and Width₅₀. The correlation between the two feature rankings was weak ($\rho_s = -0.20$). Extreme ranges were observed in figure 9.15 with the majority of features being both the lowest and the highest ranked feature for different participants.

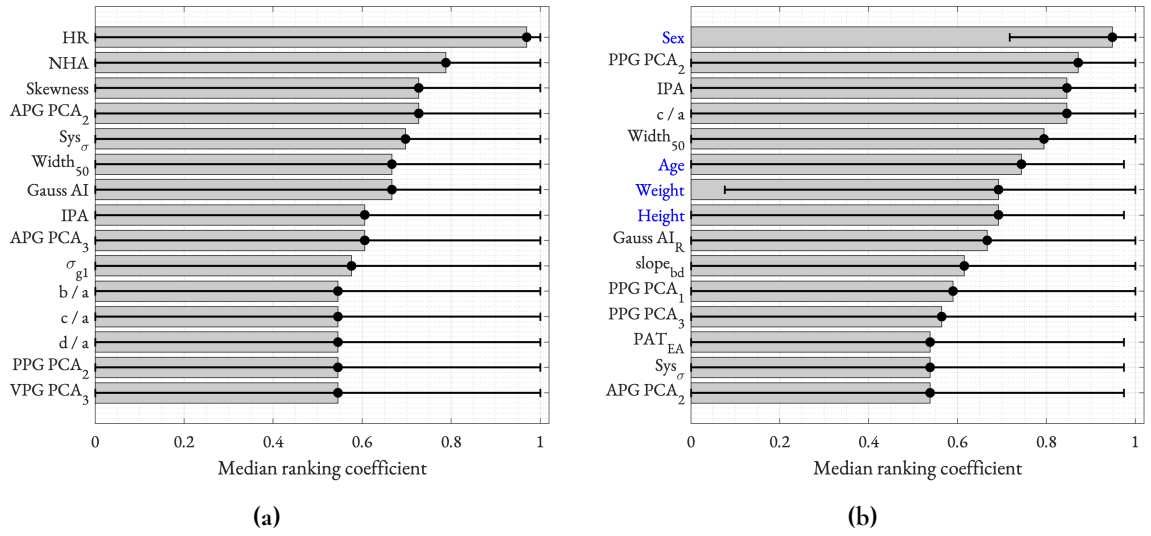


Figure 9.15: Top 15 median (across participants) SHAP (ϕ_j) values feature ranking coefficient for (a) *Individual calibration* RF_{PPG} and (b) *Hybrid calibration* ELAS+OLS_{PAT+PPG}. Features are ordered by their respective median ranking coefficients. Error bars denote the range. Demographic features are shown in blue for clarity. The Spearman's correlation coefficient between the two strategies was negative and weak: $\rho_s = -0.20$.

Figure 9.16 extends this analysis to evaluate the stability of the best-performing features from the Day₀ training set to the 'optimum' features determined when training equivalent models on Day₇ data for each participant. This comparison was performed using ρ_s . Only ELAS+OLS_{PPG+PAT} models resulted in a median ρ_s value statistically greater than zero for MAP and DBP using *individual calibration*. In general, the overall agreement was weak, with all absolute median ρ_s values below 0.3. There were no significant differences in the ρ_s values computed with *individual calibration* and *hybrid calibration* procedures.

Figures 9.17 and 9.18 provide a more in-depth analysis of the *individual calibration* RF_{PPG} model estimations for the two participants for whom the lowest (ID: 1010) and highest (ID: 1007) RMSE values were obtained. Mean $|\phi_j|$ values for the top 15 ranked features across the training set are shown. The contribution of the remaining features is denoted 'Σ Remaining'. For 1010, predictions were largely dominated by Width₅₀, with minimal contributions from 'Σ Remaining'. The sharp changes in ϕ_j values for Width₅₀ estimations are not uncommon for RF models, especially when trained on a small dataset size, due to features values at the edges of the decision tree splitting criteria [380].

For participant 1007, a much more significant contribution of 'Σ Remaining' was observed, suggesting a highly complex model that has overfit to the training data. This overfitting was likely the result of the double-cosinor model not adequately modelling participant 1007's Day₀ SBP over the 24 hours of recording (see figure 9.10). Additionally, the reference Day₀ SBP data existed in

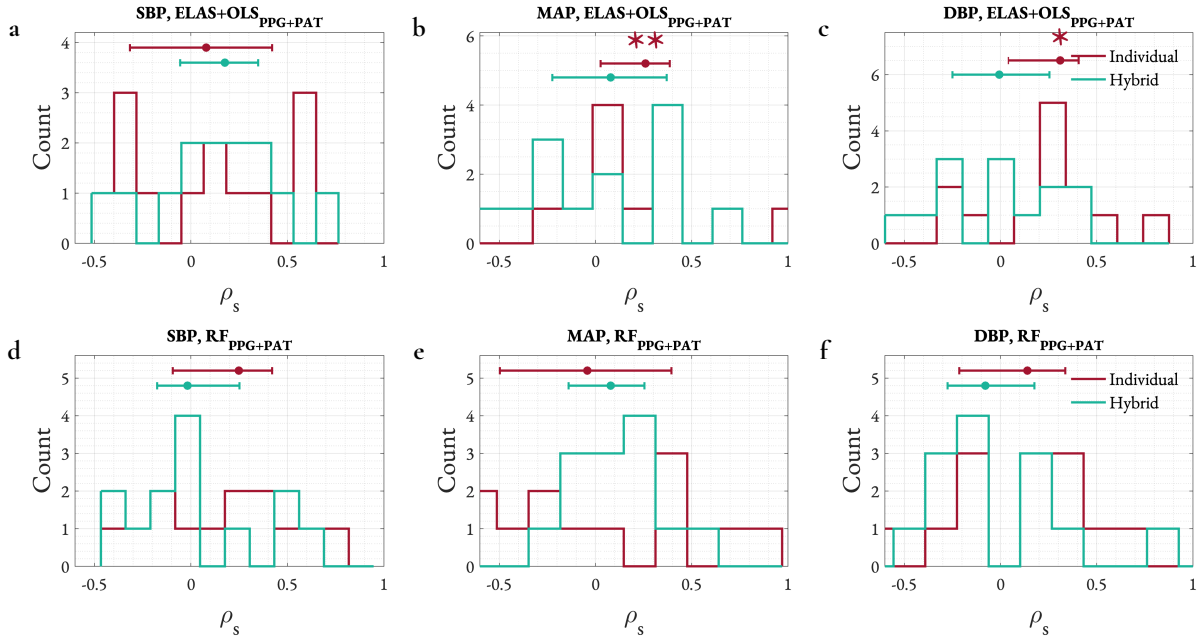


Figure 9.16: SHAP values feature rankings stability from the Day₀ training set to the 'optimum' features determined when training equivalent models on Day₇ data for each participant. The comparison of features was performed using ρ_s . The top row (a-c) shows ELAS+OLS_{PPG+PAT} models. The bottom row (d-f) shows RF_{PPG+PAT} models. The left, middle, right columns show the results for SBP, MAP, and DBP prediction respectively. Any models where the median ρ_s was statistically significantly greater than 0 are shown by: * $p < 0.05$, ** $p < 0.01$.

a tight range, 110-120 mmHg. This resulted in predictions within the same range, despite the reference Day₇ SBP values existing in the range of 85-110 mmHg.

Figures 9.19 and 9.20 provide a similar analysis of *hybrid calibration* ELAS+OLS_{PPG+PAT} model estimations for the two participants that reported the lowest (ID: 1003) and highest (ID: 1002) RMSE values respectively. For *hybrid calibration*, more complex models were expected (see figure 9.12) and the model estimates for participant 1003 were largely dominated by ' Σ Remaining'. For participant 1002, a clear calibration error of PPG PCA₂ (the most important feature for this participant) existed. The range of values for this feature in the test set fell to the extreme end of the distribution for the training set. As a result, the importance of this feature was over-inflated at estimation time.

9.9 Discussion

The nABPM study was designed to extend the analysis developed throughout this thesis towards estimation of ambulatory BP using PAT_{EA}, PPG_{fts}, and HR, 7 days after initial calibration. We developed interpretable linear and non-linear regression models implementing both *individual* and *hybrid calibration* strategies. As they are significantly more convenient for the user (often requiring only a single calibration point before BP may be recorded), *hybrid calibration* strategies

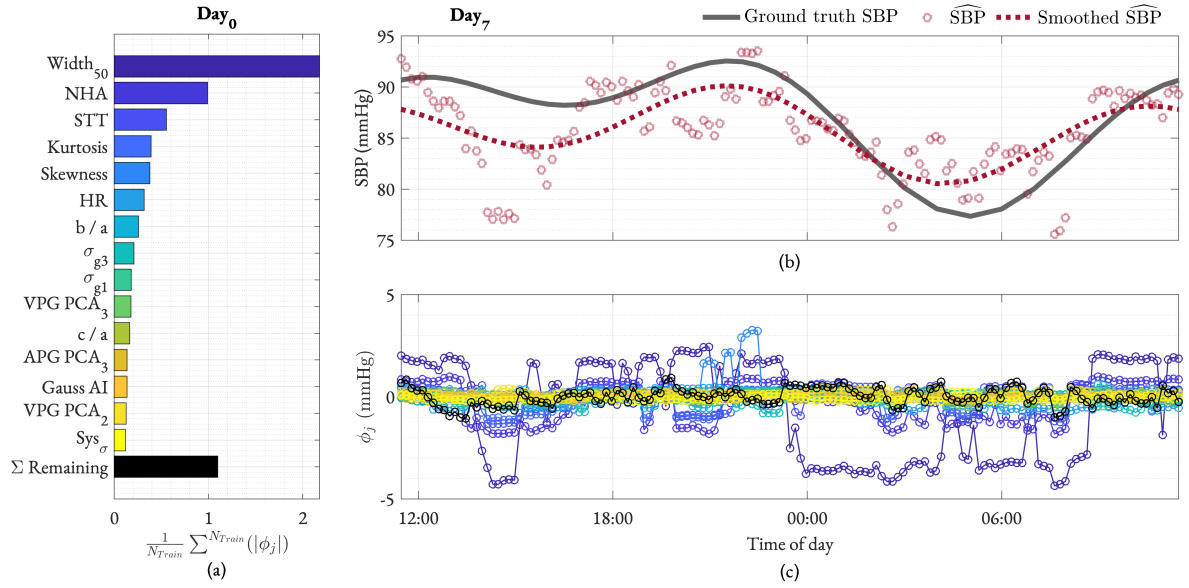


Figure 9.17: Participant ID: 1010, *individual calibration*, RF_{PPG} model. Analysis of ϕ_j in both training and test sets. (a) Shows the top 15 features ranked by mean $|\phi_j|$ across the Day₀ training set. The cumulated $|\phi_j|$ of the remaining features is denoted 'Σ Remaining'. (b) Shows the Day₇ ground truth SBP values, SBP, and the double-cosinor smoothed SBP. (c) Shows the ϕ_j values (colour-coded by (a)) for each Day₇ estimation. RMSE = 3.04 mmHg, $\rho_p = 0.83$.

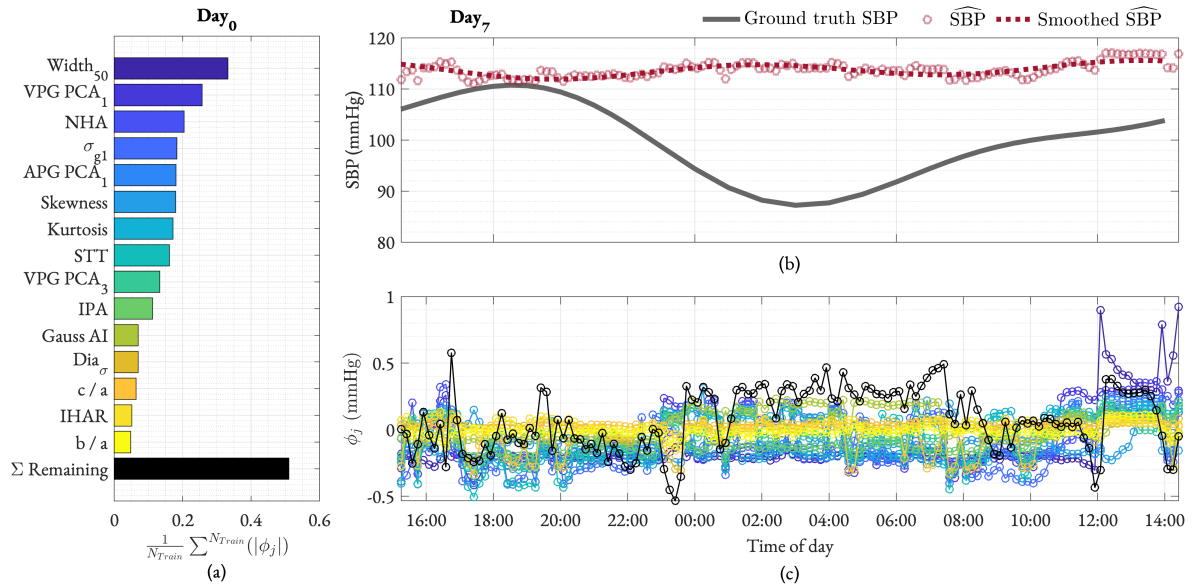


Figure 9.18: Participant ID: 1007, *individual calibration*, RF_{PPG} model. Analysis of ϕ_j in both training and test sets. (a) Shows the top 15 features ranked by mean $|\phi_j|$ across the Day₀ training set. The cumulated $|\phi_j|$ of the remaining features is denoted 'Σ Remaining'. (b) Shows the Day₇ ground truth SBP values, SBP, and the double-cosinor smoothed SBP. (c) Shows the ϕ_j values (colour-coded by (a)) for each Day₇ estimation. Note that the y-axis has a much smaller range than in figure 9.17 (c). RMSE = 14.46 mmHg, $\rho_p = -0.32$.

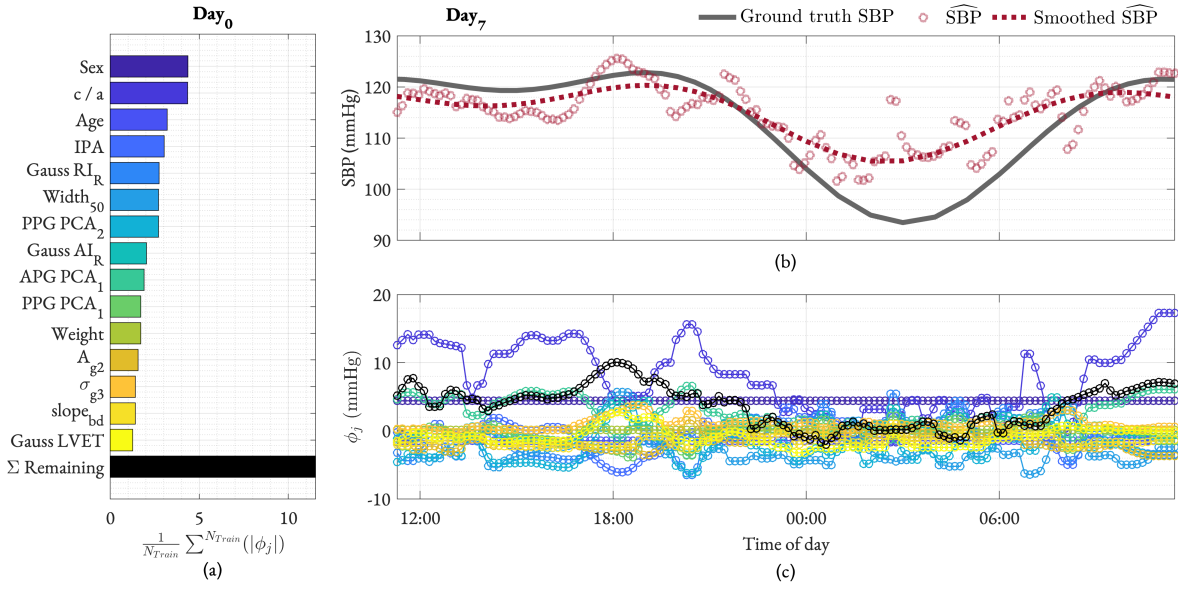


Figure 9.19: Participant ID: 1003, *hybrid calibration*, ELAS+OLS_{PPG+PAT} model. Analysis of ϕ_j in both training and test sets. (a) Shows the top 15 features ranked by mean $|\phi_j|$ across the Day₀ training set. The cumulated $|\phi_j|$ of the remaining features is denoted ' Σ Remaining'. (b) Shows the Day₇ ground truth SBP values, $\widehat{SBP}$, and the double-cosinor smoothed $\widehat{SBP}$. (c) Shows the ϕ_j values (colour-coded by (a)) for each Day₇ estimation. RMSE = 5.01 mmHg, $\rho_p = 0.95$.

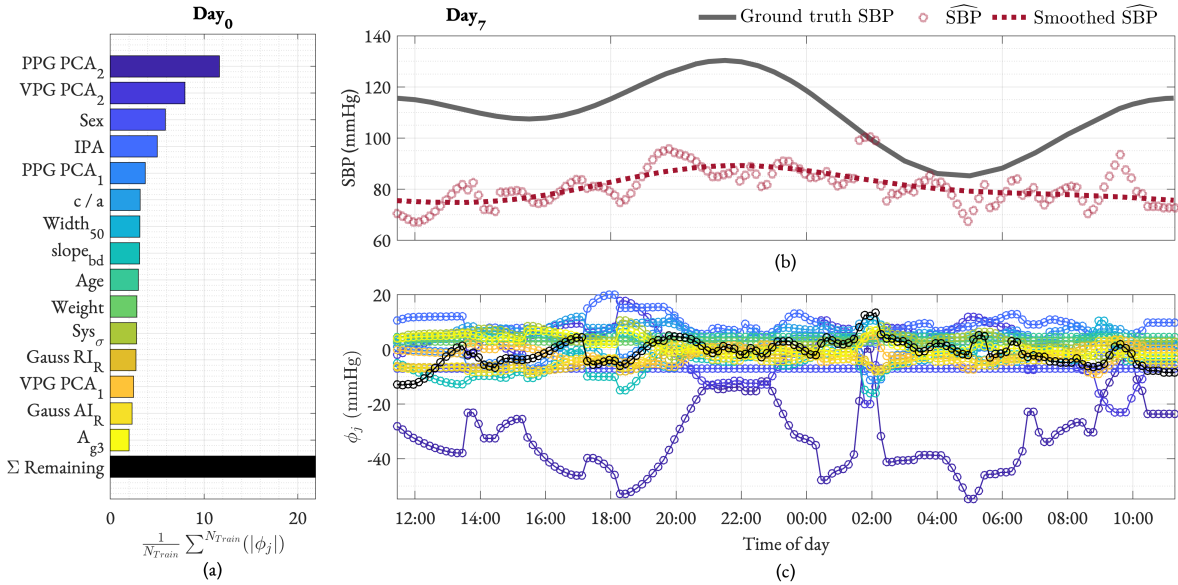


Figure 9.20: Participant ID: 1002, *hybrid calibration*, ELAS+OLS_{PPG+PAT} model. Analysis of ϕ_j in both training and test sets. (a) Shows the top 15 features ranked by mean $|\phi_j|$ across the Day₀ training set. The cumulated $|\phi_j|$ of the remaining features is denoted ' Σ Remaining'. (b) Shows the Day₇ ground truth SBP values, $\widehat{SBP}$, and the double-cosinor smoothed $\widehat{SBP}$. (c) Shows the ϕ_j values (colour-coded by (a)) for each Day₇ estimation. RMSE = 33.15 mmHg, $\rho_p = 0.57$.

are unanimously employed by commercial cuffless devices [24]. Therefore, previous equivalent studies in this field have restricted their analysis to this strategy and have not attempted to develop individual personalised regression models [381, 217].

Our analysis of the ABPM data in this study highlights the need for a cuffless solution to monitor ambulatory BP changes in the general population. For 8 sessions out of the 26 recorded (30.7%), participants reported trouble sleeping, defined as individuals who answered 'Very bad' or 'Fairly bad' on the ABPM sleep questionnaire after each session (see appendix B). This high proportion of disrupted sleep is consistent with previous reports in the literature and is one of the many factors driving intolerance for cuff-based BP monitoring [23].

Additionally, the reference ABPM time series exhibited high levels of noise, as participants lived freely during the recordings. Even during sleep, it is believed that cuff-based BP measurements may induce an arousal effect, temporarily increasing the measured physiological signal [14]. The low sampling frequency of the recorded data (every 30 minutes during the day, and every hour during the night) made detecting values corrupted by noise challenging. We addressed this by first removing observations associated with large differences between HR_{Cuff} and HR_{Ref} , and then subsequently using double-cosinor modelling to smooth high-frequency variations. We additionally smoothed the estimated BP values using the double-cosinor model.

9.9.1 Model performance

Prior work has suggested that vascular ageing progressively reduces the relevance of an initial calibration procedure, leading to the accuracy of calibrated cuffless BP monitors degrading over time. The severity of this degradation remains to be established, with commercial devices typically recommending recalibration every month [24]. However, Miao *et al.* [210] and Ding *et al.* [136] showed significant increases in cuffless BP estimation errors after 3 days and 2 weeks respectively. In this study, we evaluated BP estimation 7 days after initial calibration.

From a clinical point of view, there are a number of different use cases for cuffless BP monitoring in an ambulatory environment for the general population. For many of these uses, accurate calibration is important to ensure that absolute BP values are appropriately recorded. For example, current clinical guidelines for assessing hypertension and hypotension rely on accurate measurements in units of mmHg. In that respect, our results suggest that *individual calibration* models outperform *hybrid calibration* models, with median errors (RMSE and MAE) of the former around half of the latter ($p < 0.01$). Figure 9.13 demonstrates the consequences of inaccurate devices. Using the *hybrid calibration* model, participant 1001 would be misclassified as a 'reverse-dipper' with nocturnal hypertension, and participants 1002 and 1006 would be misclassified as

hypotensive [97]. These errors, therefore, may alter an individual's prescribed medication or may lead to medication being taken at the wrong time [57]. 1002 is analysed in further detail in figure 9.20 demonstrating that a calibration error for the most important feature PPG PCA₂ lead to a significant underestimation of SBP. This error underpins the complications of *hybrid calibration* where the subject-specific relationship between BP and the ECG and PPG cuffless surrogates may not be estimated through a separate cohort.

The larger number of features required by *hybrid* models (see figure 9.12) is indicative of the greater complexity of estimating BP using LOSOCV as opposed to personalised regression. Additionally, the performance of the *hybrid calibration* strategy may have been limited by the very small ($N_p=13$) dataset size. A larger dataset may allow for calibration gradients to be approximated through interactions with an individual's demographics. However, the validity of BP estimations that rely on user demographics (such as age, sex, height etc) has been drawn into question [25]. The Aurora project, run by a team at Microsoft [217] and previously introduced in the literature review (see section 2.9.2), covered a much larger ($N_p = 483$) number of participants whose 24-hour BP was similarly estimated using *hybrid calibration* in a LOSOCV framework. The authors implemented the same *hybrid calibration baseline reference* ELAS+OLS model (using demographics, and time of day as features) as in this chapter, and found that it outperformed (as determined by the mean and standard deviation of error) linear and non-linear models. Collectively, these results suggest that current *hybrid calibration* strategies may not be sufficient for cuffless devices to adequately estimate absolute values of ambulatory BP.

While *individual calibration* models did result in reductions in RMSE and MAE values ($p < 0.01$ for both), notable estimation errors persisted in several participants. The ISO has proposed a preliminary validation error limit of 6 mmHg for cuffless devices, although the complete validation criteria are still in the development stage [27]. For seven participants, the RMSE of the RF_{PPG} model exceeded this 6 mmHg limit. One plausible explanation for these elevated error values could be the degradation of calibration parameters over the 7-day gap between recordings. However, with only two recordings per participant, assessing the severity this degradation was outside the scope of the dataset. Nevertheless, the fact that 24-hours of ABPM data (equating to over 30 measurements per participant) was inadequate for appropriate *individual calibration* in seven of the 13 participants, coupled with the already low tolerance for ABPM [23], raises considerable practical concerns for how to appropriately calibrate accurate BP estimation models in an out-of-clinic setting.

9.9.1.1 Relative changes in blood pressure

In the absence of accurate absolute measurements of BP, there are still important clinical benefits that can result from reliably tracking relative changes in BP. For example assessments of nocturnal BP patterns; variations in BP in response to changes in diet or medication [27]; or blood pressure variability (BPV) as an early indicator of CVD [85]. The ELAS+OLS models reported similar ρ_p values for *hybrid* and *individual* calibration strategies (and non-significant differences were observed between the two), resulting in equivalent DR_{SBP} values. Therefore, the results of this study suggest that it may be possible to estimate relative changes in BP using *hybrid calibration*.

Despite this, there is little evidence that current commercial devices, which implement *hybrid calibration*, are able to reliably track relative changes in ambulatory BP. In appendix A we provide results from a small ($N_p=5$) healthy volunteer study comparing 24-hour BP from a reference ABPM and estimated BP from an Aktiia device (a commercial device that uses a PPG signal recorded in reflectance mode at the wrist to report estimates of BP) [241]. In this study, we suggest that the Aktiia SBP estimates are primarily scaled HR estimates and that the device significantly underestimates any deviations from the initial calibration value. This leads to large SBP estimation error values (median RMSE = 11.19mmHg) which we can now attribute to the use of a *hybrid calibration* strategy. Additionally, the median correlation between the estimates and the reference was 0.67, leading to an underestimation of the DR_{SBP} ($p = 0.008$). Furthermore, we demonstrated in section 7.7.2 that HR (which we propose is the signal of primary importance in the Aktiia device) on its own is insufficient for reliable tracking of BP for individuals with $\mathcal{P}_{ND}$ and $\mathcal{P}_{RD}$ phenotypes.

These conclusions were recently corroborated by Tan *et al.* [26] using a larger cohort of 42 individuals. All participants wore the Aktiia device continuously for a period ranging from 6 to 12 days. At both the study's outset and conclusion, 24-hour ABPM was conducted. The study revealed a significant ($p < 0.001$) underestimation of nocturnal blood pressure fluctuations by the Aktiia device, leading to a notably higher proportion of participants being classified as non-dippers compared to conventional ABPM. Furthermore, the study included three hypertensive participants who wore the Aktiia device continuously for one week before and two weeks after uptitration of antihypertensive medication. The analysis demonstrated that, following two weeks of uptitration, the average medication-induced decline in SBP and DBP was $-1.0/-0.8$ mmHg when assessed with Aktiia, in contrast to $-19.7/-11.5$ mmHg observed with HBPM ($p = 0.03$ for the difference). This emphasises the Aktiia device's limitations in accurately tracking medication-induced BP changes.

9.9.2 Feature importance

Figure 9.15 demonstrates the notably high inter-participant variability in feature importance (as opposed to figure 6.8) that is indicative of the stochastic nature of data collected in an ambulatory environment and the effect of limited data on feature selection and CV. Indeed, intrinsic factors such as age, sex, and autonomic nervous system tone, significantly influence BP changes over 24 hours [2, 379, 382]. In addition, BP is affected by a variety of external factors such as activity levels [383], caffeine intake [342], heart rate variability [384], and sleep/wake cycles [5]. We found that, as a result, there was little agreement between the highest-ranking features in this study and the results of the previous chapters.

Computing PAT and PPG_{fts} across 24 hours, while individuals live freely, poses significant challenges. By nature of being recorded from two devices located separately on the body, PAT in particular is highly susceptible to motion artefacts that give rise to high-frequency noise in the signal [385, 172]. Equivalently, Hasanzadeh *et al.* [212] has previously highlighted significant changes in the PPG_{fts} extracted from PPG signals corrupted with Gaussian noise at varying signal-to-noise ratios. This warranted the development of an ensemble-averaged approach in this chapter, in order to provide a smooth approximation of the underlying PAT and PPG_{fts} signals during periods of movement.

However, the underlying PAT signal is additionally impacted by changes in the contact pressure between the pulse oximeter and the index finger as well as variations in hydrostatic pressure due to posture changes [200, 222]. Additionally, the influence of the PEP as well as the hysteresis effect outlined in chapter 5 remains to be explored in an ambulatory environment. As a result, unlike chapters 7 and 8, PAT_{EA} did not significantly contribute to BP estimation for *individual calibration*, suggesting that ambulatory BP estimation may be possible using only a PPG signal. Indeed, adding PAT_{EA} to the dataset led to greater errors, although this difference was non-significant.

Significant intra-participant variability of feature importance was additionally reported, wherein there was little to no agreement between the feature ranking coefficients learned from a participant's Day₀ data to the 'optimum' features determined when training equivalent models on Day₇ data (see figure 9.16). This poor correlation between Day₀ and Day₇ features, however, did not result in an associated significant detriment in estimation performance of *individual calibration*. It is likely that several possible feature subsets could prove 'optimal' and the use of alternative ML modelling or methods for feature selection may well identify a different subset that achieves similar performance [386]. As such, the stability of different feature selection algorithms and their impact on BP estimation needs to be explored in more detail in future studies.

In the wider context of cuffless ambulatory BP estimation, these findings are particularly noteworthy. The inter-participant and intra-participant variability of feature importance may have significant implications for training models that aim to generalise to the broader population, as well as models that account for the different mechanisms through which BP may be perturbed. Figures 9.18 and 9.20 for instance demonstrate how significant estimation errors may result from errors of inappropriate feature importance learning both on an individual and a cohort level respectively.

9.9.3 Limitations

Although remote cuffless monitoring of BP has many advantages such as facilitating a unobtrusive, high-frequency, assessment of cardiovascular health a number of confounding factors must be considered when taking measurements in real-world non-laboratory scenarios.

For instance, participant adherence to the study protocol varied, resulting in differing levels of data availability reported for each participant in table 9.1. Participants were instructed to carry on with their daily activities during the study, often involving tasks such as working on a computer, cooking, and eating. Consequently, abrupt changes in the PPG waveform due to hand movements were not uncommon. These factors collectively contributed to the limited availability of PPG_{EA} , PPG_{fts} , and PAT_{EA} signals throughout the study sessions. Furthermore, there were instances where participants incorrectly positioned the PPG pulse oximeter on the wrong hand or finger for short durations. Additionally, BP was recorded at a relatively low sampling frequency (every 30 minutes during the day and every hour during the night), making it challenging to detect and filter out values corrupted by noise. These limitations necessitated the use of a double-cosinor model to attenuate high frequency variations in both the reference BP signal, and the estimated BP signal.

To manage noisy data, a simple participant activity threshold was used to label time periods as 'Active' ($\approx 10\text{-}20\%$ of each session). 'Active' periods were removed from the analysis of the PPG and ECG signals. This was effective in removing significantly noisy periods but did not take into account the impact of increasing activity levels on an individual's circadian rhythm [383] and so activity levels may play a role in cuffless ABPM.

The small number of participants recruited in this study ($N_p=13$) reflect the practical difficulties in participant recruitment and running of the study and overall it must be considered that nABPM was a feasibility study. Such small numbers, however, limit the generalisability of our conclusions. Importantly, the leave-one-subject-out cross-validation *hybrid calibration* approach was significantly confounded by the low subject numbers and any inherent diversity (including different ages, sex, ethnicity, and skin colour, all factors known to impact the PPG contour independent of BP [28]) within the cohort.

9.10 Conclusion

In this chapter, we presented the results of a healthy volunteer study designed to investigate ambulatory BP estimation 7 days after initial calibration. While *hybrid calibration* strategies are significantly more convenient (requiring only a single calibration value for an individual), they were shown to result in clinically significant errors when estimating absolute values of BP, and therefore should be used with caution. *Individual calibration*, through an ABPM training set, was shown to reduce the reported error values ($p < 0.01$), however errors greater than the 6 mmHg preliminary ISO validation limit persisted for seven participants. Moreover, we demonstrated significant inter- and intra-participant variability in feature importance and recommend that the stability of PPG_{fts} (and its impact on BP estimation) be explored in future studies.

This chapter highlights the difficulty of ambulatory BP estimation using the ECG and PPG signals. Future work will aim to expand these approaches towards greater generalisability and robustness.

10

Conclusions and future work

10.1 Introduction

The core objective of this thesis (as presented in chapter 1) was "To investigate PPG- and ECG-derived surrogate measures of ambulatory BP". Conventional ABPM measurements have significant limitations regarding: accessibility, frequency of measurements, and obtrusive cuff inflations. Cuffless BP monitoring using PPG and ECG signals aims to provide solutions to the above issues. It is thought that in doing so, and embedding the relevant technology into wearable devices or smartphones, the awareness and early treatment of hypertension, as well as reverse-dipping nocturnal patterns, in the general population would dramatically improve [27].

This chapter concludes the thesis, beginning with a review of the main results, contributions, and limitations of each chapter. We then present avenues for future work based on the overall conclusions presented throughout this thesis. Finally, we provide our concluding remarks.

10.2 Thesis review

10.2.1 Clinical datasets

In chapter 3 we introduced the three datasets employed in this thesis: the MOLLIE study, the MIMIC database, and the nABPM study. The MOLLIE study involved 26 healthy volunteers who underwent a 28-minute phenylephrine infusion that resulted in significant changes in BP values. The participant-wise median change in SBP, DBP, and MAP from their baseline values to the point of maximum infusion rate was 20.0 mmHg, 17.0 mmHg, and 15.5 mmHg respectively. As participants remained supine and still, this dataset provided an ideal setting for the initial investigation and development of algorithms for cuffless BP estimation.

MIMIC is an open-source database containing anonymised physiological data from a large, heterogeneous, ICU patient cohort [168]. Although it is not the objective of this thesis to propose cuffless BP measurement in the ICU, analysis of the MIMIC data facilitated a robust assessment of circadian rhythms and BP phenotype classification. We restricted our analysis to a subset of 1,058 patients whose vital-sign profiles were most likely to be comparable to non-hospitalised individuals.

The nABPM study was designed to extend the analysis developed throughout this thesis towards cuffless ambulatory BP estimation in healthy volunteers, 7 days after initial calibration. 13 participants took part in our study, which consisted of two 24-hour sessions spaced a week apart, during which participants carried out their regular activities while physiological data (including ABPM, PPG, ECG, and acceleration) was recorded.

10.2.2 Pulse arrival time as a surrogate measure of blood pressure

In chapter 5, we evaluated the reliability of PAT as a surrogate measure of BP in the MOLLIE study. In doing so, we developed a robust signal analysis pipeline to derive estimates of PAT and PEP, which accounted for periods of low signal quality through SQIs and a Kalman filter. PTT was calculated as the difference between PAT and PEP. We compared the performance of TT derived BP estimation using different population-based models to *a posteriori individual calibration* models.

PAT and PTT were found to have strong, negative, linear correlations with SBP ($\rho_p = -0.85$ and -0.90 , respectively), MAP ($\rho_p = -0.82$ and -0.88 , respectively), and DBP ($\rho_p = -0.83$ and -0.85 , respectively). PEP was found to be a non-negligible but relatively constant contribution to PAT values. Thus, changes in PTT accounted for most of the changes in PAT in the study.

Population-based models resulted in large error values, while *a posteriori* models produced clinically acceptable error values (RMSE < 6 mmHg). The results of a likelihood ratio test validated the need for unique, individualised TT-BP calibration slopes. We also identified a hysteresis effect whereby the TT-BP calibration slope was more shallow during the dose increase phase than the washout phase of the study ($p < 0.05$), with the effect being more pronounced for PAT than PTT, suggesting a contribution from PEP.

The conclusions of this chapter are subject to some notable limitations. Firstly, the study's participant cohort was relatively small and homogenous, comprising only 26 healthy individuals. This limited sample size, determined by practical constraints rather than power calculations, raises questions about the extent to which the findings can be extrapolated to a more diverse population. Additionally, the use of phenylephrine to perturb BP, while a controlled approach, may not fully represent the complex physiological mechanisms governing BP changes in daily life, potentially limiting the study's real-world applicability.

10.2.3 Features from the photoplethysmogram and electrocardiogram for estimating changes in blood pressure

In chapter 6, we investigated whether changes in BP (ΔBP) may be estimated through morphological features of the PPG and ECG waveforms (PPG_{fts} and ECG_{fts} respectively), using the MOLLIE

study data. We extracted a comprehensive set of PPG_{fts} and ECG_{fts}, identified from previous relevant publications. Changes in the morphology of the PPG, VPG, and APG waveforms were largely characterised by estimating the relative amplitudes and time delays of the characteristic fiducial points. The PPG morphology was further parameterised through Gaussian decomposition and principal component analysis. Changes in the morphology of the ECG waveform were characterised using complexity features.

We employed a *hybrid calibration* strategy, using a leave-one-subject-out cross-validation (LOSOVC) framework, to estimate ΔBP for each individual. Based on the results of the previous chapter, it is likely that *individual calibration* would have improved the results. However, as the recording period for each participant was short (28 minutes), it was not possible to implement such a strategy using high-dimensional, personalised, ML models in the MOLLIE dataset.

Feature collinearity was accounted for by removing features with a variance inflation factor greater than 10. The non-linear RF model significantly ($p < 0.05$) outperformed the linear LASSO+OLS model using both PPG_{fts} and ECG_{fts} across all performance metrics. Estimating ΔSBP using PPG_{fts} alone ($\rho_p = 0.86$, RMSE = 5.66 mmHg, MAE = 4.86 mmHg) performed significantly better than using ECG_{fts} alone ($\rho_p = 0.69$, RMSE = 6.79 mmHg, MAE = 5.28 mmHg), all $p < 0.001$.

The highest ranking PPG_{fts}, as assessed through SHAP values and a ranking coefficient, were associated with increased reflected wave interference driven by changes in arterial stiffness. This finding was supported by changes observed in the PPG waveform in response to the phenylephrine infusion. We concluded from this work that PPG_{fts} may be a promising indicator of ΔBP and further exploration is warranted. Our results indicate that the use of ECG_{fts} is not justified.

This chapter's limitations include dataset constraints due to a small number of participants, hindering the exploration of *individual calibration* strategies. The single-perturbation study design may introduce bias, affecting generalisability to ambulatory settings. Furthermore, the conventional systolic peak definition in PPG signals may not be suitable for scenarios with motion artefacts, and cuff accuracy can introduce consistent offsets in BP estimation.

10.2.4 Analysis of the circadian rhythm of pulse arrival time and photoplethysmogram features

In chapter 7, we investigated whether PAT and PPG_{fts} exhibited endogenous 24-hour oscillations in synchronicity with that of BP. This analysis was conducted in the MIMIC dataset, and resulted in, to our knowledge, the largest heterogeneous cohort of individuals from which cuffless surrogate BP measures were investigated over 24 hours. Statistical, double-cosinor, modelling was employed to analyse variations from the computed MESOR (the midline around which the signals oscillate).

At the population level, the circadian rhythms of HR and RR showed inconsistent relationships with the 24-hour variation of SBP across BP phenotypes. With 88.2% of patients exhibiting a 24-hour rhythm (as computed by a goodness-of-fit F-test), we concluded that PAT has its own endogenous, phenotypically modified, circadian pattern that aligns closely ($|\rho_p| > 0.80$) with that of SBP. The circadian pattern of five PPG_{fts} also reported strong ($|\rho_p| > 0.70$) linear correlations with SBP across all phenotypes. Additionally, PAT and six PPG_{fts} exhibited more than two significant ($p < 0.01$) differences in their circadian double-cosinor features between patients in the $\mathcal{P}_D$ and $\mathcal{P}_{RD}$ phenotype cohorts. Wide $CI_{95}^{\theta_{Orth}}$ and $CI_{95}^{\theta_{Bathy}}$ limits were reported for all PPG_{fts}, however, suggesting that a similar relationship was not present at an individual level.

This chapter's conclusions should be considered in light of several key limitations. The ICU environment, with its frequent medical interventions and irregular sleep patterns, may not be ideal for circadian rhythm analysis. Patients' physiological stress, short ICU stays, and potential medication effects further complicate the interpretation of circadian patterns. Moreover, distinguishing between circadian and diurnal patterns is challenging in a clinical setting with irregular schedules and medical interventions.

10.2.5 Classification of blood pressure phenotypes

In chapter 8, we investigated the utility of the PPG and ECG waveforms for classifying an individual's BP phenotype. Due to dataset requirements this analysis was also performed on the MIMIC dataset. We proposed two novel methods that overcome many of the challenges associated with ABPM in primary care settings.

Our first proposed method ('regularised *individual calibration*') used data recorded during the period between 8am and midnight (a typical monitoring period for HBPM) to train mixed-effects regression models (LMEM and MERF) for patient-specific cuffless estimation of nocturnal SBP. Non-significant differences were observed between the LMEM and MERF models in both the regression (predicting nocturnal SBP) and classification (predicting BP phenotype) analyses. The LMEM_{PPG+PAT} model reported an overall median phenotype classification performance of 0.52 in wF1, 0.54 in Acc, 0.23 in κ , and 3.6% in SMR (*severe-misclassification rate*, the proportion of patients in the $\mathcal{P}_D$ cohort that were incorrectly classified as $\mathcal{P}_{RD}$, and vice versa), all outperforming the demographic-based *baseline reference* ($p < 0.001$). However, we encountered challenges in mixed-effects modelling due to the presence of noise in the dataset that could not be accounted for when performing a *post-hoc* assessment.

Our second proposed method ('calibration-free') did not require any BP calibration measurements instead utilising circadian features of the relevant signals. The RF_{PPG+PAT} ordinal

classification model reported an overall median phenotype classification performance of 0.52 in wF1, 0.55 in Acc, 0.24 in κ , and 5.6% in *SMR*, all outperforming the demographic-based *baseline reference* ($p < 0.001$). The feature importance analysis using SHAP values showed that the model heavily relied on $\tilde{D}R$ and θ_{Ortho} across all signals. Utilising signal phases instead of amplitudes provides several advantages, such as eliminating the requirement for personalised amplitude calibration and enhancing resistance to signal noise.

The two methods reported similar classification performance metrics and non-significant differences were observed for the values of wF1, Acc, and κ . The only significant difference was reported in the *SMR*, $p < 0.001$. Both methods struggled to classify $\mathcal{P}_D$ patients accurately (recall Method 1 = 15.8%, Method 2 = 12.4%), with the majority (Method 1 = 62.5%, Method 2 = 61.1%) of 'dippers' being labelled as 'non-dippers'. Conversely, a relatively high recall (Method 1 = 71.8%, Method 2 = 69.9%) of the $\mathcal{P}_{RD}$ phenotype was observed. Early detection of this phenotype is vital to reduce the health burden of cardiovascular diseases. The calibration-free method in particular, therefore, may provide a cost-effective initial assessment of dipping status in the wider population.

The results of this chapter provide the first step towards cuffless BP phenotype estimation. We anticipate further improvements in performance through more longitudinal recordings of the relevant signals.

The chapter's limitations stem from the ICU dataset's uniqueness and constraints, which may not align with real-world scenarios. The subjectivity in labelling patient-wise BP phenotypes introduces uncertainty due to potential misclassifications caused by minor BP measurement errors. Patient selection based on circadian rhythm assumptions may bias the study towards individuals with regular rhythms, excluding those with irregular patterns. Translating methods from the ICU to at-home monitoring faces challenges like patient movement, adherence variations, and signal quality differences.

10.2.6 Towards cuffless estimation of ambulatory blood pressure

In chapter 9, we further developed the methodology implemented throughout this thesis towards the goal of cuffless ambulatory BP estimation, using the nABPM study data. In particular, we compared the performance of *individual calibration* (personalised regression) with *hybrid calibration* (leave-one-subject out) strategies for accurate estimations of both absolute values of BP and relative changes in BP in free-living individuals. ELAS+OLS and RF regression models were trained on data recorded during Day₀ and evaluated on data recorded during Day₇.

Poor-quality BP values were detected based on disagreements between HR_{Ref} and HR_{Cuff} . Subsequently, the BP time series were smoothed using double-cosinor modelling to attenuate

high-frequency components. Valid estimates for PAT_{EA} and PPG_{EA} PPG_{fts} were available for approximately 50%-90% during each 24-hour session, depending on the participant, their activity levels, and the time of day (higher availability of data at night-time).

We demonstrated significant inter-participant and intra-participant variability in feature importance and suggest that PPG_{fts} stability needs to be explored in more detail in further studies. Our analysis also found that PAT_{EA} did not significantly contribute to model estimates using either calibration method, suggesting that cuffless estimation of ambulatory BP using only the PPG signal may be feasible.

The results showed that *individual calibration* resulted in significantly ($p < 0.01$) lower errors (RMSE and MAE) for estimating absolute values of SBP in comparison to *hybrid calibration*. The absolute difference between the *individual calibration* RF_{PPG} model and the *hybrid calibration* $ELAS+OLS_{PPG+PAT}$ model (the two models that reported the lowest MAE for each calibration strategy) was 5.28 mmHg ($p = 0.0013$) for RMSE and 5.17 mmHg ($p = 0.008$) for MAE. However, both strategies were able to track relative changes in ambulatory BP equivalently and non-significant differences for ρ_p and DR_{SBP} estimation were reported. Based on the analysis performed in this chapter, we suggest that *hybrid calibration* strategies, as they currently exist, may result in clinically significant errors in BP measurements (for example, a healthy individual being classified as having nocturnal hypertension).

While *individual calibration* may improve performance it is not without its limitations, namely a requirement for multiple BP measurements from a single individual, spanning a clinically useful range of BP values. Indeed, personalised regression was not sufficient for all individuals with the RMSE for seven participants exceeding 6 mmHg.

This chapter's limitations are primarily driven by the study design and recruitment. Participant adherence varied, affecting data quality as daily activities, like computer work and cooking, caused abrupt PPG waveform changes due to hand movements. Sensor misplacement and low BP sampling frequency (30 minutes during the day, hourly at night) introduced noise, necessitating a double-cosinor model for high frequency attenuation. Noise management involved labelling 'Active' periods (≈ 10 -20% of each session) but didn't address circadian rhythm variations. A small participant group ($N_p=13$) limited generalisability, and diverse factors within the cohort complicated leave-one-subject-out cross-validation for *hybrid calibration*.

10.3 Future work

The conclusions and limitations presented in this thesis motivate several avenues for future work. We split the additional avenues of future work into four, related, research themes: "Study design and considerations", "Predictive modelling", "Developing a deeper understanding of PPG_{fts}", and "Cuffless blood pressure device considerations".

10.3.1 Study design and considerations

As a result of practical challenges surrounding study design and participant recruitment, the MOLLIE and nABPM feasibility studies were restricted to low, and relatively homogeneous, cohort sizes ($N_p = 26$ and 13 respectively). For example, the nABPM study did not include any participants whose BP increased overnight ('reverse-dippers'). It is unclear, therefore, whether the conclusions made using these studies may be generalised to the broader population. While the MIMIC dataset addressed these limitations, it only allows for a retrospective analysis of critically ill patients, who are not the target population for cuffless BP monitoring.

Future studies should aim to expand both the sample size and the diversity of the recruited cohort, ensuring a more representative distribution of patients across the three major BP phenotypes. This step is crucial for advancing cuffless BP monitoring from a consumer-grade device to a clinically relevant tool. By doing so, we can enhance the applicability of the conclusions derived from these studies to a wider and more clinically relevant population.

10.3.1.1 Recommended future studies

Based on the work concluded as part of this thesis, we recommend two future studies be conducted in order to expand the research on cuffless BP monitoring in a meaningful direction.

Firstly, the necessity for an *individual calibration* procedure significantly diminishes the convenience of cuffless BP monitoring. Moreover, current devices recommend repeating this calibration protocol every month [24], making it unlikely that the general population would adopt such devices. Therefore, it is recommended that future research further investigates the 'regularised *individual calibration*' method introduced in chapter 8 alongside the objectives of the NHS Blood Pressure @home programme [113], which aims to distribute home BP monitors to individuals with suspected hypertension. This approach may allow for the development of a personalised calibration dataset that balances accuracy and user comfort over a period equivalent to the current HBPM protocol (4-7 days, as per [102]).

Secondly, a longitudinal analysis of cuffless BP surrogates (in particular their circadian phases) may uncover early indicators of the progression of cardiovascular or neurodegenerative diseases

that are currently not feasible through ABPM. For example, notable deviations from an individual's 'typical' circadian rhythm trajectory may be an early indicator of disease progression [387]. A recent meta-analysis [388] suggests that 32% (CI_{95} : 26%-38%) of individuals exhibit inconsistent BP phenotypes from day to day (i.e. on one day they may be a 'dipper' and on another day they may be a 'non-dipper'). Cuffless BP monitoring may allow for further research on the link between these transitions and cardiovascular events over a longer period of time [389].

10.3.2 Predictive modelling

Throughout this thesis, we have restricted the use of ML models to a comparison of regularised linear models and non-linear RF models, with the aim of prioritising the interpretability of our analysis. Although this approach may have limited the performance of our models, it provided a clear understanding of the relationships between the feature set and the target values. It is worth noting, however, that numerous other data-driven models have been employed in the literature to relate PAT or PPG_{fts} to BP [9]. These approaches, if applied to our datasets, could have resulted in improved performance, as well as derived additional insights over and above those presented.

Tree-based algorithms such as RF can be further improved through the use of gradient-boosted models such as AdaBoost or XGBoost [390]. These algorithms enhance the performance of decision trees by combining weak trees in a sequential manner, whereby each new tree corrects the residual errors of the previous tree. Gradient-boosted tree algorithms are thought to produce lower errors than RF, but they also tend to be more susceptible to outliers in the data as they increasingly focus on large residual values [391].

Additionally, it is important to consider the dependence of intra-individual observations of BP. The models implemented in this thesis failed to account for the time course of BP variations. Long Short-Term Memory (LSTM) models in particular are capable of incorporating temporal information and learning long-term relationships between features and the target vector [392]. Indeed, LSTMs have been successfully applied to 24-hour BP estimation by Radha *et al.* [187] and Lee *et al.* [393].

The PAT and PPG_{fts} signals analysed in this thesis may be combined for the estimation of an individual's BP phenotype through techniques such as multivariate Gaussian process regression [394, 395], coupled with activity-based hidden Markov modelling [396]. Bayesian approaches, such as Gaussian processes, enable the incorporation of prior knowledge through parameter distributions and subsequently update this knowledge using available data to compute posterior distributions. This process can enhance circadian rhythm modelling by utilising a prior circadian profile derived from either individual ABPM data or a profile from a similar population. Bayesian methods

also provide an estimate of model uncertainty, which could provide additional information to clinicians as they interpret the output of the model.

The use of deep learning algorithms has been a recent development in the field, with studies by Slapnicar *et al.* [213], Schlesinger *et al.* [214], and Liang *et al.* [397] showing that they can outperform classical ML techniques, such as those used in this thesis. Deep learning is highly effective in detecting patterns in data and can be trained directly on raw time series data, avoiding the need for feature extraction and selection. However, it requires large amounts of data for accurate predictions, and the availability of such data for continuously estimating BP in all scenarios remains doubtful. Furthermore, the 'deeper' the model, the less interpretable it becomes as the data is transformed in increasingly abstract ways.

Further insights may be derived from additional ML applications such as novelty detection and clustering. The estimation of BP, particularly in an ambulatory environment, suffered from a high susceptibility to noisy data, largely due to the presence of motion artefacts. Novelty detection, a subset of ML that concentrates on identifying outliers in data, could be applied using regression-based methods [398].

Clustering is an unsupervised ML task that groups data into distinct clusters based on a metric of similarity [377]. There are two tasks relevant to this work that could be addressed through clustering techniques. Firstly, as the time and level of discomfort required to collect adequate data to train an *individual calibration*, personalised, model for BP estimation for each individual may be high, the incorporation of clustered individuals who behave in a similar manner may have substantial benefits. Secondly, we noted in chapter 8 that the thresholding of DR_{SBP} for ground-truth phenotype labelling may be arbitrary. Clustering algorithms may be applied to uncover data-driven circadian BP phenotypes, which may provide more meaningful insights into cardiovascular diseases [15, 55, 61].

10.3.3 Developing a deeper understanding of photoplethysmogram features

In this thesis, changes in the PPG waveforms were characterised by extracting 61 morphological PPG features from each pulse. The PPG-BP relationship was then modelled using a data-driven approach as opposed to statistical models developed based on physiological principles. This data-driven approach requires a 'high-quality' calibration training dataset to establish the PPG-BP relationship. Developing a deeper understanding of the factors impacting PPG_{fts} could help to constrain the model [190, 9]. Additionally, factors such as the PPG probe contact pressure on the skin, internal device filtering, and hydrostatic pressure changes [222, 200, 9] can impact the morphology of the PPG but are rarely considered in the modelling process.

For a few PPG_{fts}, we observed inconsistencies in their directional relationship with BP. For instance, Kurtosis showed a positive relationship with SBP in chapter 6 ($\rho_p = 0.55$), while the population-level circadian rhythm of Kurtosis exhibited a negative relationship with SBP in chapter 7 ($\rho_p = -0.37$). This discrepancy may be attributed to the different mechanisms through which BP was perturbed. To develop a generalisable approach for BP estimation, further research is necessary to understand the impact of various BP change mechanisms on the relationship between PPG and BP and to incorporate this knowledge into the modelling framework.

Furthermore, there remains a lack of consensus on the 'optimum' PPG_{fts} for adequate BP estimation. In chapter 9, we further emphasised this point by demonstrating that the 'optimal' PPG_{fts} for an individual may fluctuate on a daily basis. These findings are corroborated by the works of Sun *et al.* [215] and Lin *et al.* [315] respectively. To advance the development of a robust BP estimation technique, issues such as the stability of PPG_{fts} selection algorithms and the impact of this selection on estimation accuracy require further investigation [399, 386].

Finally, the requirement for recalibration of the PPG-BP model should also be thoroughly evaluated on both physiological [156] and experimental [210, 136] grounds.

10.3.4 Cuffless blood pressure device considerations

In this thesis, we adhered to the instrumentation requirements outlined in section 3.2 emphasising that both the PPG and ECG signals were to be recorded using clinical grade equipment, on a synchronised internal clock, in order to ensure precision in signal recording. As a result, in all studies, PPG signals were recorded using a transmission-mode finger pulse oximeter and the ECG signals were recorded using wet electrodes in electrical contact with the skin surrounding the heart. While this choice ensured signal accuracy, it did limit the portability of our data collection setup. Indeed, a number of participants reported that the Stowood device employed for the nABPM study (see figure 3.5) was too cumbersome to wear during daily tasks.

While we recommend that future studies record more longitudinal data from individuals, enhancing user comfort is a greater priority for device design. It is clear that any cuffless device should be comfortable enough for an individual to wear for at least 24-hours without significantly impeding on their daily life. In particular, a number of wearable devices can record a PPG signal in reflectance mode such as smartwatches [146, 24], rings [400, 401], and smartphone cameras [9], the latter being based on colour fluctuations in the skin (known as iPPG) [220, 263]. However, due to the smaller fraction of light reaching the photodiode, the quality of the reflectance mode PPG signal is known to be considerably lower than for transmission-mode photoplethysmography [116]. Additionally, as demonstrated by Hartmann *et al.* [402], the differences in the travelled

arterial path for the two modes gives rise to differences in the morphology of the PPG waveform. Whether the conclusions drawn from this work hold for PPG signals recorded in reflectance mode remains to be determined.

Several commercially-available wearable smartwatch devices, including the Apple Watch, Fitbit Sense and Withings ScanWatch, have the capability to capture a single-lead ECG signal using capacitive-coupling electrodes instead of wet electrodes [403, 404]. Typically, one electrode is placed on the back of the watch, in contact with the wrist, and another electrode is placed on the front of the watch. To initiate an ECG recording, the user must place their opposite hand on the front electrode and, therefore, currently, these devices are unable to passively record ECG during sleep. Alternatively, an accurate ECG recording can be achieved using a chest-worn patch like the VitalPatch [405, 406]. While this enables passive ECG recording, it poses a challenge as the ECG and PPG would be recorded through separate devices, leading to time synchronisation issues.

10.3.4.1 Other modalities

There exist several other modalities and techniques that have the potential to yield valuable insights into an individual's cardiovascular health beyond analysis of the PPG and ECG waveforms. For example, wearable devices are capable of capturing balistocardiogram (BCG) signals [407, 408], as well as applanation tonometry [217], and phonocardiogram (PCG) signals [409, 410].

10.4 Concluding remarks

To conclude, the potential for PPG and ECG waveforms to enhance cardiovascular health monitoring and address the widespread issue of CVD is significant. Throughout this thesis, signal processing and machine learning algorithms have been developed to support this assertion, and their applications have been explored across various settings.

In the controlled, single perturbation, environment of the MOLLIE study, both PAT and morphological features of the PPG enabled reliable and accurate BP measurements. This work suggested that cuffless beat-by-beat blood pressure monitoring using the ECG and PPG could be viable with appropriate calibration.

However, as we transitioned to real-world ambulatory monitoring, several challenges emerged. The nABPM study demonstrated that estimated ambulatory BP values were notably affected by motion artifacts, inherent intra- and inter-individual variability in PPG features, and inadequate training data. Furthermore, the insufficiency of 24-hour ABPM data, amounting to over 30 measurements per participant, for proper *individual calibration*, combined with a general low

tolerance for ABPM, highlights significant practical issues in developing accurate, out-of-clinic beat-by-beat BP estimation models.

Nevertheless, the potential for the ECG and PPG signals to facilitate macro assessments of cardiovascular health remains promising. For instance, within this thesis, we have introduced a novel, calibration-free method for classifying BP phenotypes using circadian features of the PPG and ECG-derived BP surrogates. While macro assessments of cardiovascular health are often overlooked by the clinical community due to the challenges of long-term BP recording outside the clinic, they still hold important prognostic value that is currently being overlooked. Given that these assessments may not require calibration and can be computed over extended periods, it is plausible that many of the limitations observed in the nABPM study concerning beat-by-beat monitoring could be addressed through this methodology.

In this concluding chapter we have outlined the recommended further technical and clinical work required for advancement in this critical field of study.

Appendices



Comparison between Aktiia and ambulatory blood pressure monitor

Contents

A.1	Introduction	259
A.2	Methods	260
A.2.1	Dataset	260
A.2.2	Data processing	260
A.3	Results	261
A.3.1	Dataset	261
A.3.2	Overview	261
A.3.3	Vital sign estimation	261
A.3.4	DR _{SBP} estimation	262
A.4	Discussion	262
A.5	Conclusions	264

A.1 Introduction

A number of companies have made commercial claims of a cuffless, non-invasive, blood pressure (BP) monitoring device. The most prominent example is the Aktiia bracelet, depicted in figure A.1(a). The Aktiia bracelet is a consumer wrist wearable proposing to enable cuffless and continuous BP measurement [24]. The wrist is illuminated using a set of light-emitting diodes (LEDs) and the light scattered by the tissues is recorded using a photodiode placed next to the LEDs (a reflection-mode PPG). From the morphology and relative timings of the averaged pulse waveforms, Aktiia derives estimates of systolic blood pressure (SBP) and diastolic blood pressure (DBP). The wrist-worn wearable optical sensor is usually calibrated with three consecutive cuff measurements (on the opposite arm) and the calibration is reported to hold for 2 or 3 weeks [411].

We ran a small healthy volunteer clinical study in order to test the performance of the Aktiia bracelet against validated reference measurements from an ambulatory blood pressure monitor (ABPM), depicted in figure A.1(b). In this appendix, we present the results of this study.



Figure A.1: Devices used in this study. (a) Aktiia bracelet, image from Aktiia website. (b) ABPM device (AND 244-1), image from AND website.

A.2 Methods

A.2.1 Dataset

The dataset employed in this paper is from our healthy volunteer study: “Reverse dipping in blood pressure and its impact on screening for hypertension in the clinic” (CUREC approval reference number R77097/RE002). Simultaneous data was gathered over a 24-hour period from a TM-2441 ABPM device (A&D Instruments Limited), set to inflate every 30 minutes during the day and every hour during the night, and an Aktiia wrist-worn blood pressure monitoring device. Recorded vital signs included SBP, DBP, and heart rate (HR). At the start of each recording session, the Aktiia device was calibrated as recommended by the manufacturer. Participants then went about their typical daily life.

A.2.2 Data processing

The Aktiia device does not record data at a regular frequency but rather at inconsistent time periods, with typically one data point within each hour. We assume that this inconsistency is driven by changes in signal quality, but there is no official documentation for this. The user is then presented with their 24-hour profile as averaged values in 2-hour bins. We, therefore, compared Aktiia and ABPM vital-signs averaged in 2-hour bins. The only exclusion criterion applied to the ABPM data was to discard any reading in excess of 180 mmHg. We additionally compared the DR_{SBP} computed from the Aktiia and the ABPM in order to answer the question of whether the Aktiia device can provide a non-invasive assessment of circadian rhythms of blood pressure. We used the following metrics to evaluate the agreement between Aktiia and reference ABPM: Pearson’s correlation coefficient (ρ_p), root mean squared error (RMSE), and mean absolute error (MAE).

Table A.1: Demographics of the participants in the study.

Descriptor	Value
Total number of participants	5
Male	3
Female	2
Age (years) ¹	29.0 (4.5)
Height (cm) ¹	168.0 (24)
Weight (kg) ¹	66.0 (27.9)
Total number of sessions	8

¹ Participant-wise median (inter-quartile range (IQR)).

A.3 Results

A.3.1 Dataset

The demographics of the participants are shown in table A.1. A total of 5 individuals made up the dataset with one individual recording 3 days of data, and 1 individual recording 2 days of data, for a total of 8 recorded 24-hour profiles.

A.3.2 Overview

Figure A.2 shows the SBP, DBP, and HR recorded from the Aktiia device and reference ABPM for a full 24-hour period from a typical participant. The ABPM profile followed a typical circadian pattern with a dip at night and a rise in the morning. There was very little variation seen in the Aktiia estimated SBP and DBP from their initial calibration value. HR computed from the Aktiia was able to follow the changes recorded from the ABPM.

A.3.3 Vital sign estimation

Table A.2 provides the results of vital sign estimation using the Aktiia bracelet. HR estimation was consistently accurate. SBP and DBP had good median ρ_p to their reference measurements, however, reported poor RMSE and MAE values. Figures A.3 to A.5 provide a closer interpretation of the results for SBP, DBP and HR respectively via a scatter and Bland-Altman analysis.

Table A.2: Performance statistics of vital sign (SBP, DBP, and HR) estimation using the Aktiia device compared to an AND 2441 (ABPM) device. Results are given as median (IQR) computed across all volunteers.

Vital sign	ρ_p	RMSE	MAE
SBP (mmHg)	0.67 (0.35)	11.19 (3.94)	8.42 (3.8)
DBP (mmHg)	0.72 (0.08)	6.45 (4.19)	5.37 (3.6)
HR (bpm)	0.86 (0.11)	4.66 (2.32)	3.89 (2.12)

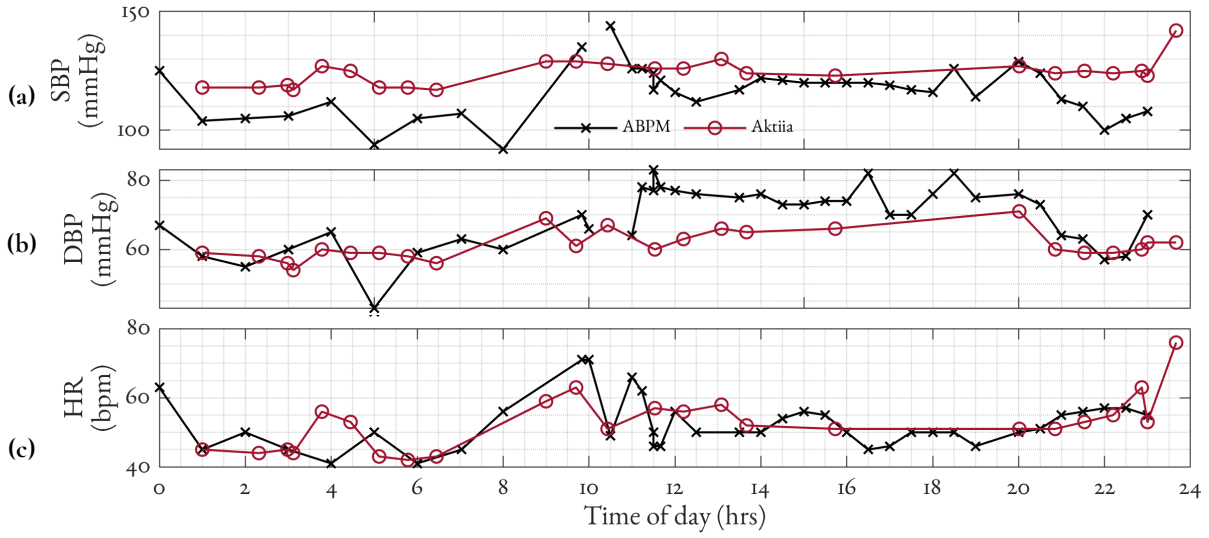


Figure A.2: 24-hour profile of vital signs recorded by the ABPM and Aktiia device. (a) SBP, (b) DBP, and (c) HR.

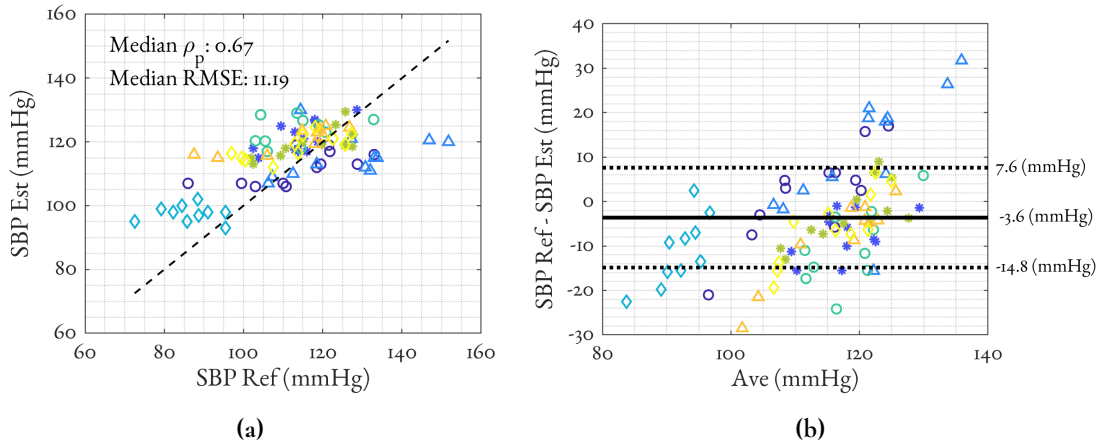


Figure A.3: Results of estimating SBP using the Aktiia bracelet. (a) Scatter plot showing a median ρ_p of 0.67 and RMSE of 11.19 mmHg. (b) Bland-Altman analysis showing a bias of -3.6 mmHg. Each volunteer is marker and colour coded.

A.3.4 DR_{SBP} estimation

Figure A.6 shows the range of DR_{SBP} values computed from the reference ABPM and the Aktiia bracelet. The Aktiia significantly ($p = 0.008$) underestimated the dip ratios.

A.4 Discussion

It was expected that the Aktiia device would accurately record HR. HR estimation via a wearable with a reflectance mode PPG signal has been around for a number of years [412]. The strong relationship found here, therefore, validated our approach in comparing the vital sign measurements from the two devices. We found high error values for both SBP and DBP estimation. Figures A.3

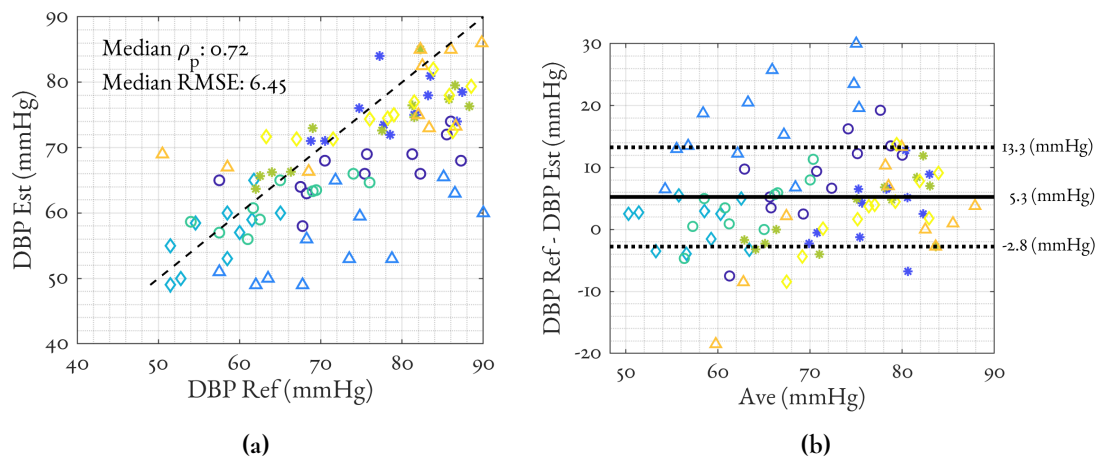


Figure A.4: Results of estimating DBP using the Aktiia bracelet. (a) Scatter plot showing a median ρ_p of 0.72 and RMSE of 6.45 mmHg. (b) Bland-Altman analysis showing a bias of 5.3 mmHg. Each volunteer is marker and colour coded.

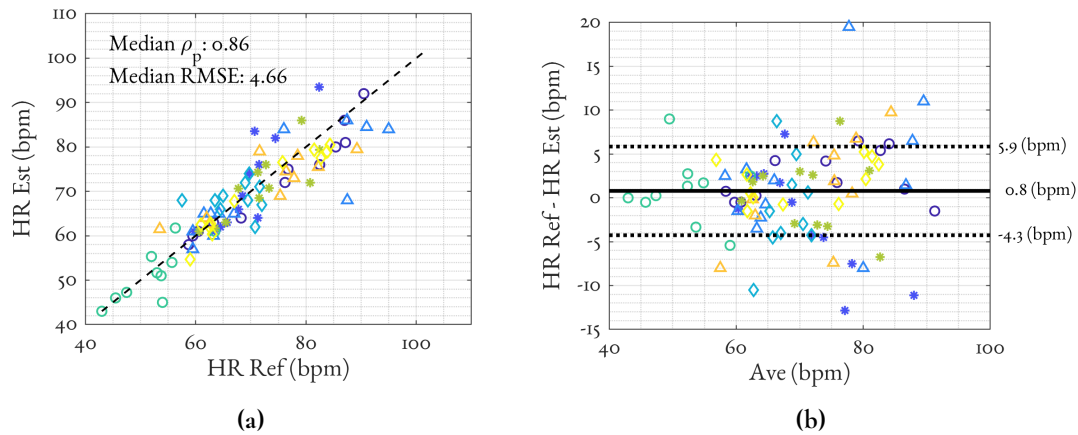


Figure A.5: Results of estimating HR using the Aktiia bracelet. (a) Scatter plot showing a median ρ_p of 0.86 and RMSE of 4.66 bpm. (b) Bland-Altman analysis showing a bias of 0.8 mmHg. Each volunteer is marker and colour coded.

and A.4 suggest that two different forms of discrepancies between the reference and estimated values exist for SBP and DBP. For SBP, the intra-volunteer variability was low and it is clear that the estimated values do not change significantly from their calibration value. Further analysis has additionally shown strong correlations (median $\rho_p = 0.93$) between estimated SBP values and Aktiia estimates for HR suggesting that for SBP the device doesn't estimate much more than a scaled HR. For DBP estimates, the correlation to HR estimates were lower (median $\rho_p = 0.67$). Figure A.4 shows much larger intra-volunteer variations suggesting that estimated values do deviate from the calibration value. However, large errors were still reported. DR_{SBP} estimation additionally showed considerable errors with dip ratios being significantly ($p = 0.008$) underestimated by a factor of at least 2.

These results underpin the recent statement from The European Society of Hypertension

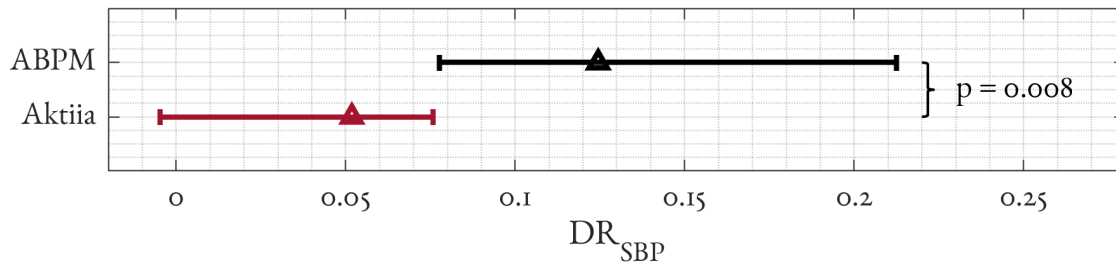


Figure A.6: DR_{SBP} estimated values using the Aktiia bracelet and the reference ABPM. The triangles show the median values, the whiskers show the range. A significant difference was observed between the two (Wilcoxon signed-rank, $p = 0.008$).

Working Group on Blood Pressure Monitoring and Cardiovascular Variability [27], which stresses that none of the currently available cuffless devices should be used for the diagnosis or management of hypertension. The Aktiia documentation suggests that the device only requires re-calibration after 2 or 3 weeks [411]. Our results show large errors 24 hours after initial calibration, we, therefore, suggest that the error values may be even larger after 3 weeks.

A.5 Conclusions

The Aktiia bracelet, reports to be a clinically validated device to measure BP. However, as mentioned in section 2.11 the current protocol (mean $\pm$ standard deviation of error $< 5 \pm 8$ mmHg) for validating BP devices are inadequate for cuffless devices and biasedly favour devices that do not predict significant changes from a calibration value in studies where BP is not significantly perturbed. The results presented here suggest that the Aktiia device should not be considered an accurate BP monitor.

B

nABPM sleep questionnaire

Investigating non-invasive methods for ambulatory blood pressure monitoring

CUREC approval reference: R80677/RE001

Eoin Finnegan
E-Mail: eoin.finnegan@eng.ox.ac.uk

Participant identifier:



Please provide approximate values for the following ...	Session date:
Time in bed (e.g. 23:00)	
How long it took to go to sleep (mins)	
Time woke up for the final time (e.g. 8:00)	
How long it took you to get out of bed (mins)	
Number of times awake during the night	
Average duration of wake ups (mins)	
How would you rate your sleep overall? 0 – “Very bad” 1 – “Fairly bad” 2 – “Fairly good” 3 – “Very good”	

Figure B.1: nABPM sleep questionnaire completed by all participants.

C

Statistical analysis

Contents

C.1	Hypothesis testing	267
C.1.1	Selecting a statistical test	267
C.1.2	Correcting for multiple hypothesis testing	269
C.2	Circular statistics	270
C.2.1	Circular measures of central tendency	270
C.2.2	Multi-sample test for equal circular median directions	271

C.1 Hypothesis testing

Hypothesis testing is a common method for assessing whether data significantly support a particular hypothesis. First, a null hypothesis (H_0) is generated which suggests the absence of a relationship or characteristic effect. Then, an alternative hypothesis (H_1) is proposed which defines the specific hypothesis being tested (for example, that the mean of a dataset is non-zero). H_0 is rejected (suggesting, but not confirming H_1) if a test statistic, formed under the null hypothesis, is greater than a critical value. The appropriate test statistic and critical value depend on the type of data and the H_0 being tested. The p -value represents the probability of the test statistic being greater than the critical value. A low p -value suggests that there is a small chance of obtaining the test data under the current null hypothesis. If the p -value of a test is less than a pre-defined threshold, the significance level, then the H_0 is rejected in favour of H_1 . Typical significance levels are 5%, 1% and 0.1%.

C.1.1 Selecting a statistical test

Appropriate selection of the statistical test is vital in order for the derived p -value to be representative of the hypothesis being tested. Hypothesis tests can be divided into parametric and non-parametric. Parametric tests make assumptions about the data, such as the assumption of Gaussianity, independence within the data, known variances, or that the data is continuous [413]. Non-parametric tests make no assumptions about the distribution of the data and are typically used when a Gaussian assumption cannot be made (for example, when the dataset size does not meet the central limit theorem threshold of 30 [413]). Additionally, statistical tests can be divided into one-tailed and two-tailed tests. One-tailed tests result from an alternative hypothesis that specifies

a direction, thus providing greater power to detect significant effects; however, if the direction is not known (as is the case for all tests computed in this thesis), a two-tailed test should be used.

C.1.1.1 One-sample t-test

The one-sample t-test tests whether the dataset, x , of length N comes from a Gaussian, or normal, distribution with mean, μ , and known variance, σ . The test statistic is:

$$t = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}} \quad (\text{C.1})$$

where $\bar{x}$ and σ are the sample mean and standard deviation respectively. The p -value is then calculated by comparing t to a t-distribution with $(N - 1)$ degrees of freedom through reference tables.

C.1.1.2 Two-sample t-test

The two-sample t-test tests whether two independent datasets, x and y , of lengths N_x and N_y respectively, come from normal distributions with equal means and unknown variances. The test statistic is:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{N_x} + \frac{\sigma_y^2}{N_y}}} \quad (\text{C.2})$$

where $\bar{x}$ and σ_x are the sample mean and standard deviation of x respectively, similarly for $\bar{y}$ and σ_y .

C.1.1.3 Wilcoxon signed-rank test

The Wilcoxon signed-rank test is the non-parametric equivalent of the t-test, which tests whether the median of a dataset x , of length N , is equal to zero. To compute the test statistic, first all values of x are ranked in ascending order. The ranking of each value of x is given by R_i . The test statistic is then computed as:

$$z = \frac{W - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}} \quad (\text{C.3})$$

where W is:

$$W = \sum_{i=1}^N \text{sign}(x_i) R_i \quad (\text{C.4})$$

and reference tables are used to convert the z -value to the corresponding p -value.

For paired measurements, where two datasets, x and y , are recorded independently and represent paired measurements, then the Wilcoxon signed-rank test tests the hypothesis that the medians of x and y are equal. A new variable, d , is introduced where $d_i = x_i - y_i$. d is then tested to have a median of zero using the same steps outlined above.

C.1.1.4 Mann-Whitney U test

The Mann-Whitney U test is the non-parametric equivalent of the two-sample t-test, which tests whether the median of two datasets x and y , of lengths N_x and N_y respectively, are equal. To compute the test statistic U , first all observations from both x and y are pooled and ranked in order. R_x is then defined as the sum of the ranks from all elements of x and vice versa for R_y . U is then defined as:

$$U = \min\left(R_x - \frac{N_x(N_x + 1)}{2}, R_y - \frac{N_y(N_y + 1)}{2}\right) \quad (\text{C.5})$$

and reference tables are used to convert U to the corresponding p -value.

C.1.1.5 Kruskal-Wallis test

The Kruskal-Wallis test is a non-parametric version of the classical one-way analysis of variance (ANOVA) test [414]. It compares the medians of two, or more, groups of data to determine if the samples come from the same population. The Kruskal-Wallis test statistic is computed using ordered ranks of data across all groups. The ranking of the i^{th} observation of the j^{th} group is given by R_{ij} . The test statistic, H is then computed as:

$$H = (N - 1) \frac{\sum_{j=1}^k n_j (\bar{R}_j - \bar{R})^2}{\sum_{j=1}^k \sum_{i=1}^{n_j} (R_{ij} - \bar{R})^2} \quad (\text{C.6})$$

where n_j is the sample size for the j^{th} group; $\bar{R}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} R_{ij}$ is the average rank of all observations in the j^{th} group; $\bar{R} = \frac{1}{2}N(N + 1)$ is the average rank of all R_{ij} ; and N is the overall sample size. Under the null hypothesis that all groups are from the same population, H is thought to follow a chi-squared distribution with $N - 1$ degrees of freedom.

C.1.2 Correcting for multiple hypothesis testing

When performing multiple hypothesis tests simultaneously, the likelihood of type I errors (rejections of true null hypotheses) increases as more tests are undertaken. Therefore, the resulting p -values are often adjusted (corrected) before they are compared to the significance levels. There is no general consensus on the optimum correction method. In this thesis, we adjust the computed p -values for multiple comparisons using the Benjamini-Hochberg method [299]. This technique aims to reduce the number of type-I errors by inflating small p values. For N multiple hypothesis tests performed with corresponding p -values, $(p_1, p_2, \dots, p_N)$, ordered in increasing order, each p -value is then corrected to:

$$p_i^{\text{corrected}} = p_i \frac{N}{i} \quad (\text{C.7})$$

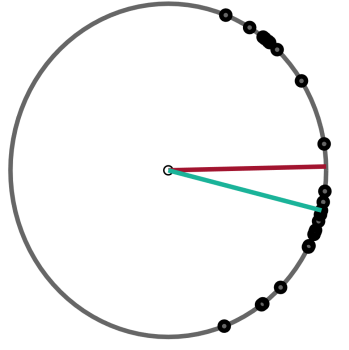


Figure C.1: Example of circular data α existing on the unit circle. The circular mean is shown in red. The circular median is shown in green.

C.2 Circular statistics

Circular variables exist in the range $[0, 2\pi]$. In contrast with a linear scale, on this scale the definition of 0 is arbitrary. For example, two angles, one just above 0 and one just below 2π follow almost exactly the same direction yet exist on opposing ends of the scale. Computing standard statistical measurements (e.g. the arithmetic mean) would produce wrong or misleading results. In this section, we outline the circular statistics relevant to this thesis. These techniques were implemented using the MATLAB toolbox CircStat [347]. We denote the direction of N_α observations as $\alpha = [\alpha_1 \dots \alpha_j \dots \alpha_{N_\alpha}]$. These observations are thought to exist on a circle of unit radius, as shown in figure C.1.

C.2.1 Circular measures of central tendency

C.2.1.1 Circular mean

Directions are transformed into unit vectors in the 2-D plane by the identity:

$$r_j = \cos \alpha_j + i \sin \alpha_j \quad (\text{C.8})$$

After this transformation, the real and imaginary terms of r_j are averaged by:

$$\bar{r} = \frac{1}{N_\alpha} \sum_{j=1}^{N_\alpha} \cos \alpha_j + i \frac{1}{N_\alpha} \sum_{j=1}^{N_\alpha} \sin \alpha_j \quad (\text{C.9})$$

$$= x + iy \quad (\text{C.10})$$

where the vector $\bar{r}$ is referred to as the mean resultant vector, the mean angular direction is given by:

$$\bar{\alpha} = \tan^{-1}\left(\frac{y}{x}\right) \quad (\text{C.11})$$

C.2.1.2 Circular median

The median direction of α , $\hat{\alpha}$, is the direction of the diameter for which half of the observations fall on either side. The median is the endpoint of the diameter closer to the circular mean of the data.

C.2.2 Multi-sample test for equal circular median directions

Similar to the Kruskal-Wallis test outlined in Appendix C.1.1.5, we can test whether multiple groups of circular data possess the same circular median direction. For S groups of data, each with a corresponding circular median, $\hat{\alpha}_i$ the null and alternative hypotheses are proposed as:

H_0 : Any of S samples share a common median direction, i.e. $\hat{\alpha}_1 = \dots \hat{\alpha}_S$

H_1 : Not all S samples share a common median direction.

The total median direction, $\hat{\alpha}$, is first computed by pooling data from all groups. Subsequently, we define m_i as the number of samples in the i^{th} group, whose angle to $\hat{\alpha}$, $d(\alpha_j^i, \hat{\alpha})$ is negative:

$$m_i = \sum_{j=1}^{n_i} [d(\alpha_j^i, \hat{\alpha}) < 0] \quad i = \{1, \dots, S\} \quad (\text{C.12})$$

where the $[\cdot]$ notation, known as Iverson brackets, results in a value of 1 when a condition is met and a value of 0 otherwise. α_j^i indicates the j^{th} sample from the i^{th} group; n_i is the number of samples in the i^{th} group. The test statistic is computed as:

$$P = \frac{N^2}{M(N-M)} \sum_{i=1}^S \left(\frac{m_i^2}{n_i} - \frac{NM}{N-m} \right) \quad (\text{C.13})$$

where $M = \sum_{i=1}^S m_i$ and N is the total number of observations. Under the null hypothesis that all groups are from the same population, P is thought to follow a chi-squared distribution with $S - 1$ degrees of freedom.

D

Linear machine learning models

Contents

D.1	Linear regression and classification methods	273
D.1.1	Ordinary least squares regression	273
D.1.2	Linear mixed effects modelling	274
D.1.3	Logistic regression	276
D.1.4	Least absolute shrinkage and selection operator (LASSO)	277
D.1.5	Elastic net regularisation	277
D.2	Evaluation approaches and metrics	277
D.2.1	Cross-validation	277
D.2.2	Regression metrics	278
D.2.3	Classification metrics	278

D.1 Linear regression and classification methods

D.1.1 Ordinary least squares regression

The simplest method for predictive modelling is ordinary least squares (OLS) regression, where the i^{th} observation of the continuous response variable y is estimated as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_M x_{iM} + \epsilon_i \quad (\text{D.1})$$

$$= \beta_0 + \sum_{j=1}^M \beta_j x_{ij} + \epsilon_i \quad (\text{D.2})$$

where β_j denotes the regression coefficients (slopes) which explain the linear relationship between the explanatory and response variables. β_0 denotes the constant intercept term. ϵ_i denotes the residuals as the discrepancies between y_i and the linear model.

Stacking all N observations together results in the matrix form for linear regression:

$$y = X\beta + \epsilon \quad (\text{D.3})$$

where ϵ is the vector of residuals, assumed to be drawn from a Normal distribution with a constant variance, $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2 I)$ where I is the identity matrix. β is the vector of coefficients (including β_0 as the first value). X is a matrix of predictors which is first concatenated with a column of ones to account for β_0 .

The optimised coefficients $\beta, \tilde{\beta}$, can be determined through maximum likelihood estimation (MLE). MLE first derives a probability distribution of the data fit over the parameter space.

Subsequently, the probability distribution is maximised for a specific parameter choice. y_i can be considered to be drawn from a Normal distribution with a mean value of $x_i\beta$ and a covariance matrix $\sigma_\epsilon^2 I$, thus:

$$p(y_i|x_i) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{(y_i - x_i\beta)^2}{2\sigma_\epsilon^2}\right) \quad (\text{D.4})$$

The likelihood function ($\mathcal{L}$) defines the likelihood that the observed data were produced by a given model as a function of the model parameters. Thus for linear regression, it can be shown that:

$$\mathcal{L}(\beta; y, X) = \prod_{i=1}^N p(y_i|x_i) \quad (\text{D.5})$$

$$\ell(\beta; y, X) = -\frac{N}{2} \ln(2\pi) - N \ln(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} (y - X\beta)^T (y - X\beta) \quad (\text{D.6})$$

where ℓ is the log-likelihood and is typically a more convenient representation of the likelihood function. Differentiating equation (D.6) with respect to β and setting the result to 0 results in the maximum likelihood estimation of β , $\tilde{\beta}$, as:

$$\tilde{\beta} = (X^T X)^{-1} X^T y \quad (\text{D.7})$$

OLS makes the following assumptions:

- Independent observations, x_i .
- Independent and identically distributed (i.i.d) residuals ϵ_i .

D.1.2 Linear mixed effects modelling

In cases involving repeated measures (such as multiple tests taken from the same group/subject), there is likely to be an inter-dependency of the observations on the particular groups and the OLS assumptions presented above are violated. For linear models, dependence between groups can be handled through linear mixed effects modelling. These models introduce group-specific differences in the linear model parameters to ensure that the correlations between the explanatory variables and the responses are not confounded by inter-group variability.

Mixed effects models are so called as they separate the terms into fixed-effects (which are equivalent to the standard linear regression model outlined above), and random-effects (which denote the additional group-specific variation).

D.1.2.1 Random intercepts model

The simplest form of the linear mixed effects model is the random intercepts model. This model allows each group, g , to have its own specific intercept which has deviated from the group intercept, β_0 , by u_0^g :

$$y_i^g = (\beta_0 + u_0^g) + \beta_1 x_{i1}^g + \cdots + \beta_M x_{iM}^g + \epsilon_i^g \quad (\text{D.8})$$

$$= \underbrace{\beta_0 + \beta_1 x_{i1}^g + \cdots + \beta_M x_{iM}^g}_{\text{fixed-effect}} + \underbrace{u_0^g + \epsilon_i^g}_{\text{random-effect}} \quad (\text{D.9})$$

Both u_0^g and ϵ_i^g are assumed to be drawn from Normal distributions, $\mathcal{N}(0, \sigma_u^2)$ and $\mathcal{N}(0, \sigma_\epsilon^2)$ respectively.

D.1.2.2 Random slopes model

The random intercepts model can be easily extended to a random slopes model by including new random effects for a number of the regression coefficients. This model allows each group, g , to have its own specific slope for the subset of parameters under interest. For example, a random slopes model which adjusts the slopes for response variable x_1 by u_1^g would be modelled as:

$$y_i^g = (\beta_0 + u_0^g) + (\beta_1 + u_1^g)x_{i1}^g + \cdots + \beta_M x_{iM}^g + \epsilon_i^g \quad (\text{D.10})$$

$$= \underbrace{\beta_0 + \beta_1 x_{i1}^g + \cdots + \beta_M x_{iM}^g}_{\text{fixed-effect}} + \underbrace{u_0^g + u_1^g x_{i1}^g + \epsilon_i^g}_{\text{random-effect}} \quad (\text{D.11})$$

In theory, we can have any number of random slopes for any of the response variables. In matrix notation the linear mixed effects model can be written as:

$$y = X\beta + Zu + \epsilon \quad (\text{D.12})$$

where u is the vector of q random effects, and Z is the design matrix defining the random effects. u and ϵ are assumed to be distributed as:

$$\begin{bmatrix} u \\ \epsilon \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & \sigma_\epsilon^2 I \end{bmatrix}\right) \quad (\text{D.13})$$

where G is the covariance matrix of the random effect parameters.

D.1.2.3 Training linear mixed-effects models

Training the linear mixed effects model requires estimation of both the fixed-effect parameters β and the random effect parameters u , or more specifically G and σ_ϵ . Again, this is performed by MLE. Equation (D.12) can be rewritten as:

$$y = X\beta + \epsilon^*, \quad \text{where} \quad \epsilon^* = Zu + \epsilon \quad (\text{D.14})$$

This is just a linear model with correlated terms since:

$$\epsilon^* \sim \mathcal{N}(0, V), \quad \text{where} \quad V = ZGZ^T + \sigma_\epsilon^2 I \quad (\text{D.15})$$

The log-likelihood function, ℓ , of β and V given the data is:

$$\ell(\beta, V; y, X) = -\frac{1}{2}(N \log(2\pi) + \log(|V|) + (y - X\beta)^T V^{-1} (y - X\beta)) \quad (\text{D.16})$$

so the MLE for (β, V) are those that maximise the right-hand side of this expression.

D.1.3 Logistic regression

Despite its name, logistic regression is used for classification tasks. The model uses the sigmoid function to map the linear combination of predictors to the probability of the target variable being a certain class. For binary classification, i.e. $y_i = \{0, 1\}$, the logistic function (which estimates the probability of $y_i = 1$) is given by:

$$p(y_i = 1|x_i, \beta) = \frac{1}{1 + \exp(-x_i\beta)} \quad (\text{D.17})$$

where again, $x_i\beta$ is the linear combination of a feature set observation and β , the parameters of the logistic regression model. Typically, a threshold of 0.5 is applied to the output $p(y_i = 1)$ to determine the classification prediction for $\hat{y}_i$.

The optimised coefficients $\beta, \tilde{\beta}$, can be determined through MLE. Logistic regression assumes y_i is drawn from a binomial distribution such that:

$$p(y_i|x_i, \beta) = p(y_i = 1|x_i, \beta)^{y_i} + p(y_i = 0|x_i, \beta)^{1-y_i} \quad (\text{D.18})$$

where $p(y_i = 0|x_i, \beta) = 1 - p(y_i = 1|x_i, \beta)$. Thus for logistic regression, it can be shown that:

$$\ell(\beta; y, X) = -\sum_{i=1}^N (y_i(x_i\beta) - \ln(1 + \exp(x_i\beta))) \quad (\text{D.19})$$

Unlike linear regression (see appendix D.1.1) where MLE takes a closed-form solution, for logistic regression the log-likelihood has to be maximised through gradient ascent which is an iterative process. At each iteration the gradient of equation (D.19) is computed and the estimate for $\tilde{\beta}$ is adjusted so as to move along the gradient. The second derivative of equation (D.19) is positive definite and so the Newton-Raphson algorithm is an accurate and fast method to find the MLE [377].

D.1.4 Least absolute shrinkage and selection operator (LASSO)

Regularisation methods attempt to mitigate against machine learning models overfitting to training data. Overfitting results in models that are not generalisable and so perform well on training data, however, have poor performance on unseen test data [377]. For linear models, regularisation penalises large coefficient values by assuming a prior distribution for the model coefficients. One such regularisation method is the Least Absolute Shrinkage and Selection Operator (LASSO) [415]. LASSO imposes the L1-norm penalty to the model coefficients using non-negative values of shrinkage parameter λ . Thus, when implementing LASSO for both linear and logistic regression, the log-likelihood is updated as:

$$\ell(\beta; y, X) = \sum_{i=1}^N \ln(p(y_i|x_i)) - \lambda \sum_{j=1}^M |\beta_j| \quad (\text{D.20})$$

λ defines how much the model is regularised. As λ increases the number of feature coefficients, β , equal to zero increases. Increasing λ is beneficial in reducing variance in a prediction model through a reduction in overfitting. However, large values of λ remove potentially important features. λ must, therefore, be carefully chosen.

D.1.5 Elastic net regularisation

Elastic-net regularisation is an additional method to reduce overfitting in linear models. Elastic net regularisation adds an additional L2-norm penalty that improves upon the performance of LASSO regularisation when the number of predictors is significantly larger than the sample size (known as the $p > n$ case) [378]. The log-likelihood when implementing elastic-net regularisation is given as:

$$\ell(\beta; y, X) = \sum_{i=1}^N \ln(p(y_i|x_i)) - \frac{\lambda}{2} \sum_{j=1}^M |\beta_j| - \frac{\lambda}{2} \sum_{j=1}^M \beta_j^2 \quad (\text{D.21})$$

Again, the value λ must be carefully chosen.

D.2 Evaluation approaches and metrics

D.2.1 Cross-validation

The generalisability of a machine learning models is typically assessed using cross-validation (CV) [377, 415]. CV splits the data into K partitions, or 'folds'. Training is performed on the $K - 1$ folds and testing on the remaining left-out fold. By repeating this procedure across all folds, the entirety of the data is used for both training and testing. A measure of the central tendency and spread of performance statistics across folds are then reported. In cases where repeated measurements

from the same subject are present in the dataset it is important to partition the data by subject. Subject-wise CV, such as Leave-One-Subject-Out CV (LOSO CV), ensures that there is no “leakage” of subject information between training and testing partitions.

D.2.2 Regression metrics

Regression models produce continuous estimates, $\hat{y}$. The following regression performance metrics were used to evaluate $\hat{y}$ against the ground truth, y : ρ_p , RMSE, and MAE.

Correlation coefficients provide a metric of the association or dependence between two variables. The values of correlation are in the range -1 to +1, with -1 signifying perfect negative correlation, +1 signifying perfect positive correlation, and values close to 0 signifying weak or no correlation between the two variables. ρ_p is used to measure the linear correlation y and $\hat{y}$. ρ_p is defined as:

$$\rho_p = \frac{\text{cov}(y, \hat{y})}{\sigma_y \sigma_{\hat{y}}} = \frac{\sum_{i=1}^N (y_i - \mu_y)(\hat{y}_i - \mu_{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \mu_y)^2 \sum_{i=1}^N (\hat{y}_i - \mu_{\hat{y}})^2}} \quad (\text{D.22})$$

where μ_y and $\mu_{\hat{y}}$ are the mean of y and $\hat{y}$ respectively, and σ_y and $\sigma_{\hat{y}}$ are the standard deviations of y and $\hat{y}$ respectively.

RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (\text{D.23})$$

where N is the number of datapoints. MAE is defined as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (\text{D.24})$$

Both MAE and RMSE characterise the average error of the prediction. As MAE reports the average magnitude of errors, its values are easily interpreted across multiple datasets. In contrast, RMSE gives a higher weight to larger error values, and so it is an important metric when large errors are undesirable [300].

D.2.3 Classification metrics

Classification models produce discrete estimates, often referred to as labels or classes. The following classification performance metrics were used: accuracy (Acc), weighted F1 score (wF1), and Cohen’s kappa (κ). Figure D.1 depicts a typical confusion matrix that summarises the agreement between predicted labels and ground truth labels. The example in figure D.1 is for binary classification

Ground truth	1	TN	FP
	0	FN	TP
		0	1
		Estimated	

Figure D.1: Example binary classification confusion matrix. A confusion matrix summarises the agreement between estimations $\hat{y}$ and y .

(label 0 or 1); however, it can be easily extended for multi-class classification. By analysing the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), a number of different performance metrics can be defined:

$$\text{Acc} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{FN} + \text{TP}} \quad (\text{D.25})$$

Acc is the simplest metric and corresponds to the percentage of correct predictions. However, in cases where there is an imbalance of classes in the dataset Acc can be an unreliable metric. For example, if a dataset contains 90% true responses, then a simple model that predicts all values to be true will report a Acc value of 90%. Therefore, metrics that are more robust to imbalances are often additionally reported, such as $F1$.

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (\text{D.26})$$

where $\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$ and $\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$. For multi-class classification, $F1$ is reported as the mean $F1$ value across all labels. To further improve robustness to imbalances, wF1 is computed by weighting the mean by the number of true instances for each label.

κ is also a robust classification measure that accounts for imbalanced data, as well as chance agreement [367]. κ is defined as:

$$\kappa = \frac{\text{Acc} - p_e}{1 - p_e} \quad (\text{D.27})$$

p_e is the estimated probability of random agreement between y and $\hat{y}$. p_e can be estimated as:

$$p_e = \frac{1}{N} \sum_k n_k^{\text{GT}} n_k^{\text{est}} \quad (\text{D.28})$$

N is the number of datapoints. n_k^{GT} and n_k^{est} denote the number of instances which are labelled as the k^{th} state in the ground-truth (GT) and estimated values. Unlike wF1, κ is a more conservative metric. Values near 1.0 denote almost perfect classification and values close to 0.0 tend towards chance agreement.

E

Supplementary results: Pulse arrival time as a surrogate measure of blood pressure

Contents

E.1	Introduction	281
E.2	Detecting the B-point in the impedance cardiogram	281
E.2.1	Introduction	281
E.2.2	Methods	282
E.2.3	Results	284
E.2.4	Discussion and conclusion	284
E.3	Estimating respiratory rate from the nasal pressure airflow signal	285
E.4	Relationship between transit times and blood pressure	286
E.5	Resampling the impedance cardiogram to 1KHz	286

E.1 Introduction

In this appendix, we provide additional methodology and analysis to supplement the work presented in chapter 5. In appendix E.2 we provide a comparative assessment of three different B-point detection algorithms available in the literature and make conclusions on the appropriate method for use in the Mapping Of Lower Limb skIn pErfusion (MOLLIE) study. In appendix E.3 we outline the methodology implemented for estimating respiratory rate from the nasal airflow signal. Finally, in appendix E.4, we show the relationship between BP and TT for all individuals in the dataset.

E.2 Detecting the B-point in the impedance cardiogram

E.2.1 Introduction

The pre-ejection period (PEP) is defined as the time delay between electrical depolarisation of the heart's left ventricle and the opening of the aortic valve to pump blood through the aorta [122]. The timing of the aortic valve opening may be located in the impedance cardiogram (ICG) by a deflection in the ICG signal known as the B-point.

However, automatically detecting the B-point has been proven to be a notoriously difficult task and remains an active area of research. This difficulty stems from the observation that intra- and inter-individual appearance of the B-point on the ICG varies and may manifest as one of the

following: a turning point, inflection, or rapid slope change [291]. Additionally, motion artefacts, muscle contractions, respiration, and different cardiovascular pathologies significantly impact the shape of the ICG waveform [291, 122]. Often, expert visual inspection is required to detect the correct B-point location, however this has obvious limitations. As a result, many different algorithms have been proposed in the literature for locating the B-point [290, 291, 292].

E.2.2 Methods

We implemented three B-point detection algorithms: Forouzanfar *et al.* [290], Arbol *et al.* [291], and Lozano *et al.* [292]. Details of these algorithms are provided in the following sections.

For each method, the PEP time series (derived using the ECG R-peak as the proximal timing) was subsequently post-processed and an estimate of the signal quality index (SQI_{PEP}) was computed using the steps outlined in section 5.5.2. Any values where SQI_{PEP} was 0 (which may be the result of: errors in ECG, errors in ICG, detected periods of transient outliers as well as PEP values outside of a physiological range) were removed from any subsequent analysis.

In the absence of a ground truth (which would typically come in the form of expert annotation), we assessed the performance of different B-point detection algorithms through visual inspection and an estimate of the level of noise present in the derived PEP time series. This analysis was performed over all participants in the MOLLIE study (see section 3.3).

An estimate of the degree of noise was computed as the standard deviation (σ) in each of the derived PEP signals. High values of σ indicate the presence of a significant level of noise. However, conversely, very low values of σ suggest that the method is unable to track the changes in PEP expected under an infusion of phenylephrine. Therefore, any conclusions drawn from analysing σ must be corroborated by visual inspection.

E.2.2.1 Arbol *et al.* [291]

Using the Arbol *et al.* method, the B-point was defined as the local maximum of d^2ICG/dt^2 within 300 ms of the previous C-point. The authors compared this approach to two other methods (zero-crossing of the ICG, and the maximum of $dICG/dt$) in a dataset of 30 individuals with ground truth provided by expert annotation. The proposed method had a significantly higher coincidence with the expert annotation (78.8%) than the other two methods (24.5% and 13.8% respectively).

E.2.2.2 Forouzanfar *et al.* [290]

Using the Forouzanfar *et al.* method, the B-point was detected as follows:

- Detect the local minimum (A-point) of the ICG waveform between the ICG onset (ECG R-peak location) and the C-point (maximum of ICG beat).
- Define H_{ICG} as the amplitude difference between the detected A-point and C-point.
- Find all the monotonically increasing segments between A-point and C-point and select the one with the highest height that starts from at least two-thirds of the C-point amplitude and reaches at least three-quarters of C-point amplitude. If no such segment is found, relax the second condition to one-half of the C-point amplitude. If no such segment is found, select the last monotonically increasing segment.
- Find all $d^2\text{ICG}/dt^2$ zero crossings over the first two-thirds of the most prominent monotonically increasing segment and discard those with $d\text{ICG}/dt > 6 \cdot H_{\text{ICG}}/f_s$, where f_s is the sampling frequency of the ICG signal.
- Find all $d^2\text{ICG}/dt^2$ maximums over the first two-thirds of the most prominent monotonically increasing segment and discard those with $d^2\text{ICG}/dt^2 < 0.002 \cdot H_{\text{ICG}}/f_s$.
- Label the last zero crossing or local maximum of $d^2\text{ICG}/dt^2$ as the B-point. If no zero crossing or local maximum exist, label the first point of the segment as B-point.

B-point detection using this method was compared to expert annotation in a sleep study dataset containing 20 healthy participants. The authors reported lower mean errors in B-point detection using their proposed method (8.6 ms) than using the Arbol *et al.* method (29.0 ms).

E.2.2.3 Lozano *et al.* [292]

The Lozano method avoids the complications associated with detecting the B-point of the ICG through a data-driven polynomial relationship which estimates PEP using the time delay from the R-peak to the C-point. Using the Lozano *et al.* method, PEP was defined using the following quadratic relationship:

$$\text{PEP} = 1.233 \cdot t_{RC} - 0.0032 \cdot t_{RC}^2 - 31.59 \quad (\text{E.1})$$

where t_{RC} is the time between ECG R-peak and ICG C-point. As discussed by Lozano *et al.* [292], the ICG beat may contain multiple maxima, in which case selecting the one with the

largest amplitude may lead to incorrect PEP estimation. We, therefore, defined the C-point as the maximum of a low-pass filtered ICG signal. A 4th-order IIR filter with a cut-off frequency of 5 Hz was subsequently used. This had the effect of combining two adjacent maxima and resulted in a less noisy PEP time series.

Lozano *et al.* reported a ρ_p value of 0.98 between PEP estimates and expert annotation PEP values in a dataset of 26 healthy volunteers.

E.2.3 Results

Table E.1 reports upon the comparison of PEP values derived from the three B-point detection algorithms analysed. ‘# detected’ defines the total number of estimated PEP values across all participants. Given that all algorithms were evaluated on the same dataset, low values of ‘# detected’ indicate that a significant proportion of PEP values have been rejected by the SQI_{PEP} for either representing spurious outliers or for being outside of physiological limits. σ provides an indication towards the level of noise present in each of the PEP time series. The Lozano method resulted in the largest set of PEP values in addition to a significantly lower σ than both the Arbol ($p < 0.001$) and Forouzanfar ($p < 0.001$) methods.

Figure E.1 illustrates the PEP time series computed by the three methods across three participants. The Lozano method shows good agreement with the overall trends of the other two methods whilst exhibiting significantly lower high-frequency variations.

Table E.1: Comparison of the three B-point detection algorithms.

Method	# detected	σ (ms) ¹
Arbol <i>et al.</i> [291]	32,550	5.86 (2.93)
Forouzanfar <i>et al.</i> [290]	34,077	8.86 (4.76)
Lozano <i>et al.</i> [292]	37,908	3.19 (1.46)

¹ Values given as participant-wise median (IQR).

E.2.4 Discussion and conclusion

In the absence of ground-truth data, we cannot conclude which of the proposed methods presents the highest accuracy for this dataset. We can, however, hypothesise that PEP signals derived using the Arbol and Forouzanfar methods, which analyse high-order derivatives of the ICG signal, clearly exhibit a large amount of variation due to errors in B-point detection. This is largely evident in figure E.1 and is further corroborated by the relatively large σ values in table E.1. The Lozano method on the other hand results in much lower variations in PEP values and appears to follow the

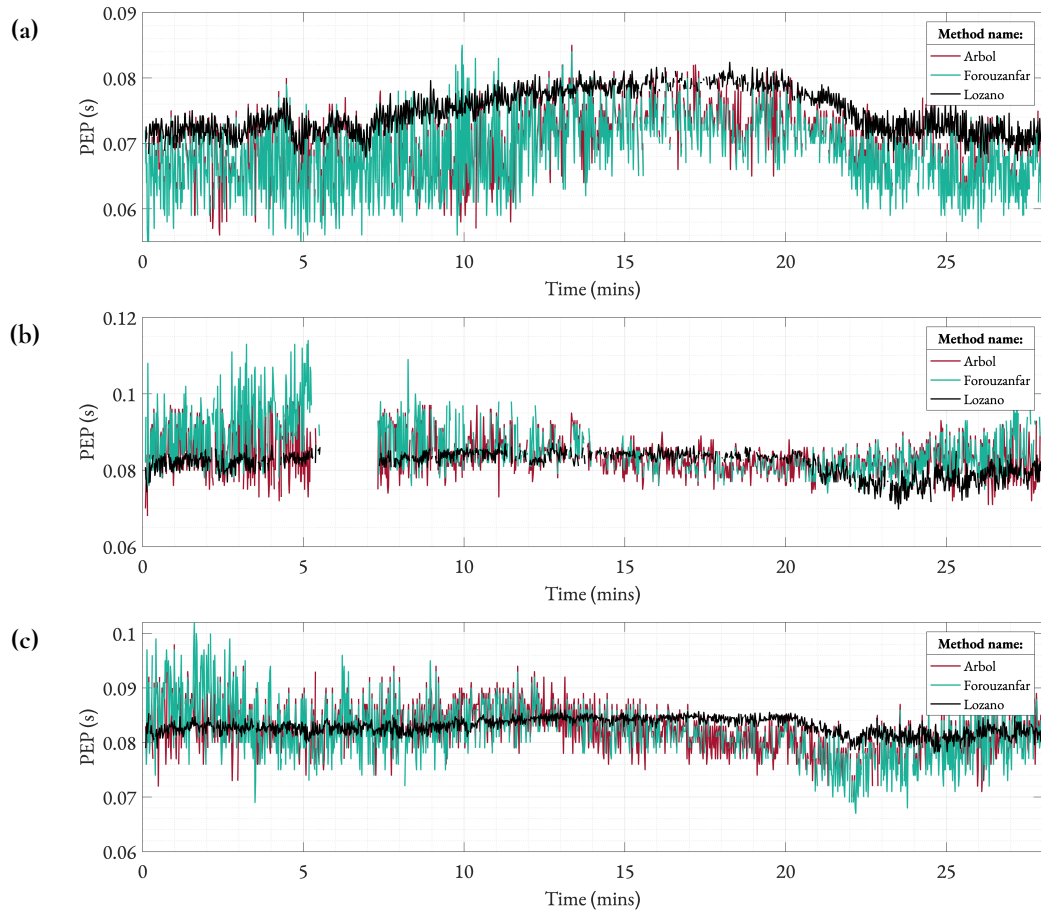


Figure E.1: The PEP time series computed by the three methods across three example participants in the MOLLIE study.

overall trend of both the Arbol and Forouzanfar methods suggesting that it is accurately tracking PEP. In addition, despite being extrapolated, PEP values computed by the Lozano method did not regularly exceed physiological limits as evidenced by the high '# detected' value.

From the analysis presented in this appendix, we concluded that estimating PEP values using the Lozano method was justified in the MOLLIE dataset.

E.3 Estimating respiratory rate from the nasal pressure airflow signal

The typical morphology of pressure airflow signal (φ_{nasal}) measured using a nasal cannula at the opening of the nostrils is shown in figure E.2. To compute an estimate of respiratory rate (RR) from φ_{nasal} we followed the work of Jorge [416] and adapted the implemented hyperparameters for our adult population.

Firstly, components outside of a physiologically accepted RR range, taken to be between 6 to 42 breaths/min [263], were removed. We, therefore, filtered the φ_{nasal} signal using a 6th-order

IIR band-pass filter with cut-off frequencies of 0.1 Hz and 0.7 Hz. The segmentation of φ_{nasal} into individual breaths was performed by adapting the Zong *et al.* [264] photoplethysmogram (PPG) segmentation algorithm, slope sum function (SSF), introduced in section 4.2.2. In order to mimic the PPG morphology, φ_{nasal} was time-reversed i.e:

$$\bar{\varphi}_{nasal}(i) = \varphi_{nasal}(N_s - i), \quad i = 1, \dots, N_s \quad (E.2)$$

where N_s is the number of samples in φ_{nasal} . We used an analysis window of 0.9 s to compute the values of the SSF for $\bar{\varphi}_{nasal}$. When the SSF signal crossed the detection threshold, a 0.7-second window around the crossing point was used to search for minimum and maximum SSF values. The initial threshold was set based on the first 15 seconds of each session.

RR was subsequently computed, at a sampling rate of 0.2 Hz, by a running window mean of the inter-breath interval between the detected breath onsets in each window. A window length of 30 seconds and step size of 5 seconds was used.

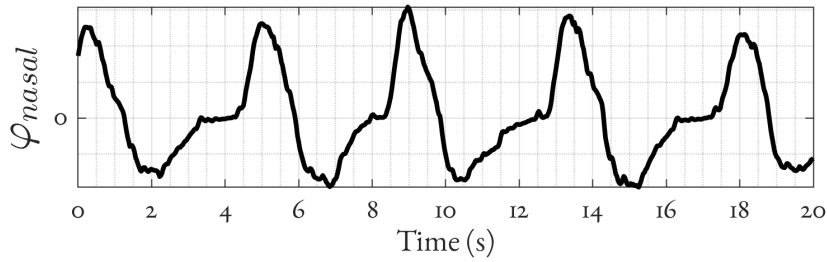


Figure E.2: Typical nasal pressure waveform for a participant breathing at 24.7 breaths/min.

E.4 Relationship between transit times and blood pressure

Tables E.2 and E.3 shows the median and IQR performance statistics for all models over all individuals in the MOLLIE dataset for MAP and DBP estimation respectively.

Figures E.3 and E.4 shows the individual relationship for PAT and PTT with MAP for all participants in the study. Note the large variations in the range of BP values for the y-axis. Figures E.5 and E.6 shows similar results for DBP.

E.5 Resampling the impedance cardiogram to 1KHz

An ICG signal recorded at 200 Hz provides a time resolution of 5 ms for detection of the B-point (point of aortic valve opening) or C-point (maximum of ICG pulse). This may lead to considerable errors in PEP estimates in our dataset. Many bioimpedance monitors will sample at 1 kHz to reduce this error. This section compares the PEP estimates used in this paper (from

Table E.2: Performance statistics of MAP estimation using the models proposed. Results are given as median (IQR) computed across all participants. Entries in bold indicate the best performance for that metric for either *a posteriori* or population-based models.

Model		ρ_p	RMSE *		MAE *	
<i>Baseline reference</i> [†]		- (-)	9.62	(4.67)	6.94	(3.97)
<i>A posteriori models</i>						
PAT	Random-intercepts	0.82 (0.19)	4.47	(2.16)	3.60	(1.70)
	Random-slopes	0.82 (0.19)	3.86	(2.49)	2.92	(1.93)
PTT	Random-intercepts	0.88 (0.16)	3.69	(1.91)	3.09	(1.68)
	Random-slopes	0.88 (0.16)	3.36 (2.02)	2.80 (1.18)		
Population-based models						
PAT	Poon [157]	0.82 (0.18)	7.40	(5.11)	5.92	(3.84)
	Gesche [160]	0.82 (0.18)	10.26	(20.66)	7.87	(17.33)
	Fung [161]	0.83 (0.19)	6.31 (2.16)	5.01 (2.49)		
PTT	Poon [157]	0.83 (0.16)	6.83	(4.53)	5.53	(3.03)
	Gesche [160]	0.84 (0.18)	516.96	(1717.41)	398.62	(1316.59)
	Fung [161]	0.85 (0.15)	23.00	(20.66)	17.88	(16.61)

* results given in units of mmHg. [†] *Baseline reference* assumed no changes in MAP for each participant from MAP^{calib}.

the original 200 Hz ICG signal) to PEP estimates from an upsampled ICG at 1 kHz via cubic spline interpolation. Figure E.7 shows the difference in PEP values from the two ICG signals for all volunteers in the dataset. Figure E.8 compares the two signals across the entire dataset. The differences were low (mean difference = 3 ms), there was a bias towards the 200 Hz ICG over estimating the PEP signal compared to the 1 kHz ICG. In general, the trends are consistent across the two signals. We conclude, therefore, that PEP estimates from the 200 Hz ICG were not a significant limitation to the results in this study.

Table E.3: Performance statistics of DBP estimation using the models proposed. Results are given as median (IQR) computed across all participants. Entries in bold indicate the best performance for that metric for either *a posteriori* or population-based models.

Model		ρ_p	RMSE *		MAE *	
Baseline reference [†]		- (-)	8.82	(4.94)	6.82	(4.01)
<i>A posteriori</i> models						
PAT	Random-intercepts	0.83 (0.33)	4.12	(2.18)	3.21	(1.76)
	Random-slopes	0.83 (0.33)	3.53	(1.81)	2.89	(1.22)
PTT	Random-intercepts	0.85 (0.13)	3.25	(1.83)	2.67	(1.55)
	Random-slopes	0.85 (0.13)	3.07 (1.30)	2.44 (1.13)		
Population-based models						
PAT	Poon [157]	0.83 (0.33)	7.98	(4.65)	6.14	(3.82)
	Gesche [160]	0.82 (0.28)	11.14	(20.19)	8.44	(16.77)
	Fung [161]	0.83 (0.32)	6.32 (2.27)	4.88 (2.56)		
PTT	Poon [157]	0.80 (0.22)	7.33	(4.63)	5.64	(3.64)
	Gesche [160]	0.78 (0.26)	521.18	(1774.94)	406.75	(1390.08)
	Fung [161]	0.81 (0.23)	24.42	(23.16)	18.44	(20.64)

* results given in units of mmHg. [†] Baseline reference assumed no changes in DBP for each participant from DBP^{calib}.

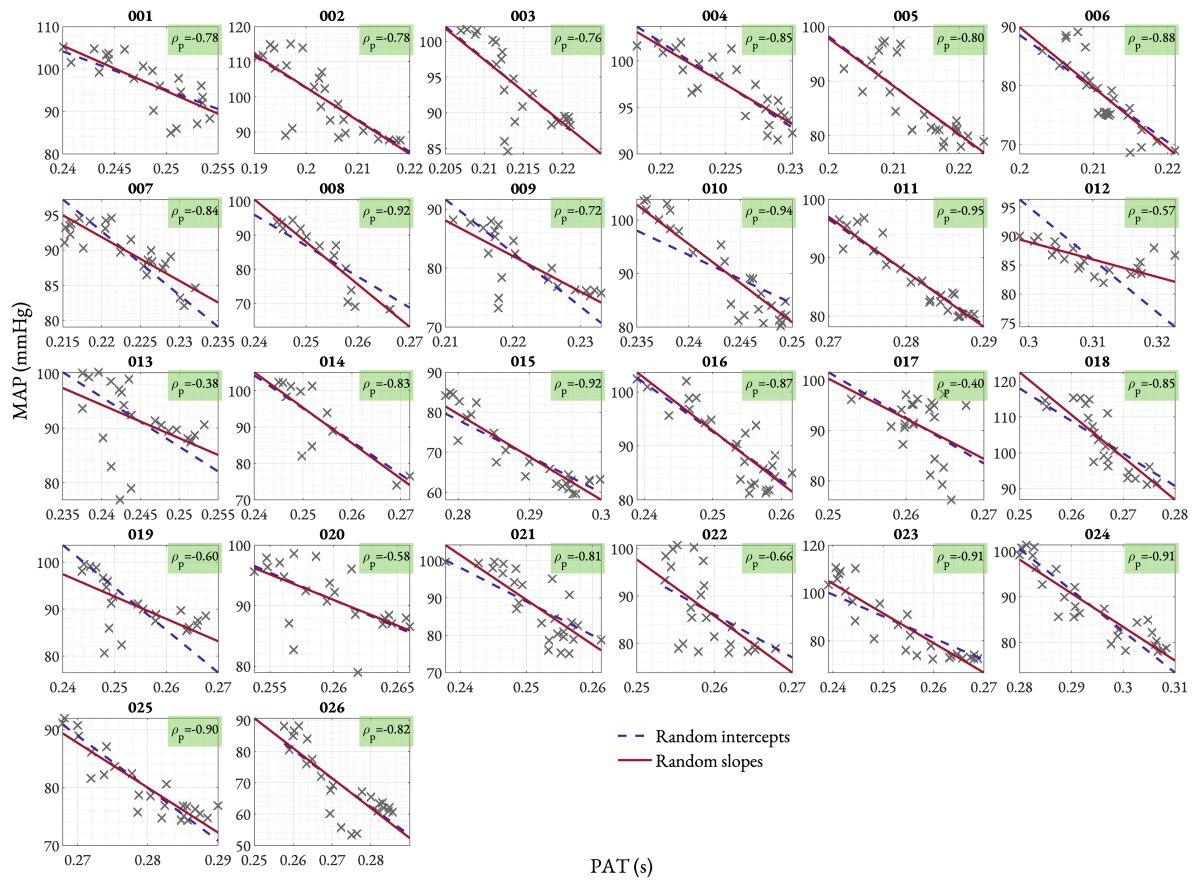


Figure E.3: Individual relationships between MAP (mmHg) and PAT (s) for all 26 participants in the MOLLIE study. The random intercepts regression line is shown in dashed blue. The random slopes regression line is shown in red.

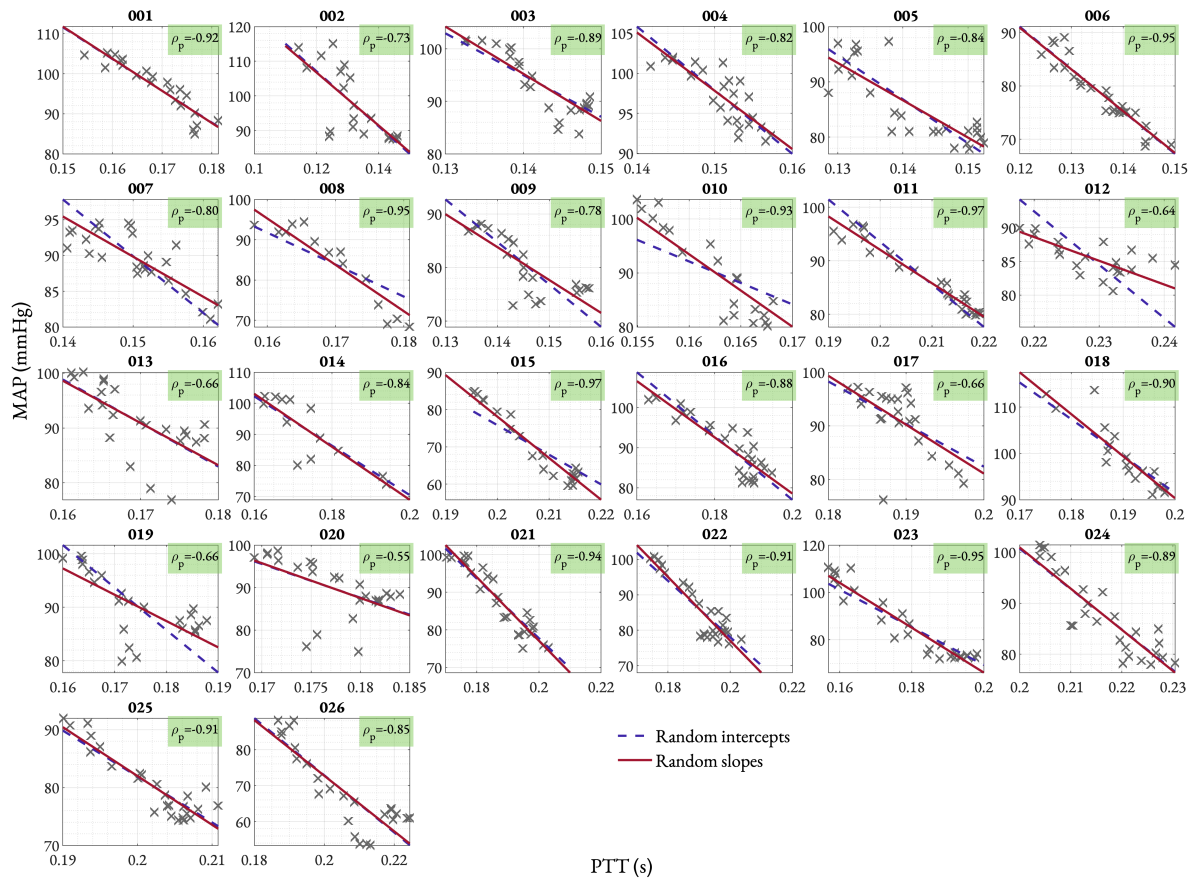


Figure E.4: Individual relationships between MAP (mmHg) and PTT (s) for all 26 participants in the MOLLIE study. The random intercepts regression line is shown in dashed blue. The random slopes regression line is shown in red.

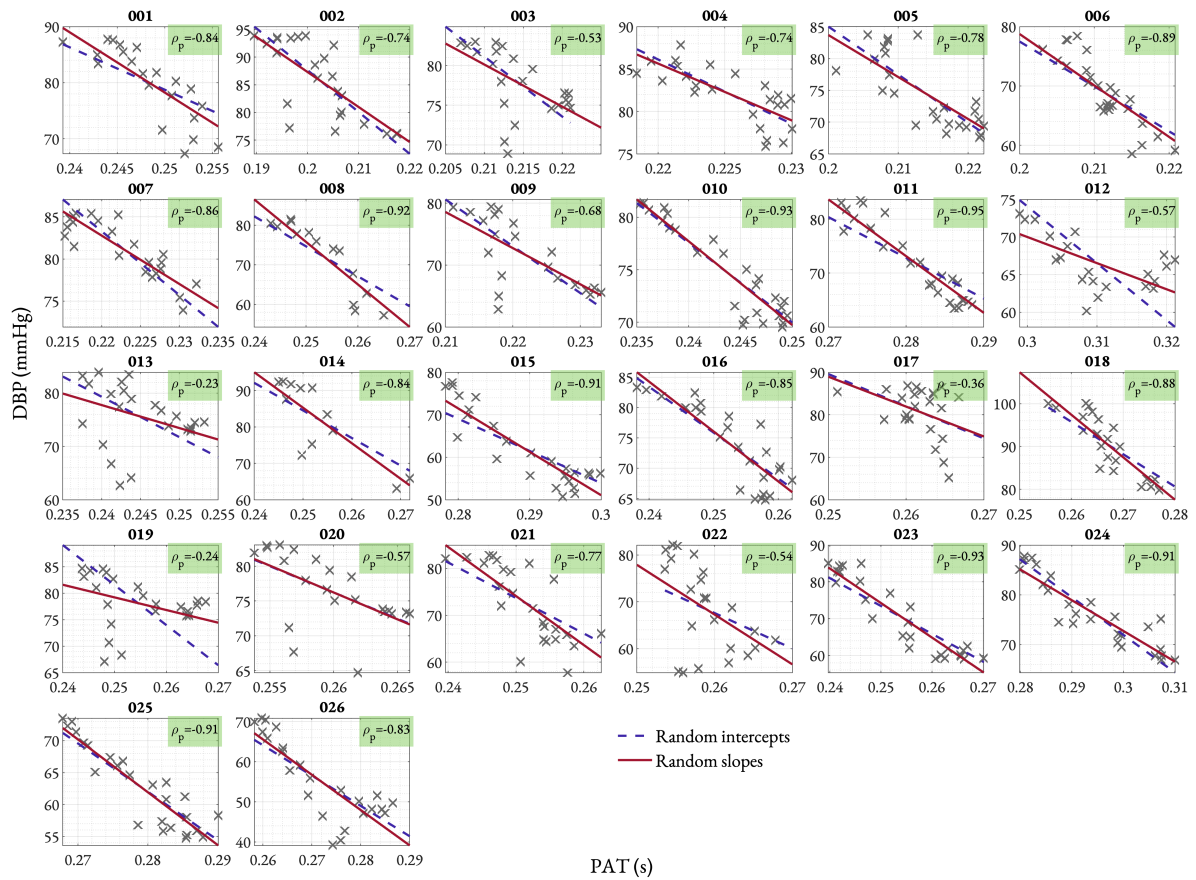


Figure E.5: Individual relationships between DBP (mmHg) and PAT (s) for all 26 participants in the MOLLIE study. The random intercepts regression line is shown in dashed blue. The random slopes regression line is shown in red.

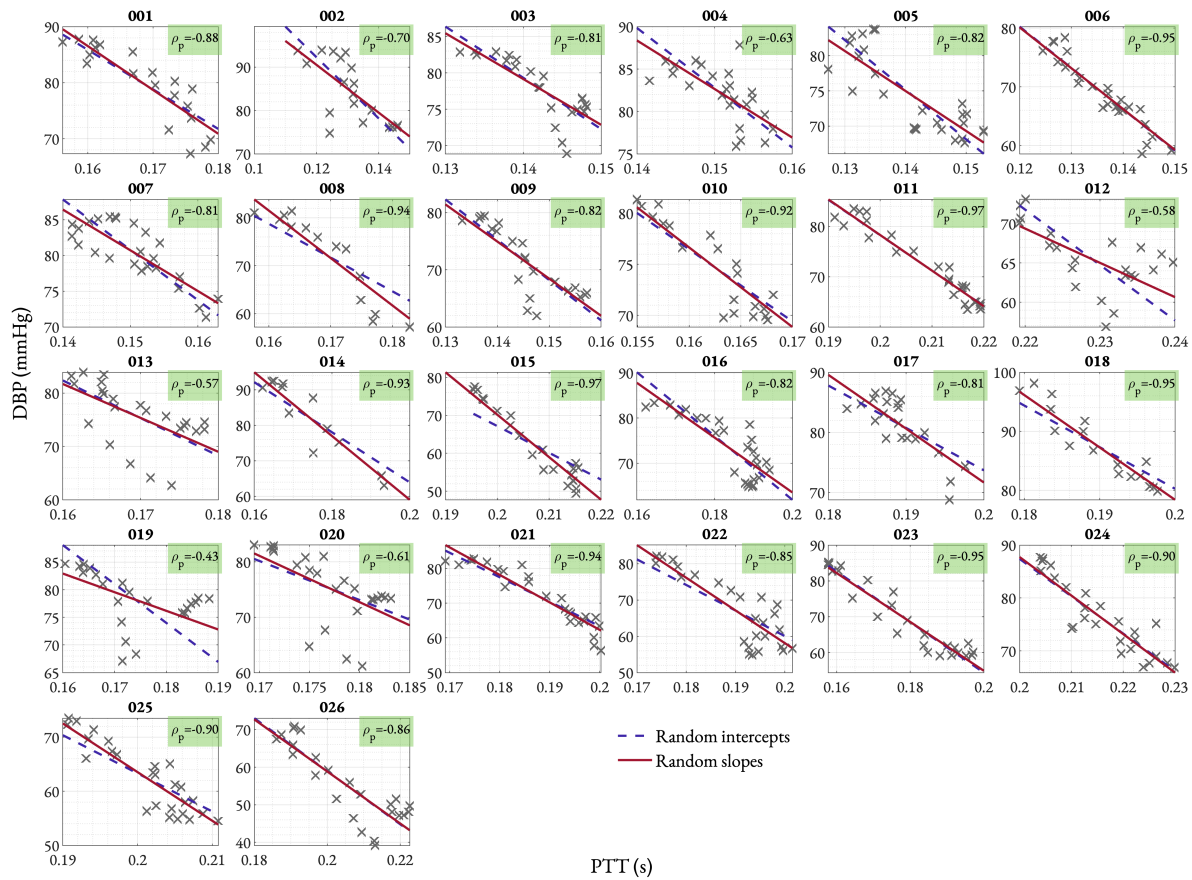


Figure E.6: Individual relationships between DBP (mmHg) and PTT (s) for all 26 participants in the MOLLIE study. The random intercepts regression line is shown in dashed blue. The random slopes regression line is shown in red.

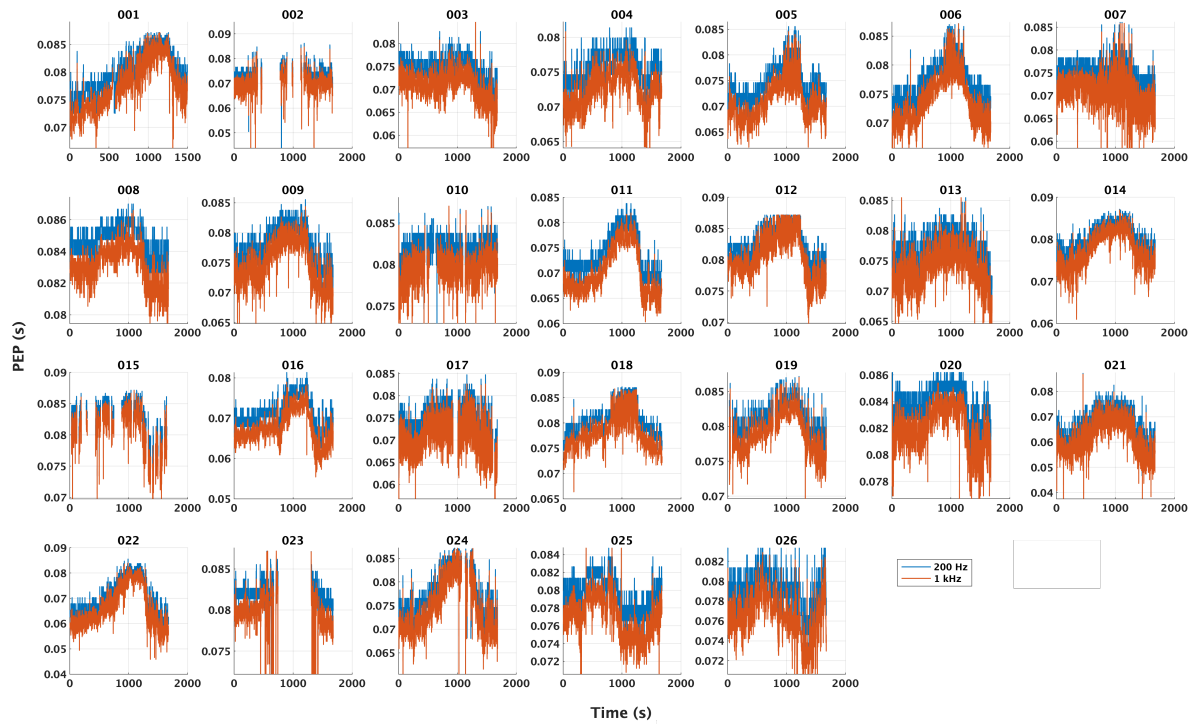


Figure E.7: PEP estimates from the original 200 Hz ICG signal (blue) compared to PEP estimates from the 1 kHz upsampled ICG signal (red) for all volunteers in the dataset.

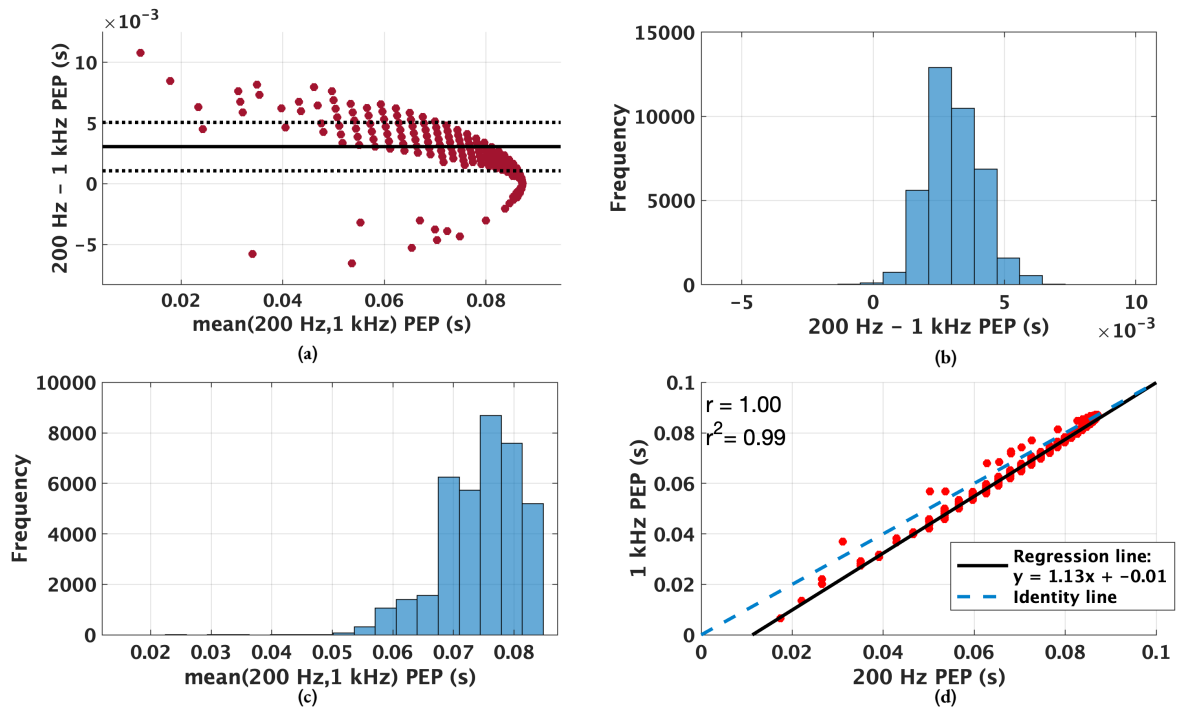


Figure E.8: Comparison of PEP estimates from the two ICG signals (original 200 Hz ICG and upsampled 1 kHz ICG). The errors were low with a slight bias where PEP estimates from the 200 Hz signal were generally higher.

Supplementary results: Features from the photoplethysmogram and electrocardiogram for estimating changes in blood pressure

F.1 Introduction

In this appendix, we provide additional results to supplement the work presented in chapter 6.

F.2 Model performance for mean arterial pressure and diastolic blood pressure estimation

Tables F.1 and F.2 shows the median and IQR performance statistics for all models across all participants in the MOLLIE dataset for ΔMAP and ΔDBP estimation respectively. Figures F.1 and F.2 shows the scatter and Bland-Altman plots for ΔMAP and ΔDBP estimation respectively.

Table F.1: Performance statistics of ΔMAP estimation using the models proposed. Results are given as median (IQR) computed across all folds of the LOSOCV. Entries in bold indicate the best performance for that metric.

Model name	ρ_p	RMSE [*]	MAE [*]
<i>Baseline reference</i>	- (-)	9.62 (4.67)	6.94 (3.97)
LASSO+OLS _{PPG}	0.81 (0.39)	6.17 (3.93)	4.77 (2.69)
RF _{PPG}	0.85 (0.21)	5.03 (3.97)	4.05 (3.72)
LASSO+OLS _{ECG}	0.69 (0.46)	6.67 (3.59)	5.53 (1.78)
RF _{ECG}	0.67 (0.46)	6.88 (2.16)	5.49 (1.49)
LASSO+OLS _{PPG+ECG}	0.80 (0.29)	6.33 (3.63)	4.73 (2.86)
RF _{PPG+ECG}	0.86 (0.23)	4.94 (3.69)	4.10 (3.77)
LASSO+OLS _{PAT}	0.77 (0.29)	6.87 (4.00)	5.14 (2.33)
RF _{PAT}	0.67 (0.24)	7.11 (3.39)	5.60 (2.77)

^{*} results given in units of mmHg. [†] *Baseline reference* assumed no changes in MAP for each participant from MAP^{calib}.

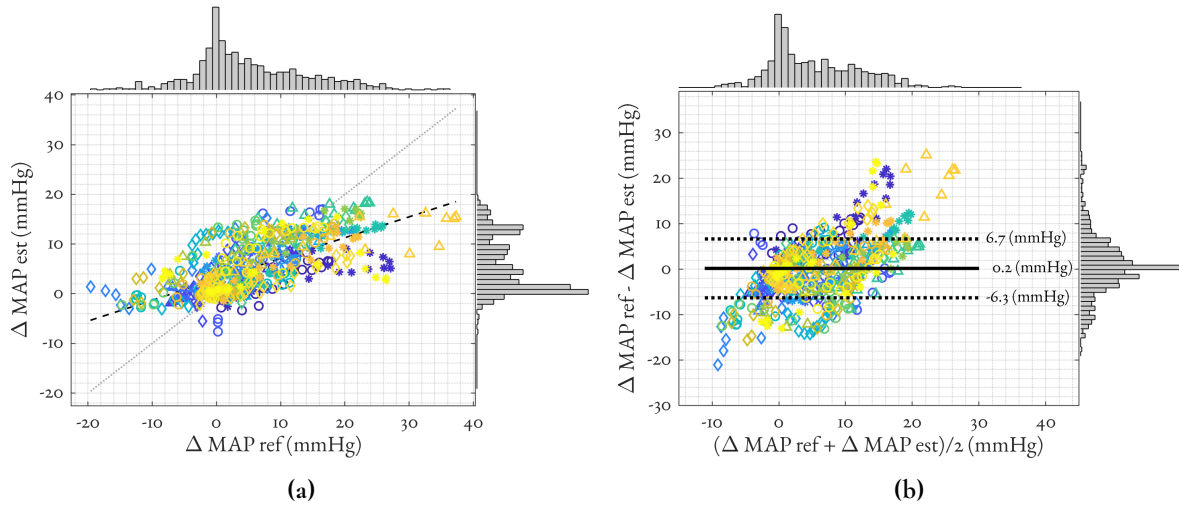


Figure F.1: Agreement between the reference ΔMAP values and the estimated ΔMAP across the cohort using $\text{RF}_{\text{PPG}} + \text{ECG}$. Data from individual participants are colour and marker-coded. (a) Scatter plot, with an overall correlation of 0.71, a median participant-wise correlation coefficient of 0.91 with a range of 0.31 to 0.97. The black striped line shows the regression line. The Grey dotted line shows the line of unity. (b) Bland-Altman analysis, with the bias of the overall error at 0.2 mmHg, with a standard deviation of 6.5 mmHg

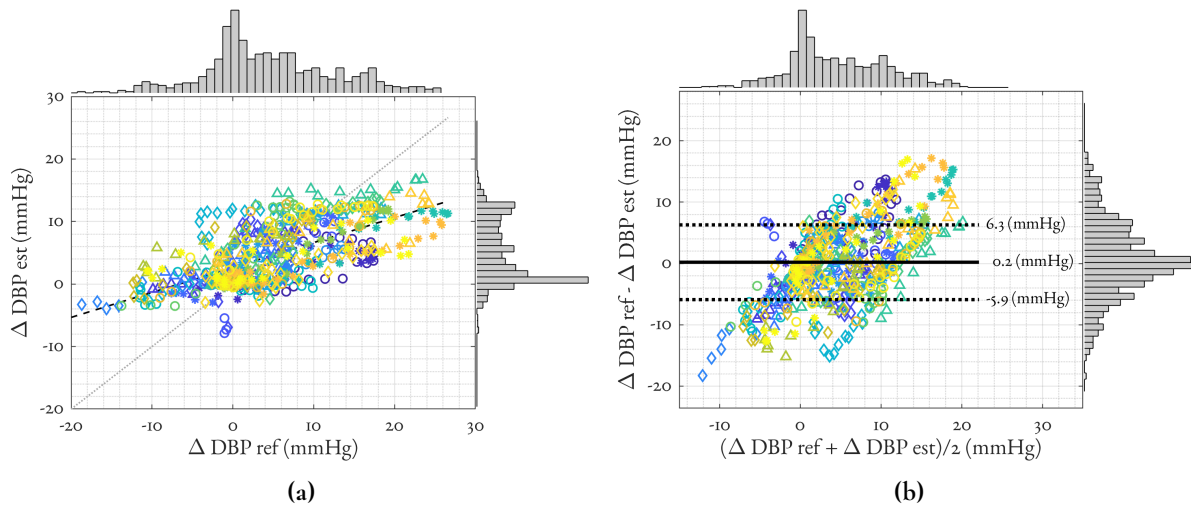


Figure F.2: Agreement between the reference ΔDBP values and the estimated ΔDBP across the cohort using $\text{RF}_{\text{PPG}} + \text{ECG}$. Data from individual participants are colour and marker-coded. (a) Scatter plot, with an overall correlation of 0.67, a median participant-wise correlation coefficient of 0.86 with a range of 0.42 to 0.89. The black striped line shows the regression line. The Grey dotted line shows the line of unity. (b) Bland-Altman analysis, with the bias of the overall error at 0.2 mmHg, with a standard deviation of 6.2 mmHg

Table F.2: Performance statistics of Δ DBP estimation using the models proposed. Results are given as median (IQR) computed across all folds of the LOSOCV. Entries in bold indicate the best performance for that metric.

Model name	ρ_p	RMSE *	MAE *
Baseline reference	- (-)	8.82 (4.94)	6.82 (4.01)
LASSO+OLS _{PPG}	0.77 (0.31)	6.64 (4.79)	5.30 (3.97)
RF _{PPG}	0.84 (0.23)	5.21 (3.65)	4.54 (3.17)
LASSO+OLS _{ECG}	0.68 (0.44)	6.36 (3.86)	5.40 (3.05)
RF _{ECG}	0.59 (0.60)	7.06 (2.66)	5.28 (2.26)
LASSO+OLS _{PPG+ECG}	0.77 (0.52)	6.95 (4.10)	5.83 (3.40)
RF _{PPG+ECG}	0.85 (0.24)	5.39 (3.68)	4.44 (3.11)
LASSO+OLS _{PAT}	0.70 (0.41)	6.43 (4.23)	5.21 (2.93)
RF _{PAT}	0.63 (0.34)	7.00 (2.95)	5.24 (2.55)

* results given in units of mmHg. † Baseline reference assumed no changes in DBP for each participant from DBP^{calib}.

F.3 Individual results

Figures F.3 and F.4 shows the individual results for BP estimation using the RF_{PPG + ECG} model across all participants in the study. Under each participant, we show the summed SHAP values for each observation grouped into PPG features (PPG_{fts}), ECG features (ECG_{fts}), and demographics. SHAP values in red indicate observations where a significant accumulated marginal contribution was reported. Features from the PPG consistently offered the largest contribution to blood pressure estimation.

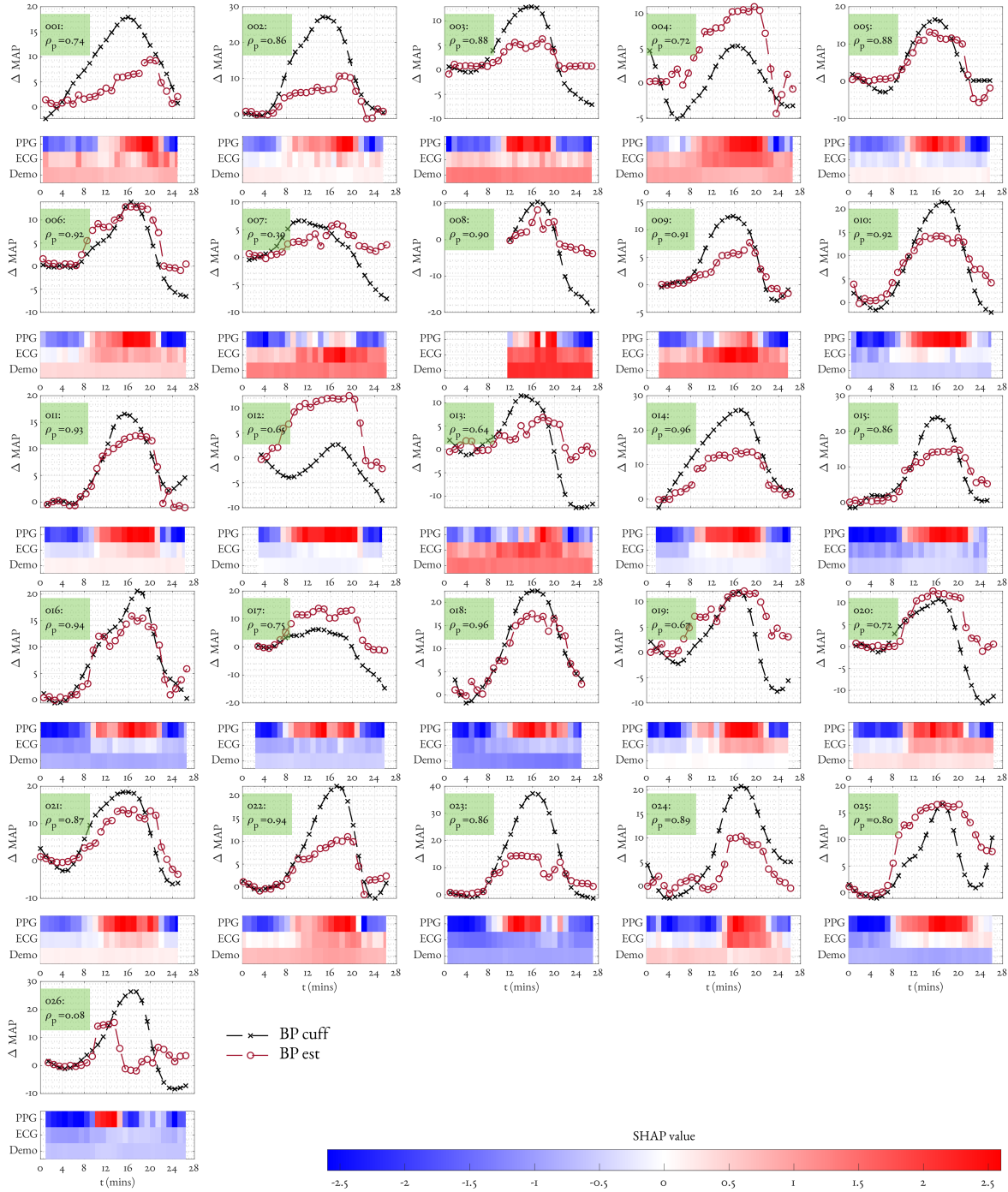


Figure F.3: Individual results of ΔMAP estimation using the $\text{RF}_{\text{PPG} + \text{ECG}}$ model across all participants in the study. Note the differences in the y-axis scale. The reference ΔMAP values from the cuff are shown in black and the estimated values are shown in red. Under each participant's data, we show the summed SHAP values for each observation grouped into PPG_{fts} , ECG_{fts} , and demographics. Red hues indicate a positive marginal contribution to the model prediction for each observation. Blue hues indicate a negative marginal contribution to the model prediction for each observation.

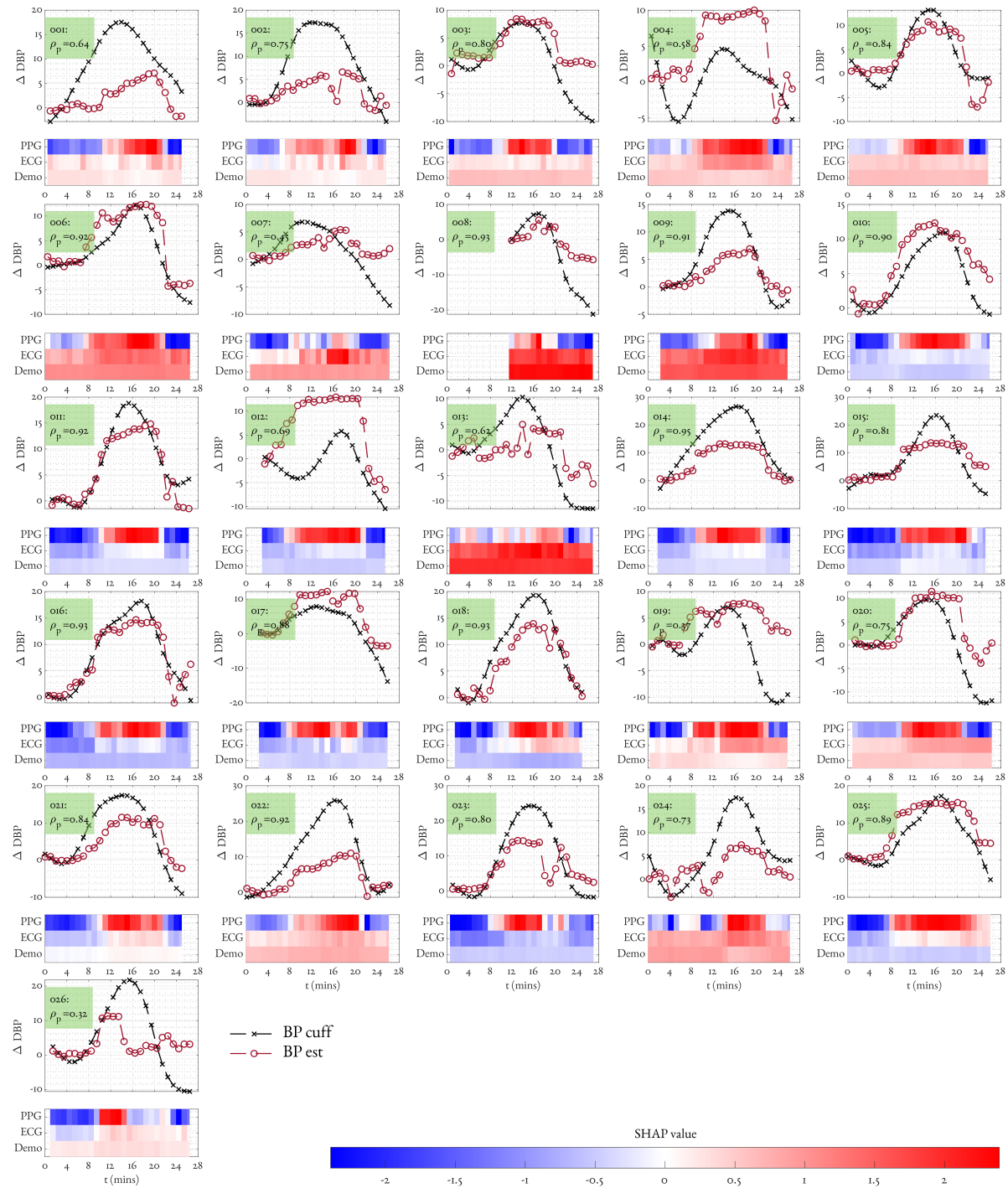


Figure F.4: Individual results of ΔDBP estimation using the $\text{RF}_{\text{PPG} + \text{ECG}}$ model across all participants in the study. Note the differences in the y-axis scale. The reference ΔDBP values from the cuff are shown in black and the estimated values are shown in red. Under each participant's data, we show the summed SHAP values for each observation grouped into PPG_{fts} , ECG_{fts} , and demographics. Red hues indicate a positive marginal contribution to the model prediction for each observation. Blue hues indicate a negative marginal contribution to the model prediction for each observation.

F.4 Individual changes in cardiac output

Figure F.5 shows the changes in cardiac output (CO) observed for all participants in the dataset across the full duration of their session. The majority of participants experienced a decrease in CO driven by a decrease in HR.

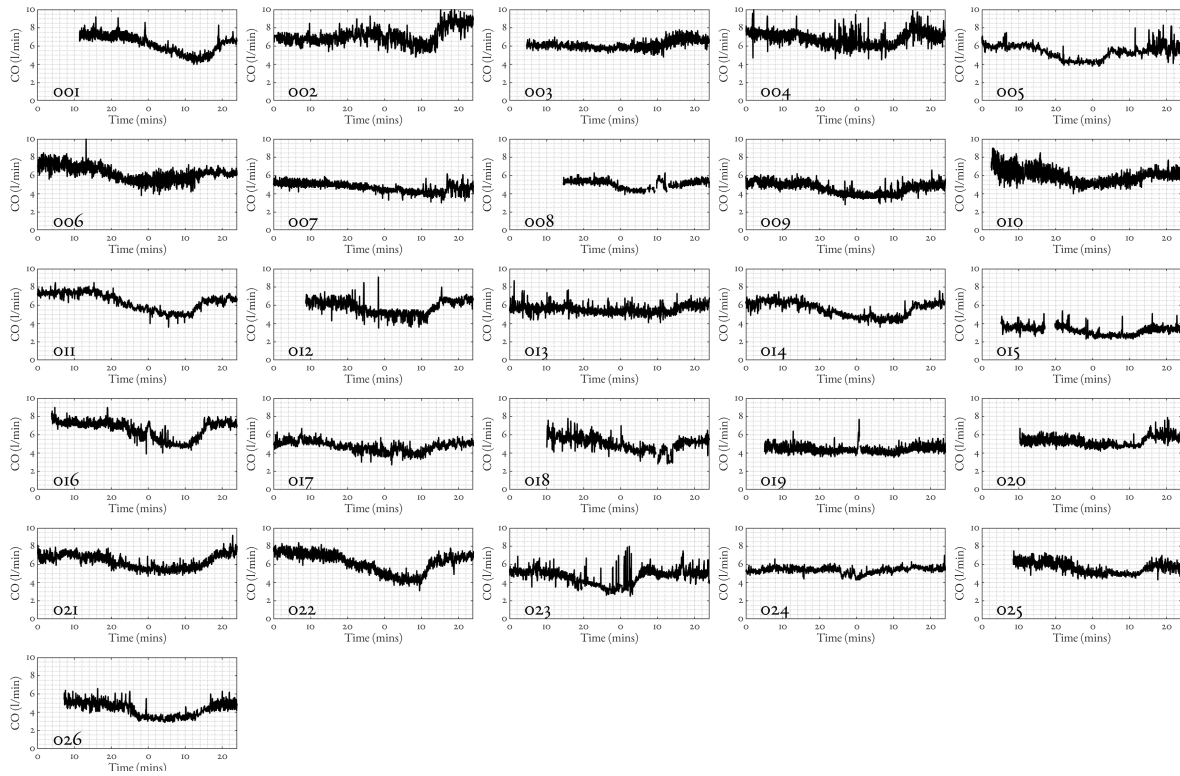


Figure F.5: Individual changes in CO observed in all participants across the full duration of their session.

Supplementary results: Classification of blood pressure phenotypes

G.1 Introduction

This appendix provides additional analysis to complement the work presented in chapter 8. Method 1, as outlined in section 8.4, estimated a patient's night-time (midnight to 8am) SBP using data collected between 8am and midnight. In this appendix, we present justification for implementing a regularised *individual calibration* approach, through the use of mixed-effects modelling, as opposed to two other modelling approaches (fixed-effects and individual).

G.2 Methods

In this appendix we analyse three different modelling approaches to map a feature set X , containing data from N_p patients, to the SBP target vector, y , using the dataset constructed in section 8.4. To ease interpretation, we restricted this analysis to the linear relationship between SBP and PAT (as opposed to multiparameter modelling including PPG_{fts} as performed in chapter 8). The three possible modelling approaches are:

- **Fixed-effects** - A global linear model, irrespective of patient.

$$y = X\beta + \epsilon \quad (\text{G.1})$$

- **Individual** - A unique linear model per patient.

$$y^p = X^p\beta^p + \epsilon, \quad p = \{1, \dots, N_p\} \quad (\text{G.2})$$

- **LMEM** - A mixed-effects linear model that accounts for the inter-dependency of multiple observations of the same patient (see appendix D.1.2). The random effect parameters, u are assumed to be drawn from a Gaussian distribution.

$$y^p = X\beta + Zu^p + \epsilon, \quad u \sim \mathcal{N}(0, G), \quad p = \{1, \dots, N_p\} \quad (\text{G.3})$$

where the superscript p denotes data from only the p^{th} patient. All other nomenclature are as defined in appendix D.

We analysed the impact of these different approaches on the modelled calibration coefficients between PAT and SBP. We then subsequently compared the regression performance of nocturnal SBP estimation using patient-wise ρ_p , RMSE, and MAE and performance metrics.

G.3 Results

Figure G.1 provides a visual comparison of the different linear models produced by the three approaches for a subset of 11 patients in the overall cohort. Fixed-effect modelling did not account for the unique and individualised relationship between PAT and SBP, and produced a global, least-squares, relationship across all patients. This model, therefore, was inadequate for a number of patients, see for example IDs: 203888, 204766, and 269759, where the fitted model was far from the training set observations.

Individual models accounted for the individualised relationship, however, for data collected in a noisy environment (such as the ICU) these models are prone to overfitting, especially for patients with very few observations. Consider ID: 2621518, who had only 2 observations of SBP and PAT during the training period between 8am and midnight. The individual linear model resulted in a, non-physiological¹, positive PAT calibration gradient, demonstrating that the approach has overfitted to noise present when recording data in the ICU.

LMEM also accounted for the individualised relationship, however, the calibration parameters were regularised to reduce overfitting by assuming a Gaussian prior of the random-effect coefficients, u . The hyperprior covariance matrix is learned across the cohort using the expectation maximisation algorithm presented in appendix D.1.2. As a result, for patients such as ID: 2621518 (who had only 2 observations of SBP and PAT), the calibration slope was regularised, resulting in a negative gradient. This effect can additionally be observed for patients such as ID: 234863, and 274771.

Figure G.2 shows the distributions of patient-wise $\Delta\text{SBP}/\Delta\text{PAT}$ (the fitted calibration slope) computed across the cohort using the three approaches. As they were fitted across the cohort, fixed-effect slopes were consistent. LMEM regularised the slopes of the individual approach resulting in a tighter distribution with fewer values at the non-physiological extremes.

Table G.1 shows the results of nocturnal SBP estimation using the three approaches when trained on PAT and SBP observations collected between 8am and midnight when data is recorded in a noisy environment. LMEM resulted in significantly ($p < 0.001$) lower error values than both the fixed-effects and individual approaches. Figure G.3 shows the RMSE error for nocturnal SBP estimation grouped by the number of observations between 8am and midnight for each patient. While LMEM resulted in lower errors across all number of observations, the most significant difference between the individual and LMEM approaches occurred for individuals who had only 2 observations of SBP and PAT during the training period.

¹Chapter 5 demonstrated that there is a strong prior knowledge that PAT is negatively correlated with SBP.

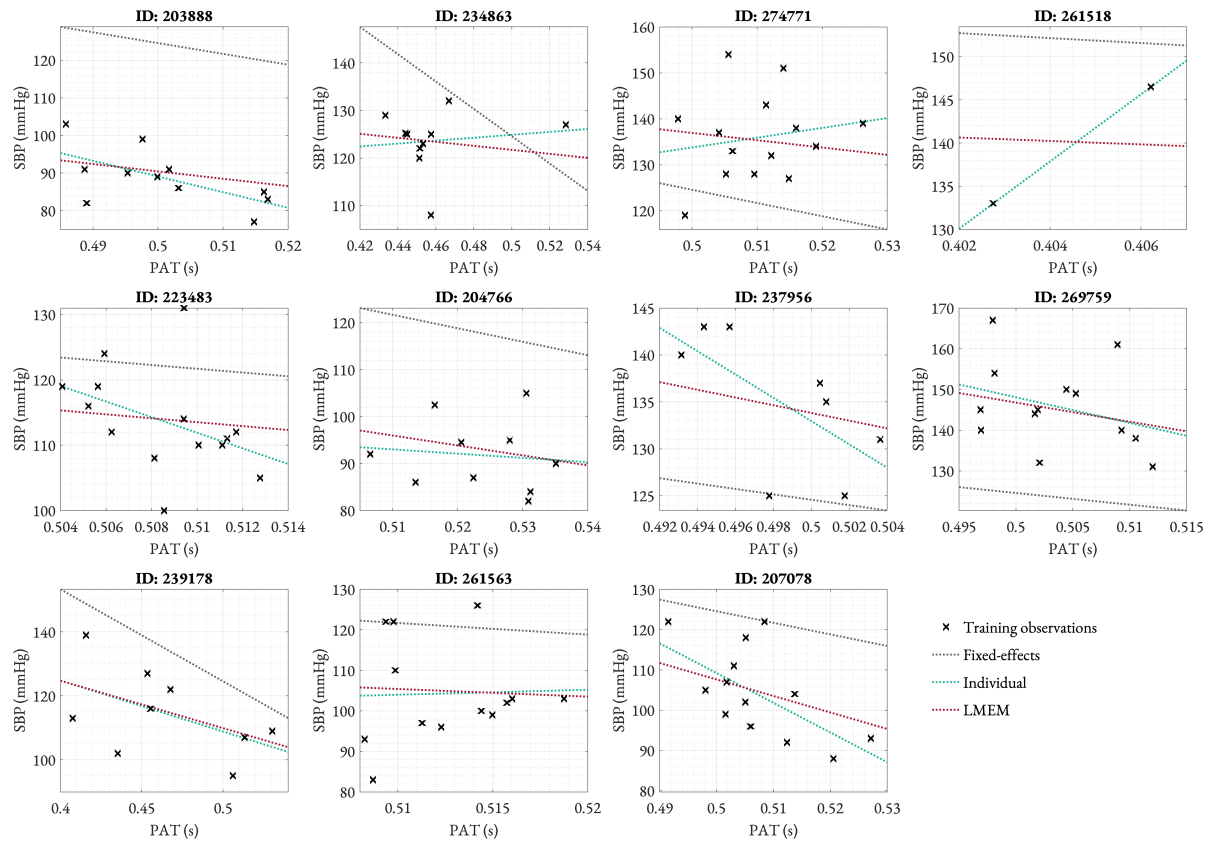


Figure G.1: Comparison of different linear models produced by the three approaches for a subset of 11 patients in the overall cohort. Fixed-effect models are produced across the cohort and therefore do not reflect the individualised nature of the PAT-BP relationship. Individual models account for this, however are prone to overfitting for patients that do not have many observations. LMEM models regularise the individual approach to reduce overfitting.

Table G.1: Performance statistics of nocturnal SBP estimation using PAT and the three approaches. Results are given as median (IQR) computed across all patients. Entries in bold indicate the best performance for that metric.

Model name	ρ_p	RMSE *	MAE *
Fixed-effects	0.34 (0.64)	15.85 (12.88)	13.99 (12.99)
Individual	0.29 (0.78)	11.07 (7.37)	9.30 (6.36)
LMEM	0.34 (0.65)	10.23 (6.22)	8.60 (5.47)

* results given in units of mmHg.

G.4 Discussion and conclusion

In this appendix, we have demonstrated the significant advantages of mixed-effects modelling for nocturnal SBP estimation using PAT observations during the period between 8am to midnight. Mixed-effects modelling was especially advantageous for reducing model overfitting for individuals with very few training observations. The results in this chapter justify our use of mixed-effects modelling for nocturnal SBP estimation using method 1 in section 8.4.

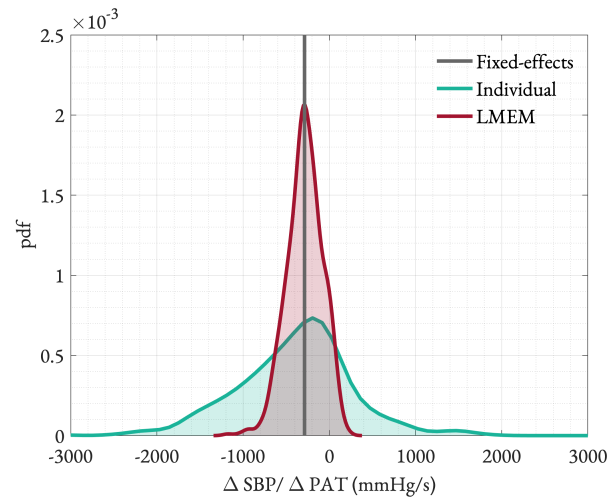


Figure G.2: Distribution of PAT-BP calibration slopes ($\Delta\text{SBP}/\Delta\text{PAT}$) modelled by the three approaches across all patients in the cohort.

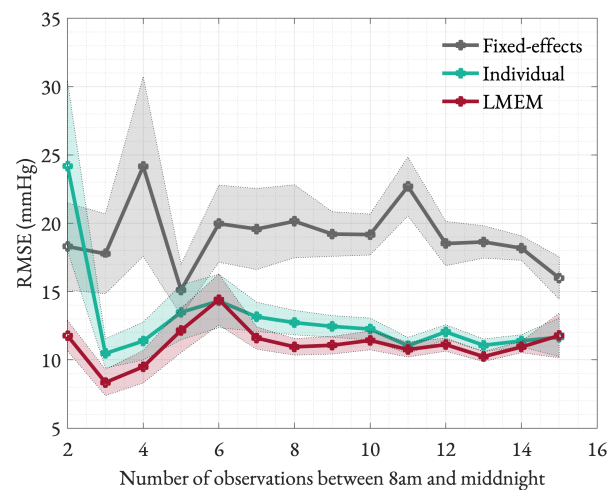


Figure G.3: RMSE nocturnal SBP estimation as a function of the number of observations a patient had between 8am and midnight (the training set). The solid line shows the patient-wise mean RMSE. The shaded area shows the patient-wise standard error around the quoted mean.

H

Supplementary results: Towards cuffless estimation of ambulatory blood pressure

H.1 Introduction

In this appendix, we provide additional results to supplement the work presented in chapter 9.

H.2 Reference blood pressure double cosinor model fits

Figure H.1 shows the results of the double-cosinor model fitted to the good-quality SBP measurements for all participant's Day₇ data. Assessing model goodness-of-fit using an F-test (see section 7.6.2) indicates that the double-cosinor model is appropriate, at the $p < 0.05$ level, for all participants apart from ID: 1013. For this participant (who reported significant difficulty sleeping due to the measurement equipment) there was significant dispersion in BP values across the 24 hours. The differences between DR_{SBP} values computed from the double-cosinor smoothed ABPM time series and the original good-quality data were found to be statistically non-significant (Wilcoxon signed-rank test: $p = 0.091$).

H.3 Model performance for mean arterial pressure and diastolic blood pressure estimation

Tables H.1 and H.2 shows the median and IQR performance statistics for all models across all participants in the nABPM dataset for ΔMAP and ΔDBP estimation respectively.

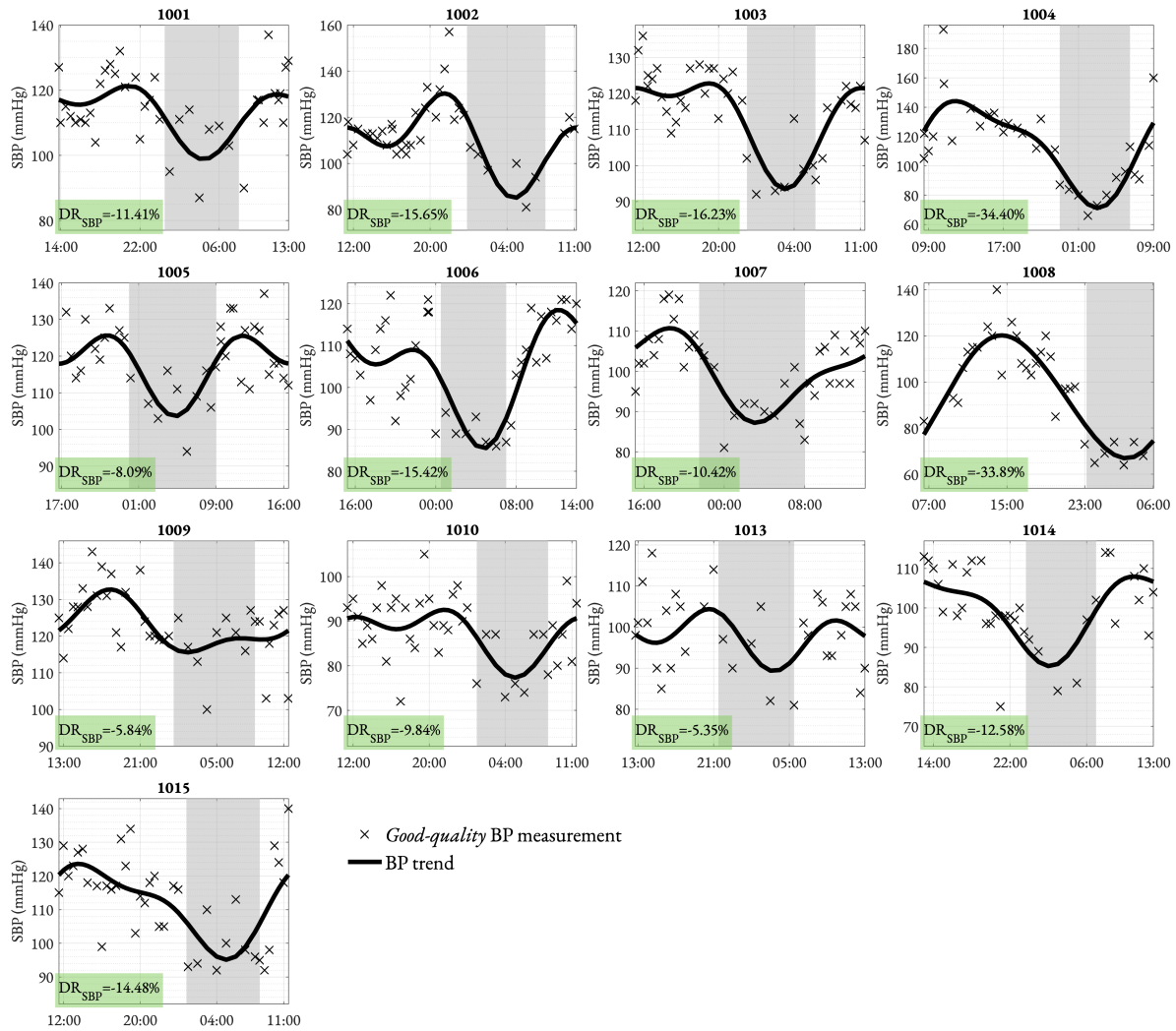


Figure H.1: Double-cosinor fitting for all participant's Day7 good-quality ABPM data to compute the reference BP time series. For one participant (ID: 1013) the double-cosinor model did not pass the goodness-of-fit F-test at the $p < 0.05$ level. The shaded area indicates self-reported sleep time. The timestamp at the left-hand edge of the plot indicates the time at which the BP monitoring started for that study participant, and it notably varies across different participants

Table H.1: Performance statistics of Day7 MAP estimation using the models proposed. DR_{MAP} errors are given as absolute differences with respect to the reference. Results are given as median (IQR) across all participants. Entries in bold indicate the best performance for that metric.

Model name	ρ_p	RMSE *	MAE *	$DR_{MAP}^\dagger$
<i>Individual calibration</i>				
Baseline reference	- (-)	8.00 (2.44)	6.49 (2.37)	16.75 (7.73)
ELAS+OLS _{PPG}	0.84 (0.17)	5.85 (3.55)	4.51 (2.97)	7.10 (8.68)
RF _{PPG}	0.87 (0.23)	5.25 (3.91)	4.15 (3.09)	9.73 (12.90)
ELAS+OLS _{PPG+PAT}	0.89 (0.15)	6.21 (5.97)	5.05 (5.81)	6.51 (9.18)
RF _{PPG+PAT}	0.86 (0.22)	5.61 (4.00)	4.50 (3.02)	9.97 (13.49)
<i>Hybrid calibration</i>				
Baseline reference ELAS+OLS	0.86 (0.15)	9.03 (10.19)	8.51 (9.59)	7.08 (5.01)
Baseline reference RF	0.88 (0.11)	6.53 (5.28)	5.37 (5.66)	6.74 (7.29)
ELAS+OLS _{PPG}	0.90 (0.28)	11.62 (15.93)	10.54 (15.35)	9.39 (13.23)
RF _{PPG}	0.82 (0.40)	7.56 (2.63)	6.76 (2.52)	13.99 (12.21)
ELAS+OLS _{PPG+PAT}	0.92 (0.23)	11.09 (15.41)	10.04 (14.95)	9.05 (14.00)
RF _{PPG+PAT}	0.86 (0.44)	7.96 (2.79)	6.72 (2.77)	13.73 (11.92)

* results given in units of mmHg. † results given in units of %.

Table H.2: Performance statistics of Day7 DBP estimation using the models proposed. DR_{DBP} errors are given as absolute differences with respect to the reference. Results are given as median (IQR) across all participants. Entries in bold indicate the best performance for that metric.

Model name	ρ_p	RMSE *	MAE *	$DR_{DBP}^\dagger$
<i>Individual calibration</i>				
Baseline reference	- (-)	8.49 (3.70)	7.00 (4.27)	16.75 (7.73)
ELAS+OLS _{PPG}	0.82 (0.20)	4.33 (3.99)	3.40 (2.78)	7.65 (8.29)
RF _{PPG}	0.88 (0.14)	5.06 (2.89)	4.05 (2.38)	9.27 (14.17)
ELAS+OLS _{PPG+PAT}	0.86 (0.15)	5.44 (3.95)	3.90 (4.20)	7.00 (6.97)
RF _{PPG+PAT}	0.88 (0.16)	5.35 (2.66)	4.01 (2.59)	9.86 (14.24)
<i>Hybrid calibration</i>				
Baseline reference ELAS+OLS	0.84 (0.14)	8.53 (11.43)	7.92 (12.15)	7.78 (6.23)
Baseline reference RF	0.89 (0.14)	7.31 (6.78)	6.94 (7.56)	7.31 (6.70)
ELAS+OLS _{PPG}	0.89 (0.55)	12.20 (13.82)	11.72 (14.33)	10.22 (10.33)
RF _{PPG}	0.80 (0.44)	7.77 (6.96)	7.07 (7.23)	14.42 (12.10)
ELAS+OLS _{PPG+PAT}	0.88 (0.52)	11.98 (17.16)	10.87 (18.47)	9.78 (10.04)
RF _{PPG+PAT}	0.70 (0.93)	7.94 (6.67)	7.24 (7.04)	14.39 (11.38)

* results given in units of mmHg. † results given in units of %.

References

- [1] H. Wang, M. Naghavi, C. Allen, R. M. Barber, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen, M. M. Coates, M. Coggeshall, L. Dandona, D. J. Dicker, H. E. Erskine, J. A. Haagsma, C. Fitzmaurice, K. Foreman, M. H. Forouzanfar, M. S. Fraser, N. Fullman, E. M. Goldberg, N. Graetz, J. A. Haagsma, S. I. Hay, C. Huynh, C. O. Johnson, N. J. Kassebaum, X. R. Kulikoff, M. Kutz, H. H. Kyu, H. J. Larson, et al. “Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015”. In: *The Lancet* 388.10053 (Oct. 2016), pp. 1459–1544.
- [2] Shaun Davidson, Mauricio Villarroel, Mirae Harford, Eoin Finnegan, Joao Jorge, Duncan Young, Peter Watkinson, and Lionel Tarassenko. “Vital-sign circadian rhythms in patients prior to discharge from an ICU: a retrospective observational analysis of routinely recorded physiological data”. In: *Critical Care* 24.1 (2020), pp. 1–13.
- [3] Lauren G Douma and Michelle L Gumz. “Circadian clock-mediated regulation of blood pressure”. In: *Free radical biology and medicine* 119 (2018), pp. 108–114.
- [4] Ethan D. Buhr and Joseph S. Takahashi. “Molecular components of the mammalian circadian clock”. In: *Handbook of Experimental Pharmacology* 217.suppl2 (2013), pp. 3–27.
- [5] Laura Catherine Armitage, Shaun Davidson, Adam Mahdi, Mirae Harford, Richard McManus, Andrew Farmer, Peter Watkinson, and Lionel Tarassenko. “Diagnosing hypertension in primary care: the importance of night-time blood pressure assessment”. In: *British Journal of General Practice* (2022).
- [6] P. Iqbal and Louise Stevenson. “Cardiovascular outcomes in patients with normal and abnormal 24-hour ambulatory blood pressure monitoring”. In: *International Journal of Hypertension* 2011 (2011).
- [7] Asadur Rahman, Arif Ul Hasan, Akira Nishiyama, and Hiroyuki Kobori. “Altered circadian timing system-mediated non-dipping pattern of blood pressure and associated cardiovascular disorders in metabolic and kidney diseases”. In: *International journal of molecular sciences* 19.2 (2018), p. 400.
- [8] Paul Muntner, Daichi Shimbo, Robert M Carey, Jeanne B Charleston, Trudy Gaillard, Sanjay Misra, Martin G Myers, Gbenga Ogedegbe, Joseph E Schwartz, and Raymond R Townsend. “Measurement of blood pressure in humans: a scientific statement from the American Heart Association”. In: *Hypertension* 73.5 (2019), e35–e66.
- [9] Ramakrishna Mukkamala, Jin-Oh Hahn, and Anand Chandrasekhar. “Photoplethysmography in noninvasive blood pressure monitoring”. In: *Photoplethysmography*. Elsevier, 2022. Chap. 11, pp. 359–400.
- [10] Richard E Klabunde. *Cardiovascular physiology concepts*. Second edition. Philadelphia, PA : Lippincott Williams & Wilkins/Wolters Kluwer, [2012] ©2012.
- [11] J. F. Augusto, Jean Louis Teboul, Peter Radermacher, and P. Asfar. “Interpretation of blood pressure signal: Physiological bases, clinical relevance, and objectives during shock states”. In: *Applied Physiology in Intensive Care Medicine 2: Physiological Reviews and Editorials*. 2012, pp. 293–301.
- [12] Sverre E. Kjeldsen. “Hypertension and cardiovascular risk: General aspects”. In: *Pharmacological Research* 129 (2018), pp. 95–99.
- [13] Katherine T Mills, Andrei Stefanescu, and Jiang He. “The global epidemiology of hypertension”. In: *Nature Reviews Nephrology* 16.4 (2020), pp. 223–237.
- [14] Oreste Marrone and Maria R Bonsignore. “Blood-pressure variability in patients with obstructive sleep apnea: current perspectives”. In: *Nat Sci Sleep* 10 (2018), pp. 229–242.

- [15] Elizabeth Silaid Muxfeldt, Claudia Regina Lopes Cardoso, and Gil Fernando Salles. *Prognostic value of nocturnal blood pressure reduction in resistant hypertension*. Tech. rep. 9. American Medical Association, 2009, pp. 874–880.
- [16] Clara K Chow, Koon K Teo, Sumathy Rangarajan, Shofiqul Islam, Rajeev Gupta, Alvaro Avezum, Ahmad Bahonar, Jephath Chifamba, Gilles Dagenais, and Rafael Diaz. “Prevalence, awareness, treatment, and control of hypertension in rural and urban communities in high-, middle-, and low-income countries”. In: *Jama* 310.9 (2013), pp. 959–968.
- [17] Bryan Williams, Giuseppe Mancina, Wilko Spiering, Enrico Agabiti Rosei, Michel Azizi, Michel Burnier, Denis L Clement, Antonio Coca, Giovanni De Simone, and Anna Dominiczak. “2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology (ESC) and the European Society of Hypertension (ESH)”. In: *European heart journal* 39.33 (2018), pp. 3021–3104.
- [18] Bernd Volker Scheer, Azriel Perel, and Ulrich J Pfeiffer. “Clinical review: complications and risk factors of peripheral arterial catheters used for haemodynamic monitoring in anaesthesia and intensive care medicine”. In: *Critical care* 6 (2002), pp. 1–7.
- [19] Heather H. Hager and Bracken Burns. *Artery Cannulation*. 2019.
- [20] Xiao Rong Ding, Ni Zhao, Guang Zhong Yang, Roderic I. Pettigrew, Benny Lo, Fen Miao, Ye Li, Jing Liu, and Yuan Ting Zhang. “Continuous Blood Pressure Measurement From Invasive to Unobtrusive: Celebration of 200th Birth Anniversary of Carl Ludwig”. In: *IEEE Journal of Biomedical and Health Informatics* 20.6 (2016), pp. 1455–1465.
- [21] Kathleen Chavanu, Jennifer Merkel, and Allen M. Quan. “Role of ambulatory blood pressure monitoring in the management of hypertension”. In: *American Journal of Health-System Pharmacy* 65.3 (2008), pp. 209–218.
- [22] Rajiv Agarwal and Robert P. Light. “The effect of measuring ambulatory blood pressure on nighttime sleep and daytime activity - Implications for dipping”. In: *Clinical Journal of the American Society of Nephrology* 5.2 (2010), pp. 281–285.
- [23] Anthony J. Viera, Kara Lingley, and Alan L. Hinderliter. “Tolerability of the Oscar 2 ambulatory blood pressure monitor among research participants: A cross-sectional repeated measures study”. In: *BMC Medical Research Methodology* 11.1 (2011), p. 59.
- [24] Anna Vybornova, Erietta Polychronopoulou, Arlene Wurznier-Ghajarzadeh, Sibylle Fallet, Josep Sola, and Gregoire Wuerzner. “Blood pressure from the optical Aktiia Bracelet: A 1-month validation study using an extended ISO81060-2 protocol adapted for a cuffless wrist device”. In: *Blood Pressure Monitoring* 30 (2021).
- [25] Ramakrishna Mukkamala, George S Stergiou, and Alberto P Avolio. “Cuffless blood pressure measurement”. In: *Annual Review of Biomedical Engineering* 24 (2022).
- [26] Isabella Tan, Sonali R Gnanenthiran, Justine Chan, Konstantinos G Kyriakoulis, Markus P Schlaich, George S Stergiou, and Aletta E Schutte. “Comparison Of 24-hour Blood Pressure Tracking Between A Commercially Available Cuffless Wearable Device And Standard Ambulatory Monitoring”. In: *Hypertension* 79.Suppl₁ (2022), A088–A088.
- [27] George S Stergiou, Ramakrishna Mukkamala, Alberto Avolio, Konstantinos G Kyriakoulis, Stephan Mieke, Alan Murray, Gianfranco Parati, Aletta E Schutte, James E Sharman, and Roland Asmar. “Cuffless blood pressure measuring devices: review and statement by the European Society of Hypertension Working Group on Blood Pressure Monitoring and Cardiovascular Variability”. In: *Journal of Hypertension* 40.8 (2022), pp. 1449–1460.
- [28] John Allen. “Photoplethysmography and its application in clinical physiological measurement”. In: *Physiological Measurement* 28.3 (2007).

- [29] Andrew T Reisner, Gari D Clifford, and Roger G Mark. "The physiological basis of the electrocardiogram". In: *Advanced methods and tools for ECG data analysis* 1 (2006), p. 25.
- [30] Angel Villegas, David McEneaney, and Omar Escalona. "Arm-ECG wireless sensor system for wearable long-term surveillance of heart arrhythmias". In: *Electronics* 8.11 (2019), p. 1300.
- [31] Ramakrishna Mukkamala, Jin Oh Hahn, Omer T. Inan, Lalit K. Mestha, Chang Sei Kim, Hakan Toreyin, and Survi Kyal. "Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice". In: *IEEE Transactions on Biomedical Engineering* 62.8 (2015), pp. 1879–1901.
- [32] Mirae Harford, Samuel H Clark, Jodie F Smythe, Stephen Gerry, Mauricio Villarroel, Joao Jorge, Sitthichok Chaichulee, Lionel Tarassenko, Duncan Young, and Peter Watkinson. "Non-invasive stroke volume estimation by transthoracic electrical bioimpedance versus Doppler echocardiography in healthy volunteers". In: *Journal of medical engineering & technology* 43.1 (2019), pp. 33–37.
- [33] Gerard J Tortora and Bryan H Derrickson. *Principles of anatomy and physiology*. John Wiley & Sons, 2018.
- [34] Philip I Aaronson, Jeremy P T Ward, and Michelle J Connolly. *The cardiovascular system at a glance*. John Wiley & Sons, 2020.
- [35] David Shier, Jackie Butler, and Ricki Lewis. *Hole's essentials of human anatomy & physiology*. McGraw-Hill Education New York, 2015.
- [36] Gail F. Baker, Gerard J. Tortora, and Nicholas Peter Anag Nostakos. *Principles of Anatomy and Physiology*. 14th ed. Vol. 76. 3. John Wiley & Sons, 1976, p. 477.
- [37] Gbenga Ogedegbe and Thomas Pickering. "Principles and techniques of blood pressure measurement". In: *Cardiology clinics* 28.4 (2010), pp. 571–586.
- [38] Anne Waugh and Allison Grant. *Ross and Wilson anatomy and physiology in health and illness*. Harcourt Publishers Limited, 2001.
- [39] Roger A L Dampney. "Resetting of the baroreflex control of sympathetic vasomotor activity during natural behaviors: description and conceptual model of central mechanisms". In: *Frontiers in neuroscience* 11 (2017), p. 461.
- [40] Claude Julien. "The enigma of Mayer waves: facts and models". In: *Cardiovascular research* 70.1 (2006), pp. 12–21.
- [41] Yue Leng, Erik S Musiek, Kun Hu, Francesco P Cappuccio, and Kristine Yaffe. "Association between circadian rhythms and neurodegenerative diseases". In: *The Lancet Neurology* 18.3 (2019), pp. 307–318.
- [42] Patricia M Wong, Brant P Hasler, Thomas W Kamarck, Matthew F Muldoon, and Stephen B Manuck. "Social jetlag, chronotype, and cardiometabolic risk". In: *The Journal of Clinical Endocrinology & Metabolism* 100.12 (2015), pp. 4612–4620.
- [43] Katja A. Lamia. *Ticking time bombs: Connections between circadian clocks and cancer*. 2017.
- [44] Martin E. Young and Molly S. Bray. "Potential role for peripheral circadian clock dyssynchrony in the pathogenesis of cardiovascular dysfunction". In: *Sleep Medicine* 8.6 (2007), pp. 656–667.
- [45] Till Roenneberg, Karla V. Allebrandt, Martha Mellow, and Céline Vetter. "Social jetlag and obesity". In: *Current Biology* 22.10 (2012), pp. 939–943.
- [46] Justine Huart, Alexandre Persu, Jean-Philippe Lengelé, Jean-Marie Krzesinski, François Jouret, and George Stergiou. "Pathophysiology of the Nondipping Blood Pressure Pattern". In: *Hypertension* (Jan. 2023).
- [47] Michael H. Smolensky, Ramón C. Hermida, Richard J. Castriotta, and Francesco Portaluppi. "Role of sleep-wake cycle on blood pressure circadian rhythms and hypertension". In: *Sleep Medicine* 8.6 (2007), pp. 668–680.

- [48] Kathryn S. Taylor, Carl J. Heneghan, Richard J. Stevens, Emily C. Adams, David Nunan, and Alison Ward. "Heterogeneity of prognostic studies of 24-hour blood pressure variability: Systematic review and meta-analysis". In: *PLoS ONE* 10.5 (2015), e0126375.
- [49] Klaus Witte, Anke Schnecko, Ruud M Buijs, Jan van der Vliet, Elizabeth Scalbert, Philippe Delagrangé, Beatrice Guardiola-Lemaitre, and B Lemmer. "Effects of SCN lesions on orcadian blood pressure rhythm in normotensive and transgenic hypertensive rats". In: *Chronobiology international* 15.2 (1998), pp. 135–145.
- [50] Eoin O'Brien, James Sheridan, and Kevin O'Malley. "Dippers and Non-Dippers". In: *The Lancet* 332.8607 (1988), p. 397.
- [51] Cesare Cuspidi, Carla Sala, Marijana Tadic, Elisa Gherbesi, Antonio De Giorgi, Guido Grassi, and Giuseppe Mancia. "Clinical and prognostic significance of a reverse dipping pattern on ambulatory monitoring: an updated review". In: *The Journal of Clinical Hypertension* 19.7 (2017), pp. 713–721.
- [52] T Quinaglia, L C Martins, V N Figueiredo, R C Santos, J C Yugar-Toledo, J F Martin, C Demacq, E Pimenta, D A Calhoun, and H Moreno. "Non-dipping pattern relates to endothelial dysfunction in patients with uncontrolled resistant hypertension". In: *Journal of human hypertension* 25.11 (2011), pp. 656–664.
- [53] William G Haynes. "Role of leptin in obesity-related hypertension". In: *Experimental physiology* 90.5 (2005), pp. 683–688.
- [54] Takashi Fujii, Takashi Uzu, Masataka Nishimura, Masanobu Takeji, Setsuko Kuroda, Satoko Nakamura, Takashi Inenaga, and Genjiro Kimura. "Circadian rhythm of natriuresis is disturbed in nondipper type of essential hypertension". In: *American journal of kidney diseases* 33.1 (1999), pp. 29–35.
- [55] Takayoshi Ohkubo, Atsushi Hozawa, Junko Yamaguchi, Masahiro Kikuya, Kaori Ohmori, Mari Michimata, Mitsunobu Matsubara, Junichiro Hashimoto, Haruhisa Hoshi, Tsutomu Araki, Ichiro Tsuji, Hiroshi Satoh, Shigeru Hisamichi, and Yutaka Imai. "Prognostic significance of the nocturnal decline in blood pressure in individuals with and without high 24-h blood pressure: The ohasama study". In: *Journal of Hypertension* 20.11 (2002), pp. 2183–2189.
- [56] Koldo Berganzo, Begoña Díez-Arrola, Beatriz Tijero, Johanne Somme, Elena Lezcano, Verónica Llorens, Iratxe Ugarriza, Roberto Giordia, J C Gómez-Esteban, and Juan J Zarranz. "Nocturnal hypertension and dysautonomia in patients with Parkinson's disease: are they related?" In: *Journal of neurology* 260.7 (2013), pp. 1752–1756.
- [57] Ramón C Hermida, Diana E Ayala, Artemio Mojón, and José R Fernández. "Influence of circadian time of hypertension treatment on cardiovascular risk: results of the MAPEC study". In: *Chronobiology international* 27.8 (2010), pp. 1629–1651.
- [58] Björn Lemmer and Martin Middeke. "A commentary on the Spanish hypertension studies MAPEC and HYGIA". In: *Chronobiology International* 37.5 (2020), pp. 728–730.
- [59] Husam M Salah and Jawahar L Mehta. "Best time for administration of antihypertensive medications: morning or evening?" In: *The Lancet* (2022).
- [60] Isla S Mackenzie, Amy Rogers, Neil R Poulter, Bryan Williams, Morris J Brown, David J Webb, Ian Ford, David A Rorie, Greg Guthrie, and J W Kerr Grieve. "Cardiovascular outcomes in adults with hypertension with evening versus morning dosing of usual antihypertensives in the UK (TIME study): a prospective, randomised, open-label, blinded-endpoint clinical trial". In: *The Lancet* (2022).
- [61] Wen-Yi Yang, Jesus D Melgarejo, Lutgarde Thijs, Zhen-Yu Zhang, José Boggia, Fang-Fei Wei, Tine W Hansen, Kei Asayama, Takayoshi Ohkubo, and Jørgen Jeppesen. "Association of office and ambulatory blood pressure with mortality and cardiovascular outcomes". In: *Jama* 322.5 (2019), pp. 409–420.

- [62] Keisuke Narita, Satoshi Hoshide, and Kazuomi Kario. “Seasonal variation in blood pressure: current evidence and recommendations for hypertension management”. In: *Hypertension Research* 44.11 (2021), pp. 1363–1372.
- [63] Anastasios Kollias, Konstantinos G Kyriakoulis, Emelina Stambolliu, Angeliki Ntineri, Ioannis Anagnostopoulos, and George S Stergiou. “Seasonal blood pressure variation assessed by different measurement methods: systematic review and meta-analysis”. In: *Journal of hypertension* 38.5 (2020), pp. 791–798.
- [64] Susan Cheng, Vanessa Xanthakis, Lisa M Sullivan, and Ramachandran S Vasan. “Blood pressure tracking over the adult life course: patterns and correlates in the Framingham heart study”. In: *Hypertension* 60.6 (2012), pp. 1393–1399.
- [65] Gary F. Mitchell, Shih Jen Hwang, Ramachandran S. Vasan, Martin G. Larson, Michael J. Pencina, Naomi M. Hamburg, Joseph A. Vita, Daniel Levy, and Emelia J. Benjamin. “Arterial stiffness and cardiovascular events: The framingham heart study”. In: *Circulation* 121.4 (2010), pp. 505–511.
- [66] Elisabete Pinto. “Blood pressure and ageing”. In: *Postgraduate medical journal* 83.976 (2007), pp. 109–114.
- [67] A. H.E.M. Maas and H. R. Franke. *Women’s health in menopause with a focus on hypertension*. Feb. 2009.
- [68] Katharina Wolf-Maier, Richard S. Cooper, José R. Banegas, Simona Giampaoli, Hans Werner Hense, Michel Joffres, Mika Katarinen, Neil Poulter, Paola Primatesa, Fernando Rodríguez-Artalejo, Birgitta Stegmayr, Michael Thamm, Jaakko Tuomilehto, Diego Vanuzzo, and Fenicia Vescio. *Hypertension Prevalence and Blood Pressure Levels in 6 European Countries, Canada, and the United States*. May 2003.
- [69] Bin Zhou, Majid Ezzati, James Benthams, Mariachiara Di Cesare, Honor Bixby, Goodarz Danaei, Melanie J Cowan, Christopher J Paciorek, Gitanjali Singh, Kaveh Hajifathalian, James E Bennett, Cristina Taddei, Ver Bilano, Rodrigo M Carrillo-Larco, Shirin Djalalinia, Shahab Khatibzadeh, Charles Lugero, Niloofar Peykari, Wan Zhu Zhang, Yuan Lu, Gretchen A Stevens, Leanne M Riley, Pascal Bovet, Paul Elliott, Dongfeng Gu, Nayu Ikeda, Rod T Jackson, Michel Joffres, Andre Pascal Kengne, Tiina Laatikainen, et al. “Worldwide trends in blood pressure from 1975 to 2015 a pooled analysis of 1479 population-based measurement studies with 19.1 million participants”. In: *The Lancet*. Vol. 389:10064, The Lancet, 2017.
- [70] Praveen Surendran, Fotios Drenos, Robin Young, Helen Warren, James P Cook, Alisa K Manning, Niels Grarup, Xueling Sim, Daniel R Barnes, and Kate Witkowska. “Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension”. In: *Nature genetics* 48.10 (2016), pp. 1151–1161.
- [71] Michael Gurven, Aaron D. Blackwell, Daniel Eid Rodríguez, Jonathan Stieglitz, and Hillard Kaplan. “Does blood pressure inevitably rise with age?: Longitudinal evidence among forager-horticulturalists”. In: *Hypertension* 60.1 (July 2012), pp. 25–33.
- [72] Bin Zhou, Pablo Perel, George A Mensah, and Majid Ezzati. “Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension”. In: *Nature Reviews Cardiology* 18.11 (2021), pp. 785–802.
- [73] Thomas G Pickering, John E Hall, Lawrence J Appel, Bonita E Falkner, John Graves, Martha N Hill, Daniel W Jones, Theodore Kurtz, Sheldon G Sheps, and Edward J Roccella. “Recommendations for blood pressure measurement in humans and experimental animals: part 1: blood pressure measurement in humans: a statement for professionals from the Subcommittee of Professional and Public Education of the American Heart Association Cou”. In: *Hypertension* 45.1 (2005), pp. 142–161.
- [74] Anna Kruyer, Nadine Soplop, Sidney Strickland, and Erin H Norris. “Chronic hypertension leads to neurodegeneration in the TgSwDI mouse model of Alzheimer’s disease”. In: *Hypertension* 66.1 (2015), pp. 175–182.

- [75] Mahesh Nirmalan and Paul M. Dark. "Broader applications of arterial pressure wave form analysis". In: *Continuing Education in Anaesthesia, Critical Care and Pain* 14.6 (2014), pp. 285–290.
- [76] S. A.A.P. Hoeksel, J. R.C. Jansen, J. A. Blom, and J. J. Schreuder. "Detection of dicrotic notch in arterial pressure signals". In: *Journal of Clinical Monitoring* 13.5 (1997), pp. 309–316.
- [77] Sandrine C Millasseau, James M Ritter, Kenji Takazawa, and Philip J Chowienczyk. "Contour analysis of the photoplethysmographic pulse measured at the finger". In: *Journal of hypertension* 24.8 (2006), pp. 1449–1456.
- [78] Mohamed Elgendi. "On the Analysis of Fingertip Photoplethysmogram Signals". In: *Current Cardiology Reviews* 8.1 (2012), pp. 14–25.
- [79] Kenji Takazawa, Nobuhiro Tanaka, Masami Fujita, Osamu Matsuoka, Tokuyu Saiki, Masaru Aikawa, Sinobu Tamura, and Chiharu Ibukiyama. "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform". In: *Hypertension* 32.2 (1998), pp. 365–370.
- [80] Agnes S. Meidert and Bernd Saugel. "Techniques for non-invasive monitoring of arterial blood pressure". In: *Frontiers in Medicine* 4.JAN (2017), p. 231.
- [81] Matthew Ward and Jeremy A. Langton. "Blood pressure measurement". In: *Continuing Education in Anaesthesia, Critical Care and Pain* 7.4 (2007), pp. 122–126.
- [82] Gert A. Van Montfrans. "Oscillometric blood pressure measurement: Progress and problems". In: *Blood Pressure Monitoring* 6.6 (2001), pp. 287–290.
- [83] Gary Drzewiecki, Rebecca Hood, and H Apple. "Theory of the oscillometric maximum and the systolic and diastolic detection ratios". In: *Annals of biomedical engineering* 22.1 (1994), pp. 88–96.
- [84] George S. Stergiou, Bruce S. Alpert, Stephan Mieke, Jiguang Wang, and Eoin O'Brien. "Validation protocols for blood pressure measuring devices in the 21st century". In: *Journal of Clinical Hypertension*. Vol. 20. 7. Blackwell Publishing Inc., July 2018, pp. 1096–1099.
- [85] Aletta E Schutte, Anastasios Kollias, and George S Stergiou. "Blood pressure and its variability: classic and novel measurement techniques". In: *Nature Reviews Cardiology* (2022), pp. 1–12.
- [86] Andreas Bur, Michael M. Hirschl, Harald Herkner, Elisabeth Oschatz, Julia Kofler, Christian Woisetschläger, and Anton N. Laggnier. "Accuracy of oscillometric blood pressure measurement according to the relation between cuff size and upper-arm circumference in critically ill patients". In: *Critical Care Medicine* 28.2 (2000), pp. 371–376.
- [87] Eoin Finnegan, Mauricio Villarreal, Shaun Davidson, Mirae Harford, Joao Jorge, and Lionel Tarassenko. "Comparing Circadian Rhythms in Blood Pressure Computed From the Arterial Line and Sphygmomanometer in the ICU". In: *2020 Computing in Cardiology*. IEEE, 2020, pp. 1–4.
- [88] Hannu Sorvoja and Risto Myllylä. "Noninvasive Blood Pressure Measurement Methods". In: *Molecular and Quantum Acoustics* 27 (2006), pp. 239–264.
- [89] Nicolas Langwieser, Luisa Prechtel, Agnes S. Meidert, Alexander Hapfelmeier, Christian Bradaric, Tareq Ibrahim, Karl Ludwig Laugwitz, Roland M. Schmid, Julia Y. Wagner, and Bernd Saugel. "Radial artery applanation tonometry for continuous noninvasive arterial blood pressure monitoring in the cardiac intensive care unit". In: *Clinical Research in Cardiology* 104.6 (2015), pp. 518–524.
- [90] Hao-Min Cheng, Dora Lang, Catalin Tufanaru, and Alan Pearson. "Measurement accuracy of non-invasively obtained central blood pressure by applanation tonometry: a systematic review and meta-analysis". In: *International journal of cardiology* 167.5 (2013), pp. 1867–1876.
- [91] J. Penáz. "Criteria for set point estimation in the volume clamp method of blood pressure measurement." In: *Physiological research / Academia Scientiarum Bohemoslovaca* 41.1 (1992), pp. 5–10.
- [92] Bernd Saugel, Ron Dueck, and Julia Y Wagner. "Measurement of blood pressure". In: *Best Practice & Research Clinical Anaesthesiology* 28.4 (2014), pp. 309–322.

- [93] Ayten Elvan-Taşpınar, Leonore A.M.Schoot Uiterkamp, J. Marko Sikkema, Michiel L. Bots, Hein A. Koomans, Hein W. Bruinse, and Arie Franx. “Validation and use of the Finometer™ for blood pressure measurement in normal, hypertensive and pre-eclamptic pregnancy”. In: *Journal of Hypertension* 21.11 (2003), pp. 2053–2060.
- [94] K. H. Wesseling. “Finapres, continuous noninvasive finger arterial pressure based on the method of Peñáz”. In: *Blood Pressure Measurements*. Springer, 1990, pp. 161–172.
- [95] B Silke and D. McAuley. “Accuracy and precision of blood pressure determination with the Finapres: An overview using re-sampling statistics”. In: *Journal of Human Hypertension* 12.6 (1998), pp. 403–409.
- [96] Kong Yien Chin and Ronney B Panerai. “Comparative study of Finapres devices”. In: *Blood Pressure Monitoring* 17.4 (2012), pp. 171–178.
- [97] David M. Reboussin, Norrina B. Allen, Michael E. Griswold, Eliseo Guallar, Yuling Hong, Daniel T. Lackland, Edgar R. Miller, Tamar Polonsky, Angela M. Thompson-Paul, Suma Vupputuri, Glenn N. Levine, Patrick T. O’Gara, Jonathan L. Halperin, Immediate Past, Sana M. Al-Khatib, Joshua A. Beckman, Kim K. Birtcher, Biykem Bozkurt, Ralph G. Brindis, Joaquin E. Cigarroa, Lesley H. Curtis, Anita Deswal, Lee A. Fleisher, Federico Gentile, Samuel Gidding, Zachary D. Goldberger, Mark A. Hlatky, John Ikonomidis, José A. Joglar, Laura Mauri, Susan J. Pressler, Barbara Riegel, Duminda N. Wijeyesundera, Mary Norine Walsh, Shalom Jacobovitz, William J. Oetgen, Mary Anne Elma, Amelia Scholtz, Katherine Sheehan, Abdul R. Abdullah, Naira Tahir, John J. Warner, Nancy Brown, Rose Marie Robertson, Gayle R. Whitman, and Jody Hundley. “Systematic review for the 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults a report of the American College of Cardiology/American Heart Association”. In: *Hypertension* 71.6 (2018), E116–E135.
- [98] Tine W. Hansen, Masahiro Kikuya, Lutgarde Thijs, Kristina Björklund-Bodegård, Tatiana Kuznetsova, Takayoshi Ohkubo, Tom Richart, Christian Torp-Pedersen, Lars Lind, Jørgen Jeppesen, Hans Ibsen, Yutaka Imai, and Jan A. Staessen. “Prognostic superiority of daytime ambulatory over conventional blood pressure in four populations: A meta-analysis of 7030 individuals”. In: *Journal of Hypertension* 25.8 (2007), pp. 1554–1564.
- [99] Alison M. Ward, Osamu Takahashi, Richard Stevens, and Carl Heneghan. “Home measurement of blood pressure and cardiovascular disease: Systematic review and meta-analysis of prospective studies”. In: *Journal of Hypertension* 30.3 (2012), pp. 449–456.
- [100] Thomas G Pickering, Daichi Shimbo, and Donald Haas. “Ambulatory blood-pressure monitoring”. In: *New England Journal of Medicine* 354.22 (2006), pp. 2368–2374.
- [101] Robert Zweiker. “Hypertension news-screen: Effect of self-monitoring and medication self-titration on systolic blood pressure in hypertensive patients at high risk of cardiovascular disease: The TASMIN-SR randomized clinical trial”. In: *Journal fur Hypertonie* 19.1 (2015), pp. 22–23.
- [102] NICE. *Information for the public: High blood pressure*. 2011. URL: <https://www.nice.org.uk/guidance/cg127/resources/information-for-the-public-high-blood-pressure-pdf> (visited on 05/09/2019).
- [103] Gülçin Kantarci. “Home and ambulatory blood pressure monitoring: When who”. In: *Kidney International Supplements*. Vol. 3. 4. Elsevier, 2013, pp. 337–339.
- [104] Mehmet Rami Helvacı and Mahmut Seyhanlı. “What a high prevalence of white coat hypertension in society!” In: *Internal medicine* 45.10 (2006), pp. 671–674.
- [105] Gbenga Ogedegbe, Charles Agyemang, and Joseph E. Ravenell. “Masked hypertension: Evidence of the need to treat”. In: *Current Hypertension Reports* 12.5 (2010), pp. 349–355.

- [106] Lindsey Elliot and Pervaiz Iqbal. “Factors associated with probability of patient rejecting a repeat 24 h ambulatory blood pressure monitoring, despite recommendation by the physician”. In: *Blood Pressure Monitoring* 8.5 (2003), pp. 191–194.
- [107] Paul Little, Jane Barnett, Lucy Barnsley, Jean Marjoram, Alex Fitzgerald-Barron, and David Mant. “Comparison of acceptability of and preferences for different methods of measuring blood pressure in primary care”. In: *British Medical Journal* 325.7358 (2002), pp. 258–259.
- [108] R. J.O. Davies, N. E. Jenkins, and J. R. Stradling. “Effect of measuring ambulatory blood pressure on sleep and on blood pressure during sleep”. In: *Bmj* 308.6932 (1994), p. 820.
- [109] Andrew Sherwood, LaBarron K Hill, James A Blumenthal, and Alan L Hinderliter. “The effects of ambulatory blood pressure monitoring on sleep quality in men and women with hypertension: dipper vs. nondipper and race differences”. In: *American Journal of Hypertension* 32.1 (2019), pp. 54–60.
- [110] Kamal Sharma and Hardik D Desai. “Role of self-measured home blood pressure monitoring (HBPM) and effectiveness of telemedicine during the era of COVID-19 pandemic”. In: *SN Comprehensive Clinical Medicine* 3.5 (2021), pp. 1071–1073.
- [111] E G Nasothimiou, N Karpettas, M G Dafni, and G S Stergiou. “Patients’ preference for ambulatory versus home blood pressure monitoring”. In: *Journal of Human Hypertension* 28.4 (2014), pp. 224–229.
- [112] Robert M Carey and Paul K Whelton. “New findings bearing on the prevention, detection and management of high blood pressure”. In: *Current Opinion in Cardiology* 36.4 (2021), pp. 429–435.
- [113] NHS England. *Home blood pressure monitoring*. URL: <https://www.england.nhs.uk/ourwork/clinical-policy/cvd/home-blood-pressure-monitoring/> (visited on 11/10/2022).
- [114] Toshiyo Tamura, Yuka Maeda, Masaki Sekine, and Masaki Yoshida. “Wearable photoplethysmographic sensors—past and present”. In: *Electronics* 3.2 (2014), pp. 282–302.
- [115] Thomas G. Mayerhöfer, Susanne Pahlow, and Jürgen Popp. *The Bouguer-Beer-Lambert Law: Shining Light on the Obscure*. Sept. 2020.
- [116] A. C.M. Dassel, R. Graaff, M. Aardema, W. G. Zijlstra, and J. G. Aarnoudse. “Effect of location of the sensor on reflectance pulse oximetry”. In: *B/OG: An International Journal of Obstetrics and Gynaecology* 104.8 (1997), pp. 910–916.
- [117] Jermana L. Moraes, Matheus X. Rocha, Glauber G. Vasconcelos, José E. Vasconcelos Filho, Victor Hugo C. de Albuquerque, and Auzuir R. Alexandria. *Advances in photoplethysmography signal analysis for biomedical applications*. 2018.
- [118] Jesse Fine, Kimberly L Branan, Andres J Rodriguez, Tananant Boonya-Ananta, Jessica C Ramella-Roman, Michael J McShane, and Gerard L Coté. “Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring”. In: *Biosensors* 11.4 (2021), p. 126.
- [119] Jiapu Pan and Willis J. Tompkins. “A Real-Time QRS Detection Algorithm”. In: *IEEE Transactions on Biomedical Engineering* BME-32.3 (1985), pp. 230–236.
- [120] Xiaochuan He. “Signal Enhancement Applied to Pulse Transit Time Measurement”. PhD thesis. 2015.
- [121] Wen Yi Yang, Blerim Mujaj, Ljupcho Efremov, Zhen Yu Zhang, Lutgarde Thijs, Fang Fei Wei, Qi Fang Huang, Aernout Luttun, Peter Verhamme, Tim S. Nawrot, José Boggia, and Jan A. Staessen. “ECG Voltage in Relation to Peripheral and Central Ambulatory Blood Pressure”. In: *American Journal of Hypertension* 31.2 (2018), pp. 178–187.
- [122] Andrew Sherwood, Michael T Allen, Jochen Fahrenberg, Robert M Kelsey, William R Lovallo, and Lorenz J P Van Doornen. “Methodological guidelines for impedance cardiography”. In: *Psychophysiology* 27.1 (1990), pp. 1–23.

- [123] Mark Ulbrich, Jens Mühlsteff, Steffen Leonhardt, and Marian Walter. "Influence of physiological sources on the impedance cardiogram analyzed using 4D FEM simulations". In: *Physiological measurement* 35.7 (2014), p. 1451.
- [124] PMJM De Vries, J W G Langendijk, and P M Kouw. "The influence of alternating current frequency on flow related admittance changes of blood: a concept for improvement of impedance cardiography". In: *Physiological Measurement* 16.1 (1995), p. 63.
- [125] Yao Xie, Rencheng Song, Dong Yang, Honglong Yu, Cuimin Sun, Qilian Xie, and Ronald X. Xu. "Motion robust ICG measurements using a two-step spectrum denoising method". In: *Physiological Measurement* 42.9 (Sept. 2021).
- [126] Mohamad Forouzanfar, Fiona C. Baker, Ian M. Colrain, and Massimiliano De Zambotti. "Automatic Artifact Detection in Impedance Cardiogram Using Pulse Similarity Index". In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Institute of Electrical and Electronics Engineers Inc., July 2019, pp. 2629–2632.
- [127] Richelle Leanne Gaw. *The effect of red blood cell orientation on the electrical impedance of pulsatile blood with implications for impedance cardiography*. 2010.
- [128] Kiichi Sagawa, Reidar K Lie, and Jochen Schaefer. "Translation of Otto frank's paper" Die Grundform des arteriellen Pulses" zeitschrift für biologie 37: 483-526 (1899)". In: *Journal of molecular and cellular cardiology* 22.3 (1990), pp. 253–254.
- [129] Max E Valentinuzzi and Alberto K Kohen. "Laplace's Law: What It Is About, Where It Comes from, and How It Is Often Applied in Physiology [Retrospectroscope]". In: *IEEE pulse* 2.4 (2011), pp. 74–84.
- [130] Rupert A Payne, Ian B Wilkinson, and David J Webb. "Arterial stiffness and hypertension: emerging concepts". In: *Hypertension* 55.1 (2010), pp. 9–14.
- [131] Guidelines Committee. "2003 European Society of Hypertension–European Society of Cardiology guidelines for the management of arterial hypertension". In: *Journal of hypertension* 21.6 (2003), pp. 1011–1053.
- [132] Alberto Avolio. "Arterial stiffness". In: *Pulse* 1.1 (2013), pp. 14–28.
- [133] P. B. Dobrin and A. A. Rovick. "Influence of vascular smooth muscle on contractile mechanics and elasticity of arteries." In: *The American journal of physiology* 217.6 (1969), pp. 1644–1651.
- [134] Nicholas R. Gaddum, Louise Keehn, Antoine Guilcher, Alberto Gomez, Sally Brett, Philipp Beerbaum, Tobias Schaeffter, and Philip Chowienzyk. "Altered dependence of aortic pulse wave velocity on transmural pressure in hypertension revealing structural change in the aortic wall". In: *Hypertension* 65.2 (2015), pp. 362–369.
- [135] L. A. Geddes. *Handbook of Blood Pressure Measurement*. Springer Science & Business Media, 1991.
- [136] Xiaorong Ding, Yuanling Zhang, and Hon Ki Tsang. "Impact of heart disease and calibration interval on accuracy of pulse transit time-based blood pressure estimation". In: *Physiological Measurement* 37.2 (2016), pp. 227–237.
- [137] Patrick Segers, Ernst R Rietzschel, and Julio A Chirinos. "How to measure arterial stiffness in humans". In: *Arteriosclerosis, thrombosis, and vascular biology* 40.5 (2020), pp. 1034–1043.
- [138] Jonathan P Mynard, Avinash Kondiboyina, Remi Kowalski, Michael M H Cheung, and Joseph J Smolich. "Measurement, analysis and interpretation of pressure/flow waves in blood vessels". In: *Frontiers in Physiology* (2020), p. 1085.
- [139] Peter H Charlton, Jorge Mariscal Harana, Samuel Vennin, Ye Li, Phil Chowienzyk, and Jordi Alastruey. "Modeling arterial pulse waves in healthy aging: a database for in silico evaluation of hemodynamics and pulse wave indexes". In: *American Journal of Physiology-Heart and Circulatory Physiology* 317.5 (2019), H1062–H1085.

- [140] E. Messas, M. Pernot, and M. Couade. "Arterial wall elasticity: State of the art and future prospects". In: *Diagnostic and Interventional Imaging* 94.5 (2013), pp. 561–569.
- [141] Ali Jadooei, O. Zaderykhin, and V. I. Shulgin. "Adaptive algorithm for continuous monitoring of blood pressure using a pulse transit time". In: *2013 IEEE 33rd International Scientific Conference Electronics and Nanotechnology, ELNANO 2013 - Conference Proceedings*. 2013, pp. 297–301.
- [142] Luc M Van Bortel, Stephane Laurent, Pierre Boutouyrie, Phil Chowienczyk, J K Cruickshank, Tine De Backer, Jan Filipovsky, Sofie Huybrechts, Francesco U S Mattace-Raso, and Athanase D Protogerou. "Expert consensus document on the measurement of aortic stiffness in daily practice using carotid-femoral pulse wave velocity". In: *Journal of hypertension* 30.3 (2012), pp. 445–448.
- [143] Chang Sei Kim, Andrew M. Carek, Omer T. Inan, Ramakrishna Mukkamala, and Jin Oh Hahn. "Ballistocardiogram-Based Approach to Cuffless Blood Pressure Monitoring: Proof of Concept and Potential Challenges". In: *IEEE Transactions on Biomedical Engineering* 65.11 (2018), pp. 2384–2391.
- [144] C Douniama, C U Sauter, and R Couronne. "Blood pressure tracking capabilities of pulse transit times in different arterial segments: a clinical evaluation". In: *2009 36th Annual Computers in Cardiology Conference (CinC)*. IEEE, 2009, pp. 201–204.
- [145] In Cheol Jeong and Joseph Finkelstein. "Introducing Contactless Blood Pressure Assessment Using a High Speed Video Camera". In: *Journal of Medical Systems* 40.4 (2016), pp. 1–10.
- [146] Venu Gopal Ganti, Andrew Carek, Brandi Nicole Nevius, James Heller, Mozziyar Etemadi, and Omer Inan. "Wearable Cuff-less Blood Pressure Estimation at Home via Pulse Transit Time". In: *IEEE Journal of Biomedical and Health Informatics* (2020).
- [147] Paul A. Obrist, Kathleen C. Light, James A. McCubbin, J. Stanford Hutcheson, and J. Lee Hoffer. "Pulse transit time: Relationship to blood pressure". In: *Behavior Research Methods & Instrumentation* 10.5 (1978), pp. 623–626.
- [148] David B. Newlin and Robert W. Levenson. "Pre-ejection Period: Measuring Beta-adrenergic Influences Upon the Heart". In: *Psychophysiology* 16.6 (1979), pp. 546–552.
- [149] Mico Yee Man Wong, Emma Pickwell-MacPherson, Yuan Ting Zhang, and Jack C Y Cheng. "The effects of pre-ejection period on post-exercise systolic blood pressure estimation using the pulse arrival time technique". In: *European journal of applied physiology* 111.1 (2011), pp. 135–144.
- [150] G. Zhang, M. Gao, D. Xu, N. B. Olivier, and R. Mukkamala. "Pulse arrival time is not an adequate surrogate for pulse transit time as a marker of blood pressure". In: *Journal of Applied Physiology* 117.10 (2011), p. 1212.
- [151] R. A. Payne, C. N. Symeonides, D. J. Webb, and S. R.J. Maxwell. "Pulse transit time measured from the ECG: An unreliable marker of beat-to-beat blood pressure". In: *Journal of Applied Physiology* 100.1 (2006), pp. 136–141.
- [152] Joel Balmer, Christopher Pretty, Shaun Davidson, Thomas Desaive, Shun Kamoi, Antoine Pironet, Philippe Morimont, Nathalie Janssen, Bernard Lambermont, and Geoffrey M Shaw. "Pre-ejection period, the reason why the electrocardiogram Q-wave is an unreliable indicator of pulse wave initialization". In: *Physiological measurement* 39.9 (2018), p. 95005.
- [153] J Muehlsteff, X A Aubert, and G Morren. "Continuous cuff-less blood pressure monitoring based on the pulse arrival time approach: The impact of posture". In: *2008 30th annual international conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2008, pp. 1691–1694.
- [154] G Cybulski, E Michalak, E Koźluk, A Piatkowska, and W Niewiadomski. "Stroke volume and systolic time intervals: beat-to-beat comparison between echocardiography and ambulatory impedance cardiography in supine and tilted positions". In: *Medical and Biological Engineering and Computing* 42.5 (2004), pp. 707–711.

- [155] Seyedmohsen Dehghanojamahalleh and Mehmet Kaya. "Sex-Related Differences in Photoplethysmography Signals Measured From Finger and Toe". In: *IEEE Journal of Translational Engineering in Health and Medicine* 7 (2019), pp. 1–7.
- [156] Ramakrishna Mukkamala and Jin Oh Hahn. "Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Predictions on Maximum Calibration Period and Acceptable Error Limits". In: *IEEE Transactions on Biomedical Engineering* 65.6 (2018), pp. 1410–1420.
- [157] C C Y Poon and Y T Zhang. "Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time". In: *2005 IEEE engineering in medicine and biology 27th annual conference*. IEEE, 2006, pp. 5877–5880.
- [158] A. S. Tijsseling and A. Anderson. "A. Isebre Moens and D.J. Korteweg: On the speed of propagation of waves in elastic tubes". In: *BHR Group - 11th International Conferences on Pressure Surges*. 2012, pp. 227–245.
- [159] D J Hughes, Charles F Babbs, L A Geddes, and J D Bourland. "Measurements of Young's modulus of elasticity of the canine aorta with ultrasound". In: *Ultrasonic Imaging* 1.4 (1979), pp. 356–367.
- [160] Heiko Gesche, Detlef Grosskurth, Gert Küchler, and Andreas Patzak. "Continuous blood pressure measurement by using the pulse transit time: comparison to a cuff-based method". In: *European journal of applied physiology* 112.1 (2012), pp. 309–315.
- [161] Parry Fung, Guy Dumont, Craig Ries, Chris Mott, and Mark Ansermino. "Continuous noninvasive blood pressure measurement by pulse transit time". In: *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*. Vol. 26 I. 2004, pp. 738–741.
- [162] Mico Yee Man Wong, Carmen Chung Yan Poon, and Yuan Ting Zhang. "An evaluation of the cuffless blood pressure estimation based on pulse transit time technique: A half year study on normotensive subjects". In: *Cardiovascular Engineering* 9.1 (2009), pp. 32–38.
- [163] Mohammad Yavarimanesh, Robert C Block, Keerthana Natarajan, Lalit K Mestha, Omer Inan, Jin-Oh Hahn, and Ramakrishna Mukkamala. "Assessment of Calibration Models for Cuff-Less Blood Pressure Measurement After One Year of Aging". In: *IEEE Transactions on Biomedical Engineering* (2021).
- [164] G. C. Jernstedt and Jeffrey P. Newcomer. "Blood pressure and pulse wave velocity measurement for operant conditioning of autonomic responding". In: *Behavior Research Methods & Instrumentation* 6.4 (1974), pp. 393–397.
- [165] Grzegorz Bilo, Eamon Dolan, Eoin O'Brien, Rita Facchetti, Davide Soranna, Antonella Zambon, Giuseppe Mancina, and Gianfranco Parati. "The impact of systolic and diastolic blood pressure variability on mortality is age dependent: Data from the Dublin Outcome Study". In: *European journal of preventive cardiology* 27.4 (2020), pp. 355–364.
- [166] Robert C Block, Mohammad Yavarimanesh, Keerthana Natarajan, Andrew Carek, Azin Mousavi, Anand Chandrasekhar, Chang-Sei Kim, Junxi Zhu, Giovanni Schifitto, and Lalit K Mestha. "Conventional pulse transit times as markers of blood pressure changes in humans". In: *Scientific Reports* 10.1 (2020), pp. 1–9.
- [167] Qiang Zhang, Xianxiang Chen, Zhen Fang, Yongjiao Xue, Qingyuan Zhan, Ting Yang, and Shanhong Xia. "Cuff-less blood pressure measurement using pulse arrival time and a Kalman filter". In: *Journal of Micromechanics and Microengineering* 27.2 (2017), p. 24002.
- [168] Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li Wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. "MIMIC-III, a freely accessible critical care database". In: *Scientific Data* 3 (2016), p. 160035.
- [169] Greg Welch and Gary Bishop. "An Introduction to the Kalman Filter". In: *In Practice* 7.1 (2006), pp. 1–16.

- [170] Joonnyong Lee, Seungman Yang, Saram Lee, and Hee Chan Kim. "Analysis of pulse arrival time as an indicator of blood pressure in a large surgical biosignal database: recommendations for developing ubiquitous blood pressure monitoring methods". In: *Journal of clinical medicine* 8.11 (2019), p. 1773.
- [171] Satu Rajala, Harri Lindholm, and Tapio Taipalus. "Comparison of photoplethysmogram measured from wrist and finger and the effect of measurement location on pulse arrival time". In: *Physiological Measurement* 39.7 (2018).
- [172] Ya Li Zheng, Bryan P. Yan, Yuan Ting Zhang, and Carmen C.Y. Poon. "An armband wearable device for overnight and cuff-less blood pressure measurement". In: *IEEE Transactions on Biomedical Engineering* 61.7 (2014), pp. 2179–2186.
- [173] Richard M Wiard, Omer T Inan, Brian Argyres, Mozziyar Etemadi, Gregory T.A. Kovacs, and Laurent Giovangrandi. "Automatic detection of motion artifacts in the ballistocardiogram measured on a modified bathroom scale". In: *Medical and Biological Engineering and Computing* 49.2 (2011), pp. 213–220.
- [174] Manuja Sharma, Karinne Barbosa, Victor Ho, Devon Griggs, Tadesse Ghirmai, Sandeep Krishnan, Tzung Hsiai, Jung-Chih Chiao, and Hung Cao. "Cuff-Less and Continuous Blood Pressure Monitoring: A Methodological Review". In: *Technologies* 5.2 (2017), p. 21.
- [175] L. Peter, N. Noury, and M. Cerny. "A review of methods for non-invasive and continuous blood pressure monitoring: Pulse transit time method is promising?" In: *Irbm* 35.5 (2014), pp. 271–282.
- [176] G Sharwood-Smith, J Bruce, and G Drummond. "Assessment of pulse transit time to indicate cardiovascular changes during obstetric spinal anaesthesia". In: *British Journal of Anaesthesia* 96.1 (2006), pp. 100–105.
- [177] Young Zoon Yoon, Jae Min Kang, Yongjoo Kwon, Sangyun Park, Seungwoo Noh, Younho Kim, Jongae Park, and Sung Woo Hwang. "Cuff-Less Blood Pressure Estimation Using Pulse Waveform Analysis and Pulse Arrival Time". In: *IEEE Journal of Biomedical and Health Informatics* 22.4 (2018), pp. 1068–1074.
- [178] Yongbo Liang, Derek Abbott, Newton Howard, Kenneth Lim, Rabab Ward, and Mohamed Elgendi. "How Effective Is Pulse Arrival Time for Evaluating Blood Pressure? Challenges and Recommendations from a Study Using the MIMIC Database". In: *Journal of Clinical Medicine* 8.3 (2019), p. 337.
- [179] Jong Yong Abdiel Foo. "Use of independent component analysis to reduce motion artifact in pulse transit time measurement". In: *IEEE Signal Processing Letters* 15 (2008), pp. 124–126.
- [180] Dilpreet Buxi, Jean Michel Redouté, and Mehmet Rasit Yuce. "A survey on signals and systems in ambulatory blood pressure monitoring using pulse transit time". In: *Physiological Measurement* 36.3 (2015), R1–R26.
- [181] Anand Chandrasekhar, Mohammad Yavarimanesh, Keerthana Natarajan, Jin-Oh Hahn, and Ramakrishna Mukkamala. "PPG sensor contact pressure should be taken into account for cuff-less blood pressure measurement". In: *IEEE Transactions on Biomedical Engineering* 67.11 (2020), pp. 3134–3140.
- [182] Braiam Escobar-Restrepo, Robinson Torres-Villa, Panayiotis A. Kyriacou, Dingchang Zheng, Fei Chen, Ramakrishna Mukkamala, and Fan Pan. "Evaluation of the linear relationship between pulse arrival time and blood pressure in ICU patients: Potential and limitations". In: *Frontiers in Physiology* 9 (2018).
- [183] Stephanie L-O Martin, Andrew M Carek, Chang-Sei Kim, Hazar Ashouri, Omer T Inan, Jin-Oh Hahn, and Ramakrishna Mukkamala. "Weighing scale-based pulse transit time is a superior marker of blood pressure than conventional pulse arrival time". In: *Scientific reports* 6.1 (2016), pp. 1–8.

- [184] Guanqun Zhang, Mingwu Gao, and Ramakrishna Mukkamala. "Robust, beat-to-beat estimation of the true pulse transit time from central and peripheral blood pressure or flow waveforms using an arterial tube-load model". In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* August 2011 (2011), pp. 4291–4294.
- [185] Gloria Martínez, Newton Howard, Derek Abbott, Kenneth Lim, Rabab Ward, and Mohamed Elgendi. "Can Photoplethysmography Replace Arterial Blood Pressure in the Assessment of Blood Pressure?" In: *Journal of Clinical Medicine* 7.10 (2018), p. 316.
- [186] Qunfeng Tang, Zhencheng Chen, Rabab Ward, Carlo Menon, and Mohamed Elgendi. "Subject-Based Model for Reconstructing Arterial Blood Pressure from Photoplethysmogram". In: *Bioengineering* 9.8 (2022), p. 402.
- [187] Mustafa Radha, Koen De Groot, Nikita Rajani, Cybele C.P. Wong, Nadjia Kobold, Valentina Vos, Pedro Fonseca, Nikolaos Mastellos, Petra A. Wark, Nathalie Velthoven, Reinder Haakma, and Ronald M. Aarts. "Estimating blood pressure trends and the nocturnal dip from photoplethysmography". In: *Physiological Measurement* 40.2 (2019), p. 025006. arXiv: 1805.09121.
- [188] Martin C Baruch, Darren E R Warburton, Shannon S D Bredin, Anita Cote, David W Gerdt, and Charles M Adkins. "Pulse decomposition analysis of the digital arterial pulse during hemorrhage simulation". In: *Nonlinear biomedical physics* 5.1 (2011), pp. 1–15.
- [189] Peter H. Charlton, Patrick Celka, Bushra Farukh, Phil Chowienzyk, and Jordi Alastruey. "Assessing mental stress from the photoplethysmogram: A numerical study". In: *Physiological Measurement* 39.5 (May 2018).
- [190] Keerthana Natarajan, Robert C Block, Mohammad Yavarimanesh, Anand Chandrasekhar, Lalit K Mestha, Omer Inan, Jin-Oh Hahn, and Ramakrishna Mukkamala. "Photoplethysmography Fast Upstroke Time Intervals Can Be Useful Features for Cuff-Less Measurement of Blood Pressure Changes in Humans". In: *IEEE Transactions on Biomedical Engineering* (2021).
- [191] Karsten Bartels, Stephen A. Esper, and Robert H. Thiele. "Blood pressure monitoring for the anesthesiologist: A practical review". In: *Anesthesia and Analgesia* 122.6 (2016), pp. 1866–1879.
- [192] S. C. Millasseau, R. P. Kelly, J. M. Ritter, and P. J. Chowienzyk. "Determination of age-related increases in large artery stiffness by digital pulse contour analysis". In: *Clinical Science* 103.4 (2002), pp. 371–377.
- [193] Uldis Rubins, A. Grabovskis, J. Grube, and I. Kukulis. "Photoplethysmography analysis of artery properties in patients with cardiovascular diseases". In: *IFMBE Proceedings*. Vol. 20 IFMBE. Springer, 2008, pp. 319–322.
- [194] Emma Von Wowern, Gerd Östling, Peter M. Nilsson, and Per Olofsson. "Digital photoplethysmography for assessment of arterial stiffness: Repeatability and comparison with applanation tonometry". In: *PLoS ONE* 10.8 (2015), e0135659.
- [195] R Couceiro, P Carvalho, R P Paiva, J Henriques, and J Muehlsteff. "Detection of motion artifact patterns in photoplethysmographic signals based on time and period domain analysis". In: *Physiological measurement* 35.12 (2014), p. 2369.
- [196] Mohamed Elgendi, Yongbo Liang, and Rabab Ward. "Toward Generating More Diagnostic Features from Photoplethysmogram Waveforms". In: *Diseases* 6.1 (2018), p. 20.
- [197] Williebosseau Murray and Patrick Anthony Foster. "The peripheral pulse wave: Information overlooked". In: *Journal of Clinical Monitoring* 12.5 (1996), pp. 365–377.
- [198] J. P. Murgo, N. Westerhof, J. P. Giolma, and S. A. Altobelli. "Aortic input impedance in normal man: Relationship to pressure wave forms". In: *Circulation* 62.1 (1980), pp. 105–116.
- [199] Kelser de Souza Kock and Jefferson Luiz Brum Marques. "Relationship between the Photoplethysmographic Parameters and Derivation from the Windkessel Model in Hemodynamics in Critically Ill Patients". In: *SM Journal of Cardiology and Cardiovascular Diseases* 4.1 (2018), pp. 1–6.

- [200] Xiao Fei Teng and Yuan Ting Zhang. "Theoretical study on the effect of sensor contact force on pulse transit time". In: *IEEE Transactions on Biomedical Engineering* 54.8 (2007), pp. 1490–1498.
- [201] Aymen A. Awad, Ala S. Haddadin, Hossam Tantawy, Tarek M. Badr, Robert G. Stout, David G. Silverman, and Kirk H. Shelley. "The relationship between the photoplethysmographic waveform and systemic vascular resistance". In: *Journal of Clinical Monitoring and Computing* 21.6 (2007), pp. 365–372.
- [202] Hangsik Shin and Se Dong Min. "Feasibility study for the non-invasive blood pressure estimation based on ppg morphology: Normotensive subject study". In: *BioMedical Engineering Online* 16.1 (2017).
- [203] Elina R.J. Seitsonen, I. K.J. Korhonen, M. J. Van Gils, M. Huiku, J. M.P. Lötjönen, K. T. Korttila, and A. M. Yli-Hankala. "EEG spectral entropy, heart rate, photoplethysmography and motor responses to skin incision during sevoflurane anaesthesia". In: *Acta Anaesthesiologica Scandinavica* 49.3 (2005), pp. 284–292.
- [204] L. Wang, Emma Pickwell-MacPherson, Y. P. Liang, and Y. T. Zhang. "Noninvasive cardiac output estimation using a novel photoplethysmogram index". In: *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*. IEEE, 2009, pp. 1746–1749.
- [205] Uldis Rubins. "Finger and ear photoplethysmogram waveform analysis by fitting with Gaussians". In: *Medical & biological engineering & computing* 46.12 (2008), pp. 1271–1276.
- [206] R. Couceiro, P. Carvalho, R. P. Paiva, J. Henriques, M. Antunes, I. Quintal, and J. Muehlsteff. "Multi-Gaussian fitting for the assessment of left ventricular ejection time from the Photoplethysmogram". In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. IEEE, 2012, pp. 3951–3954.
- [207] Paul M. Middleton, Collin H.H. Tang, Gregory S.H. Chan, Sarah Bishop, Andrey V. Savkin, and Nigel H. Lovell. "Peripheral photoplethysmography variability analysis of sepsis patients". In: *Medical and Biological Engineering and Computing* 49.3 (2011), pp. 337–347.
- [208] Meir Nitzan, Hedzer de Boer, Sergei Turivnenko, Anatoly Babchenko, and Dan Sapoznikov. "Power spectrum analysis of spontaneous fluctuations in the photoplethysmographic signal". In: *Journal of Basic and Clinical Physiology and Pharmacology* 5.3-4 (1994), pp. 269–276.
- [209] Tananant Boonya-Ananta, Andres J Rodriguez, Ajmal Ajmal, Vinh Nguyen Du Le, Anders K Hansen, Joshua D Hutcheson, and Jessica C Ramella-Roman. "Synthetic photoplethysmography (PPG) of the radial artery through parallelized Monte Carlo and its correlation to body mass index (BMI)". In: *Scientific reports* 11.1 (2021), pp. 1–11.
- [210] Fen Miao, Nan Fu, Yuan Ting Zhang, Xiao Rong Ding, Xi Hong, Qingyun He, and Ye Li. "A novel continuous blood pressure estimation approach based on data mining techniques". In: *IEEE Journal of Biomedical and Health Informatics* 21.6 (2017), pp. 1730–1740.
- [211] Juan C. Ruiz-Rodríguez, Adolf Ruiz-Sanmartín, Vicent Ribas, Jesús Caballero, Alejandra García-Roche, Jordi Riera, Xavier Nuvials, Miriam De Nadal, Oriol De Sola-Morales, Joaquim Serra, and Jordi Rello. "Innovative continuous non-invasive cuffless blood pressure monitoring based on photoplethysmography technology". In: *Intensive Care Medicine*. Vol. 39. 9. 2013, pp. 1618–1625.
- [212] Navid Hasanzadeh, Mohammad Mahdi Ahmadi, and Hoda Mohammadzade. "Blood Pressure Estimation Using Photoplethysmogram Signal and Its Morphological Features". In: *IEEE Sensors Journal* 20.8 (2019), pp. 4300–4310.
- [213] Gašper Slapničar, Nejc Mlakar, and Mitja Luštrek. "Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network". In: *Sensors* 19.15 (2019), p. 3420.

- [214] Oded Schlesinger, Nitai Vigderhouse, Yair Moshe, and Danny Eytan. "Estimation and Tracking of Blood Pressure Using Routinely Acquired Photoplethysmographic Signals and Deep Neural Networks". In: *Critical Care Explorations* 2.4 (Apr. 2020), e0095.
- [215] S. Sun, R. Bezemer, X. Long, J. Muehlsteff, and R. M. Aarts. "Systolic blood pressure estimation using PPG and ECG during physical exercise". In: *Physiological Measurement* 37.12 (2016), pp. 2154–2169.
- [216] Xiaoman Xing and Mingshan Sun. "Optical blood pressure estimation with photoplethysmography and FFT-based neural networks". In: *Biomedical Optics Express* 7.8 (2016), p. 3007.
- [217] Rebecca Mieloszyk, Hope Twede, Jonathan Lester, Jeremiah Wander, Sumit Basu, Gabe Cohn, Greg Smith, Dan Morris, Sidhant Gupta, and Desney Tan. "A comparison of wearable tonometry, photoplethysmography, and electrocardiography for cuffless measurement of blood pressure in an ambulatory setting". In: *IEEE Journal of Biomedical and Health Informatics* 26.7 (2022), pp. 2864–2875.
- [218] Manish Hosanee, Gabriel Chan, Kaylie Welykholowa, Rachel Cooper, Panayiotis A Kyriacou, Dingchang Zheng, John Allen, Derek Abbott, Carlo Menon, Nigel H Lovell, Newton Howard, Wee-Shian Chan, Kenneth Lim, Richard Fletcher, Rabab Ward, and Mohamed Elgendi. "Cuffless Single-Site Photoplethysmography for Blood Pressure Monitoring". In: *Journal of Clinical Medicine* 9.3 (2020), p. 723.
- [219] Mohamed Elgendi, Richard Fletcher, Yongbo Liang, Newton Howard, Nigel H. Lovell, Derek Abbott, Kenneth Lim, and Rabab Ward. "The use of photoplethysmography for assessing hypertension". In: *npj Digital Medicine* 2.1 (2019), p. 60.
- [220] Hong Luo, Deye Yang, Andrew Barszczyk, Naresh Vempala, Jing Wei, Si Jia Wu, Paul Pu Zheng, Genyue Fu, Kang Lee, and Zhong-Ping Feng. "Smartphone-based blood pressure measurement using transdermal optical imaging technology". In: *Circulation: Cardiovascular Imaging* 12.8 (2019), e008857.
- [221] Zeng Ding Liu, Bin Zhou, Ye Li, Min Tang, and Fen Miao. "Continuous Blood Pressure Estimation From Electrocardiogram and Photoplethysmogram During Arrhythmias". In: *Frontiers in Physiology* 11 (Sept. 2020).
- [222] Anand Chandrasekhar, Mohammad Yavarimanesh, Keerthana Natarajan, Jin-Oh Hahn, and Ramakrishna Mukkamala. "PPG sensor contact pressure should be taken into account for cuff-less blood pressure measurement". In: *IEEE Transactions on Biomedical Engineering* 67.11 (2020), pp. 3134–3140.
- [223] Mohamed Elgendi. "C, E and D Waves Detection in the Acceleration Photoplethysmogram". In: *Computer Methods and Programs in Biomedicine* 117.2 (2014), pp. 125–136.
- [224] Sanjeev Nara, Manvinder Kaur, and Kundan Lal Verma. "Novel Notch Detection Algorithm for Detection of Dicrotic Notch in PPG Signals". In: *International Journal of Computer Applications* 86.17 (2014), pp. 36–39.
- [225] Mohamed Elgendi, Ian Norton, Matt Brearley, Derek Abbott, and Dale Schuurmans. "Detection of a and b waves in the acceleration photoplethysmogram". In: *BioMedical Engineering Online* 13.1 (2014), pp. 125–136.
- [226] Xiaochuan He, Rafik A. Goubran, and Xiaoping P. Liu. "Secondary peak detection of PPG signal for continuous cuffless arterial blood pressure measurement". In: *IEEE Transactions on Instrumentation and Measurement* 63.6 (2014), pp. 1431–1439.
- [227] Viviane Timmermann, Lars A Dejgaard, Kristina H Haugaa, Andrew G Edwards, Joakim Sundnes, Andrew D McCulloch, and Samuel T Wall. "An integrative appraisal of mechano-electric feedback mechanisms in the heart". In: *Progress in biophysics and molecular biology* 130 (2017), pp. 404–417.
- [228] T Alexander Quinn and Peter Kohl. "Cardiac mechano-electric coupling: acute effects of mechanical stimulation on heart rate and rhythm". In: *Physiological reviews* 101.1 (2021), pp. 37–92.

- [229] Alistair Reed, Peter Kohl, and Rémi Peyronnet. “Molecular candidates for cardiac stretch-activated ion channels”. In: *Global Cardiology Science and Practice* 2014.2 (2014), p. 19.
- [230] T. Alexander Quinn. “The importance of non-uniformities in mechano-electric coupling for ventricular arrhythmias”. In: *Journal of Interventional Cardiac Electrophysiology* 39.1 (Jan. 2014), pp. 25–35.
- [231] A S Villecco, D De Aloysio, E De Liberali, G Ferrari, and M Mauloni. “High blood pressure and ‘ischaemic’ ECG patterns in climacteric women”. In: *Maturitas* 7.2 (1985), pp. 89–97.
- [232] Gendy Monroy Estrada, Luis Enrique Mendoza, and Valentin Molina. “Relationship of blood pressure with the electrical signal of the heart using signal processing”. In: *Tecciencia* 9.17 (2014), pp. 9–14.
- [233] Seyedeh Somayyeh Mousavi, Mohammad Hemmati, Mostafa Charmi, Maryam Moghadam, Mohammad Firouzmand, and Yadollah Ghorbani. “Cuff-less blood pressure estimation using only the ECG signal in frequency domain”. In: *2018 8th International Conference on Computer and Knowledge Engineering (ICCCKE)*. IEEE, 2018, pp. 147–152.
- [234] Monika Simjanoska, Martin Gjoreski, Matjaž Gams, and Ana Madevska Bogdanova. “Non-invasive blood pressure estimation from ECG using machine learning techniques”. In: *Sensors (Switzerland)* 18.4 (2018), p. 1160.
- [235] Monika Simjanoska, Gregor Papa, Barbara Koroušić Seljak, and Tome Eftimov. “Comparing different settings of parameters needed for pre-processing of ECG signals used for blood pressure classification”. In: *BIOSIGNALS 2019 - 12th International Conference on Bio-Inspired Systems and Signal Processing, Proceedings*. SCITEPRESS - Science and Technology Publications, 2019, pp. 62–72.
- [236] Sen Yang, Wan Suhaimizan Wan Zaki, Stephen P Morgan, Siu-Yeung Cho, Ricardo Correia, and Yaping Zhang. “Blood pressure estimation with complexity features from electrocardiogram and photoplethysmogram signals”. In: *Optical and Quantum Electronics* 52.3 (2020), p. 135.
- [237] Roberto Fabio Leonarduzzi, Gastón Schlotthauer, and María Eugenia Torres. “Wavelet leader based multifractal analysis of heart rate variability during myocardial ischaemia”. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 110–113.
- [238] Hongqiang Li, Xiuli Feng, Lu Cao, Enbang Li, Huan Liang, and Xuelong Chen. “A new ECG signal classification based on WPD and ApEn feature extraction”. In: *Circuits, Systems, and Signal Processing* 35.1 (2016), pp. 339–352.
- [239] Philippe Gosse, Emilie Jan, Paul Coulon, Antoine Cremer, Georgios Papaioannou, and Sunthareth Yeim. “ECG detection of left ventricular hypertrophy: the simpler, the better?” In: *Journal of hypertension* 30.5 (2012), pp. 990–996.
- [240] Bloomberg. *Apple Targets Watch Blood-Pressure Tool for 2024 After Snags*. 2022. URL: <https://www.bloomberg.com/news/articles/2022-04-12/apple-targets-watch-blood-pressure-feature-for-2024-after-snags> (visited on 08/29/2022).
- [241] Josep Solà, A Vybornova, Sibylle Fallet, E Olivero, B De Marco, Olivier Grossenbacher, N Ignjatovic, B Ignjatovic, M Favre-Bulle, and N Levinson. “Are cuffless devices challenged enough? Design of a validation protocol for ambulatory blood pressure monitors at the wrist: the case of the Aktiia Bracelet”. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 4437–4440.

- [242] George S. Stergiou, Bruce Alpert, Stephan Mieke, Roland Asmar, Neil Atkins, Siegfried Eckert, Gerhard Frick, Bruce Friedman, Thomas Graßl, Tsutomu Ichikawa, John P. Ioannidis, Peter Lacy, Richard McManus, Alan Murray, Martin Myers, Paolo Palatini, Gianfranco Parati, David Quinn, Josh Sarkis, Andrew Shennan, Takashi Usuda, Jiguang Wang, Colin O. Wu, and Eoin O'Brien. "A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) Collaboration Statement". In: *Hypertension* 71.3 (2018), pp. 368–374.
- [243] Mariska Weenk, Harry van Goor, Bas Frietman, Lucien JLPG Engelen, Cornelis J H M van Laarhoven, Jan Smit, Sebastian J H Bredie, and Tom H van de Belt. "Continuous monitoring of vital signs using wearable devices on the general ward: pilot study". In: *JMIR mHealth and uHealth* 5.7 (2017), e7208.
- [244] Yongbo Liang, Zhencheng Chen, Guiyong Liu, and Mohamed Elgendi. "A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China". In: *Scientific Data* 5.1 (Feb. 2018), pp. 1–7.
- [245] Yu M. Chi, Stephen R. Deiss, and Gert Cauwenberghs. "Non-contact low power EEG/ECG electrode for high density wearable biopotential sensor networks". In: *Proceedings - 2009 6th International Workshop on Wearable and Implantable Body Sensor Networks, BSN 2009*. IEEE, 2009, pp. 246–250.
- [246] K S Chong, E Zahedi, K B Gan, and M A Mohd Ali. "Evaluation of the effect of step size on delta modulation for photoplethysmogram compression". In: *Procedia Technology* 11 (2013), pp. 815–822.
- [247] Evan Richards and Christopher V. Maani. "Phenylephrine". In: *StatPearls [Internet]*. StatPearls Publishing, 2019.
- [248] Philips. *IntelliVue MX800*. URL: <https://www.philips.co.uk/healthcare/product/HC865240/intellivue-mx800-bedside-patient-monitor> (visited on 04/29/2020).
- [249] Stowood. *Stowood Visi Black Series Polygraphy*. (Visited on 04/29/2020).
- [250] Richard Harrigan, Theodore Chan, and William Brady. "The Diagnosis: Atrial Bigeminy". In: *Emergency Medicine News* 25.9 (2003).
- [251] Jean Louis Vincent. "Understanding cardiac output". In: *Critical Care* 12.4 (2008).
- [252] Jean Michel Mallion, Régis De Gaudemaris, Jean Philippe Baguet, Leila Azzouzi, Jean Louis Quesada, Christine Sauzeau, Jean Philippe Siché, Frédéric Tremel, and Sophie Boutelant. "Acceptability and tolerance of ambulatory blood pressure measurement in the hypertensive patient". In: *Blood Pressure Monitoring* 1.3 (1996), pp. 197–203.
- [253] Shaun Davidson, Mauricio Villarro, Eoin Finnegan, Mirae Harford, Joao Jorge, and Lionel Tarassenko. "Vital-Sign Synchrony as a Marker for Patient Circadian Rhythms in an Intensive Care Unit". In: *2020 Computing in Cardiology*. IEEE, 2020, pp. 1–4.
- [254] H. Lollgen and H. Drexler. "Use of inotropes in the critical care setting". In: *Critical Care Medicine* 18.1 SUPPL. (1990), p. 395.
- [255] Gerald K. McEvoy. "AHFS Drug Information". In: *Oncology Issues* 9.5 (1994), pp. 12–13.
- [256] Sultan M. Babar. "SIADH Associated With Ciprofloxacin". In: *Annals of Pharmacotherapy* 47.10 (2013), pp. 1359–1363.
- [257] Paul F White. "Propofol: pharmacokinetics and pharmacodynamics". In: *Semin Anesth* 7 (1988), pp. 4–20.

- [258] Lelia Duley, Barbara Farrell, and James P. Neilson. "Magnesium sulphate: A review of clinical pharmacology applied to obstetrics". In: *BJOG: An International Journal of Obstetrics and Gynaecology* 106.2 (1999), p. 180.
- [259] Carl G. Regårdh, Karl O. Borg, Rustan Johansson, Gillis Johnsson, and Lars Palmer. "Pharmacokinetic studies on the selective β_1 -receptor antagonist metoprolol in man". In: *Journal of Pharmacokinetics and Biopharmaceutics* 2.4 (1974), pp. 347–364.
- [260] K. O'Malley, J. L. Segal, Z. H. Israili, M. Boles, J. L. McNay, and P. G. Dayton. "Duration of hydralazine action in hypertension". In: *Clinical Pharmacology and Therapeutics* 18.5 PART 1 (1975), pp. 581–586.
- [261] Wei Zong, G. B. Moody, and D. Jiang. "A robust open-source algorithm to detect onset and duration of QRS complexes". In: *Computers in Cardiology*. Vol. 30. IEEE, 2003, pp. 737–740.
- [262] ANSI/AAMI-EC57:1998. "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms". In: *Association for the Advancement of Medical Instrumentation* 1998 (1999), p. 36.
- [263] Mauricio Villarroel, João Jorge, David Meredith, Sheera Sutherland, Chris Pugh, and Lionel Tarassenko. "Non-contact vital-sign monitoring of patients undergoing haemodialysis treatment". In: *Scientific reports* 10.1 (2020), pp. 1–21.
- [264] Wei Zong, T. Heldt, G. B. Moody, and R. G. Mark. "An open-source algorithm to detect onset of arterial blood pressure pulses". In: *Computers in Cardiology*. Vol. 30. IEEE, 2003, pp. 259–262.
- [265] Q. Li, R. G. Mark, and G. D. Clifford. "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter". In: *Physiological Measurement* 29.1 (Jan. 2008), pp. 15–32.
- [266] George B Moody and Roger G Mark. "The MIT-BIH arrhythmia database on CD-ROM and software for use with it". In: *[1990] Proceedings Computers in Cardiology*. IEEE, 1990, pp. 185–188.
- [267] Vrudhula K Murthy, Thomas M Grove, George A Harvey, and L Julian Haywood. "Clinical usefulness of ECG frequency spectrum analysis". In: *The Second Annual Symposium on Computer Application in Medical Care, 1978. Proceedings*. IEEE, 1978, pp. 610–612.
- [268] Q. Li and G. D. Clifford. "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals". In: *Physiological Measurement* 33.9 (2012), pp. 1491–1501.
- [269] Donald Berndt and James Clifford. "Using dynamic time warping to find patterns in time series". In: *Workshop on Knowledge Knowledge Discovery in Databases*. Vol. 398. 16. Seattle, WA, 1994, pp. 359–370. URL: <http://www.aaai.org/Papers/Workshops/1994/WS-94-03/WS94-03-031.pdf>.
- [270] Association for the Advancement of Medical Instrumentation. "ANSI/AAMI EC13: 2002 Cardiac monitors, heart rate meters, and alarms". In: *Arlington, Virginia, Association for the Advancement of Medical Instrumentation* (2002), p. 21.
- [271] Robin Genuer, Jean-Michel Poggi, and Christine Tuleau. "Random Forests: some methodological insights". In: *arXiv preprint arXiv:0811.3619* (2008).
- [272] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and regression trees*. Routledge, 2017.
- [273] Leo Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.
- [274] Ahlem Hajjem, François Bellavance, and Denis Larocque. "Mixed-effects random forest for clustered data". In: *Journal of Statistical Computation and Simulation* 84.6 (2014), pp. 1313–1328.
- [275] Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30 (2017).

- [276] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. “From local explanations to global understanding with explainable AI for trees”. In: *Nature machine intelligence* 2.1 (2020), pp. 56–67.
- [277] Karim Lakhal, Stephan Ehrmann, and Thierry Boulain. “Noninvasive BP Monitoring in the Critically Ill: Time to Abandon the Arterial Catheter?” In: *Chest* 153.4 (Apr. 2018), pp. 1023–1039.
- [278] D S G Pollock. *Smoothing with cubic splines*. 1993.
- [279] Peter Craven and Grace Wahba. “Smoothing noisy data with spline functions”. In: *Numerische mathematik* 31.4 (1978), pp. 377–403.
- [280] J. Muehlsteff, X. L. Aubert, and M. Schuett. “Cuffless estimation of systolic blood pressure for short effort bicycle tests: The prominent role of the pre-ejection period”. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*. Vol. 2006. Conf Proc IEEE Eng Med Biol Soc, 2006, pp. 5088–5092.
- [281] Mitra Jabalameli, Safoura Rouholamin, and Fatemeh Gourtanian. “A comparison of the effects of fentanyl and remifentanyl on nausea, vomiting, and pain after cesarean section”. In: *Iranian Journal of Medical Sciences* 36.3 (2011), pp. 183–187.
- [282] Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. “Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median”. In: *Journal of Experimental Social Psychology* 49.4 (2013), pp. 764–766.
- [283] L. Tarassenko, L. Mason, and N. Townsend. “Multi-sensor fusion for robust computation of breathing rate”. In: *Electronics Letters* 38.22 (2002), pp. 1314–1316.
- [284] Abraham Savitzky and Marcel J E Golay. “Smoothing and differentiation of data by simplified least squares procedures.” In: *Analytical chemistry* 36.8 (1964), pp. 1627–1639.
- [285] Sophocles J Orfanidis. *Introduction to signal processing*. Prentice-Hall, Inc., 1995.
- [286] Neal B Gallagher. “Savitzky-Golay Smoothing and Differentiation Filter”. In: *Eigenvector Research Incorporated* (2020).
- [287] Jianwen Luo, Kui Ying, Ping He, and Jing Bai. “Properties of Savitzky–Golay digital differentiators”. In: *Digital Signal Processing* 15.2 (2005), pp. 122–136.
- [288] Mohamad Forouzanfar, Fiona C Baker, Massimiliano de Zambotti, Corey McCall, Laurent Giovannardi, and Gregory T A Kovacs. “Toward a better noninvasive assessment of preejection period: A novel automatic algorithm for B-point detection and correction on thoracic impedance cardiogram.” eng. In: *Psychophysiology* 55.8 (Aug. 2018), e13072.
- [289] Shafa-at Ali Sheikh, Amit J Shah, Oleksiy Levantsevych, Majd Soudan, Mhmtjamil I Alkhalaf, Ali Bahrami Rad, Omer T Inan, and Gari D Clifford. “An open-source automated algorithm for removal of noisy beats for accurate impedance cardiogram analysis”. In: *Physiological Measurement* (2020).
- [290] Mohamad Forouzanfar, Fiona C Baker, Ian M Colrain, Aimée Goldstone, and Massimiliano de Zambotti. “Automatic analysis of pre-ejection period during sleep using impedance cardiogram.” eng. In: *Psychophysiology* 56.7 (July 2019), e13355.
- [291] Javier Rodríguez Árbol, Pandelis Perakakis, Alba Garrido, José Luis Mata, M. Carmen Fernández-Santaella, and Jaime Vila. “Mathematical detection of aortic valve opening (B point) in impedance cardiography: A comparison of three popular algorithms”. In: *Psychophysiology* 54.3 (Mar. 2017), pp. 350–357.
- [292] David L Lozano, Greg Norman, Dayan Knox, Beatrice L Wood, Bruce D Miller, Charles F Emery, and Gary G Berntson. “Where to B in dZ/dt”. In: *Psychophysiology* 44.1 (2007), pp. 113–119.
- [293] Michael Richter and Guido H.E. Gendolla. “The heart contracts to reward: Monetary incentives and preejection period”. In: *Psychophysiology* 46.3 (May 2009), pp. 451–457.

- [294] Michael Richter, Antonia Friedrich, and Guido H.E. Gendolla. "Task difficulty effects on cardiac activity". In: *Psychophysiology* 45.5 (Sept. 2008), pp. 869–875.
- [295] Gerard Cybulski, Gerard Młyńczak, Marcel Żyliński, Anna Strasz, Anna Gasiorowska, and Wiktor Niewiadomski. "The quality of automatic artifact identification in ambulatory impedance cardiography monitoring". In: *EMBECE & NBC 2017*. Springer, 2017, pp. 165–168.
- [296] Claude E Shannon. "Communication in the presence of noise". In: *Proceedings of the IRE* 37.1 (1949), pp. 10–21.
- [297] Qing Liu, Bryan P. Yan, Cheuk Man Yu, Yuan Ting Zhang, and Carmen C.Y. Poon. "Attenuation of systolic blood pressure and pulse transit time hysteresis during exercise and recovery in cardiovascular patients". In: *IEEE Transactions on Biomedical Engineering* 61.2 (2014), pp. 346–352.
- [298] Anthony V Incognito, Milena Samora, Andrew D Shepherd, Roberta A Cartafina, Gabriel M N Guimarães, Mauricio Daher, Lauro C Vianna, and Philip J Millar. "Sympathetic arterial baroreflex hysteresis in humans: different patterns during low-and high-pressure levels". In: *American Journal of Physiology-Heart and Circulatory Physiology* 319.4 (2020), H787–H792.
- [299] Yoav Benjamini and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the Royal statistical society: series B (Methodological)* 57.1 (1995), pp. 289–300.
- [300] Tianfeng Chai and Roland R Draxler. "Root mean square error (RMSE) or mean absolute error (MAE)". In: *Geoscientific Model Development Discussions* 7.1 (2014), pp. 1525–1534.
- [301] Xiaorong Ding, Bryan P. Yan, Walter Karlen, Yuan Ting Zhang, and Hon Ki Tsang. "Pulse transit time based respiratory rate estimation with singular spectrum analysis". In: *Medical and Biological Engineering and Computing* 58.2 (Feb. 2020), pp. 257–266.
- [302] James S. Dean II and Anil Kumar Reddy Reddivari. *Alpha 1 Receptor Agonists*. StatPearls Publishing, June 2019. URL: <https://www.ncbi.nlm.nih.gov/books/NBK551698/%20http://www.ncbi.nlm.nih.gov/pubmed/31869162>.
- [303] Tonous N Silfani, Chao-Yin Chen, Rahul D Patil, and Stephen E DiCarlo. "Vascular smooth muscle and exercise". In: *Sports Medicine, Training and Rehabilitation* 8.4 (Apr. 1998), pp. 301–320. URL: <https://doi.org/10.1080/15438629909512535>.
- [304] René van Lien, Nienke M Schutte, Jan H Meijer, and Eco J C de Geus. "Estimated preejection period (PEP) based on the detection of the R-wave and dZ/dt-min peaks does not adequately reflect the actual PEP across a wide range of laboratory and ambulatory conditions". In: *International Journal of Psychophysiology* 87.1 (2013), pp. 60–69.
- [305] Peter H Charlton, Drew A Birrenkott, Timothy Bonnici, Marco A F Pimentel, Alistair E W Johnson, Jordi Alastruey, Lionel Tarassenko, Peter J Watkinson, Richard Beale, and David A Clifton. "Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review". In: *IEEE reviews in biomedical engineering* 11 (2017), pp. 2–20.
- [306] V Gemignani, E Bianchini, F Faita, M Giannoni, E Pisanisi, E Picano, and T Bombardini. "Assessment of cardiologic systole and diastole duration in exercise stress tests with a transcutaneous accelerometer sensor". In: *2008 Computers in Cardiology*. IEEE, 2008, pp. 153–156.
- [307] J. M. Padilla, E. J. Berjano, J. Sáiz, L. Fácila, P. Díaz, and S. Mercé. "Assessment of relationships between blood pressure, pulse wave velocity and digital volume pulse". In: *Computers in Cardiology*. Vol. 33. IEEE, 2006, pp. 893–896.
- [308] Jae Mok Ahn. "New aging index using signal features of both photoplethysmograms and acceleration plethysmograms". In: *Healthcare Informatics Research* 23.1 (2017), pp. 53–59.
- [309] Paul S Addison. "Slope transit time (STT): A pulse transit time proxy requiring only a single signal fiducial point". In: *IEEE Transactions on Biomedical Engineering* 63.11 (2016), pp. 2441–2444.

- [310] Yongqiang Lyu, Xiaomin Luo, Jun Zhou, Chun Yu, Congcong Miao, Tong Wang, Yuanchun Shi, and Ken-ichi Kameyama. "Measuring photoplethysmogram-based stress-induced vascular response index to assess cognitive load and stress". In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 2015, pp. 857–866.
- [311] L Wang and Y T Zhang. "A novel photoplethysmogram index for total peripheral resistance after bicycle exercise". In: (2008).
- [312] K. Pilt, K. Meigas, R. Ferenets, K. Temitski, and M. Viigimaa. "Photoplethysmographic signal waveform index for detection of increased arterial stiffness". In: *Physiological Measurement* 35.10 (2014), pp. 2027–2036.
- [313] Xiaoman Xing, Zhimin Ma, Mingyou Zhang, Xi Gao, Ying Li, Mingxuan Song, and Wen-Fei Dong. "Robust blood pressure estimation from finger photoplethysmography using age-dependent linear models". In: *Physiological measurement* 41.2 (2020), p. 25007.
- [314] Alfredo Ortiz, Cristina Gherasim, Mario Manana, Carlos J Renedo, L Ignacio Eguiluz, and Ronnie J M Belmans. "Total harmonic distortion decomposition depending on distortion origin". In: *IEEE Transactions on Power delivery* 20.4 (2005), pp. 2651–2656.
- [315] Wan-Hua Lin, Xiangxin Li, Yuanheng Li, Guanglin Li, and Fei Chen. "Investigating the physiological mechanisms of the photoplethysmogram features for blood pressure estimation". In: *Physiological measurement* 41.4 (2020), p. 44003.
- [316] Ricardo Couceiro, P Carvalho, R P Paiva, J Henriques, I Quintal, M Antunes, J Muehlsteff, C Eickholt, C Brinkmeyer, and M Kelm. "Assessment of cardiovascular function from multi-Gaussian fitting of a finger photoplethysmogram". In: *Physiological measurement* 36.9 (2015), p. 1801.
- [317] Josep Solà and Ricard Delgado-Gonzalo. *The Handbook of Cuffless Blood Pressure Monitoring: A Practical Guide for Clinicians, Researchers, and Engineers*. Springer Nature, 2019.
- [318] Lu Wang, Lisheng Xu, Shuting Feng, Max Q.H. Meng, and Kuanquan Wang. "Multi-Gaussian fitting for pulse waveform using Weighted Least Squares and multi-criteria decision making method". In: *Computers in Biology and Medicine* 43.11 (2013), pp. 1661–1672.
- [319] Takumi Nagasawa, Kaito Iuchi, Ryo Takahashi, Mari Tsunomura, Raquel Pantojo de Souza, Keiko Ogawa-ochiai, Norimichi Tsumura, and George C. Cardoso. "Blood Pressure Estimation by Photoplethysmogram Decomposition into Hyperbolic Secant Waves". In: *Applied Sciences (Switzerland)* 12.4 (Feb. 2022), p. 1798.
- [320] PAUL Kligfield and PETER Okin. "Effect of ventricular function on left ventricular ejection time in aortic stenosis." In: *Heart* 42.4 (1979), pp. 438–441.
- [321] Hervé Abdi and Lynne J. Williams. "Principal component analysis". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 2.4 (2010), pp. 433–459.
- [322] Tomoyuki Higuchi. "Approach to an irregular time series on the basis of the fractal theory". In: *Physica D: Nonlinear Phenomena* 31.2 (1988), pp. 277–283.
- [323] Achmad Rizal and Sugondo Hadiyoso. "ECG signal classification using Hjorth Descriptor". In: *2015 International conference on automation, cognitive science, optics, micro electro-mechanical system, and information technology (ICACOMIT)*. IEEE, 2015, pp. 87–90.
- [324] Kenneth Brezinski. *ECG Beat Classification Using Higuchi Fractal Dimension and Heart Rate Variability*. Tech. rep. University of Manitoba, 2019.
- [325] Carsten F Dormann, Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R García Marquéz, Bernd Gruber, Bruno Lafourcade, and Pedro J Leitão. "Collinearity: a review of methods to deal with it and a simulation study evaluating their performance". In: *Ecography* 36.1 (2013), pp. 27–46.

- [326] Jong Hae Kim. "Multicollinearity and misleading statistical results". In: *Korean journal of anesthesiology* 72.6 (2019), p. 558.
- [327] Charles Spearman. "The proof and measurement of association between two things." In: (1961).
- [328] Krzysztof Narkiewicz, Catherine A Pesek, Masahiko Kato, Bradley G Phillips, Diane E Davison, and Virend K Somers. "Baroreflex control of sympathetic nerve activity and heart rate in obstructive sleep apnea". In: *Hypertension* 32.6 (1998), pp. 1039–1043.
- [329] Eoin Finnegan, Shaun Davidson, Mirae Harford, João Jorge, Peter Watkinson, Duncan Young, Lionel Tarassenko, and Mauricio Villarroel. "Pulse arrival time as a surrogate of blood pressure". In: *Scientific reports* 11.1 (2021), pp. 1–21.
- [330] Maxime Cannesson, Zhongping Jian, Guo Chen, Trung Q Vu, and Feras Hatib. "Effects of phenylephrine on cardiac output and venous return depend on the position of the heart on the Frank-Starling relationship". In: *Journal of applied physiology* 113.2 (2012), pp. 281–289.
- [331] Yongwei Zhu, Haihong Zhang, and Cuntai Guan. "On robust extraction of pulse transit time from multimodal pulsatile signals". In: *International Conference on Digital Signal Processing, DSP*. Vol. 2015-Sept. IEEE, 2015, pp. 403–406.
- [332] Frank C. Bennis, Carola van Pul, Jarno J.L. van den Bogaart, Peter Andriessen, Boris W. Kramer, and Tammo Delhaas. "Artifacts in pulse transit time measurements using standard patient monitoring equipment". In: *PLoS ONE* 14.6 (2019), e0218784.
- [333] Yu-Ting Lin, Yu-Lun Lo, Chen-Yun Lin, Martin G Frasch, and Hau-Tieng Wu. "Unexpected sawtooth artifact in beat-to-beat pulse transit time measured from patient monitor data". In: *PloS one* 14.9 (2019), e0221319.
- [334] Joel Balmer, Rachel Smith, Christopher G Pretty, Thomas Desaive, Geoff M Shaw, and J Geoffrey Chase. "Accurate end systole detection in dicrotic notch-less arterial pressure waveforms". In: *Journal of clinical monitoring and computing* (2020), pp. 1–10.
- [335] Germaine Cornelissen. "Cosinor-based rhythmometry". In: *Theoretical Biology and Medical Modelling* 11.1 (2014), pp. 1–24.
- [336] Marcelo E Katz, Fernando Margulis, Rubén Schiavelli, Pablo Arias, Geoffrey A Head, Diego A Golombek, and Ahmed Bahammam. "Disruption of transitional stages in 24-h blood pressure recording in renal transplant recipients". In: *Frontiers in neuroscience* (2012).
- [337] J M Madden, L. D. Browne, X. Li, P. M. Kearney, and A. P. Fitzgerald. "Morning surge in blood pressure using a random-effects multiple-component cosinor model". In: *Statistics in Medicine* 37.10 (2018), pp. 1682–1695.
- [338] Sky McKinley and Megan Levine. "Cubic spline interpolation". In: *College of the Redwoods* 45.1 (1998), pp. 1049–1060.
- [339] Jamie M Madden, Xia Li, Patricia M Kearney, Kate Tilling, and Anthony P Fitzgerald. "Exploring diurnal variation using piecewise linear splines: an example using blood pressure". In: *Emerging themes in epidemiology* 14.1 (2017), pp. 1–11.
- [340] Geoffrey A Head, Christopher M Reid, Louise M Shiel, Garry L Jennings, and Elena V Lukoshkova. "Rate of Morning Increase in Blood Pressure Is Elevated in Hypertensives". In: *American Journal of Hypertension* 19.10 (2006), pp. 1010–1017.
- [341] Ikuya Nakamura, Kazuo Tsuzuki, and Shigemitsu Ito. "Twenty-four hour monitoring of blood pressure and heart rate in patients with chronic renal failure or renal transplant recipients: Analysis by the cosinor method". In: *Pediatrics International* 37.1 (Feb. 1995), pp. 52–57.
- [342] J R Fernández, R C Hermida, and A Mojón. "Chronobiological analysis techniques. Application to blood pressure". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367.1887 (2009), pp. 431–445.

- [343] Franz Halberg. "Temporal coordination of physiologic function". In: *Cold Spring Harbor symposia on quantitative biology*. Vol. 25. Cold Spring Harbor Laboratory Press, 1960, pp. 289–310.
- [344] Franz Halberg, Yung Liang Tong, and E A Johnson. "Circadian system phase—an aspect of temporal morphology; procedures and illustrative examples". In: *The cellular aspects of biorhythms*. Springer, 1967, pp. 20–48.
- [345] José R Fernández, Artemio Mojón, and Ramón C Hermida. "Comparison of parameters from rhythmometric models with multiple components on hybrid data". In: *Chronobiology international* 21.3 (2004), pp. 469–484.
- [346] Tim Hesterberg. "Bootstrap". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 3.6 (2011), pp. 497–526.
- [347] Philipp Berens. "CircStat: a MATLAB toolbox for circular statistics". In: *Journal of statistical software* 31 (2009), pp. 1–21.
- [348] Alistair E W Johnson, Andrew A Kramer, and Gari D Clifford. "A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy". In: *Critical care medicine* 41.7 (2013), pp. 1711–1718.
- [349] Nishtha Vaidya, Atanu Kumar Pati, and Arti Parganiha. "Circadian variability and nocturnal dipping pattern in blood pressure in young normotensive subjects". In: *Biological Rhythm Research* 43.5 (2012), pp. 485–496.
- [350] Helen McKenna, Gijsbertus T J van der Horst, Irwin Reiss, and Daniel Martin. "Clinical chronobiology: a timely consideration in critical care medicine". In: *Critical Care* 22.1 (2018), p. 124.
- [351] Vasilios Papaioannou, Alexandre Mebazaa, Benoît Plaud, and Matthieu Legrand. "Chronomics' in ICU: circadian aspects of immune response and therapeutic perspectives in the critically ill". In: *Intensive care medicine experimental* 2.1 (2014), pp. 1–14.
- [352] Christina M Spengler, Charles A Czeisler, and Steven A Shea. "An endogenous circadian rhythm of respiratory control in humans". In: *The Journal of physiology* 526.3 (2000), pp. 683–694.
- [353] Martial M Massin, Krystel Maeyns, Nadia Withofs, Françoise Ravet, and Paul Gérard. "Circadian rhythm of heart rate and heart rate variability". In: *Archives of disease in childhood* 83.2 (2000), pp. 179–182.
- [354] Manuel Wallbach, Ellen Born, Ann-Kathrin Schäfer, and Michael J Koziolk. "Effect of baroreflex activation therapy on dipping pattern in patients with resistant hypertension". In: *The Journal of Clinical Hypertension* (2022).
- [355] Yu-Sok Kim, Shyrin C A T Davis, Wim J Stok, Frans J van Ittersum, and Johannes J van Lieshout. "Impaired nocturnal blood pressure dipping in patients with type 2 diabetes mellitus". In: *Hypertension Research* 42.1 (2019), pp. 59–66.
- [356] Terry B J Kuo, Chun-Yu Chen, Yi Ping Wang, Yi-Yun Lan, Kai-Ho Mak, Guo-She Lee, and Cheryl C H Yang. "The role of autonomic and baroreceptor reflex control in blood pressure dipping and nondipping in rats". In: *Journal of Hypertension* 32.4 (2014), pp. 806–816.
- [357] Sarah L Stevens, Sally Wood, Constantinos Koshari, Kathryn Law, Paul Glasziou, Richard J Stevens, and Richard J McManus. "Blood pressure variability and cardiovascular disease: systematic review and meta-analysis". In: *bmj* 354 (2016).
- [358] Eeva P Juhanoja, Teemu J Niiranen, Jouni K Johansson, Pauli J Puukka, Lutgarde Thijs, Kei Asayama, Ville L Langén, Atsushi Hozawa, Lucas S Aparicio, and Takayoshi Ohkubo. "Outcome-driven thresholds for increased home blood pressure variability". In: *Hypertension* 69.4 (2017), pp. 599–607.
- [359] Rianne A A De Heus, Christophe Tzourio, Emily Jo Lynn Lee, Melissa Opozda, Andrew D Vincent, Kaarin J Anstey, Albert Hofman, Kazuomi Kario, Simona Lattanzi, and Lenore J Launer. "Association between blood pressure variability with dementia and cognitive impairment: a systematic review and meta-analysis". In: *Hypertension* 78.5 (2021), pp. 1478–1489.

- [360] Michael E Ernst, Joanne Ryan, Enayet K Chowdhury, Karen L Margolis, Lawrence J Beilin, Christopher M Reid, Mark R Nelson, Robyn L Woods, Raj C Shah, and Suzanne G Orchard. "Long-term blood pressure variability and risk of cognitive decline and dementia among older adults". In: *Journal of the American Heart Association* 10.13 (2021), e019613.
- [361] Meghan Prin and Hannah Wunsch. "International comparisons of intensive care: informing outcomes and improving standards". In: *Current opinion in critical care* 18.6 (2012), p. 700.
- [362] Hannah Wunsch, Derek C Angus, David A Harrison, Walter T Linde-Zwirble, and Kathryn M Rowan. "Comparison of medical admissions to intensive care units in the United States and United Kingdom". In: *American journal of respiratory and critical care medicine* 183.12 (2011), pp. 1666–1673.
- [363] Harriët M R van Goor, Kim van Loon, Martine J M Breteler, Cornelis J Kalkman, and Karin A H Kaasjager. "Circadian patterns of heart rate, respiratory rate and skin temperature in hospitalized COVID-19 patients". In: *Plos one* 17.7 (2022), e0268065.
- [364] Xinxia Li, Chenghong Hu, Airon Meng, Yanmei Guo, Yang Chen, and Rongqing Dang. "Heart rate variability and heart rate monitoring of nurses using PPG and ECG signals during working condition: A pilot study". In: *Health science reports* 5.1 (2022), e477.
- [365] Eibe Frank and Mark Hall. "A simple approach to ordinal classification". In: *European conference on machine learning*. Springer, 2001, pp. 145–156.
- [366] Haibo He and Edwardo A Garcia. "Learning from imbalanced data". In: *IEEE Transactions on knowledge and data engineering* 21.9 (2009), pp. 1263–1284.
- [367] Jacob Cohen. "A coefficient of agreement for nominal scales". In: *Educational and psychological measurement* 20.1 (1960), pp. 37–46.
- [368] Daniel Berrar. *Cross-Validation*. 2019.
- [369] Charles E McCulloch. "Maximum likelihood algorithms for generalized linear mixed models". In: *Journal of the American statistical Association* 92.437 (1997), pp. 162–170.
- [370] Thomas Rutten, Jan Willem Bikker CQM, Edwin van den Heuvel, and Ir Michiel Paus CQM. "Mixed-effects random forest model for quantifying relations in clustered data". In: (2021).
- [371] Anna Gottard, Giulia Vannucci, Leonardo Grilli, and Carla Rampichini. "Mixed-effect models with trees". In: *Advances in Data Analysis and Classification* (July 2022), pp. 1–31.
- [372] Natalie A Bello, Byron JAEGER, John N BOOTH III, Marwah Abdalla, D Edmund Anstey, Daniel N Pugliese, Cora E Lewis, Samuel S Gidding, Donald Lloyd-Jones, and Joseph E SCHWARTZ. "Associations of Awake and Asleep Blood Pressure and Blood Pressure Dipping with Abnormalities of Cardiac Structure: The Coronary Artery Risk Development in Young Adults (CARDIA) Study". In: *Journal of hypertension* 38.1 (2020), p. 102.
- [373] Tine W. Hansen, Yan Li, José Boggia, Lutgarde Thijs, Tom Richart, and Jan A. Staessen. *Predictive role of the nighttime blood pressure*. 2011.
- [374] F Portaluppi and L Montanari. "Consistency of circadian blood pressure pattern assessed by non-invasive monitoring and cosinor analysis in hospitalized hypertensive patients." In: *Acta cardiologica* 43.5 (1988), pp. 605–613.
- [375] Emily Lam, Shahrose Aratia, Julian Wang, and James Tung. "Measuring heart rate variability in free-living conditions using consumer-grade photoplethysmography: Validation study". In: *JMIR Biomedical Engineering* 5.1 (2020), e17355.
- [376] Kazuomi Kario, Satoshi Hoshide, Kimiyo Saito, Keiko Sato, Haruna Hamasaki, Hiromi Suwa, and Naoko Tomitani. "Validation of the TM-2441 ambulatory blood pressure measurement device according to the ISO 81060-2: 2013 standard". In: *Blood Pressure Monitoring* 24.1 (2019), pp. 38–41.

- [377] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.
- [378] Hui Zou and Trevor Hastie. “Regularization and variable selection via the elastic net”. In: *Journal of the royal statistical society: series B (statistical methodology)* 67.2 (2005), pp. 301–320.
- [379] Adam Mahdi, Peter Watkinson, Richard J McManus, and Lionel Tarassenko. “Circadian blood pressure variations computed from 1.7 million measurements in an acute hospital setting”. In: *American journal of hypertension* 32.12 (2019), pp. 1154–1161.
- [380] Adele Cutler, D Richard Cutler, and John R Stevens. “Random forests”. In: *Ensemble machine learning*. Springer, 2012, pp. 157–175.
- [381] M Radha, G Zhang, J Gelissen, K de Groot, R Haakma, and R M Aarts. “Arterial path selection to measure pulse wave velocity as a surrogate marker of blood pressure”. In: *Biomedical Physics & Engineering Express* 3.1 (2017), p. 015022.
- [382] Domenic A Sica and Dawn K Wilson. “Sodium, Potassium, the Sympathetic Nervous System, and the Renin—Angiotensin System”. In: *Blood pressure monitoring in cardiovascular medicine and therapeutics* (2001), pp. 171–189.
- [383] Helen Jones, Greg Atkinson, Andrew Leary, Keith George, Michael Murphy, and Jim Waterhouse. “Reactivity of ambulatory blood pressure to physical activity varies with time of day”. In: *Hypertension* 47.4 (2006), pp. 778–784.
- [384] Elisa Mejía-Mejía, Karthik Budidha, Panayiotis A Kyriacou, and Mohammad Mamouei. “Comparison of pulse rate variability and morphological features of photoplethysmograms in estimation of blood pressure”. In: *Biomedical Signal Processing and Control* 78 (2022), p. 103968.
- [385] Yang Wang, Zhiwen Liu, and Shaodong Ma. “Cuff-less blood pressure measurement from dual-channel photoplethysmographic signals via peripheral pulse transit time with singular spectrum analysis”. In: *Physiological Measurement* 39.2 (2018).
- [386] Sarah Nogueira, Konstantinos Sechidis, and Gavin Brown. “On the stability of feature selection algorithms”. In: *The Journal of Machine Learning Research* 18.1 (2017), pp. 6345–6398.
- [387] Marco A.F. Pimentel, David A. Clifton, Lei Clifton, Peter J. Watkinson, and Lionel Tarassenko. “Modelling physiological deterioration in post-operative patient vital-sign data”. In: *Medical and Biological Engineering and Computing* 51.8 (2013), pp. 869–877.
- [388] Yacong Bo, Kin On Kwok, Vincent Chi Ho Chung, Chun Pong Yu, Kelvin Kam Fai Tsoi, Samuel Yeung Shan Wong, and Eric Kam Pui Lee. “Short-term reproducibility of ambulatory blood pressure measurements: a systematic review and meta-analysis of 35 observational studies”. In: *Journal of hypertension* 38.11 (Nov. 2020), pp. 2095–2109.
- [389] Alan L. Hinderliter, Faye S. Rutledge, James A. Blumenthal, Gary Koch, Michael A. Hussey, William K. Wohlgenuth, and Andrew Sherwood. “Reproducibility of blood pressure dipping: Relation to day-to-day variability in sleep quality”. In: *Journal of the American Society of Hypertension* 7.6 (Nov. 2013), pp. 432–439.
- [390] Alexey Natekin and Alois Knoll. “Gradient boosting machines, a tutorial”. In: *Frontiers in Neurobotics* 7.DEC (Dec. 2013), p. 21.
- [391] Martin Sewell. “Ensemble Learning”. In: (2011).
- [392] Sepp Hochreiter and Jürgen Schmidhuber. “LSTM can solve hard long time lag problems”. In: *Advances in Neural Information Processing Systems*. Vol. 9. 1997, pp. 473–479.
- [393] Dongseok Lee, Hyunbin Kwon, Dongyeon Son, Heesang Eom, Cheolsoo Park, Yonggyu Lim, Chulhun Seo, and Kwangsuk Park. *Beat-to-beat continuous blood pressure estimation using bidirectional long short-term memory network*. Dec. 2021.

- [394] Lei Clifton, David A Clifton, Marco A.F. Pimentel, Peter J Watkinson, and Lionel Tarassenko. "Gaussian process regression in vital-sign early warning systems". In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. 2012, pp. 6161–6164.
- [395] Soojeong Lee, Gyanendra Prasad Joshi, Anish Prasad Shrestha, Chang Hwan Son, and Gangseong Lee. "Cuffless Blood Pressure Estimation with Confidence Intervals using Hybrid Feature Selection and Decision Based on Gaussian Process". In: *Applied Sciences (Switzerland)* 13.2 (Jan. 2023), p. 1221.
- [396] Qi Huang, Dwayne Cohen, Sandra Komarzynski, Xiao Mei Li, Pasquale Innominato, Francis Lévi, and Bärbel Finkenstädt. "Hidden Markov models for monitoring circadian rhythmicity in telemetric activity data". In: *Journal of the Royal Society Interface* 15.139 (2018).
- [397] Yongbo Liang, Zhencheng Chen, Rabab Ward, and Mohamed Elgendi. "Photoplethysmography and deep learning: Enhancing hypertension risk stratification". In: *Biosensors* 8.4 (2018).
- [398] Marco A.F. Pimentel, David A. Clifton, Lei Clifton, and Lionel Tarassenko. *A review of novelty detection*. June 2014.
- [399] Payam Refaeilzadeh, Lei Tang, and Huan Liu. "On comparison of feature selection algorithms". In: 2007.
- [400] S. Rhee, B. H. Yang, and H. H. Asada. "Artifact-resistant power-efficient design of finger-ring plethysmographic sensors". In: *IEEE Transactions on Biomedical Engineering* 48.7 (2001), pp. 795–805.
- [401] D. C. Zheng and Y. T. Zhang. "A Ring-type Device for the Noninvasive Measurement of Arterial Blood Pressure". In: *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*. Vol. 4. 2003, pp. 3184–3187.
- [402] Vera Hartmann, Haipeng Liu, Fei Chen, Qian Qiu, Stephen Hughes, and Dingchang Zheng. "Quantitative comparison of photoplethysmographic waveform characteristics: Effect of measurement site". In: *Frontiers in Physiology* 10 (Mar. 2019).
- [403] Alessandro Zompanti, Anna Sabatini, Simone Grasso, Giorgio Pennazza, Giuseppe Ferri, Gianluca Barile, Massimo Chello, Mario Lusini, and Marco Santonico. "Development and test of a portable ecg device with dry capacitive electrodes and driven right leg circuit". In: *Sensors* 21.8 (Apr. 2021), p. 2777.
- [404] Nino Isakadze and Seth S. Martin. *How useful is the smartwatch ECG?* Oct. 2020.
- [405] Carlos Areia, Louise Young, Sarah Vollam, Jody Ede, Mauro Santos, Lionel Tarassenko, and Peter Watkinson. "Wearability testing of ambulatory vital sign monitoring devices: prospective observational cohort study". In: *JMIR mHealth and uHealth* 8.12 (2020), e20214.
- [406] Carlos Morgado Areia, Mauro Santos, Sarah Vollam, Marco Pimentel, Louise Young, Cristian Roman, Jody Ede, Philippa Piper, Elizabeth King, and Owen Gustafson. "A chest patch for continuous vital sign monitoring: Clinical validation study during movement and controlled hypoxia". In: *Journal of Medical Internet Research* 23.9 (2021), e27547.
- [407] Chang Sei Kim, Andrew M. Carek, Ramakrishna Mukkamala, Omer T. Inan, and Jin Oh Hahn. "Ballistocardiogram as proximal timing reference for pulse transit time measurement: Potential for cuffless blood pressure monitoring". In: *IEEE Transactions on Biomedical Engineering* 62.11 (2015), pp. 2657–2664.
- [408] Sungtae Shin, Azin Mousavi, Sophia Lyle, Elisabeth Jang, Peyman Yousefian, Ramakrishna Mukkamala, Dae-Geun Jang, Ui Kun Kwon, Younho Kim, and Jin-Oh Hahn. "Posture-Dependent Variability in Wrist Ballistocardiogram-Photoplethysmogram Pulse Transit Time: Implication to Cuff-Less Blood Pressure Tracking". In: *IEEE Transactions on Biomedical Engineering* (2021).

- [409] Tai Le, Floranne Ellington, Tao-Yi Lee, Khuong Vo, Michelle Khine, Sandeep Kumar Krishnan, Nikil Dutt, and Hung Cao. "Continuous non-invasive blood pressure monitoring: a methodological review on measurement techniques". In: *IEEE Access* 8 (2020), pp. 212478–212498.
- [410] Jingyu Choi, Younghwan Kang, Jaesoon Park, Yeunho Joung, and Chiwan Koo. "Development of Real-Time Cuffless Blood Pressure Measurement Systems with ECG Electrodes and a Microphone Using Pulse Transit Time (PTT)". In: *Sensors* 2023, Vol. 23, Page 1684 23.3 (Feb. 2023), p. 1684.
- [411] Josep Sola, Meritxell Cortes, David Perruchoud, Bastien De Marco, Melvin D Lobo, Cyril Pellaton, Gregoire Wuerzner, Naomi D L Fisher, and Jay Shah. "Guidance for the Interpretation of Continual Cuffless Blood Pressure Data for the Diagnosis and Management of Hypertension". In: *Frontiers in Medical Technology* (2022), p. 25.
- [412] Benjamin W Nelson, Carissa A Low, Nicholas Jacobson, Patricia Areán, John Torous, and Nicholas B Allen. "Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research". In: *NPJ digital medicine* 3.1 (2020), pp. 1–9.
- [413] Michael R Harwell. "Choosing between parametric and nonparametric tests". In: *Journal of Counseling & Development* 67.1 (1988), pp. 35–38.
- [414] William H Kruskal and W Allen Wallis. "Use of ranks in one-criterion variance analysis". In: *Journal of the American statistical Association* 47.260 (1952), pp. 583–621.
- [415] Sylvain Arlot and Alain Celisse. "A survey of cross-validation procedures for model selection". In: *Statistics surveys* 4 (2010), pp. 40–79.
- [416] Joao Jorge. *Non-contact monitoring of respiration in the neonatal intensive care unit*. 2018.