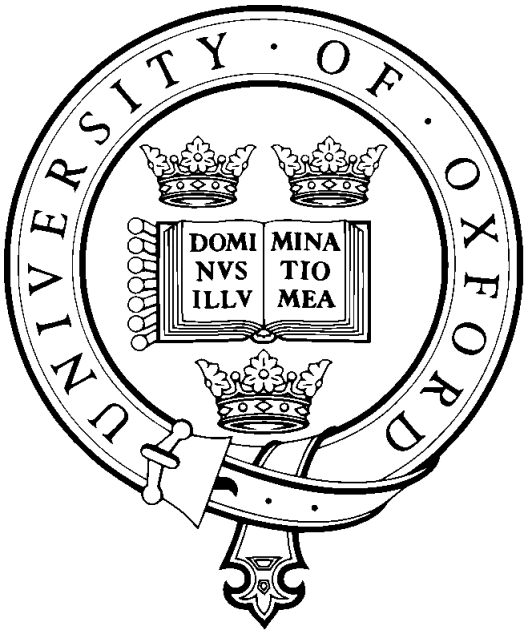
Respiratory Quality Indices for Automated Monitoring of Respiration from Sensor Data



Drew A. Birrenkott

Trinity College

Supervised by

Professor David A. Clifton

Submitted: Michaelmas Term 2018

This thesis is submitted to the Department of Engineering Science, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Declaration

The work contained herein is entirely my own work except where otherwise indicated.

Drew A. Birrenkott
11 January 2019

Acknowledgements

So many people have helped me in both large and small ways during my DPhil journey, and I am thankful to every one of them. Firstly, I must thank my supervisor, Professor David Clifton. Your faith in my abilities from the very beginning has always given me the motivation to work harder and think deeper. Thank you to Dr. Marco Pimentel, you have always been a source of great experience and knowledge. Thank you also to Dr. Peter Watkinson for always bringing a clinician's eye to my work.

Thank you to everyone, old and new, on Time-Series Terrace that I have had the opportunity to share desk space, lunch, coffee, and ideas with, especially Dr. Kate Niehaus, Dr. Rebecca Pullon, Dr. Johanna (Ernst) O'Donnell, Farah Colchester, Sarah-Jane Rodgers, Farah Shamout, Yanting Shen, and Heloise Greeff. I am truly a better engineer and person for having had the opportunity to work with so many wonderful people. Thank you also to all of the other members of the continually growing Computational Health Informatics laboratory and the Biomedical Signal Processing & m-Health group including Dr. Tingting Zhu, Dr. Peter Charlton, Dr. Hamish Tomlinson, Dr. Glen Wright Colopy, Dr. Mauro Santos, Peter Gyring, and so many others. All of your insights and ideas have continually pushed the boundaries of my knowledge.

I am especially grateful to the housemates and friends with whom I have had the opportunity to celebrate many successes and relied on to deal with challenges. To my housemates at Bullingdon Road, Adam Mastroianni, Field Brown, Alex Diaz, Vinay Nayak, and Brian McGrail, thank you for making sure our living room was always a lively and inviting place to decompress. To my housemates at Juxon Street, Aidan Daly, Mike Norton, Phil Maffettone, and Ben Daus-Haberle, thank you for filling the darkest parts of my DPhil with light. Thank you also to Andrew Hovel, Sam Polstein, Moizza Sarwar, Aurora Griffin, Lien Davidson, Sai Gourisankar, Kieran Fitzpatrick, Dylan Smith, Mark Jbeily, Helen Hansen Jbeily, Jonathan Pedde, and Erin Hylton, you all have made my life in Oxford and beyond an adventure. Thank you to everyone at the Old Bookbinders, especially Josh Mullett-Sadones, Annie Cutting, and of course, Belle. They told me that when I moved to the UK I would have to find a pub to be my home away from home. I am glad I found mine.

Thank you to the Rhodes Trust for funding this work and for giving me the opportunity to meet so many amazing people during my time in Oxford. Thank you especially to Mary Eaton, Bob Wylie, Dawn Wylie, and the Rhodes House Porters—you have helped me out of more tough spots than I care to count. Thank you also to everyone at Trinity College and especially the Trinity College Porters. You have always made Trinity a welcoming place to work, study, and socialise.

My deepest thanks goes to my family. Thank you to all of you for always sending your love when I was so far away. Thank you to my brothers, Brant and Scott, for never letting me forget that no matter what I do, I am still the middle child. Thanks, of course, to you, Mom and Dad. I have always felt your constant love and support in everything I have done. And to Morgan, thank you for always being there, I love you.

Finally, thanks to you, the reader. I learned a lot from writing this thesis, I hope you learn at least a little by reading it.

Respiratory Quality Indices for Automated Monitoring of Respiration from Sensor Data

Drew A. Birrenkott
Thesis submitted for the degree of Doctor of Philosophy
Trinity College
Michaelmas Term 2018

Abstract

Abnormal respiratory rate (RR) is known to be one of the most clinically effective predictors of catastrophic decline. Despite this, RR is often the least monitored and most inaccurately measured vital sign. This is primarily because of the lack of a non-invasive, robust, automated method for estimating RR. It has previously been shown that the amplitude modulation (AM), frequency modulation (FM), and baseline wander (BW) of both the electrocardiogram (ECG) and photoplethysmogram (PPG) contain respiratory waveform information. However, these respiratory modulations are driven by the physiologic interrelationship between the cardiovascular and respiratory systems and may or may not be present based on a patient's characteristics and condition. Despite this, current methods for RR estimation from ECG and PPG do not account for this physiologic variability. The investigations in this thesis describe the development and evaluation of respiratory quality indices (RQIs), a novel method for evaluating the presence or absence of physiologically important respiratory information from the AM, FM, and BW extracted from the ECG and PPG. This work is conducted in three unique data sets, CapnoBase, MIMIC-III, and Dialysis III, all of which represent important, different patient populations. Five initial RQIs are described based on five signal processing techniques: fast Fourier transform (FFT), autocorrelation, cosine correlation, autoregression, and Hjorth parameters. Of these, the individual RQIs based on the FFT, autocorrelation, and autoregression are deemed to be good predictors of the presence of respiratory waveform data. The three individual RQIs are used to derive three fusion RQIs based on two supervised learning algorithms: linear regression and support vector regression (SVR) and one unsupervised learning algorithm: principal component analysis (PCA). Both the linear regression and PCA fusion RQIs are accurate and robust. The linear regression fusion RQI is used in the development of an RR estimation algorithm, termed RQIFusion, which achieves highly accurate and more complete RR estimates than existing methods. In the Dialysis III data set, implementation of RQIFusion improved RR estimation error by between 1.35 to 2.29 breaths per minute (brpm) to achieve RR estimation errors between 2.18 to 3.46 brpm, depending on the RR estimation algorithm employed. These results represent a marked improvement in RR estimation and indicate the importance of conducting respiratory quality analysis using RQIs on respiratory modulations extracted from ECG and PPG prior to RR estimation.

Related Publications

Journals and Conference Proceedings

D.A. Birrenkott, M.A.F. Pimentel, P.J. Watkinson, and D.A. Clifton, "A robust fusion model for estimating respiratory rate from photoplethysmography and electrocardiography," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 2033-2041, 2018.

P. Charlton, **D.A. Birrenkott**, T. Bonnici, M.A.F. Pimentel, A.E.W. Johnson, J. Alastruey, L. Tarassenko, P.J. Watkinson, R. Beale, and D.A. Clifton, "Breathing rate estimation from the electrocardiogram and photoplethysmogram: a review," *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 2-20, 2017.

M.A.F. Pimentel, A.E.W. Johnson, P. Charlton, **D.A. Birrenkott**, P.J. Watkinson, L. Tarassenko, and D.A. Clifton, "Towards a robust estimation of respiratory rate from pulse oximeters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914-1923, 2017.

D.A. Birrenkott, M.A.F. Pimentel, P.J. Watkinson, and D.A. Clifton, "Robust estimation of respiratory rate via ECG- and PPG-derived respiratory quality indices," in *the Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 676-679, 2016.

Patents

D.A. Clifton, **D.A. Birrenkott**, M.A.F. Pimentel, and L. Tarassenko (Oxford, UK), "Method of determining the frequency of a periodic physiological process of a subject and a device and system for determining the frequency of a periodic physiological process of a subject," WIPO Patent WO2017194915, 2017. *Technology exclusively licensed to Sensyne Health PLC in 2018.*

Contents

1 Introduction **1**
- 1.1 Importance of Respiratory Rate Monitoring . . . 1
- 1.2 Current RR Monitoring Methods . . . 3
- 1.3 Assessing the Quality of Biomedical Signals . . . 5
- 1.4 Structure of Thesis . . . 6

2 Respiratory Physiology **8**
- 2.1 Physiological Interactions of the Cardiovascular and Respiratory Systems 9
  - 2.1.1 Respiratory System Modulations of the Cardiovascular System by Physical Means . . . 9
  - 2.1.2 Respiratory System Modulations of the Cardiovascular System via Autonomic Nervous System Control . . . 11
- 2.2 Monitoring the Cardiovascular and Respiratory Systems . . . 13
  - 2.2.1 Impedance Plethysmography . . . 14
  - 2.2.2 Capnography . . . 14
  - 2.2.3 The PPG . . . 15
  - 2.2.4 The ECG . . . 24
- 2.3 Conclusion . . . 30

3 Literature Review **32**
- 3.1 Previous Work . . . 34
- 3.2 Module I: Waveform Extraction . . . 35
  - 3.2.1 Feature-Based Methods . . . 35
  - 3.2.2 Frequency-Based Methods . . . 39
- 3.3 Module II: Extraction of Respiratory Rate . . . 43
  - 3.3.1 Time-Domain Methods . . . 44
  - 3.3.2 Frequency-Domain Methods . . . 45
- 3.4 Module III: Respiratory Rate Fusion . . . 48
- 3.5 Checkpoint I: Signal Quality Index (SQI) . . . 50
- 3.6 Checkpoint II: Respiratory Quality Index (RQI) . . . 51
- 3.7 Single-Step Methods . . . 53
- 3.8 Conclusion . . . 54

4 Data Sets **55**
- 4.1 CapnoBase Database . . . 56
- 4.2 MIMIC-III Database . . . 57
- 4.3 Oxford Dialysis III Database . . . 59

4.4 Data Set Distinctions . . . 60
4.4.1 Age Range . . . 61
4.4.2 Reason for Admission . . . 62
4.5 Conclusion . . . 63

5 Individual RQIs **64**
5.1 Techniques . . . 65
5.1.1 The Fourier Transform . . . 65
5.1.2 Autocorrelation . . . 65
5.1.3 Cross-Correlation . . . 67
5.1.4 Autoregression . . . 68
5.1.5 Hjorth Parameters . . . 72
5.2 Methods . . . 73
5.2.1 Pre-Processing . . . 74
5.2.2 Individual RQIs . . . 77
5.2.3 RQI Performance Evaluation Methods . . . 81
5.3 Results . . . 92
5.3.1 FFT RQI ($Q_{S,1}$) . . . 93
5.3.2 Autocorrelation RQI ($Q_{S,2}$) . . . 94
5.3.3 Cosine-Correlation RQI ($Q_{S,3}$) . . . 95
5.3.4 Autoregression RQI ($Q_{S,4}$) . . . 97
5.3.5 Hjorth Complexity RQI ($Q_{S,5}$) . . . 97
5.3.6 Inter-RQI Comparison . . . 99
5.4 Discussion . . . 101
5.5 Conclusion . . . 103

6 Fusion RQIs **105**
6.1 Techniques . . . 106
6.1.1 Linear Regression . . . 106
6.1.2 Support Vector Regression . . . 109
6.1.3 Principal Component Analysis . . . 115
6.2 Methods . . . 118
6.2.1 Pre-Processing . . . 119
6.2.2 Fusion RQIs . . . 120
6.2.3 Post-Processing . . . 124
6.2.4 Fusion RQI Performance Evaluation Methods . . . 125
6.3 Results . . . 127
6.3.1 Data Standardisation . . . 128
6.3.2 Linear Regression Fusion RQI ($Q_{F,1}$) . . . 129
6.3.3 Support Vector Regression Fusion RQI ($Q_{F,2}$) . . . 133
6.3.4 Principal Component Analysis Fusion RQI ($Q_{F,3}$) . . . 138
6.4 Discussion . . . 143
6.5 Conclusion . . . 147

7 RQI-Based RR Estimation and Time-Series Analysis 149
7.1 Techniques . . . 149
7.1.1 Median Filtering . . . 150
7.1.2 Cubic Spline Interpolation . . . 151
7.2 Assessment of RQIFusion . . . 154
7.2.1 Methods . . . 154
7.2.2 Results . . . 158
7.3 RQI Implementation in RR Estimation Work Flow . . . 165
7.3.1 Methods . . . 166
7.3.2 Results . . . 178
7.4 Discussion . . . 197
7.5 Conclusion . . . 204

8 Conclusion 206
8.1 Future Work . . . 210
8.1.1 EWS Systems . . . 210
8.1.2 Wearable Technologies . . . 211
8.2 Conclusion . . . 214

Appendix 216

A Additional Results 216
A.1 Comparison of RR Estimation Methods . . . 216
A.2 Robustness Measurement of Individual RQIs . . . 217
A.3 Time-Series Analysis . . . 220

References 226

List of Figures

2.1 Idealised PPG waveform and features. 17
2.2 Absorption coefficient vs. wavelength of light for Hb and HbO_2. 20
2.3 Respiratory modulations of the PPG waveform. 23
2.4 Idealized ECG waveform and features. 28
2.5 Respiratory modulations of the ECG waveform. 30

3.1 Respiratory waveform and RR extraction framework. 33

4.1 Relative frequency of age distributions for for $\mathcal{D}_c$, $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, $\mathcal{D}_m$, and $\mathcal{D}_d$ data sets. 62

5.1 Performance of the $Q_{S,1}$ individual RQI in the combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. 94
5.2 Performance of the $Q_{S,2}$ individual RQI in the combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. 95
5.3 Performance of the $Q_{S,3}$ individual RQI in the combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. 96
5.4 Performance of the $Q_{S,4}$ individual RQI in the combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. 98
5.5 Performance of the $Q_{S,5}$ individual RQI in the combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. 99
5.6 Frequencies with which individual RQIs perform within 10% and 20% of expected performance. 101

6.1 MSE of training and validation folds for linear regression fusion RQI ($Q_{F,1}$). 129
6.2 Performance of the $Q_{F,1}$ fusion RQI in the training and validation data set, $\mathcal{D}_{c,1}$. 131
6.3 Performance of the $Q_{F,1}$ fusion RQI in the test data set, $\mathcal{D}_{c,2}$. 132
6.4 Performance of the $Q_{F,1}$ fusion RQI in the test data set, $\mathcal{D}_m$. 133
6.5 Robustness comparison of $Q_{F,1}$ to $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$. 134
6.6 MSE of training and validation folds for coarse grid search for fusion RQI SVR ($Q_{F,2}$). 135
6.7 MSE of training and validation folds for fine grid search for fusion RQI SVR ($Q_{F,2}$). 135
6.8 Performance of the $Q_{F,2}$ fusion RQI in the training and validation data set, $\mathcal{D}_{c,1}$. 136
6.9 Performance of the $Q_{F,2}$ fusion RQI in the test data set, $\mathcal{D}_{c,2}$. 138

6.10 Performance of the $Q_{F,2}$ fusion RQI in the test data set, $\mathcal{D}_m$. 139
6.11 Robustness comparison of $Q_{F,2}$ to $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$. 140
6.12 Performance of the $Q_{F,3}$ fusion RQI in the training and validation data set, $\mathcal{D}_{c,1}$. 141
6.13 Performance of the $Q_{F,3}$ fusion RQI in the test data set, $\mathcal{D}_{c,2}$. 142
6.14 Performance of the $Q_{F,3}$ fusion RQI in the test data set, $\mathcal{D}_m$. 143
6.15 Robustness comparison of $Q_{F,3}$ to $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$. 144

7.1 MAE of RR estimates for windows vs. percentage of windows discarded for RQIFusion with varying thresholds, T, and comparator methods. . . . 159
7.2 MAE of RR estimates and percentage of windows discarded for RQIFusion with varying thresholds, T_{LT}, T_{HT}, and T_{SF}. 161
7.3 Time-series performance of $\text{RQIFusion}_{\text{HT}}$ compared to Smart Fusion, modified Smart Fusion, and R_{ref} for two subjects in $\mathcal{D}_{c,2}$. 164
7.4 Relative frequency of windows discarded for $\text{RQIFusion}_{\text{LT}}$, $\text{RQIFusion}_{\text{HT}}$, Smart Fusion, and modified Smart Fusion in $\mathcal{D}_c$ and $\mathcal{D}_m$. 165
7.5 Performance and window discard measures of RQIFusion with five exemplar values for T. 180
7.6 120-minute time-series of the RR difference compared to the reference for the $\text{SQI/RQIFusion}_{\text{HT}}$ and the RR average methods for $\mathcal{D}_d$ subjects 1 to 4. 186
7.7 120-minute time-series of the RR difference compared to the reference for the $\text{SQI/RQIFusion}_{\text{HT}}$ and the RR Average methods for $\mathcal{D}_d$ subjects 5 to 8. 187
7.8 120-minute time-series performance of $\text{SQI/RQIFusion}_{\text{HT}}$ compared to RR average method for $\mathcal{D}_d$ data set subject 2. 190
7.9 120-minute time-series performance of $\text{SQI/RQIFusion}_{\text{HT}}$ compared to RR average method for $\mathcal{D}_d$ data set subject 3. 192
7.10 120-minute time-series performance of $\text{SQI/RQIFusion}_{\text{HT}}$ compared to RR average method for $\mathcal{D}_d$ data set subject 8. 193
7.11 120-minute time-series performance of $\text{SQI/RQIFusion}_{\text{HT}}$ compared to RR average method for $\mathcal{D}_d$ data set subject 1. 195

8.1 Diagram of platform integrating RQI RR estimation methodology into mobile technologies. 215

A.1 Frequencies with which $Q_{S,1}$ falls within expected decile based on rank ordering of RR error. 218
A.2 Frequencies with which $Q_{S,2}$ falls within expected decile based on rank ordering of RR error. 218
A.3 Frequencies with which $Q_{S,3}$ falls within expected decile based on rank ordering of RR error. 219
A.4 Frequencies with which $Q_{S,4}$ falls within expected decile based on rank ordering of RR error. 219
A.5 Frequencies with which $Q_{S,5}$ falls within expected decile based on rank ordering of RR error. 220
A.6 Time-series performance of $\text{RQIFusion}_{\text{HT}}$ compared to Smart Fusion, modified Smart Fusion, and R_{ref} for eight subjects in $\mathcal{D}_{c,2}$. 221
A.7 120-minute time-series performance of $\text{SQI/RQIFusion}_{\text{HT}}$ compared to RR average method for $\mathcal{D}_d$ data set subject 4. 222

A.8 120-minute time-series performance of SQI/RQIFusion$_{\mathrm{HT}}$ compared to RR average method for $\mathcal{D}_d$ data set subject 5. 223
A.9 120-minute time-series performance of SQI/RQIFusion$_{\mathrm{HT}}$ compared to RR average method for $\mathcal{D}_d$ data set subject 6. 224
A.10 120-minute time-series performance of SQI/RQIFusion$_{\mathrm{HT}}$ compared to RR average method for $\mathcal{D}_d$ data set subject 7. 225

List of Tables

4.1 Physiological information on $\mathcal{D}_c$, $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, $\mathcal{D}_m$, and $\mathcal{D}_d$ data sets 57

5.1 Spearman's Rank Order Correlations between individual RQIs ($Q_{S,[1...5]}$) and control RQIs ($Q_{C,[1...2]}$). 100

6.1 Values for zero-mean, unit-variance normalisation of the individual RQIs ($Q_{S,[1,2,3]}$). 128
6.2 Eigenvectors and eigenvalues of PCA decomposition of $Q_{S,[1,2,4]}$ in the $\mathcal{D}_{c,1}$ data set. 139
6.3 Comparative performance statistics for $Q_{F,1}$ and $Q_{F,3}$ in the $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$ data sets. 147

7.1 Equivalent percentage of windows discarded (Dis. %) or equivalent MAE for RQIFusion vs. comparator methods in $\mathcal{D}_{c,1}$. 160
7.2 Performance of exemplar threshold models (median, MAE, and % discard) compared to existing fusion methods for all three data sets, $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, and $\mathcal{D}_m$. 162
7.3 Median (IQR), MAE (± SD), and % of windows discarded for ECG and PPG Smart Fusion compared to $\text{RQIFusion}_{\text{DIS}}$. 163
7.4 NEWS2 Scores for specified physiological parameters. 175
7.5 Summary statistics for RR reference in $\mathcal{D}_d$ subjects. 179
7.6 Number and type of modulations kept for individual subjects in the $\mathcal{D}_d$ data set using T_{HT}. 181
7.7 Performance of implementations of five RR estimation methods with varying SQI and RQI implementations in the $\mathcal{D}_d$ data set. 183
7.8 Performance of peak-zero RR estimation method with varying SQI and RQI implementations for individual $\mathcal{D}_d$ subjects. 184
7.9 Analysis of the F1 Score for $\text{SQI/RQIFusion}_{\text{HT}}$ implementation compared to RR Average implementation for individual subjects in the $\mathcal{D}_d$ data set. 196

A.1 Evaluation of the RR error and % of windows discarded in the $\mathcal{D}_c$, $\mathcal{D}_m$, and $\mathcal{D}_d$ data sets using five RR estimation algorithms. 217

Abbreviations

AIC	Akaike's Information Criterion
AM	Amplitude Modulation
AR	Autoregression
AV	Atrioventricular
BP	Blood Pressure
bpm	Beats Per Minute
brpm	Breaths Per Minute
BW	Baseline Wander
CCA	Canonical Correlation Analysis
CDM	Complex Demodulation
COPD	Chronic Obstructive Pulmonary Disease
CSD	Correntropy Spectral Density
CVN	Cardiac Vagal Neurone
CWT	Continuous Wavelet Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiography
EMD	Empirical Mode Decomposition
EMG	Electromyography
$ETCO_2$	End-Tidal CO_2
EWS	Early Warning Score
FFCDM	Fixed Frequency Complex Demodulation
FFT	Fast Fourier Transform
FM	Frequency Modulation
FN	False Negative
FP	False Positive

GP	Gaussian Process
Hb	Haemoglobin
HbO_2	Oxyhaemoglobin
HD	Haemodialysis
HR	Heart Rate
HRV	Heart Rate Variability
ICU	Intensive Care Unit
IMF	Intrinsic Mode Function
IMS	Incremental Merge Segmentation
IP	Impedance Plethysmography
iPPG	Non-Contact Imaging PPG
IQR	Interquartile Range
kPCA	Kernel Principal Component Analysis
LA	Left Arm
LED	Light-Emitting Diode
LL	Left Leg
LR	Linear Regression
MAE	Mean Absolute Error
MAP	Mean Arterial Blood Pressure
ML	Machine Learning
MSE	Mean-Square Error
MSPCA	Multi-Scale Principal Component Analysis
NEWS2	National Early Warning Score 2
OPS	Optimal Parameter Search
PC	Principal Component
PCA	Principal Component Analysis
PPG	Photoplethysmography
PSD	Power Spectral Density
PTT	Pulse Transit Time
PWV	Pulse Width Variability
RA	Right Arm
RAP	Ridge Amplitude Perturbation
RBF	Radial Basis Function

RFP	Ridge Frequency Perturbation
RL	Right Leg
RQI	Respiratory Quality Index
RR	Respiratory Rate
RSA	Respiratory Sinus Arrhythmia
SA	Sinoatrial
SBP	Systolic Blood Pressure
SD	Standard Deviation
SpO_2	Oxygen Saturation
SQI	Signal Quality Index
SV	Stroke Volume
SVD	Singular Value Decomposition
SVR	Support Vector Regression
SWFD	Secondary Wavelet Feature Decoupling
TP	True Positive
TTI	Transthoracic Impedance
TVSINCOS	Time-Varying Sine Cosine
VFCDM	Variable-Frequency Complex Demodulation
WHD	Wearable Health Device

Chapter 1

Introduction

1.1 Importance of Respiratory Rate Monitoring

Changes in a patient's vital signs have long been considered one of the key indicators of a patient's health status and have been used for identifying physiological deterioration [1]. Measuring a patient's vital signs is one of the first interactions a clinician has with the patient, and the outcome typically dictates subsequent actions [2]. While abnormal vital signs cannot usually be used to render specific diagnoses, they are indicative of general malaise. In the presence of other, more condition-specific medical information, abnormal vital signs make a diagnosis more likely than if the vital signs were normal [3]. It has been shown that alterations and derangements in vital signs can be used as an early indicator of critical deterioration up to eight hours prior to an adverse event [4, 5, 6, 7].

The most commonly-accepted vital signs are heart rate (HR), blood pressure (BP), temperature, and respiratory rate (RR) [8]. More recently, some authors have argued that additional vital signs such as oxygen saturation (SpO_2), level of pain, level of consciousness, and urine output might complement these traditional vital signs and serve as indicators of clinical deterioration [3, 8]. In the past 10 years, a substantial amount

of research looking to quantify the level of patient derangement based on vital signs, and determine when that derangement indicates a need for further intervention, has led to the development of Early Warning Score (EWS) systems [7, 9, 10, 11, 12]. EWS systems have been developed both heuristically and based on statistical approaches. EWS systems function by assigning a score to a patient based on their current vital sign measures. In the event that the score exceeds a pre-determined threshold, further action is taken. However, in order for these EWS systems to be optimally effective they rely on high-quality and timely updates for each of the vital signs included in the EWS system. EWS systems are typically used with patients who are outside the intensive care unit (ICU), and where the usual method of monitoring such patients is by manual observation.

One of the most critical decisions in designing an EWS model is determining which vital signs are the most relevant indicators of patient deterioration. From previous research, it has been well-verified that of all the vital signs, RR provides the earliest and most accurate warning for catastrophic deterioration [4, 11, 13, 14, 15]. A study conducted by Schein *et al.* (1990) found that, of patients who suffered from catastrophic deterioration in the hospital, 60% presented with abnormal RR before deterioration. Additionally, the study showed that for 38% of all patients suffering from catastrophic deterioration, abnormal RR was the only abnormal vital sign before the event [4]. It can therefore be argued that RR is one of the most, if not the most, important vital sign to monitor.

Despite the well-known importance of RR in determining a patient's health status, it is the often the least monitored vital sign [16]. Even when RR is recorded, it is often improperly measured or recorded incorrectly [16, 17, 18]. This is mainly due to three

factors: lack of time, lack of knowledge, and lack of technology [17].

1. **Lack of Time:** Hospital staff must monitor increasingly large numbers of patients in most developed health care systems and may not have time to conduct lengthy manual observations at high frequency, as is necessary when measuring RR [17, 19]. Current methods for monitoring RR are discussed in Section 1.2.
2. **Lack of Knowledge:** Medical training often fails to stress the importance of monitoring RR. As such, measuring RR can be overlooked [17].
3. **Lack of Technology:** While technologies to non-invasively monitor HR, BP, temperature, and SpO_2, among others, have been developed and are widely used, technologies developed for monitoring RR have not seen the same widespread use [20]. As this thesis will show, this lack of technology is in part due to the fact that RR is very difficult to measure in a manner that is suitably robust for use in clinical practice.

1.2 Current RR Monitoring Methods

In the vast majority of instances outside the ICU, RR is still monitored via manual observation of the movement of the chest cavity [2]. The clinical standard for manually observing RR is to count the number of breaths taken by a patient in an interval of between 15-60 seconds and multiply up to calculate breaths per minute (brpm) [2]. However, this technique has a number of disadvantages. First, it has been shown that the length of time over which the breaths are counted is critical. One study found large discrepancies in recorded RR for the same patients when nurses counted breaths using a 15-second window and medical students counted breaths using a 60-second window [21]. An additional disadvantage, caused by the length of the counting window, is

the resolution of the RR. For example, when breaths are counted for 15 seconds, the breathing rate is quantised in values of 4 brpm. This represents poor resolution when the typical RR is expected to be between 12 and 20 brpm [3, 22]. There is also a significant "white coat" effect where patients have deranged RR when they become aware that a clinician is monitoring their RR [21]. Finally, despite its importance, RR is often under-monitored. Some authors have shown that RR is monitored, at best, once during every 8-10 hour hospital shift while the the clinical standard in most UK hospitals is measurement every 4 hours. Any deterioration that occurs between these measurements will go unnoticed [23, 24, 25].

The lack of reliability of manually-observed RR has led to a number of technologies for monitoring RR automatically. The primary methods for monitoring RR, and the methods currently in clinical use, are principally focused on two aspects of respiration: (1) the movement of the chest wall during respiration or (2) the flow of air into and out of the mouth and nose. Methods pertaining to the movement of the chest wall include: impedance plethysymography (IP) [26] and transthoracic impedance (TTI) [27]. Methods pertaining to the flow of air into and out of the nose and mouth include: capnography [28], flow thermography [29], and spirometry [30].[1] While these methods are generally robust and can be considered the "gold standards" in RR monitoring, they are not widely used across all hospital and clinical settings—they are typically reserved for use during surgery and in the most acute wards, such as the ICU [31, 32, 33]. This is because these technologies are either too invasive or cumbersome for long-term use, too costly for widespread use, or because they have been purpose-built for specific applications [31, 32, 33, 34].

[1]Both IP and capnography will be used to obtain "gold standard" reference RR in this thesis. A more thorough explanation of the principles of IP and capnography is presented in Chapter 2.

These shortcomings have led to a second class of methods for monitoring RR based on deriving RR as a "secondary signal" from primary signals already nearly ubiquitously collected in health care settings. These "secondary signals" are derived largely from methods originally designed to monitor the heart rate or blood flow such as photoplethymography (PPG) [35], electrocardiography (ECG) [36], and BP [37]. Additionally, some methods have been designed to extract the RR from other signals such as electromyography (EMG) [38]. While all of these methods have previously demonstrated efficacy, the most promising are those that use the ECG and PPG because of the low cost of acquiring these signals [39, 40], their wide availability in clinical settings [41, 42, 43], and the well-understood physiological relationship of respiratory modulation on both signals. The derivation of RR from these two signals will be the primary focus of this thesis.

While many methods have been shown to estimate RR from the ECG and PPG under ideal conditions, they are not yet suitably robust for clinical implementation [44]. This is largely because the secondary modulations of the ECG and PPG from which RR is estimated are patient-specific, may appear and disappear for an individual through time, and vary with gender, age, pre-existing health conditions, level of hydration, body position, and the magnitude of the RR itself [33, 45, 46].

1.3 Assessing the Quality of Biomedical Signals

The estimation of the quality of a signal is a widely-understood necessity in biomedical signal processing [47]. Due to the large number of biological processes occurring at any given time and the high likelihood for signals to be corrupted due to (1) motion artefact and (2) misplaced or disconnected leads, a large sub-field of biomedical engineering has

become devoted to developing signal quality indices (SQIs) [47]. SQIs are metrics that assess the overall quality of a signal. Often for a signal to be used for a clinically-relevant purpose, a scheme is adopted whereby data are used if the SQI values exceed a pre-determined threshold. The development of SQIs for ECG, in particular, has been a major focus [47, 48, 49, 50]. However, to date, very little work has been undertaken for creating SQIs for secondary modulations of these signals, such as those used to estimate RR. Estimation of signal quality for secondary signals is particularly important because even if the primary signal (ECG and PPG) is of high-quality (e.g., is non-artefactural), the secondary signal (corresponding to respiration) may not be present due to various physiological factors, as described in Chapter 2. Therefore, the derivation and validation of respiratory-specific quality indices, hereafter termed respiratory quality indices (RQIs), are necessary. The creation of such RQIs is the primary subject matter in this thesis.

1.4 Structure of Thesis

In Chapter 2, the physiology of the cardiorespiratory system and the known effects of the RR on the HR are described. Additionally, the mechanisms by which the ECG and PPG measure physiological phenomena are explained. A review of the current state-of-the-art methods for extracting RR as a secondary signal from the ECG and PPG are presented in Chapter 3, along with a discussion concerning the performance of these methods. Chapter 4 introduces the three data sets used by the work described in this thesis: CapnoBase [51], MIMIC-III [52, 53, 54], and the Oxford Dialysis Study [55]. In Chapter 5, the development and performance of five individual RQIs is described. Chapter 6 describes the development and validation of fused RQIs from the individual RQIs. In

Chapter 7, the best-performing fusion RQI from Chapter 6 is used in conjunction with an SQI to develop a high-fidelity RR estimation algorithm that is validated on the clinically-relevant Oxford Dialysis Study data. Finally, Chapter 8 summarises the current work, discusses the key findings and limitations of this approach, and proposes areas for future work.

Chapter 2

Respiratory Physiology

The primary role of the cardiovascular and respiratory systems is to carry oxygen and nutrients to (and carry carbon dioxide and other waste products from) the organ systems of the body. While the cardiovascular and respiratory systems are distinct, they are anatomically and functionally coupled and operate in series with each other [56]. Despite the alveolocapillary barrier being the only direct interaction between these two systems, their functions are highly interconnected via physical mechanisms within the thoracic cavity and via nervous system control [56]. The physical and nervous system mechanisms by which the cardiovascular and respiratory system feedback to each other are of particular importance when observing how the action of one system affect the measurement of the other system. In the work described by this thesis, we are interested in observing how the respiratory system modulates commonly-acquired cardiovascular signals.

The focus of this chapter is twofold. The first section will explore the physiological inter-regulation of the cardiovascular and respiratory systems via physical mechanisms and nervous system control. The second section will describe the methods that are used to obtain the ECG and PPG waveforms as well as the IP and capnography waveforms. It

will also explore how the physiological interactions described in the first section manifest in the ECG and PPG and can be used to obtain the respiratory waveform from the ECG and PPG.

2.1 Physiological Interactions of the Cardiovascular and Respiratory Systems

The cardiovascular and respiratory systems function in tandem via both physical mechanisms and nervous system control. The physical mechanisms occur as a result of changes in intrathoracic pressure during respiration. The nervous system control mechanisms operate under the control of the respiratory control centre in the medulla oblongata [57].

2.1.1 Respiratory System Modulations of the Cardiovascular System by Physical Means

Spontaneous respiration, or normal breathing, is the result of changing the pressure differential between the intrathoracic cavity and the atmosphere. At rest, the intrathoracic cavity has a slightly negative pressure differential compared to atmospheric pressure (which ensures that the lungs do not collapse). However, during inspiration, the negative pressure differential is increased by increasing the volume of the intrathoracic cavity via contraction of either or both of the intercostal muscles and diaphragm. Chest breathing is the result of intercostal contraction, while abdominal breathing is the result of diaphragm contraction [58]. According to Boyle's law, a change in intrathoracic volume causes an inverse, proportional change in pressure. The result is that the highly elastic and compliant lungs fill with atmospheric air to compensate for the change, which results in inspiration [58]. Exhalation is the opposite process by which the contracted

muscles relax, causing the intrathoracic cavity to return to its normal volume, and the air that entered the lungs during inspiration to be expired [58].

While the primary function of the changing intrathoracic volume is to force air into and out of the lungs, because much of the cardiovascular system is also located within the intrathoracic cavity, the changes in pressure differentials during respiration have effects on the function of the cardiovascular system [57, 59, 60, 61]. These effects can be grouped into three primary categories: (1) increased venous return to the thoracic cavity [57, 59], (2) changes in left and right ventricular stroke volumes [57, 59, 61], and (3) translational and rotational motion of the heart [60].

Instantaneous Variations in Venous Return During inspiration, the drop in intrathoracic pressure results in increased venous return from the peripheral venous system to the thorax [57]. This is the result of distension of both the pulmonary vessels and the right atrium of the heart, which results in increased flow of blood into the thoracic cavity [57, 59]. The effect is further exacerbated by expansion of the diaphragm, which increases the pressure on the intra-abdominal veins causing a further increased venous return from these vessels [57]. While it might be expected that the flow caused by increased intra-abdominal pressure would be bidirectional, this effect is minimised due to valves within the venous system, which ensures uni-directional flow into the thoracic cavity [57].

Instantaneous Variations in Stroke Volume The stroke volume (SV) of the heart is the volume of blood that is pumped out of the left ventricle on every heartbeat. The SV is reduced during inspiration due to a number of factors. The first factor is that blood pools in the pulmonary vessels due to the previously-noted distensions of

the pulmonary vessels during inspiration, which causes a reduction in ventricular filling [57, 59, 61]. Secondly, the magnitude of the pressure that the heart must work against in the beginning of systole, known as the afterload, is higher during inspiration, which causes a decrease in SV [59]. One reason for this may be the displacement of the septum between the left and right ventricles due to right ventricular expansion during inspiration. This reduces left ventricular compliance and ultimately SV [57]. Finally, heart rate variability (HRV) causes a minor reduction in SV. HRV is a phenomenon in which the HR is increased during inspiration and decreased during exhalation [61]. While the phenomenon is driven by nervous system control and will be described in greater detail in Section 2.1.2, the increased HR during inspiration decreases the ventricular filling and subsequently the SV [61].

Translational and Rotational Motion of Heart Due to the proximity of the heart to the lungs, the heart is subject to both rotation and translation within the thoracic cavity during inspiration. Rotation is counter-clockwise and can be as large as 30 mm [60]. There is also notable translational motion in the longitudinal axis (12.4 $\pm$ 5.9 mm, inferior) and the anterior-posterior axis (4.3 $\pm$ 3.7 mm, anterior). There is also a small amount of translational motion in the transverse axis (2.0 $\pm$ 2.1, right lateral) [60].

2.1.2 Respiratory System Modulations of the Cardiovascular System via Autonomic Nervous System Control

The autonomic nervous system plays a major role in the coordination of the HR and RR. The primary mechanism by which the RR modulates the HR is respiratory sinus arrhythmia (RSA) [46]. RSA is a phenomenon in which the beat-to-beat heart rate

interval is decreased (i.e., the heart beats faster) during inspiration and the beat-to-beat heart rate interval is increased (i.e., the heart beats slower) during expiration [46]. This phenomenon may also be referred to less specifically as HRV. RSA occurs via two different autonomic nervous system receptors: the arterial baroreceptors and pulmonary stretch receptors.

The two most active arterial baroreceptors are in the aortic arch and the carotid artery, and serve the purpose of maintaining the arterial blood pressure. When the arterial blood pressure is decreased, the arterial baroreceptors are activated and cause an increase in heart rate [61]. This mechanism serves as a means by which RR causes RSA because the arterial blood pressure is reduced during inspiration [61, 62].

In addition to activation of the arterial baroreceptors, the pulmonary stretch receptors of the lungs are activated during inspiration. The purpose of these receptors is to modulate the intercostal muscles and diaphragm so as to avoid lung over-inflation; however, these receptors also send inhibitor signals to the cardiac vagal neurones (CVNs) within the medulla oblongata [46]. The CVNs normally send inhibitory signals to the heart to slow the heart rate, but the signals from the pulmonary stretch receptors inhibit CVN signalling, which increases the HR during inspiration [57].

It has been hypothesised that the function of RSA is to improve the efficiency of gas exchange. This is because a brief increase in heart rate during inspiration should lead to an increase in the instantaneous blood volume which is ideal when the lungs are maximally inflated and allows for maximal gas exchange at the alveoli [46]. This hypothesis is supported by a study that found that by artificially inducing RSA, O_2 uptake was increased by 4% compared to when RSA was not present [46].

One well-studied phenomenon of RSA is that its prevalence and magnitude is known

to diminish due to a number of factors including: poor cardiovascular function and disease, old age (RSA is prominent in infancy but declines thereafter), and poor physical fitness (athletes and those who routinely exercise have higher prevalence of RSA than those who do not) [46].

In addition to RSA, the sympathetic nervous system is also thought to modulate venous return to the thoracic cavity during inspiration through vasoconstriction of the veins in the periphery. This effect leads to reduced blood volumes in the systemic veins during inspiration as compared to exhalation [63].

2.2 Monitoring the Cardiovascular and Respiratory Systems

The ECG and PPG are two common techniques used for monitoring patient status in the hospital [64]. While both technologies have multiple uses, one of the most common is obtaining the cardiac waveform. The ECG contains more specific information on general cardiac function than the PPG [65]. In addition to the ECG and PPG, IP and capnography can be used to obtain respiratory status [26, 28]. As previously described, both IP and capnography are rarely used outside of acute hospital settings. They are nonetheless considered the "gold standard" in respiratory monitoring and serve to provide the reference RR throughout this thesis and therefore will be briefly described. Additionally, a number of respiratory modulations of the cardiac signal can be observed in both the ECG and PPG waveforms [45, 66]. The methods for measuring the ECG and PPG as well as these respiratory modulations will be described in detail.

2.2.1 Impedance Plethysmography

IP can be used to measure physiologic volume changes through changing electrical impedance [67]. IP can be used to accurately estimate blood volume changes based on Swanson's derivation, but since the lungs lack a consistent and regular shape, the precise lung volume cannot be calculated using IP [67]. However, given that the chest cavity changes in volume with every inspiration, ventilation produces a measurable change in impedance through the chest cavity which correlates to the respiratory waveform [67]. Respiratory IP is most commonly measured by passing a high frequency current (often 100 kHz) across the chest using two electrodes on either side of the chest and measuring the varying impedance change between the two electrodes [67]. Specifically, as the resistance of the chest cavity increases due to the increased volume of air in the chest, the impedance of the chest also increases [67].

2.2.2 Capnography

Capnography can be used to measure end-tidal CO_2 ($ETCO_2$), SpO_2, RR, and HR [68]. Capnography specifically measures the partial pressure of CO_2 in expired air using infrared radiation [69]. Clinical capnography relies on either a nasal cannula or mask which diverts and measures the partial pressure of CO_2 in a small amount of expired air. The partial pressure is detected using an infrared sensor which emits infrared light and measures how much is absorbed [69]. This technique can be used to measure a number of gases, including CO_2 and N_2O. In general capnography, a wavelength of 4.28 μm is used, as this wavelength is specifically absorbed by CO_2 [69]. By measuring the change in absorption over time, the respiratory waveform is obtained and RR can be calculated [69].

2.2.3 The PPG

The PPG waveform was first described by Hertzman in 1937. The PPG quantifies the changes in blood volume in a particular location of the body by using the changes in light absorption due to variations in blood flow [70]. The original device used for obtaining the PPG was the photoelectric plethysmograph; however, advances in the measurement methods led to the development of the pulse oximeter which was first described by Aoyagi *et al.* (1974) and became commercially available in 1983 [70, 71, 72]. Commercially available pulse oximeters are widely used in clinical settings. While the PPG signal itself is rarely clinically used, the waveform is the basis from which both the SpO_2 and HR are derived [72]. More recently, interest in using the PPG to derive vital signs has been renewed as it has been shown that the PPG can be obtained using CCD and CMOS cameras via non-contact imaging PPG (iPPG) [72, 73, 74].

The standard, clinical pulse oximeter device consists of a light source and a photodetector. The light source is most commonly a red (i.e., 660 nm) or infra-red (i.e., 880-940 nm) light-emitting diode (LED) [75, 76]. Other wavelengths of light, such as green (i.e., 525 nm), have been successfully used, and have even shown improved performance at some tasks such as monitoring heart rate [77]. Most commercial pulse oximeters monitor two wavelengths of light, because the magnitude of the PPGs of the two different types of light can be used to estimate SpO_2 [78, 79]. The other key feature of the pulse oximeter is the photodetector. The photodetector can either be placed next to the light source or on the opposite side of the site that is being monitored [80, 81]. The first pulse oximeter orientation is termed "reflectance" pulse oximetery and works by measuring the back-scattering of the light source from the LED to the photodetector. The benefit

of reflectance pulse oximetry is that the plethysmogram may be measured at almost any body location [40, 80]. The second pulse oximetry orientation is called "transmission" pulse oximetry. Transmission pulse oximetry is by far the more commonly-used configuration and works by collecting the light that travels through the tissue to the photodetector [78]. One of the challenges of transmission pulse oximetry is the limited number of sites at which it can be used. The light source and the photodetector must be opposite one another and any site that is too thick or dense will not have any measurable light transmission. The technique is therefore limited to use with body locations where the required light transmission length is quite small. These sites include the fingers, toes, earlobes, and nasal septum [40, 80, 82].

The PPG signal is acquired from the pulse oximeter based on the varying absorption of light at the monitoring site due to volumetric and physical changes. The amount of light absorbed is dependent on the composition of the tissue through which the light source is being passed, which includes: skin, bone, cartilage, adipose tissue, and blood [78]. The total light absorption of these tissues can be measured using the Beer-Lambert Law [76]:

$$A_{total} = \sum_{i}^{N} E_i C_i L_i \tag{2.1}$$

where A_{total} is the total amount of light absorbed, E_i is the tissue-specific absorption coefficient at a specified wavelength of light, C_i is the density of the tissue, L_i is the path length, and i represents the N different types of tissue in the area being measured [76]. The photodetector measures the inverted value of A_{total} (only light that is not absorbed by the body is observed by the photodetector). The PPG waveform is the inverted value of the light intensity measured by the photodetector, A_{total}, through time (thus during

periods of higher light absorption, the amplitude of the PPG waveform is higher) [79].

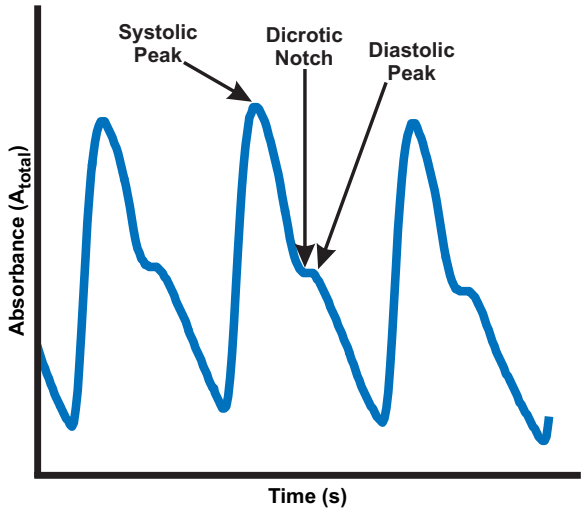


Figure 2.1: *Idealised PPG waveform. The anacrotic phase occurs during systole and begins at the waveform trough and continues to the systolic peak. The catacrotic phase occurs during diastole and starts at the systolic peak and ends at the waveform trough. Both the dicrotic notch and diastolic peak occur during the catacrotic phase [83].*

Figure 2.1 shows an idealised PPG waveform where the amplitude of the waveform is the value of A_{total} [84]. The PPG waveform is a measure of changes in blood vessel volume at the monitoring site [83]. Despite similar morphologies, the PPG waveform is distinct from the arterial pressure pulse waveform. The relationship of these two waveforms has been quantified by Millasseau *et al.* 2000 [85]. The PPG waveform consists of two phases, the anacrotic and catacrotic phases [83]. The anacrotic phase is the systolic phase and starts at the PPG trough and ends at the PPG peak (known as the systolic peak). The anacrotic phase is the diastolic phase which starts at the peak and continues to the trough. The two morphological features of the anacrotic phase are the diastolic peak and the dicrotic notch. In the PPG, the dicrotic notch has been observed to measure the reflective pressure from the periphery (as opposed to the closing of the aortic valve as seen in the arterial pressure pulse waveform) [76, 83]. The location of the dicrotic notch has been used as a measure of vasomotor tone [76].

Equation 2.1 represents a linear model that explains the light absorption by the body. Since the PPG is used to monitor the dynamically changing features of absorption, the equation can be simplified:

$$A_{total} = (ECL)_{arterial} + (ECL)_{venous} + (ECL)_{other} \tag{2.2}$$

where the only dynamically-changing values are $(ECL)_{arterial}$ and $(ECL)_{venous}$, which represent the arterial and venous blood flow, respectively [78]. $(ECL)_{other}$ represents the absorption of all other tissues. While $(ECL)_{other}$ varies based both on monitoring location, PPG method, and from patient to patient, it is assumed to be constant for each monitoring location and can be filtered out by removing all components of A_{total} that are not cardiosynchronous [79, 86]. As a result, the PPG waveform is not an absolute measure, but rather a measure of the dynamic changes in the light absorption and as such the waveform is unit-less [86].

The PPG waveform is often even further simplified and understood as consisting of an AC and DC component where the AC component consists of the absorption due to $(ECL)_{arterial}$ and the DC component consists of the absorption due to both $(ECL)_{venous}$ and $(ECL)_{other}$ [79]. In this representation, the AC component of the PPG is the more dynamically fluctuating signal as it represents the arterial pulsations most closely associated with the heartbeat. The fluctuations in the DC component represent the more slowly oscillating blood flows in the venous return along with any other oscillations that might be observed in the body such as vasomotor activity, vasoconstrictor waves, thermoregulation, and Mayer waves [40, 79, 82].

One further complication of acquiring the PPG is the selection of the wavelength of

light to be used. The light absorption pattern of blood varies based on the wavelength of light and the oxygen state of the haemoglobin protein. The haemoglobin proteins in the red blood cells are present in the body in one of two states: (deoxygenated) haemoglobin (Hb) and oxyhaemoglobin (HbO_2) [67, 78, 87]. In the range of 500 nm to 1000 nm, Hb and HbO_2 have significantly different absorption patterns but do have isosbestic points at ∼800 nm and at multiple wavelengths below 600 nm (Figure 2.2) [87, 88, 89]. The isosbestic points are important because they serve as references where light absorption is independent of oxygen state of the haemoglobin protein [89].

In measuring the PPG, often the selection of the wavelength of light is purpose-dependent. For some applications, particularly measuring SpO_2, the optimal wavelength is between 600 nm and 1000 nm. This is because the attenuation of the light due to body tissues is reduced in this range [87]. The other critical feature of measuring SpO_2 is that it requires measurement of light absorption at two different wavelengths in order to differentiate absorption due to Hb and absorption due to HbO_2. In early oximeters, the isosbestic point was often used for one of the measurements; however, this is not necessary and present technologies often rely on one wavelength in the red spectrum (i.e., 660 nm) and one in the infrared spectrum (i.e., 910 nm) [78, 87, 89]. Use of the PPG to measure other biological activity, such as the HR, does not require such careful wavelength selection. Wavelengths in the blue, green, red, and infrared spectra have all been used to measure HR [77, 90]. Wavelengths in the red and infrared spectrum are most often used for measuring HR in hospital settings [40]. However, green light is often preferred in monitoring HR during daily activities as it is less susceptible to motion artefact when measuring reflectance PPG. It is hypothesised that this is because the higher light wavelengths penetrate deeper into the tissue and measure pulsations in the

larger arterioles, which are more prone to motion artefact. The lower wavelength lights do not penetrate as deeply and measures pulsations in the smaller arterioles, which are less prone to motion artefact [90].

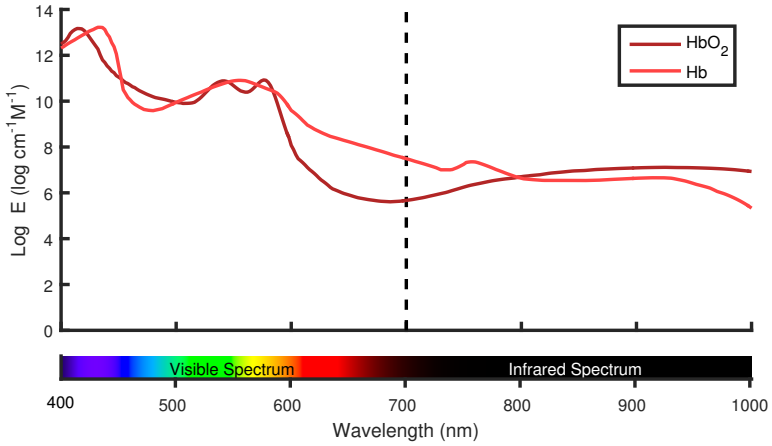


Figure 2.2: *Absorption coefficient, E (log scale) vs. wavelength of light for Hb and HbO_2 in water for the visible (400 nm - 700 nm) and infrared (700 nm - 1000 nm) spectra. Isosbestic points can be seen between below 600 nm and at ∼800 nm [88].*

One of the biggest challenges in obtaining the PPG signal is poor signal quality. The two biggest causes of poor signal quality are motion artefact [78, 91, 92, 93, 94] and ambient light entering the photodetector [78, 95]. In addition to these, additional sources of error include [78, 80]:

- thermal injuries (which affect blood perfusion)
- poor sensor placement (placement over a large artery results in poor signal quality and placement on the thoracic cavity results in multiple periodic artefacts due to the physical motion of the chest cavity by the heartbeat and respiration)
- poor sensor attachment (which can result in either the sensor being too tight, which reduces blood perfusion at the monitoring site, or too loose which introduces both random motion artefact and ambient light noise)
- dark skin pigmentation
- nail polish at the monitoring site

In order to address these challenges, there has been great interest in signal processing techniques to improve the quality of PPG signals, especially the development of SQIs capable of estimating when a PPG signal is artefactual [96, 97].

Despite these challenges, and because of its relative ease of acquisition, the PPG is widely collected throughout the hospital. While monitoring SpO_2 remains the principal use of the PPG in most settings, interest in the cardiac waveform obtained from the PPG has resulted in a number of novel uses for the PPG [76]. These uses include: measuring vascular tone due to attenuation of the sympathetic nervous system during general anaesthesia; measuring blood pressure in conjunction with a blood pressure cuff by observing when arterial pulsations return in the PPG (a technique which might be useful in noisy environments where use of a stethoscope is not possible); measurement of regional tissue perfusion; monitoring and observing cardiac arrhythmias; measuring vascular elasticity through the rate of change of the arterial pulsations; and using the features of the dicrotic notch to monitor vasomotor tone [76]. There has also been growth in using the cardiac waveform of the PPG to extract the respiratory waveform due to the physiological modulations of the cardiac signal by respiration that have been described [44].

Respiratory Modulations Present in the PPG

No fewer than six general approaches for measuring respiration from the cardiac waveform of the PPG have been described. The most prevalent in the literature are those associated with amplitude modulation (AM), frequency modulation (FM), and baseline wander (BW) [45]. These respiratory modulations occur as a result of the interconnections of the cardiovascular and respiratory systems described previously.

Amplitude Modulation The AM of the PPG signal is the difference in the amplitude of the PPG signal at the start (trough) and peak of a pulsation (Figure 2.3). This respiratory modulation is a result of the reduced cardiac output during inspiration compared to during exhalation [45, 66]. The peak-to-trough distance is smaller during inhalation than exhalation. Depending on the location of the PPG sensor, there can be a delay between the actual respiration and the observed respiratory signal at the periphery due to the pulse transit time (PTT) [98].

Frequency Modulation The FM of the PPG is the result of the variations in the instantaneous heart rate due to RSA. This can be measured as the length of time between any known PPG feature (peaks, troughs, or the dicrotic notch, among others). It is most commonly measured based on the peaks of the PPG [45, 57, 66] (Figure 2.3). As described, RSA occurs due to autonomic nervous system modulation of the heart rate due to both pulmonary stretch receptors and arterial baroreceptors.

Baseline Wander The BW of the PPG, which is also referred to as the respiratory induced intensity variation, is the result of the light absorption due to all tissue components except the arterial pulsations. This modulation is therefore not a direct measure of the heartbeat, but the result of observing the changes in venous return during respiration due to changes in intrathoracic pressure. It is hypothesised that sympathetic nervous system mechanisms causing venous vasoconstriction during inspiration exacerbate this effect [63, 99]. The BW is measured as the variation in the waveform between a set of features. The respiratory waveform associated with the baseline wander has been measured using the upper envelope (peaks) [49], the lower envelope (troughs) [100], and the mean envelope (average of peaks and troughs) [37, 101].

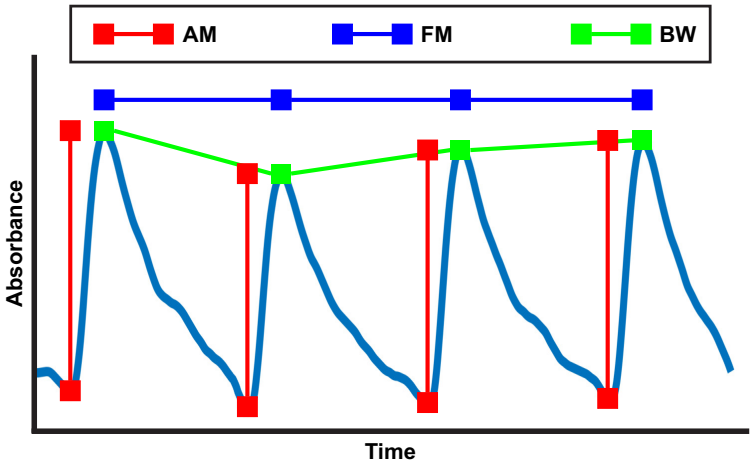


Figure 2.3: *Common respiratory modulations present in the PPG including the AM (red), FM (blue), and BW (green). The BW is only shown as the upper envelope for clarity; however, it can also be measured as the lower and mean envelopes.*

The respiratory modulations of the cardiac signal of the PPG are dependent on large variations in pulse size which are dependent on the placement of the pulse oximeter. The respiratory modulations of the PPG can be obtained from the finger, wrist, forearm, shoulder, and forehead. The PPG signal obtained at the forearm provides the best respiratory waveform while the finger provides the worst respiratory modulation [102]. In addition, the finger provides the best cardiac waveform while the forearm provides the worst cardiac waveform [102]. Additionally, the finger is a challenging location from which to extract the respiratory modulation from the PPG because the prevalence of the respiratory modulations is reduced due to arterial elasticity (the variations in the arterial pulsations are minimised by the time they are observed in the finger). These variations are reduced further when the arm is elevated above the heart [82, 103]. Despite this, the vast majority of algorithms for extracting RR from the PPG have been developed using PPG signals obtained from finger probes [78].

2.2.4 The ECG

The contractions of the heart occur as a result of a number of highly correlated electrical events throughout the various tissues of the heart [65]. These electrical events result in an instantaneously changing electrical potential within the thoracic cavity, the observation of which is fairly uniform from heartbeat to heartbeat [65]. This electrical profile can be observed by measuring the electrical potential using electrodes placed on the surface of the thoracic cavity or in the extremities, however, the magnitude of the potential decreases the further the electrodes are moved away from the heart [65]. The method by which the varying electrical potentials in the thoracic cavity can be used to monitor cardiac function was first defined in 1887 using a mercury capillary electrometer [104]. This concept was termed electrocardiography and was further improved with the introduction of the string galvanometer [105].

Over the next 50 years, ECG technology continued to improve using a standard set of four electrodes. These electrodes are placed on the four extremities: left arm (LA), right arm (RA), left leg (LL), and right leg (RL). In this configuration, all of the potentials between the LA, RA, and LL are recorded while the RL electrode is used as ground [65]. The potential between any two electrodes is termed a lead where the standard leads from these electrodes are as follows [65]:

- **Lead I:** LA to RA
- **Lead II:** LL to RA
- **Lead III:** LL to LA

Based on these three leads, one of the most significant advances was the description of "Einthoven's triangle", a mathematical explanation of the relationship of the potentials

recorded at each one of these leads. The leads are assumed to make an equilateral triangle where the potential of each lead is a vector that is the scalar product of the actual cardiac potential vector and the unit vector in the direction of the lead. Thus, each lead represents the magnitude of the cardiac vector in the direction of that lead [106]. From this, the actual magnitude and direction of the cardiac potential vector can be described using any two sets of leads [65]. This is important because, once the actual cardiac potential vector is obtained, it is possible to observe a number of cardiac abnormalities based on the orientation of the cardiac potential vector over the course of a cardiac cycle [65].

The most commonly employed ECG configuration is the 12-lead ECG, which is an expansion of the 3-lead system already defined. The additional leads are unipolar leads which measure the electrical potential from a single electrode to Wilson's Central Terminal. Wilson's Central Terminal is the sum of the electrical potentials from Leads I, II, and III and is assumed to be a zero potential terminal at the centre of the heart [65]. The nine additional leads include three unipolar extremity leads and six precordial leads [65, 107]. The extremity leads (aVL, aVR, and aVF) are the potentials from each of the three non-ground extremities (LA, RA, LL) to a modified version of Wilson's Central Terminal which does not include the potential of the electrode currently in consideration (i.e., aVL is the LA electrode to Wilson's Central Terminal where in this instance Wilson's Central Terminal only includes the RA and LL electrodes) [65, 107]. The six precordial leads are measured based on the potential between six chest electrodes and Wilson's Central Terminal where the six chest electrodes are placed at predefined anatomical chest locations [107]. Each of these leads is termed V1-V6 moving from left to right across the chest [107].

The benefit of the 12-lead ECG is that the full complement of leads can be used to observe the specific electrical potential in different parts of the heart; however, the basic features of the ECG waveform are the same across all leads for healthy subjects [65]. The ECG waveform consists of the electrical polarisations in the heart that are the precursors to the contractile events that make up the heartbeat. These electrical events are divided into two main components which are related to their physical counterparts but slightly time-shifted: electrical systole, the cardiac contractile phase, and electrical diastole, the cardiac filling phase [65].

Electrical systole begins with activation (depolarisation) of the sinoatrial (SA) node in the upper wall of the right atrium [108]. The depolarisation of the SA node does not cause a large deflection in the electrical potential in the thoracic cavity to be observed in the ECG, but after SA node activation the wave of depolarisation flows from the SA node to the atria and finally to the ventricles [65, 108]. The atrial depolarisation is the first event in the ECG that can be observed and is termed the P-wave (Figure 2.4) [67, 108]. Physiologically, it is essential that the P-wave occurs prior to ventricular activation as it ensures that the blood in the atria is pumped into the ventricles prior to ventricular contraction. This delay, termed the P-R segment, is the result of a conduction delay in the atrioventricular (AV) node, which is responsible for conducting the electrical impulse from the atrium to the ventricles (Figure 2.4) [65]. After the P-R segment, ventricular contraction begins. Ventricular activation consists of three events, the Q-, R-, and S-waves, which are together are termed the QRS complex (Figure 2.4) [65]. The QRS complex can be broken down into three electrical vectors. The first vector occurs during the Q- and R-waves (between 0-15 ms of the entire QRS complex) and consists of septal activation. Septal activation is the outward spread of the electrical

signal from the AV node across the cardiac septum which divides the left and right ventricles. The second vector of the QRS complex (between 20-60 ms) is the activation of the left and right ventricular myocardium. After 30 ms, all of the separately-activated regions have converged, leading to a closed activation surface which leads to the third activation vector of the QRS complex—the latter corresponds to the activation of the basal right and left ventricles and septum and proceeds in an epicardial (outside the heart) direction [65, 108]. The S-wave exhibits a downward deflection in the ECG which represents an overshoot in the depolarisation followed by a rapid repolarisation at the end of the ventricular activation [65]. After the initial, rapid repolarisation, a depolarization plateau is maintained for 200 to 300 ms, known as the S-T segment, after which the T-wave can be observed (Figure 2.4) [65]. The T-wave is the primary ventricular repolarisation event (Figure 2.4) [108]. The atrial repolarisation cannot be observed as it occurs concurrently with the QRS complex, which has a much greater magnitude [65]. Finally, a U-wave is occasionally observed. The U-wave is a result of the repolarisation of the ventricular papillary muscles [65]. The time between the end of the T-wave and the next P-wave is known as electrical diastole [65].

The primary use of the ECG is to observe the presence of cardiac conditions including: bradycardia (slowed HR), tachycardia (increased HR), arrhythmia (early heartbeats due to premature activation of the AV node or myocardium), pre-excitation (syncope or palpitations), scarring from previous myocardial infarction, abnormal electrical conduction through the atria or ventricles, atrial or ventricular enlargement, myocardial stretching, ischemia, electrolyte imbalance, hypothermia, and congenital ionic channel diseases [65, 108, 109]. Numerous protocols exist to observe these conditions using ECG, even when they are rarely or minimally present. They have been designed to induce car-

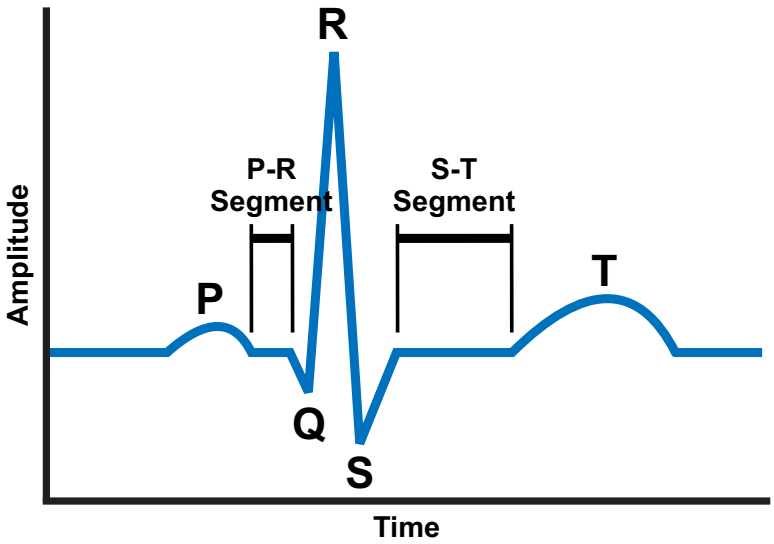


Figure 2.4: *Common waves and features of a single heartbeat in the standard ECG including: P-wave (atrial activation), P-R Segment, Q-, R-, and S-waves (QRS complex, ventricular activation), S-T segment, and T-wave (ventricular repolarisation) [65, 108].*

diac stress in controlled settings and include: the tilt-table test, carotid sinus massage, and chemically- or physically-induced cardiac stress testing [108]. The ECG can also observe sporadic cardiac events that occur in ambulatory patients outside a hospital setting using devices such as Holter monitors [108].

Respiratory Modulations Present in the ECG

More recently, much interest has arisen in deriving the RR from the ECG, via respiratory modulations of the cardiac waveform. Despite recent interest, the presence of respiratory-related modulations in the ECG is not a recent discovery and the respiratory modulation of the ECG due to cardiac rotation was first described as early as 1950 [106]. The first algorithm for estimating RR from the ECG based on cardiac rotation was described by Moody *et al.* (1985) and was based on using two orthogonal ECG chest leads to obtain the angle of the mean cardiac electrical axis through time. The variations of the angle of the mean cardiac electrical axis through time were shown to correlate to the RR [36].

Further work on obtaining respiratory modulations from the ECG has focused on single-lead ECGs. Similarly to the PPG, three respiratory modulations of the ECG can be observed based on the AM, FM, and BW [110]. The following is a physiological explanation for the presence of these modulations:

Amplitude Modulation and Baseline Wander It is well established that both amplitude deflection and baseline wander are present in the ECG due to respiration [106]. Amplitude deflection is caused by cardiac rotation within the thoracic cavity which changes the magnitude of the ECG projected onto each lead [106] (Figure 2.5). The initial exploration of this phenomenon made three important observations: (1) algorithms based on the changes in the R-wave amplitude were preferred over those based on changes in the T-wave amplitude as the R-wave is more pronounced; (2) compared to Leads I and III, the deflection of Lead II was minimised; and (3) it was observed that the deflection was hardly noticeable in many individuals [106]. In addition to the effects of cardiac rotation on amplitude modulation and baseline wander, the motion of the electrodes with respect to the heart due to the expansion and contraction of the thoracic cavity during respiration also factors into the observed AM and BW signals [36, 111]. Further, despite previous findings, a number of studies have had success at deriving the amplitude-based respiratory modulation from the ECG using Lead II [112, 113].

Frequency Modulation The frequency modulation of the ECG is believed to be derived from RSA, the same mechanism that causes frequency modulation in the PPG [114]. In contrast to the AM and BW of the ECG, previous work has shown that the RSA feature of the ECG is most prevalent in Lead II of the ECG [115] (Figure 2.5).

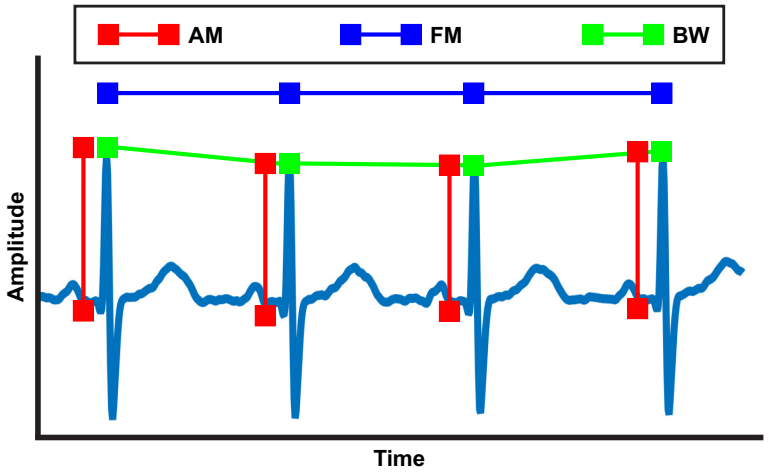


Figure 2.5: *Common respiratory modulations present in the ECG: AM, FM, and BW.*

The respiratory modulations of the ECG are based on the general features of the ECG, and so it is possible to derive the previously-described modulations from any single lead ECG [112, 113, 116, 117, 118]. In fact, for many publicly-available data sets in which a single ECG waveform is available and in which this has been used to derive RR, the actual ECG lead was not stated; however, it can generally be assumed that it is one of the bipolar leads (Leads I-III) and often Lead II, as this lead most directly aligns with the actual cardiac electrical axis [119, 120, 121, 122].

2.3 Conclusion

The present chapter describes the physiological mechanisms, both via physical phenomenon and nervous system control, by which the respiratory system influences the cardiovascular system. Furthermore, this chapter describes the process by which the PPG and ECG, the two primary waveforms used in this thesis are obtained and the challenges presented with the acquisition of these signals. Finally, based on the described physiology, the means by which the respiratory system modulations the signals acquired by the PPG and ECG are described. The following chapter presents a review

of the signal processing methods that have been previously described for accurately extracting the respiratory waveform and RR from both the PPG and ECG.

Chapter 3

Literature Review

In the previous chapter, the physiological mechanisms by which respiration modulates the ECG and PPG were described. Furthermore, the manifestations of these mechanisms in the ECG and PPG signals—amplitude modulation (AM), frequency modulation (FM), and baseline wander (BW)—were described. Despite the simple mechanism of observing the respiration modulations in the ECG and PPG, the robust extraction of these modulations has proven far more complicated. The first extraction methods were described in the 1980s. Since then, a myriad of signal processing and machine learning methods, ranging from simplistic to highly complex, have been described. The goal of these developments is to create methods that are accurate and robust enough for clinical implementation.

To date, the majority of methods by which the respiratory waveform and RR have been extracted from the ECG and PPG can be classified as three separate modules: (I) extraction of respiratory waveform (Section 3.2), (II) extraction of RR (Section 3.3), and (III) fusion of RR estimates from multiple modulations (Section 3.4, Figure 3.1). In addition to this, two signal quality checkpoints exist: Checkpoint I before Module I using SQIs to assess the quality of the original PPG or ECG input (Section 3.5)

and Checkpoint II before Module II using RQIs to assess the quality of the respiratory waveform prior to estimation of the RR (Section 3.6, Figure 3.1). The focus of this thesis is on Checkpoint II as the current literature contains a wide array of effective methods for each of the three modules and Checkpoint I. However, prior to this work the number of RQI methods (i.e., those algorithms at Checkpoint II) previously described is sparse and those methods that exist lack both effectiveness and robustness.

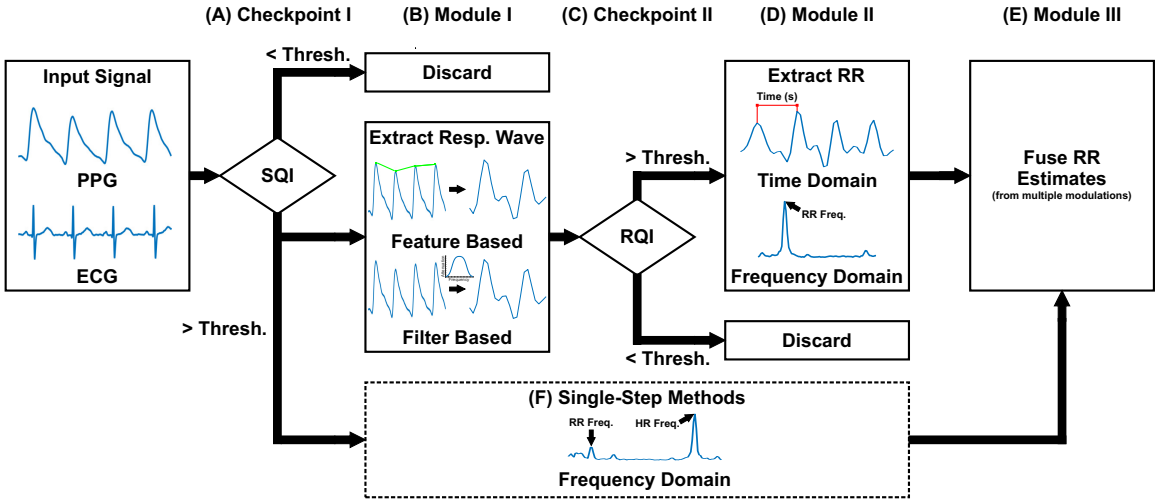


Figure 3.1: *The proposed respiratory waveform and RR extraction framework in which all previous methods for estimating RR from ECG and PPG can be categorised. The framework consists of three modules: (B) the respiratory waveform extraction module, (D) the RR extraction module, and (E) the RR fusion module. The framework also consists of two checkpoints: (A) the SQI checkpoint and (C) the RQI checkpoint. Finally, the framework also accounts for the small class of methods in which the RR is extracted directly from the ECG or PPG in the (F) single-step methods.*

While much of the previous literature on extracting the RR from ECG and PPG does not specifically consider the framework in Figure 3.1, the majority of the literature fits within this framework. The methods described often fall into one specific module or, in some cases, multiple methods are described which pertain to separate modules. In this chapter, each section will present the previous work and methods that have been used for a specific module or checkpoint. In a small number of cases, the methods described cannot be categorised into a single distinct module, and these instances will be presented

in an additional section on single-step methods at the end of the chapter (i.e., box F in Figure 3.1).

3.1 Previous Work

A search for previous publications that present novel methods for extracting either the respiratory waveform or RR found 124 unique publications. While the selection of these works does not represent an exhaustive systematic selection, the publications represent the subset of publications from those found by Charlton *et al.* (2017) which describes novel methods for extracting the respiratory waveform (or RR) from ECG and PPG signals [39]. The original set of 196 publications described by Charlton *et al.* (2017) was found through database searches in Google Scholar, IEEE Xplore, PubMed, Science Direct, and Scopus using a wide-array of search terms for publications related to respiration and the ECG or PPG. In addition to this, publications not found in the databases, but familiar to the authors of that work, were added [39]. The 124 publications analysed here represent a subset of the original 196 publications in which only publications describing a specific respiratory waveform or RR estimation method were considered.

Of these 124 publications, 51 (41.1%) considered only ECG, 63 (50.8%) considered only PPG, and 10 (8.1%) considered both ECG and PPG. Additionally, of the three possible ECG and PPG modulations of respiration previously described, AM was considered in 30 (24.2%) publications, FM in 47 (37.9%), and BW in 97 (78.2%). The largest number of publications considered methods to extract respiration or RR from the BW of the PPG, corresponding to 63 (50.8%) publications in total. In addition to this, only 9 (7.3%) publications looked at methods that extracted respiration or RR from AM, FM, and BW. Finally, 22 (17.7%) publications looked at methods other than

AM, FM, BW for extracting respiration or RR. These extraction methods most commonly included: PTT [98, 123, 124, 125], QRS area [33, 126, 127, 128, 129, 130], EMG [131, 132], and Pulse Width Variability (PWV) [123, 125, 126, 133, 134, 135].

Based on the framework described, 108 (87.1%) publications included a method related to Module I, the extraction of the respiratory waveform, 77 (62.1%) included a method related to Module II, the extraction of the RR from the respiratory waveform, 21 (16.9%) included a method for Module III, fusing multiple respiratory waveforms or RR, 2 (1.6%) employed an SQI prior to Module I, and 4 (3.2%) employed an RQI prior to Module II. Additionally, 16 (12.9%) publications used a single-step method in which the method simultaneously extracted the respiratory waveform and estimated the RR.

3.2 Module I: Waveform Extraction

The first step in the framework shown in Figure 3.1 is the extraction of the respiratory waveform from either or both the ECG and PPG, based on at least one of the modulations previously described. Of the 108 publications in which a stand-alone respiratory waveform was extracted, all of the methods by which this has been done fall into two specific classes: (1) feature-based and (2) frequency-based. Of the waveform extraction methods, feature-based methods were the most common with 71 (65.7%) publications using a feature-based method and 33 (28.7%) publications using a frequency-based method. Additionally, 6 (5.5%) publications used both feature- and frequency-based methods.

3.2.1 Feature-Based Methods

The general method for the feature-based respiration extraction methods is: (1) ECG or PPG pre-processing, (2) feature extraction, and (3) post-processing analysis.

The first step, pre-processing, generally includes either or both signal down-sampling and digital filtering at the extrema; however, the specific down-sampling frequencies and filter types and cut-off frequencies vary widely between studies.

The second step is feature-based extraction. Generally, this is an algorithm that annotates specific features of the original ECG or PPG time-series. Most often the feature-based extraction method begins with a pre-defined peak detection algorithm. The most commonly employed peak detectors were the Pan-Tompkins algorithm [117, 136, 137, 138, 139, 140, 141], the Hamilton-Tompkins algorithm [33, 142], and wavelet-based QRS detectors [125, 127, 143] for ECG and a peak-to-peak detector from arterial pulsations [81] for PPG; however, the majority of publications did not specifically state the peak detector that was used. In addition to this, a number of authors either manually annotated or corrected peaks [128, 129, 144, 145].

For both FM and BW respiration extraction methods, peak detection is sufficient for extraction of the modulation from the ECG or PPG. However, for AM and some of the other modulations previously described, further signal features, such as troughs or mid-points are required. A number of authors used windows with a pre-set size centred around the detected peaks to search for additional features [45, 126, 132, 146] while occasionally authors described unique algorithms capable of detecting other necessary features in addition to peaks [147]. Further, some authors employed feature correction algorithms [130].

The third and final step in the feature-based extraction of the respiratory waveform is post-processing analysis. Currently in the literature, there are three distinct methods by which this has been done: (1) time-series re-sampling, (2) empirical mode decomposition (EMD), and (3) principal component analysis (PCA).

(1) Time-Series Re-sampling. Time-series re-sampling is the most common method for extracting the respiratory waveform. The extraction of the respiratory-modulated feature and feature timestamp creates an unevenly sampled time-series, which is often not useful for further analysis. Thus, the majority of authors evenly re-sample the time-series and bandpass filter the waveform within a pre-determined respiratory waveform frequency range in order to extract a high-quality respiratory waveform. While not all authors explicitly state their time-series re-sampling method, the most common reported re-sampling frequency is 4 Hz [97, 113, 123, 134, 144, 148, 149, 150, 151] and the most common re-sampling techniques are linear [110, 114, 143, 151, 152], cubic spline [118, 123, 127, 130, 132, 137, 145, 146, 153, 154, 155, 156], and polynomial interpolation [157].

(2) Empirical Mode Decomposition (EMD). EMD is a time-series analysis method that was first proposed by Huang *et al.* (1998) to observe non-linear and non-stationary properties [158]. The method allows the ECG and PPG signals to be decomposed into the intrinsic oscillations, one of which is assumed to be the respiratory oscillation within the signal [138]. EMD extracts a number of Intrinsic Mode Functions (IMFs) from the original waveform in an iterative manner. This is done by identifying the waveform extrema (both minimum and maximum values), finding an upper and lower envelope, finding the mean of the two envelopes, and extracting the waveform created from this mean as the first IMF. This IMF is subtracted from the original signal and the same process is repeated for all subsequent IMFs for as many iterations as necessary [138].

A handful of variations on the EMD have been used to extract the respiratory waveform from the time-series of the ECG and PPG. The original work presenting the use of

EMD for respiration waveform extraction by Campolo *et al.* (2011) selected the respiratory IMF by inspection [138]. Madhav *et al.* (2011) found the frequency spectrum of all IMFs and added any IMFs that had a dominant frequency within a pre-determined RR frequency range [159]. Fusco *et al.* (2015) found that the first and second dominant frequencies in the PPG signal were the HR and RR, respectively, and thus assumed the second IMF represented the respiratory waveform [101]. Kumarand and Singh (2015) applied empirical EMD and canonical correlation analysis (CCA) in which a series of EMDs were conducted after adding white noise to the original signal. This was used to determine which series of IMFs were the true signal and not white noise (including the white noise present in the original signal) [160]. These IMFs were summed to create a cleaned ECG and the respiratory waveform was taken as the upper envelope of this signal.

(3) Principal Component Analysis. PCA is a method that is most often used to reduce the dimensionality of large, multi-dimensional data sets. This is done by finding a set of orthogonal vectors where each subsequent vector explains as much of the remaining variation in the data as possible. This results in a set of eigenvectors and an associated set of eigenvalues which are proportional to the variance accounted for by its associated eigenvector [37]. It is often found that the majority of the variance in a data set can be explained by only the first few eigenvectors [37].

This technique was originally applied to the extraction of the respiratory waveform of the PPG by Madhav *et al.* (2011) [37]. After locating the peaks within the PPG signal, Singular Value Decomposition (SVD) was used to determine the periodicity around each peak. The length of the periodicity was used to create an $M \times N$ data matrix in which M represented the number of peaks and N was the length of the periodicity (it

was thus a matrix of all the individual PPG peaks). PCA was applied to this matrix and the first three eigenvectors were extracted and used as respiratory waveforms [37]. Further extrapolations of this method have included multi-scale PCA (MSPCA) and kernel PCA (kPCA). The MSPCA method was applied to the series of wavelet coefficients obtained from wavelet decomposition of individual pulse segments and only used wavelet coefficients with eigenvalues above a set threshold to reconstruct the respiratory waveform [161]. The kPCA method was applied by taking the matrix of peaks and transforming them using either a polynomial or radial basis kernel function. PCA was conducted on the transformed matrix and the first eigenvector (or principal component) was re-transformed back to the original domain and taken as the respiratory waveform [139].

3.2.2 Frequency-Based Methods

The frequency-based methods rely on frequency characteristics as opposed to time-series features to extract the respiratory waveform from the ECG and PPG. The most widely used transform for converting a signal from the time-domain to the frequency-domain is the Fourier transform. However, as a standalone method, it does not fit into the current framework as it is not a method for extracting the respiratory waveform from the ECG and PPG but is rather a method for directly extracting the RR from the ECG and PPG. Therefore, FFT-based ECG and PPG analysis techniques will be discussed in Section 3.7 as a single-step method. The frequency-based methods presented in this section will specifically consider methods aimed at extracting the respiratory waveform prior to any extraction of the RR. The four most common methods for doing this are: (1) digital filtering, (2) the wavelet transform, (3) signal demodulation, and (4) signal

down-sampling.

(1) Digital Filtering. Digital filtering is by far the most commonly-used frequency-based method for extracting the respiratory waveform. In a typical ECG or PPG, the cardiac rhythm is the dominant frequency in the signal. By filtering out the expected cardiac rhythm, the secondary frequencies in the signal become more prevalent. It is often assumed that the respiratory waveform is the most prevalent secondary frequency. Digital filtering takes advantage of this by filtering out only the range of possible cardiac frequencies thus, in the most ideal circumstances, leaving a respiratory waveform free from cardiac artefacts.

While the basic technique for extracting the respiratory waveform from the ECG and PPG using digital filters is well-defined, there are at least three key filter selection criteria: (1) length of analysis window and windowing function, (2) filter type, and (3) filter cut-off frequencies. The length of time in which the original ECG or PPG is segmented prior to filter application is variable between studies, but the range is most commonly between 30 seconds [162, 163] and 2 minutes [75, 164]. Additionally, some authors choose to apply a windowing function prior to filtering with the most common choice being the Kaiser window [164, 165]. The range of filter types that have been used varies widely with Chebyshev [166], Bessel [162, 167], and Butterworth [168, 169, 170] filters being most common. The final aspect is the cut-off frequencies. Many authors have used low-pass filters, only focusing on removing high frequency oscillations [163, 165, 166] while others have used bandpass filters where the low frequency cut-offs are between 0.10 to 0.15 Hz [169, 170]. The high frequency cut-offs observed range from 0.28 to 1.25 Hz [165, 166]; however, the cut-off range must be altered when considering paediatric cases as the respiratory frequency is higher [164]. Additionally, some authors have de-

signed algorithms that adjust the cut-off frequencies of the filter based on the subject's current heart rate as an increase in cardiac frequency is correlated with an increase in respiratory frequency [162, 171]. Recently, some work has focused on more complex filtering techniques, such as homomorphic filtering, which are capable of more accurately de-convolving the frequency components, particularly the HR and RR frequencies, of the ECG and PPG [172].

(2) Wavelet Analysis. While wavelet analysis has more commonly been used to extract the RR from the respiratory waveform, the application of the wavelet transform for extracting the respiratory frequency from the ECG and PPG has been implemented in a number of cases. The first application was in 2002 [112]. The wavelet transform represents an improvement to the more standard spectral techniques because the method allows for the interpretation of the signal in the time and frequency spaces simultaneously [173]. This therefore allows the method to observe small, transient, changes in respiratory frequency [173]. The primary principle of wavelet analysis for extracting the respiratory frequency from the ECG and PPG is that the original signal is iteratively transformed by a wavelet function where the wavelet function is varied based on a dilation parameter (representing changes in the characteristic frequency of the wavelet function) and a location parameter (representing time shifts in the wavelet function). This allows the method to provide both time and frequency information simultaneously [112]. Using a selected respiratory band of interest in the wavelet space, the amplitude-time signal (in addition to the frequency-time signal) can be extracted as the respiratory waveform [174, 175].

The most prevalent method for extracting the respiratory waveform from the ECG and PPG using wavelet analysis was proposed by Addison *et al.* 2003 and uses the

continuous wavelet transform (CWT) with a Morelet wavelet [174]. Furthermore, this method computes two secondary decompositions from the primary decomposition based on the location of the primary respiratory component in what the authors term Secondary Wavelet Feature Decoupling (SWFD). These two secondary decompositions have been defined as the ridge amplitude (RAP) and ridge frequency perturbation (RFP) signals [173, 174, 175, 176, 177]. Additional wavelet analysis methods for extracting the respiratory waveform have used the discrete wavelet transform (DWT) [112, 178] and wavelets other than the more prevalent Morlet wavelet such as the biorthogonal wavelet basis [142].

(3) Demodulation. A small number of authors have used demodulation techniques which treat the respiratory waveform as the carrier wave of the ECG and PPG. The goal is to extract this underlying carrier from these signals. The most prevalent demodulation technique that has been employed is variable-frequency complex demodulation (VFCDM) [179, 180]. The latter consist of five steps. The first is to define a bank of low pass filters with small passbands and slightly shifting centre-frequencies. The second step is to apply these filters to the original ECG or PPG waveform and extract the dominant frequency within each bandwidth using fixed frequency complex demodulation (FFCDM). The third step is to decompose the signal into sinusoidal modulations using complex demodulation (CDM). The fourth step is to calculate the instantaneous frequencies and amplitudes of the sinusoidal modulations. The final step is to obtain the time-frequency representation of the signal using the estimates from the third and fourth steps. This can then be used to extract the respiratory waveform [179, 180]. In addition to VFCDM, Nguyen *et al.* (2011) have employed AM-FM demodulation to extract the respiratory waveform [181].

(4). Signal Down-Sampling. This method has been employed by several authors prior to application of more complex techniques to extract the RR. The method down-samples the signal to a low enough frequency to ensure that the cardiac synchronous pulsatile component of the ECG or PPG is no longer dominant and therefore the underlying respiratory waveform dominates the signal [182, 183].

3.3 Module II: Extraction of Respiratory Rate

The second step in the framework shown in Figure 3.1 is the extraction of the RR. This module is applied specifically to respiratory waveforms. Of the 124 publications analysed, 78 (62.9%) of all publications included an RR extraction technique. Publications that did not include an RR extraction technique either employed correlation statistics with a ground truth respiratory waveform to confirm the accuracy of their respiratory waveform extraction method or employed a single-step method in which Modules I and II could not be separately categorised. Of particular note is that many of the techniques used for RR extraction also have applications to respiratory waveform extraction and the basic details of the method have already been described in Section 3.2. Similar to Module I, there are two categories of RR extraction methods: (1) time-domain methods and (2) frequency-domain methods. Of the 78 publications, time-domain methods for estimating RR are seen in 21 (26.9%) instances and frequency-domain methods in 56 (71.8%) instances, and both time- and frequency-domain methods are seen in one publication (1.3%).

3.3.1 Time-Domain Methods

Similar to the feature-based methods in Module I, the time-domain methods for extracting the RR from the respiratory waveform require the identification of features within the respiratory signal. While the peaks are the most commonly located feature [75, 98, 124, 137, 143, 155, 171, 172, 178, 184, 185, 186], additional methods have located peaks and troughs [118, 145, 187], troughs alone [188], signal zero-crossing points [112, 189], and signal periodicity [190]. The feature finding algorithms are often simpler than those described in Module I and focus on identifying local maxima and minima. A number of authors implemented pre-set threshold and rule-based systems to discard features that do not meet specific criteria. These include selecting only maxima that are above a certain threshold of the average amplitude of the waveform [118, 143, 145] and discarding signals where the peak exceeds a threshold amplitude (as this is assumed to be an underlying noise component and not respiration) [145]. Once the respiration features have been found, there are many ways in which authors have actually determined the RR. The most common is to take a window of data and use either the number of breaths [75, 124] or the average duration of breath intervals in that window [137, 155, 186, 188] to calculate the RR. Most commonly windows of one minute duration have been used [155, 170, 188], but window lengths ranging from 45 seconds [143] to ten minutes [189] have been used. Other authors have varied this technique. Nakajima *et al.* 1996 used the median of a series of breathing intervals to calculate RR [171]. Chan *et al.* 2013 used sliding windows to calculate the RR as the average of the instantaneous RR between successive QRS peaks within windows. As a quality measure, the top and bottom decile of the instantanous estimates were discarded [143]. Finally, unlike the other time-domain

RR estimates, one author Pimentel *et al.* (2013) fitted Gaussian Processes (GPs) to two derived respiratory waveforms using a periodic covariance function which used the periodicity of the fitted function to determine the RR. This represents a unique time-domain RR estimation method because it does not rely on the presence and location of specific features (i.e., peaks, troughs, and zero-crossing points) but rather relies on the overall periodicity of the respiratory waveform. Additionally, the method provides the probability of the accuracy of the RR estimate, a result which is rarely seen in other RR estimation techniques [190].

3.3.2 Frequency-Domain Methods

The frequency domain methods for extracting the RR from the respiratory waveform include many of the same methods used to extract the respiratory waveform in Module I. The goal of the frequency-domain methods in Module II is to extract a single frequency value that can be used as the RR. The frequency-domain methods applied to extract the RR include: (1) the Fourier transform, (2) autoregression (AR), (3) time-frequency methods (including the wavelet, Hilbert, and short-time frequency transforms), and (4) Filtering methods.

(1) Fourier Transform. Fourier-transform based methods are the largest group of methods that have been employed to calculate the RR from respiratory waveforms. Most often authors employ the fast Fourier transform (FFT) which is a method that decomposes the time-based respiratory component into its component parts in the frequency-domain, and where the power spectral density (PSD) corresponding to the FFT represents the power of a given frequency in the signal. In the simplest case, the FFT is used to calculate the PSD of the respiratory waveform and the frequency with the largest

magnitude in the PSD is taken as the RR.

While many authors do not specify the details of their FFT method for estimating RR, only slight variations exist in the implementations that are described. One such variation is the implementation of a windowing function, such as the Hamming window, prior to taking the FFT [66, 144]. Another method [125] used a technique which the authors term "peak conditioning" in which they compute the PSD of sliding, overlapping windows. For each PSD they find all of the points that are at least 0.75 times the magnitude of the largest peak and take the value of the RR to be the point that is closest to the previous point.

(2) Autoregression. AR is a method, similar to the FFT, which can be used to decompose a time-series into its frequency components and provide the relative prevalence of those components in the underlying signal. The AR model is often preferred over the FFT because, unlike the FFT, the frequency estimates provided by the AR model are not discretised [142]. An AR model is formulated as a linear regression problem in which a set number of previous points in a time-series are used to predict the current point. The number of previous points used to formulate the AR model define the model order and are an important design selection criterion. Depending on the model order, the AR model output is either a series of complex-conjugate pole pairs equal to the model order (even order) or a series of complex-conjugate pole pairs equal to $n-1$ of the model order and a model zero (odd order). The magnitude of a pole corresponds to the strength of that pole in the time-series and the pole phase angle corresponds to the pole frequency in the PSD [182]. Therefore, a number of authors have used both the AR poles [182, 191] and the PSD derived from the transfer function of the AR [164, 192] to estimate RR.

The first applications of AR for estimating RR were in 2002 by Meste *et al.* (2002) and Thayer *et al.* (2002) [157, 192]. Since then, authors have implemented a number of AR variations for estimating RR. The two biggest variations in AR modelling are the algorithm used to obtain the AR poles and the order of the model. The two most common AR algorithms used are Burg's algorithm [163, 191] and the Yule-Walker algorithm [142]. The other major selection criteria is the model order. Some authors have used pre-set model orders [142, 182] while others have used ideal model order selection methods such as Akaike's Information Criterion (AIC) [33, 157] and the Optimal Parameter Search (OPS) method [191, 193]. The final major AR selection element has been the determination of the RR frequency. While some authors have focused on methods based on the model poles, others have focused on the PSD. When considering the AR poles, the most common and simplest method is to select the pole with the largest frequency within the RR range of interest [163, 191]. However some authors have designed additional methods such as finding all poles that are at least 0.95 times the magnitude of the largest pole and selecting the pole with the smallest angle (and thus smallest RR) within the RR range of interest [182] or using particle filtering to use prior information to predict the most likely RR pole [193]. For selecting the RR from the PSD, selecting the highest magnitude peak in the PSD is most common [157]; however, one method, termed *ARSpec* involves taking the PSD for all AR model orders from 1 to 20 and finding the median PSD based on the panel of 20 PSDs. The peak of the median PSD is taken as the RR frequency [164].

(3) Time-Frequency Methods. In addition to the wavelet transform, a small number of authors have used variations of time-frequency techniques to extract the RR from the respiratory waveform such as: the Hilbert transform to estimate the instan-

taneous respiratory frequency [127], cross-time frequency analysis [133], and quadratic time-frequency distribution analysis [134].

(4) Filtering. While digital filtering methods are far more common for extracting the respiratory waveform from the ECG or PPG, some authors have developed novel techniques to use filters to extract the RR from the respiratory waveform. Varanini *et al.* (1991) used adaptive filtering on two separate respiratory waveforms to find the filter that minimized the mean-square error (MSE) which meant that the correlated components of the signal were cancelled out [116]. The transfer function of the ideal filter could be used to determine the RR [116]. Mirmohamadsadeghi *et al.* (2015) used a bank of notch filters with small passbands and small centre-frequency steps and found the filter output that minimised the output power of the filtered signal [150]. The centre-frequency of this filter was used as the RR frequency [150].

3.4 Module III: Respiratory Rate Fusion

Due to the presence of multiple respiratory modulations in the ECG and PPG, the final module in extracting the RR from the ECG and PPG is the fusion of the estimates from multiple modules. Compared to Modules I and II, there is a significant reduction in the prevalence of RR fusion methods as most authors simply compare the results from the various respiratory modulations and their relative performance. However, within the last ten years there has been an observable expansion of methods for fusing RR estimates. The methods by which this is most often done are: (1) basic statistical techniques, (2) best-value selection, and (3) machine learning techniques.

(1) Basic Statistical Techniques. A number of statistical methods have been employed to fuse multiple RR estimates. These include the mean, median, and summa-

tion. The most basic technique that has been used is to average multiple RR estimates [149, 150]. Additionally, a number of authors have implemented weighted averaging techniques where the weights are obtained using RQI values, the methods for obtaining which will be described in Section 3.6 [33, 143, 194]. Currently, one of the most prevalent fusion techniques is SmartFusion. The SmartFusion method averages the available RR estimates for small time windows and only accepts the RR as accurate when the standard deviation (SD) amongst the estimates is > 4. In instances where this is not the case, no RR estimate is given [66, 195]. In addition to techniques focused on the mean, some authors have used the median of multiple RR estimates [133]. Finally, the summation of the normalised PSDs for multiple RR estimates has also been used to estimate a fused PSD in which the maximum value of the PSD is taken as the RR frequency [148].

(2) Best-Value Selection. The premise of best-value selection is to define some criteria for determining the highest-quality RR estimate from a series of estimates. This method is not truly a fusion method as multiple estimates from the same time point are not combined into a single estimate; however, some implementations are constantly updating so that the best method updates from window to window. Best-value selection has been done using GPs which provide both an RR estimate and a probability estimate for multiple respiratory modulations, from which the RR estimate with the highest probability estimate is taken [190]. A similar strategy was applied for AR modelling in which the phase angle of the AR pole with the large magnitude, across all modulations, was taken as the RR estimate [110].

(3) Machine Learning (ML). A small number of machine learning methods have been implemented to fuse multiple RR estimates. Johansson (2003) used a neural

network with 25 inputs, 1 hidden layer (5 nodes) and two output nodes (inspiration and expiration). The 25 inputs to the network were the five most recent values for five PPG derived features and the outputs provide waveforms of the estimated inspiration and expiration waveforms for 2 minute segments [189]. An extension of this neural network by Ahlstrom *et al.* (2004) sought to fuse RR estimates from ECG and PPG. The neural network used a similar structure with a single output for respiration maxima and the PPT, and RSA derived respiratory waveforms for ECG and PPG [98]. Finally, Zhu *et al.* (2014) used a Bayesian Continuous Valued Label Aggregator (BCLA) to fuse multiple RR estimates from the PPG into a single estimate [151].

3.5 Checkpoint I: Signal Quality Index (SQI)

All of the algorithms designed to extract the RR from the ECG and PPG assume that the underlying ECG and PPG waveforms are high-quality and relatively artefact free. As a result these algorithms are often ineffective in the presence of poor quality signals [47]. While low quality signals can often be improved through techniques to remove signal noise and corruption, there is a point beyond which the noise and corruption make quality analysis of the ECG or PPG no longer possible. As a result, a major topic of signal processing has been the development of SQIs capable of determining if a segment exceeds the threshold beyond which further analysis is no longer useful. Two authors, Karlen *et al.* (2013) and Peláez-Coca *et al.* (2013) have applied SQIs and artefact detection algorithms to the ECG and PPG signals prior to extracting the RR from the ECG and PPG. The method described by Karlen *et al.* (2013) seeks to detect artefactual PPG peaks using an Incremental Merge Segmentation (IMS) algorithm which extracts the pulse amplitude, maximum and minimum pulse intensity, and pulse period

for every PPG peak and compares the values for each peak to the expected values given the entire series of peaks to determine whether or not the peak is artefactual or not [66]. While the authors do not use this analysis to discard any segments or windows from analysis, an extension of this method implementing it as an SQI is possible. Peláez-Coca *et al.* (2013) calculate the first and second Hjorth parameters for segments of PPG data. These parameters represent the expected centre-frequency and half bandwidth of the signals, respectively. They discard segments of data when either of these parameters is outside of a pre-set threshold [134].

Despite the fact that very few SQIs have been implemented in applications specifically seeking to extract the RR from the ECG and PPG, there is a wide array of SQI methods that have been developed by authors looking to improve the robustness of HR estimation methods of the ECG and PPG [47]. These techniques include both time- and frequency-domain techniques and have been developed for numerous purposes including: arrhythmia detection, m-Health, robust HR estimation, and assessment of overall signal quality [47]. While most of the RR estimation algorithms published to date have little need for implementing an SQI checkpoint prior to conducting Module I because the data sets being used contain high-quality signals, the importance of the SQI checkpoint will be paramount in real-time clinical applications where the data collection is less likely to be closely monitored.

3.6 Checkpoint II: Respiratory Quality Index (RQI)

While SQIs are important because they assess the quality of the raw ECG and PPG waveforms, when seeking to extract the RR from the ECG and PPG, they are not sufficient to ensure that the respiratory modulations being extracted are truly present. Even

in situations where the ECG and PPG are minimally corrupted, it may not be possible to extract the RR because the presence and quality of the respiratory modulations vary based on a number of physiological factors such as gender, age, pre-existing health conditions, level of hydration, body position and the value of the RR itself [33, 45, 46]. However, the implementation of respiratory quality indices (RQIs) as a second checkpoint after the extraction of the respiratory waveforms can address this problem.

To date very few authors have developed and implemented RQIs and indeed the identification of a separate set of methods (and the name, RQI) is unique to this thesis. The current RQI methods focus on extracting some pertinent feature or features of the respiratory waveform and based on the value of these features determining some value, often from 0 to 1, on which to assess the quality of the respiratory waveform. The first example of this was Chan *et al.* (2013). In this work, the authors extracted four features from 45 second segments of the respiratory waveform: (1) the coefficient of variation, which is the standard deviation of the signal divided by the mean, (2) the mean peak-to-peak amplitude, (3) the coefficient of variation between minima, and (4) the ratio of true maxima and minima to local maxima and minima. These values were then fused in order to create a single RQI metric ranging from 0 to 1 [143]. Another method proposed by Lazaro *et al.* (2014) uses "peak condition" averaging in which the PSDs for the different respiratory waveforms are taken and where the peak area in the PSD peak must exceed a certain percentage of the total PSD area in order to be considered. All of the respiratory waveforms that are sufficiently peaked are then averaged and used to find the RR frequency [148]. Finally, Nemati *et al.* (2014) used the third Hjorth parameter,"purity," on respiratory waveform segments to assess the periodicity of the waveform where more periodic waveforms were given higher RQI values. The RQI values

were used as independent weighting factors in the Kalman filter-based fusion models described by the authors [33]. These authors show that the implementation of their RQIs improves the overall RR estimate. However, the development and improvement of RQIs remains an understudied element of the estimation of RR from ECG and PPG.

3.7 Single-Step Methods

While the majority of the methods for extracting the RR from the ECG and PPG can be fit into the proposed RR extraction framework (Figure 3.1), there remain some methods that have been described that extract the RR directly from the ECG and PPG. The underlying similarity of these methods is that they are all frequency-based methods that are designed to extract the BW respiratory modulation as it is an underlying respiratory frequency in the ECG and PPG. The challenge of these methods is that they must address specifically how the RR frequency is extracted because, without any prior extraction of the respiratory waveform alone, the HR frequency will be the dominant frequency component in the signal. There are two primary methods by which this challenge has been addressed: observation of the dominant frequency component that falls within a pre-set respiratory frequency range [196, 197, 198] or searching for the second-most dominant frequency component [199, 200]. The most common single-step method is the FFT [35, 115, 199, 201]; however, numerous other methods and variations have been applied including: short-time FFT [196], AR modelling [183, 197], Gaussian basis decomposition [202], correntropy-based spectral density (CSD) analysis [198], and time-varying sine cosine (TVSINCOS) decomposition [203]. While the SQI checkpoint can still be applied to single-step methods, one of the critical shortcomings of single-step methods is that, because no respiratory waveform is derived in the method, the RQI

checkpoint cannot be implemented.

3.8 Conclusion

The current chapter provides both a review and a framework (Figure 3.1) for understanding the previous methods that have been used for extracting the respiratory waveform and RR from the ECG and PPG. The framework described presents three "modules" and two "quality checkpoints" that can be implemented in order to estimate RR from ECG and PPG. To date, no authors have implemented all three modules and two checkpoints together. The application of the best-performing methods for each module and checkpoint throughout the literature in tandem present the opportunity to improve RR estimation. Furthermore, while methods have been described for the signal quality checkpoints, the number of methods is far sparser, particularly for the RQI checkpoint, than for any of the estimation modules. Due to this, the development of highly robust RQIs has the potential to greatly improve RR estimation methods. It is this pursuit that is the primary goal of this thesis. The following chapter presents the databases that will be used to develop, train, test, and evaluate the novel RQIs developed in this thesis.

Chapter 4

Data Sets

The investigations described in this thesis have been conducted using subsets of three independent data sets: CapnoBase ($\mathcal{D}_c$) [51, 66], MIMIC-III ($\mathcal{D}_m$) [52, 53, 54], and Oxford Dialysis III ($\mathcal{D}_d$) [55]. The first two data sets, $\mathcal{D}_c$ and $\mathcal{D}_m$, have been used to train and validate the RQI-based RR estimation method described in Chapters 6 and 7, and $\mathcal{D}_d$ has been used to demonstrate the efficacy of the RQI-based RR estimation method for early estimation of patient deterioration prior to known critical events.

The three data sets are independent and two are currently publicly available.[1] The data sets have all been collected at different sites and comprise different patient populations (as described later). Each data set contains three waveforms: a standard respiratory waveform (either capnography or IP), ECG, and PPG. For each data set, the standard respiratory waveform was manually scored in order to obtain a reliable ground truth.

Manually assessing the respiratory waveforms to obtain the RR ground truth was necessary in order to minimise the inaccuracy of the RR estimate obtained from the respiratory waveform. As a result of the data requirements described, the data used in these

[1] http://www.capnobase.org for $\mathcal{D}_c$ and http://www.robots.ox.ac.uk/~davidc for $\mathcal{D}_m$, the complete MIMIC-III database can be found at [54].

investigations were subject to a number of limitations: (1) only data sets with non-mobile patients were used, as no data sets containing a reliable respiratory waveform, ECG, and PPG were available, (2) only data with the highest-quality standard respiratory waveforms (as described later) were used, as it was impossible to accurately manually score poor-quality respiratory waveforms, and (3) the number of patients and the length of the recordings of the data sets available was small as only a limited number of data were available (because manually scoring respiratory waveforms is a time-consuming process). The data used in the investigations described here represent one of the largest and most diverse accumulations of data that have highly accurate manually-scored ground truth respiratory waveforms.

4.1 CapnoBase Database

The CapnoBase data set was collected at the British Columbia Children's Hospital and St. Paul's Hospital in Vancouver, Canada by Karlen *et al.* (2010) [51, 66]. The initial data set consists of data extracted from 59 paediatric surgeries (BC Children's) and 35 adult surgeries (St. Paul's). All of the patients in the data set were undergoing elective surgery and routine anaesthesia and were either breathing spontaneously or artificially ventilated. For all patients, capnography, standard transmission-mode finger PPG, and lead II ECG were collected. The capnography was collected at 25 Hz, while the ECG and PPG were collected at 300 Hz. The capnography and PPG were subsequently up-sampled to 300 Hz.

The authors of the original study randomly selected 8-minute segments from 42 (29 paediatric and 13 adult cases) of the initial 84 recordings, which have been made publicly available. Additionally, the authors of the original work manually scored the start of

inspiration and expiration for every breath in the capnography data, the peaks and artefacts in the PPG data, and the peaks in the ECG data. The 42 segments described represent the $\mathcal{D}_c$ data set used in this thesis. The physiological characteristics of the patients can be seen in Table 4.1 with the exception of gender, which was not reported by the original authors.

Table 4.1: *Physiological information on $\mathcal{D}_c$, $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, $\mathcal{D}_m$, and $\mathcal{D}_d$ data sets*

Data Set	Pt. Type	n	Median Age (Range)	% Female	% Ventilated	Rec. Length. (min)
$\mathcal{D}_c$	Paediatric	29	8 (1 - 16)	-	44.8%	8
	Adult	13	47 (32 - 63)	-	69.2%	8
	All	42	14 (1 - 63)	-	52.4%	8
$\mathcal{D}_{c,1}$	Paediatric	21	7 (1 - 16)	-	38.1%	8
	Adult	11	43 (32 - 63)	-	63.6%	8
	All	32	15 (1 - 63)	-	46.9%	8
$\mathcal{D}_{c,2}$	Paediatric	8	11.5 (2 - 16)	-	62.5%	8
	Adult	2	55.5 (52 - 59)	-	100.0%	8
	All	10	13 (2 - 59)	-	70.0%	8
$\mathcal{D}_m$	Adult	53	68 (19 - 91.4)	60.4%	0.0%	8
$\mathcal{D}_d$	Adult	8	75.5 (59 - 87)	25.0%	0.0%	120

For this work, $\mathcal{D}_c$ was randomly subdivided into training and validation ($\mathcal{D}_{c,1}$) and test ($\mathcal{D}_{c,2}$) sets. $\mathcal{D}_{c,1}$ contains 32 segments (21 paediatric and 11 adult, 76.2% of the data overall) and $\mathcal{D}_{c,2}$ contains 10 segments (8 paediatric and 2 adult, 23.8% of the data overall). The physiological characteristics of each subset can be seen in Table 4.1.

4.2 MIMIC-III Database

The MIMIC-III data set was collected from 2001 to 2007 at the Beth Israel Deaconess Medical Center in Boston, MA USA in four ICU units within the hospital: medical, surgical, cardiac, and cardiac surgery. The MIMIC-III Waveform Database contains 22,317 sets of continuous recordings [52, 53, 54]. Each recording contains a number of continuous waveforms which may include ECG (multiple leads), standard transmission-

mode finger PPG, IP, and BP (multiple locations) all sampled at 125 Hz. In addition, a number of numerics were also collected or derived from the waveforms including: BP, CO, CO_2, HR, RR, SpO_2, ST, and temperature. These numerics were collected less frequently and non-uniformly. 10,282 of the recordings were matched with the MIMIC-III Clinical Database and thus the recordings are associated with additional patient information such as demographics, provider collected vital signs, medications, and lab results.

For this analysis, 8-minute waveform segments from 53 unique patients were extracted from the MIMIC-III Waveform Database. 8-minute segments were chosen to provide consistency with the $\mathcal{D}_c$ data set. This subset of the MIMIC-III data set comprises the $\mathcal{D}_m$ data set. The 53 recordings in the MIMIC-III $\mathcal{D}_m$ data set were selected from the 22,317 recordings in the entire MIMIC-III Waveform Data Set based on the following criteria:

1. Records with matching information from the MIMIC-III Clinical Database (N = 10,282)
2. Records containing, at a minimum, IP, ECG, and PPG lead II (N = 1,123)
3. Records that contain, at a minimum, recordings of 68 minutes in length (N = 1,102)[2]
4. Records with estimated RR error in the IP of < 1 brpm (N = 53)[3]

[2]Each 8-minute recording represents the 60th to 68th minute of data. This allowed for the first 60 minutes to be discarded as they were more likely to be artefactual or contain excessive signal noise during monitor set up and calibration.

[3]The RR error in the IP was obtained by estimating the RR using two independent estimators, a peak detection (time-domain) algorithm and an FFT (frequency-domain) algorithm. The RR was calculated using each estimator in 15 consecutive, non-overlapping 32-second windows for each 8-minute recording. The RR error was the absolute difference between the RR estimates for each window and the overall RR error for an 8-minute segment represented the mean absolute error (MAE) for the entire recording. While these RR estimates were not used as the gold standard RR for the IP data, their agreement was considered a suitable proxy for selecting the highest quality respiratory waveforms prior to manual scoring.

The demographic data from the 53 patients in $\mathcal{D}_m$ are shown in Table 4.1. After initial selection of the 53 segments comprising the $\mathcal{D}_m$ data set, the peaks of the IP waveforms were manually scored by two researchers (this author and M.A.F. Pimentel) to provide two unique sets of annotations for respirations (where the peaks are assumed to be the end of inspiration and the start of exhalation). These annotations were used to obtain a gold standard RR using the algorithm described in Chapter 5.

4.3 Oxford Dialysis III Database

The Oxford Dialysis III data set was collected at the Churchill Hospital, Oxford, UK and was originally described by MacEwen (2016) [55]. The $\mathcal{D}_d$ data set consists of 635 recordings from 58 patients during haemodialysis (HD) treatment. Patients were initially recruited into the study if they had been on HD for at least one month, were over 18 years of age, and were capable of giving informed consent. For each patient and recording, basic physiological information was collected in addition to continuous waveform data for IP, ECG, and PPG. IP was sampled at 64 Hz and ECG and PPG were sampled at 256 Hz.

For this analysis a subset of the Oxford Dialysis III data set was extracted, $\mathcal{D}_d$, which consists of 8 unique recordings of at least two and a half hours in duration in which the end point of the segment is an intradialytic hypotensive event. The selection criteria for this subset of data is described below:

1. Records containing an intradialytic hypotensive event defined as either (1) not having a record of systolic blood pressure (SBP) at baseline (prior to bleed out) and a mean arterial blood pressure (MAP) < 60 mmHg or (2) a decrease in SBP by $> 20\%$ from baseline and an MAP < 60 mmHg ($N = 193$ sessions from 31

unique patients)

2. Records containing at least 2.5 hours of waveform data prior to the intradialytic hypotensive event ($N = 35$ sessions from 18 unique patients)[4]
3. Only results with a linear regression fusion RQI value ($Q_{F,1}$) higher than the high-threshold (T_{HT}) based on the methods described in Chapters 5 and 6 ($N = 15$ sessions from 8 unique patients)
4. Select only the session with the highest value for $Q_{F,1}$ for each unique patient ($N = 8$ sessions)

The demographic data from the 8 patients in $\mathcal{D}_d$ are shown in Table 4.1. After initial selection of the 8 unique 2 hour sessions (after removal of the first 30 minutes of recording), the peaks of the IP waveforms were manually scored by two researchers (this author and S.J. Rodgers[5]) to obtain a gold standard RR using the same method that was used for the $\mathcal{D}_m$ data set and described in Chapter 5.

4.4 Data Set Distinctions

While all of the data sets used in this work have many similar features, there are two distinctions that make them ideal for the purpose that the data set was used. These are:

[4]This was selected such that a 2 hour recording could be obtained after discarding the first 30 minutes of data due to artefact and increased noise during device setup and calibration. While 30 minutes is considered less optimal than the prior hour that is discarded in $\mathcal{D}_m$, this was chosen because the lengths of the recordings in the Oxford Diaylsis III data set are significantly shorter than those in the MIMIC-III data set.

[5]The process of manually scoring peaks in reference waveforms in the $\mathcal{D}_m$ and $\mathcal{D}_d$ data sets was methodologically identical and both were scored by this author. The second scorer was different in the two data sets, M.A.F. Pimentel in $\mathcal{D}_m$ and S.J. Rodgers in $\mathcal{D}_d$. The percentage of windows in which the RR estimates based on the scorers annotations did not agree within $\pm$ 2 brpm in the two data sets was 7.3% in $\mathcal{D}_m$ and 17.0% in $\mathcal{D}_d$. This difference could reflect a systematic difference between the second scorers but could also reflect a difference in the signal quality of the reference waveforms between the two data sets. However, given that the primary interest in manually scoring peaks was to obtain a reliable gold standard reference RR and because in both data sets the peaks were scored independently by two researchers and only those windows in which the two researchers agreed were kept, the reference RR values in both $\mathcal{D}_m$ and $\mathcal{D}_d$ are believed to be of similarly high fidelity and suitable for use as gold standard RR reference estimates.

(1) population age range and (2) reason for initial admission and recording termination.

4.4.1 Age Range

The age ranges for the three data sets: $\mathcal{D}_c$, $\mathcal{D}_m$, and $\mathcal{D}_d$ vary and together span ages encompassing early infancy to beyond the average life-expectancy. The age range for each data set and subset is shown in Figure 4.1. The $\mathcal{D}_c$ data set contains the largest age range, from 1 to 63. $\mathcal{D}_c$, specifically $\mathcal{D}_{c,1}$, was selected to be used as the training and validation data set because it encompasses a wide age range. $\mathcal{D}_c$ was randomly subdivided into $\mathcal{D}_{c,1}$ and $\mathcal{D}_{c,2}$. As a result, any variation in the age ranges of these two subsets is the result of random variation. The $\mathcal{D}_m$ data set has ages ranging from 19 to 91.4. Given that this data set is comprised of a general population of hospitalised adults, this age range spans from early adulthood to beyond the average life-expectancy and skews towards an elderly population, as would be expected in an inpatient hospital setting. Finally, the age range of $\mathcal{D}_d$ data set, 59 to 87, skews heavily towards an elderly population. This is a reflection that the entirety of this population have been diagnosed with chronic conditions requiring haemodialysis. Often conditions requiring haemodialysis onset late in adulthood. Together, these three data sets encompass a wide variety of ages and reflect the importance of RR monitoring among many different populations.

It is of particular note that the $\mathcal{D}_c$ data sets contain paediatric as well as adult patients, while $\mathcal{D}_m$ and $\mathcal{D}_d$ data sets only contain adult patients. The inclusion of paediatric patients in the $\mathcal{D}_c$ data set is important as this data set represents the training and validation data set and thus a broader age distribution may increase robustness especially since normal RR varies significantly between paediatric and adult patients.

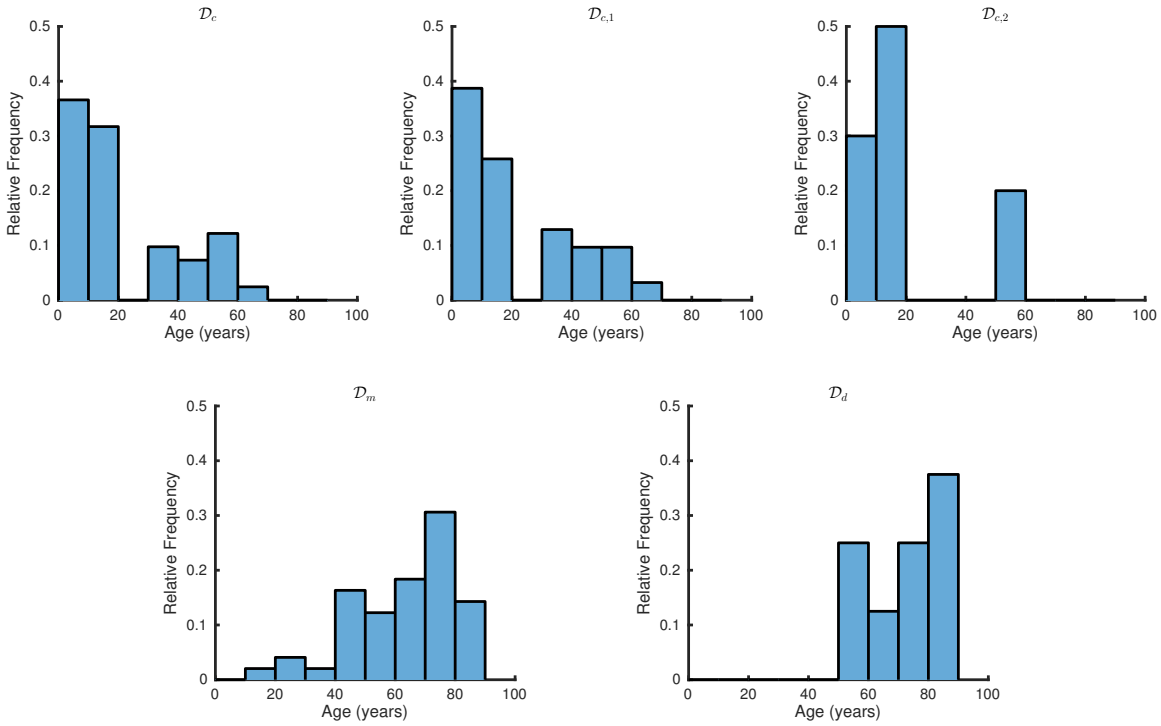


Figure 4.1: *Relative frequency of age distributions for $\mathcal{D}_c$, $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, $\mathcal{D}_m$, $\mathcal{D}_d$ data sets.*

4.4.2 Reason for Admission

The $\mathcal{D}_c$ data set contained patients that were admitted for elective surgery and routine anaesthesia. This is an ideal data set for model training as these data are highly likely to have high signal quality and thus excessive signal noise will not be a factor in training the model. Conversely, the $\mathcal{D}_m$ data set contains patients admitted to the ICU and are much more likely to have large noise artefacts. As a result, this is a suitable test data set for the model. Furthermore, the segments in these data sets were not terminated for any known physiological meaning thus it is acceptable to assume that over the course of the 8-minute segment a critical physiological change was unlikely. Finally, the $\mathcal{D}_d$ data set is collected on patients who were initially stable but decline until a critical event occurs. As a result, the ability of the model to observe clinical deterioration can be studied using this data set.

4.5 Conclusion

This chapter presents the three data sets that have been used in this thesis. These data sets have been derived from three different databases CapnoBase, MIMIC-III, and Dialysis III. Together these three databases represent a broad subject population including: paediatric, adult, and geriatric subjects; healthy and hospitalized subjects; and chronically- and acutely-ill subjects. Each of these databases contains continuous monitoring waveforms for ECG, PPG, and either IP or capnography. The CapnoBase database has been subdivided into two data sets, $\mathcal{D}_{c,1}$, a 32 subject training and validation data set, and $\mathcal{D}_{c,2}$, a ten subject test data set. The $\mathcal{D}_m$ data set is a test data set consisting of 53 subjects selected from the MIMIC-III database. Finally, the $\mathcal{D}_d$ data set is an eight subject test data set selected from the Dialysis III database. The remainder of this thesis uses these data sets to describe the development and validation of RQIs for the improvement of RR monitoring using ECG and PPG.

Chapter 5

Individual RQIs

The preceding chapters established the clinical need for robust RR estimation. A number of physiological mechanisms modulate the ECG and PPG waveforms, and the respiratory waveform can be extracted from the ECG and PPG as secondary modulations. Despite an extensive literature on methods for extracting the RR from the ECG and PPG, methods with clinical accuracy and robustness still do not exist. In seeking to improve the estimation of the RR waveform from the ECG and PPG, the following chapters of this thesis introduce, test, and validate a novel method for assessing the presence and quality of the respiratory modulations extracted from the ECG and PPG which can be implemented in tandem with existing RR estimation algorithms. The aim is to move the field closer to a system that is sufficiently robust for use in clinical practice.

The present chapter focuses on the development and validation of five individual RQI techniques based on: the FFT ($Q_{S,1}$), autocorrelation ($Q_{S,2}$), cosine cross-correlation ($Q_{S,3}$), AR ($Q_{S,4}$), and Hjorth complexity ($Q_{S,5}$). The first section of this chapter is a technical review which will introduce and explain the principal signal processing techniques from which each RQI is derived. The second section will explain how these techniques are applied to the ECG and PPG-derived respiratory waveforms to obtain

RQIs. This will include waveform extraction and pre-processing, the application of our RQI techniques, and our RQI testing and validation methods. The final section of this chapter presents the performance of the individual RQIs.

5.1 Techniques

5.1.1 The Fourier Transform

The Fourier transform is one of the most commonly-used and most robust signal processing tools. It decomposes a time-domain signal into its frequency components [204], and its basic principle was described in Section 3.3.2. While the Fourier transform can be implemented for both discrete and continuous signals, the implementations are different. The most common implementation of the Discrete Fourier Transform (DFT) used in digital signal processing is the Fast-Fourier Transform (FFT) which will be used in this thesis [204]. A thorough exploration of the FFT can be found in Lyons (2011) [204], which is not repeated here for brevity, due to the standard nature of the FFT in most fields of engineering.

5.1.2 Autocorrelation

The autocorrelation function measures the similarity of a signal to itself at different points in time based on a time shift, k, called the *lag* [205]. For a discrete signal, the autocorrelation can be measured at every possible lag, k, up to $(N - k) > 0$ where N represents the total number of points in the signal, and k is an integer that represents shifting the signal by $\frac{k}{f_s}$ seconds, where f_s is the signal sampling frequency [206]. The autocorrelation is a biased estimator and makes the assumption that the signal being analysed is stochastic [206]. The autocorrelation can be defined for discrete signals as

[205]:

$$r[k] = \frac{1}{N} \sum_{n=1}^{N-k} x[n]x[n+k] \tag{5.1}$$

where $r[k]$ is the autocorrelation value at lag k, N is the number of samples in the signal, $x[n]$ is the original signal and $x[n+k]$ is the signal shifted by lag, k. The autocorrelation function can be used to obtain the autocorrelation matrix, $\mathbf{R}$, defined as:

$$\mathbf{R} = \begin{bmatrix} r[0] & r[-1] & \dots & r[1-k] \\ r[1] & r[0] & \dots & \vdots \\ \vdots & \dots & \dots & r[-1] \\ r[k-1] & \dots & r[1] & r[0] \end{bmatrix} \tag{5.2}$$

and the autocorrelation vector, $\mathbf{r}$, defined as:

$$\mathbf{r} = \begin{bmatrix} r[1] & r[2] & \dots & r[k] \end{bmatrix}^{\mathrm{T}} \tag{5.3}$$

The autocorrelation matrix, $\mathbf{R}$ and the autocorrelation vector, $\mathbf{r}$ can be used within the Yule-Walker equation for autoregression, which will be discussed in Section 5.1.4 [205].

Depending on the desired output of the autocorrelation function, the basic autocorrelation equation in 5.1 can be modified to include zero-mean, unit-variance normalisation as well as a correction for sample bias by dividing the sum of the autocorrelations for each data point by $N-1$ instead of N [207]. As a result the updated autocorrelation function is:

$$r[k] = \frac{1}{s^2(N-1)} \sum_{n=1}^{N-k} (x[n] - \bar{x})(x[n+k] - \bar{x}) \tag{5.4}$$

where s^2 is the unbiased sample variance, $s^2 = \frac{1}{(N-1)} \sum_{n=1}^{N} (x[n] - \bar{x})^2$, and where $\bar{x}$ is the sample mean. This can be simplified to:

$$r[k] = \frac{\sum_{n=1}^{N-k} (x[n] - \bar{x})(x[n+k] - \bar{x})}{\sum_{n=1}^{N} (x[n] - \bar{x})^2} \tag{5.5}$$

The modified autocorrelation, Equation 5.5, is the equation that will be used in this work.

5.1.3 Cross-Correlation

The cross-correlation function is identical to the autocorrelation function except that instead of a single waveform being correlated with itself, two signals are correlated with each other [206]. The function for the cross-correlation for two waveforms of identical length N is:

$$r[k] = \sum_{n=1}^{N-k} x[n] y[n+k] \tag{5.6}$$

where $x[n]$ and $y[n]$ represent the two waveforms and k represents the lag between the two signals. Similar to the autocorrelation, the cross-correlation can be modified from the original cross-correlation in Equation 5.6 to give an unbiased or biased estimate. The function can also be normalised such that the cross-correlation at lag $k = 0$ is $r[0] = 1$ [208]. The cross-correlation function used in this thesis is the unbiased estimator with zero-mean, unit-variance normalisation of each waveform:

$$r[k] = \frac{1}{N-1} \sum_{n=1}^{N-k} \frac{(x[n] - \bar{x})}{s_x^2} \frac{(y[n+k] - \bar{y})}{s_y^2} \tag{5.7}$$

where $\bar{x}$ and $\bar{y}$ are the sample means, and where s_x^2 and s_y^2 are the sample variances of $x[n]$ and $y[n]$, respectively.

5.1.4 Autoregression

The basic principles of AR modelling were described in Section 3.3.2. An AR model is a method for analysing the underlying frequency content and the power spectral density of a time-series waveform. An AR model is advantageous because the resulting spectral estimation is a continuous function, even for a discrete time-series input. Furthermore, the model only relies on the assumptions that the underlying time-series is both stationary and stochastic. However, AR modelling requires complex model identification, often making it more difficult to apply than the Fourier transform [205].

The basic concept in AR modelling is that the n^{th} data point in a time-series waveform, $x[n]$, can be estimated using a weighted sum of the previous i data points and an error term, ε. Therefore the AR model can be defined as [205]:

$$x[n] = \sum_{i=1}^{M} a_i x[n-i] + \varepsilon[n] \tag{5.8}$$

where M is the AR model order and represents the number of previous values that are used to estimate the current value, and where a_i is the unique weight for each of the previous values.

Therefore, in order to obtain the AR model for a time-series, the model order M must be set, and the AR model must be solved to obtain $\mathbf{a}$, the vector of weights and $\boldsymbol{\varepsilon}$, the vector of error terms. The optimal AR model for a given model order can be solved in one of two commonly-used ways: the covariance method and the Yule-Walker equation. The goal of both methods is to determine the optimal weight values, $\mathbf{a}_{opt}$ such

that the values in $\boldsymbol{\varepsilon}$ are minimised [205].

The covariance method minimises the values of $\boldsymbol{\varepsilon}$ using the direct least-squares technique to find the optimal values for $\mathbf{a}$, called $\mathbf{a}_{opt}$. The minimised terms of $\boldsymbol{\varepsilon}$ represent the values where the mean squared error (MSE) between the actual values in the time series and the autoregression estimates are minimised. In the covariance method, this relationship is found by finding the values of $\boldsymbol{\varepsilon}$ that are orthogonal to $\mathbf{X}$. $\mathbf{X}$ is an $M \times M$ matrix consisting of M column vectors where each vector contains the M previous terms for each value $x[n]$ for $n = 1 \dots M$. The relationship of $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ can be best expressed in matrix notation as [205]:

$$\mathbf{X}^T \boldsymbol{\varepsilon} = 0 \tag{5.9}$$

where $\mathbf{X}^{\mathrm{T}}$ is the transpose of the matrix $\mathbf{X}$. This equation can be solved to give the optimal weight terms, $\mathbf{a}_{opt}$:

$$\mathbf{a}_{opt} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{x} \tag{5.10}$$

where $\mathbf{x}$ represents the vector of the $x[n]$ terms for $n = 1 \dots M$.

The other method for determining the optimal weight values, the Yule-Walker equation, is similar to the covariance method except that it equates $\mathbf{X}^T\mathbf{X}$ to $\mathbf{R}$, the autocorrelation matrix and $\mathbf{X}^{\mathrm{T}}\mathbf{x}$ to $\mathbf{r}$, the autocorrelation vector such that [205]:

$$\mathbf{a}_{opt} = \mathbf{R}^{-1}\mathbf{r} \tag{5.11}$$

where the methods for finding both $\mathbf{R}$ and $\mathbf{r}$ were defined in Section 5.1.2. The solutions for the AR model found using the covariance method and the Yule-Walker equation differ slightly, but the differences are usually not substantial. In this thesis, the Yule-Walker

was selected as the preferred method for solving AR models.

In addition to solving the AR model for both $\mathbf{a}_{opt}$ and $\boldsymbol{\varepsilon}$, the optimal model order M must must be determined. Many criteria for determining the prediction error exist including Akaike's Information Criterion (AIC), the Optimal Parameter Search (OPS), Final Prediction Error, Criterion Autoregressive Transfer Function, and Rissanen Minimum Description Length. The most commonly-used criterion is AIC which is defined as [205]:

$$AIC(M) = Nlog(\sigma_p^2) + 2M \tag{5.12}$$

where N is the sample size, σ_p^2 is the prediction error ($\boldsymbol{\varepsilon}$) variance, and M is the model order. The value of M that minimises AIC is this optimal model order. As defined by the AIC, the optimal model order is weighted by two competing factors: model complexity and model fit. The reason that both model complexity and model fit are necessary in selecting the optimal value for M is because M can be any value from 1 to N. The maximum value of M is N because this represents the case where every previous data point is being used to predict the next data point; therefore, a requirement for the selection of the model order is $M \leq N$. However, if M were N, the model would have minimised σ_p^2, but the model would be over-fit to the specific input data and would not be applicable to test data on which the model was not trained. For this reason, in the AIC, $2M$ is added as a penalty term which reduces the chance that the model over-fits the training data. Thus, the optimal model order is one that neither maximises fit nor minimises model order, but rather finds a reasonable "optimal" fit for a reasonably minimal model order.

The above description of the AR model provides the optimal recreation of the original

waveform, however the output is still in the time-domain. The AR model must be transformed into the frequency domain to provide both the frequency content and power spectral density (PSD) estimate. This is done by taking the z-transform of the original AR model in Equation 5.8 such that [205]:

$$X[z] = \sum_{i=1}^{M} a_i X[z] z^{-i} + W[z] \tag{5.13}$$

where $X[z]$ is the z-transform values of $x[n]$, a_i are the weight terms, z^{-i} is the position of the sample values in the sequence, and $W[z]$ are the z-transform values of the error terms, ε. This z-transformed model of the AR can be used to generate the model transfer function:

$$H(z) = \frac{1}{1 - \sum_{i=1}^{M} a_i z^{-1}} \tag{5.14}$$

where $H(z)$ is the all-pole transfer function of the AR model and where z can be substituted by $e^{i\omega T}$ such that:

$$H(\omega) = \frac{1}{1 - \sum_{i=1}^{M} a_i e^{-i\omega T}} \tag{5.15}$$

and where the roots of the denominator of the transfer function represent the complex poles of the AR model in the complex plane (and which appear as conjugate pairs). These roots can be used to determine the magnitude and frequency of the PSD peaks from the complex plane. They can also be used to generate the PSD by multiplying the square of the transfer function, $H(\omega)$ with the prediction error variance, σ_p^2, such that

the PSD of the AR model is [205]:

$$PSD = \frac{\sigma_p^2}{|1 - \sum_{i=1}^{M} a_i e^{-ji\omega T}|^2} \tag{5.16}$$

5.1.5 Hjorth Parameters

The Hjorth parameters are a set of three statistical parameters used in the time-domain to extract a set of feature descriptors of the signal. The three Hjorth parameters were first described by Hjorth (1970) and include: activity, mobility, and complexity [209]. Activity is a measure of the variance of the mean power of the signal. Mobility is a measure of the standard deviation (SD) of the signal slope and can be more simply understood as a measure of the mean frequency. Complexity is a measure of the sinusoidality of the waveform, where signals that more closely match the ideal sinusoid have lower complexity values than those signals that do not.

The three Hjorth parameters are calculated using the zero'th (m_0), second (m_2), and fourth (m_4) moments of the waveform, which are variance of the signal, and its first and second derivative, respectively, and when assuming zero-mean unit-variance normalisation can be calculated as:

$$m_0 = \frac{1}{N} \sum_{i=0}^{N} x[n]^2 \tag{5.17}$$

$$m_2 = \frac{1}{N} \sum_{i=0}^{N-1} (x[n+1] - x[n])^2 \tag{5.18}$$

$$m_4 = \frac{1}{N} \sum_{i=0}^{N-2} (x[n+2] - 2x[n+1] + x[n]))^2 \tag{5.19}$$

where $x[0] = 0$ and N is the length of the signal. Using these moments, the Hjorth parameters can be calculated as [209]:

$$HP_{activity} = m_0 \tag{5.20}$$

$$HP_{mobility} = \left(\frac{m_2}{m_0}\right)^{\frac{1}{2}} \tag{5.21}$$

$$HP_{complexity} = \frac{(m_4/m_2)^{\frac{1}{2}}}{(m_2/m_0)^{\frac{1}{2}}} \tag{5.22}$$

The primary Hjorth parameter considered in this thesis is complexity, because of its ability to measure the sinusoidality of a signal. The less sinusoidal a signal, the higher the value of its complexity, where the complexity of a perfectly sinusoidal signal is unity [209].

5.2 Methods

Five unique RQIs were derived from the five signal processing techniques described in the previous section. The techniques used to calculate the RQIs did not require training to set ideal parameters; because of this, division of the available data into training-and-validation ($\mathcal{D}_{c,1}$) and test ($\mathcal{D}_{c,2}$ and $\mathcal{D}_m$) sets was not necessary. As a result, the five individual RQIs were validated using a single superset containing the data from both $\mathcal{D}_c$ and $\mathcal{D}_m$. This section consists of three subsections: (1) the pre-processing steps that were used to conduct the respiratory waveforms from the ECG and PPG, (2) the description and application of the derived RQIs, and (3) the methods that were applied to test the performance of the RQIs.

5.2.1 Pre-Processing

Prior to applying the RQIs, the respiratory waveforms needed to be extracted from the ECG and PPG waveforms in $\mathcal{D}_c$ and $\mathcal{D}_m$. The pre-processing included three steps: (1) data-windowing, (2) respiratory waveform extraction, and (3) waveform processing, including: data interpolation, re-sampling, and filtering.

(1) Data-Windowing

As described in Chapter 4, the $\mathcal{D}_c$ data set consisted of 42 recordings and the $\mathcal{D}_m$ consisted of 53 recordings, all of 8-minutes in duration and containing concurrent ECG, PPG, and respiratory reference waveform. It was determined that the use of an SQI (as described in Section 3.5) was not necessary as the data in $\mathcal{D}_c$ and $\mathcal{D}_m$ were convenience-selected for high-quality waveforms as described in Chapter 4. Prior to respiratory waveform extraction, the 8-minute segments of data were divided into 32s non-overlapping windows such that each recording resulting in 15 windows. $\mathcal{D}_c$ therefore consisted of 630 windows ($\mathcal{D}_{c,1}$ consisted of 480 and $\mathcal{D}_{c,2}$ consisted of 150) and $\mathcal{D}_m$ consisted of 795 windows. As previously described, windows for extracting the respiratory waveform are often between 30 seconds [163] and 2 minutes [164]. The 32s window was selected for two reasons. The first is to allow for an even division of the 8-minute segments (i.e., the entire available waveform is used and exactly 15 windows are created). The second was that it was assessed that the smaller the window of data that was used, the stronger was the assumption that the respiratory waveform was stationary (i.e., it is much less likely that the RR varies extensively for smaller window sizes). The windows were extracted to be non-overlapping so that each window contained a unique waveform.

(2) Respiratory Waveform Extraction

Using the 32s windows, the respiratory waveforms from the ECG and PPG were extracted. It was determined that because true frequency-based methods for extracting the respiratory waveform are only capable of extracting the BW modulation, a feature-based extraction technique (summarised in Section 3.2.1) was preferred in order to allow uniform extraction of all three modulations: AM, FM, and BW.

The feature-based extraction method that was implemented was not identical for both ECG and PPG but consisted of the same steps: (1) peak detection, (2) trough detection, (3) peak-correction. The specific algorithms for both ECG and PPG are described below:

ECG Feature Extraction The R-peaks were detected in the ECG windows using the search-back algorithm defined by Pimentel *et al.* (2015) [49]. The troughs corresponding to each R-peak were assessed as being the local minima of the waveform in the 0.2s preceding an R-peak. After finding the troughs, a peak correction algorithm was implemented to determine if a more accurate local maxima existed $\pm$ 0.2s around the peak determined by the search-back algorithm. Using the detected peaks and troughs, the time-points and amplitudes of each feature were extracted.

PPG Feature Extraction The peaks and troughs in the PPG waveform were determined using a multi-element process. The first element was to apply a bandpass filter to the waveform with a passband from 0.05 to 4.00 Hz. This removed any low-frequency baseline fluctuation in the signal and was empirically determined to improve the PPG peak detection algorithm. Using the filtered signal, the PPG peaks were found using the peak delineator algorithm defined in Li *et al.* (2010) [45]. The troughs were

found by searching for the local minima in the 0.2s preceding a detected peak; the same method was used for the ECG waveform. Additionally, the PPG peak locations were corrected using the same ± 0.2s search window around a detected peak. The detected peaks and troughs were then used to obtain the time-points and amplitudes of each feature, however, the feature amplitudes that were extracted were for the *original, unfiltered* signal.

For each window, the peaks and troughs were used to generate a vector pair of data such that the peaks are represented as $(t_{pk}, y_{pk})_{i...N}$ and the troughs are represented as $(t_{tr}, y_{tr})_{i...N}$, where t_{pk} and t_{tr} are the time-points and y_{pk} and y_{tr} are the amplitudes of the peaks and troughs, respectively, and where N represents the total number of peaks and troughs in a given window. These vector pairs were used to extract the AM, FM, and BW modulations of each window where:

$$y_{AM,i} = (t_{pk,i}, y_{pk,i} - y_{tr,i})_{i...N} \tag{5.23}$$

$$y_{FM,i} = (t_{pk,i}, t_{pk,i+1} - t_{pk,i})_{i...N} \tag{5.24}$$

$$y_{BW,i} = (t_{pk,i}, y_{pk,i})_{i...N} \tag{5.25}$$

(3) Waveform Processing

The extraction of the AM, FM, and BW modulations resulted in 6 respiratory modulations (3 ECG and 3 PPG) for each concurrent 32s window in the $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. Each modulation was fitted with a cubic spline interpolation, down-sampled to 4 Hz, and filtered using a 5^{th}-order Butterworth bandpass filter with a 0.083 to 1.000 Hz passband (representing the assumed physiological limits of the RR, 5 to 60 brpm).

5.2.2 Individual RQIs

Each of the five RQIs proposed earlier was applied to every respiratory modulation extracted from the $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets, and designed such that as the value of the RQI increases, the likelihood that a given modulation contains a true respiratory waveform increases.

FFT RQI ($Q_{S,1}$)

The FFT RQI ($Q_{S,1}$) is applied by obtaining the PSD of the FFT after applying mean-normalisation and a Hamming window to the respiratory modulation. The result is a 128-point FFT where each bin covers a frequency band of $\frac{1}{32}$ Hz. The first half of the PSD ($X_{i=1...64}$) is used to calculate $Q_{S,1}$ which is defined as:

$$Q_{S,1} = \frac{\hat{X}}{\sum_j X_j} \tag{5.26}$$

where:

$$\hat{X} = \max\left\{ \sum_{k=z-1}^{z} X_k \middle| (z-1), (z) \in [2\ 32] \right\} \tag{5.27}$$

and where $j \in [3\ 32]$. Here the 128-point FFT has components with indices [2 32] corresponding to the range of plausible respiratory frequencies [0.083 1.000] Hz. $Q_{S,1}$ thus defines the ratio of (the maximum power in any 2-component subrange corresponding to a window range of 0.031 Hz) to (the total power in the range of respiratory frequencies).

Autocorrelation RQI ($Q_{S,2}$)

The autocorrelation RQI ($Q_{S,2}$) was determined by calculating the autocorrelation (Equation 5.28) of the respiratory waveform for every sample lag within the relevant physiologic

respiratory range ($k = 4 \ldots 48$) where each lag denotes a phase shift of $\frac{4}{k}$ Hz. Therefore $Q_{S,2}$ is defined as:

$$Q_{S,2} = \max_{k=4\ldots48} \left\{ \frac{\sum_{n=1}^{N-k}(x[n] - \bar{x})(x[n+k] - \bar{x})}{\sum_{n=1}^{N}(x[n] - \bar{x})^2} \right\} \tag{5.28}$$

Cosine Correlation RQI ($Q_{S,3}$)

The cosine correlation RQI ($Q_{S,3}$) was based on the cosine correlation SQI defined by Springer (2015) [210]. The principle of the cosine correlation is to find how similarly a particular waveform correlates to a perfect cosine wave. This is done using three steps: (1) calculate the signal autocorrelation, (2) calculate the frequency that corresponds to the highest signal autocorrelation, and (3) calculate the correlation between the autocorrelation function and the specified cosine waveform. In the first step, the autocorrelation function is calculated for the respiratory waveform using Equation 5.5. In the second step, the cosine function that best fits the respiratory waveform was calculated by summing the value of the waveform autocorrelation at a window centred around each of the first three peaks of the test cosine function. The assumption is that the overall sum of the autocorrelation values from these windows would be maximised when the underlying waveform had a primary oscillation that most resembled the test cosine function. This was calculated as:

$$\epsilon_n = \max_{n=4\ldots40} \sum_{j=1}^{3} \sum_{k=j(n-2)}^{j(n+2)} (r_k) \tag{5.29}$$

where ϵ_n is the term that is maximised when the optimised cosine function is found, j is the current cosine peak that is being considered (where in this instance the total number of peaks considered is 3), k and r_k are the lag and the autocorrelation value at lag k,

respectively, as seen in Equation 5.5, and n represents the range of time lags within the physiologically relevant respiratory range, where in this case $n = 4 \dots 40$. To further explain the importance of n, if the signal were a perfect sinusoid, the maximum value for n would represent the point in which the lagged signal was lagged by one period from the original signal. $2n$ and $3n$ would represent the point in which the lagged signal was lagged by two and three periods, respectively, from the original signal. Given this, it should be noted that for the cosine correlation to span a breathing range of 5 to 60 brpm with a 4 Hz sampling frequency, consistent with the range seen in Section 5.2.2, n should span from $n = 1 \dots 48$. However, because the cosine correlation implementation described requires adding values of the autocorrelation up to $3n+2$ and the sample signal consists of 124 data points, the maximum allowable value for n is 40. This means that for values of $n = 4 \dots 40$, the respiratory range considered ranges from 6 to 60 brpm. Ultimately, it was determined that the model was more robust by incorporating three cosine peaks into the function as opposed to two at the loss of being able to calculate the cosine correlation for a respiratory frequency of 0.083 Hz. Once the optimal value for ϵ is determined, it can be used to back-calculate the optimal cosine oscillation frequency, where $f_{cos} = \frac{f_s}{n}$ where f_s is the waveform sampling frequency, which in this instance is 4 Hz. Once the optimal cosine function frequency is determined, $Q_{S,3}$ can be calculated using Equation 5.7 with $k = 0$ such that:

$$Q_{S,3} = \frac{1}{N-1} \sum_{n=1}^{N} \frac{(x[n] - \bar{x})}{s_x^2} \frac{(y[n] - \bar{y})}{s_y^2} \tag{5.30}$$

where $x[n]$ is the respiratory waveform, $y[n]$ is the ideal cosine waveform, N is the total number of samples in both waveforms, $\bar{x}$ and $\bar{y}$ are the means, and s_x^2 and s_y^2 are the

variances of the respiratory waveform and the ideal cosine waveform, respectively.

Autoregression RQI ($Q_{S,4}$)

The autoregression RQI ($Q_{S,3}$) was calculated by fitting series of AR models to the $N = 128$ points of each respiratory modulation using the Yule-Walker method (Equation 5.11). The AR models are fit for model orders $M = 1 \ldots 30$ and the optimal model order for each individual respiratory modulation is determined via the AIC (Equation 5.12) as recommended by [205]. While other selection criteria exist, AIC was exclusively used in defining $Q_{S,4}$. The transfer function associated with the fitted AR process was used to obtain the power spectral estimate of the process at $i = 481$ evenly spaced points ($X_{i=1\ldots481}$) from 0.0 to 2.0 Hz (the Nyquist frequency of the time-series) using Equation 5.16. Therefore, each X_i represents a frequency band of $\frac{2}{480}$. This selection allows for the exact range of 0.083 Hz to 1.0 Hz (corresponding to 5 to 60 brpm) to be considered. Similar to $Q_{S,1}$, $Q_{S,4}$ was calculated as:

$$Q_{S,4} = \frac{\hat{X}}{\sum_j X_j} \tag{5.31}$$

where:

$$\hat{X} = \max \left\{ \sum_{k=z-4}^{z+4} X_k \middle| (z-4), (z+4) \in [21\ 241] \right\} \tag{5.32}$$

and where $j \in [21\ 241]$. Here the 481-point spectrum has components with indices [21 241] corresponding to the range of plausible respiratory frequencies [0.83 1.00] Hz. $Q_{S,4}$ thus defines the ratio of (the maximum power in any 9-component sub-range, corresponding to a window range of 0.03 Hz, of these respiratory frequencies) to (the total power in the range of respiratory frequencies).

Hjorth Complexity RQI ($Q_{S,5}$)

The Hjorth complexity RQI ($Q_{S,5}$) is calculated based on the equation for the third Hjorth parameter (Equation 5.33). However, because the third Hjorth parameter is large when a waveform is most volatile and is closest to one when it is "smoothest" and most similar to a sine wave, $Q_{S,5}$ was calculated as the inverse of Hjorth complexity:

$$Q_{S,5} = \frac{(m_2/m_0)^{\frac{1}{2}}}{(m_4/m_2)^{\frac{1}{2}}} \tag{5.33}$$

5.2.3 RQI Performance Evaluation Methods

The performances of the individual RQIs previously described ($Q_{S,[1...5]}$) were evaluated to assess both their accuracy and robustness. In this work, accuracy is defined as describing how well the RQIs assess the quality of the respiratory information on average. Robustness is defined as describing the stability of the RQIs in aggregate (i.e., does a high-performing RQI always work well or are there cases where it is errant). Prior to evaluating the RQIs, the estimated RR from each respiratory waveform modulation for both ECG and PPG and for the gold standard respiratory waveform (either capnography or IP) were extracted. Next, two "comparison" RQIs were extracted to evaluate the generalised performance of $Q_{S,[1...5]}$.

The accuracy of the individual RQIs ($Q_{S,[1...5]}$) was measured by assessing the correlation between the RQIs and the RR error. The best individual RQI has the strongest negative correlation between RQI and error. This means that the highest-quality respiratory modulations (signified as being those which have the smallest absolute RR error between the modulation RR estimate and the gold standard reference RR) have the

highest magnitude RQI values. The robustness will be assessed as a measurement of the spread of the accuracy of the RQIs. Both accuracy and robustness are critical measures of an RQI's performance where the best RQIs have optimal accuracy and robustness.

RR Extraction

In order to evaluate the RQIs, it was necessary to obtain known quality labels for the data superset containing both $\mathcal{D}_c$ and $\mathcal{D}_m$. In order to obtain these quality labels, the RR from each extracted respiratory modulation and the RR from the gold standard reference respiratory waveform were estimated. The RR for each modulation, AM_{ECG} ($R_{M,1}$), FM_{ECG} ($R_{M,2}$), BW_{ECG} ($R_{M,3}$), AM_{PPG} ($R_{M,4}$), FM_{ECG} ($R_{M,5}$), and BW_{ECG} ($R_{M,6}$), was extracted using the *ARSpec* RR estimation algorithm described in Shah *et al.* (2015).[1]

The RR from the reference waveforms (R_{ref}) was estimated using the two sets of respiratory annotations described in Section 4.1 for $\mathcal{D}_c$ and in Section 4.2 for $\mathcal{D}_m$. The RR for each set of annotations for each 32s window of the reference waveform was estimated as:

$$RR_{ref,j} = \frac{60(N-1)}{\sum_{n=1}^{N-1}(t_{n+1} - t_n)} \tag{5.34}$$

where j is one of the two sets of annotations, N is the total number of annotations for a given window, t_n and t_{n+1} are the time labels (in seconds) of the n^{th} and $(n+1)^{th}$ annotations, and $RR_{ref,j}$ is in brpm. Using the two sets of RR estimations, RR_{ref} was

[1]The algorithm in Shah *et al.* (2015) was chosen as it has been proven to produce accurate RR estimations from high-quality respiratory waveforms. Furthermore, the algorithm was compared to other commonly-used RR estimation algorithms where the variation between algorithms was negligible. The results from this exploration can be seen in Appendix A.

estimated as:

$$RR_{ref} = \begin{cases} \frac{RR_{ref,1}+RR_{ref,2}}{2} & |RR_{ref,2} - RR_{ref,1}| \ < \ 2 \\ \\ \emptyset & |RR_{ref,2} - RR_{ref,1}| \ \geq \ 2 \end{cases} \tag{5.35}$$

where a value for RR_{ref} was only obtained when the difference between the two RR estimations ($RR_{ref,j}$) is < 2 brpm. Otherwise, the window (and all ECG and PPG modulations derived from the window) is discarded as no reliable reference RR can be obtained. This procedure resulted in 620 windows (98.4%) being retained in $\mathcal{D}_c$ (470, 97.9%, in $\mathcal{D}_{c,1}$ and 150, 100.0%, in $\mathcal{D}_{c,2}$) and 737 (92.7%) in $\mathcal{D}_m$.

Using both RR estimations for each window and modulation, RR_{ref} and $RR_{M,[1...6]}$, the absolute error in the estimation between the reference and the RR derived from the respiratory modulations was estimated as:

$$E_{M,i,j} = |RR_{ref,i} - RR_{M,i,j}| \tag{5.36}$$

where $RR_{ref,i}$ is the reference RR estimate for the i^{th} window, $RR_{M,i,j}$ is the RR estimate for the j^{th} modulation, in the i^{th} window, and $E_{M,i,j}$ is the absolute error in the RR estimates for a given modulation and window.

Comparison RQIs

In order to compare the five individual RQIs, two comparison methods have been developed, $Q_{C,[1...2]}$.

SQI Comparison ($Q_{C,1}$) is an SQI that assesses the quality of an ECG or PPG waveform by estimating the agreement of the detected peaks using two independent

peak detectors. The inclusion of $Q_{C,1}$ as a comparison is intended to emulate a situation in which an SQI would be applied to the data prior to the application of an RQI. If an RQI does not outperform the ability of the SQI to discriminate between high-quality respiratory waveforms, it is assumed that the improvement in RR estimation by the addition of an RQI checkpoint would be negligible as the SQI would eliminate the same waveforms as the RQI.

The SQI that has been applied uses the commonly-used F1 score where two peak detectors have been used to find the peaks of the ECG and PPG ($y_{pk,i,[1...2]}$), and where the set of peaks from one peak detector, $y_{pk,i,1}$, are considered to be the "true" peaks and the set of peaks from the other peak detector, $y_{pk,i,2}$, are the "test" peaks that are compared to the true peaks. For the ECG waveforms the true peaks, $y_{pk,i,1}$, are detected using the search-back algorithm previously described in Section 5.2.1 [49]. The ECG test peaks are detected using an algorithm described by Pimentel *et al.* (2015) and which is based on the Pan-Tompkins algorithm [49, 211]. For the PPG waveforms, the true peaks, $y_{pk,i,1}$, are detected using the peak delineator algorithm previously described in Section 5.2.1 [45]. The test peaks are detected based on the portable detector described by Lao *et al.* (2012) [212].

Using the time locations, $t_{pk,i,2}$, of the test peaks, each test peak is classified as being either a true positive (TP) or a false positive (FP). A true positive is an instance where the time location of the test peak is within ± 150ms of a true peak, and a false positive is an instance where the time location of the test peak is not within ± 150ms of a true peak. Additionally, any true peak can be classed as being a false negative (FN) if one of two conditions exist: (1) there is not a test peak within ± 150 ms of the true peak or (2) there is a test peak within ± 150 ms of the true peak, but there is another true peak

that is closer to the test peak than the current true peak being considered. Therefore, all of the test peaks are classed as either TP or FP and any true peak that is unassociated with a test peak is classed as FN.

Using these classes, the F1 score, which is a measure of how well the two peak detectors agree, can be calculated as a combined measure of the precision and recall. Precision is defined as:

$$P_r = \frac{TP}{TP + FP} \tag{5.37}$$

which is a measure of the portion of how many of the "test" peaks are associated with actual locations of true peaks. This is also termed the positive-predictive value. Recall is defined as:

$$R_e = \frac{TP}{TP + FN} \tag{5.38}$$

where recall is the proportion of true peaks that have been associated with test peaks. This is also termed the sensitivity. Based on these two measures, the F1 score can be calculated as:

$$F1 = 2 \times \frac{P_r \times R_e}{P_r + R_e} \tag{5.39}$$

where the maximum value of the F1 score is $F1 = 1$ when there is perfect precision and recall. The F1 score can be further rearranged and used as a comparison metric to the RQIs:

$$Q_{C,1} = \frac{2TP}{TP + FN + FP} \tag{5.40}$$

where the ECG and PPG waveforms with the best peak agreement have the highest values for $Q_{C,1}$ and are assumed to be the best quality ECG and PPG waveforms.

Ideal Control ($Q_{C,2}$) is an idealised RQI control where the value of $Q_{C,2}$ is the inverse of the calculated RR error, $E_{M,i,j}$ such that:

$$Q_{C,2} = \frac{1}{E_{M,i,j}} \tag{5.41}$$

where the result of this is that the windows with the smallest RR error have the highest value for $Q_{C,2}$. This is considered an idealised control because it emulates the situation where there is perfect negative correlation between decreasing RR error and increasing RQI value. This represents perfect knowledge of the error, and therefore of the reference, which is of course impossible. $Q_{C,2}$ represents the best feasible RQI (which is not realisable for the aforementioned reasons).

Accuracy Measures

The accuracy of the RQIs is a measure of how well each of the individual RQIs predicts the error in the RR estimation. The individual RQIs have been designed such that the highest magnitude RQI values should correlate to RR errors that are close to zero, while RQI values of lower magnitudes (bounded by zero) should correlate to RR errors that are large. Three methods have been used to assess the accuracy of the RQIs: the mean absolute error (MAE), a novel metric called the $50th$ Percentile Mean Error Comparison (E_{50}), based on the MAE at varying points, and Spearman's rank order correlation.

Mean Absolute Error is an aggregate measure of the absolute error between a ground-truth reference value and an estimated value, in the case of this analysis, $RR_{ref,i}$ and $RR_{M,i,j}$, respectively. MAE is calculated using the average error between all values of the ground truth and the estimated value, which in this analysis is the average value

for $E_{M,i,j}$ (defined in Equation 5.36) for all values of i and j such that:

$$MAE = \frac{\sum_{i=1}^{N} \sum_{j=1}^{6} E_{M,i,j}}{N} \tag{5.42}$$

where N is the total number of windows. However, the key question being explored is not how accurate $R_{M,i,j}$ is in comparison to $R_{ref,i}$, which is what $E_{M,i,j}$ assesses, but rather how accurate $R_{M,i,j}$ is in comparison to $R_{ref,i}$ when a given RQI value ($Q_{S,[1...5]}$ or ($Q_{C,[1...2]}$) is high. Therefore, it is necessary to look at the MAE when only particular subsets of the data are considered, where the subsets are selected based on the RQI values such that:

$$MAE_T = \frac{\sum_{i \in d} \sum_{j \in d} E_{M,i,j}}{D} \tag{5.43}$$

where:

$$d = i, j \mid Q_{[S|C],i,j} \geq T \tag{5.44}$$

and where d is the subsets of windows (a particular set of values for i and j that are considered when the RQI value, $Q_{[S|C],i,j}$ for S or C[2], is greater than the threshold, T where $0 \leq T \leq 1$ and where D is the total number of windows for a given value of T. One aspect of the selection of the value for T is that it can be consistently picked such that $\frac{D}{N}$ is a series of values decreasing at an equal increment for increasingly more restrictive values for T. The result of this is a series of values for MAE_T where each change in the value of T represents an equal step in the proportion of the data that were retained, such that each MAE_T can be understood as the MAE value when a set percentage of the data were retained (or, conversely, discarded) where the retention of

[2]In Equation 5.44, $[S|C]$ indicates S or C as opposed to a mathematical conditional statement.

those data is based on the threshold value of $Q_{[S|C],i,j}$ alone.[3]

50*th* Percentile Mean Error Comparison (E_{50}) is a unique metric that has been defined based on the difference between (1) the MAE prior to the application of any threshold, T, and (2) after a threshold, T, has been applied which retains exactly 50% of the data, such that:

$$E_{50} = MAE_{T_1} - MAE_{T_2} \tag{5.45}$$

where MAE_{T_1} is the MAE value where 100% of data are retained (i.e. threshold T_1 is zero) and MAE_{T_2} is the MAE where threshold T_2 retains only 50% of the data. The E_{50} measure is meant to observe the rate at which the breathing error decreases as the RQI values become more stringent across data sets regardless of the starting error in the data set. Furthermore, the standard error of the E_{50} is obtained as the standard error of the difference of means [213]:

$$SE_{50} = \sqrt{\frac{s_{T_1}^2}{N_{T_1}} + \frac{s_{T_2}^2}{N_{T_2}}} \tag{5.46}$$

where $s_{T_1}^2$ and $s_{T_2}^2$ are the sample variances when thresholds T_1 and T_2, respectively, are applied. N_{T_1} and N_{T_1} are the total number of windows when thresholds T_1 and T_2, respectively, are applied.

Spearman's Rank Order Correlation Spearman's rank order correlation is a measure of the correlation, ρ, of two observations in a data set that are expected to be monotonically related but not linearly related. As such it is equivalent to Pearson's

[3]In this analysis, each window (i) and each modulation (j) are considered independently, thus as has been previously noted, the number of data points is 8122 for the $\mathcal{D}_c$ and $\mathcal{D}_m$ superset.

correlation except that, prior to observing the correlation, the values in both sets of data are transformed so the data are ranked in even intervals (typically steps of one) from smallest to largest.[4] In the case of this analysis, the first set of observations is $E_{M,i,j}$ and the second set set of observations is one of the RQIs, $Q_{[S|C],i,j}$. Spearman's rank order correlation was selected because the correlation between the error and the RQIs is expected to be monotonically decreasing (represented by negative values for ρ) where the value of the RQI increases as the value of the error decreases. Furthermore, the RQIs were not designed to linearly correlate to the error and as such, Spearman's rank order correlation is preferred over Pearson's correlation as linearity is not a requirement in the correlation. Spearman's rank order correlation can be calculated as:

$$\rho = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)} \tag{5.47}$$

where $d_i = rg(E_{M,i,j}) - rg(Q_{[S|C],i,j})$, and where $rg(E_{M,i,j})$ and $rg(Q_{[S|C],i,j})$ are the vectors of $E_{M,i,j}$ and $Q_{[S|C],i,j}$ sorted by their rank values (i.e., highest value to lowest value), respectively.[5]

Robustness Measures

While the accuracy measures described in the previous section (MAE, E_{50}, SE_{50} and Spearman's Rank Order Correlation) seek to capture how accurately the RQIs for a window correlate with the known RR error, the robustness measures seek to capture the precision of the RQI values. This can be understood as trying to determine the spread

[4] When two or more values are equal they are assigned equal ranks by averaging the set of ranks that are associated with that set of values the rank ordering then continues as if this set of ranks had been normally assigned (i.e. if the values associated with ranks 5 and 6 are equal, both values are assigned ranks of 5.5 and the next rank assigned is 7). This ensures that the maximum rank value for the two sets of data are the same.

[5] rg indicates a rank-ordered series.

of the RQIs for the entire data set, or a subset of the data sets where the RR errors are very similar. There are two approaches to capturing the robustness of each individual RQI that were used: standard deviation of the MAE, and the spread of the RQI values for data quantiles where quantiles are derived based on RR error.

Standard Deviation of MAE Standard deviation is a common method for looking at the spread of a data set where the goal of the measure is to observe the degree to which individual samples vary with respect to the sample mean. The standard deviation for the MAE is calculated as:

$$\sigma = \sqrt{\frac{\sum_{i \in d} \sum_{j \in d} (E_{M,i,j} - MAE_T)^2}{D - 1}} \tag{5.48}$$

where d, D, and T are defined in equation 5.44 as the subset of the data where a progressively more stringent threshold, T, is applied. It is expected that the standard deviation decreases as the threshold, T, gets more stringent, as this suggests that as the RQI values get higher, they are more selectively choosing the subset of the data with lower RR errors and thus there is a smaller spread of RR errors in the more stringent subset of data.

Average RQI Values for Quantiles The RQIs have been designed so that the increasing magnitude of the RQI correlates with decreasing RR error. It is expected that when the data set is split evenly into five quantiles based on RR error (with quantile 1 representing the data with the highest RR error and quantile 5 representing the data with the lowest RR error) that important trends for the average and standard deviations of the RQI values in each quantile will be observed. It is hypothesised that the average RQI values between quantiles will be monotonically increasing (indicating that as RR error

decreases, RQI value increases). The significance of this trend is tested using Welch's t-test among quantile groups.[6] It is also hypothesised that the standard deviation of the RQI values between quantiles will be of similarly small magnitudes, indicating that the RR error and the RQI values were able to equivalently sort the data by quality. In this analysis the data were split into five quantiles. This selection was made so that a sufficient amount of data were available in each quantile (∼1624 data points are included in each quantile) but also so that the spread of RR errors in the quantiles was small enough that it could be assumed that the variation in the RR errors was nearly negligible. (Some variance in the RR errors in each quantile is expected and this has the effect of slightly increasing the RQI standard deviation higher than it otherwise would be if a subset of data with identical RR errors were available.) The average and standard deviations for each quantile can be calculated as:

$$\bar{Q}_{[S|C],g} = \frac{\sum_{g=1}^{G} Q_{[S|C],g}}{G} \tag{5.49}$$

where g is the varying subset of i and j for each quantile, 1 to 5, and G is the total number of individual values in each quantile. Further, the standard deviation in each quantile can be calculated as:

$$s_{Q_{[S|C],g}} = \sqrt{\frac{\sum_{g=1}^{G}(Q_{[S|C],g} - \bar{Q}_{[S|C],g})^2}{G-1}} \tag{5.50}$$

The methods for comparing the performance of the individual RQIs can be under-

[6] Welch's t-test is preferred in this instance over the more common Student's t-test as the former does not assume equal variance (where in this analysis the hypothesis is specifically that each quantile does not have equal variance). Further, for this analysis, Welch's t-test is only computed between quantiles 1 and 3 and quantiles 3 and 5 for simplicity. Welch's t-test is further described in Welch 1947 [214].

stood to be observing two important features of the RQIs: their accuracy and precision. Accuracy measures assess how well, in aggregate, the RQI values for each RQI method estimate the RR error between the derived respiratory waveform and the reference RR. Accuracy is measured using the MAE, E_{50}, Spearman's Rank Order Correlation, and Welch's t-test for error-derived quantiles. Precision measures how closely RQIs for similar RR errors cluster together, where the stronger the clustering of the RQIs, the more precise the RQI measure. This is measured using the standard deviation of the individual RQIs as well as the magnitude of the standard deviation for the error-derived RQI quantiles.

5.3 Results

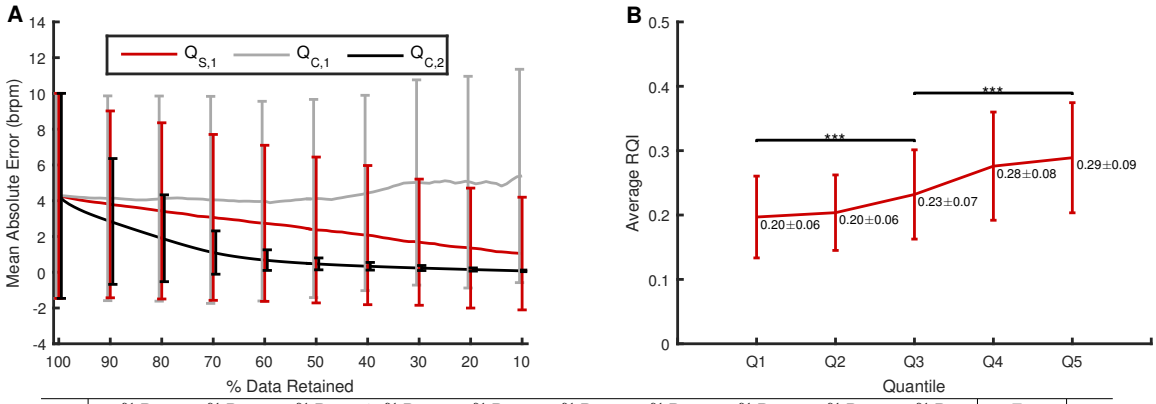
The performance of the individual RQIs ($Q_{S,[1...5]}$) are assessed against the two control RQIs ($Q_{C,[1...2]}$) where the first control, $Q_{C,1}$, represents the ability of a standard SQI applied to either the ECG or PPG and the second control, $Q_{C,2}$, represents the ideal RQI in which the RQI prediction correlates exactly to the RR error. Thus it is expected that the best-performing RQIs will perform more similarly to $Q_{C,2}$ and less similarly to $Q_{C,1}$ and that any RQI that is outperformed by $Q_{C,1}$ is no better than an estimation with no prior knowledge of the respiratory waveform information.

None of the RQIs, ($Q_{S,[1...5]}$), required training and validation to set their parameters, and so the analysis of the individual RQIs was conducted on the combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data set with a total of 8122 unique observations (and where each of the six modulations observed for each window were treated independently). The results are first described for each individual RQI ($Q_{S,[1...5]}$) and then all of the individual RQIs are compared to one another.

5.3.1 FFT RQI ($Q_{S,1}$)

Figure 5.1 shows the results for $Q_{S,1}$ compared to $Q_{C,[1...2]}$ These results show that $Q_{S,1}$ is both an accurate and robust RQI. The MAE of the subset windows monotonically decreases as more windows are progressively discarded based on their RQI values (Figure 5.1A). Furthermore, the E_{50} shows a decrease in MAE of nearly 2 brpm (1.91) when 50% of the windows are retained according to RQI. This is compared to $Q_{C,1}$ where nearly no improvement is seen and $Q_{C,2}$ where the E_{50} is 3.80 (Figure 5.1, table). Further accuracy measures support these findings with Spearman's ρ indicating a moderately strong negative correlation of RQI value to RR error and the average RQI values for five quantiles selected based on RR error showing a monotonic increase in the RQI values as the RR error in the quanta decreases. Additionally, according to Welch's t-test, the difference in RQI value between quantiles 1 and 3 and quantiles 3 and 5 are significant to $p < 0.001$ (Figure 5.1B).

In addition to being an *accurate* RQI measure, $Q_{S,1}$ is also a moderately *robust* measure. Similar to the MAE, standard deviation of the MAE also monotonically decreases when the windows are discarded based on a increasingly more stringent RQI threshold. This indicates that as the RQI threshold gets more stringent, the range of RR errors observed in the remaining windows decreases (Figure 5.1A). Additionally, when the windows are divided into quantiles based on RQI error, it can be observed that the RQI values for each quantiles have similar standard deviation values ranging from 0.06 to 0.09 (Figure 5.1B).



	100% Ret.	90% Ret.	80% Ret.	70% Ret.	60% Ret.	50% Ret.	40% Ret.	30% Ret.	20% Ret.	10% Ret.	E_{50}	ρ
$Q_{S,1}$	4.27 ± 5.73	3.80 ± 5.22	3.43 ± 4.93	3.07 ± 4.64	2.74 ± 4.36	2.36 ± 4.08	2.08 ± 3.89	1.68 ± 3.53	1.35 ± 3.35	1.05 ± 3.15	1.91 ± 0.79	-0.48
$Q_{C,1}$	4.27 ± 5.73	4.14 ± 5.73	4.12 ± 5.74	4.05 ± 5.79	3.98 ± 5.58	4.12 ± 5.54	4.43 ± 5.46	5.02 ± 5.74	5.04 ± 5.92	5.39 ± 5.97	0.15 ± 0.92	-0.10
$Q_{C,2}$	4.27 ± 5.73	2.84 ± 3.52	1.90 ± 2.43	1.10 ± 1.21	0.68 ± 0.58	0.47 ± 0.33	0.34 ± 0.21	0.24 ± 0.14	0.15 ± 0.09	0.08 ± 0.05	3.80 ± 0.61	-1.00

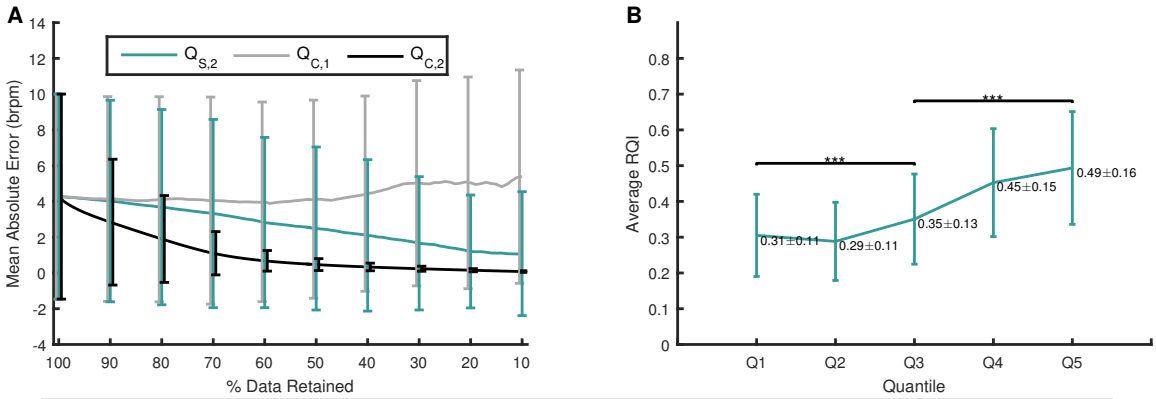
Figure 5.1: *Performance of $Q_{S,1}$ in combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. (A) MAE of $Q_{S,1}$ and controls as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

5.3.2 Autocorrelation RQI ($Q_{S,2}$)

Figure 5.2 shows the results for $Q_{S,2}$ compared to $Q_{C,1}$ and $Q_{C,2}$. Like $Q_{S,1}$, the $Q_{S,2}$ RQI is both accurate and robust. The MAE of $Q_{S,2}$ for higher cut-off thresholds of the RQI is monotonically decreasing (Figure 5.2A) and the E_{50} (1.78) is substantially higher than $Q_{C,2}$ (0.15) (Figure 5.2, table). The average RQIs when the quantiles are sorted by RR error show a generally monotonically increasing trend except for Q1 and Q2 where Q1 (0.31) is slightly larger than Q2 (0.29) (Figure 5.2B). This suggests that $Q_{S,2}$ might not be as effective as other RQIs for the poorest quality windows; however, Welch's t-test between quantiles 1 and 3 and 3 and 5 are significantly different for p<0.001 indicating a generally increasing value for the RQIs across the quantiles.

$Q_{S,2}$ also shows good robustness with the standard deviation of the MAE being

generally monotonically decreasing as more data are discarded (Figure 5.2A). However, this trend is not observed between the 20% and 10% retention rates where the standard deviation grows. This suggests that while both accuracy and precision are generally increasing as fewer data are retained, from the 20% to 10% retention rates, precision does not increase which would be indicative of some windows with high RR error that have high RQI values. The standard deviations of the average RQI values also demonstrate good robustness with a small range of standard deviations across quantiles (from 0.11 to 0.16) (Figure 5.2B).



	100% Ret.	90% Ret.	80% Ret.	70% Ret.	60% Ret.	50% Ret.	40% Ret.	30% Ret.	20% Ret.	10% Ret.	E_{50}	ρ
$Q_{S,2}$	4.27 ± 5.73	4.02 ± 5.64	3.68 ± 5.46	3.32 ± 5.26	2.82 ± 4.76	2.49 ± 4.56	2.10 ± 4.23	1.66 ± 3.73	1.20 ± 3.16	1.08 ± 3.47	1.78 ± 0.83	-0.49
$Q_{C,1}$	4.27 ± 5.73	4.14 ± 5.73	4.12 ± 5.74	4.05 ± 5.79	3.98 ± 5.58	4.12 ± 5.54	4.43 ± 5.46	5.02 ± 5.74	5.04 ± 5.92	5.39 ± 5.97	0.15 ± 0.92	-0.10
$Q_{C,2}$	4.27 ± 5.73	2.84 ± 3.52	1.90 ± 2.43	1.10 ± 1.21	0.68 ± 0.58	0.47 ± 0.33	0.34 ± 0.21	0.24 ± 0.14	0.15 ± 0.09	0.08 ± 0.05	3.80 ± 0.61	-1.00

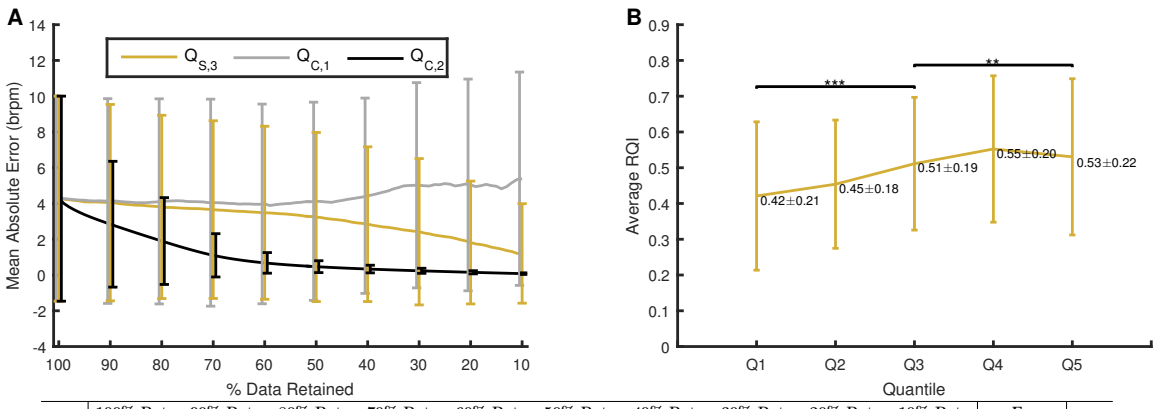
Figure 5.2: *Performance of $Q_{S,2}$ in combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. (A) MAE of $Q_{S,2}$ and controls as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

5.3.3 Cosine-Correlation RQI ($Q_{S,3}$)

Figure 5.3 shows the results for $Q_{S,3}$. These results demonstrate that $Q_{S,3}$ is moderately accurate and lacks robustness. While the MAE for increasingly stringent value for the RQI are applied monotonically decreases (Figure 5.3A) and E_{50} is substantially lower

than $Q_{C,2}$ (Figure 5.3, table), the decreasing trend only shows substantial improvements in the MAE at the $\leq 50\%$ retention rate. Additionally, Spearman's ρ shows only a small correlation between the RQI and the RR error (Figure 5.3, table) and the quantiles selected by RR error show that the average RQI value drops from the quantile 4 to quantile 5 and the difference between quantile 3 and 5 is only significant to $p < 0.01$, which is poor given the size of the data sets (Figure 5.3B).

Additionally, $Q_{S,3}$ lacks robustness. While the standard deviations of the MAE as the RQI threshold is increased show a monotonically decreasing trend (Figure 5.3A), the standard deviation of the quantiles is approximately 50% the size of the average RQI values, indicating a fairly wide spread of RQI values in each quantile (Figure 5.3B).



	100% Ret.	90% Ret.	80% Ret.	70% Ret.	60% Ret.	50% Ret.	40% Ret.	30% Ret.	20% Ret.	10% Ret.	E_{50}	ρ
$Q_{S,3}$	4.27 ± 5.73	4.05 ± 5.49	3.81 ± 5.12	3.66 ± 4.97	3.48 ± 4.84	3.24 ± 4.73	2.84 ± 4.33	2.42 ± 4.09	1.82 ± 3.43	1.21 ± 2.78	1.03 ± 0.85	-0.24
$Q_{C,1}$	4.27 ± 5.73	4.14 ± 5.73	4.12 ± 5.74	4.05 ± 5.79	3.98 ± 5.58	4.12 ± 5.54	4.43 ± 5.46	5.02 ± 5.74	5.04 ± 5.92	5.39 ± 5.97	0.15 ± 0.92	-0.10
$Q_{C,2}$	4.27 ± 5.73	2.84 ± 3.52	1.90 ± 2.43	1.10 ± 1.21	0.68 ± 0.58	0.47 ± 0.33	0.34 ± 0.21	0.24 ± 0.14	0.15 ± 0.09	0.08 ± 0.05	3.80 ± 0.61	-1.00

Figure 5.3: *Performance of $Q_{S,3}$ in combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. (A) MAE of $Q_{S,3}$ and controls as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

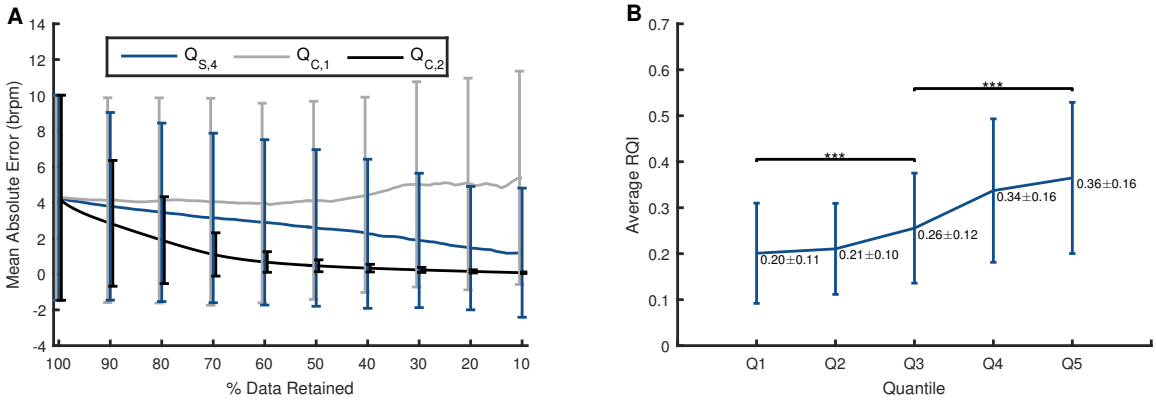
5.3.4 Autoregression RQI ($Q_{S,4}$)

Figure 5.4 shows that $Q_{S,4}$ is accurate and fairly robust. The MAE when the RQI threshold is increased monotonically decreases (Figure 5.4A) and the E_{50} is substantially larger (1.69) than the E_{50} for $Q_{C,1}$ and Spearman's ρ for $Q_{S,4}$ (-0.44). This indicates a moderate negative correlation between RQI and RR error (Figure 5.4, table). Additionally, the trend in the average RQI value when the windows are split into quantiles based on RR errors show a monotonically increasing trend (Figure 5.4B). Interestingly, it appears that the average RQI for quantiles 1 and 2 and quantiles 4 and 5 are quite similar, but the magnitude difference between quantile 2 and quantile 4 is quite substantial. Additionally, Welch's t-test shows that quantiles 1 and 3 and quantiles 3 and 5 are significantly different for $p < 0.001$.

While $Q_{S,4}$ shows good accuracy, the measure appears to be only moderately robust. The trend in the standard deviation of the MAE for more stringent RQI thresholds demonstrates robustness as it is monotonically decreasing, except for the standard deviations between 80% and 90% discard levels (Figure 5.4A). However, the standard deviations of the windows divided into quantiles based on RR error show good robustness for quantiles 1, 2, and 3; it also shows fairly poor robustness for quantiles 4 and 5 as the standard deviation is fairly large compared to the average RQI value for both of these quantiles (Figure 5.4B). This demonstrates a large spread in RQI values for the windows with the smallest RR error.

5.3.5 Hjorth Complexity RQI ($Q_{S,5}$)

Figure 5.5 shows that $Q_{S,5}$ is neither accurate nor robust. The trend in the MAE for increasing the RQI threshold monotonically increases and the value exceeds $Q_{C,1}$ until



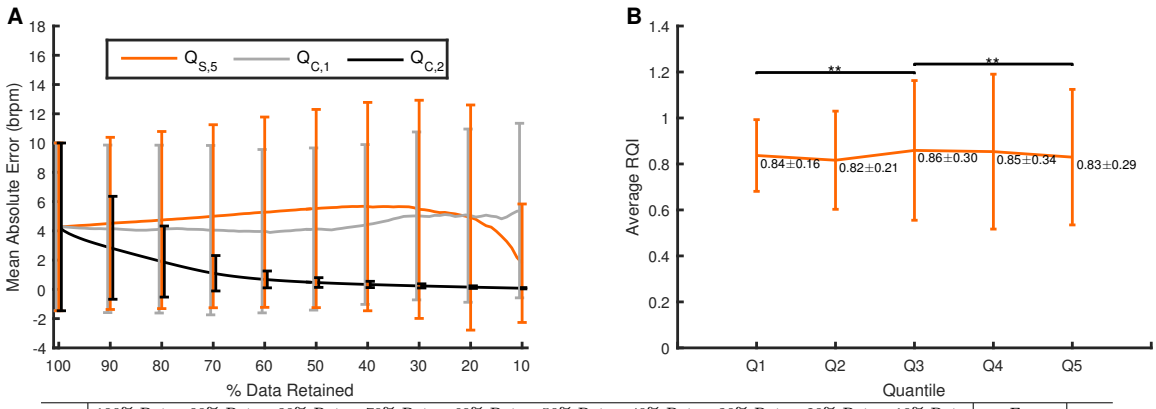
	100% Ret.	90% Ret.	80% Ret.	70% Ret.	60% Ret.	50% Ret.	40% Ret.	30% Ret.	20% Ret.	10% Ret.	E_{50}	ρ
$Q_{S,4}$	4.27 ± 5.73	3.79 ± 5.25	3.46 ± 4.99	3.14 ± 4.74	2.89 ± 4.62	2.58 ± 4.38	2.26 ± 4.17	1.88 ± 3.76	1.45 ± 3.45	1.20 ± 3.61	1.69 ± 0.82	-0.45
$Q_{C,1}$	4.27 ± 5.73	4.14 ± 5.73	4.12 ± 5.74	4.05 ± 5.79	3.98 ± 5.58	4.12 ± 5.54	4.43 ± 5.46	5.02 ± 5.74	5.04 ± 5.92	5.39 ± 5.97	0.15 ± 0.92	-0.10
$Q_{C,2}$	4.27 ± 5.73	2.84 ± 3.52	1.90 ± 2.43	1.10 ± 1.21	0.68 ± 0.58	0.47 ± 0.33	0.34 ± 0.21	0.24 ± 0.14	0.15 ± 0.09	0.08 ± 0.05	3.80 ± 0.61	-1.00

Figure 5.4: *Performance of $Q_{S,4}$ in combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. (A) MAE of $Q_{S,4}$ and controls as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates p <0.001, ** indicates p <0.01, and * indicates p <0.05). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

the 20% retention rate (Figure 5.5A). Additionally, E_{50} is less than $Q_{C,1}$ and negative (-1.25) indicating that the MAE at the 50% retention rate is worse than the MAE when all data are retained. Further, Spearman's ρ does not indicate a trend between RR error and RQI (Figure 5.5, table). Finally, the average RQI value when the windows are divided into quantiles based on RR error shows no trend and indicates that the value of the RQI does not depend on the value of the RR error (Figure 5.5B). Further, the value of Welch's t-test between quantiles 1 and 3 and quantiles 3 and 5 is only significant for $p < 0.01$, which is poor for such a large data set.

In addition to lacking accuracy and having no measurable correlation with the RR error, $Q_{S,5}$ lacks robustness. This is demonstrated by the nearly monotonically increasing value of the standard deviation of the MAE in Figure 5.5A (the exception is between the 20% and 10% retention rates). Additionally, the magnitude of the standard deviation

of the RQI values when the data set is divided into quantiles based on RR error is quite significant, indicating a lack of robustness as the range of RQI values in each quantile is large (Figure 5.5B).



	100% Ret.	90% Ret.	80% Ret.	70% Ret.	60% Ret.	50% Ret.	40% Ret.	30% Ret.	20% Ret.	10% Ret.	E_{50}	ρ
$Q_{S,5}$	4.27 ± 5.73	4.51 ± 5.88	4.74 ± 6.05	5.00 ± 6.26	5.27 ± 6.50	5.52 ± 6.77	5.66 ± 7.12	5.47 ± 7.46	4.91 ± 7.69	1.79 ± 4.05	-1.25 ± 1.04	0.16
$Q_{C,1}$	4.27 ± 5.73	4.14 ± 5.73	4.12 ± 5.74	4.05 ± 5.79	3.98 ± 5.58	4.12 ± 5.54	4.43 ± 5.46	5.02 ± 5.74	5.04 ± 5.92	5.39 ± 5.97	0.15 ± 0.92	-0.10
$Q_{C,2}$	4.27 ± 5.73	2.84 ± 3.52	1.90 ± 2.43	1.10 ± 1.21	0.68 ± 0.58	0.47 ± 0.33	0.34 ± 0.21	0.24 ± 0.14	0.15 ± 0.09	0.08 ± 0.05	3.80 ± 0.61	-1.00

Figure 5.5: *Performance of $Q_{S,5}$ in combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. (A) MAE of $Q_{S,5}$ and controls as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

5.3.6 Inter-RQI Comparison

In addition to the previous five sections (Sections 5.3.1 to 5.3.5) which described the results for each of the individual RQIs ($Q_{S,[1...5]}$) compared to the control RQIs ($Q_{C,[1...2]}$), the following section compares the results of the individual RQIs. The individual RQIs are compared using two analyses: the first is Spearman's Rank order correlation amongst all of the individual RQIs to each other RQI and the control RQIs (Table 5.1) for the combined $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. This analysis demonstrates the similarities between the performances of each individual RQI to each of the other RQIs, where positive values

indicate a strong positive correlation between two RQIs. The results show that $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$ are the most highly correlated individual RQIs, and that $Q_{S,5}$ and $Q_{C,1}$ show very poor correlations with the remaining individual RQIs. This result is consistent with what would be expected based on the performance of the individual RQIs.

Table 5.1: *Spearman's Rank Order Correlations between individual RQIs ($Q_{S,[1...5]}$) and control RQIs ($Q_{C,[1...2]}$).*

	$Q_{S,1}$	$Q_{S,2}$	$Q_{S,3}$	$Q_{S,4}$	$Q_{S,5}$	$Q_{C,1}$	$Q_{C,2}$
$Q_{S,1}$	**1.00**						
$Q_{S,2}$	0.78	**1.00**					
$Q_{S,3}$	0.43	0.38	**1.00**				
$Q_{S,4}$	0.83	0.77	0.43	**1.00**			
$Q_{S,5}$	-0.43	-0.25	-0.27	-0.36	**1.00**		
$Q_{C,1}$	0.01	0.04	0.04	0.03	0.05	**1.00**	
$Q_{C,2}$	0.48	0.49	0.24	0.45	-0.16	0.10	**1.00**

The second analysis compares the performance of the individual RQIs by measuring the frequency with which the individual RQIs (when ranked for the entirety of the $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets) are sufficiently close to the rankings of the windows when they are ranked based on RR error (Figure 5.6). This analysis is done by finding the absolute difference between the ranking of the individual RQI and the RR error and determining the frequency with which the absolute difference of the individual RQI is within 10% of the error (difference in ranking between individual RQI and RR error is less than 812, dark bars of Figure 5.6) and between 20% of the error (difference in ranking between individual RQI and RR error is between 813 and 1624, light bars in Figure 5.6).

These results show that $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$ outperform both $Q_{S,2}$ and $Q_{S,3}$. $Q_{S,4}$ has the highest frequency of windows that fall within 10% of the ranking of the RR error (0.30) and $Q_{S,2}$ has the highest frequency of windows that fall within 20% of the RR error (0.53, both bars combined); however, the performances of $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$ are similar. Additionally, these results show that for every individual RQI, the frequency of

windows within 10% of the RR error is greater than the frequency of windows between 10% and 20% of the error. Furthermore, this trend is preserved for each subsequent 10% band where the frequency at the next 10% interval decreases compared to the previous band as shown in Appendix A. As a result, the frequency for which the rankings of the windows for the individual RQIs is between 80% and 100% away from the RR error (absolute difference values between 6499 and 8122) is extremely small (0.008 for $Q_{S,1}$, 0.008 for $Q_{S,2}$, 0.026 for $Q_{S,3}$, 0.012 for $Q_{S,4}$ and 0.052 for $Q_{S,5}$, with both $Q_{S,3}$ and $Q_{S,5}$ being larger than $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$).

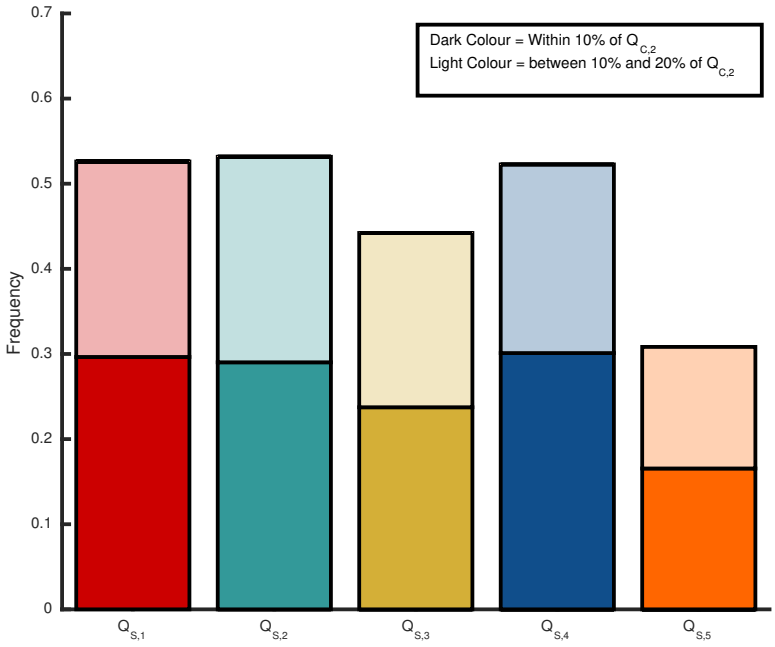


Figure 5.6: *Frequencies in which the rank orderings each of the individual RQIs, $Q_{S,1}$ (red), $Q_{S,2}$ (teal), $Q_{S,3}$ (gold), $Q_{S,4}$ (blue), and $Q_{S,5}$ (orange), falls within the 10% (difference of 0 to 812, dark bands), and 20% (difference between 813 and 1624, light bars) of the rank ordering of the RR error in the data superset containing $\mathcal{D}_c$ and $\mathcal{D}_m$.*

5.4 Discussion

The best-performing RQIs are those that are both accurate and robust. The most accurate RQIs are those that show the strongest ability to estimate which windows likely contain respiratory information and those that do not. The most robust RQIs are

those that show the greatest precision by demonstrating that windows with similar RR error have similar RQI values. Both features are critical for assessing whether a given RQI can be used as a predictive marker for a valid respiratory waveform. This analysis shows that $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$ meet this criteria and are both accurate and robust while $Q_{S,3}$ shows adequate accuracy but lacks sufficient robustness, and $Q_{S,5}$ lacks both accuracy and robustness.

Of the set of RQIs, the most accurate and robust is $Q_{S,1}$; however, $Q_{S,2}$ and $Q_{S,4}$ are only slightly less accurate and robust. The superior accuracy of $Q_{S,1}$ is demonstrated by the slope of the decline in the MAE as the RQI threshold is increased compared to the other RQIs (Figure 5.1A). The accuracy of $Q_{S,1}$ is also demonstrated by having the highest magnitude E_{50} and the second-best Spearman's ρ ($Q_{S,2}$ has a slightly better correlation, -0.49 for $Q_{S,2}$ compared to -0.48 for $Q_{S,1}$). Additionally, $Q_{S,1}$ shows high robustness with the standard deviation of the MAE for $Q_{S,1}$ being the only RQI with a completely monotonically decreasing trend. This is also demonstrated by the relatively small magnitude of the standard deviation of the average RQI when the windows are divided into quantiles based on RR error (Figure 5.1B).

While $Q_{S,1}$ is the best-performing RQI based on accuracy and robustness alone, $Q_{S,2}$ and $Q_{S,4}$ also show high accuracy and robustness and all three of these RQIs vastly outperform $Q_{S,3}$, $Q_{S,5}$, and $Q_{C,1}$. Figure 5.1 shows that the frequency with which $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$ generate an RQI that ranks a window within 20% of the ground truth ranking of the performance of that window based on RR error is above 0.50. Furthermore, the performances of $Q_{S,1}$, $Q_{S,2}$, $Q_{S,4}$, and even $Q_{S,3}$ are a strong indication of the ability of RQIs to assess the quality of the extracted respiratory waveform compared to an SQI ($Q_{C,2}$). This is a strong indication of the value of the addition of an RQI step

after respiratory waveform extraction but prior to RR estimation.

Given that $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$ all provide accurate and robust RQI values, an important point to consider is whether these RQIs can be used singularly or whether their performance improves as a result of fusion. The correlation between each of these RQIs is shown in Table 5.1. These results indicate a strong, but not identical correlation between these three RQIs. This suggests that these RQIs provide similar, reliable RQI estimates for the easiest cases of both very good and very poor respiratory waveforms; however, they are not always in agreement on the most challenging cases. This result and the results from Figure 5.6 suggest that there is still an opportunity to improve these results. The value of a panel of methods has been suggested by the results of Chemla *et al.* (2005) which found that when detecting the RR in diabetic patients, FFT-based methods were more accurate while in other classes of patients, AR-based methods were preferred [215]. Based on the established viability of a panel of methods, the following chapter explores methods by which the three best-performing individual RQIs, $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$, can be fused into one single fusion RQI. This is explored using both supervised and unsupervised learning approaches.

5.5 Conclusion

This chapter presented the results for five individual RQIs, $Q_{S,[1...5]}$ based on the FFT, autocorrelation, cosine-correlation, autoregression, and Hjorth complexity. These RQIs were compared to a standard SQI method, the F1 Score, which was measured prior to respiratory waveform extraction, as well as the ideal RQI performance derived from the RR error. The results clearly showed an improvement of most of the individual RQIs over the SQI. This is a strong indication that the RQIs are better suited to assess

respiratory waveform quality than SQIs, and that implementation of an RQI step after respiratory waveform extraction but prior to RR estimation might improve overall RR estimation techniques. This analysis also assessed the individual RQIs for both accuracy and robustness and the results showed that $Q_{S,1}$, based on the respiratory waveform FFT, was the most accurate and robust method. $Q_{S,2}$ and $Q_{S,4}$ were also highly accurate and robust, and may contain additional information that is not observed in the respiratory waveforms by the FFT in the most challenging cases. As a result of this, it is believed that methods for fusing the best-performing individual RQIs from this chapter into a single fused RQI offers the possibility to further improve RQI performance. The description and analysis of both supervised and unsupervised learning methods for RQI fusion is the subject of the subsequent chapter.

Chapter 6

Fusion RQIs

The prior chapter presented the development of five individual RQIs, $Q_{S,[1...5]}$, derived from common signal processing techniques: FFT, autocorrelation, cross-correlation (cosine-correlation RQI), AR, and Hjorth complexity. Analysis of these RQIs indicated that only the RQIs based on the FFT ($Q_{S,1}$), autocorrelation ($Q_{S,2}$), and autoregression ($Q_{S,4}$) were both accurate and robust, while the cosine correlation RQI ($Q_{S,3}$) and Hjorth complexity RQI ($Q_{S,5}$) lacked the necessary accuracy and robustness to serve as confidence measures that a given respiratory modulation contains physiologically relevant respiratory information.

The results from the previous chapter also indicated that while $Q_{S,[1,2,4]}$ were accurate and robust, they were not identically so. An analysis of Spearman's correlation of these RQIs indicated that the maximum correlation between them was 0.83 ($Q_{S,1}$ and $Q_{S,4}$) and that the minimum correlation was 0.77 ($Q_{S,2}$ and $Q_{S,4}$). However, the performance of these RQIs was similar (as shown in Figure 5.6) which demonstrates that $Q_{S,[1,2,4]}$ have similar frequencies at which the ranking of their RQI values fall within 20% of the "true" RQI ranking from the "ideal" RQI, $Q_{C,2}$. These findings motivate the present chapter which seeks to develop a method to fuse the multiple individual RQIs that have

been obtained for a given time window and RR modulation into a single RQI value. It is hypothesised that fused RQIs may be *less* accurate than the highest-quality individual RQI for any given *single* observation (as the lesser-quality RQIs will also influence the fusion RQI), but generally, RQI robustness will be improved for *all* observations compared to any individual RQI. The loss of accuracy in lieu of improved robustness is the preferred outcome when no *a priori* information on the waveforms, data sets, or patient populations is available prior to using the RQIs in RR estimation.

This chapter presents the development of three fusion RQIs, $Q_{F,1...3}$ based on two supervised fusion methods: linear regression (LR) and support vector regression (SVR) and one unsupervised fusion method: principal component analysis (PCA). This chapter will first describe the general techniques of these three methods; the specific methods that will be used to fuse $Q_{S,[1,2,4]}$ using these techniques will follow, and the results of each fusion method and a discussion will conclude this chapter.

6.1 Techniques

6.1.1 Linear Regression

Linear regression is a common supervised machine learning technique that uses a training set of observations, $\{x_{n,1} \dots x_{n,D}\}$ to predict a target variable, t_n where $n = 1 \dots N$ is the number of specific instances in the training set and D is the specific set of multi-dimensional observations used to predict each value of t_n [216]. While there are many variations of linear models, the key feature of a linear regression model is that the function, y, that is used to estimate t is based on a linear set of parameters, w, and the values of x. However, linear regression models do not require that the model be

defined based on a linear relationship of the values of x. Many linear regression models transform the values of x using basis functions such that x is defined in the model as $\phi(x)$. Thus, the simplest linear regression models can be represented as:

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^{M} w_j \phi_j(\mathbf{x}) \tag{6.1}$$

where $y(x, w)$ is the function that predicts t, M is the total number of parameters in the model, w_j are the model parameters, and $\phi_j(x)$ represents each x transformed by the chosen basis function. Some of the most common basis functions are the linear, polynomial, sigmoidal, logistic sigmoid, Fourier, and wavelet functions [216].

Additionally, many linear regression models incorporate a bias parameter, w_0, which adjusts for a fixed offset between the observed values for x and t. The bias parameter is most often incorporated into the model as a dummy basis function where $\phi_0(x) = 1$ and the linear regression model becomes:

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\mathbf{x}) \tag{6.2}$$

The notation for the linear regression function can be simplified using vector notation where $\mathbf{w} = (w_0, \ldots, w_{M-1})^{\mathrm{T}}$ and $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_{M-1})^{\mathrm{T}}$ and thus

$$y(\mathbf{x}, \mathbf{w}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}) \tag{6.3}$$

Based on 6.3, the goal of linear regression is to find the optimal values for $\mathbf{w}$ such that the difference between $y(\mathbf{x}, \mathbf{w})$ and t_n is minimised [216]. A number of error functions can be used to examine this relationship, but one of the most commonly used is the

sum-of-squares error function which is defined as

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 \tag{6.4}$$

and where the gradient of the sum-of-squares error function is

$$\nabla E_D(\mathbf{w}) = \sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}\boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \tag{6.5}$$

Therefore, the sum-of-squares function can be minimized when $\nabla E_D(\mathbf{w}) = 0$ and the values for $\mathbf{w}$ can be found by solving this function which when rearranged shows

$$\mathbf{w}^{\mathrm{T}} = \left(\sum_{n=1}^{N}\boldsymbol{\phi}(\mathbf{x}_n)\boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}}\right)^{-1} t_n\boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \tag{6.6}$$

and when $\boldsymbol{\phi}(\mathbf{x}_n) = \boldsymbol{\Phi}$, which is an $N \times M$ matrix, called the design matrix, and $t_n = \mathbf{t}$, the vertical vector of target values, becomes

$$\mathbf{w}^{\mathrm{T}} = (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t} \tag{6.7}$$

However, one of the problems with solving linear regression models using 6.7 is over-fitting the model to the training set [216]. This is most often solved by using regularised least-squares which incorporates a regularisation coefficient, λ, as an additional error term,

$$E_W(\mathbf{w}) = \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} \tag{6.8}$$

The use of $E_W(\mathbf{w})$ is known as weight decay because depending on the value of λ, it forces the values of $\mathbf{w}$ towards 0 unless supported by the data. Ultimately, the use of

$E_W(\mathbf{w})$ updates the overall error function to

$$E(\mathbf{w}) = E_D(\mathbf{w}) + E_W(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} \tag{6.9}$$

When the gradient of 6.9 is set to 0 and solved for $\mathbf{w}^{\mathrm{T}}$, the updated equation for $\mathbf{w}^{\mathrm{T}}$ is

$$\mathbf{w}^{\mathrm{T}} = (\lambda\mathbf{I} + \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{t} \tag{6.10}$$

However, the introduction of λ now means that the linear regression model is a system of equations with one more variable than there are equations [216]. This challenge can be overcome by subdividing the training set into training and validation subsets where the training set is use to train a discrete number of linear regression models for a discrete set of values for λ. Each of these models is applied to the validation set and the model with the smallest $E(\mathbf{w})$ on the validation set is used to select the ideal parameters for $\mathbf{w}$ and λ. This is the preferred method as it allows λ to serve as a penalty term for overfitting the model to the training set but also tests the overfitting assumption of λ on the independent validation set. This training-validation technique can be further extended using cross-fold validation.

6.1.2 Support Vector Regression

SVR, similar to linear regression, is a supervised machine learning technique that uses a set of training observations, $\{x_{n,1} \ldots x_{n,D}\}$ to predict the target variable, t_n and obtain a regression function, $y(n)$. However, unlike linear regression, the support vector regression solution is a sparse solution, meaning that only a particular subset of values from $\{x_{n,1} \ldots x_{n,D}\}$ is used to obtain the regression function, $y(n)$ [216]. The benefit of the

sparse solution is that it improves the regression function when the actual relationship between the observations is loosely-coupled. However, because the specific subset of values from $\{x_{n,1} \dots x_{n,D}\}$ that is used to train the model are only those that are outside a certain error margin of the regression function, $y(n)$, support vector regression models have a tendency to overfit the training set.

Compared to Equation 6.9, the error function for linear regression, the error function in support vector regression can be obtained by replacing the least-squares error term in the linear regression error function, $\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2$, with the ϵ-insensitive error function, $E_\epsilon(y(x_n) - t)$, which can be defined as [217]

$$E_\epsilon(y(x) - t) = \begin{cases} 0, & \text{if } |y(x) - t| < \epsilon; \\ |y(x) - t| - \epsilon, & \text{otherwise} \end{cases} \tag{6.11}$$

such that the SVR error is defined as:

$$E = C \sum_{n=1}^{N} E_\epsilon(y(x_n) - t_n) + \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} \tag{6.12}$$

where $y(x)$ is given by Equation 6.2 and C is the inverse of the previous regularisation parameter, λ, and has been moved outside of the error term by convention [216].

Based on Equations 6.11 and 6.12, the SVR function must be solved by optimising both ϵ and C. However, the system of equations cannot be solved by simple substitution because Equation 6.11 is an inequality equation which must be transformed into an equality equation using slack variables. This can be done by transforming and simplify-

ing the inequality in Equation 6.11, $|y(x) - t| < \epsilon$, into its two component equations:

$$t_n < y(x) + \epsilon \tag{6.13}$$

$$t_n > y(x) - \epsilon \tag{6.14}$$

and introducing two slack variable, ξ_n and $\hat{\xi}_n$, where $\xi_n > 0$ only when $t_n > y(x) + \epsilon$ and is zero otherwise, and $\hat{\xi}_n > 0$ only when $t_n < y(x) - \epsilon$ and is zero otherwise [216]. Thus, considering the constraints on ξ_n and $\hat{\xi}_n$ and Equations 6.13 and 6.14, the entire set of values for t_n can be described by the two equations:

$$t_n \leq y(x) + \epsilon + \xi_n \tag{6.15}$$

$$t_n \geq y(x) - \epsilon + \hat{\xi}_n \tag{6.16}$$

These two slack variables, which define the absolute distance t_n can both lie outside of the ϵ-tube. They can be used to substitute Equation 6.11 into Equation 6.12 to remove the inequality and make the SVR error function solvable such that:

$$E = C \sum_{n=1}^{N} (\xi_n + \hat{\xi}_n) + \frac{1}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} \tag{6.17}$$

However, 6.17 must be minimised with respect to four constraints: (1) $\xi_n \geq 0$, (2) $\hat{\xi}_n \geq 0$, (3) $\epsilon + \xi_n + y(x) - t_n \geq 0$ from Equation 6.15, and (4) $\epsilon + \hat{\xi}_n - y(x) - t_n \geq 0$ from Equation 6.16. Given these constraints, the optimisation problem can be solved

using Lagrange multipliers, $\mu_n \geq 0$, $\hat{\mu}_n \geq 0$, $a_n \geq 0$, and $\hat{a}_n \geq 0$, such that:

$$L = C\sum_{n=1}^{N}(\xi_n + \hat{\xi}_n) + \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} - \sum_{n=1}^{N}\mu_n\xi_n - \sum_{n=1}^{N}\hat{\mu}_n\hat{\xi}_n \\ - \sum_{n=1}^{N}a_n(\epsilon + \hat{\xi}_n - y(x) - t_n) - \sum_{n=1}^{N}\hat{a}_n(\epsilon + \hat{\xi}_n - y(x) - t_n) \quad (6.18)$$

where the equation $y(x) = \mathbf{w}^{\mathrm{T}}\phi(x) + b$ is substituted for $y(x)$ (where this equation is identical to Equation 6.3 except that the bias parameter, b, has been explicitly removed from the vector calculations) [216]. This allows for the minimum of the Lagrangian derivatives to be found where:

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \sum_{n=1}^{N}(a_n - \hat{a}_n)\phi(x) = \mathbf{w} \quad (6.19)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{n=1}^{N}(a_n - \hat{a}_n) = 0 \quad (6.20)$$

$$\frac{\partial L}{\partial \xi_n} = 0 \rightarrow a_n + \mu_n = C \quad (6.21)$$

$$\frac{\partial L}{\partial \hat{\xi}_n} = 0 \rightarrow \hat{a}_n + \hat{\mu}_n = C \quad (6.22)$$

Using Equations 6.19, 6.20, 6.21, and 6.20, Equation 6.18 can be simplified to:

$$L = -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}(a_n - \hat{a}_n)k(x_n, x_m) - \epsilon\sum_{n=1}^{N}(a_n - \hat{a}_n) + \sum_{n=1}^{N}(a_n - \hat{a}_n)t_n \quad (6.23)$$

where the kernel $k(x_n, x_m)$ has been substituted for $\phi(x)^{\mathrm{T}}\phi(x')$ by Mercer's theorem. Additionally, given the constraints on the Lagrange multipliers and Equations 6.21 and

6.22, a_n and $\hat{a}_n$ are further constrained such that $0 \leq a_n \leq C$ and $0 \leq \hat{a}_n \leq C$. Therefore, based on the Karush-Kuhn-Tucker conditions [216] the following equalities must hold:

$$a_n(\epsilon + \xi_n + y_n - t_n) = 0 \tag{6.24}$$

$$\hat{a}_n(\epsilon + \hat{\xi}_n + y_n + t_n) = 0 \tag{6.25}$$

$$(C - a_n)\xi_n = 0 \tag{6.26}$$

$$(C - \hat{a}_n)\hat{\xi}_n = 0 \tag{6.27}$$

The condition in Equation 6.24 demonstrates that a_n is only non-zero when $\epsilon + \xi_n + y_n - t_n = 0$, indicating a condition where t_n lies either on or above the ϵ-tube (where the ϵ-tube is the range of values $y_n \pm \epsilon$). Conversely, the condition in Equation 6.25 shows that $\hat{a}_n$ is only non-zero when $\epsilon + \xi_n - y_n + t_n = 0$, the condition where t_n lies on or below the ϵ-tube. Furthermore, given the specific constraints on a_n and $\hat{a}_n$ being non-zero, there exists no situation in which both a_n and $\hat{a}_n$ are both non-zero for any specific t_n. Furthermore, only when either a_n or $\hat{a}_n$ is non-zero is t_n one of the support vectors used to solve the SVR. Thus when a_n is non-zero, t_n lies either above or on the upper border of the ϵ-tube, when a_n and $\hat{a}_n$ are both zero, t_n lies within (but not on) the ϵ-tube, and when $\hat{a}_n$ is non-zero, t_n lies either below or on the lower border of the ϵ-tube. Thus the inequality constraints described in Equations 6.15 and 6.16 have been transformed into equality constraints [216].

The regression function can be rewritten as a function of a_n and $\hat{a}_n$ using Equation 6.19 such that:

$$y(x) = \sum_{n=1}^{N}(a_n - \hat{a}_n)k(\mathbf{x}, \mathbf{x}_n) + b \tag{6.28}$$

where b can be solved by substitution using either Equation 6.24 or 6.25 such that

$$b = t_n - \epsilon - \sum_{m=1}^{N}(a_m - \hat{a}_m)k(\mathbf{x}_n, \mathbf{x}_m) \tag{6.29}$$

This means that the SVR model can be solved by determining the value for either a_n and $\hat{a}_n$ for only those values of t_n that lie outside of the ϵ-tube. Similar to the linear regression, the model penalty term, C must be determined prior to solving the model and is often solved empirically through testing a range of values. Additionally, ϵ must also be determined prior to solving the model where many authors have suggested methods for determining the ideal value for ϵ [218, 219, 220, 221].

Similar to the linear regression model, the SVR can be used to model non-linear relationships between $\mathbf{x}$ and t_n using kernel functions. The most common kernel functions for SVR are the radial basis function (RBF), the polynomial function, and the sigmoid function where $\phi(x)$ represents the kernel transformed matrix of values for $\mathbf{x}$ [221]. The most common kernel function for SVR is the RBF which is defined as:

$$\phi(x) = e^{-\gamma|\mathbf{x}_i - \mathbf{x}_j|} \tag{6.30}$$

where γ is the kernel parameter and where $\gamma > 0$. The RBF kernel is often used because the situation where $\phi(x) = x$, the linear kernel, is a special case of the RBF kernel and the sigmoid kernel has similar performance to the RBF kernel [222]. Additionally, the RBF kernel is often preferred over the polynomial kernel, as the polynomial kernel requires two kernel parameters where the RBF kernel only requires a single kernel parameter [221].

The introduction of a kernel function in the SVR requires further model optimisation because, in addition to determining the SVR model parameters, C and ϵ, the kernel parameters must also be determined. The number of model parameters that must be empirically determined in the SVR introduces a higher likelihood than simpler models for overfitting the SVR model to the training data compared to the test data. This can be overcome by careful selection of the model parameters through testing of both the model and kernel parameters using both grid-search and cross-validation [221].

A grid-search is conducted by solving the SVR for a range of values for necessary model and kernel parameters where each potential value for one parameter is tested for every possible value of the other parameters. The optimal parameters are determined as those that result in the model with the smallest value for the loss function described in Equation 6.23 for a held-out validation set.

The robustness of the grid-search method can be improved by introducing cross-validation where the training data are subdivided into n folds where every permutation of $n-1$ folds are used to train the SVR model using the grid-search method and the held out fold is used to determine the values for the loss function. The grid of loss function values for each held-out fold are summed, resulting in a combined grid-search loss function matrix which can be used to select the optimal model and kernel parameters. The goal of both grid-search and cross-validation in training the SVR model is to reduce the probability of model overfitting and increase SVR robustness [221].

6.1.3 Principal Component Analysis

Unlike linear regression and SVR, PCA is an unsupervised method for dimensionality reduction. Where LR and SVR used a set of observations, $\{x_{n,1} \dots x_{n,D}\}$ to predict

the target variable, t_n, the goal of PCA is to reduce and transform $\{x_{n,1} \dots x_{n,D}\}$ to $\{x'_{n,1} \dots x'_{n,M}\}$ where $D \geq M$ and where each transformed value, $x'_{n,M}$ is a projection of $\{x_{n,1} \dots x_{n,D}\}$ onto a D-dimensional vector. The first of these vectors is the first principal component (PC) and is defined as the vector that encompasses the maximum variance for the entire sets of observations [216]. Each subsequent PC is orthogonal to all previous PCs and explains the maximal variance remaining between observations. Thus, if $M = D$, M is a D-dimensional space in which all of the observations have only been translated (as opposed to having reduced dimensionality).

Mathematically, PCA can be solved via the maximal variance formulation or the minimum-error formulation where the results from both methods yield the same PCs [216]. For simplicity, only the maximal variance formulation will be considered here. Like the previous methods, PCA uses the set of observations, $\{x_{n,1} \dots x_{n,D}\}$ where D describes the multi-dimensional observations for a single instance. However, unlike the previously described methods, PCA is an unsupervised method where the target variable, t_n, is not required to find the PCs. PCA is an iterative process where each PC is found based on three requirements: (1) each new PC must be orthogonal to all previous PCs, (2) each new PC is optimised to find the maximal variance among the data, and (3) the total number of PCs cannot exceed D, the dimensionality of the original data. Thus the first step in PCA is to find the first PC where the only requirement is that the PC must be optimised to find the maximal variance among the data as the other two requirements are inherently met. If $\mathbf{u}_1$ is a D-dimensional unit vector (such that $\mathbf{u}_1\mathbf{u}_1^{\mathrm{T}} = 1$), $\mathbf{u}_1$ is the first PC if and only if the variance between $\mathbf{x}$ is maximised when projected onto $\mathbf{u}_1$.

This requires finding the covariance matrix, S, of $\mathbf{x}$ which is defined as:

$$S = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^{\mathrm{T}} \tag{6.31}$$

where N is the total number of observations, $\mathbf{x}_n$ is the vector of observations for a single instance, and $\bar{\mathbf{x}}$ is a vector of the sample mean of each observation type for every instance defined as:

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n \tag{6.32}$$

Given both the sample mean, $\bar{\mathbf{x}}$ and the co-variance matrix, S, the variance of the data projected onto $\mathbf{u}_1$ must be maximised where the variance of the projected data are defined as:

$$V = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{u}_1^{\mathrm{T}}\mathbf{x}_n - \mathbf{u}_1^{\mathrm{T}}\bar{\mathbf{x}}) = \mathbf{u}_1^{\mathrm{T}}\mathbf{S}\mathbf{u}_1 \tag{6.33}$$

Given the definition of the variance of $\mathbf{x}_n$ projected onto $\mathbf{u}_1$, the first PC is found by maximising the variance, $\mathbf{u}_1^{\mathrm{T}}\mathbf{S}\mathbf{u}_1$, with respect to $\mathbf{u}_1$; however, $\mathbf{u}_1$ must be constrained such that $|\mathbf{u}_1| < \infty$. As previously stated, this can most easily be done by ensuring that $\mathbf{u}_1\mathbf{u}_1^{\mathrm{T}} = 1$. This can be done by introducing a Lagrange multiplier, α_1, and maximising the constrained equation:

$$\mathbf{u}_1^{\mathrm{T}}\mathbf{S}\mathbf{u}_1 - \alpha_1(1 - \mathbf{u}_1\mathbf{u}_1^{\mathrm{T}}) \tag{6.34}$$

taking the derivative of this equation with respect to $\mathbf{u}_1$ and finding the zero-crossing points gives the equation:

$$\mathbf{S}\mathbf{u}_1 - \alpha_1\mathbf{u}_1 = 0 \tag{6.35}$$

which can be simplified to:

$$\mathbf{u}_1^{\mathrm{T}}\mathbf{S}\mathbf{u}_1 = \alpha_1 \tag{6.36}$$

Based on Equation 6.36, the variance is optimised when $\mathbf{u}_1$ is equivalent to the eigenvector with the largest eigenvalue, α_1. Given the first PC, $\mathbf{u}_1$, every subsequent PC up to D can be found where the additional constraints previously described are met. The result is a matrix of PCs (eigenvectors), $\mathbf{u}_{1\ldots M}$ and a vector of eigenvalues, $\alpha_{1\ldots M}$, where $M \leq D$ and where the eigenvalue of each PC describes the total variance described by that PC. Based on this understanding, there are numerous methods to conduct the PCA eigenvector and eigenvalue decomposition [216].

6.2 Methods

The goal of RQI fusion presented in this chapter is to combine the three best individual RQIs from the previous chapter, the FFT RQI ($Q_{S,1}$), the autocorrelation RQI ($Q_{S,2}$), and the autoregression RQI ($Q_{S,4}$), into a single value which accurately and robustly estimates the quality of the underlying respiratory signal. Three different techniques (as described in Section 6.1) were used to create fusion RQIs: linear regression, support vector regression, and principal component analysis resulting in three unique fusion RQIs, the LR-based RQI ($Q_{F,1}$), the SVR-based RQI ($Q_{F,2}$), and the PCA-based RQI ($Q_{F,3}$). The following section describes the methods that were used to implement each fusion RQI including: (1) the pre-processing of the individual RQIs and data sets prior to application of the fusion RQIs, (2) the description and application of the fusion RQIs, (3) post-processing of the fusion RQIs, and (4) the methods implemented to assess the individual RQIs.

6.2.1 Pre-Processing

Prior to the implementation of the RQI fusion methods, two pre-processing steps were required: (1) division of the data sets into training, validation, and test sets; and (2) standardisation of the individual RQIs.

The first pre-processing step was the division of the data sets into training, validation, and test sets. Unlike in Chapter 5 where all of the data sets were treated as one large combined data set for analysis, division of the data sets is essential in the creation of the fusion RQIs because the fusion RQIs have parameters that must be optimised using available data. It is crucial that the data be subdivided into training, validation, and test sets prior to determining the optimal parameters for the fusion models as this will ensure optimal generalisability of the fusion models.

As was previously described in Sections 4.1 and 4.2, the CapnoBase and MIMIC-III data sets were subdivided into three data sets, $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, and $\mathcal{D}_m$, where $\mathcal{D}_{c,1}$ consists of 32 subjects (21 paediatric, 11 adult) from the CapnoBase database, $\mathcal{D}_{c,2}$ consists of 10 subjects (8 paediatric, 2 adult) from the CapnoBase database, and $\mathcal{D}_m$ consists of 53 subjects from the MIMIC-III database. Additionally, because linear regression and support vector regression are both supervised learning methods, $\mathcal{D}_{c,1}$ was divided further for cross-validation. For creation of the cross-validation folds, each time window and respiratory modulation was treated as a separate observation with three individual RQI measurements ($Q_{S,[1,2,4]}$) for each observation. This resulted in a total of 2816 observations for $\mathcal{D}_{c,1}$. The set $\mathcal{D}_{c,1}$ was randomly subdivided into five folds, $\mathcal{D}_{c,1,1\ldots5}$ with each fold containing 563 observations ($\mathcal{D}_{c,1,5}$ contained 564 observations). These subdivided folds were used to create five different training ($\mathcal{D}_{c,1,t_{1\ldots5}}$) and validation

($\mathcal{D}_{c,1,v_{1\ldots5}}$) sets where:

$$\mathcal{D}_{c,1,t_n} = \mathcal{D}_{c,1} \setminus \mathcal{D}_{c,1,n} \tag{6.37}$$

and

$$\mathcal{D}_{c,1,v_n} = \mathcal{D}_{c,1,n} \tag{6.38}$$

where $n = 1 \ldots 5$, $\mathcal{D}_{c,1,t_n}$ is a training data set consisting of the combination of four folds and $\mathcal{D}_{c,1,v_n}$ is a validation data set consisting of the held-out fold.

The second data pre-processing step was normalisation of the individual RQIs. While by design the individual RQIs range from 0 to 1 with lower values indicating poorer quality respiratory information, normalisation prior to fusion was implemented to ensure that none of the RQIs were over- or under-weighted due to variations on the range and magnitude of the RQIs compared to each other. This was done using a zero-mean, unit-variance normalisation where:

$$Q'_{S,i,j} = \frac{(Q_{S,i,j} - \bar{Q}_{S,i})}{\sigma_{Q_{S,i}}} \tag{6.39}$$

where i represents the individual RQI, j represents a specific modulation and window, $Q'_{S,i,j}$ is the normalised RQI, $Q_{S,i,j}$ is the original RQI, $\bar{Q}_{S,i}$ is the mean of all of the RQIs for a given individual RQI, and $\sigma_{Q_{S,i}}$ is the standard deviation for a given individual RQI. Additionally, both $\bar{Q}_{S,i}$ and $\sigma_{Q_{S,i}}$ were determined based on $\mathcal{D}_{c,1}$ alone and used for the normalisation of $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, and $\mathcal{D}_m$.

6.2.2 Fusion RQIs

Using the pre-processed individual RQIs, $Q'_{S,1}$, $Q'_{S,2}$, and $Q'_{S,4}$ as the inputs, the three fusion methods were applied to obtain three fusion RQIs, LR RQI ($Q_{F,1}$), SVR RQI

($Q_{F,2}$), and PCA RQI ($Q_{F,3}$). The implementation of these methods as RQIs is described in the following sections.

LR RQI ($Q_{F,1}$)

As described in Section 6.1.1, LR requires four important components: (1) a training set of observations, (2) a set of target variables, (3) selection of a basis function, and (4) optimisation of the regularisation coefficient (α). For development of $Q_{F,1}$, the training set of observations was the matrix of normalised individual RQIs $[Q'_{S,1,j}Q'_{S,2,j}Q'_{S,4,j}]$ previously described and the target variable was the RR error ($E_{M,i,j}$) defined in Equation 5.36. Additionally, the linear basis function, where the individual RQIs were not transformed, was selected. Finally, a range of 16 heuristically-obtained values for the regularisation coefficient were used. These values were determined through experimentation as those in which $E(\mathbf{w})$ was minimised within the bounds of the range of regularisation coefficients tested for all of the models trained. The value for the regularisation parameters are $\alpha_1 = 0$, $\alpha_2 = 0.001$, and the remaining fourteen values were defined as:

$$\underset{n=3\ldots16}{\alpha_n} = 2\alpha_{n-1} \tag{6.40}$$

In order to determine the optimal regularisation parameter, an LR model was trained on each training set created using cross-validation, $\mathcal{D}_{c,1,t_n}$ for $n = 1\ldots5$, for each regularisation coefficient, $\alpha_{1\ldots16}$. The specific LR model for each training set and regularisation coefficient was applied to the associated validation set, $\mathcal{D}_{c,1,v_n}$ for $n = 1\ldots5$, and the sum-of-squares error defined in Equation 6.9 was calculated for each model. The error term obtained for each model on each validation set was summed for each optimisation

parameter where the optimisation parameter was determined as the value that minimised the sum of errors across all five validation sets. This optimisation parameter was then applied to the entire $\mathcal{D}_{c,1}$ in order to obtain the LR RQI coefficients, $\mathbf{w}$.

SVR RQI ($Q_{F,2}$)

Like LR and as described in Section 6.1.2, SVR requires the selection of a number of key parameters: (1) the training set of observations, (2) the set of target variables, (3) selection of the kernel function, and (4) selection or optimisation for C, ϵ, and the kernel parameters. As in LR, the training set of observations was $[Q'_{S,1,j}Q'_{S,2,j}Q'_{S,4,j}]$ and target variable was the RR error ($E_{M,i,j}$). The RBF kernel was selected for the reasons described in Section 6.1.2 which requires optimisation for a single kernel parameter, γ. $\epsilon = 0.001$ as suggested by [221] and thus the SVR model only needed to be optimised for two parameters, C and γ. This optimisation was performed using grid-search.

The grid search method used a coarse grid search paired with a fine grid search centred around the best value from the coarse grid search. The range of values for the initial course grid search for C and γ were set as:

$$\underset{a=-5...15}{C_a} = 2^a \tag{6.41}$$

and

$$\underset{b=-15...5}{\gamma_b} = 2^b \tag{6.42}$$

where the range for a and b were suggested by [221, 223]. The 441 combinations for C and γ were applied to each of the five training sets, $\mathcal{D}_{c,1,t_n}$, and the resulting SVR models were applied to the held-out validation sets, $\mathcal{D}_{c,1,v_n}$, and the optimal values for

C and γ from the coarse grid search were determined as those which minimised the sum of the errors across all five validation sets. Using the optimal values for a and b from the coarse grid search, denoted as a^* and b^*, a fine grid search was applied where:

$$\underset{n=\{-0.75,-0.50,-0.25,0,0.25,0.50,0.75\}}{C_{a^*+n}} = 2^{a^*+n} \tag{6.43}$$

and

$$\underset{n=\{-0.75,-0.50,-0.25,0,0.25,0.50,0.75\}}{\gamma_{b^*+n}} = 2^{b^*+n} \tag{6.44}$$

Using the same method as the coarse grid search, the 49 combinations of values for C and γ were used to train SVR models on all of the training sets, $\mathcal{D}_{c,1,t_n}$, and each model was applied to its associated held-out validation set, $\mathcal{D}_{c,1,v_n}$. The optimal values for C and γ, denoted as C_{opt} and γ_{opt} were determined as those which minimised the sum of the errors across all five validation sets. Using these parameters for C and γ, the final fusion SVR RQI was trained on the entire $\mathcal{D}_{c,1}$ data set.

PCA RQI ($Q_{F,3}$)

Unlike LR and SVR, PCA is an unsupervised learning method meaning that the only required input for PCA is the training set of observations, $[Q'_{S,1,j}Q'_{S,2,j}Q'_{S,4,j}]$. Additionally, because PCA does not require learning of any model-specific parameters, training using cross-fold validation is unnecessary. Therefore, the PCA decomposition of $\mathcal{D}_{c,1}$ was obtained using the maximal variance technique described in Section 6.1.3. The transformation of $[Q'_{S,1,j}Q'_{S,2,j}Q'_{S,4,j}]$ using the first PC eigenvector was taken to be the PCA RQI ($Q_{F,3}$).

6.2.3 Post-Processing

After training of the three fusion RQI models, $Q_{F,1}$, $Q_{F,2}$, and $Q_{F,3}$, two post-processing steps were required: (1) application of the models to the test sets, $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$, and transformation of the fusion RQIs into standardised values.

The first post-processing step was the application of the optimised models to the two test sets. The models that were obtained in Section 6.2.2 using $\mathcal{D}_{c,1}$ and the optimised parameters for α ($Q_{F,1}$), C, γ ($Q_{F,2}$), and the eigenvector of the first PC ($Q_{F,3}$) were applied to the two test data sets.

In addition to applying the fusion models to the test data sets, it was necessary to standardise the outputs for the fusion RQI models. This was necessary for two reasons: first, the models were trained to estimate RR error, E, such that a low value for E maps to a low value for $Q_{F,i}$. However, the RQIs are meant to be inversely related to the E such that low values of E map to high values for $Q_{F,i}$. Second, the fusion models did not bound the RQI output such that the range of RQI values could theoretically range from $-\infty$ to ∞ which makes the RQI values more difficult to interpret. Therefore, both $Q_{F,1}$ and $Q_{F,2}$ were transformed to their additive inverses and all three fusion RQIs were transformed using a logistic function.[1] The RQI standardisation transformation is defined as:

$$Q'_{F,i} = \frac{1}{1 + e^{(Q_{F,i} - \bar{Q}_{F,i})}} \tag{6.45}$$

for $i = 1...3$, where i represents each fusion model, $Q_{F,i}$ is the vector of all fusion RQIs, $\bar{Q}_{F,i}$ is the mean of the fusion RQIs from the $\mathcal{D}_{c,1}$ data set, and $Q'_{F,i} \in [0\ 1]$ is the

[1] This was not required for $Q_{F,3}$ as PCA is unsupervised and was obtained directly from $Q'_{S,1,j}$, $Q'_{S,2,j}$, and $Q'_{S,4,j}$ without input from E and the individual RQIs are already properly mapped to E.

bounded fused RQI.[2]

6.2.4 Fusion RQI Performance Evaluation Methods

As with the individual RQIs, the fusion RQIs were evaluated for both accuracy and robustness. The fusion RQIs were evaluated against the two control RQIs: the SQI comparison RQI ($Q_{C,1}$) and the ideal control RQI ($Q_{C,2}$). The fusion RQIs were also compared to their individual RQI inputs, $Q_{S,[1,2,4]}$. It is not expected that the fusion RQIs are more accurate than the most accurate individual RQI but rather that the fusion RQIs are more robust across data sets than any of the individual RQIs.

Unlike the previous analysis of the individual RQIs, the fusion RQIs cannot be evaluated on the superset of data which includes both $\mathcal{D}_c$ and $\mathcal{D}_m$. Since training the fusion models required training and validation using $\mathcal{D}_{c,1}$, this data subset alone cannot be used to assess performance. Additionally, since the goal is to assess accuracy and robustness across different data sets and given the important variation in patient population, $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$ have been separated for this analysis.

The techniques for assessing accuracy in the fusion RQIs are the same as those used for assessing the individual RQIs: MAE, E_{50}, SE_{50}, and Spearman's rank order correlation, ρ. The robustness is also assessed using the same techniques, the standard deviation of the MAE and the average RQI value between quantiles with their statistical differences being assessed using Welch's t-test. In addition to these measures for robustness, an analysis was conducted to examine the robustness of the fusion RQIs in comparison to the individual RQIs. This is critical because, as was previously described, the primary hypothesised benefit of the fusion RQIs compared to the individual RQIs is

[2]The term used in the exponential for $Q_{F,3}$ is $(-1 \times (Q_{F,i} - \bar{Q}_{F,i}))$ reflecting the fact that the additive inverse of $Q_{F,3}$ is not required.

that they will improve robustness. The robustness of the fusion RQIs compared to the individual RQIs was explored through the method described in the following section.

Comparison of Robustness in Fusion RQIs versus Individual RQIs

Given the previous discussion about the expected non-linear but monotonic relationship between the RQIs and the RR error in Section 5.2.3, it is not possible to make a direct comparison of the values of the RQIs. However, as in the previous section, it is possible to make a comparison using rank ordering. Given that the ideal RQI would have the highest values when the error is lowest, all of the fusion and individual RQIs have been rank ordered from 1 to N where N is the total number of observations in a data set and each RR modulation for a given time window is considered a separate observation ($\mathrm{rg}Q_{S,1}$, $\mathrm{rg}Q_{S,2}$, $\mathrm{rg}Q_{S,4}$, $\mathrm{rg}Q_{F,1}$, $\mathrm{rg}Q_{F,2}$, and $\mathrm{rg}Q_{F,3}$). The values for the RR error ($E_{M,i,j}$) have been similarly ranked where the observations with the highest RR error are given the lowest rank and the values with the highest RR error are given the highest rank.[3] This equates to the inverse rank ordering of $Q_{C,1}$ which will be defined as $\mathrm{rg}Q_{C,1}$. The result of this is that the most robust RQIs are those where the ranking of the RQIs for all observations is closest to the RR error ranking.

The robustness between the fusion RQIs and the individual RQIs is compared using the absolute difference in the rank between each RQI and the RR error rank defined as:

$$|\mathrm{rg}Q_{C,1} - \mathrm{rg}Q_{[S|F],[1...4]}| \tag{6.46}$$

[3]The rank ordering of the RR error is inverted compared to Spearman's Rank Order correlation where the highest ranks were associated with the highest RR error, thus if Spearman's correlation were conducted using these ranks, the magnitudes would be the same but the directionality would be reversed.

Using the results from Equation 6.46, the worst-performing RQI for each observation can be selected as the RQI which has the largest magnitude of difference. In order to compare the fusion RQIs to the individual RQIs, an analysis of each fusion RQI was done where the worst-performing RQI for each observation was selected where one of the fusion RQIs and all three individual RQIs were included in each analysis. In each of these analyses, the number and total percentage of instances where each RQI was the worst-performing and the mean of the differences when an RQI was the worst-performing was compared to the mean of the differences for the other three RQIs combined. The expected outcomes of this analysis are twofold. First, it shows the number of times that an RQI is outperformed by the others. Second, it shows how badly on average that RQI performs when it is the worst-performing RQI. Based on this analysis, the most robust RQI should have the fewest number of instances where it is the worst-performing RQI and an average difference value that is similar to the average difference value for the other three RQIs.

6.3 Results

The results of training and validation process for each of the fusion RQIs are first presented for each RQI followed by the performance of the fusion RQI in the training, validation, and test sets. The performances of the fusion RQIs, $Q_{F,[1...3]}$, are assessed compared to both the control RQIs, $Q_{C,[1,2]}$ and the individual RQIs that comprise the fusion RQIs, $Q_{S,[1,2,4]}$. The RQIs are assessed using the same accuracy and robustness measures as were used to assess the individual RQIs. Additionally, the fusion RQIs are also assessed for their robustness in comparison to the individual RQIs.

The evaluation of the fusion RQIs is conducted separately in each data set, $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$,

$\mathcal{D}_m$, unlike the evaluation of the individual RQIs. This is necessary for two reasons. The first is that $\mathcal{D}_{c,1}$ was used as the training and validation data for $Q_{F,1}$ and $Q_{F,2}$, whereas the individual RQIs need no training sets. The second is that the CapnoBase and MIMIC-III data sets represent different patient populations (as described in Chapter 4) and thus differentiating the data sets in the analysis of the fusion RQI allowed for a more robust analysis of the fusion models as it is possible to see the applicability of the models trained on the CapnoBase data set ($\mathcal{D}_{c,1}$) on a held-out data set ($\mathcal{D}_{c,2}$) and a completely independent data set ($\mathcal{D}_m$). The robustness analysis of the fusion RQIs compared to the individual RQIs was not conducted in the $\mathcal{D}_{c,1}$ data set but was conducted for the $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$ data sets for each of the three fusion RQIs. For brevity, these results are only shown for $\mathcal{D}_m$ as the results for $\mathcal{D}_{c,2}$ are all quantitatively similar.

6.3.1 Data Standardisation

As described in Section 6.2.1, prior to fusion analysis, the three individual RQIs of interest ($Q_{S,[1,2,4]}$) were normalised for the $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, and $\mathcal{D}_m$ data sets using the zero-mean, unit-variance method shown in Equation 6.39 where all of the coefficients were determined using only the $\mathcal{D}_{c,1}$ data set. The coefficients for that normalisation are shown in Table 6.1.

Table 6.1: *Values for zero-mean, unit-variance normalisation of the individual RQIs ($Q_{S,[1,2,3]}$) based on their means and standard deviations in the $\mathcal{D}_{c,1}$ data set.*

	Mean	**St. Dev.**
$\boldsymbol{Q_{S,1}}$	0.258	0.087
$\boldsymbol{Q_{S,2}}$	0.305	0.157
$\boldsymbol{Q_{S,4}}$	0.415	0.164

6.3.2 Linear Regression Fusion RQI ($Q_{F,1}$)

The $Q_{F,1}$ linear regression model was trained on $\mathcal{D}_{c,1}$ using five-fold cross validation (as described in Section 6.2.1) using the regularised parameters, $Q'_{S,[1,2,4]}$ and Equation 6.10. Training was required in order to determine the optimal regularisation parameter, λ. The range of values for λ tested is described in Section 6.2.2 but ranges from 0 to 16. The model coefficients were first determined on the five training folds and then each model was applied to the four held out validation folds. The combined average mean squared error values for the training and validation folds are shown in Figure 6.1.

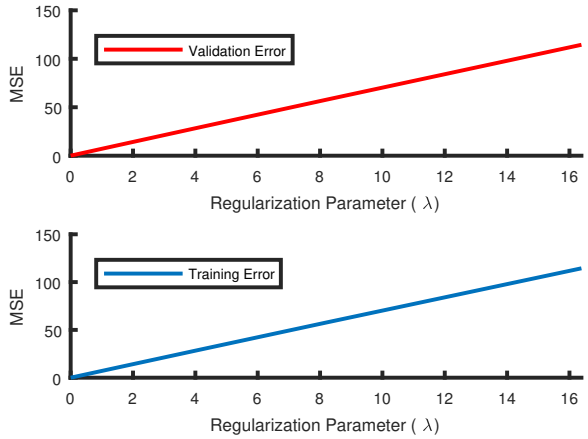


Figure 6.1: *Average mean squared error (MSE) values for the validation and training folds of the $\mathcal{D}_{c,1}$ data set to determine the optimal value of α for the $Q_{F,1}$ fusion RQI.*

Based on these results, the optimal regularisation parameter was determined to be $\lambda = 0$, indicating that no model regularisation was necessary. Using this regularisation parameter, the linear regression model was trained on the entire $\mathcal{D}_{c,1}$ data set to determine the model coefficients. The linear regression model for $Q_{F,1}$ is:

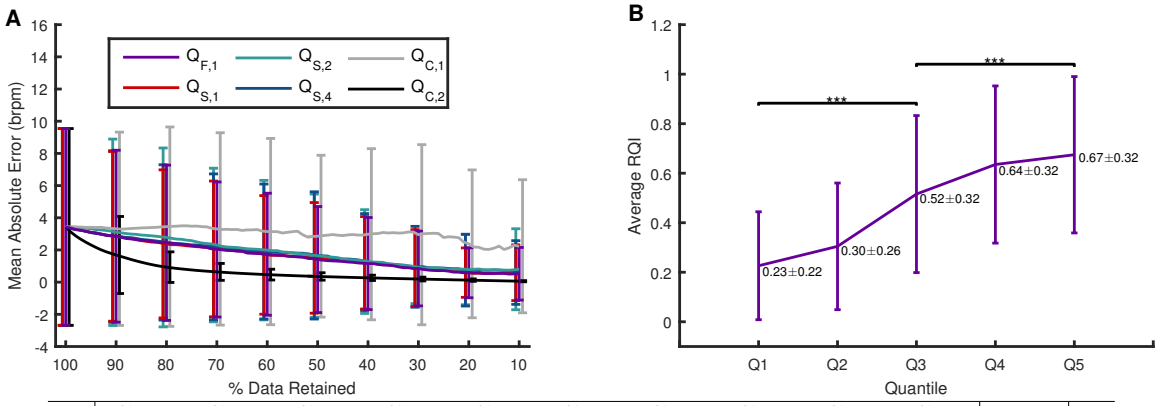
$$Q_{F,1} = 3.43 - 1.40Q'_{S,1} - 0.40Q'_{S,2} - 0.54Q'_{S,4} \tag{6.47}$$

However, as described in Section 6.2.3, the values for $Q_{F,1}$ were corrected using an

inverted logistic function to both invert and bound the RQI values.

The performance of $Q_{F,1}$ in $\mathcal{D}_{c,1}$, the training and validation set, is shown in Figure 6.2 and in the two test sets, $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$, in Figures 6.3 and 6.4, respectively. The robustness analysis described in Section 6.2.4 for $Q_{F,1}$ in the $\mathcal{D}_m$ data set is shown in Figure 6.5. As expected, the performance of $Q_{F,1}$ in the training and validation set slightly outperforms its three individual component RQIs ($Q_{S,[1,2,4]}$) in both accuracy and robustness. The E_{50} is slightly greater than any of the component RQIs and Spearman's ρ indicates a moderately strong negative correlation comparable to $Q_{S,2}$, the individual RQI with the strongest ρ in the $\mathcal{D}_{c,1}$ data set. Furthermore, the robustness measures indicate strong performance with the RQI values in each of the quantiles monotonically increasing with the comparisons between quantiles 1 and 3 and quantiles 3 and 5 being significant to p $<$0.001 (Figure 6.2).

Given that $\mathcal{D}_{c,1}$ is both the training and validation set, this performance is expected. The results on the two test sets, $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$, indicate similar albeit moderately less accurate and robust performance. These results do suggest that $Q_{F,1}$ is both an accurate and robust fusion of $Q_{S,[1,2,3]}$ and the model was not overfitted to the $\mathcal{D}_{c,1}$ data set. Figure 6.3 shows that all of the accuracy metrics of $Q_{F,1}$ are comparable to the individual RQI with the best accuracy metric. For example, the E_{50} of $Q_{S,4}$ (1.09) is the best in $\mathcal{D}_{c,1}$ and $Q_{F,1}$ has the second best E_{50} (1.04). The same trend is also seen in $\mathcal{D}_m$ where the ranked correlation with the RR error is equally highest in $Q_{S,3}$ and $Q_{F,1}$ (Figure 6.4). Additionally, robustness measures in Figures 6.3 and 6.4 show generally decreasing standard deviations as the worst-performing data are discarded and show statistically relevant changes in the average RQIs in the average quantile values, especially between quantile 3 and quantile 5. Overall, these results support $Q_{F,1}$ as being a moderately

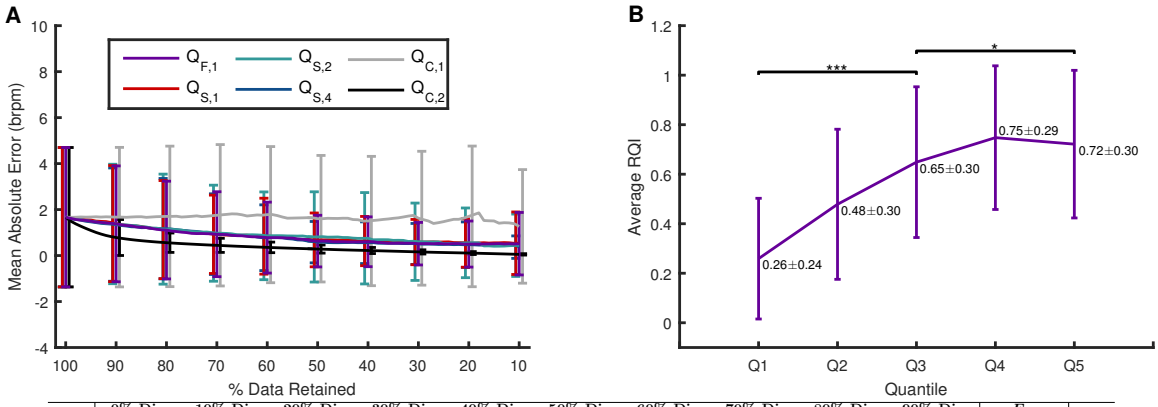


	0% Dis.	10% Dis.	20% Dis.	30% Dis.	40% Dis.	50% Dis.	60% Dis.	70% Dis.	80% Dis.	90% Dis.	E_{50}	ρ
$Q_{F,1}$	3.43 ± 6.11	2.85 ± 5.34	2.45 ± 4.83	2.03 ± 4.20	1.73 ± 3.79	1.40 ± 3.30	1.16 ± 2.86	0.85 ± 2.32	0.58 ± 1.55	0.52 ± 1.63	2.03 ± 1.00	-0.52
$Q_{S,1}$	3.43 ± 6.11	2.85 ± 5.27	2.38 ± 4.60	2.06 ± 4.22	1.69 ± 3.68	1.51 ± 3.44	1.20 ± 2.87	0.88 ± 2.39	0.59 ± 1.54	0.59 ± 1.73	1.92 ± 1.01	-0.50
$Q_{S,2}$	3.43 ± 6.11	3.10 ± 5.79	2.78 ± 5.56	2.31 ± 4.77	1.99 ± 4.33	1.64 ± 3.83	1.28 ± 3.22	0.93 ± 2.26	0.78 ± 2.19	0.80 ± 2.52	1.79 ± 1.05	-0.52
$Q_{S,4}$	3.43 ± 6.11	2.84 ± 5.33	2.47 ± 4.83	2.20 ± 4.53	1.90 ± 4.19	1.66 ± 3.95	1.25 ± 3.01	0.95 ± 2.51	0.75 ± 2.23	0.60 ± 1.98	1.77 ± 1.06	-0.48
$Q_{C,1}$	3.43 ± 6.11	3.32 ± 6.01	3.45 ± 6.20	3.31 ± 5.98	3.14 ± 5.79	2.86 ± 5.03	2.98 ± 5.32	2.95 ± 5.60	2.38 ± 4.59	2.23 ± 4.13	0.57 ± 1.17	-0.04
$Q_{C,2}$	3.43 ± 6.11	1.69 ± 2.39	0.93 ± 0.95	0.64 ± 0.52	0.47 ± 0.33	0.35 ± 0.23	0.27 ± 0.17	0.19 ± 0.12	0.12 ± 0.08	0.06 ± 0.04	3.07 ± 0.84	-1.00

Figure 6.2: *Performance of $Q_{F,1}$ in the $\mathcal{D}_{c,1}$ training and validation data set. (A) MAE of $Q_{F,1}$, controls ($Q_{C,[1,2]}$), and individual RQIs ($Q_{S,[1,2,4]}$) as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

accurate and robust fusion RQI which performs on par with the best-performing RQI for any single metric and data set. Additionally, because the analysis was conducted on two held-out data sets, the results suggest that the results are robust and do not overfit the training and validation set.

Figure 6.5 shows the robustness analysis of $Q_{F,1}$ compared to $Q_{S,[1,2,4]}$ in $\mathcal{D}_m$. Figure 6.5, Panels A, B, and C show all instances where $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$, respectively were the worst-performing RQIs based on the absolute difference of the ranks between the RQIs and the RR error. A key observation from these results is that each of the individual RQIs was the worst performer in roughly a similar number of observations (29.5% for $Q_{S,1}$, 32.8% for $Q_{S,2}$, and 33.7% for $Q_{S,4}$) indicating that within the $\mathcal{D}_m$ data set no one individual RQI was the *de facto* best-performer and that in a roughly equal

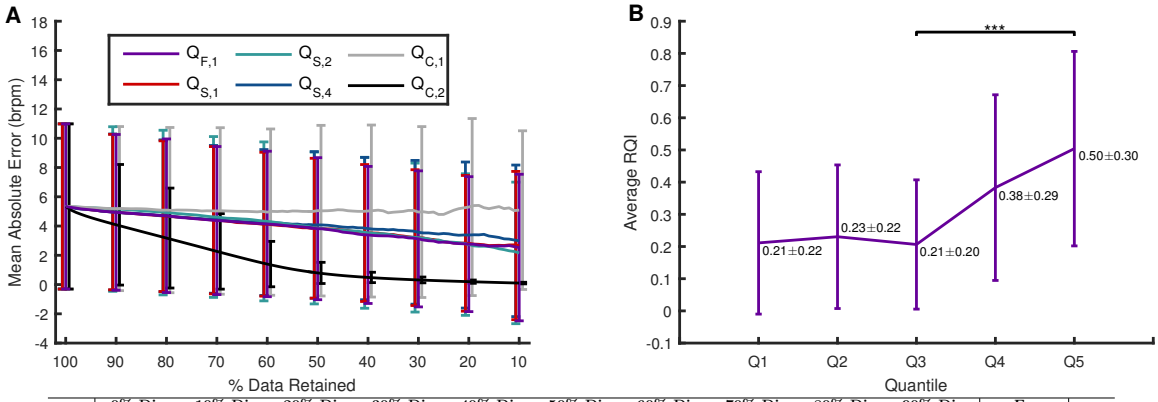


	0% Dis.	10% Dis.	20% Dis.	30% Dis.	40% Dis.	50% Dis.	60% Dis.	70% Dis.	80% Dis.	90% Dis.	E_{50}	ρ
$Q_{F,1}$	1.67 ± 3.03	1.38 ± 2.52	1.11 ± 2.12	0.93 ± 1.85	0.78 ± 1.54	0.63 ± 1.12	0.60 ± 1.08	0.52 ± 0.93	0.50 ± 1.00	0.52 ± 1.36	1.04 ± 0.61	-0.49
$Q_{S,1}$	1.67 ± 3.03	1.39 ± 2.52	1.13 ± 2.13	0.93 ± 1.70	0.84 ± 1.65	0.68 ± 1.17	0.63 ± 1.07	0.59 ± 0.98	0.55 ± 1.05	0.54 ± 1.36	0.99 ± 0.61	-0.44
$Q_{S,2}$	1.67 ± 3.03	1.37 ± 2.60	1.15 ± 2.39	0.97 ± 2.09	0.86 ± 1.91	0.81 ± 1.96	0.75 ± 1.99	0.60 ± 1.68	0.55 ± 1.52	0.45 ± 1.36	0.86 ± 0.70	-0.54
$Q_{S,4}$	1.67 ± 3.03	1.35 ± 2.46	1.17 ± 2.18	0.93 ± 1.75	0.78 ± 1.43	0.58 ± 0.90	0.56 ± 0.91	0.51 ± 0.89	0.47 ± 0.99	0.37 ± 0.49	1.09 ± 0.59	-0.50
$Q_{C,1}$	1.67 ± 3.03	1.67 ± 3.03	1.70 ± 3.06	1.75 ± 3.07	1.78 ± 2.96	1.60 ± 2.75	1.50 ± 2.81	1.62 ± 2.91	1.70 ± 3.06	1.27 ± 2.47	0.06 ± 0.81	0.03
$Q_{C,2}$	1.67 ± 3.03	0.78 ± 0.78	0.56 ± 0.42	0.44 ± 0.31	0.36 ± 0.23	0.28 ± 0.18	0.22 ± 0.13	0.16 ± 0.09	0.11 ± 0.06	0.05 ± 0.03	1.38 ± 0.55	-1.00

Figure 6.3: *Performance of $Q_{F,1}$ in the $\mathcal{D}_{c,2}$ test data set. (A) MAE of $Q_{F,1}$, controls ($Q_{C,[1,2]}$), and individual RQIs ($Q_{S,[1,2,4]}$) as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

number of instances each individual RQI was the worst. Additionally, the results show that the fusion RQI is the worst-performing RQI in far fewer instances (4.0%), which is a strong indication of its improved robustness compared to the individual RQIs. It can be further seen when comparing the average rank difference of an individual RQI compared to the RR error ideal rank (the mean of the absolute rank difference in all instances where an RQI was the worst performer) to the average rank difference of the three other RQIs compared to the RR error ideal rank (the aggregate mean of the absolute rank difference of these RQIs when they are not the worst-performer) is that the difference in these averages is much larger for all of the individual RQIs than it is for $Q_{F,1}$. This is a strong indicator of robustness because it means that $Q_{F,1}$ is most likely to be the worst-performing RQI only when the three individual RQIs are all performing poorly. Given that $Q_{F,1}$ relies on its three component parts, this shows that $Q_{F,1}$ is robust against poor

performance of at least one of the individual RQIs, but not all three.

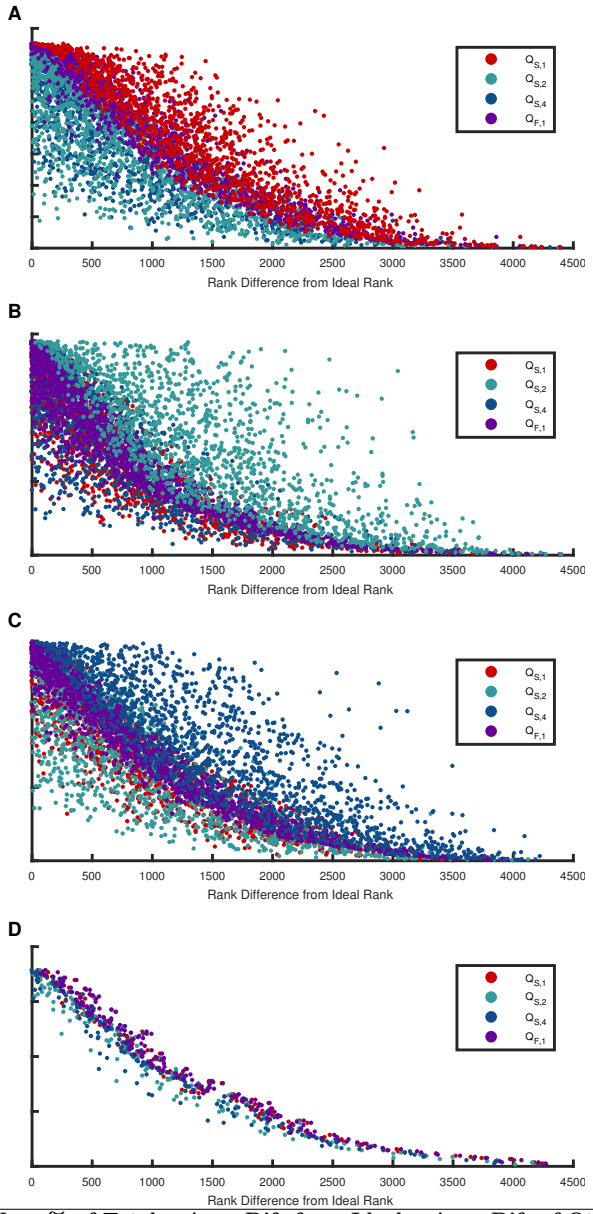


	0% Dis.	10% Dis.	20% Dis.	30% Dis.	40% Dis.	50% Dis.	60% Dis.	70% Dis.	80% Dis.	90% Dis.	E_{50}	ρ
$Q_{F,1}$	5.34 ± 5.65	4.94 ± 5.32	4.70 ± 5.25	4.37 ± 5.06	4.14 ± 4.97	3.82 ± 4.85	3.39 ± 4.68	3.12 ± 4.65	2.76 ± 4.61	2.53 ± 5.01	1.52 ± 0.99	-0.39
$Q_{S,1}$	5.34 ± 5.65	4.95 ± 5.31	4.67 ± 5.15	4.40 ± 5.03	4.13 ± 4.92	3.85 ± 4.78	3.52 ± 4.68	3.21 ± 4.64	2.82 ± 4.64	2.67 ± 5.07	1.49 ± 0.98	-0.37
$Q_{S,2}$	5.34 ± 5.65	5.16 ± 5.62	4.92 ± 5.63	4.62 ± 5.50	4.32 ± 5.43	3.87 ± 5.19	3.54 ± 5.16	3.21 ± 5.09	2.73 ± 4.84	2.16 ± 4.84	1.47 ± 1.03	-0.38
$Q_{S,4}$	5.34 ± 5.65	4.96 ± 5.33	4.70 ± 5.19	4.46 ± 5.05	4.24 ± 4.99	4.08 ± 5.01	3.83 ± 4.86	3.57 ± 4.91	3.38 ± 4.99	2.99 ± 5.18	1.26 ± 1.01	-0.33
$Q_{C,1}$	5.34 ± 5.65	5.19 ± 5.61	5.09 ± 5.65	5.03 ± 5.69	4.95 ± 5.68	5.05 ± 5.83	5.03 ± 5.88	4.96 ± 5.84	5.30 ± 6.05	5.09 ± 5.42	0.29 ± 1.10	-0.11
$Q_{C,2}$	5.34 ± 5.65	4.09 ± 4.12	3.18 ± 3.42	2.27 ± 2.58	1.40 ± 1.55	0.80 ± 0.72	0.49 ± 0.35	0.32 ± 0.20	0.20 ± 0.11	0.10 ± 0.06	4.54 ± 0.70	-1.00

Figure 6.4: *Performance of $Q_{F,1}$ in the $\mathcal{D}_m$ test data set. (A) MAE of $Q_{F,1}$, controls ($Q_{C,[1,2]}$), and individual RQIs ($Q_{S,[1,2,4]}$) as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

6.3.3 Support Vector Regression Fusion RQI ($Q_{F,2}$)

As described in Section 6.2.2, two parameters for $Q_{F,2}$, the support vector regression RQI model, need to be empirically determined from the training and validation set $\mathcal{D}_{c,1}$. These two parameters are: C, the penalty term (or regularisation parameter) and γ, the regularisation parameter of the RBF kernel controlling the radius of influence of a single training observation. The values for C and γ were determined empirically using five-fold cross-validation with both a coarse and a fine grid search. For the initial coarse grid search, 16 values for C and 21 values for γ were used as described in Equations 6.41 and 6.42. The average MSE for the sum of all of the training and validation sets for the grid search is shown as a heat map in Figure 6.6 where the lowest MSE in the validation



	N	% of Total	Avg. Dif. from Ideal	Avg. Dif. of Other Three
$\boldsymbol{Q_{S,1}}$	1299	29.5	1557.8 ± 848.2	1010.6 ± 832.0
$\boldsymbol{Q_{S,2}}$	1446	32.8	1555.0 ± 875.5	851.7 ± 801.4
$\boldsymbol{Q_{S,4}}$	1484	33.7	1659.9 ± 909.9	1000.3 ± 839.9
$\boldsymbol{Q_{F,1}}$	178	4.0	1416.3 ± 1000.8	1283.9 ± 981.6

Figure 6.5: *Robustness comparison of $Q_{F,1}$ (purple) to $Q_{S,1}$ (red), $Q_{S,2}$ (teal), and $Q_{S,4}$ (blue) in the $\mathcal{D}_m$ data set. The x-axis represents the rank difference ($|rgQ_{C,1} - rgQ_{X,Z}|$) and the y-axis represents sorting all observations based on the average difference of the three best-performing RQIs. (A) The rank difference of $Q_{F,1}$ and $Q_{S,[1,2,4]}$ where $Q_{S,1}$ is the worst-performing RQI. (B) The rank difference of $Q_{F,1}$ and $Q_{S,[1,2,4]}$ where $Q_{S,2}$ is the worst-performing RQI. (C) The rank difference of $Q_{F,1}$ and $Q_{S,[1,2,4]}$ where $Q_{S,4}$ is the worst-performing RQI. (D) The rank difference of $Q_{F,1}$ and $Q_{S,[1,2,4]}$ where $Q_{F,1}$ is the worst-performing RQI. (Table) The number of instances and percentage of instances in which a given RQI was the worst-performer in the data set and the average rank difference and standard deviation of that RQI compared to the average rank difference and standard deviation of the three best-performing RQIs.*

set was 35.26 when $C = 8$ and $\gamma = 8$.

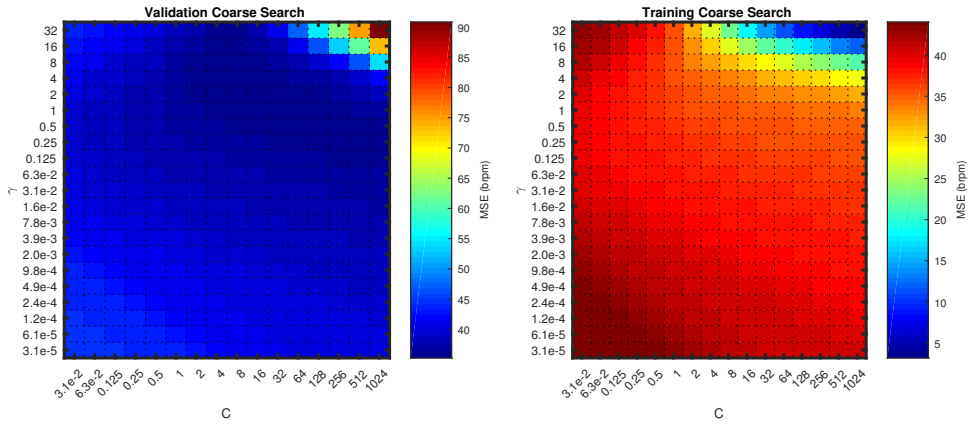


Figure 6.6: *Average MSE values of the coarse grid search for the validation and training folds of the $\mathcal{D}_{c,1}$ data set to determine the optimal values of C and γ for the $Q_{F,2}$ fusion RQI.*

Following the coarse grid search, the optimal values for C and γ (those which minimised the MSE in the coarse search) were used to centre a fine grid search according to Equations 6.43 and 6.44. Similar results to the coarse grid search for the fine grid search are shown in Figure 6.7. These results show that the minimal MSE in the validation sets is 35.20 when $C = 8$ and $\gamma = 9.5$. These values of C and γ were used to train the SVR model for $Q_{F,2}$ on the entire $\mathcal{D}_{c,1}$ training and validation set. This model was applied to the test sets, $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$. The fine grid search suggests that the optimal model parameters were determined as neither C nor γ were edge values.

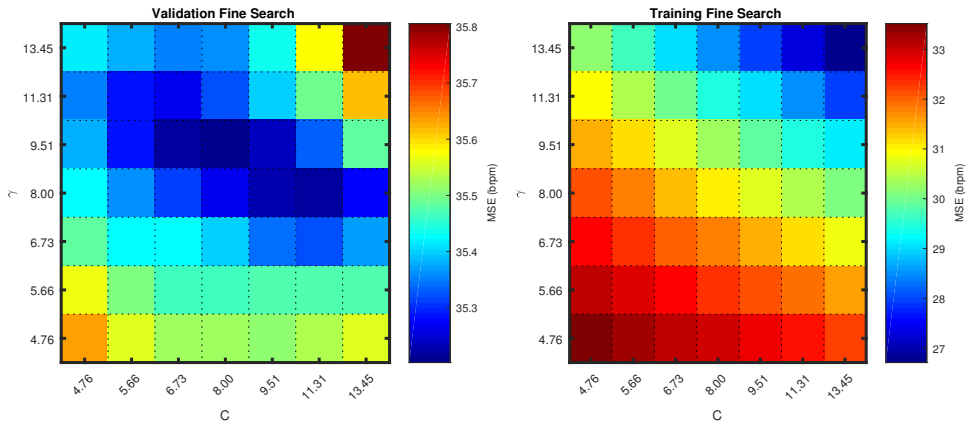
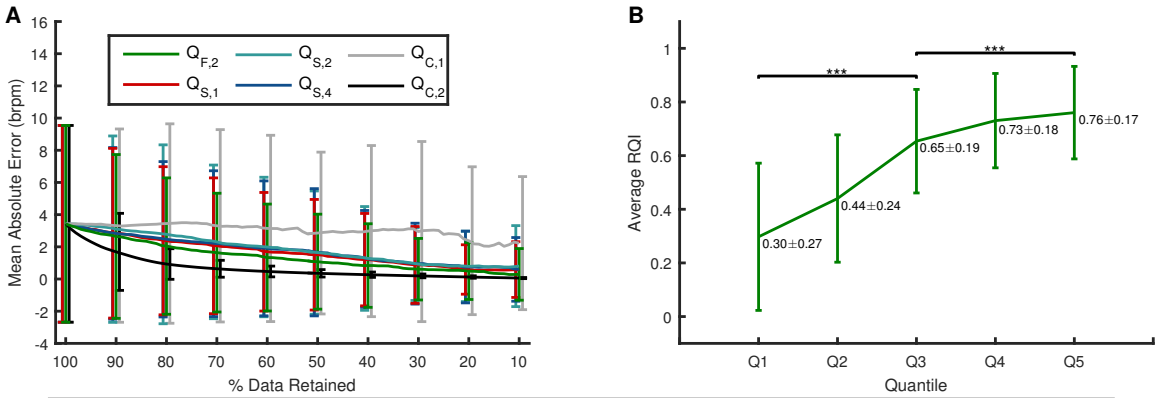


Figure 6.7: *Average MSE values of the fine grid search for the validation and training folds of the $\mathcal{D}_{c,1}$ data set to determine the optimal values of C and γ for the $Q_{F,2}$ fusion RQI*

The performance results for $Q_{F,2}$ are shown in Figure 6.8 for $\mathcal{D}_{c,1}$, Figure 6.9 for $\mathcal{D}_{c,2}$, and Figure 6.10 for $\mathcal{D}_m$. The robustness analysis is shown in Figure 6.11.

Like $Q_{F,1}$, the results for $Q_{F,2}$ in the training and validation set, $\mathcal{D}_{c,1}$, indicate that the SVR model outperforms the individual RQIs (Figure 6.8). In fact, $Q_{F,2}$ outperforms the individual RQIs on all accuracy and robustness measures. The E_{50} and ρ for $Q_{F,1}$ in $\mathcal{D}_{c,1}$ were 2.35 and -0.70, respectively, which are nearly comparable to 3.07 and -1.00, the values for $Q_{C,2}$, the ideal control.



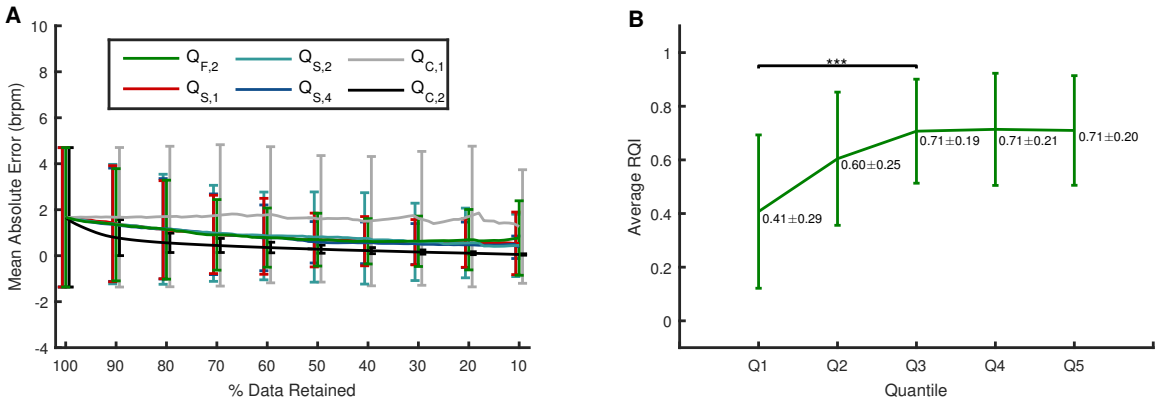
	0% Dis.	10% Dis.	20% Dis.	30% Dis.	40% Dis.	50% Dis.	60% Dis.	70% Dis.	80% Dis.	90% Dis.	E_{50}	ρ
$Q_{F,2}$	3.43 ± 6.11	2.64 ± 5.09	2.05 ± 4.24	1.64 ± 3.69	1.33 ± 3.32	1.08 ± 2.95	0.84 ± 2.60	0.61 ± 1.91	0.49 ± 1.77	0.29 ± 1.60	2.35 ± 0.97	-0.70
$Q_{S,1}$	3.43 ± 6.11	2.85 ± 5.27	2.38 ± 4.60	2.06 ± 4.22	1.69 ± 3.68	1.51 ± 3.44	1.20 ± 2.87	0.88 ± 2.39	0.59 ± 1.54	0.59 ± 1.73	1.92 ± 1.01	-0.50
$Q_{S,2}$	3.43 ± 6.11	3.10 ± 5.79	2.78 ± 5.56	2.31 ± 4.77	1.99 ± 4.33	1.64 ± 3.83	1.28 ± 3.22	0.93 ± 2.26	0.78 ± 2.19	0.80 ± 2.52	1.79 ± 1.05	-0.52
$Q_{S,4}$	3.43 ± 6.11	2.84 ± 5.33	2.47 ± 4.83	2.20 ± 4.53	1.90 ± 4.19	1.66 ± 3.95	1.25 ± 3.01	0.95 ± 2.51	0.75 ± 2.23	0.60 ± 1.98	1.77 ± 1.06	-0.48
$Q_{C,1}$	3.43 ± 6.11	3.32 ± 6.01	3.45 ± 6.20	3.31 ± 5.98	3.14 ± 5.79	2.86 ± 5.03	2.98 ± 5.32	2.95 ± 5.60	2.38 ± 4.59	2.23 ± 4.13	0.57 ± 1.17	-0.04
$Q_{C,2}$	3.43 ± 6.11	1.69 ± 2.39	0.93 ± 0.95	0.64 ± 0.52	0.47 ± 0.33	0.35 ± 0.23	0.27 ± 0.17	0.19 ± 0.12	0.12 ± 0.08	0.06 ± 0.04	3.07 ± 0.84	-1.00

Figure 6.8: *Performance of $Q_{F,2}$ in the $\mathcal{D}_{c,1}$ training and validation data set. (A) MAE of $Q_{F,2}$, controls ($Q_{C,[1,2]}$), and individual RQIs ($Q_{S,[1,2,4]}$) as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

However, as for $Q_{F,1}$, it is expected that the fusion model will outperform the individual RQIs on the training and validation set. The results for $Q_{F,2}$ on the two test sets show that the SVR model trained for $Q_{F,2}$ is not generalisable to unseen data sets. The results for $\mathcal{D}_{c,2}$, the CapnoBase test set, show performance that is comparable to the worst-performing RQI for different accuracy and robustness measures and only shows

statistically significant difference in the average RQI values between quantiles 1 and 3 (Figure 6.9). Furthermore, the results from the $\mathcal{D}_m$ data set show results that are the inverse of the expected for a high-performing fusion RQI model. The E_{50} and ρ are negative indicating that higher RQI values correlate to higher RR errors. This trend is also seen when considering robustness. The standard deviation increases as more data are discarded. It is also seen in the quantile analysis where the largest average RQI values are seen in quantile 1 (3.78) and the smallest RQI values are seen in quantile 5 (1.58). This downward trend is statistically significant to $p < 0.001$ between quantiles 1 and 3 and quantiles 3 and 5 (Figure 6.10). These results suggest that despite fitting the optimal parameters to the SVR model, $Q_{F,2}$ is overfitted to the training set $\mathcal{D}_{c,1}$ and is not generalisable to data sets unseen by the model. While the results for $\mathcal{D}_{c,2}$ indicate poor performance of the RQI, they do show the expected trend where higher RQIs have lower RR error. $\mathcal{D}_m$ shows an inverse trend. The discrepancy in these results may be due to the fact that $\mathcal{D}_{c,2}$ is drawn from the same population as the training and validation data. The results are not nearly as poor for $\mathcal{D}_{c,2}$ as for $\mathcal{D}_m$, which is drawn from an entirely different patient population.

The robustness analysis in Figure 6.11 supports the conclusions about $Q_{F,2}$ drawn from the previous analyses. These results show that for the $\mathcal{D}_m$ data set, $Q_{F,2}$ is the worst-performing RQI in 49.9% of the observations. It also shows that when considering the average difference from ideal for $Q_{F,2}$ when it is the worst-performing RQI compared to the average of the individual RQIs, the distance between these values is larger than for any one of the individual RQIs. This indicates that in these instances, $Q_{F,2}$ is poorly mapping to the individual RQIs. Based on all of these analyses of $Q_{F,2}$, it appears that $Q_{F,2}$ is overfitted to the training and validation data set and performs poorly on the two



	0% Dis.	10% Dis.	20% Dis.	30% Dis.	40% Dis.	50% Dis.	60% Dis.	70% Dis.	80% Dis.	90% Dis.	E_{50}	ρ
$Q_{F,2}$	1.67 ± 3.03	1.34 ± 2.44	1.13 ± 2.15	0.91 ± 1.53	0.79 ± 1.29	0.70 ± 1.15	0.63 ± 0.99	0.63 ± 1.10	0.70 ± 1.31	0.77 ± 1.62	0.97 ± 0.61	-0.42
$Q_{S,1}$	1.67 ± 3.03	1.39 ± 2.52	1.13 ± 2.13	0.93 ± 1.70	0.84 ± 1.65	0.68 ± 1.17	0.63 ± 1.07	0.59 ± 0.98	0.55 ± 1.05	0.54 ± 1.36	0.99 ± 0.61	-0.44
$Q_{S,2}$	1.67 ± 3.03	1.37 ± 2.60	1.15 ± 2.39	0.97 ± 2.09	0.86 ± 1.91	0.81 ± 1.96	0.75 ± 1.99	0.60 ± 1.68	0.55 ± 1.52	0.45 ± 1.36	0.86 ± 0.70	-0.54
$Q_{S,4}$	1.67 ± 3.03	1.35 ± 2.46	1.17 ± 2.18	0.93 ± 1.75	0.78 ± 1.43	0.58 ± 0.90	0.56 ± 0.91	0.51 ± 0.89	0.47 ± 0.99	0.37 ± 0.49	1.09 ± 0.59	-0.50
$Q_{C,1}$	1.67 ± 3.03	1.67 ± 3.03	1.70 ± 3.06	1.75 ± 3.07	1.78 ± 2.96	1.60 ± 2.75	1.50 ± 2.81	1.62 ± 2.91	1.70 ± 3.06	1.27 ± 2.47	0.06 ± 0.81	0.03
$Q_{C,2}$	1.67 ± 3.03	0.78 ± 0.78	0.56 ± 0.42	0.44 ± 0.31	0.36 ± 0.23	0.28 ± 0.18	0.22 ± 0.13	0.16 ± 0.09	0.11 ± 0.06	0.05 ± 0.03	1.38 ± 0.55	-1.00

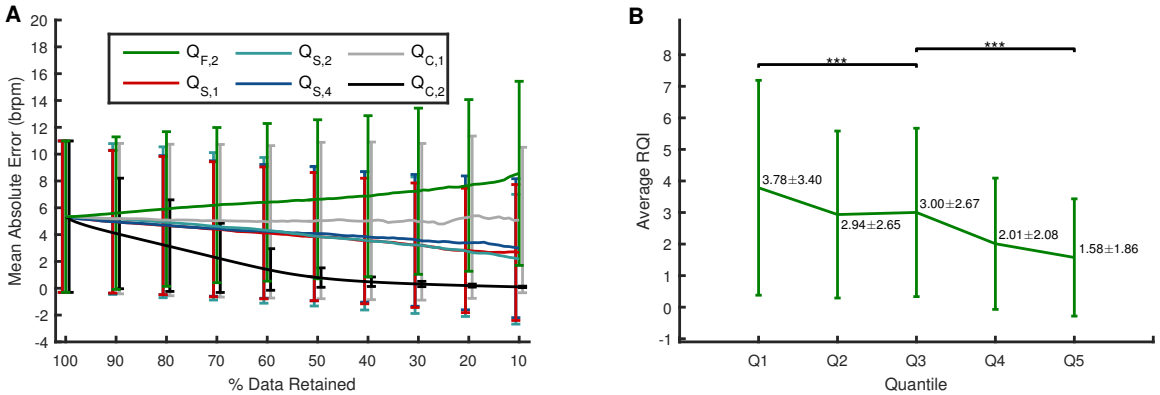
Figure 6.9: *Performance of $Q_{F,2}$ in the $\mathcal{D}_{c,2}$ test data set. (A) MAE of $Q_{F,2}$, controls ($Q_{C,[1,2]}$), and individual RQIs ($Q_{S,[1,2,4]}$) as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

tests sets making it a poorly generalisable fusion RQI.

6.3.4 Principal Component Analysis Fusion RQI ($Q_{F,3}$)

As the only unsupervised RQI fusion model, $Q_{F,3}$, based on the PCA, only depends on the training and validation set observations from $\mathcal{D}_{c,1}$ as described in Section 6.2.2. The model was trained using the maximal variance formulation described in Section 6.1.3. The first PC was used to define $Q_{F,3}$. The eigenvectors and eigenvalues of the model can be seen in Table 6.2. The first PC is a good fusion of all three individual RQIs as 89.06% of the variance between all of the individual RQIs was explained by the first PC.

The performance results for $Q_{F,3}$ are shown in Figure 6.12 for $\mathcal{D}_{c,1}$, in Figure 6.13 for $\mathcal{D}_{c,2}$, and in Figure 6.14 for $\mathcal{D}_m$. The robustness comparison results are shown in Figure 6.15.



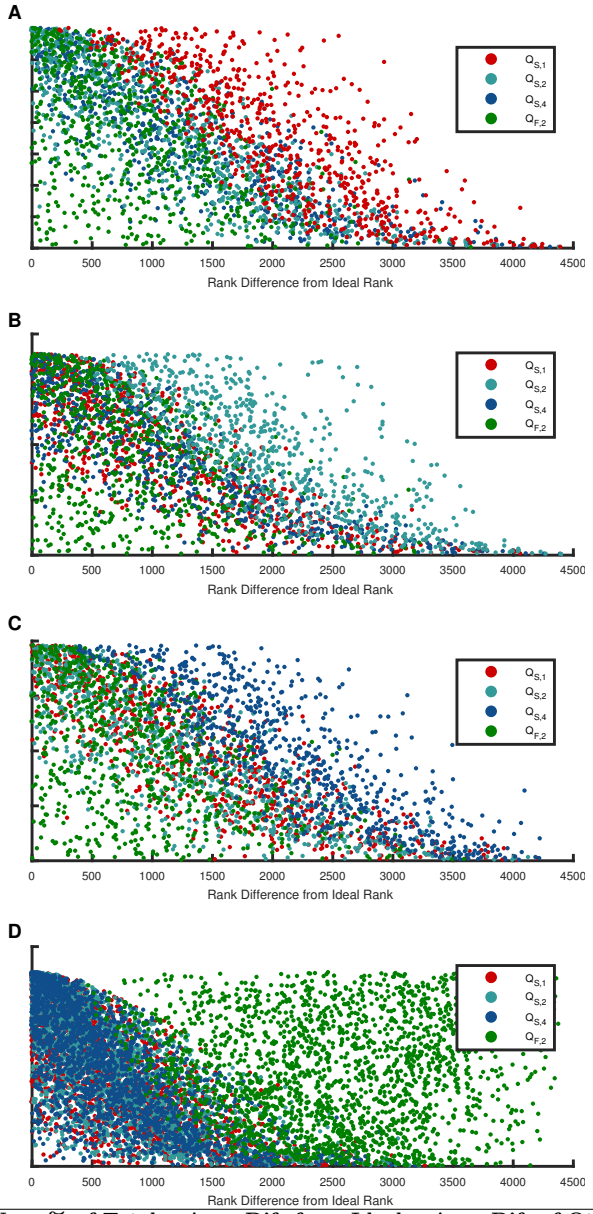
	0% Dis.	10% Dis.	20% Dis.	30% Dis.	40% Dis.	50% Dis.	60% Dis.	70% Dis.	80% Dis.	90% Dis.	E_{50}	ρ
$Q_{F,2}$	5.34 ± 5.65	5.60 ± 5.69	5.91 ± 5.77	6.20 ± 5.79	6.41 ± 5.88	6.63 ± 5.93	6.86 ± 6.01	7.23 ± 6.19	7.67 ± 6.40	8.57 ± 6.86	-1.30 ± 1.11	0.32
$Q_{S,1}$	5.34 ± 5.65	4.95 ± 5.31	4.67 ± 5.15	4.40 ± 5.03	4.13 ± 4.92	3.85 ± 4.78	3.52 ± 4.68	3.21 ± 4.64	2.82 ± 4.64	2.67 ± 5.07	1.49 ± 0.98	-0.37
$Q_{S,2}$	5.34 ± 5.65	5.16 ± 5.62	4.92 ± 5.63	4.62 ± 5.50	4.32 ± 5.43	3.87 ± 5.19	3.54 ± 5.16	3.21 ± 5.09	2.73 ± 4.84	2.16 ± 4.84	1.47 ± 1.03	-0.38
$Q_{S,4}$	5.34 ± 5.65	4.96 ± 5.33	4.70 ± 5.19	4.46 ± 5.05	4.24 ± 4.99	4.08 ± 5.01	3.83 ± 4.86	3.57 ± 4.91	3.38 ± 4.99	2.99 ± 5.18	1.26 ± 1.01	-0.33
$Q_{C,1}$	5.34 ± 5.65	5.19 ± 5.61	5.09 ± 5.65	5.03 ± 5.69	4.95 ± 5.68	5.05 ± 5.83	5.03 ± 5.88	4.96 ± 5.84	5.30 ± 6.05	5.09 ± 5.42	0.29 ± 1.10	-0.11
$Q_{C,2}$	5.34 ± 5.65	4.09 ± 4.12	3.18 ± 3.42	2.27 ± 2.58	1.40 ± 1.55	0.80 ± 0.72	0.49 ± 0.35	0.32 ± 0.20	0.20 ± 0.11	0.10 ± 0.06	4.54 ± 0.70	-1.00

Figure 6.10: *Performance of $Q_{F,2}$ in the $\mathcal{D}_m$ test data set. (A) MAE of $Q_{F,2}$, controls ($Q_{C,[1,2]}$), and individual RQIs ($Q_{S,[1,2,4]}$) as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

Table 6.2: *Eigenvectors and eigenvalues of PCA decomposition of $Q_{S,[1,2,4]}$ in the $\mathcal{D}_{c,1}$ data set used for determining the $Q_{F,3}$ fusion RQI.*

	PC1	**PC2**	**PC3**
	Eigenvector		
$\boldsymbol{Q_{S,1}}$	0.3387	-0.1840	0.9227
$\boldsymbol{Q_{S,2}}$	0.6813	0.7244	-0.1056
$\boldsymbol{Q_{S,4}}$	0.6490	-0.6644	-0.3707
Eigenvalues	89.06	8.33	2.61

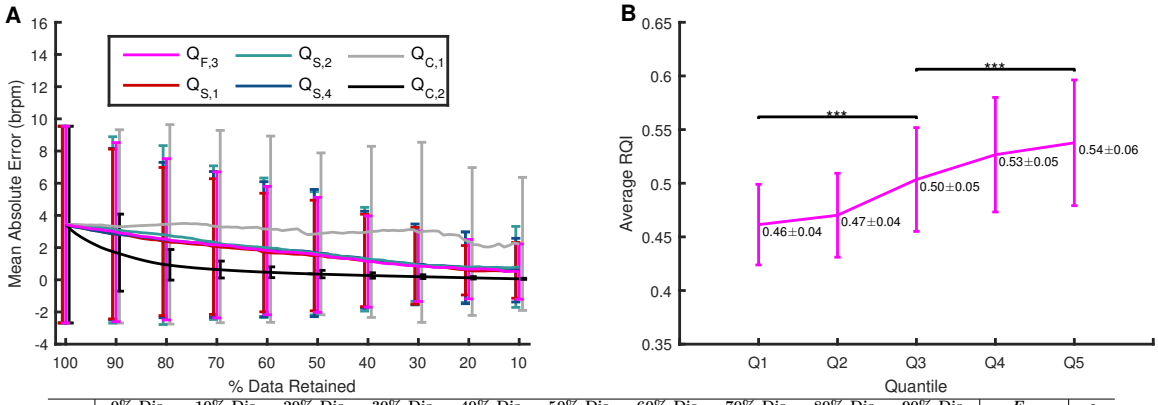
The performance results for $Q_{F,3}$ in $\mathcal{D}_{c,1}$ (Figure 6.12) show performance similar to the performance seen for $Q_{F,1}$ in $\mathcal{D}_{c,1}$. Where $Q_{F,3}$ does not outperform the individual RQIs, it is only slightly outperformed by a single individual RQI. E_{50} for $Q_{F,3}$ is 1.88 brpm compared to 1.92 for $Q_{S,1}$, 1.79 for $Q_{S,2}$, and 1.77 for $Q_{S,4}$. Furthermore, Spearman's ρ for $Q_{F,3}$ is -0.53, which outperforms all of the individual RQIs. The robustness measures for $Q_{F,3}$ in $\mathcal{D}_{c,1}$ are similarly high with the standard deviations steadily decreasing as



	N	% of Total	Avg. Dif. from Ideal	Avg. Dif. of Other Three
$\boldsymbol{Q_{S,1}}$	699	15.9	2086.4 ± 785.4	1170.1 ± 851.7
$\boldsymbol{Q_{S,2}}$	725	16.5	2052.1 ± 806.9	1066.1 ± 831.5
$\boldsymbol{Q_{S,4}}$	782	17.7	2180.3 ± 821.3	1202.3 ± 853.7
$\boldsymbol{Q_{F,2}}$	2201	49.9	2581.0 ± 840.2	702.1 ± 569.2

Figure 6.11: *Robustness comparison of $Q_{F,2}$ (green) to $Q_{S,1}$ (red), $Q_{S,2}$ (teal), and $Q_{S,4}$ (blue) in the $\mathcal{D}_m$ data set. The x-axis represents the rank difference ($|rgQ_{C,1} - rgQ_{X,Z}|$) and the y-axis represents sorting all observations based on the average difference of the three best-performing RQIs. (A) The rank difference of $Q_{F,2}$ and $Q_{S,[1,2,4]}$ where $Q_{S,1}$ is the worst-performing RQI. (B) The rank difference of $Q_{F,2}$ and $Q_{S,[1,2,4]}$ where $Q_{S,2}$ is the worst-performing RQI. (C) The rank difference of $Q_{F,2}$ and $Q_{S,[1,2,4]}$ where $Q_{S,4}$ is the worst-performing RQI. (D) The rank difference of $Q_{F,2}$ and $Q_{S,[1,2,4]}$ where $Q_{F,2}$ is the worst-performing RQI. (Table) The number of instances and percentage of instances in which a given RQI was the worst-performer in the data set and the average rank difference and standard deviation of that RQI compared to the average rank difference and standard deviation of the three best-performing RQIs.*

the worst data are discarded and with the quantile analysis of the RQI values showing statistically different values (for p <0.001) between both quantiles 1 and 3 and quantiles 3 and 5.

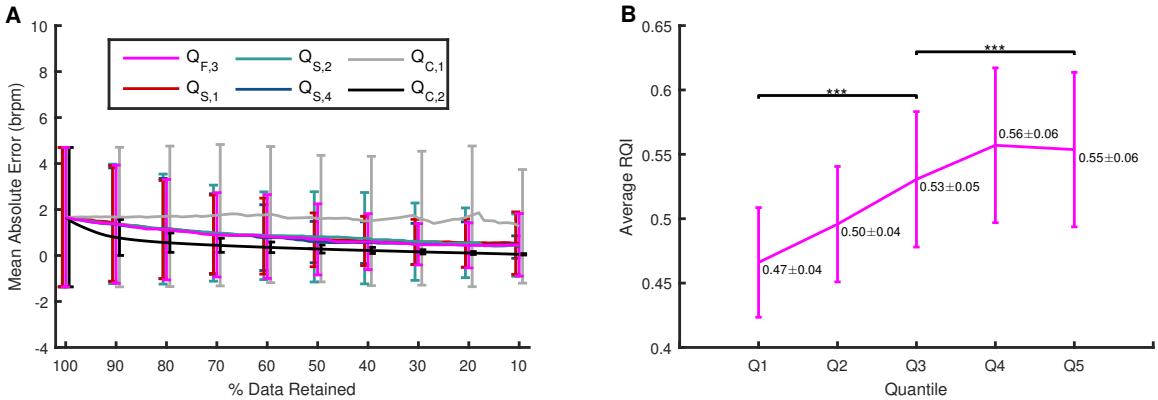


	0% Dis.	10% Dis.	20% Dis.	30% Dis.	40% Dis.	50% Dis.	60% Dis.	70% Dis.	80% Dis.	90% Dis.	E_{50}	ρ
$Q_{F,1}$	3.43 ± 6.11	2.96 ± 5.57	2.52 ± 5.01	2.17 ± 4.54	1.81 ± 3.99	1.55 ± 3.57	1.14 ± 2.84	0.84 ± 2.19	0.66 ± 1.85	0.51 ± 1.72	1.88 ± 1.02	-0.53
$Q_{S,1}$	3.43 ± 6.11	2.85 ± 5.27	2.38 ± 4.60	2.06 ± 4.22	1.69 ± 3.68	1.51 ± 3.44	1.20 ± 2.87	0.88 ± 2.39	0.59 ± 1.54	0.59 ± 1.73	1.92 ± 1.01	-0.50
$Q_{S,2}$	3.43 ± 6.11	3.10 ± 5.79	2.78 ± 5.56	2.31 ± 4.77	1.99 ± 4.33	1.64 ± 3.83	1.28 ± 3.22	0.93 ± 2.26	0.78 ± 2.19	0.80 ± 2.52	1.79 ± 1.05	-0.52
$Q_{S,4}$	3.43 ± 6.11	2.84 ± 5.33	2.47 ± 4.83	2.20 ± 4.53	1.90 ± 4.19	1.66 ± 3.95	1.25 ± 3.01	0.95 ± 2.51	0.75 ± 2.23	0.60 ± 1.98	1.77 ± 1.06	-0.48
$Q_{C,1}$	3.43 ± 6.11	3.32 ± 6.01	3.45 ± 6.20	3.31 ± 5.98	3.14 ± 5.79	2.86 ± 5.03	2.98 ± 5.32	2.95 ± 5.60	2.38 ± 4.59	2.23 ± 4.13	0.57 ± 1.17	-0.04
$Q_{C,2}$	3.43 ± 6.11	1.69 ± 2.39	0.93 ± 0.95	0.64 ± 0.52	0.47 ± 0.33	0.35 ± 0.23	0.27 ± 0.17	0.19 ± 0.12	0.12 ± 0.08	0.06 ± 0.04	3.07 ± 0.84	-1.00

Figure 6.12: *Performance of $Q_{F,3}$ in the $\mathcal{D}_{c,1}$ training and validation data set. (A) MAE of $Q_{F,3}$, controls ($Q_{C,[1,2]}$), and individual RQIs ($Q_{S,[1,2,4]}$) as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates p <0.001, ** indicates p <0.01, and * indicates p <0.05). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

The results in the two test sets are slightly less accurate and robust but suggest that $Q_{F,3}$ is a suitable fusion RQI. Accuracy results in $\mathcal{D}_{c,2}$ show that the ranked correlation for $Q_{F,3}$ ($\rho = -0.53$) is only outperformed by $Q_{S,2}$ and that E_{50} is quite good, compared to $Q_{C,2}$, but is outperformed by both $Q_{S,1}$ and $Q_{S,4}$. The robustness measures show a generally decreasing trend in standard deviations as the worst windows based on RQI value are discarded and the quantile analysis shows statistically relevant differences (for p <0.001) between quantiles 1 and 3 and quantiles 3 and 5 (Figure 6.13). Results for $\mathcal{D}_m$ show slightly worse results than for $\mathcal{D}_{c,2}$ but are still accurate and robust. The E_{50} value (1.41) is outperformed by $Q_{S,1}$ (1.49) and $Q_{S,2}$, but only slightly, and the ρ value is

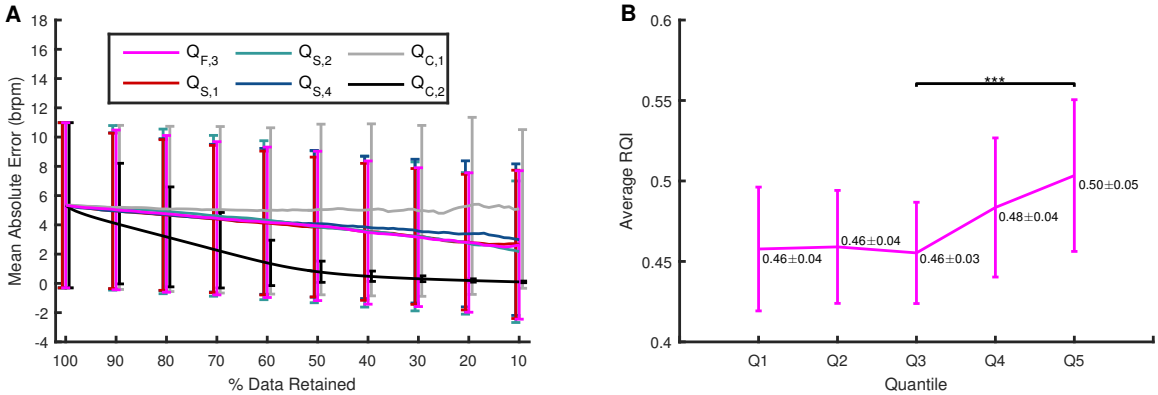
equal to the highest performing individual RQI (-0.38, $Q_{S,2}$). The robustness measures show that the RQI may not be particularly robust when considering the lowest values of the RQI as there is no noted difference between the quantile values in quantile 1 and 3. There is a statistically significant difference (p <0.001) between quantile 1 and 3, indicating discrimination between the highest RQI values.



	0% Dis.	10% Dis.	20% Dis.	30% Dis.	40% Dis.	50% Dis.	60% Dis.	70% Dis.	80% Dis.	90% Dis.	E_{50}	ρ
$Q_{F,1}$	1.67 ± 3.03	1.36 ± 2.57	1.12 ± 2.19	0.90 ± 1.83	0.83 ± 1.82	0.70 ± 1.55	0.60 ± 1.22	0.49 ± 0.90	0.44 ± 0.98	0.46 ± 1.36	0.97 ± 0.65	-0.53
$Q_{S,1}$	1.67 ± 3.03	1.39 ± 2.52	1.13 ± 2.13	0.93 ± 1.70	0.84 ± 1.65	0.68 ± 1.17	0.63 ± 1.07	0.59 ± 0.98	0.55 ± 1.05	0.54 ± 1.36	0.99 ± 0.61	-0.44
$Q_{S,2}$	1.67 ± 3.03	1.37 ± 2.60	1.15 ± 2.39	0.97 ± 2.09	0.86 ± 1.91	0.81 ± 1.96	0.75 ± 1.99	0.60 ± 1.68	0.55 ± 1.52	0.45 ± 1.36	0.86 ± 0.70	-0.54
$Q_{S,4}$	1.67 ± 3.03	1.35 ± 2.46	1.17 ± 2.18	0.93 ± 1.75	0.78 ± 1.43	0.58 ± 0.90	0.56 ± 0.91	0.51 ± 0.89	0.47 ± 0.99	0.37 ± 0.49	1.09 ± 0.59	-0.50
$Q_{C,1}$	1.67 ± 3.03	1.67 ± 3.03	1.70 ± 3.06	1.75 ± 3.07	1.78 ± 2.96	1.60 ± 2.75	1.50 ± 2.81	1.62 ± 2.91	1.70 ± 3.06	1.27 ± 2.47	0.06 ± 0.81	0.03
$Q_{C,2}$	1.67 ± 3.03	0.78 ± 0.78	0.56 ± 0.42	0.44 ± 0.31	0.36 ± 0.23	0.28 ± 0.18	0.22 ± 0.13	0.16 ± 0.09	0.11 ± 0.06	0.05 ± 0.03	1.38 ± 0.55	-1.00

Figure 6.13: *Performance of $Q_{F,3}$ in the $\mathcal{D}_{c,2}$ test data set. (A) MAE of $Q_{F,3}$, controls ($Q_{C,[1,2]}$), and individual RQIs ($Q_{S,[1,2,4]}$) as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates p <0.001, ** indicates p <0.01, and * indicates p <0.05). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

The comparative robustness analysis of $Q_{F,3}$ in $\mathcal{D}_m$ shows similar results as the results in $Q_{F,1}$ (Figure 6.15). The results show that $Q_{S,[1,2,4]}$ are the worst performers at similar rates and that when one of $Q_{S,[1,2,4]}$ is the worst performer, the degree to which it underperforms (compared to the other RQIs) is similar. $Q_{F,3}$, however, is the worst-performing RQI in only 4.9% of instances and has a much smaller difference between its average rank difference when it is the worst RQI compared to the average rank difference of the three individual RQIs in those instances. These comparative robustness results along



	0% Dis.	10% Dis.	20% Dis.	30% Dis.	40% Dis.	50% Dis.	60% Dis.	70% Dis.	80% Dis.	90% Dis.	E_{50}	ρ
$Q_{F,1}$	5.34 ± 5.65	5.02 ± 5.46	4.75 ± 5.36	4.45 ± 5.24	4.18 ± 5.14	3.93 ± 5.10	3.47 ± 4.89	3.16 ± 4.74	2.79 ± 4.76	2.63 ± 5.07	1.41 ± 1.02	-0.38
$Q_{S,1}$	5.34 ± 5.65	4.95 ± 5.31	4.67 ± 5.15	4.40 ± 5.03	4.13 ± 4.92	3.85 ± 4.78	3.52 ± 4.68	3.21 ± 4.64	2.82 ± 4.64	2.67 ± 5.07	1.49 ± 0.98	-0.37
$Q_{S,2}$	5.34 ± 5.65	5.16 ± 5.62	4.92 ± 5.63	4.62 ± 5.50	4.32 ± 5.43	3.87 ± 5.19	3.54 ± 5.16	3.21 ± 5.09	2.73 ± 4.84	2.16 ± 4.84	1.47 ± 1.03	-0.38
$Q_{S,4}$	5.34 ± 5.65	4.96 ± 5.33	4.70 ± 5.19	4.46 ± 5.05	4.24 ± 4.99	4.08 ± 5.01	3.83 ± 4.86	3.57 ± 4.91	3.38 ± 4.99	2.99 ± 5.18	1.26 ± 1.01	-0.33
$Q_{C,1}$	5.34 ± 5.65	5.19 ± 5.61	5.09 ± 5.65	5.03 ± 5.69	4.95 ± 5.68	5.05 ± 5.83	5.03 ± 5.88	4.96 ± 5.84	5.30 ± 6.05	5.09 ± 5.42	0.29 ± 1.10	-0.11
$Q_{C,2}$	5.34 ± 5.65	4.09 ± 4.12	3.18 ± 3.42	2.27 ± 2.58	1.40 ± 1.55	0.80 ± 0.72	0.49 ± 0.35	0.32 ± 0.20	0.20 ± 0.11	0.10 ± 0.06	4.54 ± 0.70	-1.00

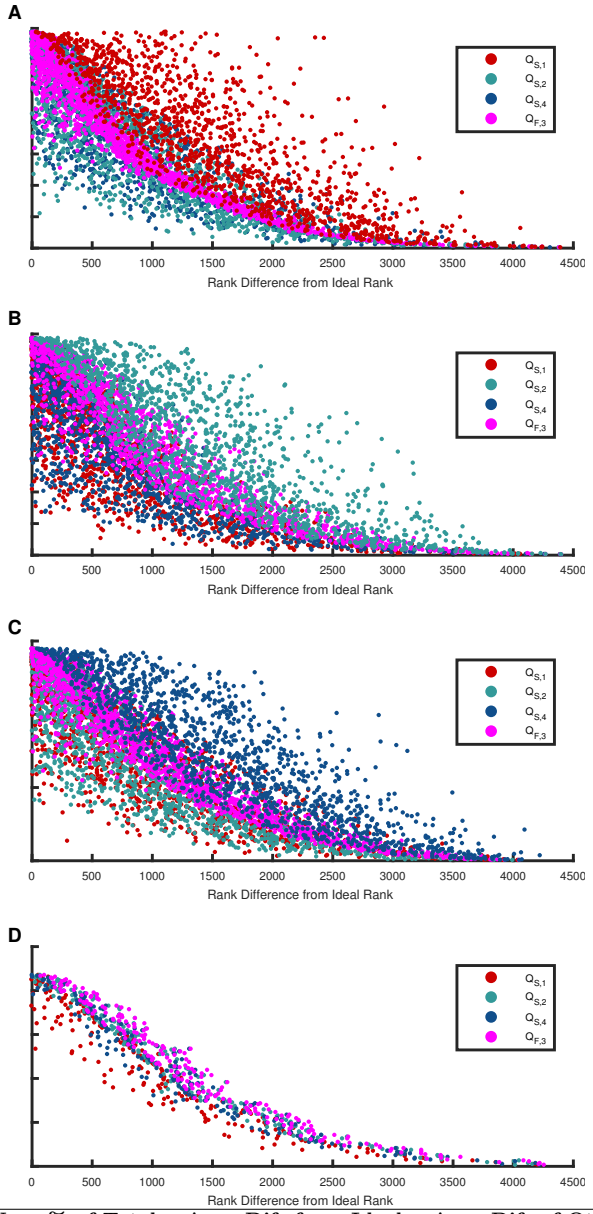
Figure 6.14: *Performance of $Q_{F,3}$ in the $\mathcal{D}_m$ test data set. (A) MAE of $Q_{F,3}$, controls ($Q_{C,[1,2]}$), and individual RQIs ($Q_{S,[1,2,4]}$) as windows are discarded based on comparison of RQIs to a threshold, T. Note, the horizontal axis shows the percentage of data retrained, which decreases as the value of T increases. Error bars represent standard deviation of the MAE at selected data retention percentages. (B) The average and standard deviation RQI values based on quantiles selected based on RR error. Bars represent results of Welch's t-test for Q1 vs. Q3 and Q3 vs. Q5 (*** indicates $p < 0.001$, ** indicates $p < 0.01$, and * indicates $p < 0.05$). (Table) The values of the MAE and standard deviations of the MAE shown in (A), the E_{50}, and Spearman's Rank Order Correlation.*

with the previous accuracy and robustness results suggest that $Q_{F,3}$ is a high-quality fusion RQI method.

6.4 Discussion

The motivation for creating fusion RQIs from the best-performing individual RQIs ($Q_{S,[1,2,4]}$) is that while the individual RQIs all seek to capture the quality of the underlying respiratory modulation of both the ECG and PPG, they do so using different signal processing techniques which vary the type of information that they capture. As a result, there is a benefit to considering all of the individual models together through a single, fusion RQI metric.

The goal of RQI fusion presents two primary challenges. The first is how to maintain adequate accuracy and robustness when one or more of the individual RQIs may be



	N	% of Total	Avg. Dif. from Ideal	Avg. Dif. of Other Three
$\boldsymbol{Q_{S,1}}$	1375	31.2	1556.6 ± 859.3	913.2 ± 816.8
$\boldsymbol{Q_{S,2}}$	1371	31.1	1549.4 ± 874.8	911.5 ± 819.2
$\boldsymbol{Q_{S,4}}$	1444	32.8	1669.8 ± 907.8	1025.2 ± 847.7
$\boldsymbol{Q_{F,3}}$	217	4.9	1459.5 ± 946.5	1278.7 ± 931.2

Figure 6.15: *Robustness comparison of $Q_{F,3}$ (pink) to $Q_{S,1}$ (red), $Q_{S,2}$ (teal), and $Q_{S,4}$ (blue) in the $\mathcal{D}_m$ data set. The x-axis represents the rank difference ($|rgQ_{C,1} - rgQ_{X,Z}|$) and the y-axis represents sorting all observations based on the average difference of the three best-performing RQIs. (A) The rank difference of $Q_{F,3}$ and $Q_{S,[1,2,4]}$ where $Q_{S,1}$ is the worst-performing RQI. (B) The rank difference of $Q_{F,3}$ and $Q_{S,[1,2,4]}$ where $Q_{S,2}$ is the worst-performing RQI. (C) The rank difference of $Q_{F,3}$ and $Q_{S,[1,2,4]}$ where $Q_{S,4}$ is the worst-performing RQI. (D) The rank difference of $Q_{F,3}$ and $Q_{S,[1,2,4]}$ where $Q_{F,3}$ is the worst-performing RQI. (Table) The number of instances and percentage of instances in which a given RQI was the worst-performer in the data set and the average rank difference and standard deviation of that RQI compared to the average rank difference and standard deviation of the three best-performing RQIs.*

neither accurate nor robust. The second is how to create a generalisable model that can be applied broadly. The present chapter set out to create and test multiple fusion RQIs to both capture all of the information available in the three best-performing individual RQIS and address these two primary challenges previously noted.

The results of this fusion analysis show that two of the fusion methods achieve the goals described. Both $Q_{F,1}$ and $Q_{F,3}$, based on linear regression and principal component analysis, respectively, are capable of fusing the three individual RQIs in a way that is both accurate and robust across multiple test data sets. The remaining fusion RQI, $Q_{F,2}$, does not address the second challenge noted above in that the model performs exceptionally on the training and validation set but that the results are not generalisable. The results for $Q_{F,1}$ and $Q_{F,3}$ do address both of the primary challenges described.

Firstly, the accuracy rates for both $Q_{F,1}$ and $Q_{F,3}$ are comparable to the best individual RQI in almost all of the data sets. It is unsurprising, however, that there is a slight loss of general accuracy in the fusion model compared to the best-performing individual RQI. Given that the fusion models are based on the individual RQIs, it is to be expected that even in the best fusion models, the worst-performing RQIs in any observation will factor into determining the fusion RQI and thus will decrease the overall accuracy of the fusion RQI compared to the best-performing RQI. The more important measure (and principal reason for implementing the fusion RQI) is the lack of knowledge about which individual RQI will be the worst-performing for any given observation. As can be noted in Figures 6.5, 6.11, and 6.15, each of the individual RQIs is the worst-performing for a given observation at a nearly equal rate. Further, as observed in the average rank difference of the individual RQI when it is the worst-performing compared to the other RQIs, it appears that there are cases where each individual RQI is far worse than all

of the other RQIs. This means that in some cases each individual RQI performs exceptionally but in other cases it performs poorly. These results also show that for the entire set of individual RQIs, these cases are not the same for all of the individual RQIs. This variability reinforces the purpose of the fusion RQIs: to serve as a modulator of robustness. Fusing all of the individual RQIs shows that while the accuracy of the fusion RQI might be slightly degraded, the fusion RQIs do not have exceptionally bad cases like the individual RQIs. Given the need for consistently high-quality respiratory estimation in a clinically relevant RR estimation model, the choice of improved robustness over slightly degraded accuracy is a key feature of the fusion RQIs.

The second challenge of creating the fusion RQIs is the generalisability across data sets. Figures 6.2, 6.8, and 6.12 show that the fusion RQIs performed as well as or better than the best-performing individual RQI in the $\mathcal{D}_{c,1}$ training and validation data set. However, the key consideration in generalisability is how the fusion data sets perform on unseen test data sets. In order to ensure generalisability of the models, the fusion RQIs were tested on two independent test data sets ($\mathcal{D}_{c,2}$ and $\mathcal{D}_m$) acquired from the CapnoBase and MIMIC-III databases. $Q_{F,1}$ and $Q_{F,3}$ performed well in both of these data sets, suggesting the generalisability of these models. Another important observation is that $Q_{F,1}$ and $Q_{F,2}$ generally performed better on the $\mathcal{D}_{c,2}$ than the $\mathcal{D}_m$ data set. This is unsurprising given that $\mathcal{D}_{c,2}$ was drawn from the same population and database as the training and validation set and thus it is likely that these observations closely resembled the observations in $\mathcal{D}_{c,1}$. $\mathcal{D}_m$ is instead completely independent of $\mathcal{D}_{c,1}$, and a far more clinically relevant data set, as the data acquisition is far less controlled and the underlying data generally have more signal noise. Given this, and despite the fact that the results for $Q_{F,1}$ and $Q_{F,3}$ do not perform as well in $\mathcal{D}_m$ versus $\mathcal{D}_{c,2}$, the performance

of these two fusion RQIs in this data set is quite accurate and robust and is a strong indicator of the general applicability of these models.

Based on the results described, both $Q_{F,1}$ and $Q_{F,3}$ appear to be accurate, robust, and generalisable fusion RQIs of $Q_{S,[1,2,4]}$. However, $Q_{F,1}$ performs slightly better than $Q_{F,3}$ in both accuracy and robustness. A comparison of E_{50} and Spearman's ρ for $Q_{F,1}$ and $Q_{F,3}$ is shown in Table 6.3. Therefore, for the time-series analysis presented in Chapter 7, is the only RQI that will be considered is $Q_{F,1}$.

Table 6.3: *Comparative performance statistics for $Q_{F,1}$ and $Q_{F,3}$ in the $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$ data sets.*

		$Q_{F,1}$	$Q_{F,3}$
$\mathcal{D}_{c,2}$	E_{50}	1.04 ± 0.39	0.97 ± 0.42
	ρ	-0.49	-0.53
$\mathcal{D}_m$	E_{50}	1.52 ± 0.43	1.41 ± 0.44
	ρ	-0.39	-0.38
	% Worst-Perform	4.0	4.9

6.5 Conclusion

This chapter presented the creation and analysis of three fusion RQIs that took as inputs the three best individual RQIs from Chapter 5 based on the FFT ($Q_{S,1}$), autocorrelation ($Q_{S,2}$), and autoregression ($Q_{S,4}$). The fusion RQIs models created were based on linear regression ($Q_{F,1}$), support vector regression ($Q_{F,2}$), and principal component analysis ($Q_{F,3}$). The models were created using the $\mathcal{D}_{c,1}$ data set as a training and validation set and the $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$ data sets as test sets. The motivation for creating fusion RQIs was to improve on the overall robustness compared to the individual RQIs while maintaining accuracy and generalisability. Both $Q_{F,1}$ and $Q_{F,3}$ met this standard while $Q_{F,2}$ lacked generalisability. Notably, both $Q_{F,1}$ and $Q_{F,3}$ demonstrated accuracy and robustness in

$\mathcal{D}_m$, a clinically relevant data set where the observations were drawn from a completely different population than the training and validation sets. Due to the slightly improved performance of $Q_{F,1}$ over $Q_{F,3}$, $Q_{F,1}$ will be used for further analysis.

In order to demonstrate the clinical relevance of the RQIs that has been described and analysed in Chapters 5 and 6, Chapter 7 will discuss the creation of an RR estimation algorithm based on $Q_{F,1}$ which will improve on existing RR estimation algorithms for both individual observations and through a time-series of observations. Chapter 7 will also discuss the ability of the described RR estimation algorithm to predict clinical decompensation due to cardiac or respiratory failure.

Chapter 7

RQI-Based RR Estimation and Time-Series Analysis

This chapter presents the validation of the use of RQIs to improve the continuous estimation of respiratory rate. The discussion is split into two parts. The first part demonstrates a method, which will be termed RQIFusion, by which the RQI values can be used to obtain one estimate of RR for a single time window from multiple RR modulations. This method is compared to previously-defined methods of RR fusion.[1] The analysis is conducted using the data from $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, and $\mathcal{D}_m$. The second part validates the use of the RQI as a quality checkpoint in the RR estimation methodology presented in Figure 3.1, as well as the use of the RQI for real-time, continuous clinical monitoring of RR. This analysis is conducted using the Dialysis III data set ($\mathcal{D}_d$) that was described in Section 4.3.

7.1 Techniques

Previous chapters have presented techniques for the frequency analysis of time-series and for multi-factor fusion models. The current chapter presents and describes methods for the filtering, smoothing, and estimating the RR of continuous, non-uniformly sampled

[1] These methods are described in Section 3.4.

time-series information. Two techniques will be described: median-filtering and cubic spline interpolation.

7.1.1 Median Filtering

Median filters or smoothers, first popularised in signal processing by Gallagher and Wise (1982), are a non-linear method for time-series noise reduction [224, 225]. The median filter works by processing a vector of data points centred around x_n and selecting the median amongst those data points to represent the value of x_n. Thus for a time-series of data, the set of vectors of data points can be defined as:

$$\mathbf{x}_n = \{x_{(n-N_L)} \dots x_n \dots x_{(n-N_R)}\} \tag{7.1}$$

where $\mathbf{x}_n$ represents the vector of values centred around n. N_L and N_R represent the lower and upper bounds of the median filter, respectively [225]. In order to obtain median-filtered values at all time points in the original signal, a series of values must be appended to both ends of the signal that are of length N_L at the beginning and length N_R at the end [225]. Most often these points are set to the value of the beginning and end points such that:

$$x_{(1-N_L)\dots 1} = x_1 \tag{7.2}$$

and

$$x_{n\dots(n+N_R)} = x_n \tag{7.3}$$

Thus, the entire time-series used to create the vectors for median filtering is $x_{(1-N_L)\dots(n+N_R)}$. Most often in median filtering, $N_L = N_R$ and thus the size of the median vector is $N_L + N_R + 1$ [225]. In real-time applications, this is not possible. As a result, $N_R = 0$

and the selection of the median point for x_n only relies on x_n and N_L previous points. Given the median vector, $\mathbf{x}_n$ the filtered value of x_n, called y_n is defined as:

$$y_n = \tilde{\mathbf{x}}_n \tag{7.4}$$

where $\tilde{\mathbf{x}}_n$ is the median of $\mathbf{x}_n$ where $\mathbf{x}_n$ is the mean of the two median points if N_L+N_R+1 is odd [225].

Median filters are particularly useful when the underlying signal is expected to contain a series of constant regions with transient noise [225]. Conceptually in median filtering, the signal can be subdivided into four distinct regions: (1) root regions, (2) edge regions, (3) impulses, and (4) oscillations. Root regions are defined as regions of size at least $N_L + N_R + 1$ where all values of y_n for that region are equal. Edge regions are the regions between two root regions. Impulses are regions that have a series of values not longer than $N_L + N_R$ that dissect two identical root regions. Oscillations are regions that are not root regions, edge regions, or impulses [225]. Median filters work best when the underlying signal is expected to be comprised of root regions, edge regions, and slowly moving oscillations, because median filters will remove impulses and fast oscillations.

7.1.2 Cubic Spline Interpolation

Spline interpolation is another method of providing a continuous, smooth time-series. Spline interpolation relies on fitting an m-order polynomial between each pair of data points where the cubic polynomial is the most commonly used [226]. Cubic spline interpolation is preferred to fitting a high-degree, continuous m-order polynomial as it avoids Runge's phenomenon [227].

Cubic spline interpolation relies on having a known set of data pairs, $\{x_n, y_n\}$, where each x_n is unique. Given this, the cubic spline interpolation can be given as:

$$S(x) = \begin{cases} C_1(x), & x_0 \leq x \leq x_1 \\ C_n(x), & x_{n-1} \leq x \leq x_n \\ C_N(x), & x_{N-1} \leq x \leq x_N \end{cases} \tag{7.5}$$

where each C_n represents a unique cubic function between each data pair of the form:

$$C_n(x) = a_n x^3 + b_n x^2 + c_n x + d_n \tag{7.6}$$

In order to ensure the spline interpolation between points is a smooth function, two conditions are placed on the cubic functions:

$$C'_n(x_n) = C'_{n+1}(x_n) \tag{7.7}$$

and

$$C''_n(x_n) = C''_{n+1}(x_n) \tag{7.8}$$

that is, these conditions constrain the first and second derivatives of the cubic spline functions to be equal to the first and second derivatives of the next cubic spline function at the edge points, x_n [226]. The result is that the entire system of equations results in N equations with $4N$ unknowns (each cubic spline coefficient) and $4N - 2$ known constraints (the two known end point constraints for each cubic spline, $\{x_n, y_n\}$, and the two conditions defined in Equations 7.7 and 7.8). In order to solve the cubic spline

system two more constraints are required. These constraints are based on the beginning and end conditions of the cubic spline system at x_0 and x_N where three options for these constraints are:

$$C_1'(x_0) = C_N'(x_N) = 0 \tag{7.9}$$

or

$$C_1''(x_0) = C_N''(x_N) = 0 \tag{7.10}$$

or

$$C_1'''(x_0) = C_2'''(x_1) \tag{7.11}$$

$$C_{N-1}'''(x_{N-1}) = C_N'''(x_N) \tag{7.12}$$

which are known as clamped (Equation 7.9 and Equation 7.10) or not-a-knot (Equations 7.11 and 7.12) conditions. The first condition (Equation 7.9) constrains the slope of the first and last cubic splines at x_0 and x_N to be zero. The second constraint (Equation 7.10) forces the concavity of the functions to be zero at x_0 and x_N. The third constraint (Equations 7.11-7.12) forces continuity of the third derivative between the first and second and the penultimate and ultimate data points. This means that the first two and last two intervals each represent an individual polynomial function [226, 228]. Given these constraints, the set of cubic spline functions, $S(x)$, can be solved as a system of linear equations. There are two primary shortcomings of the cubic spline method. The first is that, because it relies on knowing the entire set of data prior to fitting the cubic spline functions, it is less useful in real-time applications unless the data set is windowed or more advanced cubic spline methods are employed [229, 230]. The second is that the spline fit is constrained to go through the known data points and does not permit

the realistic assumption that these data points may have some underlying measurement error.

7.2 Assessment of RQIFusion

Previous work has shown that our RQI methodology is capable of evaluating the underlying quality of the respiratory information in the respiratory modulations extracted from the ECG and PPG. However, this metric is not clinically relevant if it cannot be used to improve RR estimation. The present analysis aims to demonstrate that an RQI-based RR estimation method, hereafter termed RQIFusion, presents a marked improvement to the current standards of RR estimation fusion. This will be demonstrated not only on an individual window basis but also for individual subject time-series.

7.2.1 Methods

RR Estimation

The previous work has described a highly accurate and robust fusion RQI based on linear regression ($Q_{F,1}$). From the feature-based RR modulation extraction method described in Section 5.2.1, each window of data contains six respiratory modulations: AM, FM, and BW from both ECG and PPG. Therefore, in order to obtain a single RR estimate per time window, the six RR estimates obtained from the six RR modulations must be fused. Given the known RR modulation quality information available in the RQIs, the most intuitive way to achieve RR estimation fusion is through a weighted average based on RQI values. This can be defined as:

$$R_{F,1} = \frac{\sum Q'_{F,1,j} R_{M,j}}{\sum Q'_{F,1,j}} \tag{7.13}$$

where $Q'_{F,1,j}$ is the transformed LR fusion RQI described in Equation 6.45 for each of the modulations $j = 1 \ldots 6$. $R_{M,j}$ is the RR estimate of the given modulation, j, using the *ARSpec* algorithm, and $R_{F,1}$ is the weighted average RR estimate based on the linear regression fusion RQI for a given time window.

An extension of this weighted average is that the RR estimate can be obtained using the RQIs that exceed a cut-off threshold. This technique, RQIFusion, can be defined as:

$$R_{F,i} = \frac{\sum Q'_{F,1,j_T} R_{M,j_T}}{\sum Q'_{F,1,j_T}} \tag{7.14}$$

where all of the variables are the same as Equation 7.13, except that T represents a threshold for the values $Q'_{F,1,j}$ where only those modulations for which the RQIs are above the threshold are included in the weighted average. When $T = 0$ (RQIFusion_0), the result is equivalent to the simple weighted average technique in Equation 7.13. The hypothesised trade-off is that as T increases, not only is the RR error expected to decrease, but so does the number of windows without at least one RQI modulation value where $Q'_{F,i,j} \geq T$. In all windows for which $Q'_{F,i,j} < T$, no RR estimation is obtained. Tuning the value of T can therefore be based on the accuracy requirements of the specific monitoring application. For example, one might accept fewer windows having an RR estimate recorded as long as these estimates were highly reliable.

RQIFusion will be compared to two control RR fusion estimation methods. The first of these controls is the average of the six RR estimations, $R_{M,j}$, without consideration of the RQI values (RR Average). The second method is based on the work by Karlen *et al.* (2013), which has been termed Smart Fusion [66]. The latter is meant to improve RR estimation by only providing RR fusion estimates when the agreement between the panel

of RR estimates for a window, $R_{M,j}$, is sufficiently high. The agreement between these estimates is determined via the standard deviation between the panel of estimates such that only windows where the standard deviation is above 4 brpm are considered to be in sufficient agreement. For windows that are deemed acceptable, the RR fusion estimate is obtained via an average of all of the available RR modulation estimates [66]. The Smart Fusion method was initially described using only the RR modulations obtained from PPG. In order to allow a fair comparison of the Smart Fusion method to the RQI-based methods, three Smart Fusion variations will be considered. The first variation, "Smart Fusion" is based on the original work where only the PPG RR estimations, $R_{M,j}$ for $j = 4 \dots 6$, are used. The second is "modified Smart Fusion" where the original technique is used but the method is applied to the entire panel of $R_{M,j}$ for $j = 1 \dots 6$. The final is "ECG Smart Fusion" which is identical to the original method except that only ECG modulations, $R_{M,j}$ for $j = 1 \dots 3$, are considered.

Threshold Determination

The value of T for RQIFusion can be arbitrarily set. To show the broad applicability of using different values of T, this analysis makes use of five different thresholds. The first three thresholds are based on specific selection parameters determined from the $\mathcal{D}_{c,1}$ training and validation data sets. The first threshold value, called the high threshold, T_{HT} ($\text{RQIFusion}_{\text{HT}}$), is the value of T in $\mathcal{D}_{c,1}$ where 90% of all windows have at least one modulation with a value above T_{HT} and thus have a fused RR estimate $R_{F,1}$. The second threshold, called the low threshold, T_{LT} ($\text{RQIFusion}_{\text{LT}}$), is the lowest value of T such that 100% of all windows in $\mathcal{D}_{c,1}$ have at least one modulation with a value above T_{LT}. The third threshold, called the Smart Fusion threshold, T_{SF} ($\text{RQIFusion}_{\text{SF}}$), is the

threshold where the percentage of windows with RR estimates in $\mathcal{D}_{c,1}$ is equivalent to the percentage of windows with RR estimates using the Smart Fusion method. The final two thresholds are determined based on the specific data set with which they are being applied, and are meant to provide direct comparisons to (1) RR Average, (2) Smart Fusion, and (3) modified Smart Fusion. These thresholds are specifically determined in the $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$ data sets. The first threshold is the value at which T results in an RR error that is equivalent to the comparator RR fusion estimation method (1) - (3), T_{MAE} ($\text{RQIFusion}_{\text{MAE}}$). The second is the threshold value at which T results in a data discard percentage that is equivalent to the data discard percentage of the comparator RR fusion estimation method (1) - (3), T_{DIS} ($\text{RQIFusion}_{\text{DIS}}$). In addition, the weighted average threshold, RQIFusion_0, will also be considered.

Subject Specific Analysis

The described methods will be applied to the individual subjects in the $\mathcal{D}_{c,2}$ data set. Firstly, the individual 8 minute time-series data for each subject in the $\mathcal{D}_{c,2}$ data set will be considered and compared to the reference RR, Smart Fusion, and modified Smart Fusion. This analysis will specifically make use of $\text{RQIFusion}_{\text{HT}}$ in order to ensure a consistent methodology across the subject population of the RQI-based RR estimate. The $\text{RQIFusion}_{\text{HT}}$ waveform is smoothed using a three-point median filter where $N_L = 2$ and $N_R = 0$ in order to simulate a real-time application. Two data points will be appended to the beginning of the signal, as described in Equation 7.1.1 so that the filter will provide a value at the first time point of the signal. In instances where estimates do not exist for a time point, no median filter value will be given (this will constitute a gap in monitoring).

Secondly, we will calculate the per-subject frequency with which windows are dropped by various RR estimation methods, including Smart Fusion, modified Smart Fusion, $\text{RQIFusion}_{\text{LT}}$, and $\text{RQIFusion}_{\text{HT}}$ in $\mathcal{D}_c$ and $\mathcal{D}_m$. In order to obtain a more robust estimate of the expected frequency with which windows are discarded in the $\mathcal{D}_c$ data set, four unique test sets of ten subjects each are extracted.[2] The linear regression RQI method described in Section 6.2.2 is applied to each of the training and validation sets which remain after each unique test set is extracted. These unique coefficients are applied to their respective held-out test sets and used to obtain the LR fusion RQI, $Q'_{F,1}$, for each modulation in the test sets. Based on the RR estimates and the values for $Q'_{F,1}$, RR estimates are obtained using Smart Fusion, modified Smart Fusion, $\text{RQIFusion}_{\text{LT}}$, and $\text{RQIFusion}_{\text{HT}}$. The number of windows dropped for each subject is found based on each method (i.e. each subject has 15 unique windows, the number of these windows for which an RR estimate is not available is determined based on each method). The frequency with which a given number of windows is dropped amongst all patients is found. For both $\mathcal{D}_c$ and $\mathcal{D}_m$ the frequency with which windows are dropped is smoothed for visualisation by applying a 100-point linear interpolation followed by a 3-point median filter with the same parameters as described previously.

7.2.2 Results

Comparison of RQIFusion to Control RR Estimation

Figure 7.1 shows the mean absolute RR error compared to the percentage of windows discarded (these are windows where all six RR modulations had RQI values below T) of RQIFusion (defined in Equation 7.14) where the threshold values, T, are not shown but

[2] The first test set used is the original $\mathcal{D}_{c,2}$ test set.

are increased for $T \in [0\ 1]$ for $\mathcal{D}_{c,1}$. These results are compared to the static values of the MAE and the percentage of discarded windows for the control RR fusion estimation methods: RR Average, Smart Fusion, and modified Smart Fusion.

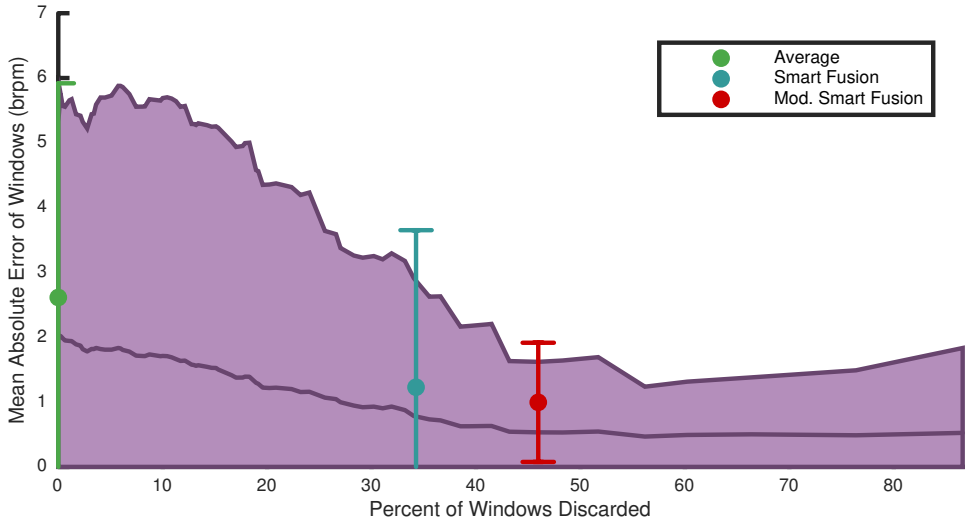


Figure 7.1: *MAE of RR estimates for windows vs. the percentage of windows discarded for RQIFusion as threshold T is increased in the $\mathcal{D}_{c,1}$ data set (middle purple line). Shaded purple area shows MAE $\pm$ SD. Results are shown for the average RR estimate for a window (green), Smart Fusion (blue), and modified Smart Fusion (red). Upper and lower bounds represent the standard deviation of the MAE over all windows, shown plotted around the MAE (purple).*

This analysis shows that RQIFusion outperforms all three comparator RR fusion estimation methods. When the percentage of windows discarded in RQIFusion is equivalent to any of the comparator methods, the MAE is lower. At the equivalent discard MAE, the discard percentage of RQIFusion is lower.

This trend is further demonstrated in Table 7.1 where the static values for the MAE and percentage of windows discarded using the control RR fusion estimation methods in $\mathcal{D}_{c,1}$ are shown in the first two columns. In the third column in Table 7.1, the percentage of windows discarded in RQIFusion is fixed to that of the control method, $\text{RQIFusion}_{\text{DIS}}$, and the corresponding MAE is shown. In the fourth column of Table 7.1, the complement is shown where the MAE of RQIFusion is fixed based on the control method, $\text{RQIFusion}_{\text{MAE}}$, and the percentage of windows discarded for the RQI-based

estimation method is shown.[3] It can be observed from Table 7.1 that when one feature, discard percentage or MAE, is used to fix T in RQIFusion, the value of the other feature in the RQI method outperforms the control, in most cases significantly. For example, when the value for T is set such that the percentage of windows discarded is 46.0%, which represents $\text{RQIFusion}_{\text{DIS}}$ fixed to the percentage of windows discarded by modified Smart Fusion, the MAE of $\text{RQIFusion}_{\text{DIS}}$ is 0.47 brpm less than modified Smart Fusion (MAE of 0.53 compared to 1.00).

Table 7.1: *Equivalent percentage of windows discarded (Dis. %) or equivalent MAE for RQIFusion vs. comparator methods in $\mathcal{D}_{c,1}$.*

			RQI-based Estimation	
Current Methods	**MAE**	**Dis. %**	**Equiv. MAE (Fixed Dis. %)**	**Equiv. Dis. % (Fixed MAE)**
RR Average	2.61	0.0%	1.98	
SmartFusion	1.23	34.3%	0.79	20.9%
Mod. SmartFusion	1.00	46.0%	0.53	27.0%

Determination of RQI Thresholds

As described in Section 7.2.1, three thresholds, T_{HT}, T_{LT}, and T_{SF}, were determined based on the percentage of windows that were discarded in the $\mathcal{D}_{c,1}$ training and validation data set. The value for $T_{HT} = 0.44$ was set as the value in which 90% of all windows had an RR estimate. The value of $T_{LT} = 0.11$ was set as the value where 100% of all windows had an RR estimate. Finally, the value of $T_{SF} = 0.84$ was set as the value where the percentage of windows without estimates was equivalent to the percentage seen when applying Smart Fusion to the data set. This discard rate was 34.3%, which can be noted in Figure 7.1.[4] These results, shown in Figure 7.2, confirm that as T is

[3] Because the $\text{RQIFusion}_{\text{MAE}}$ has a lower MAE than RR Average, no value can be shown for $\text{RQIFusion}_{\text{MAE}}$ compared to RR Average.

[4] The percentage of windows discarded refers specifically to the number of windows where at least one respiratory modulation has a fusion RQI value greater than T. It is not a measure of the number of modulations that have been discarded.

increased, the MAE in brpm decreases while the percentage of windows without RR estimates increases. These results also show that the selection of the RQI threshold has a meaningful effect on both the RR error and the number of windows for which there are RR estimates.

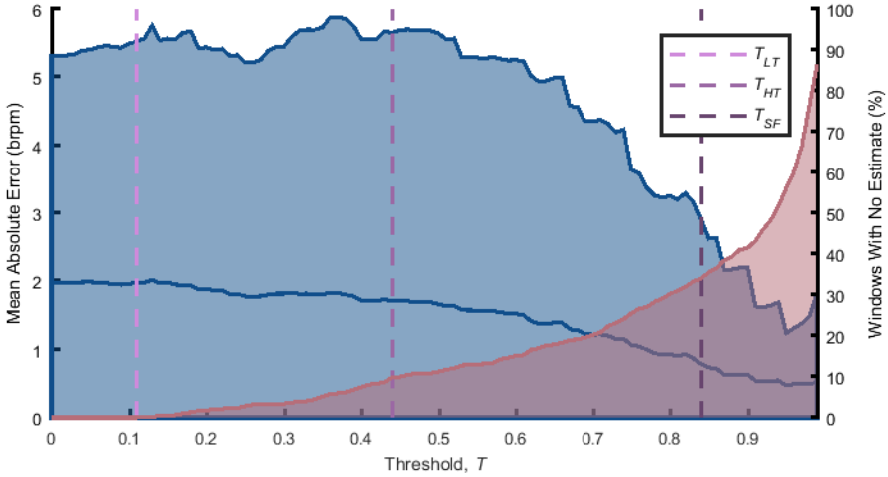


Figure 7.2: *MAE (middle blue line, left axis) ± SD (shaded blue area) and the number of windows with no estimate (red line, right axis) for $\mathcal{D}_{c,1}$ vs. increasing values of threshold T (horizontal axis). The dashed lines represent the values of T_{LT} (light purple, leftmost), T_{HT} (mid-purple, centre-most), and T_{SF} (dark purple, rightmost).*

General Applicability of RQI Thresholds

With the selection of the exemplar RQI threshold values using $\mathcal{D}_{c,1}$, this analysis shows the performance of these three exemplar thresholds for RR estimation in RQIFusion: $\text{RQIFusion}_{\text{HT}}$, $\text{RQIFusion}_{\text{LT}}$, and $\text{RQIFusion}_{\text{SF}}$ along with RQIFusion_0 in $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, and $\mathcal{D}_m$ compared RR Average, Smart Fusion, and modified Smart Fusion. These results are shown in Table 7.2.

These results show that the trade-offs demonstrated in Figure 7.2 in the $\mathcal{D}_{c,1}$ apply to both $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$ as well. When comparing one of the control methods to any of the thresholds it can be observed that even in instances where a control method achieves either a lower error or a lower discard rate, the threshold method has vastly

outperformed the control method in the other metric. For example, while modified Smart Fusion has a slightly lower MAE than RQIFusion$_0$ in $\mathcal{D}_m$ (3.33 ± 3.00 vs. 3.68 ± 3.42), the percentage of windows discarded using modified Smart Fusion is 73.0% while no windows are discarded using RQIFusion$_0$. This observation is particularly important as it applies to both test sets. In particular, it applies to $\mathcal{D}_m$ which is completely independent of $\mathcal{D}_c$ and contains a more clinically relevant patient population.

Table 7.2: *Performance of exemplar threshold models (median, MAE, and % discard) compared to existing fusion methods for all three data sets, $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, and $\mathcal{D}_m$.*

	$\mathcal{D}_{c,1}$			$\mathcal{D}_{c,2}$			$\mathcal{D}_m$		
	Median (Q1-Q2)	MAE (± SD)	% Dis.	Median (IQR)	MAE (± SD)	% Dis.	Median (IQR)	MAE (± SD)	% Dis.
Avg.	1.34 (0.63 - 3.26)	2.61 ± 3.30	0.0%	0.81 (0.39 - 1.39)	1.06 ± 0.93	0.0%	3.54 (1.72 - 5.88)	4.04 ± 2.90	0.0%
SF	0.62 (0.26 - 1.18)	1.23 ± 2.42	34.0%	0.46 (0.23 - 0.92)	0.68 ± 0.61	18.0%	2.83 (0.66 - 8.51)	4.61 ± 4.50	58.8%
Mod. SF	0.82 (0.41 - 1.31)	1.00 ± 0.92	46.0%	0.65 (0.31 - 1.08)	0.76 ± 0.56	26.7%	2.20 (1.21 - 4.51)	3.33 ± 3.00	73.0%
T_0 (0.00)	0.80 (0.31 - 1.87)	1.98 ± 3.34	0.0%	0.56 (0.24 - 0.88)	0.70 ± 0.61	0.0%	2.49 (0.98 - 5.63)	3.68 ± 3.42	0.0%
T_{LT} (0.11)	0.77 (0.31 - 1.75)	1.98 ± 3.55	0.0%	0.63 (0.31 - 1.20)	0.86 ± 0.72	0.0%	3.09 (1.23 - 5.86)	3.97 ± 3.45	1.2%
T_{HT} (0.44)	0.49 (0.24 - 1.08)	1.71 ± 3.94	10.0%	0.43 (0.25 - 0.77)	0.71 ± 0.89	1.3%	0.72 (0.27 - 4.91)	3.12 ± 4.39	23.2%
T_{SF} (0.84)	0.34 (0.17 - 0.62)	0.79 ± 2.13	34.0%	0.34 (0.21 - 0.57)	0.53 ± 0.82	14.7%	0.41 (0.15 - 1.07)	2.38 ± 5.09	66.3%

Consideration of ECG and PPG Alone

In some medical applications it is foreseeable that waveform data will only be available from one modality, ECG or PPG. Given that the Smart Fusion method was implemented using only PPG data, this analysis presents a direct comparison of Smart Fusion with the RQI-based fusion estimation method when both methods only consider ECG or PPG. Table 7.3 shows the comparative performance of Smart Fusion and RQIFusion in the $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, and $\mathcal{D}_m$ data sets when only one waveform modality is considered. In each analysis, the percentage of data discarded using the single modality in Smart Fusion is used to fix the data discarded in RQI Fusion (this is a single modality modified version of RQIFusion$_{\text{DIS}}$).[5] The exemplar thresholds (T_{HT}, T_{LT}, and T_{SF}) were not applied

[5]For the ECG analysis, SmartFusion was applied to the three ECG modulations alone. While this was not done in the original SmartFusion publication, Karlen *et al.* (2013) [66], this was done here to allow for a consistent ECG-alone comparison.

in this analysis as their performance was originally determined when both ECG and PPG modalities were available. The more direct analysis of fixing the percentage of windows discarded to Smart Fusion was deemed more relevant. These results show that, in all cases, $\text{RQIFusion}_{\text{DIS}}$ has lower MAE and median error than ECG or PPG Smart Fusion. This suggests that, even in situations where the data available are constrained, RQIFusion outperforms Smart Fusion. In many instances, the differences in performance are very stark. In particular this can be seen in the consideration of the PPG in $\mathcal{D}_m$ where $\text{RQIFusion}_{\text{DIS}}$ reduces the MAE by over twice as much as PPG Smart Fusion (2.28 ± 3.83 vs. 4.79 ± 4.55).

Table 7.3: *Median (IQR), mean absolute error (± SD), and % of windows discarded for ECG and PPG Smart Fusion and the corresponding threshold value T for $RQIFusion_{DIS}$ when % discard values are set equal and either PPG modulations or ECG modulations are considered in $\mathcal{D}_{c,1}$, $\mathcal{D}_{c,2}$, and $\mathcal{D}_m$.*

Data Set	Modality	Method	Median (IQR)	MAE (± SD)	% Dis.
$\boldsymbol{\mathcal{D}_{c,1}}$	**ECG**	**SF (ECG Only)**	0.72 (0.33 - 1.61)	1.40 ± 2.44	34.5%
$\boldsymbol{\mathcal{D}_{c,1}}$	**ECG**	$\boldsymbol{T = 0.57}$	0.45 (0.23 - 0.95)	1.23 ± 3.12	34.3%
$\boldsymbol{\mathcal{D}_{c,1}}$	**PPG**	**SF (PPG Only)**	0.62 (0.25 - 1.34)	1.44 ± 2.54	34.7%
$\boldsymbol{\mathcal{D}_{c,1}}$	**PPG**	$\boldsymbol{T = 0.75}$	0.34 (0.16 - 0.66)	0.84 ± 2.19	33.8%
$\boldsymbol{\mathcal{D}_{c,2}}$	**ECG**	**SF (ECG Only)**	0.63 (0.22 - 1.30)	0.92 ± 0.96	24.7%
$\boldsymbol{\mathcal{D}_{c,2}}$	**ECG**	$\boldsymbol{T = 0.71}$	0.47 (0.21 - 0.70)	0.68 ± 1.29	24.0%
$\boldsymbol{\mathcal{D}_{c,2}}$	**PPG**	**SF (PPG Only)**	0.47 (0.22 - 1.15)	0.87 ± 0.99	18.0%
$\boldsymbol{\mathcal{D}_{c,2}}$	**PPG**	$\boldsymbol{T = 0.69}$	0.37 (0.17 - 0.67)	0.52 ± 0.62	18.0%
$\boldsymbol{\mathcal{D}_m}$	**ECG**	**SF (ECG Only)**	1.40 (0.45 - 6.04)	3.57 ± 4.39	55.2%
$\boldsymbol{\mathcal{D}_m}$	**ECG**	$\boldsymbol{T = 0.43}$	0.63 (0.26 - 4.10)	3.13 ± 5.16	55.1%
$\boldsymbol{\mathcal{D}_m}$	**PPG**	**SF (PPG Only)**	3.59 (0.68 - 8.18)	4.79 ± 4.55	58.9%
$\boldsymbol{\mathcal{D}_m}$	**PPG**	$\boldsymbol{T = 0.68}$	0.43 (0.18 - 1.35)	2.28 ± 3.83	58.6%

Subject Specific Comparison

In addition to the aggregate analyses presented in the previous sections, this section considers the subject-specific analyses. The first analysis, shown in Figure 7.3, considers

the RR estimates obtained in two exemplar subjects in the $\mathcal{D}_{c,2}$ data set using Smart Fusion, modified Smart Fusion, and RQIFusion$_{\text{HT}}$.[6] These results show that when considering a subject-specific time-series, RQIFusion$_{\text{HT}}$ is more accurate compared to the reference and discards fewer windows. The two exemplar cases in Figure 7.3 are particularly important because they demonstrate this effectiveness in two particular cases. The first case in Figure 7.3 (left panel) shows the accuracy of RQIFusion$_{\text{HT}}$ compared to the reference when the RR is highly variable. The second case, shown in Figure 7.3 (right panel), shows that RQIFusion$_{\text{HT}}$ has highly improved time-series completeness compared to the Smart Fusion and modified Smart Fusion cases. For this time-series, both Smart Fusion and modified Smart Fusion discard nine of 15 windows while RQIFusion$_{\text{HT}}$ does not discard any.

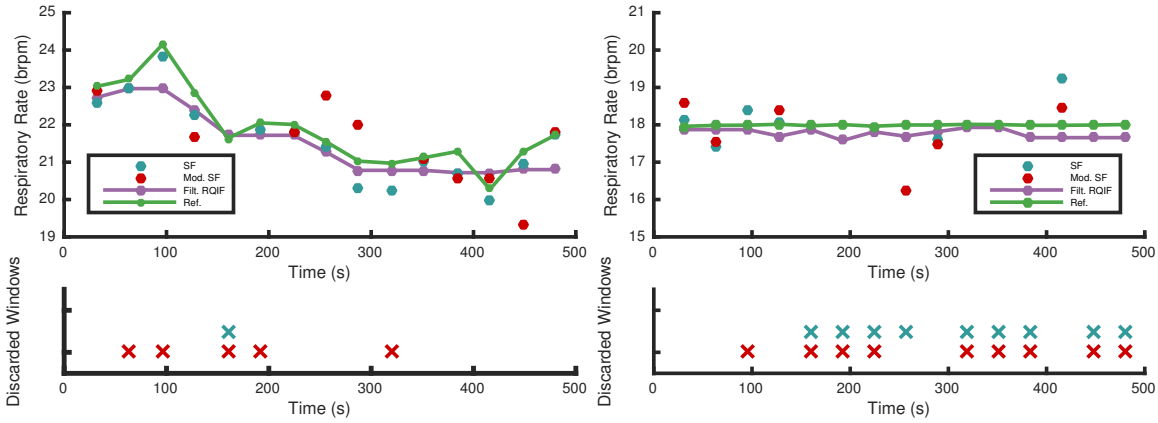


Figure 7.3: *(Upper row) Time-series of RR estimates for Smart Fusion, modified Smart-Fusion, RQIFusion$_{HT}$, and the R_{ref} values for two subjects in data set $\mathcal{D}_{c,2}$: subject 23 (left panel) and subject 103 (right panel). (Lower row) Corresponding plots showing those windows for which each method did not produce an RR estimate.*

The second subject-specific comparison, shown in Figure 7.4, shows the relative frequency of the number of windows discarded per subject in the $\mathcal{D}_c$ (left panel) and $\mathcal{D}_m$ (right panel) data sets. These results show that both RQIFusion$_{\text{LT}}$ and RQIFusion$_{\text{HT}}$

[6] The data from the remaining subjects $\mathcal{D}_{c,2}$ can be seen in Figure A.6 in Appendix A. The waveforms for the subjects in $\mathcal{D}_{c,1}$ and $\mathcal{D}_m$ are not shown for brevity.

discard fewer windows than both Smart Fusion and modified Smart Fusion in both $\mathcal{D}_c$ and $\mathcal{D}_m$. Further, for both RQIFusion$_{\text{LT}}$ and RQIFusion$_{\text{HT}}$, the relative frequency with which more than ten windows discarded is nearly zero. This means that for the majority of patients, it can be expected that RQIFusion$_{\text{LT}}$ or RQIFusion$_{\text{HT}}$ will provide RR estimates for, at a minimum, one-third of all windows. Another trend shown in Figure 7.4 is that as the threshold is increased from RQIFusion$_{\text{LT}}$ to RQIFusion$_{\text{HT}}$, the relative frequency with which larger numbers of windows are discarded increases. This is to be expected given that the higher threshold will discard more modulations and is a further example of the tunable nature of T.

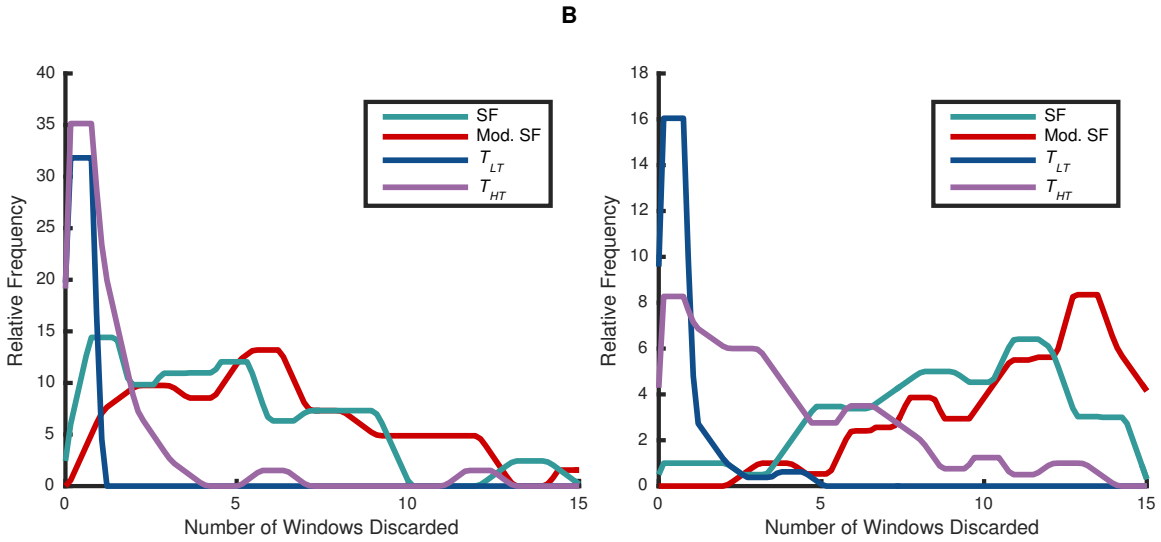


Figure 7.4: *The relative frequency of windows being discarded for individual subjects in $\mathcal{D}_c$ (left panel) and $\mathcal{D}_m$ (right panel), using modified SmartFusion (red), SmartFusion (turquoise), and RQIFusion$_{LT}$ (blue), and RQIFusion$_{HT}$ (purple).*

7.3 RQI Implementation in RR Estimation Work Flow

The previous analysis showed that the performance of RQIFusion is better than the control methods in two distinct data sets ($\mathcal{D}_c$ and $\mathcal{D}_m$) even when considering various

parameters and constraints. However, this analysis still relies on a number of specific features that may bias the performance of RQIFusion. One of these features is that the previous analysis relied on the *ARSpec* algorithm for all RR estimation [164]. Given that the fusion RQI was trained to reduce error between a standard and the *ARSpec* estimation, it is possible that the RQIs are biased to only observe when the *ARSpec* estimation is likely to have a highly accurate estimation as opposed to observing when a respiratory modulation contains clinically-relevant respiratory data. Another feature is that although the data sets used have been collected on subjects that are actually undergoing procedures or are hospitalised, the data do not truly represent a clinically-relevant time-series. The data series per subject in both $\mathcal{D}_c$ and $\mathcal{D}_m$ only represents a time-series of 8 minutes and often this data set was collected early during the monitoring period when data collection was likely most accurate and least noisy.

In order to address these questions about the general applicability of the RQI method in an actual clinical implementation of an algorithm to estimate RR, the present analysis incorporates RQIs into the proposed clinical RR estimation work flow presented in Figure 3.1. Additionally, this analysis is conducted in a previously unused data set, $\mathcal{D}_d$ which was initially described in Section 4.3.

7.3.1 Methods

Data Set

The $\mathcal{D}_d$ data set was extensively described in Section 4.3. This data set represents an ideal data set for the current analysis for four reasons. The first is that a clinically-relevant time-series of data, consisting of two hours of recording followed by a critical event, was collected for each patient. The second relates to the data collection context.

The data was collected during dialysis, a necessary clinical procedure that resulted in a critical event for the subject, thus having a robust assessment of the RR for these subjects would likely represent a clinical benefit. The third is that the previous analyses have relied heavily on the $\mathcal{D}_c$ and $\mathcal{D}_m$ data sets. Achieving similar RR estimation quality using a previously unused data set strengthens the argument for the general applicability of RQIs. Finally, because the collected $\mathcal{D}_d$ waveforms included ECG, PPG, and a respiration belt, all six RR modulation could be obtained along with a hand-scored RR estimation reference. Based on these reasons, the $\mathcal{D}_d$ data set was deemed a good candidate data set for the implementation of RQIs into the RR estimation work flow.

RR Estimation Work Flow

The RR estimation work flow, as shown in Figure 3.1, relies on three distinct signal processing modules and two data-quality checkpoints. Prior to use of the proposed method, pre-processing of the data and a quality-check of the RR estimation reference was implemented. The implementation of each of these steps is described in the following sections.

Pre-processing Prior to implementation of the RR estimation work flow, the data from each of the eight subjects in the $\mathcal{D}_d$ data set were first pre-processed in two steps. The first pre-processing step was to window the data. The same method as was previously described in Section 5.2.1 was used to window the data into 32-second non-overlapping windows. This resulted in 225 unique windows per subject and a total of 1800 windows for the entire $\mathcal{D}_d$ data set. The second pre-processing step was to obtain the RR estimation reference for each window. This was done using the same expert scoring method of the respiration belt data as described for the $\mathcal{D}_m$ data set in Section

5.2.3. This resulted in 83% of all windows in $\mathcal{D}_d$ having RR reference estimations (range among subjects: 64.4% to 96.4%).

Checkpoint I: SQI After pre-processing the data into 32-second windows, the first data checkpoint is the implementation of a signal quality index (SQI) applied to ECG and PPG data. While an existing SQI based on the F1 score was implemented in Section 5.2.3, this SQI is not ideal for this application as the values of the SQI are continuous and no suggested cut-off threshold for discarded windows is available. The present analysis makes use of a modification of the adaptive template matching SQI for ECG and PPG described by Orphanidou *et al.* (2015) [231]. This method relies on both a rule-based and template-matching approach where windows are given a binary score of either "good" quality or "bad" quality. Initially, the peaks in the ECG and PPG are extracted. In this implementation, the ECG peaks are extracted using the same search-back algorithm that was previously implemented in Section 5.2.1 and the PPG peaks are extracted using the same filtering and peak delineator algorithm that was implemented in Section 5.2.1. Following this, the peaks were used to assess three criteria for "good" quality windows:

1. Windows must have a calculated heart rate between 40 and 180 bpm.
2. Windows must not have any gaps between peaks of > 3 seconds.
3. The ratio of the maximum beat-to-beat interval to the minimum beat-to-beat interval must be less than 2.2.[7]

Following the rule-based criteria, median template matching was applied to all remaining "good" candidate windows. Median template matching required finding the median

[7] As described by Orphanidou *et al.* (2015) [231], a value of 2.2 was used as it restricts "good" windows to those where the HR varies by less than 10% within the window and allows for a skipped beat.

beat-to-beat interval in a window and extracting each peak in a window where the size of the extracted peak is the median beat-to-beat interval centred around each peak. The mean of this collection of peaks for a window is then used to obtain the average peak. The Pearson's correlation for each peak and the average peak is obtained and only those windows where the average correlation amongst all of the peaks is greater than a specified threshold are deemed to be "good" windows. The threshold applied was 0.66 based on the results described by Orphanidou *et al.* (2015) [231].

Module I: Extract Respiratory Waveform After the SQI checkpoint, the first module is the extraction of the respiratory waveforms. As in previous analyses, three respiratory waveforms based on AM, FM, and BW are extracted from both the ECG and PPG. This is done using the same peaks detected from the SQI checkpoint and the extraction method and waveform processing methods described in Section 5.2.1. The result is six respiratory modulations, three ECG and three PPG, for each 32-second window all evenly sampled at 4 Hz.

Checkpoint II: RQI With the respiratory waveforms extracted from the previous modulation, the fusion RQI, $Q_{F,1}$ was obtained for each of the six respiratory modulations. The fusion RQI was determined based on the individual RQIs, $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,4}$ based on both the individual and fusion RQI techniques described in Chapters 5 and 6. Additionally, in order to determine "good" and "bad" windows, the value for T, the RQI threshold, was set to be $T = 0.44$, the determined value of T_{HT} from the previous analysis.

Module II: RR Estimation All previous RR estimation has relied on the use of *ARSpec* as an exemplar "state-of-the-art" RR estimation algorithm [164]. However, the

RQI method is a novel pre-processing step prior to RR estimation that we hypothesise has general applicability to alternative RR estimation algorithms. In this analysis, the performance of five state-of-the-art RR estimation algorithms based on both frequency and time-domain techniques with the variable implementation of the SQI and the RQI will be compared. These algorithms are termed: (1) *ARSpec* [164], (2) Fourier transform estimation [66], (3) autocorrelation estimation [145], (4) count-origin estimation [145], and (5) peak-detection and zero-crossing estimation [232].

(1) ARSpec: The *ARSpec* is a frequency-based RR estimation technique which relies on taking the power spectral density (PSD) of the autoregression of a signal. The method finds the PSD for all autoregression model orders from 1 to 20 and finds the median PSD of this panel of 20. The peak of the median PSD is taken to represent the respiratory frequency and ultimately the RR [164].

(2) Fourier Transform Estimation: Many implementations of the Fourier transform have been used to estimate RR. The present analysis makes use of the implementation described by Karlen *et al.* (2013) [66]. In this implementation, the windows are first filtered using a high pass filter to remove the DC component of the signal and then down-sampled to 4 Hz. In the present work flow, these steps were completed during the RR modulation extraction. The RR is extracted by applying a Hamming window followed by taking the Fourier transform of the windows. The RR frequency is determined as the largest frequency peak in the power spectrum within what Karlen *et al.* (2013) deemed the relevant respiratory range (0.067-1.08 Hz or 4-65 breaths/min). This frequency is used to obtain the RR in brpm [66].

(3) Autocorrelation Estimation: The autocorrelation estimation method that was implemented is based on the "ACF-adv" method described by Schäfer and Kratky (2008)

[145]. The method first finds the autocorrelation of the windowed respiratory modulation for all lags $> n/2$. The autocorrelation spectrum is then transformed using the Fourier transform. The power spectrum of the autocorrelation values is used to determine the RR where the median of the spectral power components between 0.1 to 0.5 Hz is used as a lower threshold cut-off and anything above that threshold is expected to be relevant to the RR. Any spectral power between 0.1 to 0.5 Hz and above the median is incorporated into an RR estimate where the power spectral values for each frequency are used to obtain the weighted average of the above median frequencies. The weighted average frequency is used to obtain the RR [145].

(4) Count-Origin Estimation: Unlike the previous three estimation methods, the count-origin estimation method described by Schäfer and Kratky (2008) is a time-domain estimation method consisting of five steps [145]. The first step is to filter the windowed respiratory modulation with a 10th-order bandpass Butterworth filter with cut-off frequencies of 0.1 to 0.5 Hz. The second step is identification of the peaks and troughs of the respiratory modulation. This is done by observing the local maxima and minima. The local maxima are identified as those points in which the value of the previous point $(n-1)$ and the next point $(n+1)$ are less than the value of the current point (n) and the local minima are those points (n) where the value of the previous point $(n-1)$ and the next point $(n+1)$ are greater. The third step is to keep only maxima which have values above $0.2Q_3$ where Q_3 represents the value of the 75th percentile of all detected maxima. This step is important as it discards maxima that are likely not breaths while not allowing extraordinarily large breaths to influence the threshold. The fourth step is to find breath cycles which are defined as the instances where two subsequent local maxima only have a single local minima between them. The fifth step is to use the

average length between all of the breath cycles for a window to determine the RR [145].

(5) Peak Detection & Zero-Crossing Estimation: The peak detection & zero-crossing estimation algorithm is a second time-domain estimation technique based on the work of Fleming (2010) [232]. This method relies on peak and trough detection by finding local minima and maxima using the same approach as the second step of the count-origin estimation method. Following local minima and maxima detection, the true respiratory peaks and troughs are determined through several steps. The first step is to take the average amplitude of the signal and discard any peaks that do not fall above that average and discard any troughs that fall below that average. The second is to eliminate any peaks or troughs that occur within 0.5 seconds of the previous peak or trough (this is physiologically acceptable as a cycle of peaks and troughs every 0.5 seconds would result in an RR of 60 brpm which is physiologically unlikely). The third step is the zero-crossing step (meaning that the signal has to cross its average value) in which maxima or minima that are not preceeded by their opposite (i.e., a maxima proceeded by another maxima, not a minima) are discarded. The average length of time between all peaks in a window is then used to calculate the RR [232].

Module III: RR Fusion The final module in the RR estimation work flow is fusion of the RR estimates for all of the modulations that were retained after the SQI and RQI checkpoints. This estimation was done using the weighted averaging technique described in Equation 7.12 where the value of $T = 0.44$ (T_{HT}). The RR estimation obtained when implementing all described components of the RR estimation work flow will be termed $\text{SQI/RQIFusion}_{\text{HT}}$. In further analysis, $\text{SQI/RQIFusion}_{\text{HT}}$ will be compared to three other implementations of the RR estimation work flow that do not make use of all elements of the work flow. The first is the RR Average method previously described,

which does not use the SQI or RQI checkpoints and which provides an average RR estimate of all six RR modulations. The second is SQI Average which makes use of the SQI checkpoint to discard poor-quality windows and then averages the RR estimates from all remaining modulations. (In instances where no modulations are kept, no RR estimate is provided for the given window.) The third is RQIFusion$_{\text{HT}}$ which is identical to the previously described method in that it uses T_{HT} to discard poor-quality windows and provides a weighted average of the remaining modulation's RR estimates. The SQI checkpoint is not used in this implementation.

Time-Series Analysis

The RR estimation methods described in the previous section will result in a time-series of RR estimates over the two-hour recording period of the subjects in the $\mathcal{D}_d$ data set. These estimates will also be used to estimate the difference in the RR estimates from the SQI/RQIFusion$_{\text{HT}}$ and RR Average methods compared to the reference. Different from absolute error (shown in Equation 5.36), the RR difference for these time-series estimates will be obtained as the difference between the given RR estimate and the reference:

$$\delta_{E,i} = RR_{E,i} - RR_{ref,i} \tag{7.15}$$

where $RR_{ref,i}$ is the reference RR estimate for the i^{th} window, $RR_{E,i}$ is the RR estimate for the given estimation method, E, either SQI/RQIFusion$_{\text{HT}}$ or RR average, and $\delta_{E,i}$ is the difference between the estimation method and the reference.[8]

Additionally, the subjects in the $\mathcal{D}_d$ data set all had critical events at the end of the

[8]Throughout this chapter, references to RR error will refer to the absolute error, defined here as the absolute difference between the reference RR for a window and the RR estimate from either SQI/RQIFusion$_{\text{HT}}$ or the RR average for that window. References to RR difference will refer to the difference between an RR estimate and the reference for specific windows as shown in Equation 7.15.

two-hour monitoring session. Given this, studying not just the individual windows but the time-series estimates in these subjects offers key insights into how RR estimation from ECG and PPG can be used for patient monitoring.

One shortcoming of the time-series estimates from these various implementations (except RR Average) is that the RR estimate is not guaranteed to update with a new estimate in every window. This is because the SQI and RQI checkpoints might deem all modulations in a window to be artefactual and thus no updated estimate will be available. Therefore, in order to obtain a smooth RR waveform, both median-filtering and cubic spline interpolation will be used to smooth the RR estimate waveforms and cubic spline interpolation will be used to smooth the RR differences. The median-filtering and cubic spline techniques are described in Sections 7.1.1 and 7.1.2 respectively.

Three point median-filtering was applied to the RR waveforms where $N_L = 2$ and $N_R = 0$ in order to simulate a real-time application. Thus two data points will be appended to the beginning of the signal, as described in Equation 7.1.1, and the filter will provide a value at the first time point of the signal. Additionally, in instances where estimates do not exist for a time point, the median filter value will be based on the values that do exist and in instances where no estimates exist within the median filter range, no estimate will be given (this will constitute a gap in monitoring).

Cubic spline interpolation will also be applied to the RR estimates. Unlike median-filtering, cubic spline interpolation relies on knowing both end points and is therefore not a real-time smoothing mechanism. In order to solve the cubic spline function, the end conditions must be specified. To ensure a smooth function at the end points, the not-a-knot end conditions are used (Equations 7.11 and 7.12). Finally, the cubic spline interpolation will only interpolate between points that have values, which means that

the length of the splines will not be uniform.

The time-series RR estimates obtained in $\mathcal{D}_d$ from the smoothing functions can be used to observe if and when subjects became physiologically deranged. The current clinical standard for detecting patient derangement is the use of early warning scores (EWS). The most commonly-used EWS system in the UK is the National Early Warning Score 2 (NEWS2) which is based on calculating a risk score for patients based on their observed vital signs [233]. The NEWS2 score provides both an individual vital-sign risk score and an aggregate risk score. While the NEWS2 scores are most often determined after manual observation of the vital signs, they can also be applied to data acquired from continuous monitors. The risk scores associated with RR, HR, BP, temperature, and SpO_2 for NEWS2 can be seen in Table 7.4 [233]. For NEWS2, any individual vital sign score over 3 indicates a low to medium clinical risk. An aggregate score for all NEWS2 vital signs (including RR, HR, SBP, Temp., SpO_2, consciousness, and whether the patient is using supplemental oxygen) that lies between 5-6 is deemed to be a medium risk, and a score over 6 is a high risk.

Table 7.4: *NEWS2 Scores for specified physiological parameters [233].*

	Score						
Physiologic Parameter	**3**	**2**	**1**	**0**	**1**	**2**	**3**
RR (brpm)	≤ 8		9 - 11	12 - 20		21 - 24	≥ 25
HR (bpm)	≤ 40		41 - 50	51 - 90	91 - 110	111 - 130	≥ 131
SBP (mmHg)	≤ 90	91 - 100	101 - 110	111 - 219			≥ 220
Temp. (°C)	≤ 35.0		35.1 - 36.0	36.1 - 38.0	38.1 - 39.0	≥ 39.1	
SpO_2 (%)	≤ 91	92 - 93	94 - 95	≥ 96			

Furthermore, because the $\mathcal{D}_d$ data set contains both an RR reference and the RR estimates from the various work flow implementations, the NEWS2 score can be used to obtain "episodes of derangement" which can be compared with the RR reference and the

comparator methods: SQI/RQIFusion$_{\text{HT}}$ and RR Average. Comparison of the correlation of the "episodes of derangement" between the RR reference and SQI/RQIFusion$_{\text{HT}}$ or RR average can be used to observe if the SQI and RQI checkpoints are better able to identify to actual "episodes of derangement." The ECG data in the $\mathcal{D}_d$ data set can also be used to estimate HR. This will be done using using the time difference between ECG peaks detected in Section 7.3.1 to determine the average HR in each time window. This can be used to observe "episodes of derangement" in the HR which can be compared to the "episodes of derangement" in the RR waveform and used to assess which vital sign is more effective for determining derangement.

True RR "episodes of derangement" were identified in the median-filtered reference RR waveform and defined as distinct low RR episodes (with RR falling below 8 brpm) and high RR episodes (with RR rising above 20 brpm). According to Table 7.4, these were all events where the NEWS2 score for RR was non-zero. Due to the nature of the three-point median filter, at least two RR values within the three point window (1 minute and 36 seconds in length) had to be below the low threshold (for a low RR episode) or above the high threshold (for a high RR episode). The result of using the median filter to identify the start of the events was an increased likelihood that a low or high RR event was actually a physiologic event. A low or high RR event was defined as persisting until two subsequent median filtered values were above the low threshold (for low RR events) or above the high threshold (for high RR events). The importance of the "two subsequent point" requirement was to ensure that any of the RR estimates that factored into a previous below- or above-threshold event could not contribute to a new event (i.e., if the third subsequent median filtered point after an event was above or below the threshold, none of the points in that event would have contributed to the

estimation in the previous event).

Given this as the definition of an episode of derangement in the reference, all of the true positive (TP), false positive (FP) and false negative (FN) events in the SQI/RQIFusion$_{\mathrm{HT}}$ and RR Average estimation techniques were identified. A TP event was defined as an instance where at least one median-filtered RR estimate occurred within ± 2 data points of an episode of derangement as determined in the reference waveform. For a TP event, the ± 2 data point window was appended on both ends of a true episode of derangement to allow for a small variability (1 minute and 4 seconds) in when the episode of derangement began and ended in the RR estimation work flows and the reference. A FN event was defined as an instance where there was an episode of derangement (± the 2-point window) in which an RR estimation work flow did not have a median filtered value below the low threshold or above the high threshold. A FP event was defined as an instance where an RR estimation work flow had a median filtered data point below the low threshold or above the high threshold that was outside of an episode of derangement (± the 2-point window). Further, all instances of TPs had to agree with whether it was a low RR or high RR event. In the event where there was disagreement between the reference and the RR estimation, it was defined as a FP event. For example, if the reference had a high RR episode of derangement and the RR estimation had an RR estimate below the low threshold during the period of that episode, that point was deemed to be a FP, not a TP; if during that episode there were no other values in the RR estimate that were above the high threshold, the episode was deemed to be a FN.

The definitions of TPs, FPs, and FNs were used to calculate the precision, recall, and F1 score for the SQI/RQIFusion$_{\mathrm{HT}}$ and RR average techniques compared to the reference

RR. The precision, recall, and F1 scores have previously been defined in equations 5.37, 5.38, and 5.39, respectively.

7.3.2 Results

Data Acquisition and Reference Estimation

Table 7.5 shows the waveform summary statistics for the RR reference for the eight selected subjects of the $\mathcal{D}_d$ data set. The statistics show the average RQI ($Q_{F,1}$) of the reference RR waveform amongst the 225 windows for each subject, the percentage of windows where a reference value was obtained based on the criteria in Section 5.2.3 and the average RR ($\pm$ SD) for those windows. The average percentage of windows kept amongst all of the subjects was 83.0% with the largest percentage of windows with an RR estimate seen in subject 1 (96.4% of windows) and the smallest seen in subject 6 (64.4% of windows). Additionally, the correlation between the RQI and the percentage of windows kept amongst all subjects was $p = 0.692$. This indicates good performance of the RQI on the reference data as it demonstrates that the RQI is able to assess which subjects have the best data as indicated by having a larger proportion of their windows with reference RR values. Finally, the average RR in all individual subjects was between 12 and 20 indicating that according to NEWS2 scoring, all subjects were in normal respiration for the majority of the recordings.

RQIFusion Analysis in $\mathcal{D}_d$ Data Set

The RQIFusion method was applied to the $\mathcal{D}_d$ data set using five values for T centred around the original value of T_{HT} in increments of 0.10 and the *ARSpec* RR estimation implementation. The selected values were $T = 0.64$, 0.54, 0.44, 0.34, and 0.24. The

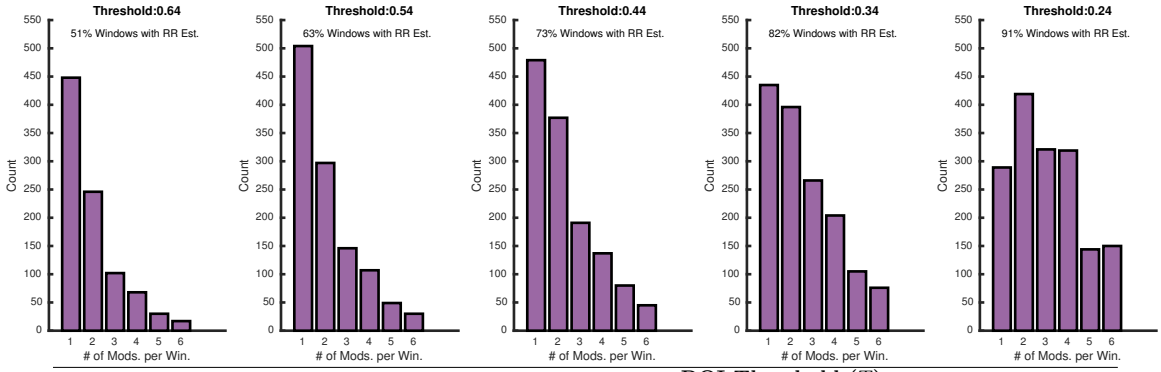
Table 7.5: *Summary statistics, including RQI value ($Q_{F,1}$) percentage of windows with estimates, and the average RR (± SD), for RR reference in $\mathcal{D}_d$ subjects.*

Subject	Avg. RQI	% Win.	Avg. RR (± SD)
1	0.84	96.4%	18.91 ± 2.02
2	0.75	94.2%	15.40 ± 2.24
3	0.60	84.9%	16.70 ± 1.67
4	0.57	77.3%	17.84 ± 3.65
5	0.50	85.3%	15.89 ± 2.61
6	0.49	64.4%	11.02 ± 2.39
7	0.46	73.8%	17.00 ± 2.52
8	0.44	87.6%	15.25 ± 1.87
Avg.	**0.58**	**83.0%**	**16.15 ± 3.18**

results from the application of the varying values for T for the entire $\mathcal{D}_d$ data set considered together can be seen in Figure 7.5. These results show that the total number of windows with an RR estimate (these are windows in which at least one RR modulation had an RQI value greater than or equal to the given threshold) is nearly perfectly negatively correlated with the threshold value ($p = -0.998$), thus as the threshold decreases, the number of windows with an estimate increases. Additionally, the RR MAE increases as the threshold decreases. This further demonstrates the trend in a new data set that as the threshold is decreased (allows more modulations to pass) the number of windows with RR estimates increases but so does the error. This trend can be further examined using the histograms in Figure 7.5 which show the counts of the number of modulations that went into an RR estimate for RQIFusion (as any one window with an estimate can be composed of the estimates from between one and six modulations). These histograms show that not only does the number of windows with estimates increase as the value of T is decreased but, as should be expected, so do the number of modulations that make up those estimates. The fact that more likely-erroneous modulations go into RR estimates for specific windows as T decreases is likely the primary factor that causes the

RR error to increase at lower values of T.

Figure 7.5 also shows the frequency for each value of T for which certain RR modulations were above that value. This shows that at all thresholds, the ECG baseline wander modulation is the most likely modulation to be above the threshold. The PPG frequency modulation and ECG frequency modulation are most likely to be discarded. This is consistent with prior knowledge. It is well-understood that the quality of frequency modulation is heavily affected by exogenous conditions [46]. Importantly, these results also show that neither ECG nor PPG is the more favoured method for obtaining usable RR modulations. This is an important finding as it demonstrates that both ECG and PPG have value in improving RR estimation and that if one of the methods is not available, the other can still provide an adequate estimation.



	RQI Threshold (T)				
	0.64	**0.54**	**0.44**	**0.34**	**0.24**
PPG Amp. Mod.	390	488	606	752	917
PPG Freq. Mod.	165	251	316	418	582
PPG Baseline Wan.	317	418	537	658	837
ECG Amp. Mod.	245	359	486	636	862
ECG Freq. Mod.	181	250	314	419	606
ECG Baseline Wan.	472	623	765	939	1182
Tot. Wins. With Est.	911	1133	1309	1482	1642
# Wins. Dropped (No Reference)	113	158	190	232	262
RR MAE (± SD)	3.27 ± 4.36	3.53 ± 4.38	3.65 ± 4.27	3.74 ± 4.19	3.83 ± 4.03

Figure 7.5: *Performance of RQIFusion using five exemplar values for T centred around T_{HT} using the ARSpec RR estimation method for the $\mathcal{D}_d$ data set. Figure (top) shows the percentage of windows with RR estimates and the number of modulations that make up those estimates for each value of T. Table (bottom) shows the number of times each modulation was above the specified threshold and the total number of windows dropped and the resulting RR MAE (± SD) from all of the windows kept.*

This phenomenon is further explored in Table 7.6, which shows the number and type

of modulations kept in each individual subject when T is set to the value of the threshold for T_{HT}. This clearly demonstrates that in some subjects (subject 1) ECG is preferred to PPG, in others (subject 3) PPG is preferred to ECG, and in others (subject 8), both are preferred. This further demonstrates the value of both ECG and PPG in estimating RR. It also shows that cases where one method is preferred to the other are often patient specific. One of the reasons for the improved performance of one method over the other may be lead/probe placement.

Table 7.6: *Number and type of modulations kept for each individual subject in the $\mathcal{D}_d$ data set when the threshold is set to T_{HT}.*

Subject	**ECG**			**PPG**		
	AM	**FM**	**BW**	**AM**	**FM**	**BW**
1	115	79	129	51	59	56
2	15	3	19	10	16	175
3	1	4	8	63	0	78
4	65	32	65	62	50	76
5	17	47	20	20	13	39
6	144	5	73	113	31	102
7	87	25	62	37	33	83
8	162	121	161	130	112	156
Total	**606**	**316**	**537**	**486**	**314**	**765**

RQI/SQI Work Flow Implementation

The SQI/RQI work flow as described in Section 7.3.1 is implemented in six steps:

1. Pre-processing
2. SQI Checkpoint
3. Respiratory Modulation Extraction
4. RQI Checkpoint
5. RR Estimation
6. RR Fusion

Table 7.7 shows the four different variations on the full implementation: RR Average, SQI Average, $\text{RQIFusion}_{\text{HT}}$, and $\text{SQI/RQIFusion}_{\text{HT}}$ for five different state-of-the-art RR estimation methods: *ARSpec*, FFT estimation, autocorrelation count-orig, and combined peak-zero estimation.

The results in Table 7.7 show the percentage of windows with RR estimates and the RR MAE ($\pm$ SD) for each work flow implementation and RR estimation method for the $\mathcal{D}_d$ data set in aggregate. The results show that in almost all instances SQI Average and $\text{RQIFusion}_{\text{HT}}$ reduce the MAE compared to RR Average while providing RR estimates for $\sim$75% of the windows. (The exceptions are the SQI Average with count-origin estimation and $\text{RQIFusion}_{\text{HT}}$ with the FFT estimation.) Even more critical is that the $\text{SQI/RQIFusion}_{\text{HT}}$ decreases the MAE, in all cases, compared to the other three implementations, for all RR estimation methods with the MAE ranging from 2.18 ± 3.33 to 3.46 ± 4.02. In addition to this, the percentage of windows discarded using $\text{SQI/RQIFusion}_{\text{HT}}$ compared to either the SQI or RQI alone is nearly doubled. This is important as it shows that the SQI and RQI are functioning independently of each other and are being used to discard different sets of data windows. This result demonstrates the necessity of both quality indices for the most accurate and robust RR estimation.

The results of the state-of-the-art RR estimation algorithms in Table 7.7 show that the autocorrelation estimate achieves the lowest error and the count-origin estimation method has the highest error, but all of the RR estimation methods provide similar results. This provides evidence that the RQIFusion method is robust across estimation techniques and is not reliant on the use of *ARSpec* for accurate RR estimation. Based on these results, it was determined that the combined peak-zero method would be implemented for further analysis. This was chosen because the difference between

the MAE for the count-origin method is only slightly higher than the autocorrelation estimation (2.35 vs. 2.18, respectively) but that the standard deviation of the combined peak-zero method is substantially smaller than all of the other methods meaning that the population of estimates is more tightly clustered to the reference than other methods.

Table 7.7: *Results from implementation of RR Average, SQI Average, RQIFusion$_{HT}$, and SQI/RQIFusion$_{HT}$ work flows using five state-of-the-art RR estimation methods: ARSpec, FFT estimation, autocorrelation estimation, count-orig estimation, and combined peak-zero estimation. Results for each RR estimation work flow are for the entire $\mathcal{D}_d$ data set and show the percentage of windows with RR estimates and the RR MAE (± SD) compared to the reference RR.*

Est. Method	No SQI/RQI		SQI		RQI		SQI/RQI	
	% Win	RR MAE	% Win	RR MAE	% Win	RR MAE	% Win	RR MAE
ARSpec	100.0	3.86 ± 3.20	76.6	3.19 ± 2.90	74.6	3.66 ± 4.29	50.7	2.21 ± 3.20
FFT	100.0	3.59 ± 3.12	76.6	2.95 ± 2.75	74.6	3.67 ± 4.35	50.7	2.25 ± 3.24
Autocorrelation	100.0	3.76 ± 3.26	76.6	3.06 ± 2.91	74.6	3.66 ± 4.46	50.7	2.18 ± 3.33
Count-Orig	100.0	5.75 ± 4.77	76.6	6.34 ± 5.08	69.7	3.52 ± 3.67	47.3	3.46 ± 4.02
Combined Peak-Zero	100.0	3.89 ± 2.88	76.6	3.84 ± 2.97	74.6	3.05 ± 3.01	50.7	2.35 ± 2.48

The results for the combined peak-zero method for all four work flow implementations in the individual subjects in the $\mathcal{D}_d$ data set can be seen in Table 7.8. This shows that in six of eight cases SQI/RQIFusion$_{\text{HT}}$ improved RR estimation compared to RR Average. The decrease in the MAE in these cases ranges from 3.36 brpm to 0.03 brpm. In the cases where SQI/RQIFusion$_{\text{HT}}$ does not improve compared to RR Average (subjects 1 and 4), the increase in the MAE is minimal (0.04 brpm and 0.35 brpm, respectively).

The results in Table 7.8 also show that in some instances where the SQI and RQI improve RR estimation, the RQI is the critical checkpoint in improving estimation (subjects 2, 3, 6, and 7), but in other instances, both the SQI and RQI appear to be critical in tandem (subject 5). This observation supports the importance of both the SQI and RQI as quality checkpoints. Another effect observed in these results is that the number of windows kept per subject is highly variable. This is a good indication that the SQI

and RQI are considering the underlying quality of the waveform for the given subject as many of the factors that affect signal and waveform quality would be expected to be present for the entirety of the recording for a subject as opposed to being present only sporadically during a recording.

Table 7.8: *Results of implementation of RR Average, SQI Average, RQIFusion$_{HT}$, and SQI/RQIFusion$_{HT}$ using the combined peak-zero estimation method for each individual subject in the $\mathcal{D}_d$ data set. Results show the percentage of windows with RR estimates and the RR MAE (± SD) compared the RR reference.*

Subject	No SQI/RQI		SQI		RQI		SQI/RQI	
	% Win	RR MAE	% Win	RR MAE	% Win	RR MAE	% Win	RR MAE
1	100.0	3.54 ± 2.41	35.5	2.95 ± 2.45	89.9	4.73 ± 3.45	19.4	3.58 ± 3.46
2	100.0	5.74 ± 2.55	97.6	5.29 ± 2.55	81.1	2.06 ± 1.99	78.8	2.01 ± 1.98
3	100.0	4.73 ± 2.30	82.2	4.53 ± 2.58	45.0	2.15 ± 1.84	32.5	2.19 ± 1.84
4	100.0	4.61 ± 2.61	66.7	3.75 ± 2.46	77.6	6.14 ± 3.67	35.6	4.96 ± 3.52
5	100.0	2.26 ± 1.75	42.2	2.90 ± 2.01	50.0	2.55 ± 2.15	8.9	1.86 ± 1.33
6	100.0	7.11 ± 3.08	100.0	7.10 ± 3.20	84.1	3.68 ± 2.28	82.1	3.75 ± 2.35
7	100.0	2.68 ± 1.97	99.4	2.81 ± 2.08	74.7	1.71 ± 2.35	64.5	1.69 ± 2.33
8	100.0	1.08 ± 0.99	100.0	1.03 ± 0.93	93.9	1.10 ± 1.11	91.9	1.05 ± 1.07

Finally, two cases (subject 6 and subject 8) show two important features in the SQI/RQIFusion$_{\mathrm{HT}}$ implementation when considering the worst and the best input waveforms. Subject 6 represents a case with a highly error-prone waveform. The MAE in RR Average of 7.11 brpm is not clinically usable. However, with the implementation of SQI/RQIFusion$_{\mathrm{HT}}$, the MAE is decreased by nearly twofold to 3.75 brpm. While this error is still sub-optimally accurate, the average is within the range of a clinically relevant RR error (approximately 4 brpm, as a human might measure) and represents an improvement on the expected error from manually assessed RRs [21]. Thus the implementation of the SQI and RQI checkpoints improves the estimations for subject 6 such that the estimates change from being clinically irrelevant to clinically useful. Conversely, subject 8 represents a case where the waveforms are highly accurate. When neither the

SQI nor RQI is implemented and 100% of all modulations and windows are kept, the MAE is 1.08 brpm. In this case, implementation of the SQI and RQI slightly improves the MAE from 1.08 brpm to 1.05 brpm, but more critically, it only discards 8.1% of windows. This demonstrates that in the best cases when the SQI and RQI are not necessary, the SQI and RQI do not needlessly discard data and maintain the fidelity of the RR estimate. It can be noted that subject 5 might be the counterpoint to this argument where the MAE in RR Average is 2.26; however, with SQI/RQIFusion$_{\mathrm{HT}}$, the MAE drops to 1.86, but where 91.1% of the windows are discarded. This may represent a case where the baseline RQI threshold value (T_{HT}) is unsuitable and that a higher threshold will result in slightly higher errors but will retain more windows. However, this approach is application-specific, as even with only 8.9% of windows providing RR estimates, the RR will be updated on average every 12 minutes, which is typically suitable for all but the most critical clinical applications.

Time-Series Analysis

The previous section presented the results of the RR estimation work flow when considering both $\mathcal{D}_d$ and each subject in aggregate. The results in the present section consider the *individual* time-series estimates for each subject in $\mathcal{D}_d$. These results are the most relevant to clinical applications, because episodes of clinical derangement represent individual isolated events as opposed to aggregate measures. Thus, it is important that the RR estimation work flow presented is both accurate in aggregate and on a window-by-window basis.

Figures 7.6 and 7.7 show the RR difference and cubic spline interpolation between these points for both RQIFusion$_{\mathrm{HT}}$ and RR Average for every subject in the $\mathcal{D}_d$ data

set. These results in general show that while the RQIFusion$_{\text{HT}}$ has less estimates than RR Average, the estimates are closer to the reference RR.

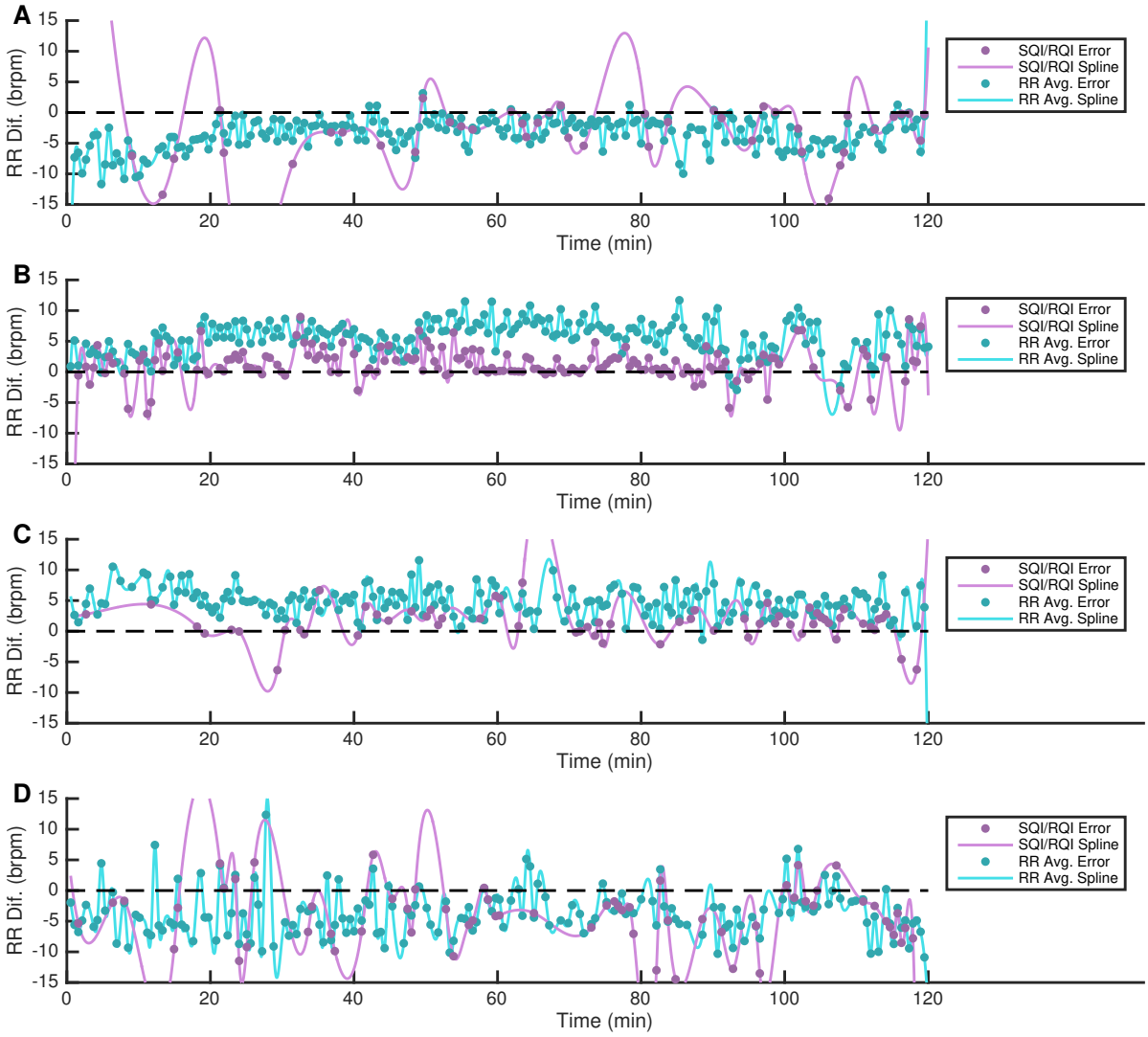


Figure 7.6: *120-minute time-series of RR difference compared to reference for the SQI/RQIFusion$_{HT}$ method (purple points) and RR average method (blue points). Cubic spline interpolation of errors is shown for SQI/RQIFusion$_{HT}$ (light purple line) and RR average (teal line). Results shown for four subjects in the $\mathcal{D}_d$ data set, (A) subject 1, (B) subject 2, (C) subject 3, (D) subject 4.*

Specifically, these results show three important trends. The first trend is that for subjects 1, 3, 4, and 5 (Figures 7.6A, 7.6C, 7.6D, and 7.7A) where SQI/RQIFusion$_{\text{HT}}$ discarded greater than 50% of the windows, the remaining windows with estimates are distributed across the entire 120-minute monitoring period. This means that even for

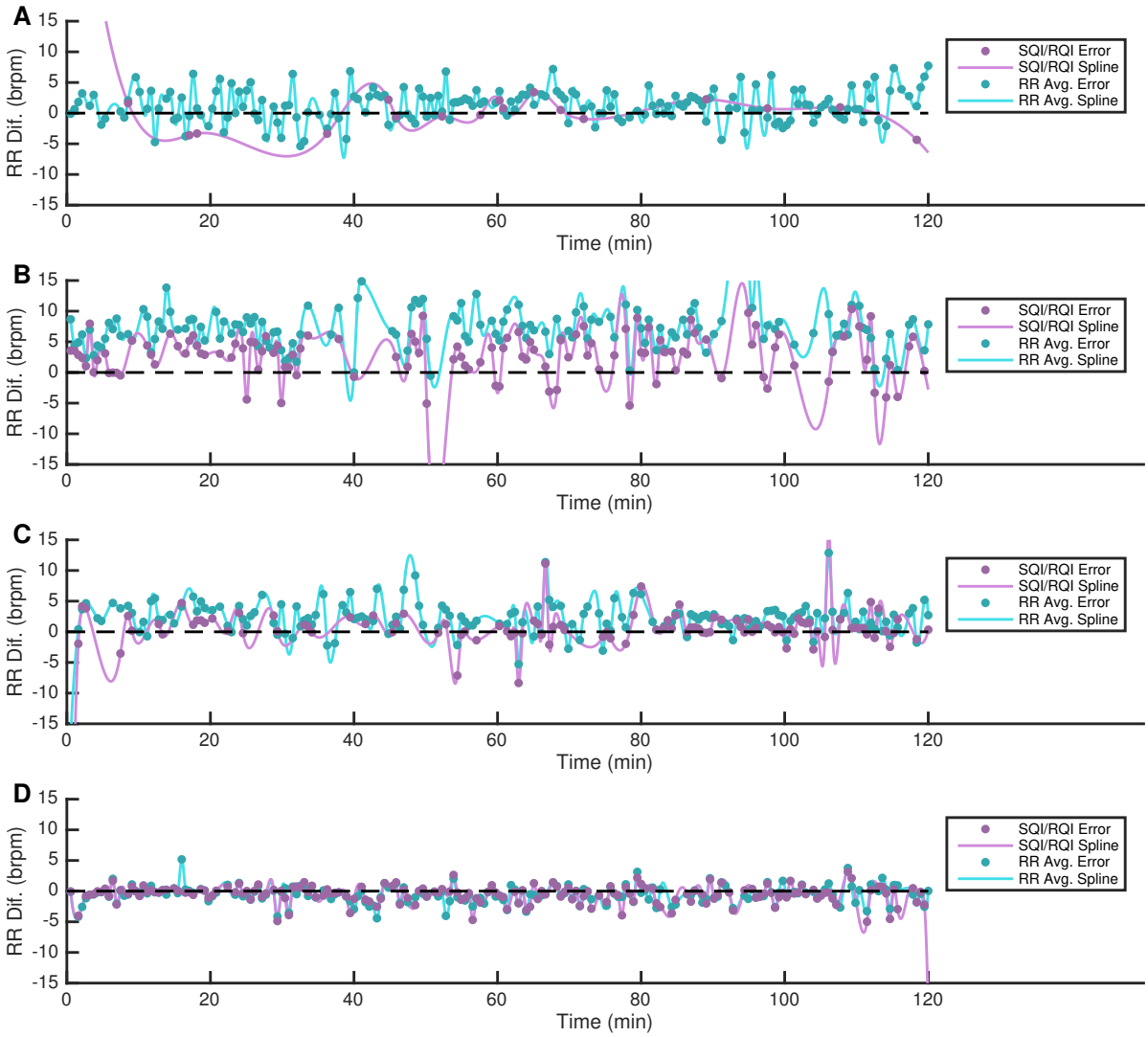


Figure 7.7: *120-minute time-series of RR difference compared to reference for the SQI/RQIFusion$_{HT}$ method (purple points) and RR average method (blue points). Cubic spline interpolation of errors is shown for SQI/RQIFusion$_{HT}$ (light purple line) and RR average (teal line). Results shown for four subjects in the $\mathcal{D}_d$ data set, (A) subject 5, (B) subject 6, (C) subject 7, (D) subject 8.*

these subjects, obtaining a continuously updating RR estimate is possible. The second trend is that for all subjects, SQI/RQIFusion$_{\text{HT}}$ both overestimates and underestimates RR in at least some instances compared to the reference. This is an important finding because it shows that SQI/RQIFusion$_{\text{HT}}$ does not have a systematic positive or negative estimation bias. The third finding is that the RR difference compared to the reference for SQI/RQIFusion$_{\text{HT}}$ is rarely greater than the RR difference compared to the reference

for RR Average. This is an important finding because results from Table 7.8 show that in the majority of cases the MAE for SQI/RQIFusion$_{\mathrm{HT}}$ is less than RR Average. The results in Figures 7.6 and 7.7 further support this and show that, nearly uniformly for individual windows, SQI/RQIFusion$_{\mathrm{HT}}$ improves the RR estimation.

Figures 7.8, 7.9, 7.10, and 7.11 show the three-point median filter smoothing and cubic spline interpolation of the RR estimates for the RR reference, RQIFusion$_{\mathrm{HT}}$, and RR Average, in addition to the HR for subject 2, 3, 8, and 1, respectively. Additionally, the specific modulations that were discarded in each time window by the RR work flow with the SQI and RQI are shown, labelled by which checkpoint threshold they fell below. Episodes of derangement (including the $\pm$ 2 point window) as determined in the RR reference are shown on all RR waveforms. The specific median filtered data points that fall below the low threshold or above the high threshold are specifically annotated. Similar results for the remaining subjects in $\mathcal{D}_d$ data set are shown in Figures A.7 to A.10 in Appendix A.

Subjects 1, 2, 3, and 8 were specifically chosen because they demonstrate four different categories of performance of RQIFusion$_{\mathrm{HT}}$ in the $\mathcal{D}_d$ data set. The first category are instances where RQIFusion$_{\mathrm{HT}}$ resulted in the most accurate RR estimates with a small percentage of windows discarded. This is is the optimal use case for RQIFusion$_{\mathrm{HT}}$ and is exemplified by subject 2 (Figure 7.8), but also subjects 6 (Figure A.9) and 7 (Figure A.10). Figure 7.8 shows a number of important features. The first is that the SQI/RQIFusion$_{\mathrm{HT}}$ implementation more accurately assesses the RR compared to the RR average implementation. It can be observed in Figure 7.8 that the RR Average implementation consistently overestimates the actual RR (third panel compared to fourth panel) while the SQI/RQIFusion$_{\mathrm{HT}}$ implementation more accurately follows the trends

seen in the RR reference (first panel compared to fourth panel). The systematic overestimation observed in RR Average for subject 2 can be attributed to the PPG frequency modulation which, for this subject, drastically overestimates the RR for many windows but has a consistently low RQI value. The SQI/RQIFusion$_{\mathrm{HT}}$ implementation removes this poor quality modulation from consideration. Furthermore, the SQI/RQIFusion$_{\mathrm{HT}}$ implementation, while not precisely tracking the NEWS2 events observed in the reference, is far more accurate and has far fewer false alarms compared to the RR average implementation (this trend is quantitatively assessed for all subjects in Table 7.9). Figure 7.8 also demonstrates the value of both the SQI and RQI checkpoints. It shows that modulations are discarded due to deficiencies in either and both checkpoints (second panel). Finally, Figure 7.8 shows that the RR estimate is a far more valuable estimator of derangement than the HR estimate. In subject 2, the NEWS2 score for HR was 1 during the entire period of monitoring while the NEWS2 score for the RR estimate was only above 0 at the start of monitoring and 27 minutes before the episode of derangement. This finding is consistent with previous research which has shown that RR is one of the most useful vital signs when monitoring for episodes of derangement [4].

The second category of SQI/RQIFusion$_{\mathrm{HT}}$ performance are those cases where the method improved RR estimation compared to RR Average, but a large percentage of windows were discarded. These cases demonstrate the usefulness of SQI/RQIFusion$_{\mathrm{HT}}$ when the RR waveforms are of poor quality. They also demonstrate the balance that must be struck between data retention and estimation accuracy resulting from the selection of the RQI threshold. This is exemplified by subject 3 (Figure 7.9) and subject 5 (Figure A.8). Similar to subject 2, the RR Average for subject 3 systematically overestimates the RR. In this case, this is due to cases of overestimation by all three PPG

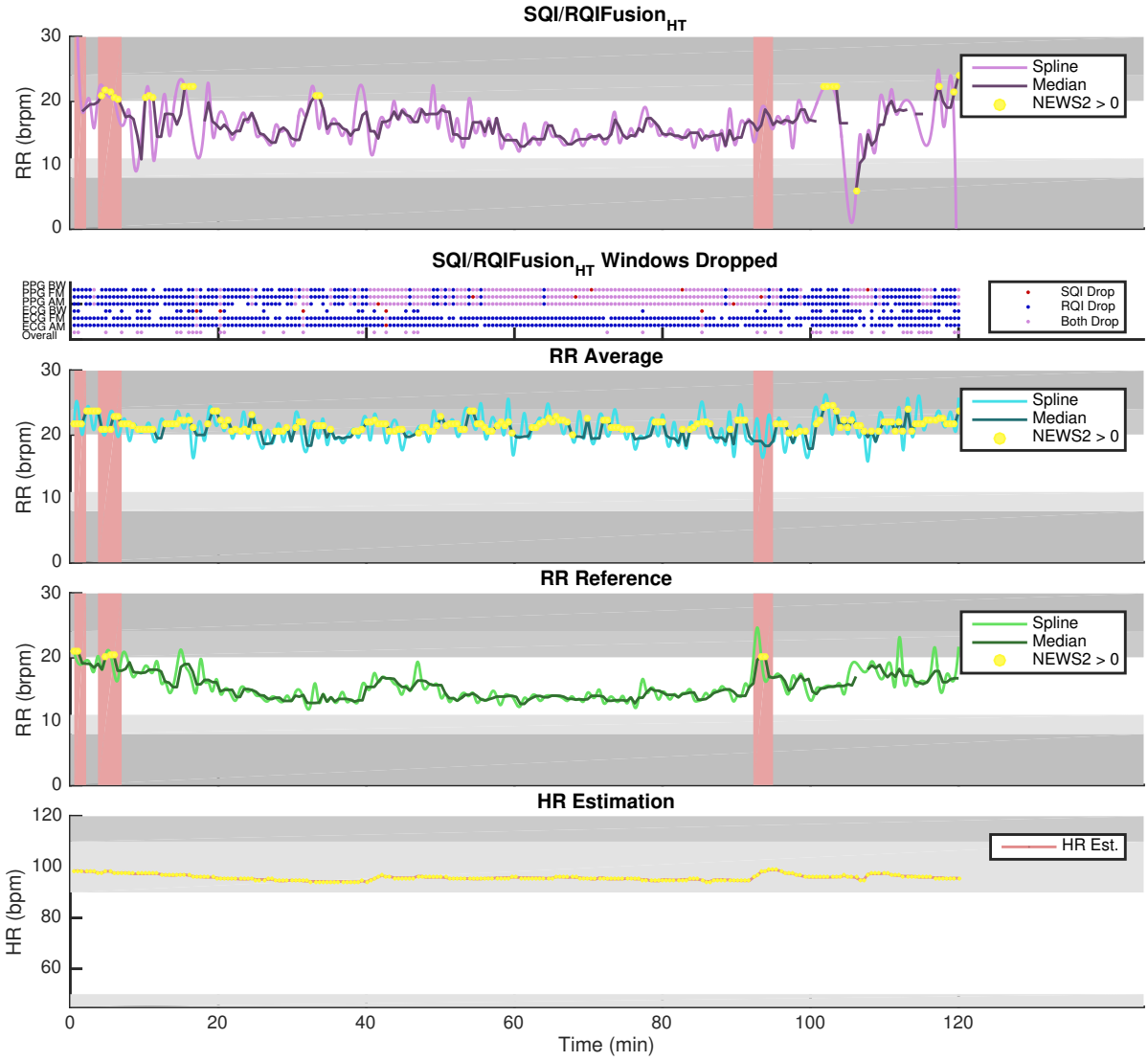


Figure 7.8: *Time-series RR and HR waveforms for two hours of monitoring of subject 2 in the $\mathcal{D}_d$ data set. Upper panel is the RR waveform from SQI/RQIFusion$_{HT}$. Second panel (from top) shows the the specific modulations (PPG BW, PPG FM, PPG AM, ECG BW, ECG FM, and ECG AM) that are below the discard threshold for the SQI (red dots), RQI (blue dots), or both the SQI and RQI (purple dots) for each window. Bottom row of panel shows if a window was dropped (all six modulations dropped). Third panel (from top) is the RR waveform from the RR Average implementation. The fourth panel (from top) is the RR reference waveform. The bottom panel is the HR estimation obtained from ECG. In all RR waveform panels the lightest coloured waveforms (light purple, teal, and green) represent the cubic spline interpolations of the RR estimates, the darker coloured waveforms (purple, teal, and green) represent the three-point median filter of the RR estimates. The darkest coloured points (dark purple, teal, and green) represent the RR estimates obtained from median filtering. In the HR waveform the red line represents the HR estimates. The darkest grey bars, grey bars, and lightest grey bars represent RR or HR regions (depending on waveform) where the NEWS2 score is 3, 2, or 1, respectively. NEWS2 scores of 0 on the waveform plots are represented by white regions. The rose regions on all of RR waveforms represent critical events as detected by the RR reference as defined in Section 7.3.1. Yellow points in RR and HR waveforms represent all instances where the given waveform had an RR or HR estimate with a NEWS2 score exceeding 0.*

modulations and the ECG frequency modulation. All of these modulations, on average, have low RQIs and are almost entirely discarded by SQI/RQIFusion$_{\text{HT}}$. The estimates obtained for this subject from the PPG were poor which may be because of poor place-

ment of the PPG probe. Due to the poor quality of the PPG and ECG frequency modulation, the majority of estimates for this subject were derived from the ECG amplitude modulation and baseline wander alone. It is not surprising that many windows were discarded as during any instance when both of these modulations was not reliable, an estimate could not be obtained. The major result exemplified by subject 3 is that even in instances of poor data quality, reliable RR estimates can be obtained during the entire 120-minute monitoring period. While the SQI/RQIFusion$_{\mathrm{HT}}$ estimations still recorded more events with NEWS2 scores above 0 than the reference, it did not systematically overestimate the RR, as did RR Average. It is possible that further threshold tuning, while further reducing the number of windows with estimates, might improve overall monitoring accuracy.

The third category of RQIFusion$_{\mathrm{HT}}$ performance are cases where the method produced minimal improvements in the accuracy of the RR estimates but had little data loss. The exemplar case for this category is subject 8 (Figure 7.10). In this case, RQIFusion$_{\mathrm{HT}}$ offers little benefit compared to RR Average. However, because there is very little data lost (8.1% of windows are discarded), implementation of RQIFusion$_{\mathrm{HT}}$ does not degrade the continuity of the RR estimates. This shows the performance of RQIFusion$_{\mathrm{HT}}$ when all of the inputs are of high-quality. As expected, much of the data is retained and the highly-accurate RR estimates are maintained. This category of performance demonstrates that when the data is already of high-quality, RQIFusion$_{\mathrm{HT}}$ cannot further improve it but neither is implementation of RQIFusion$_{\mathrm{HT}}$ a hindrance. The high-quality of the ECG and PPG data used to obtain the estimates shown in Figure 7.10 is further demonstrated from examining the cause for windows to be dropped for subject 8. For subject 8, it was far more likely that windows were dropped due to a low RQI

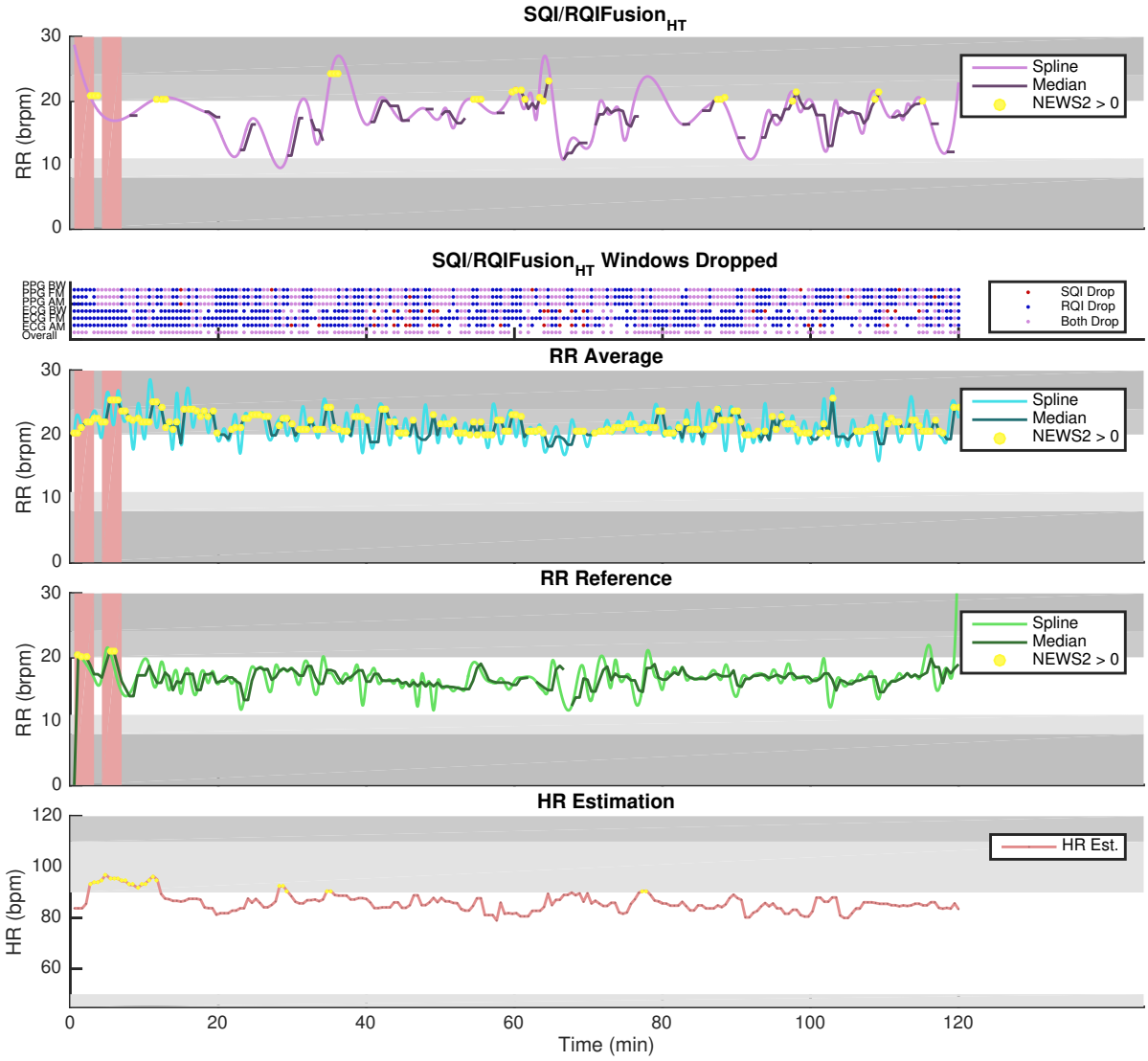


Figure 7.9: *Time-series RR and HR waveforms for two hours of monitoring in subject 3 in the $\mathcal{D}_d$ data set. Upper panel is the RR waveform from SQI/RQIFusion$_{HT}$. Second panel (from top) shows the the specific modulations (PPG BW, PPG FM, PPG AM, ECG BW, ECG FM, and ECG AM) that are below the discard threshold for the SQI (red dots), RQI (blue dots), or both the SQI and RQI (lavender dots) for each window. Bottom row of panel shows if a window was dropped (all six modulations dropped). Third panel (from top) is the RR waveform from the RR Average implementation. The fourth panel (from top) is the RR reference waveform. The bottom panel is the HR estimation obtained from ECG. In all RR waveform panels the lightest coloured waveforms (light lavender, teal, and green) represent the cubic spline interpolations of the RR estimates, the darker coloured waveforms (lavender, teal, and green) represent the three-point median filter of the RR estimates. The darkest coloured points (dark lavender, teal, and green) represent the RR estimates obtained from median filtering. In the HR waveform the red line represents the HR estimates. The darkest grey bars, grey bars, and lightest grey bars represent RR or HR regions (depending on waveform) where the NEWS2 score is 3, 2, or 1, respectively. NEWS2 scores of 0 on the waveform plots are represented by white regions. The rose regions on all of RR waveforms represent critical events as detected by the RR reference as defined in Section 7.3.1. Yellow points in RR and HR waveforms represent all instances where the given waveform had an RR or HR estimate with a NEWS2 score exceeding 0.*

value as opposed to a low SQI value (as shown in Figure 7.10, second panel). The high-quality of PPG and ECG data is further exemplified by observing the comparison of the NEWS2 events compared to the reference. The reference does not have any NEWS2

events with scores greater than 0. This same outcome is seen in both RQIFusion$_{\text{HT}}$ and RR Average where there are no events with NEWS2 scores above 0.

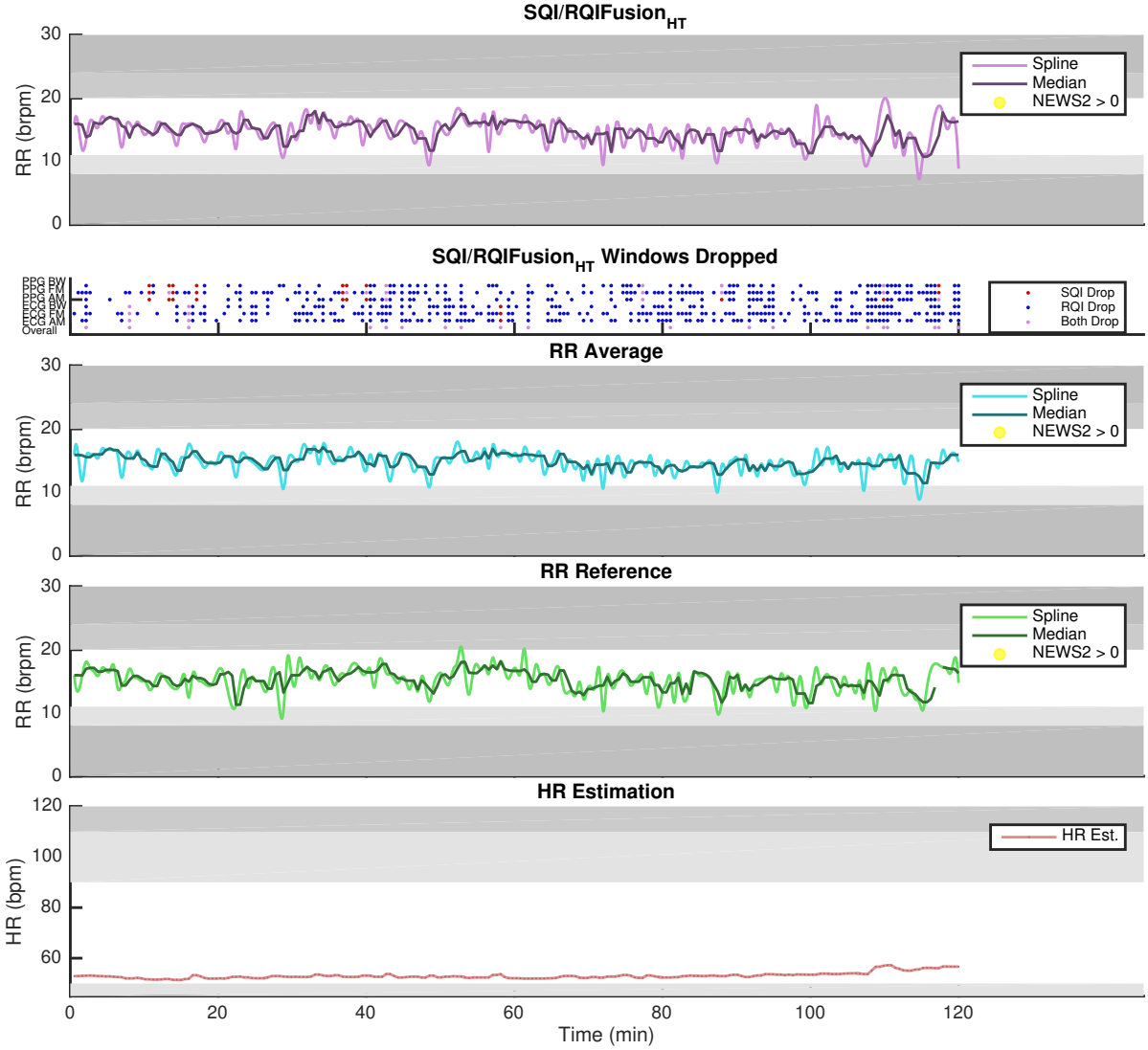


Figure 7.10: *Time-series RR and HR waveforms for two hours of monitoring in subject 8 in the $\mathcal{D}_d$ data set. Upper panel is the RR waveform from SQI/RQIFusion$_{HT}$. Second panel (from top) shows the the specific modulations (PPG BW, PPG FM, PPG AM, ECG BW, ECG FM, and ECG AM) that are below the discard threshold for the SQI (red dots), RQI (blue dots), or both the SQI and RQI (lavender dots) for each window. Bottom row of panel shows if a window was dropped (all six modulations dropped). Third panel (from top) is the RR waveform from the RR Average implementation. The fourth panel (from top) is the RR reference waveform. The bottom panel is the HR estimation obtained from ECG. In all RR waveform panels the lightest coloured waveforms (light lavender, teal, and green) represent the cubic spline interpolations of the RR estimates, the darker coloured waveforms (lavender, teal, and green) represent the three-point median filter of the RR estimates. The darkest coloured points (dark lavender, teal, and green) represent the RR estimates obtained from median filtering. In the HR waveform the red line represents the HR estimates. The darkest grey bars, grey bars, and lightest grey bars represent RR or HR regions (depending on waveform) where the NEWS2 score is 3, 2, or 1, respectively. NEWS2 scores of 0 on the waveform plots are represented by white regions. The rose regions on all of RR waveforms represent critical events as detected by the RR reference as defined in Section 7.3.1. Yellow points in RR and HR waveforms represent all instances where the given waveform had an RR or HR estimate with a NEWS2 score exceeding 0.*

The final category of SQI/RQIFusion$_{\text{HT}}$ is those where no improvement in the RR estimation is observed but there is a large percentage of the data discarded. These are

the difficult cases where SQI/RQIFusion$_{\text{HT}}$ under performs from expected. The exemplar case of this category is subject 1 (Figure 7.11). This is also seen in subject 4 (Figure A.7). In the case of subject 1, the RR Average underestimates the RR reference. The systematic underestimation of the RR is also seen in subject 4. The underestimation of the RR for subject 1 is seen in all of the PPG and ECG modulations. However, the average RQI values for all of the modulations are consistently low meaning that there are only a few windows where estimates are made by SQI/RQIFusion$_{\text{HT}}$. One plausible explanation for this finding is that there is another signal with a frequency lower than the RR waveform is also present in the extracted ECG and PPG respiratory waveforms and that it is causing systematic underestimation of the RR. In this case, as would be expected, SQI/RQIFusion$_{\text{HT}}$ is assigning low RQI values to these windows and they are being discarded. However, some of the RQIs are still high enough to pass the threshold which is resulting in SQI/RQIFusion$_{\text{HT}}$ giving a small number of estimates that also appear to underestimate the reference (Figure 7.6A). For cases such as this, the threshold for SQI/RQIFusion may need to be further increased; however, with the already low number of windows retained, this may not be possible. Interestingly, despite these findings, SQI/RQIFusion$_{\text{HT}}$ does appear to perform better at observing NEWS2 events than RR Average. Of the seven NEWS2 events with scores greater than 0 for the reference, the SQI/RQIFusion$_{\text{HT}}$ has NEWS2 scores above 0 for two of them while RR Average never has a NEWS2 score above 0.

Table 7.9 shows the counts for the TPs, FPs, and FNs in the RR Average and SQI/RQIFusion$_{\text{HT}}$ work flows compared to the RR reference and the calculation of the precision, recall, and F1 score. Results are shown for each individual subject in the $\mathcal{D}_d$ data set and in aggregate amongst all subjects. The results in Table 7.9 show that

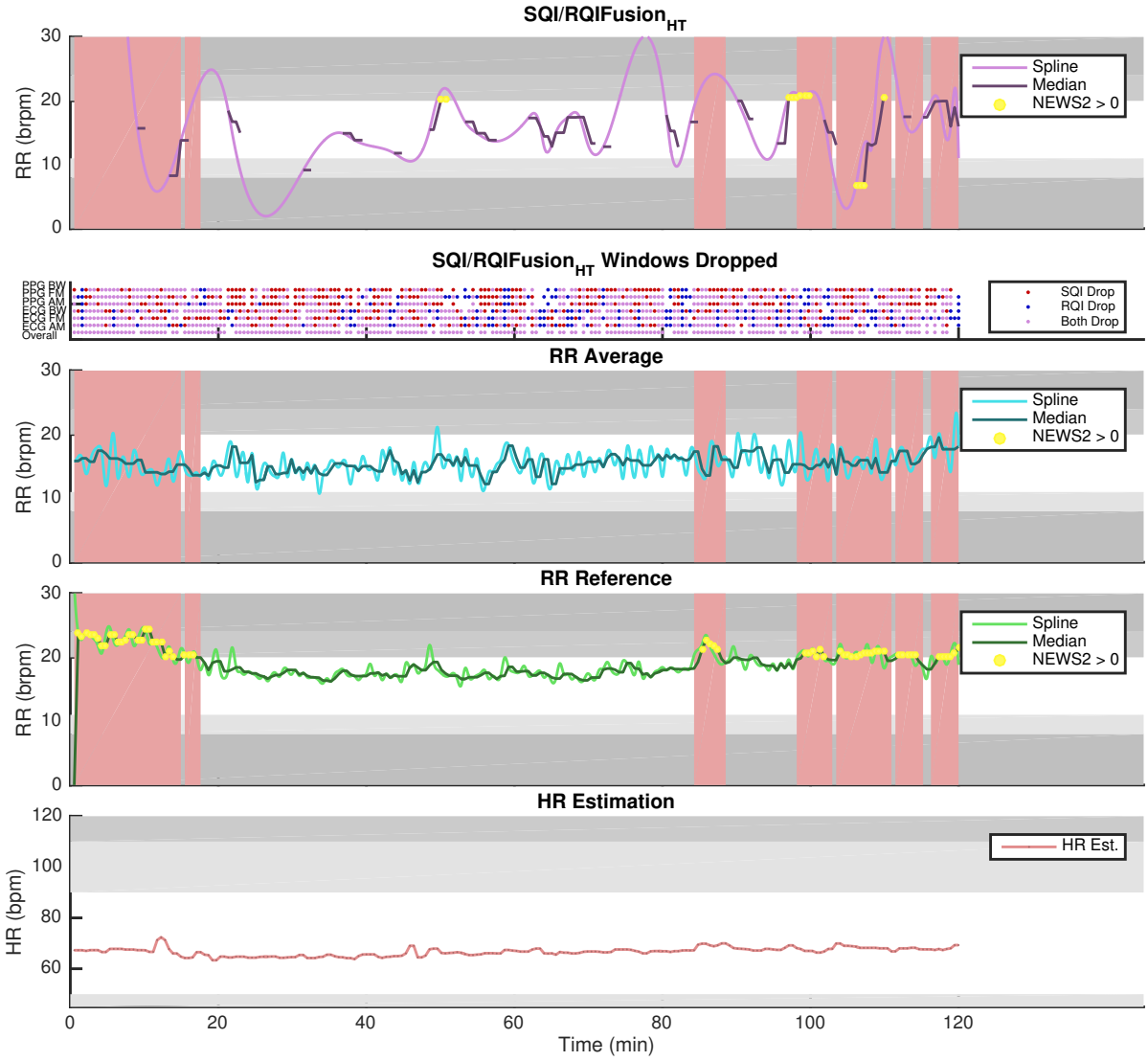


Figure 7.11: *Time-series RR and HR waveforms for two hours of monitoring in subject 1 in the $\mathcal{D}_d$ data set. Upper panel is the RR waveform from SQI/RQIFusion$_{HT}$. Second panel (from top) shows the the specific modulations (PPG BW, PPG FM, PPG AM, ECG BW, ECG FM, and ECG AM) that are below the discard threshold for the SQI (red dots), RQI (blue dots), or both the SQI and RQI (lavender dots) for each window. Bottom row of panel shows if a window was dropped (all six modulations dropped). Third panel (from top) is the RR waveform from the RR Average implementation. The fourth panel (from top) is the RR reference waveform. The bottom panel is the HR estimation obtained from ECG. In all RR waveform panels the lightest coloured waveforms (light lavender, teal, and green) represent the cubic spline interpolations of the RR estimates, the darker coloured waveforms (lavender, teal, and green) represent the three-point median filter of the RR estimates. The darkest coloured points (dark lavender, teal, and green) represent the RR estimates obtained from median filtering. In the HR waveform the red line represents the HR estimates. The darkest grey bars, grey bars, and lightest grey bars represent RR or HR regions (depending on waveform) where the NEWS2 score is 3, 2, or 1, respectively. NEWS2 scores of 0 on the waveform plots are represented by white regions. The rose regions on all of RR waveforms represent critical events as detected by the RR reference as defined in Section 7.3.1. Yellow points in RR and HR waveforms represent all instances where the given waveform had an RR or HR estimate with a NEWS2 score exceeding 0.*

in the overall F1 score for SQI/RQIFusion$_{\text{HT}}$ is twice the magnitude (0.12) of the F1 score for RR Average (0.06). While the number of TPs and FNs detected by both methods is similar, the number of FPs is far higher for RR Average. The result of this

is that the recall of both methods is similar but that the precision of SQI/RQIFusion$_{\mathrm{HT}}$ is higher. In five instances, the F1 score for individual subjects using SQI/RQIFusion$_{\mathrm{HT}}$ is higher than the RR Average implementation. As previously described, this is due to an increased precision value.

Table 7.9: *Analysis of F1 Score for the RR Average and SQI/RQIFusion$_{HT}$ implementations compared to the RR reference per subject in the $\mathcal{D}_d$ data set. Results show the number of TPs, FPs, FNs, precision, recall, and F1 score for each subject.*

Subject	SQI/RQI						No SQI/RQI					
	TP	**FP**	**FN**	**Precision**	**Recall**	**F1 Score**	**TP**	**FP**	**FN**	**Precision**	**Recall**	**F1 Score**
1	2	13	5	0.13	0.29	0.18	0	0	7	0.00	0.00	0.00
2	1	18	2	0.05	0.33	0.09	2	157	1	0.01	0.67	0.02
3	1	26	1	0.04	0.50	0.07	2	171	0	0.01	1.00	0.02
4	1	39	18	0.03	0.05	0.03	0	2	19	0.00	0.00	0.00
5	0	0	6	0.00	0.00	0.00	4	3	2	0.57	0.67	0.62
6	6	9	15	0.40	0.29	0.33	1	39	20	0.03	0.05	0.03
7	1	12	7	0.08	0.13	0.10	6	40	2	0.13	0.75	0.22
8	0	3	0	0.00	0.00	0.00	0	0	0	0.00	0.00	0.00
Total	12	120	54	0.09	0.18	0.12	15	412	51	0.04	0.23	0.06

Results also show that for monitoring instances of clinical derangement SQI/RQIFusion$_{\mathrm{HT}}$ is more useful than the HR estimate. For nearly all subjects, the HR estimation is constantly in a range providing a NEWS2 score of 0 or 1 and it rarely deviates out of that range. While this may be physiologically accurate, it is not useful for detecting potential acute critical events as the lack of deviation in the signal means that there is no difference in measurement between an event and a non-event.

Finally, these results demonstrate the use of the three-point median filter as a real-time smoothing function for continuous monitoring. While all subjects have some periods of drop out in their estimation of RR, for all but one subject, these periods of drop out are limited to less than 10 minutes in all instances. (Subject 5 has two instances where periods of drop out last longer than 10 minutes, one period lasting 17 minutes and another period lasting 16 minutes; all other periods of drop out are less than 10

minutes.)

7.4 Discussion

The aim of the present chapter was twofold. The first aim was to develop an RR estimation algorithm, termed RQIFusion, which made use of the RQI method developed in the previous chapters. The method makes use of both an RQI threshold to save only the highest efficacy modulations for the RR estimation in each window and weighted-averaging of the above-threshold modulations. The second aim was to apply RQIFusion to a previously untested two-hour time-series of ECG and PPG waveforms in clinical subjects undergoing dialysis who underwent an adverse event during the session (which occurs at the end of the monitoring period). This was meant to validate RQIFusion as a clinically relevant and useful technique for automated RR monitoring.

The first analysis, focusing on the development of RQIFusion, compared the RR estimation method to three standard RR fusion methods. These included: averaging, Smart Fusion, and modified Smart Fusion. Results shown in Figure 7.1 and Table 7.1 show that RQIFusion is capable of outperforming all of the state-of-the-art methods. Particularly, these results showed that when a threshold for RQIFusion is selected that holds either the percentage of windows without RR estimates or the RR error constant compared to the standard method being compared, RQIFusion always had a lower value for the unconstrained parameter.

These results also demonstrated one of the greatest benefits but also one of the greatest challenges of RQIFusion: threshold tuning. RQIFusion makes use of a cut-off threshold to discard poor-quality windows; however, this threshold can be adjusted where any adjustment has a consequent trade-off. When the threshold is relaxed, more

modulations are factored into the weighted-average step of the RR estimation and more windows have RR estimates, but the RR error accordingly goes up. This is most clearly demonstrated in Figure 7.2. The benefit of this is that the threshold can be modified for specific purposes or for specific patients. If, for a given application, more continuous RR estimation is preferred over lower error, the threshold can be increased. The converse is also true, if lower RR error is preferred with a less frequently-estimated RR. To clearly demonstrate this, three thresholds were selected based on criteria defined in the $\mathcal{D}_{c,1}$ training and the validation data set. These thresholds, T_{LT}, T_{HT}, and T_{SF} were applied to $\mathcal{D}_{c,1}$ as well as two test sets, $\mathcal{D}_{c,2}$ and $\mathcal{D}_m$. The results, shown in Table 7.2, demonstrate that in nearly all instances across all data sets a similar RR error can be achieved compared to the comparator methods, but that a much lower number of windows need to be discarded. The result of this is that RQIFusion is an equally robust, if not improved, RR estimation method with increased estimation continuity. These results are further considered in instances where only the ECG or PPG waveform is available. In all of these instances and across all data sets, RQIFusion performs better than Smart Fusion.

Finally, RQIFusion has been validated as a highly robust RR estimation technique for individual subject time-series. Figure 7.3 shows that in two exemplar cases, RQIFusion is both more accurate and continuous than both the Smart Fusion and modified Smart Fusion methods.

Overall, the construction and validation of RQIFusion demonstrates that the RQIs can be effectively implemented to improve RR estimation compared to current state-of-the-art methods. This was demonstrated for a number of different cases, considerations, and data sets where all the results demonstrated a consistent improvement in accuracy,

robustness, and continuity of estimates. The RQIFusion method however presents the primary consideration of threshold selection where different parameters must be considered when selecting the threshold.

The promising results of RQIFusion motivated the second analysis, which aimed to use and validate RQIFusion in an RR estimation work flow that included both the SQI and RQI checkpoints along with RR modulation extraction, RR estimation, and RR fusion. In order to demonstrate clinical relevance, this analysis was implemented in the $\mathcal{D}_d$ data set, a clinical data set of subjects receiving dialysis, where all recordings lasted two hours.

The initial analysis of the RQI in the $\mathcal{D}_d$ data set looked at the general performance of the RQI. In the reference RR waveform, it was observed that the average RQI obtained across the entire windowed data set was positively correlated with the percentage of windows where a reference RR estimate was obtained from manual, expert analysis. This is a strong indicator of the validity of the RQI in assessing the underlying quality of respiratory modulations. The general performance of the RQI for use in RR estimation from the ECG and PPG in the $\mathcal{D}_d$ data set was also assessed using RQIFusion. In particular, the choice of the threshold for RQIFusion was again examined where a panel of threshold values surrounded by T_{HT} were tested. The results from this analysis, shown in Figure 7.5, demonstrate that as with previous analyses, as the threshold is increased, the number of windows without RR estimates increases, but the RR error concurrently decreases. However, what was also studied in this analysis is the types of modulations that are most often the highest quality (i.e., have the highest RQI). It was observed that PPG amplitude modulation and ECG baseline wander are the modulations that are most often kept. This is a critical finding as it indicates that both ECG and PPG

play an important role in RR estimation. It was also shown that, even in cases with the most stringent threshold, there are still instances where all six modulations are used to make an RR estimate (indicating excellent RQI performance in all six modulations). Finally, these results also validated the use of T_{HT} as a suitable threshold in the majority of cases. In the entire $\mathcal{D}_{c,1}$ data set, the use of T_{HT} in RQIFusion obtained an average RR error of 3.65 ± 4.27 while providing and RR estimate for 73% of windows. Thus, it is recommended that the T_{HT} threshold (0.44) be used in the majority of RR estimation using RQIFusion and that use of different thresholds only be considered in exceptional cases.

The next analysis in the $\mathcal{D}_d$ data set focused on the implementation of an RR estimation work flow (as shown in Figure 3.1) which makes use of the RQI. In this work flow, an SQI is initially used to drop poor quality windows based on the ECG or PPG. Next, the six RR modulations are extracted and assessed for quality using the RQI. The RR for each retained modulation is then estimated and RQIFusion is applied to obtain the final RR estimate per time window. The aim of this analysis was not only to show the effective implementation of RQIs and RQIFusion into a plausible clinical work flow, but also to demonstrate the RQIs robustness when different RR estimation implementations are considered. Five different RR estimation implementations were used—three frequency-based: (1) *ARSpec*, (2) FFT estimation, and (3) autocorrelation estimation; and two time-series based: (4) count-origin estimation, and (5) peak detection & zero-crossing estimation. Results from this analysis on the whole $\mathcal{D}_d$ data set show a number of key results (as shown in Table 7.7). The first is that the performance of the RQI and RQIFusion is not biased towards *ARSpec*. Given that the determination of the RQI was trained using RR estimates from *ARSpec* it was possible that the RQI

would only perform best when the RR for windows was estimated using *ARSpec*. This would reduce the general applicability of the RQI checkpoint as a universally important step in improving RR estimation. This however was not the case as all of the other RR estimation methods assessed had similar performance when the RQI was incorporated. The other important finding from these results is that both the SQI and RQI seem to be critical in improving RR estimation. On average, in the SQI Average and RQIFusion$_{\text{HT}}$ implementations, one quarter of the windows were discarded as being of poor quality and RR errors decreased compared to the RR Average implementation. Using SQI/RQIFusion$_{\text{HT}}$, this number increased to approximately half and RR errors were further lowered. What this demonstrates is that not only are the SQI and RQI methods discarding bad windows, but they are almost exclusively discarding different windows. This strongly suggests that in order to best improve RR estimation, both the SQI and RQI should be used in tandem.

The final analysis in $\mathcal{D}_d$ made use of SQI/RQIFusion$_{\text{HT}}$ with RR estimation being conducted using the combined peak-zero estimation method (due to the low standard deviation of the method) in the individual subjects in the data set. Results from the individual subjects showed variability between subjects.

The plots of the RR differences between the reference and SQI/RQIFusion$_{\text{HT}}$ and RR Average in Figures 7.6 and 7.7 demonstrated that the difference between SQI/RQIFusion$_{\text{HT}}$ and the reference is smaller than the difference between RR Average and the reference.

The plots of the individual waveforms for subjects showed four different categories of outcomes for SQI/RQIFusion$_{\text{HT}}$. The four categories were: (1) instances where SQI/RQIFusion$_{\text{HT}}$ improved RR estimation with few windows discarded (subjects 2,

6 and 7), (2) instances where SQI/RQIFusion$_{\mathrm{HT}}$ improved RR estimation but many windows were discarded (subjects 3 and 5), (3) instances where SQI/RQIFusion$_{\mathrm{HT}}$ did not improve RR estimation but also did not discard many windows (subject 8), and (4) instances where SQI/RQIFusion$_{\mathrm{HT}}$ did not improve RR estimation and discarded many windows (subjects 1 and 4).

The first three categories showed the versatile performance of SQI/RQIFusion$_{\mathrm{HT}}$. In the first category, there were some erroneous RR modulations, often stemming from the poor quality of one of the RR modulations. In these cases, SQI/RQIFusion$_{\mathrm{HT}}$ successfully discarded that modulation and estimated from the remaining modulations which is why RR estimates improved but few windows were discarded. In the second category, many windows were discarded but RR estimates improved with the use of SQI/RQIFusion$_{\mathrm{HT}}$. These cases represented those in which many of the extracted modulations were erroneous and thus were discarded but that when an accurate modulation existed, an improved RR estimate could be made. The third category represents the instances where SQI/RQIFusion$_{\mathrm{HT}}$ is presented with already high-quality data. In these instances, the RR estimation could not be improved, but because of its uniformly high-quality, SQI/RQIFusion$_{\mathrm{HT}}$ did not discard very many windows.

The fourth category demonstrates the challenging cases for the SQI/RQIFusion$_{\mathrm{HT}}$. In both of these cases, it appears that the RR estimations from RR Average underestimated the reference. As a result, many of the windows were discarded as all of the modulation were poor. However, when some modulations had high RQI values, the RR estimations were generally still underestimating the reference. It is this subset of cases for which further consideration and fine tuning of SQI/RQIFusion$_{\mathrm{HT}}$ is necessary.

Overall, this analysis demonstrates that the RR for six of the eight subjects in the

$\mathcal{D}_d$ data set could have been successfully monitored using SQI/RQIFusion$_{\mathrm{HT}}$. The other two cases might have been able to be monitored with fine tuning of the RQI threshold. However, the clinical challenge presented by this is that without *a priori* information about RR, it is not possible to know for which patients the threshold needs to be fine-tuned for. One of the potential explanations for the results observed in subjects 1 and 4 is that there was an additional low-frequency signal that was present in the RR modulations that caused an underestimation of the RR. Further evaluation of these cases may elucidate the cause and nature of this artefact.

Despite the challenges presented by subjects 1 and 4; however, SQI/RQIFusion$_{\mathrm{HT}}$ is an improvement on RR Average in all cases. In subjects 2, 3, 5, 6, 7, and 8, SQI/RQIFusion$_{\mathrm{HT}}$ either outperformed or performed on par with RR Average. Even in subjects 1 and 4 in which it was not able to improve the RR estimation, the fact that high levels of data were being discarded is an indication that the extracted respiratory waveforms may be low-quality. An observation that is not possible using the RR Average method.

The analysis of the NEWS2 scores showed that the general improvement in RR estimation is beneficial in detecting critical events. This was done using the reference RR to determine when critical RR events occur, based on the NEWS2 score of the RR and then observing if SQI/RQIFusion$_{\mathrm{HT}}$ compared to the RR Average implementation detected events at a similar point in time. The results were twofold. Both methods were equally successful at detecting true positive events, but SQI/RQIFusion$_{\mathrm{HT}}$ reduced the number of false negatives by nearly fourfold. This is a very important finding in clinical settings as it has been widely reported that false alarms in hospitals can account for between 72% and 99% of all alarms. The result of the high number of

false alarms is provider alarm fatigue which is a critical patient safety concern [91, 234]. The results of the individual time-series for the $\mathcal{D}_d$ subjects show that the main effect of SQI/RQIFusion$_{\text{HT}}$ is decreasing the RR estimate compared to when they are not used. Prior to implementation of the SQI and RQI, it appears that the RR estimates systematically overestimate the RR reference. These results also support the claim that RR is the most important vital sign preceding a catastrophic decline [4, 11, 13, 14, 15]. In comparing the SQI/RQIFusion$_{\text{HT}}$ waveform to the HR estimation waveform, it is clear for the vast majority of patients in $\mathcal{D}_d$ that the HR is relatively constant throughout the entire monitoring session. For all but two subjects, the HR waveform indicates an HR that either gives a constant NEWS2 score of 0 or 1. This finding is not useful as a marker for potential decline as a constant score does not indicate any change in overall well-being. The RR estimates for all patients give a much more descriptive picture of the overall patient status. In almost all cases, the patient's RR fluctuates between NEWS2 score categories including time periods of slowed and increased RR. While the meaning of these results is purely observational, given that it is known that all of these subjects had a critical event at the end of the two hour monitoring period, it is a very interesting finding. A limitation of this finding however is that it cannot be compared to instances where patients received the entire dialysis treatment and did not have a critical event at the end.

7.5 Conclusion

The results in this chapter show that the RQI is an important element in clinically relevant RR estimation from ECG and PPG. It was demonstrated that a relatively conservative threshold (T_{HT}) is capable of providing high-quality RR estimation. It was

also demonstrated that the implementation of the RQI into a complete RR work flow in a two-hour time-series vastly improves RR estimation at all time points compared to when an RQI is not implemented. These results further show the unique importance of both SQIs and RQIs in improving RR estimation. These results show that a complete RR work flow incorporating an SQI and the RQI in a two-hour time-series greatly improves RR estimation at all time points as indicated by the reduced differences in the RR estimates compared to the reference when the SQI and RQI are implemented compared to when they are not. Additionally, RR estimation using both an SQI and RQI is more effective at reducing false alarms in monitoring compared to when neither checkpoint is used. For eight 120-minute recordings, 412 false alarms were observed when no SQI or RQI were implemented compared to 120 when both an SQI and RQI were implemented. Finally, these results show that, in the majority of cases, the RR is a better indicator of patient decline than is HR. The eight 120-minute patient recordings of subjects that had critical events demonstrated that the RR was a more fluctuating and volatile vital sign compared to the RR. Therefore, in these subjects it would likely be an alarm due to changes in respiratory status as opposed to cardiac status which would serve as the first indicator of deterioration.

Chapter 8

Conclusion

We here review the purpose of this thesis, before presenting our concluding remarks. Monitoring of vital signs is one of the most important indicators of a patient's health status [1]. Among the most commonly measured vital signs, RR is not only the most critical but also often the most sporadically and inaccurately monitored [4, 11, 13, 14, 15, 16]. Despite available RR monitoring technologies, such as IP, TTI, capnography, flow thermography, and spirometry, RR is still most often collected manually, a practice which introduces high error into RR measurements [2]. Furthermore, because of their cost and invasive nature, purpose-built RR technologies remain reserved for the most critical ICU cases [31, 32, 33].

Numerous efforts to estimate RR using secondary signals from commonly-collected measurements such as ECG, PPG, and BP have been shown to be effective in controlled settings; however, results of previous studies of these methods lack adequate clinical robustness [44].

One of the principal reasons these methods lack robustness is that in the majority of all previous work, all of the respiratory waveforms extracted from the primary signal (ECG, PPG, or BP) have been assumed to be true respiratory waveforms. This assump-

tion however fails to acknowledge that the respiratory waveform, which can be obtained from the primary signal, is based on important physiologic modulations which *may or may not be present* regardless of the quality and reliability of the primary signal itself.

The aim of this thesis was to demonstrate that when using respiratory waveforms derived from ECG and PPG signals, it is not only necessary to measure the signal quality of the primary ECG and PPG signal using an SQI, but that RR analysis is drastically improved by estimating the likelihood that the extracted respiratory waveform *contains physiologically-relevant respiratory information* using an RQI. The work presented demonstrated that incorporation of RQIs into RR estimation from ECG and PPG signals exhibits a number of advantages and is capable of providing substantially more accurate and complete RR monitoring compared to other available methods.

The key physiological feature that allows respiratory waveforms to be extracted from ECG and PPG signals is the modulation of the (1) respiratory system on (2) the cardiovascular system. The physiologic interrelationship between these two systems (and how this can be exploited for estimating the likelihood that an extracted respiratory waveform contains physiologic respiratory information with no other *a priori* information) was explored in Chapter 2. Principally, these modulations can be considered in two categories: (1) physical means, by which the proximity of the expanding and deflating lungs exert an effect on the cardiac output; and (2) nervous-system means, by which interplay between the body's systemic needs at a given instant drive the relationship between RR and HR.

Chapter 3 presented a novel framework, shown in Figure 3.1, for understanding the interconnection between all previous work in the literature. This framework described three estimation "modules" and two quality "checkpoints" that are critical for estimating

RR from ECG and PPG. All of the methods in the first and second modules (extraction of the respiratory waveform and extraction of the RR from the respiratory waveform, respectively) can be broken down into two classes: time-domain and frequency-domain methods. The third module, RR fusion, has been employed by far fewer authors and as such the methods employed feature far less variability. The two data-quality check points are implementation of SQIs and our RQIs. However, it is also observed that only three examples of what can be construed as RQIs have previously been described, demonstrating a key motivation for improved RQI methods.

Individual RQIs, $Q_{S,1} \dots Q_{S,5}$, were based on five signal processing techniques: FFT, autocorrelation, cosine cross-correlation, autoregression, and Hjorth complexity, in Chapter 5. The results demonstrated both accuracy and robustness for the individual RQIs based on the FFT ($Q_{S,1}$), autocorrelation ($Q_{S,2}$), and autoregression ($Q_{S,4}$). Each of these individual RQIs out-performed the standard F1 score SQI and was able to predict the quality of the respiratory information in the respiratory waveforms in both the held-out data sets. Furthermore, each of these methods was implemented to capture different anticipated features of the respiratory waveforms. Comparative studies of these methods in HRV show that each provides different information in regards to the quality of a respiratory modulation, despite methodological similarities [215].

Chapter 6 sought to use the three best individual RQIs from Chapter 5 to create a single fusion RQI. Three methods, two supervised and one unsupervised, were tested. The supervised methods included a method based on linear regression ($Q_{F,1}$) and support vector regression ($Q_{F,2}$). The unsupervised method was based on principal component analysis ($Q_{F,3}$). The results from this analysis demonstrated that both $Q_{F,1}$ and $Q_{F,3}$ had high accuracy and robustness. The more complex, non-linear SVR method, $Q_{F,2}$

does not improve results and is a poor estimator of respiratory quality in held-out data. Additional analysis comparing each of the fusion RQIs to the three input individual RQIs showed that both $Q_{F,1}$ and $Q_{F,3}$ improve on the individual RQIs. Each individual RQI has instances where it wrongly predicts the presence or absence of high-quality respiratory information, but it is only in rare scenarios that all three do this. As a result, the fusion RQIs are more robust to this variability and provide more reliable estimates of the quality of the underlying respiratory information compared to the individual RQIs when considered in isolation.

Finally, a methodology using $Q_{F,1}$ was defined to estimate the RR from all six of the respiratory modulations obtained from both ECG and PPG in a single time window in Chapter 7. This methodology is predicated on two factors. The first is a variable threshold, which can be used to discard respiratory modulations deemed to be of poor quality. The second is a weighted averaging of respiratory estimates based on their RQI values. The performance of RQIFusion is evaluated using two analyses. The first analysis compares RQIFusion to two current methods for fusing multiple RR estimates into a single estimate: averaging and Smart Fusion [66]. These results show that for both individual windows and for individual subjects RQIFusion is able to provide more accurate estimates at more time points than both comparator methods. The final analysis implements $Q_{F,1}$ as the RQI checkpoint and RQIFusion as module III of the RR estimation work flow described in Figure 3.1. These results showed that, regardless of the RR estimation method used for module II, implementation of both $Q_{F,1}$ as a quality checkpoint and RQIFusion is capable of improving RR estimation accuracy. These results also demonstrated the co-dependency of both the SQI and the RQI in selecting the highest-quality respiratory waveforms, demonstrating the importance of RQIs for

RR estimation. These results are shown in the $\mathcal{D}_d$ data set. Chapter 7 also shows that simply averaging RR estimates across modulations without implementation of an SQI or RQI systematically overestimates the RR while an implementation using both an SQI and RQI increases the accuracy of RR estimation and improves the ability to observe critical respiratory events.

8.1 Future Work

The performance of RQIs presents multiple opportunities for implementation into on-going research on vital-sign monitoring, and providing early warning of deterioration as well as opportunities for continued testing and improvement of the RQI methodology. Specifically, two areas of future work include (1) use of RQIs for improvements to EWS systems, and (2) implementation of RQIs for RR monitoring in mobile settings (such as wearables).

8.1.1 EWS Systems

As was described in Chapter 7, the current clinical standard for EWS monitoring in the UK is the National Early Warning Score 2 (NEWS2) system [233]. In addition to NEWS2, there are a number of other EWS systems including ViEWs [235], mCEWS [236], and MEWS [11]. Despite being derived from a variety of both heuristic and machine learning methods, all of these systems remain dependent on manual clinical measurement of the vital signs. While it has been observed that electronic EWS systems, which provide automatic EWS calculations after input of the vital signs by clinicians, can improve the level of completeness of EWS measurements, there often remain many missing observations, particularly in measuring RR [237, 238].

Given the intermittent nature of EWS measurements by clinicians, there is potential for implementation of continuous EWS measurement from continuous, automatic vital-sign monitoring. Despite this, there remain few examples of implementations of continuous, automatic EWS measurements, one of the most prominent being cCEWS [239]. One reason for this is that a poor implementation of the technology required for continuous, automated monitoring has the potential to worsen patient care [240]. This is primarily attributable to a lack of a robust method for estimating RR from the waveform data most often collected in the hospital [241]. As noted by Pimentel (2015), a robust method for estimating RR from ECG and PPG holds great promise for continuous health status monitoring systems as it will provide additional channels (in conjunction with manual monitoring) of RR which can be used as an input in multi-source estimates of patient status [241].

Given the limitation of continuous EWS monitoring due to the lack of a robust and continuous RR estimate, as described by Pimentel (2015) [241], the present work, which demonstrates an RR estimation technique with accuracy and robustness suitable for clinical use, presents the opportunity to incorporate RQI-based RR monitoring into continuous EWS monitoring systems. The availability of continuous, robust RR estimates also increases the potential for more complex, non-linear EWS systems which are based, for example, on Gaussian processes such as those described by Kapoor *et al.* (2005) and Xiao *et al.* (2013) [242, 243].

8.1.2 Wearable Technologies

Introduced in the late 1990s, wearable health devices (WHDs) have seen a massive growth in use as the current technology revolution has resulted in miniaturised devices

that are both reliable and adaptive [244]. WHDs are increasingly used for both fitness and health status monitoring [244]. Demand for vital-sign monitoring using WHDs for health status monitoring have continued to grow as these devices are capable of monitoring vital signs over days and weeks for patients outside hospitals and clinics. This has ensured better diagnosis and recovery from medical interventions [244, 245, 246].

Interest in monitoring vital signs via WHDs has mainly focused on the traditional vital signs: HR, BP, temperature, SpO_2, and RR [8, 244]. RR has been of particular interest not only because it is often the earliest and most accurate predictor of clinical deterioration, but also because of the need for continuous, ambulatory monitoring for the detection of sleep apnoea, chronic obstructive pulmonary disease (COPD), and asthma [244]. Furthermore, continuous, ambulatory monitoring of RR is of particular interest for children with pulmonary disease [8, 244, 247].

The challenge of monitoring RR in WHDs, however, is the same challenge that is present in monitoring RR in hospital settings: the available technologies, such as IP, elastomeric plethysmography, and respiratory inductive plethysmography, all require restrictive devices to be placed on the torso. Therefore, the ability to unobtrusively measure ECG and PPG and use those primary signals to obtain respiratory waveforms to estimate RR is of great importance. Despite advances in measuring ECG using minimally-invasive flexible cardiac sensors placed on the chest, these technologies have rarely been implemented outside of cardiac pathophysiology studies of short duration (24-48 hours) [248]. Monitoring of PPG in ambulatory devices such as wristbands is far more common [248].

Despite the growing regularity with which PPG is monitored with WHDs at the wrist, the use of the PPG signal obtained from WHDs at this location on the body

must overcome a number of challenges before it can be used for medical monitoring of RR. The first of these challenges is overcoming signal artefact. Signal artefact in the PPG obtained from WHDs is derived from multiple sources. Because the WHD uses the body potential as ground voltage, the ground is floating and is subject to body motion-related variability [249]. In addition to this, WHDs rely on reflectance to measure PPG. Reflectance PPG is much more susceptible to signal artefact than transmission PPG because motion of the sensor, pressure variability between the sensor and the skin, and changes in impedance between the skin and the sensor can all introduce noise [249, 250]. Finally, measurement location also introduces signal artefact. The wrist is known to have low perfusion and as a result suffers from lower PPG pulse amplitude than other anatomical locations [80]. The result of this is a signal that is more prone to being corrupted by signal noise [249].

The second challenge in using wrist-worn WHDs to obtain reliable RR estimations from PPG waveforms is the computational requirements. Conventional analytical methods often have high processing requirements, large power consumption, and over-sized, burdensome devices [249]. A robust and reliable wrist-worn WHD for medical monitoring must overcome this computational challenge via lightweight computation.

Efforts to address these challenges are ongoing and the implementation of RQIFusion for RR estimation described in this thesis could be adapted to address these concerns. The RQIs described in this thesis are not designed to address signal artefact introduced due to motion. Rather, this is the primary aim of the SQI checkpoint. While the SQIs presented and used in this thesis are robust for hospitalised or immobile patients, their use in clinical mobile or ambulatory settings is largely unproven. However, development of robust SQIs for mobile and ambulatory monitoring applications is ongoing [251].

One of the challenges that RQIFusion could be adapted to address is high computational requirements and energy consumption. Unlike Smart Fusion, which was designed for optimal power efficiency in mobile applications, RQIFusion was not designed to take into consideration computing requirements or power efficiency measures for mobile and ambulatory applications [66]. However, one potential method for addressing this is a three-tiered data analysis platform shown in Figure 8.1. The first tier of the platform represents the WHD itself. This is the "first line of defence" and represents the most computationally simple estimation algorithm. For example, in instances of RR estimation, the platform may only extract one RR modulation from the PPG and make use of two individual RQIs. In instances where the RQIs agree, RR could be estimated, but in instances where they do not agree, the PPG waveform for the time window in question is passed to a mobile platform (such as a nearby smartphone) via a mobile transfer connection (such as Bluetooth). The larger computational capacity of the mobile platform would allow it to conduct more computationally complex implementations such as the full RR estimation workflow presented in this thesis. This RR estimation can then be passed back to the sensor and additionally, if necessary, can be passed up to the third tier: the central server via mobile Wi-Fi or a broadband cellular network connection. The central server would then be responsible for higher-order computational integration, such as integration into continuous EWS monitoring as described in Section 8.1.1.

8.2 Conclusion

The work of this thesis presents a novel contribution to RR estimation from ECG and PPG waveforms by defining and validating a pre-processing step in RR estimation. These RQIs can quantify the underlying quality of the physiologic respiratory informa-

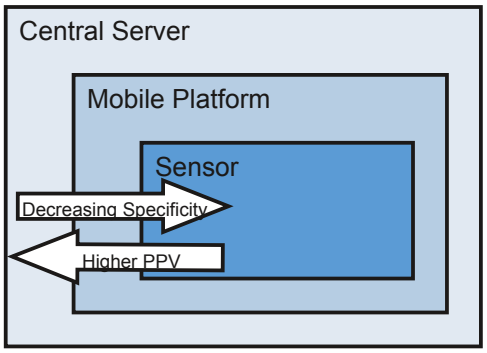


Figure 8.1: *Diagram of a platform for integrating necessity for computational efficiency and accurate and robust RR estimation in mobile health technologies.*

tion contained within respiratory modulations extracted from the ECG and PPG waveforms. This was performed through the implementation of signal processing methods to extract relevant quality measures, and through supervised and unsupervised machine learning methods to fuse these quality measures into a single metric. This metric was validated on three unique and independent patient populations. Furthermore, this metric, along with a method for fusing multiple available RR estimates, was implemented in conjunction with other critical signal quality and RR extraction and estimation algorithms to provide a single robust RR estimate. While further work and validation is required before RQIs can be implemented into applications for patient care, the work described herein describes an important addition to RR estimation.

Appendix A

Additional Results

A.1 Comparison of RR Estimation Methods

Throughout this thesis, *ARSpec* has been used as an exemplar "state-of-the-art" RR estimation algorithm for estimating the RR from respiratory modulations [164]. As a result, the RQIs have been trained and validated on the assumption that *ARSpec* is an accurate RR estimation method. The results in Table A.1 show the percentage of windows discarded (due to the algorithms inability to produce an estimate for that modulation) and the MAE (± SD) for all of the windows and modulations in $\mathcal{D}_c$, $\mathcal{D}_m$, and $\mathcal{D}_d$ for five "state-of-the-art" RR estimation algorithms: (1) *ARSpec* [164], (2) Fourier transform estimation [66], (3) autocorrelation estimation [145], (4) count-origin estimation [145], and (5) peak-detection and zero-crossing estimation [232]. These results show that there is not a uniform best-performer across all three data sets. It is shown that *ARSpec* performs uniformly well and consistently with the other estimation algorithms across the data sets which supports its use as the exemplar algorithm for training and validation of the RQIs. The relationship between the five "state-of-the-art RR estimation algorithms is further explored in Section 7.3.2.

Table A.1: *Evaluation of the RR error (± SD) and % of windows discarded for the combined set of PPG and ECG RR modulations for five RR estimation algorithms: ARSpec, FFT, autocorrelation, count-orig, and combined peak-zero in the $\mathcal{D}_c$, $\mathcal{D}_m$, and $\mathcal{D}_d$ data sets.*

Est. Method	$\mathcal{D}_c$		$\mathcal{D}_m$		$\mathcal{D}_d$	
	% Win	RR Error	% Win	RR Error	% Win	RR Error
ARSpec	100.0%	3.00 ± 5.58	100.0%	5.34 ± 5.65	100.0%	3.86 ± 3.20
FFT	100.0%	3.30 ± 5.72	100.0%	5.11 ± 5.57	100.0%	3.59 ± 3.12
Autocorrelation	100.0%	2.90 ± 5.66	100.0%	4.91 ± 5.51	100.0%	3.76 ± 3.26
Count-Orig	96.9%	6.69 ± 6.38	98.3%	6.66 ± 6.17	100.0%	5.75 ± 4.77
Combined Peak-Zero	100.0%	4.61 ± 4.77	100.0%	4.32 ± 4.15	100.0%	3.89 ± 2.88

A.2 Robustness Measurement of Individual RQIs

Figures A.1, A.2, A.3, A.4, and A.5 present the entire results for Figure 5.6. These results show that each of the individual RQIs has a monotonically decreasing trend in the frequency with which the rank of the RQI falls within the given decile of the rank of the RR error. This result indicates that there are few RQI ranks that are exceptionally different from the RR error rank. It is also noteworthy that the steepness of the downward trend is greatest for $Q_{S,1}$, $Q_{S,2}$, and $Q_{S,3}$, the three best-performing individual RQIs.

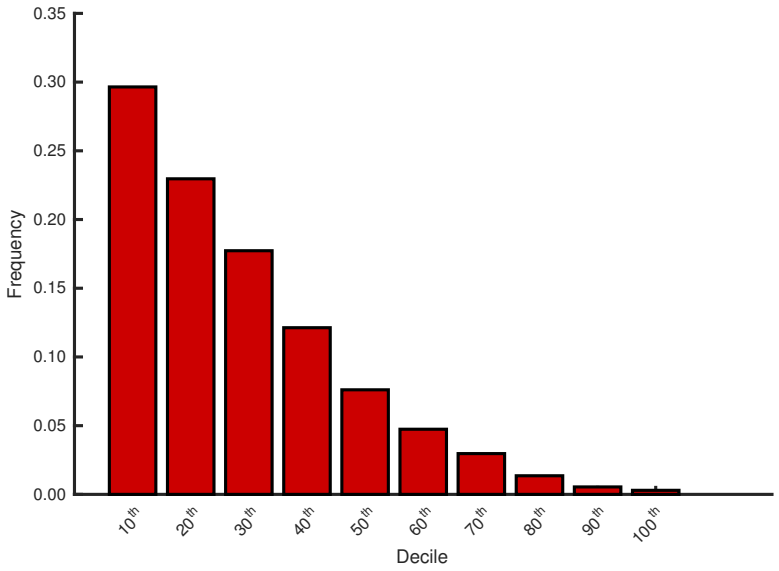


Figure A.1: *Frequencies in which the rank orderings of each window of $Q_{S,1}$ falls within the expected decile based on the rank ordering of the RR error in the data supra-set containing $\mathcal{D}_c$ and $\mathcal{D}_m$.*

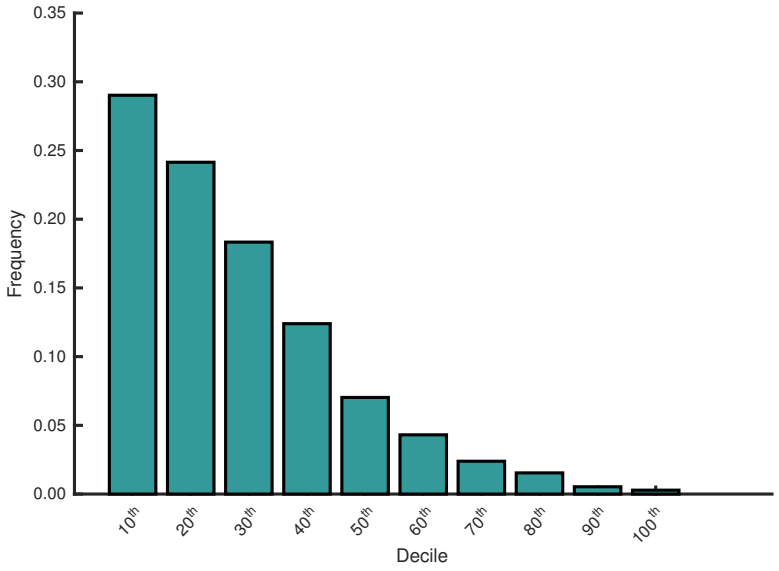


Figure A.2: *Frequencies in which the rank orderings of each window of $Q_{S,2}$ falls within the expected decile based on the rank ordering of the RR error in the data supra-set containing $\mathcal{D}_c$ and $\mathcal{D}_m$.*

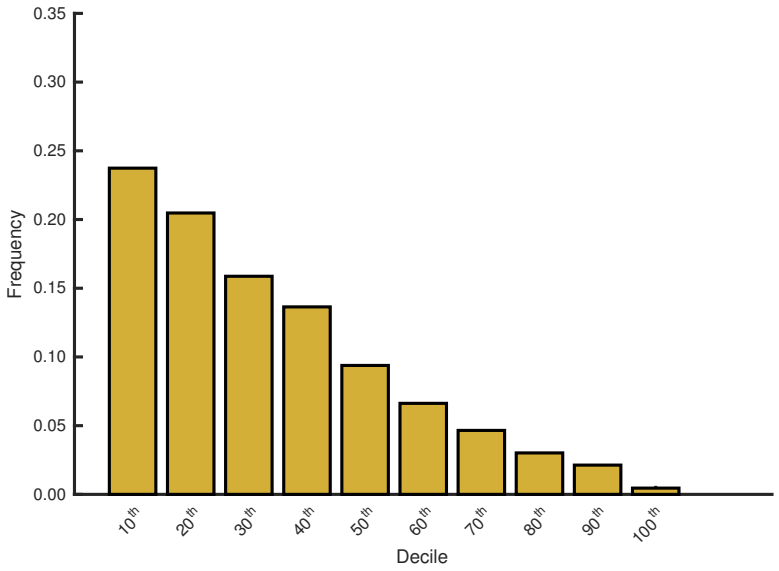


Figure A.3: *Frequencies in which the rank orderings of each window of $Q_{S,3}$ falls within the expected decile based on the rank ordering of the RR error in the data supra-set containing $\mathcal{D}_c$ and $\mathcal{D}_m$.*

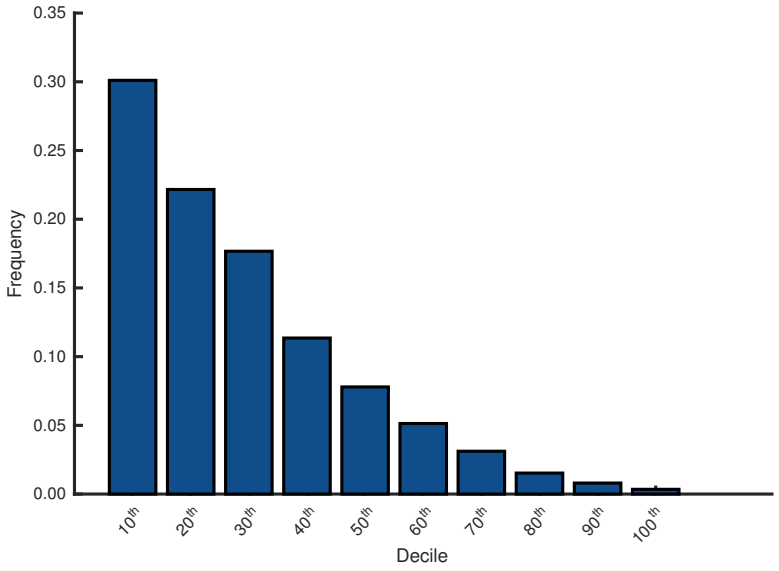


Figure A.4: *Frequencies in which the rank orderings of each window of $Q_{S,4}$ falls within the expected decile based on the rank ordering of the RR error in the data supra-set containing $\mathcal{D}_c$ and $\mathcal{D}_m$.*

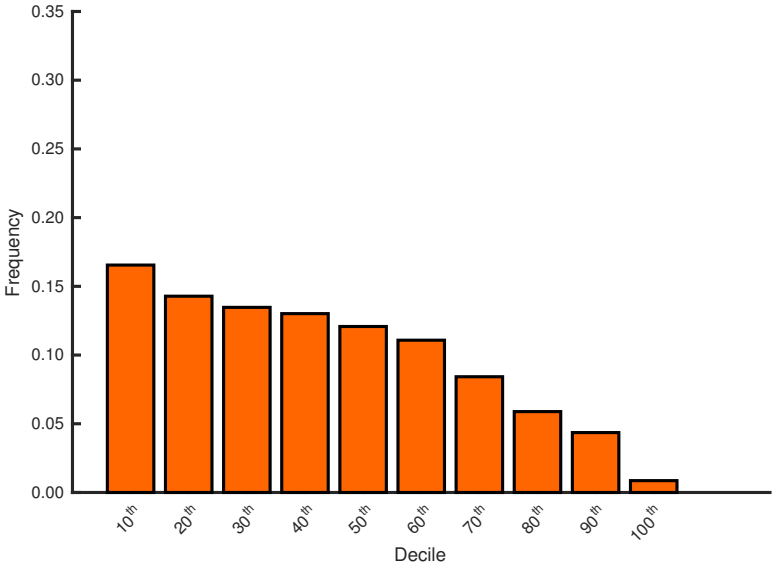


Figure A.5: *Frequencies in which the rank orderings of each window of $Q_{S,5}$ falls within the expected decile based on the rank ordering of the RR error in the data supra-set containing $\mathcal{D}_c$ and $\mathcal{D}_m$.*

A.3 Time-Series Analysis

Chapter 7 presents time-series analyses for subjects in both the $\mathcal{D}_{c,2}$ and $\mathcal{D}_d$ data sets. The time-series analyses for the remaining subjects in those data sets is presented here. Figure A.6 shows the time-series RR estimates for the the eight remaining subjects in the $\mathcal{D}_{c,2}$ data set. These results are analogous to the results shown in Figure 7.3 for the remaining $\mathcal{D}_{c,2}$ subjects. Figures A.7, A.8, A.9, and A.10 show the analogous results to Figures 7.8, 7.9, 7.10, and 7.11 for subjects 4, 5, 7, and 7 in the $\mathcal{D}_d$ data set.

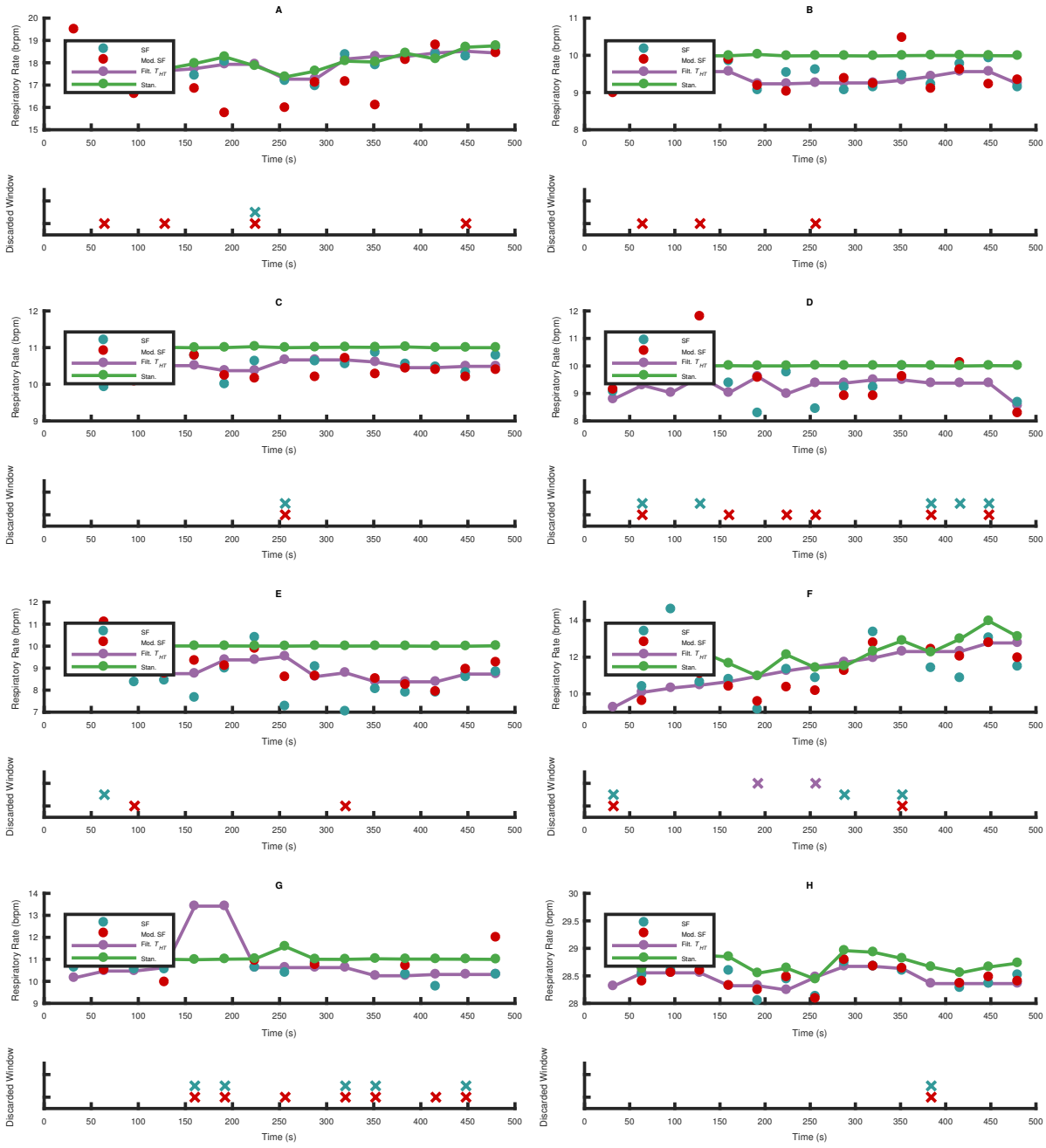


Figure A.6: *(Upper row) Time-series of RR estimates for modified SmartFusion, SmartFusion, the HT-based proposed method, and the R_{ref} reference values for subjects in data set $\mathcal{D}_{c,2}$ not shown in in figure 5: (A) subject 15, (B) subject 123, (C) subject 128, (D) subject 148, (E) subject 149, (F) subject 150, (G) subject 313, and (H) subject 322. (Lower row) Corresponding plots showing those windows for which each method did not produce an RR estimate.*



Figure A.7: *Time-series RR and HR waveforms for two hours of monitoring in subject 4 in the $\mathcal{D}_d$ data set. Upper panel is the RR waveform from SQI/RQIFusion$_{HT}$. Second panel (from top) shows the the specific modulations (PPG BW, PPG FM, PPG AM, ECG BW, ECG FM, and ECG AM) that are below the discard threshold for the SQI (red dots), RQI (blue dots), or both the SQI and RQI (lavender dots) for each window. Bottom row of panel shows if a window was dropped (all six modulations dropped). Third panel (from top) is the RR waveform from the RR Average implementation. The fourth panel (from top) is the RR reference waveform. The bottom panel is the HR estimation obtained from ECG. In all RR waveform panels the lightest coloured waveforms (light lavender, teal, and green) represent the cubic spline interpolations of the RR estimates, the darker coloured waveforms (lavender, teal, and green) represent the three-point median filter of the RR estimates. The darkest coloured points (dark lavender, teal, and green) represent the RR estimates obtained from median filtering. In the HR waveform the red line represents the HR estimates. The darkest grey bars, grey bars, and lightest grey bars represent RR or HR regions (depending on waveform) where the NEWS2 score is 3, 2, or 1, respectively. NEWS2 scores of 0 on the waveform plots are represented by white regions. The rose regions on all of RR waveforms represent critical events as detected by the RR reference as defined in Section 7.3.1. Yellow points in RR and HR waveforms represent all instances where the given waveform had an RR or HR estimate with a NEWS2 score exceeding 0.*

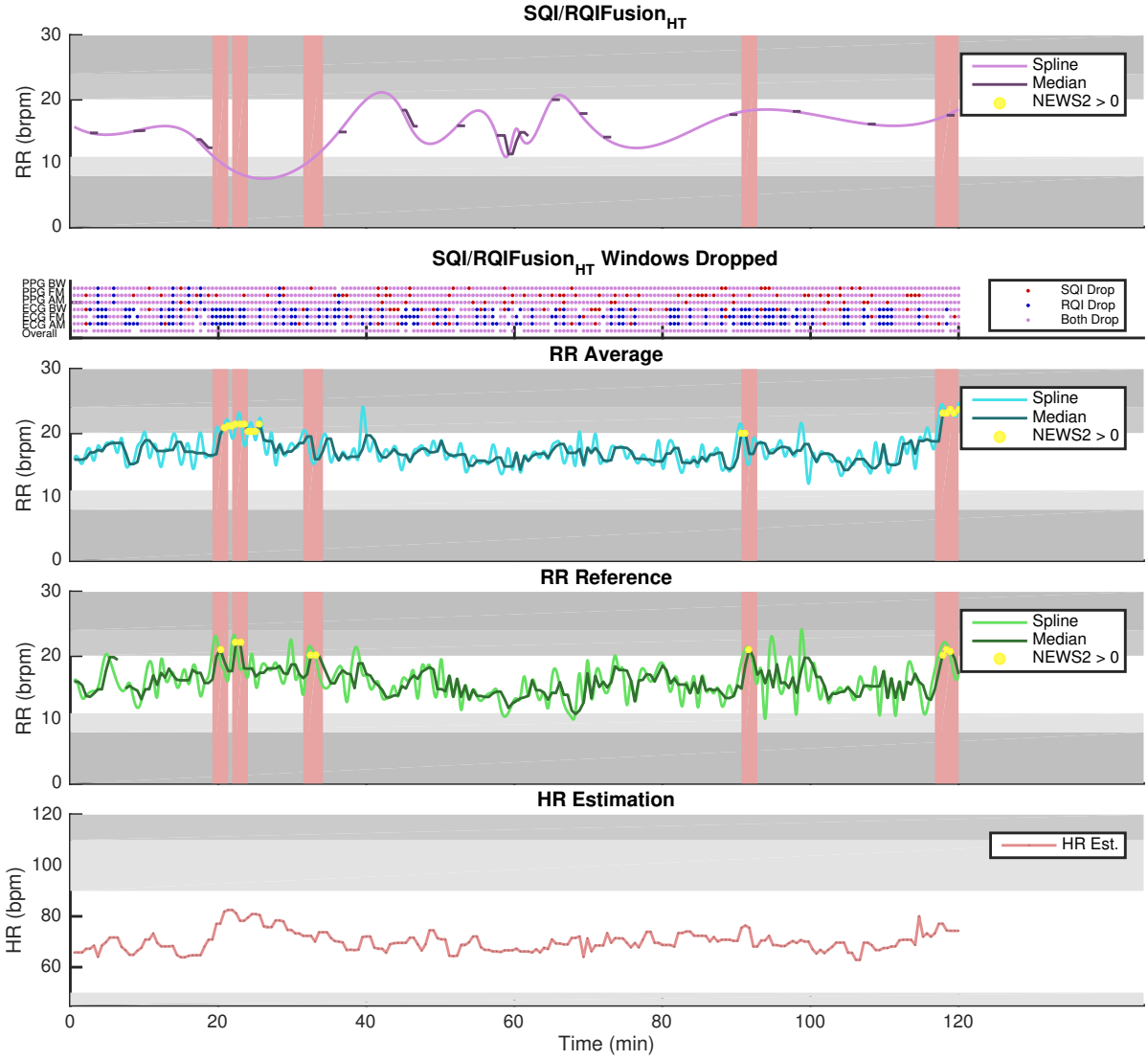


Figure A.8: *Time-series RR and HR waveforms for two hours of monitoring in subject 5 in the $\mathcal{D}_d$ data set. Upper panel is the RR waveform from SQI/RQIFusion$_{HT}$. Second panel (from top) shows the the specific modulations (PPG BW, PPG FM, PPG AM, ECG BW, ECG FM, and ECG AM) that are below the discard threshold for the SQI (red dots), RQI (blue dots), or both the SQI and RQI (lavender dots) for each window. Bottom row of panel shows if a window was dropped (all six modulations dropped). Third panel (from top) is the RR waveform from the RR Average implementation. The fourth panel (from top) is the RR reference waveform. The bottom panel is the HR estimation obtained from ECG. In all RR waveform panels the lightest coloured waveforms (light lavender, teal, and green) represent the cubic spline interpolations of the RR estimates, the darker coloured waveforms (lavender, teal, and green) represent the three-point median filter of the RR estimates. The darkest coloured points (dark lavender, teal, and green) represent the RR estimates obtained from median filtering. In the HR waveform the red line represents the HR estimates. The darkest grey bars, grey bars, and lightest grey bars represent RR or HR regions (depending on waveform) where the NEWS2 score is 3, 2, or 1, respectively. NEWS2 scores of 0 on the waveform plots are represented by white regions. The rose regions on all of RR waveforms represent critical events as detected by the RR reference as defined in Section 7.3.1. Yellow points in RR and HR waveforms represent all instances where the given waveform had an RR or HR estimate with a NEWS2 score exceeding 0.*



Figure A.9: *Time-series RR and HR waveforms for two hours of monitoring in subject 6 in the $\mathcal{D}_d$ data set. Upper panel is the RR waveform from SQI/RQIFusion$_{HT}$. Second panel (from top) shows the the specific modulations (PPG BW, PPG FM, PPG AM, ECG BW, ECG FM, and ECG AM) that are below the discard threshold for the SQI (red dots), RQI (blue dots), or both the SQI and RQI (lavender dots) for each window. Bottom row of panel shows if a window was dropped (all six modulations dropped). Third panel (from top) is the RR waveform from the RR Average implementation. The fourth panel (from top) is the RR reference waveform. The bottom panel is the HR estimation obtained from ECG. In all RR waveform panels the lightest coloured waveforms (light lavender, teal, and green) represent the cubic spline interpolations of the RR estimates, the darker coloured waveforms (lavender, teal, and green) represent the three-point median filter of the RR estimates. The darkest coloured points (dark lavender, teal, and green) represent the RR estimates obtained from median filtering. In the HR waveform the red line represents the HR estimates. The darkest grey bars, grey bars, and lightest grey bars represent RR or HR regions (depending on waveform) where the NEWS2 score is 3, 2, or 1, respectively. NEWS2 scores of 0 on the waveform plots are represented by white regions. The rose regions on all of RR waveforms represent critical events as detected by the RR reference as defined in Section 7.3.1. Yellow points in RR and HR waveforms represent all instances where the given waveform had an RR or HR estimate with a NEWS2 score exceeding 0.*

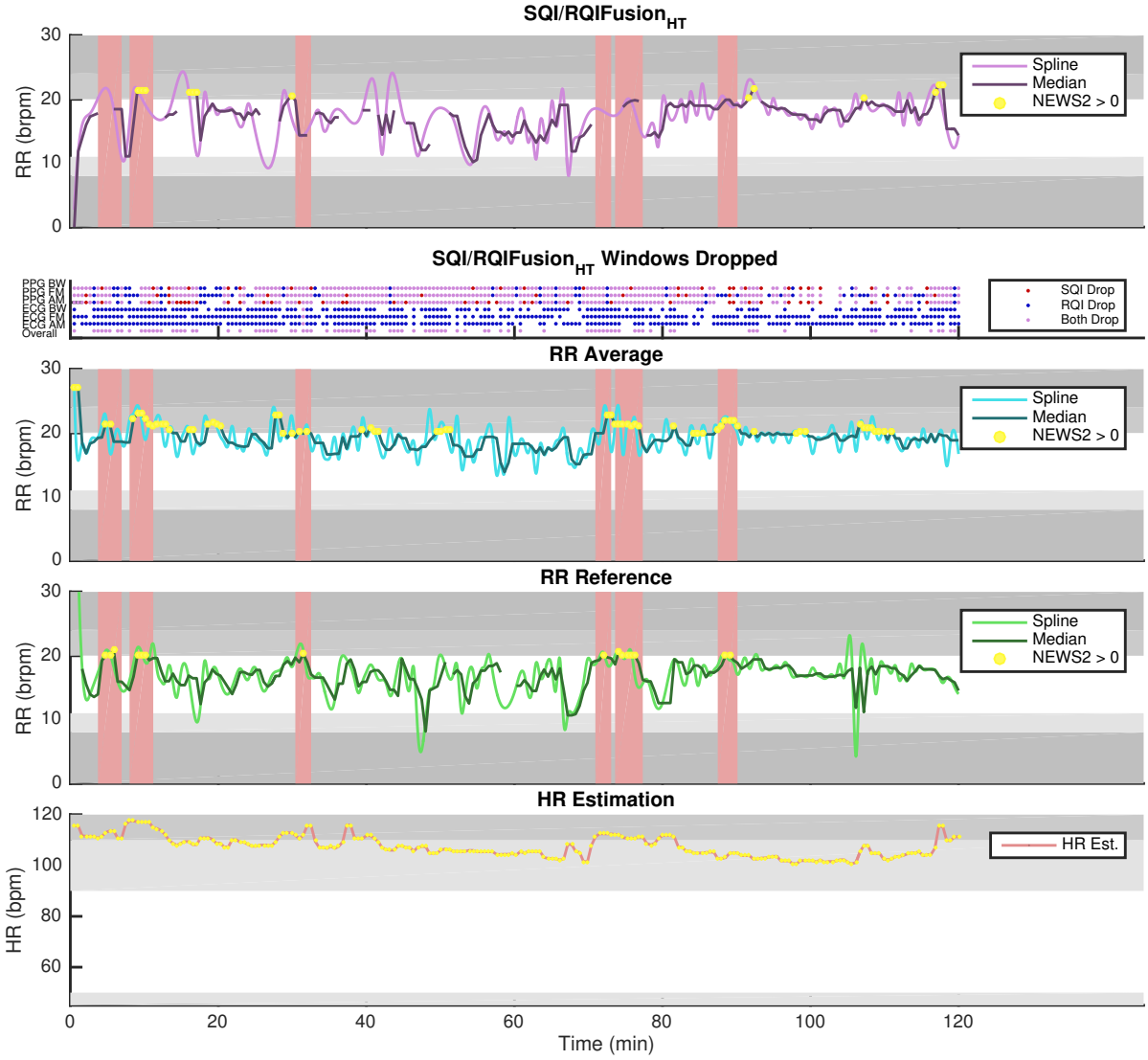


Figure A.10: *Time-series RR and HR waveforms for two hours of monitoring in subject 7 in the $\mathcal{D}_d$ data set. Upper panel is the RR waveform from SQI/RQIFusion$_{HT}$. Second panel (from top) shows the the specific modulations (PPG BW, PPG FM, PPG AM, ECG BW, ECG FM, and ECG AM) that are below the discard threshold for the SQI (red dots), RQI (blue dots), or both the SQI and RQI (lavender dots) for each window. Bottom row of panel shows if a window was dropped (all six modulations dropped). Third panel (from top) is the RR waveform from the RR Average implementation. The fourth panel (from top) is the RR reference waveform. The bottom panel is the HR estimation obtained from ECG. In all RR waveform panels the lightest coloured waveforms (light lavender, teal, and green) represent the cubic spline interpolations of the RR estimates, the darker coloured waveforms (lavender, teal, and green) represent the three-point median filter of the RR estimates. The darkest coloured points (dark lavender, teal, and green) represent the RR estimates obtained from median filtering. In the HR waveform the red line represents the HR estimates. The darkest grey bars, grey bars, and lightest grey bars represent RR or HR regions (depending on waveform) where the NEWS2 score is 3, 2, or 1, respectively. NEWS2 scores of 0 on the waveform plots are represented by white regions. The rose regions on all of RR waveforms represent critical events as detected by the RR reference as defined in Section 7.3.1. Yellow points in RR and HR waveforms represent all instances where the given waveform had an RR or HR estimate with a NEWS2 score exceeding 0.*

References

[1] T. Ahrens, "The most important vital signs are not being measured," *Australian Critical Care*, vol. 21, no. 1, pp. 3–5, 2008.

[2] D. L. Simel, "Approach to the Patient: History and Physical Examination," in *Goldman's Cecil Medicine*. L. Goldman and A. I. Schafer, Eds. Philadelphia, PA: Elsevier/Saunders, 24th ed., 2012, pp. 22-27.

[3] D. L. Schriger, "Approach to the Patient with Abnormal Vital Signs," in *Goldman's Cecil Medicine*. L. Goldman and A. I. Schafer, Eds. Philadelphia, PA: Elsevier/Saunders, 24th ed., 2012, pp. 27-30.

[4] R. M. Schein, N. Hazday, M. Pena, B. H. Ruben, and C. L. Sprung, "Clinical antecedents to in-hospital cardiopulmonary arrest," *Chest*, vol. 98, no. 6, pp. 1388–1392, 1990.

[5] C. Franklin and J. Mathew, "Developing strategies to prevent inhospital cardiac arrest: Analyzing responses of physicians and nurses in the hours before the event," *Critical Care Medicine*, vol. 22, no. 2, pp. 244–247, 1994.

[6] M. D. Buist, G. E. Moore, S. A. Bernard, B. P. Waxman, J. N. Anderson, and T. V. Nguyen, "Effects of a medical emergency team on reduction of incidence of and mortality from unexpected cardiac arrests in hospital: Preliminary study," *BMJ*, vol. 324, no. 7334, pp. 387–390, 2002.

[7] G. B. Smith, D. R. Prytherch, P. Schmidt, P. I. Featherstone, D. Knight, G. Clements, and M. A. Mohammed, "Hospital-wide physiological surveillance—A new approach to the early identification and management of the sick patient," *Resuscitation*, vol. 71, no. 1, pp. 19–28, 2006.

[8] M. Elliott and A. Coventry, "Critical care: The eight vital signs of patient monitoring," *British Journal of Nursing*, vol. 21, no. 10, pp. 621–625, 2012.

[9] R. J. M. Morgan, F. Williams, and M. M. Wright, "An early warning scoring system for detecting developing critical illness," *Clinical Intensive Care*, vol. 8, no. 2, p. 100, 1997.

[10] C. Stenhouse, S. Coates, M. Tivey, P. Allsop, and T. Parker, "Prospective evaluation of a modified early warning score to aid earlier detection of patients developing critical illness on a general surgical ward," *British Journal of Anaesthesia*, vol. 84, no. 5, p. 663, 2000.

[11] C. P. Subbe, R. G. Davies, E. Williams, P. Rutherford, and L. Gemmell, "Effect of introducing the modified early warning score on clinical outcomes, cardiopulmonary arrests and intensive care utilisation in acute medical admissions," *Anaesthesia*, vol. 58, no. 8, pp. 797–802, 2003.

[12] C. P. Subbe, M. Kruger, P. Rutherford, and L. Gemmel, "Validation of a modified early warning score in medical admissions," *QJM*, vol. 94, no. 10, pp. 521–526, 2001.

[13] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, "Respiratory rate: The neglected vital sign," *Medical Journal of Australia*, vol. 188, no. 11, pp. 657–659, 2008.

[14] J. F. Fieselmann, M. S. Hendryx, C. M. Helms, and D. S. Wakefield, "Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients," *Journal of General Internal Medicine*, vol. 8, no. 7, pp. 354–360, 1993.

[15] D. R. Goldhill, "Preventing surgical deaths: Critical care and intensive care outreach services in the postoperative period," *British Journal of Anaesthesia*, vol. 95, no. 1, pp. 88–94, 2005.

[16] C. H. Van Leuvan and I. Mitchell, "Missed opportunities? An observational study of vital sign measurements," *Critical Care Resuscitation*, vol. 10, no. 2, pp. 111–115, 2008.

[17] J. Hogan, "Why don't nurses monitor the respiratory rates of patients?," *British Journal of Nursing*, vol. 15, no. 9, pp. 489–492, 2006.

[18] T. J. Hodgetts, G. Kenward, I. G. Vlachonikolis, S. Payne, and N. Castle, "The identification of risk factors for cardiac arrest and formulation of activation criteria to alert a medical emergency team," *Resuscitation*, vol. 54, no. 2, pp. 125–131, 2002.

[19] C. Butler-Williams, N. Cantrill, and S. Maton, "Increasing staff awareness of respiratory rate significance," *Resuscitation*, vol. 62, no. 2, pp. 137–141, 2005.

[20] S. E. Bedell, T. L. Delbanco, E. F. Cook, and F. H. Epstein, "Survival after cardiopulmonary resuscitation in the hospital," *New England Journal of Medicine*, vol. 309, no. 10, pp. 569–576, 1983.

[21] P. Jevon, "How to ensure patient observations lead to prompt identification of tachypnoea," *Nursing Times*, vol. 106, no. 2, pp. 12–14, 2010.

[22] *Advanced Life Support Manual*. London, UK: Resuscitation Council (UK), 7th ed., 2016.

[23] K. van Loon, B. van Zaane, E. J. Bosch, C. J. Kalkman, and L. M. Peelen, "Non-invasive continuous respiratory monitoring on general hospital wards: A systematic review," *PLOS One*, vol. 10, no. 12, p. e0144626, 2015.

[24] D. R. Goldhill, S. A. White, and A. Sumner, "Physiological values and procedures in the 24 h before ICU admission from the ward," *Anaesthesia*, vol. 54, no. 6, pp. 529–534, 1999.

[25] F. McGain, M. A. Cretikos, D. Jones, S. Van Dyk, M. D. Buist, H. Opdam, V. Pellegrino, M. S. Robertson, and R. Bellomo, "Documentation of clinical review and vital signs after major surgery," *Medical Journal of Australia*, vol. 189, no. 7, pp. 380–383, 2008.

[26] K. Konno and J. Mead, "Measurement of the separate volume changes of rib cage and abdomen during breathing," *Journal of Applied Physiology*, vol. 22, no. 3, pp. 407–422, 1967.

[27] J. J. Freundlich and J. C. Erickson, "Electrical impedance pneumography for simple nonrestrictive continuous monitoring of respiratory rate, rhythm and tidal volume for surgical patients," *Chest*, vol. 65, no. 2, pp. 181–184, 1974.

[28] R. Ortega, C. Connor, S. Kim, R. Djang, and K. Patel, "Monitoring ventilation with capnography," *New England Journal of Medicine*, vol. 367, no. 19, p. e27, 2012.

[29] M. K. Marks, M. South, and B. G. Carter, "Measurement of respiratory rate and timing using a nasal thermocouple," *Journal of Clinical Monitoring*, vol. 11, no. 3, pp. 159–164, 1995.

[30] L. Bernstein, J. L. D'Silva, and D. Mendel, "The effect of the rate of breathing on the maximum breathing capacity determined with a new spirometer," *Thorax*, vol. 7, no. 3, pp. 255–262, 1952.

[31] K. Miyasaka, Y. Kondo, T. Suzuki, H. Sakai, and M. Takata, "Toward better home respiratory monitoring: A comparison of impedance and inductance pneumography," *Pediatrics International*, vol. 36, no. 3, pp. 307–310, 1994.

[32] A. H. Taenzer, J. B. Pyke, S. P. McGrath, and G. T. Blike, "Impact of pulse oximetry surveillance on rescue events and intensive care unit transfers: A before-and-after concurrence study," *Anesthesiology*, vol. 112, no. 2, pp. 282–287, 2010.

[33] S. Nemati, A. Malhotra, and G. D. Clifford, "Data fusion for improved respiration rate estimation," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, p. 926305, 2010.

[34] I. M. Cheifetz and N. R. MacIntyre, "Respiratory controversies in the critical care setting. Conference summary," *Respiratory Care*, vol. 52, no. 5, pp. 636–644, 2007.

[35] H. Ugnell, L. G. Lindberg, and P. A. Oberg, "Monitoring of heart- and respiration rates using a fiber optic sensor," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 13, pp. 1566–1567, 1991.

[36] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, "Derivation of respiratory signals from multi-lead ECGs," in *the Proceedings of Computers in Cardiology (CinC)*, vol. 12, pp. 113–116, 1985.

[37] K. V. Madhav, M. R. Ram, E. H. Krishna, K. N. Reddy, and K. A. Reddy, "Extraction of respiratory activity from PPG and BP signals using principal component analysis," in *the Proceedings of the 2011 International Conference on Communications and Signal Processing*, pp. 452–456.

[38] E. Helfenbein, C. S. Chien, S. Babaeizadeh, P. Dasari, A. Könik, J. M. Mukherjee, M. King, and L. Shao, "ECG/EMG-derived respiration during cardiac magnetic resonance imaging," *Journal of Electrocardiology*, vol. 45, no. 6, pp. 693–694, 2012.

[39] P. H. Charlton, D. A. Birrenkott, T. Bonnici, M. A. F. Pimentel, A. E. Johnson, J. Alastruey, L. Tarassenko, P. J. Watkinson, R. Beale, and D. A. Clifton, "Respiratory rate estimation from the electrocardiogram and photoplethysmogram: A review," *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 2–20, 2018.

[40] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, pp. R1–R39, 2007.

[41] A. Jubran, "Pulse oximetry," *Critical Care*, vol. 3, no. 2, pp. R11–R17, 1999.

[42] A. Jubran, "Pulse oximetry," *Critical Care*, vol. 19, no. 1, p. 272, 2015.

[43] K. E. Sandau, M. Funk, A. Auerbach, G. W. Barsness, K. Blum, M. Cvach, R. Lampert, J. L. May, G. M. McDaniel, M. V. Perez, S. Sendelbach, C. E. Sommargren, and P. J. Wang, "Update to practice standards for electrocardiographic monitoring in hospital settings: A scientific statement from the american heart association," *Circulation*, vol. 136, pp. e273–e344, 2017.

[44] P. H. Charlton, T. Bonnici, L. Tarassenko, J. Alastruey, D. A. Clifton, R. Beale, and P. J. Watkinson, "Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: Technical and physiological determinants," *Physiological Measurement*, vol. 38, no. 5, pp. 669–690, 2017.

[45] J. Li, J. Jin, X. Chen, W. Sun, and P. Guo, "Comparison of respiratory-induced variations in photoplethysmographic signals," *Physiological Measurement*, vol. 31, no. 3, pp. 415–425, 2010.

[46] F. Yasuma and J. Hayano, "Respiratory sinus arrhythmia: Why does the heartbeat synchronize with respiratory rhythm?," *Chest*, vol. 125, no. 2, pp. 683–690, 2004.

[47] N. Gambarotta, F. Aletti, G. Baselli, and M. Ferrario, "A review of methods for the signal quality assessment to improve reliability of heart rate and blood pressures derived parameters," *Medical & Biological Engineering & Computing*, vol. 54, no. 7, pp. 1025–1035, 2016.

[48] 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," *Physiological Measurement*, vol. 29, no. 1, pp. 15–32, 2008.

[49] M. A. F. Pimentel, M. D. Santos, D. B. Springer, and G. D. Clifford, "Heart beat detection in multimodal physiological data using a hidden semi-Markov model and signal quality indices," *Physiological Measurement*, vol. 36, no. 8, pp. 1717–1727, 2015.

[50] G. D. Clifford, J. Behar, Q. Li, and I. Rezek, "Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms," *Physiol Meas*, vol. 33, no. 9, pp. 1419–33, 2012.

[51] W. Karlen, M. Turner, E. Cooke, G. Dumont, and J. M. Ansermino, "Capnobase: Signal database and tools to collect, share and annotate respiratory signals," in *the Proceedings of the Annual Meeting of the Society for Technology in Anesthesia (STA)*, p. 25, 2010.

[52] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "Physiobank, Physiotoolkit, and Physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–E220, 2000.

[53] A. E. W. Johnson, T. J. Pollard, L. Shen, L. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovitz, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, p. 160035, 2016.

[54] T. J. Pollard and A. E. W. Johnson, "The MIMIC-III Clinical Database," 2016. [Online]. Available: https://physionet.org/physiobank/database/mimic3cdb/. [Accessed on: 9 January 2019].

[55] C. MacEwen, "Can data fusion techniques predict adverse physiological events during haemodialysis," DPhil thesis, University of Oxford, 2016. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:1ef92d5d-920d-4ff4-b368-5e892527e675. [Accessed on: 6 January 2019].

[56] M. Thiriet, *Anatomy and Physiology of the Circulatory and Ventilatory Systems*. New York, NY: Springer, 2014.

[57] D. J. Meredith, D. A. Clifton, P. J. Charlton, J. Brooks, C. W. Pugh, and L. Tarassenko, "Photoplethysmographic derivation of respiratory rate: A review of relevant physiology," *Journal of Medical Engineering & Technology*, vol. 36, no. 1, pp. 1–7, 2012.

[58] J. Blom, *Monitoring of Respiration and Circulation*. Boca Raton, FL: CRC Press, 2004.

[59] A. J. Buda, M. R. Pinsky, N. B. Ingels, G. T. Daughters, E. B. Stinson, and E. L. Alderman, "Effect of intrathoracic pressure on left ventricular performance," *New England Journal of Medicine*, vol. 301, no. 9, pp. 453–459, 1979.

[60] K. McLeish, D. L. G. Hill, D. Atkinson, J. M. Blackall, and R. Razavi, "A study of the motion and deformation of the heart due to respiration," *IEEE Transactions on Medical Imaging*, vol. 21, no. 9, pp. 1142–1150, 2002.

[61] K. Toska and M. Eriksen, "Respiration-synchronous fluctuations in stroke volume, heart rate and arterial pressure in humans," *Journal of Physiology*, vol. 472, pp. 501–512, 1993.

[62] M. Piepoli, P. Sleight, S. Leuzzi, F. Valle, G. Spadacini, C. Passino, J. Johnston, and L. Bernardi, "Origin of respiratory sinus arrhythmia in conscious humans. An important role for arterial carotid baroreceptors," *Circulation*, vol. 95, no. 7, pp. 1813–1821, 1997.

[63] M. Nitzan, I. Faib, and H. Friedman, "Respiration-induced changes in tissue blood volume distal to occluded artery, measured by photoplethysmography," *Journal Biomedical Optics*, vol. 11, no. 4, p. 040506, 2006.

[64] K. Davidson and V. Barber, "Electronic monitoring of patients in general wards," *Nursing Standard*, vol. 18, no. 49, pp. 42–46, 2004.

[65] J. G. Webster, Ed., *Medical Instrumentation: Application and Design*. Hoboken, NJ: John Wiley & Sons Inc., 4th ed., 2010.

[66] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter respiratory rate estimation from the photoplethysmogram," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1946–1953, 2013.

[67] J. G. Webster, Ed., *Bioinstrumentation*. Hoboken, NJ: John Wiley & Sons Inc., 2004.

[68] J. L. Woomer and D. A. Berkheimer, "Using capnography to monitor ventilation," *Nursing*, vol. 33, no. 4, pp. 42–43, 2003.

[69] D. Wood, "Capnography: principles, applications and uses in trauma and emergencies," MPhil thesis, University of Cape Town, 2004. [Online]. Available: https://open.uct.ac.za/bitstream/handle/11427/2878/thesis_hsf_2004_wood_d.pdf?sequence=1. [Accessed on: 6 January 2019].

[70] A. B. Hertzman, "Photoelectric plethysmography of the fingers and toes in man," *Experimental Biology and Medicine*, vol. 37, no. 3, pp. 529–534, 1937.

[71] T. Aoyagi, K. M, Y. K, and W. S, "Improvement of the earpiece oximeter," in *Abstracts of the 13th annual meeting of the Japanese Society of Medical Electronics and Biological Engineering*, pp. 90–91, 1974.

[72] A. A. Kamshilin and N. B. Margaryants, "Origin of photoplethysmographic waveform at green light," *Physics Procedia*, vol. 86, pp. 72–80, 2017.

[73] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and autoregressive models," *Physiological Measurement*, vol. 35, pp. 807–831, 2014.

[74] T. Wu, V. Blazek, and H. H. Schmitt, "Photoplethysmography imaging: A new noninvasive and noncontact method for mapping of the dermal perfusion changes," in *Procedings SPIE 4163, Optical Techniques and Instrumentation for the Measurement of Blood Composition, Structure, and Dynamics*, pp. 62–70, 2000.

[75] L. Nilsson, A. Johansson, and S. Kalman, "Monitoring of respiratory rate in postoperative care using a new photoplethysmographic technique," *Journal of Clinical Monitoring and Computing*, vol. 16, no. 4, pp. 309–315, 2000.

[76] K. Shelley and S. Shelley, "Pulse Oximeter Waveform: Photoelectric Plethysmography," in *Clinical Monitoring: Practical Applications for Anesthesia and Critical Care*. C. L. Lake, R. L. Hines, and C. D. Blitt, Eds. Philadelphia PA: Elsevier/Saunders, 2001, pp. 420-428.

[77] Y. Maeda, M. Sekine, T. Tamura, A. Moriya, T. Suzuki, and K. Kameyama, "Comparison of reflected green light and infrared photoplethysmography," *the Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2008, pp. 2270–2272, 2008.

[78] J. H. van Oostrom, B. Fuehrlein, and R. J. Melker, "Pulse Oximetry and Plethysmography," in *Civetta, Taylor and Kirby's Critical Care*. A. Gabrielli, A. J. Layon, and M. Yu, Eds. Philadelphia, PA: Lippincott Williams & Wilkins, 4th ed., 2008, pp. 217-222.

[79] A. Reisner, P. A. Shaltis, D. McCombie, and H. H. Asada, "Utility of the photoplethysmogram in circulatory monitoring," *Anesthesiology*, vol. 108, no. 5, pp. 950–958, 2008.

[80] A. B. Hertzman, "The blood supply of various skin areas as estimated by the photoelectric plethysmograph," *American Journal of Physiology*, vol. 124, no. 2, pp. 328–340, 1938.

[81] J. C. Dorlas and J. A. Nijboer, "Photo-electric plethysmography as a monitoring device in anaesthesia. Application and interpretation," *British Journal of Anaesthesia*, vol. 57, no. 5, pp. 524–530, 1985.

[82] K. H. Shelley, D. H. Jablonka, A. A. Awad, R. G. Stout, H. Rezkanna, and D. G. Silverman, "What is the best site for measuring the effect of ventilation on the pulse oximeter waveform?," *Anesthesia & Analgesia*, vol. 103, no. 2, pp. 372–377, 2006.

[83] E. M., "On the analysis of fingertip photoplethysmogram signals," *Current Cardiology Reviews*, vol. 8, no. 1, pp. 14–25, 2012.

[84] M. Nirmalan and P. Dark, "Broader applications of arterial pressure wave form analysis," *Continuing Education in Anaesthesia, Critical Care & Pain*, vol. 14, no. 6, pp. 285–290, 2015.

[85] S. Millasseau, F. Guigui, R. Kelly, K. Prasad, J. Cockcroft, and J. Ritter, "Noninvasive assessment of the digital volume pulse. comparison with the peripheral pressure pulse," *Hypertension*, vol. 36, no. 6, pp. 952–956, 2000.

[86] K. H. Shelley, "Photoplethysmography: Beyond the calculation of arterial oxygen saturation and heart rate," *Anesthesia & Analgesia*, vol. 105, no. 6, pp. S31–S36, 2007.

[87] N. Townsend, "Pulse Oximetry," Medical Electronics, University of Oxford, 2001. [Online]. Available: https://www.robots.ox.ac.uk/~neil/teaching/lectures/med_elec/notes6.pdf. [Accessed on: 20 February 2019].

[88] S. Prahl, "Tabulated molar extinction coefficient for hemoglobin in water," 1998. [Online]. Available: http://omlc.org/spectra/hemoglobin/summary.html. [Accessed on: 23 May 2017].

[89] "Principles of pulse oximetry," AnaesthesiaUK, 2004. [Online]. Available: http://www.frca.co.uk/article.aspx?articleid=332. [Accessed on: 20 February 2019].

[90] J. Lee, K. Matsumura, K. Yamakoshi, P. Rolfe, S. Tanaka, and T. Yamakoshi, "Comparison between red, green and blue light reflection photoplethysmograph for heart rate monitoring during motion," in *the Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1724–1727.

[91] S. T. Lawless, "Crying wolf: False alarms in a pediatric intensive care unit," *Critical Care Medicine*, vol. 22, no. 6, pp. 981–985, 1994.

[92] J. T. Moller, T. Pedersen, L. S. Rasmussen, P. F. Jensen, B. D. Pedersen, O. Ravlo, N. H. Rasmussen, K. Espersen, N. W. Johannessen, J. B. Cooper, J. S. Gravenstein, B. Chraemmer-Jørgensen, F. Wiberg-Jørgensen, M. Djernes, L. Heslet, and S. H. Johansen, "Randomized evaluation of pulse oximetry in 20,802 patients: I. Design, demography, pulse oximetry failure rate, and overall complication rate," *Anesthesiology*, vol. 78, no. 3, pp. 436–444, 1993.

[93] D. L. Reich, A. Timcenko, C. A. Bodian, J. Kraidin, J. Hofman, M. DePerio, S. N. Konstadt, T. Kurki, and J. B. Eisenkraft, "Predictors of pulse oximetry data failure," *Anesthesiology*, vol. 84, no. 4, pp. 859–864, 1996.

[94] W. B. Runciman, R. K. Webb, L. Barker, and M. Currie, "The australian incident monitoring study. The pulse oximeter: Applications and limitations–An analysis of 2000 incident reports," *Anaesthesia and Intensive Care*, vol. 21, no. 5, pp. 543–550, 1993.

[95] N. S. Trivedi, A. F. Ghouri, N. K. Shah, E. Lai, and S. J. Barker, "Effects of motion, ambient light, and hypoperfusion on pulse oximeter function," *Journal of Clinical Anesthesia*, vol. 9, no. 3, pp. 179–183, 1997.

[96] M. A. F. Pimentel, E. W. Johnson, P. H. Charlton, D. A. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton, "Towards a robust estimation of respiratory rate from pulse oximeters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, 2017.

[97] C. Orphanidou, T. Bonnici, P. Charlton, D. A. Clifton, D. Vallance, and L. Tarassenko, "Signal quality indices for the electrocardiogram and phtoplethysmogram: Derivation and applications to wireless monitoring," *IEEE Journal of Biomedical Health Informatics*, vol. 19, no. 3, pp. 832–838, 2014.

[98] C. Ahlstrom, A. Johansson, T. Lanne, and P. Ask, "A respiration monitor based on electrocardiographic and photoplethysmographic sensor fusion," in *the Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2311–2314, 2004.

[99] M. Leier, G. Jervan, and W. Stork, "Respiration signal extraction from photoplethysmogram using pulse wave amplitude variation," in *the Proceedings of the 2014 IEEE International Conference on Communications (ICC)*, pp. 3535–3540.

[100] W. B. Murray and P. A. Foster, "The peripheral pulse wave: Information overlooked," *Journal of Clinical Monitoring*, vol. 12, no. 5, pp. 365–377, 1996.

[101] A. Fusco, D. Locatelli, F. Onorati, G. C. Durelli, and M. D. Santambrogio, "On how to extract breathing rate from PPG signal using wearable devices," in *the Proceedings of the 2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*.

[102] L. Nilsson, T. Goscinski, S. Kalman, L. G. Lindberg, and A. Johansson, "Combined photoplethysmographic monitoring of respiration rate and pulse: A comparison between different measurement sites in spontaneously breathing subjects," *Acta Anaesthesiologica Scandinavica*, vol. 51, no. 9, pp. 1250–1257, 2007.

[103] M. Hickey, J. P. Phillips, and P. A. Kyriacou, "The effect of vascular changes on the photoplethysmographic signal at different hand elevations," *Physiological Measurement*, vol. 36, no. 3, pp. 425–440, 2015.

[104] A. D. Waller, "A demonstration on man of electromotive changes accompanying the heart's beat," *Journal Physiology*, vol. 8, no. 5, pp. 229–234, 1887.

[105] W. Einthoven, "Un nouveau galvanometer," *Arch. Neerl. Sc. Ex. Nat.*, no. 6, pp. 625–633, 1901.

[106] W. Einthoven, G. Fahr, and A. de Waart, "On the direction and manifest size of the variations of potential in the human heart and on the influence of the position of the heart on the form of the electrocardiogram," *American Heart Journal*, vol. 40, no. 2, pp. 163–211, 1950.

[107] F. N. Wilson, C. E. Kossmann, E. Burch, George, E. Golberger, A. Graybiel, H. H. Hecht, F. D. Johnston, E. Lepeschkin, and G. B. Myers, "Recommendations for standardization of electrocardiographic and vectorcardiographic leads," *Circulation*, vol. 10, no. 4, pp. 564–573, 1954.

[108] F. G. Cosío, J. Palacios, A. Pastor, and A. Núñez, "The Electrocardiogram," in *The ESC Textbook of Cardiovascular Medicine*. J. A. Camm, T. F. Lüscher, and P. W. Serruys, Eds. Oxford UK: Oxford University Press, 2nd ed., 2009.

[109] M. Al Ghatrif and J. Lindsay, "A brief review: History to understand fundamentals of electrocardiography," *Journal of Community Hospital Internal Medicine Perspectives*, vol. 2, no. 1, p. 14383, 2012.

[110] C. Orphanidou, S. Fleming, S. A. Shah, and L. Tarassenko, "Data fusion for estimating respiratory rate from a single-lead ECG," *Biomedical Signal Processing and Control*, vol. 8, no. 1, pp. 98–105, 2013.

[111] A. Travaglini, C. Lamberti, J. DeBie, and M. Ferri, "Respiratory signal derived from eight-lead ECG," in *the Proceedings from Computers in Cardiology (CinC)*, vol. 25, pp. 65–68, 1998.

[112] W. J. Yi and K. S. Park, "Derivation of respiration from ECG measured without subject's awareness using wavelet transform," in *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society*, pp. 130–131, 2002.

[113] N. A. Nayan, N. S. Risman, and R. Jaafar, "Breathing rate estimation from a single-lead electrocardiogram acquisition system," *International Journal of Applied Engineering Research*, vol. 10, no. 17, pp. 38154–38158, 2015.

[114] B. F. Womack, "The analysis of respiratory sinus arrhythmia using spectral analysis and digital filtering," *IEEE Transactions on Biomedical Engineering*, vol. 18, no. 6, pp. 399–409, 1971.

[115] R. Pallas-Areny, J. Colominas-Balague, and F. J. Rosell, "The effect of respiration-induced heart movements on the ECG," *IEEE Transactions on Biomedical Engineering*, vol. 36, no. 6, pp. 585–590, 1989.

[116] M. Varanini, M. Emdin, F. Allegri, M. Raciti, F. Conforti, A. Macerata, A. Taddei, R. Francesconi, G. Kraft, and A. L. Abbate, "Adaptive filtering of ECG signal for deriving respiratory activity," in *the Proceedings of Computers in Cardiology (CinC)*, vol. 17, pp. 621–624, 1990.

[117] S. Ding, X. Zhu, W. Chen, and D. Wei, "Derivation of respiratory signal from single-channel ECGs based on source statistics," *International Journal of Bioelectromagnetism*, vol. 6, no. 2, pp. 43–49, 2004.

[118] B. Mazzanti, C. Lamberti, and J. de Bie, "Validation of an ECG-derived respiration monitoring method," in *the Proceedings of Computers in Cardiology (CinC)*, vol. 30, pp. 613–616, 2003.

[119] S. Meek and F. Morris, "Introduction. I—Leads, rate, rhythm, and cardiac axis," *BMJ*, vol. 324, no. 7334, pp. 415–418, 2002.

[120] N. Iyengar, C. K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *American Journal of Physiology*, vol. 271, no. 4, pp. R1078–R1084, 1996.

[121] Y. Ichimaru and G. B. Moody, "Development of the polysomnographic database on CD-ROM," *Psychiatry and Clinical Neurosciences*, vol. 53, no. 2, pp. 175–177, 1999.

[122] J. Francis, "ECG monitoring leads and special leads," *Indian Pacing Electrophysiology Journal*, vol. 16, no. 3, pp. 92–95, 2016.

[123] J. Lázaro, E. Gil, R. Bailón, and P. Laguna, "Deriving respiration from the pulse photoplethysmographic signal," in *the Proceedings of Computing in Cardiology (CinC)*, vol. 38, pp. 713–716, 2011.

[124] D. Wu, P. Yang, G. Z. Liu, and Y. T. Zhang, "Automatic estimation of respiratory rate from pulse transit time in normal subjects at rest," in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, pp. 779–781.

[125] J. Lázaro, E. Gil, R. Bailón, A. Mincholé, and P. Laguna, "Deriving respiration from photoplethysmographic pulse width," *Medical & Biological Engineering & Computing*, vol. 51, no. 1-2, pp. 233–242, 2013.

[126] G. D. Furman, Z. Shinar, A. Baharav, and S. Akselrod, "Electrocardiogram derived respiration during sleep," in *the Proceedings of Computers in Cardiology (CinC)*, vol. 32, pp. 351–354, 2005.

[127] S. B. Park, Y. S. Noh, S. J. Park, and H. R. Yoon, "An improved algorithm for respiration signal extraction from electrocardiogram measured by conductive textile electrodes using instantaneous frequency estimation," *Medical & Biological Engineering & Computing*, vol. 46, no. 2, pp. 147–158, 2008.

[128] P. Langley, E. J. Bowers, and A. Murray, "Principal component analysis as a tool for analyzing beat-to-beat changes in ECG features: Application to ECG-derived respiration," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 821–829, 2010.

[129] A. Sobron, I. Romero, and T. Lopetegi, "Evaluation of methods for estimation of respiratory frequency from the ECG," in *the Proceedings of Computing in Cardiology (CinC)*, vol. 37, pp. 513–516, 2010.

[130] L. S. Correa, E. Laciar, A. Torres, and R. Jane, "Performance evaluation of three methods for respiratory signal estimation from the electrocardiogram," in *the Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4760–4763, 2008.

[131] E. Helfenbein, S. Babaeizadeh, J. Lindauer, and S. Zhou, "Electrocardiogram/electromyogram-derived respiration in the presence of simulated respiratory disease," *Journal of Electrocardiology*, vol. 44, no. 6, pp. 751–752, 2011.

[132] E. Helfenbein, R. Firoozabadi, S. Chien, E. Carlson, and S. Babaeizadeh, "Development of three methods for extracting respiration from the surface ECG: A review," *Journal of Electrocardiology*, vol. 47, no. 6, pp. 819–825, 2014.

[133] M. Orini, M. D. Pelaez-Coca, R. Bailón, and E. Gil, "Estimation of spontaneous respiratory rate from photoplethysmography by cross time-frequency analysis," in *the Proceedings of Computing in Cardiology (CinC)*, vol. 38, pp. 661–664, 2011.

[134] M. D. Pelez-Coca, M. Orini, J. Lázaro, R. Bailón, and E. Gil, "Cross time-frequency analysis for combining information of several sources: Application to estimation of spontaneous respiratory rate from photoplethysmography," *Computational and Mathematical Methods in Medicine*, vol. 2013, p. 631978, 2013.

[135] J. Lázaro, Y. Nam, E. Gil, P. Laguna, and K. H. Chon, "Respiratory rate derived from smartphone-camera-acquired pulse photoplethysmographic signals," *Physiological Measurement*, vol. 36, no. 11, pp. 2317–2333, 2015.

[136] S. P. Arunachalam and L. F. Brown, "Real-time estimation of the ECG-derived respiration (EDR) signal using a new algorithm for baseline wander noise removal," in *the Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5681–5684.

[137] R. Ruangsuwana, G. Velikic, and M. Bocko, "Methods to extract respiration information from ECG signals," in *the Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 570–573.

[138] M. Campolo, D. Labate, F. La Foresta, F. Morabito, A. Lay-Ekuakille, and P. Vergallo, "ECG-derived respiratory signal using empirical mode decomposition," in *the Proceedings of the 2011 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 399–403.

[139] D. Widjaja, C. Varon, A. C. Dorado, J. A. K. Suykens, and S. Van Huffel, "Application of kernel principal component analysis for single-lead-ECG-derived respiration," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 1169–1176, 2012.

[140] A. Shayei, S. P. Ehsani, and M. Shabany, "Efficient implementation of real-time ECG derived respiration system using cubic spline interpolation," in *the Proceedings of the 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, pp. 1083–1086.

[141] H. H. Kuo, N. H. Peng, T. Kao, and W. C. Hu, "Using ECG surface electrodes in measurement of respiration rate for preterm infants," in *the Proceedings of the 2014 IEEE International Symposium on Bioelectronics and Bioinformatics (IEEE ISBB 2014)*.

[142] C. Orphanidou, O. Brain, J. Feldmar, S. Khan, J. Price, and L. Tarassenko, "Spectral fusion for estimating respiratory rate from the ECG," in *the Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine*, 2009.

[143] A. M. Chan, N. Ferdosi, and R. Narasimhan, "Ambulatory respiratory rate detection using ecg and a triaxial accelerometer," in *the Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4058–4061, 2013.

[144] W. Karlen, C. J. Brouse, E. Cooke, J. M. Ansermino, and G. A. Dumont, "Respiratory rate estimation using respiratory sinus arrhythmia from photoplethysmography," in *in the Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1201–1204.

[145] A. Schäfer and K. W. Kratky, "Estimation of breathing rate from respiratory sinus arrhythmia: Comparison of various methods," *Annals of Biomedical Engineering*, vol. 36, no. 3, pp. 476–485, 2008.

[146] S. Babaeizadeh, S. H. Zhou, S. D. Pittman, and D. P. White, "Electrocardiogram-derived respiration in screening of sleep-disordered breathing," *Journal of Electrocardiology*, vol. 44, no. 6, pp. 700–706, 2011.

[147] W. Karlen, A. Garde, D. Myers, C. Scheffer, J. M. Ansermino, and G. A. Dumont, "Estimation of respiratory rate from photoplethysmographic imaging videos compared to pulse oximetry," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1331–1338, 2015.

[148] J. Lázaro, Y. Nam, E. Gil, P. Laguna, and K. H. Chon, "Smartphone-camera-acquired pulse photoplethysmographic signal for deriving respiratory rate," in *the Proceedings of the 8th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO)*, pp. 121–122, 2014.

[149] L. Mirmohamadsadeghi and J. Vesin, "Real-time multi-signal frequency tracking with a bank of notch filters to estimate the respiratory rate from the ECG," *Physiological Measurement*, vol. 37, no. 9, pp. 1573–1587, 2016.

[150] L. Mirmohamadsadeghi and J. M. Vesin, "Estimating the real-time respiratory rate from the ECG with a bank of notch filters," in *the Proceedings of Computing in Cardiology (CinC)*, vol. 42, pp. 581–584, 2015.

[151] T. Zhu, M. A. F. Pimentel, G. D. Clifford, and D. A. Clifton, "Bayesian fusion of algorithms for the robust estimation of respiratory rate from the photoplethysmogram," in *the Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6138–6141, 2015.

[152] A. Kikta and P. Augustyniak, "Comparing methods of ECG respiration signals derivation based on measuring the amplitude of QRS complexes," *Journal of Medical Informatics & Technologies*, vol. 11, pp. 155–164, 2007.

[153] R. Firoozabadi, E. D. Helfenbein, and S. Babaeizadeh, "Monitoring respiration rate in sleep-disordered breathing patients using chest belts or ECG," *Journal of Electrocardiology*, vol. 47, no. 6, p. 908, 2014.

[154] L. Mirmohamadsadeghi and J. Vesin, "Respiratory rate estimation from the ECG using an instantaneous frequency tracking algorithm," *Biomedical Signal Processing and Control*, vol. 14, pp. 66–72, 2014.

[155] S. Sarkar, S. Bhattacherjee, and S. Pal, "Extraction of respiration signal from ECG for respiratory rate estimation," in *the Proceedings of the Michael Faraday IET International Summit 2015*, pp. 336–340.

[156] K. T. Sweeney, D. Kearney, T. E. Ward, S. Coyle, and D. Diamond, "Employing ensemble empirical mode decomposition for artifact removal: Extracting accurate respiration rates from ECG data during ambulatory activity," in *the Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 977–980, 2013.

[157] O. Meste, G. Blain, and S. Bermon, "Analysis of the respiratory and cardiac systems coupling in pyramidal exercise using a time-varying model," in *the Proceedings of Computers in Cardiology (CinC)*, vol. 29, pp. 429–432, 2002.

[158] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," in *the Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995, 1998.

[159] K. V. Madhav, M. R. Ram, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "Estimation of respiration rate from ECG, BP and PPG signals using empirical mode decomposition," in *the Proceedings of the 2011 IEEE International Instrumentation and Measurement Technology Conference.*

[160] V. Kumarand and G. Singh, "Estimation of respiration rate from ECG using canonical components analysis and ensemble empirical mode decomposition," *International Journal of Bio-Science and Bio-Technology*, vol. 7, no. 3, pp. 139–146, 2015.

[161] K. V. Madhav, M. Raghuram, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "Use of multi scale PCA for extraction of respiratory activity from photoplethysmographic signals," in *the 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, pp. 1784–1787.

[162] A. Johansson, P. A. Oberg, and G. Sedin, "Monitoring of heart and respiratory rates in newborn infants using a new photoplethysmographic technique," *Journal of Clinical Monitoring and Computing*, vol. 15, no. 7-8, pp. 461–467, 1999.

[163] S. Fleming, L. Tarassenko, M. Thompson, and D. Mant, "Non-invasive measurement of respiratory rate in children using the photoplethysmogram," in *the Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1886–1889, 2008.

[164] S. A. Shah, S. Fleming, M. Thompson, and L. Tarassenko, "Respiratory rate estimation during triage of children in hospitals," *Journal of Medical Engineering & Technology*, vol. 39, no. 8, pp. 514–524, 2015.

[165] D. Wertheim, C. Olden, E. Savage, and P. Seddon, "Extracting respiratory data from pulse oximeter plethysmogram traces in newborn infants," *Archives of Disease in Childhood—Fetal and Neonatal Edition*, vol. 94, no. 4, pp. F301–F303, 2009.

[166] K. Nakajima, T. Tamura, T. Ohta, H. Miike, and P. A. Oberg, "Photoplethysmographic measurement of heart and respiratory rates using digital filters," in *the Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1006–1007, 1993.

[167] L. Nilsson, A. Johansson, and S. Kalman, "Respiratory variations in the reflection mode photoplethysmographic signal. Relationships to peripheral venous pressure," *Medical & Biological Engineering & Computing*, vol. 41, no. 3, pp. 249–254, 2003.

[168] H. Ugnell, "The respiratory synchronous photoplethysmographic signal. Its dependence on light wavelength and sample volume," *Medical & Biological Engineering & Computing*, vol. 34, pp. 275–276, 1996.

[169] A. Canu, M. Canu, S. Marinkovic, S. Faul, and E. Popovici, "Respiration rate calculation using low power DSP processor and SpO_2 sensor," in *the Proceedings of the 2011 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 517–520.

[170] S. S. Chreiteh, B. Belhage, K. Hoppe, J. Branebjerg, R. Haahr, S. Duun, and E. V. Thomsen, "Estimation of respiratory rates based on photoplethysmographic measurements at the sternum," in *the Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6570–6573, 2015.

[171] K. Nakajima, T. Tamura, and H. Miike, "Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique," *Medcal Engineering & Physics*, vol. 18, no. 5, pp. 365–372, 1996.

[172] H. Sharma, K. K. Sharma, and O. L. Bhagat, "Respiratory rate extraction from single-lead ECG using homomorphic filtering," *Computers in Biology and Medicine*, vol. 59, pp. 80–86, 2015.

[173] P. A. Leonard, D. A. Clifton, P. S. Addison, J. N. Watson, and T. Beattie, "An automated algorithm for determining respiratory rate by photoplethysmogram in children," *Acta Paediatrica*, vol. 95, no. 9, pp. 1124–1128, 2006.

[174] P. S. Addison and J. N. Watson, "Secondary wavelet feature decoupling (SWFD) and its use in detecting patient respiration from the photoplethysmogram," in *the*

Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2602–2605, 2003.

[175] D. A. Clifton, J. G. Douglas, P. S. Addison, and J. N. Watson, "Measurement of respiratory rate from the photoplethysmogram in chest clinic patients," *Journal Clinical Monitoring and Computing*, vol. 21, no. 1, pp. 55–61, 2007.

[176] P. Leonard, N. R. Grubb, P. S. Addison, D. A. Clifton, and J. N. Watson, "An algorithm for the detection of individual breaths from the pulse oximeter waveform," *Journal Clinical Monitoring and Computing*, vol. 18, no. 5-6, pp. 309–312, 2004.

[177] P. A. Leonard, J. G. Douglas, N. R. Grubb, D. A. Clifton, P. S. Addison, and J. N. Watson, "A fully automated algorithm for the determination of respiratory rate from the photoplethysmogram," *Journal of Clinical Monitoring and Computing*, vol. 20, no. 1, pp. 33–36, 2006.

[178] A. E. Santo and C. Carbajal, "Respiration rate extraction from ECG signal via discrete wavelet transform," in *the Proceedings of the 2nd Circuits and Systems for Medical and Environmental Applications Workshop (CASME)*, 2010.

[179] K. H. Chon, S. Dash, and K. Ju, "Estimation of respiratory rate from photoplethysmogram data using time-frequency spectral estimation," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2054–2063, 2009.

[180] S. Dash, K. H. Shelley, D. G. Silverman, and K. H. Chon, "Estimation of respiratory rate from ECG, photoplethysmogram, and piezoelectric pulse transducer signals: A comparative study of timefrequency methods," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 5, pp. 1099–1107, 2010.

[181] Q. V. Nguyen, R. Le Page, J. Goujon, L. Poffo, and M. Thual, "Comparison of two methods for demodulation of pulse signals—Application in case of central sleep apnea," *Journal of Science and Technology*, vol. 49, no. 1, pp. 0866–708X, 2011.

[182] S. G. Fleming and L. Tarassenko, "A comparison of signal processing techniques for the extraction of breathing rate from the photoplethysmogram," in *Proceedings of the World Academy of Science: Engineering & Technology*, vol. 24, p. 276, 2007.

[183] J. Lee, J. P. Florian, and K. H. Chon, "Respiratory rate extraction from pulse oximeter and electrocardiographic recordings," *Physiological Measurement*, vol. 32, no. 11, pp. 1763–1773, 2011.

[184] C. L. Mason and L. Tarassenko, "Quantitative assessment of respiratory derivation algorithms," in *the Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1998–2001, 2001.

[185] D. Cysarz, R. Zerm, H. Bettermann, M. Frühwirth, M. Moser, and M. Kröz, "Comparison of respiratory rates derived from heart rate variability, ECG amplitude, and nasal/oral airflow," *Annals of Biomedical Engineering*, vol. 36, no. 12, pp. 2085–2094, 2008.

[186] R. A. Cernat, S. I. Ciorecan, C. Ungureanu, J. Arends, R. Strungaru, and G. M. Ungureanu, "Recording system and data fusion algorithm for enhancing the estimation of the respiratory rate from photoplethysmogram," in *the Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5977–5980, 2005.

[187] Y. Yoshida, K. Yokoyama, and N. Ishii, "Real-time continuous estimation of respiratory frequency during sleep based on heart rate time series," in *the Proceedings*

of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 648–651, 2007.

[188] J. Y. Foo and S. J. Wilson, "Estimation of breathing interval from the photoplethysmographic signals in children," *Physiological Measurement*, vol. 26, no. 6, pp. 1049–1058, 2005.

[189] A. Johansson, "Neural network for photoplethysmographic respiratory rate monitoring," *Medical & Biological Engineering & Computing*, vol. 41, no. 3, pp. 242–248, 2003.

[190] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "Probabilistic estimation of respiratory rate using Gaussian processes," in *the Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2902–2905, 2013.

[191] J. Lee and K. H. Chon, "Respiratory rate extraction via an autoregressive model using the optimal parameter search criterion," *Annals of Biomedical Engineering*, vol. 38, no. 10, pp. 3218–3225, 2010.

[192] J. F. Thayer, J. J. Sollers, E. Ruiz-Padial, and J. Vila, "Estimating respiratory frequency from autoregressive spectral analysis of heart period," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 4, pp. 41–45, 2002.

[193] J. Lee and K. H. Chon, "An autoregressive model-based particle filtering algorithms for extraction of respiratory rates as high as 90 breaths per minute from pulse oximeter," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 9, pp. 2158–2167, 2010.

[194] P. S. Addison, J. N. Watson, M. L. Mestek, and R. S. Mecca, "Developing an algorithm for pulse oximetry derived respiratory rate (RR_{oxi}): A healthy volunteer study," *Journal of Clinical Monitoring and Computing*, vol. 26, no. 1, pp. 45–51, 2012.

[195] W. Karlen, A. Garde, D. Myers, C. Scheffer, J. M. Ansermino, and G. A. Dumont, "Respiratory rate assessment from photoplethysmographic imaging," in *the Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5397–5400, 2014.

[196] K. H. Shelley, A. A. Awad, R. G. Stout, and D. G. Silverman, "The use of joint time frequency analysis to quantify the effect of ventilation on the pulse oximeter waveform," *Journal of clinical monitoring and computing*, vol. 20, no. 2, pp. 81–87, 2006.

[197] J. Lee and K. H. Chon, "Time-varying autoregressive model-based multiple modes particle filtering algorithm for respiratory rate extraction from pulse oximeter," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 790–794, 2011.

[198] A. Garde, W. Karlen, J. M. Ansermino, and G. A. Dumont, "Estimating respiratory and heart rates from the correntropy spectral density of the photoplethysmogram," *PLoS One*, vol. 9, no. 1, p. e86427, 2014.

[199] C. Wang, Z. Li, and X. Wei, "Monitoring heart and respiratory rates at radial artery based on PPG," *Optik*, vol. 124, no. 19, pp. 3954–3956, 2013.

[200] N. Daimiwal, M. Sundhararajan, and R. Shriram, "Respiratory rate, heart rate and continuous measurement of BP using PPG," in *the Proceedings of the 2014 International Conference on Communication and Signal Processing*, pp. 999–1002.

[201] T. Kagawa, A. Kawamoto, and N. Nakajima, "System for simultaneous measurement of breathing rate and heart rate using photoplethysmogram," in *the Proceedings of the 8th International Conference on Body Area Networks*, pp. 519–525, 2013.

[202] D. Li, H. Zhao, and S. Dou, "A new signal decomposition to estimate breathing rate and heart rate from photoplethysmography signal," *Biomedical Signal Processing and Control*, vol. 19, pp. 89–95, 2015.

[203] B. Yang and K. H. Chon, "A novel approach to monitor nonstationary dynamics in physiological signals: Application to blood pressure, pulse oximeter, and respiratory data," *Annals of Biomedical Engineering*, vol. 38, no. 11, pp. 3478–3488, 2010.

[204] R. G. Lyons, *Understanding Digital Signal Processing*. Upper Saddle River, NJ: Prentice Hall, 3rd ed., 2011.

[205] R. Takalo, H. Hytti, and H. Ihalainen, "Tutorial on univariate autoregressive spectral analysis," *Journal Clinical Monitoring and Computing*, vol. 19, no. 6, pp. 401–410, 2005.

[206] D. Rowell, "2.161 Signal Processing: Continuous and Discrete," MIT OpenCourseWare, 2008. [Online]. Available: https://ocw.mit.edu/courses/mechanical-engineering/2-161-signal-processing-continuous-and-discrete-fall-2008/lecture-notes/lecture_21.pdf. [Accessed on: 2 August 2017].

[207] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ: John Wiley & Sons Inc., 4th ed., 2008.

[208] J. R. Buck, M. M. Daniel, and C. Singer, Andrew, *Comptuter Exploration in Signals and Systems Using MATLAB*. Upper Saddle River, NJ: Prentice Hall, 2nd ed., 2002.

[209] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.

[210] D. B. Springer, "Mobile phone-based rheumatic heart disease detection," DPhil thesis, University of Oxford, 2015. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:5ec8c818-dafb-4571-8198-97607f8d0451. [Accessed on: 6 January 2019].

[211] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230–236, 1985.

[212] C. K. Lao, U. K. Che, W. Chen, S. H. Pun, P. U. Mak, F. Wan, and M. I. Vai, "Portable heart rate detector based on photoplethysmography with android programmable devices for ubiquitous health monitoring system," *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, vol. 2, no. 1, pp. 18–26, 2012.

[213] M. L. Samuels, J. A. Witmer, and A. A. Schaffner, *Statistics for the Life Sciences*. Upper Saddle River, NJ: Prentice Hall, 4th ed., 2012.

[214] B. L. Welch, "The generalization of 'student's' problem when several different population variances are involved," *Biometrika*, vol. 34, no. 1-2, pp. 28–35, 1947.

[215] D. Chemla, J. Young, F. Badilini, P. Maison-Blanche, H. Affres, Y. Lecarpentier, and P. Chanson, "Comparison of fast Fourier transform and autoregressive spectral analysis for the study of heart rate variability in diabetic patients," *International Journal of Cardiology*, vol. 104, no. 3, pp. 307–313, 2005.

[216] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.

[217] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY: Springer, 1995.

[218] D. Mattera and S. Haykin, "Support Vector Machines for Dynamic Reconstruction of a Chaotic System," in *Advances in Kernel Methods: Support Vector Learning*. C. J. C. Burges, B. Schölkopf, and A. J. Smola, Eds. Boston, MA: MIT Press, 1997, pp. 211-242.

[219] A. J. Smola, N. Murata, B. Schölkopf, and K. R. Müller, "Asymptotically optimal choice of ϵ-loss for support vector machines," in *the Proceedings of the International Conference on Artificial Neural Networks (ICANN)*, pp. 105–110, 1998.

[220] J. T. Kwok, "Linear dependency between ϵ and the input noise in ϵ-support vector regression," in *the Proceedings of the International Conference on Artificial Neural Networks (ICANN)*, pp. 405–410, 2001.

[221] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," National Taiwan University, 2007. [Online]. Available: https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf. [Accessed on: 6 January 2019].

[222] H. T. Lin and C. J. Lin, "A Study on Sigmoid Kernels for SVM and the Training of non-PSD Kernels by SMO-type Methods," National Taiwan University, 2003. [Online]. Available: http://www.work.caltech.edu/~htlin/publication/doc/tanh.pdf. [Accessed on: 6 January 2019].

[223] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," National Taiwan University, 2003. [Online]. Available: https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf. [Accessed on: 6 January 2019].

[224] N. Gallagher and G. Wise, "Theoretical analysis of the properties of the median filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, 1982.

[225] G. R. Arce, *Nonlinear Signal Processing: A Statistical Approach*. Hoboken, NJ: John Wiley & Sons Inc., 2005.

[226] T. Young and M. J. Mohlenkamp, *Introduction to Numerical Methods and MATLAB Programming for Engineers*. Athens, OH: Department of Mathematics, Ohio University, 2014.

[227] J. F. Epperson, "On the Runge example," *The American Mathematical Monthly*, vol. 94, no. 4, pp. 329–341, 1987.

[228] R. H. Bartels, J. C. Beatty, and B. Barsky, *An Introduction to Splines for Use in Computer Graphics and Geometric Modeling*. San Mateo, CA: Morgan Kaufmann Publishers, Inc., 1987.

[229] J. M. de Carvalho and J. V. Hanson, "Real-time interpolation with cubic splines and polyphase networks," *Canadian Electrical Engineering Journal*, vol. 11, no. 2, pp. 64–72, 1986.

[230] O. Guven, A. Eftekhar, W. Kindt, and T. Constandinou, "Computationally efficient real-time interpolation algorithm for non-uniform sampled biosignals," *Healthcare Technology Letters*, vol. 3, no. 2, pp. 105–110, 2016.

[231] C. Orphanidou, T. Bonnici, P. Charlton, D. A. Clifton, D. Vallance, and L. Tarassenko, "Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 832–838, 2015.

[232] S. Fleming, "Measurement and fusion on non-invasive vital signs for routine triage of acute paediatric illness," DPhil thesis, University of Oxford, 2010. [Online]. Available: http://www.ibme.ox.ac.uk/research/biomedical-signal-processing-instrumentation/prof-l-tarassenko/publications/pdf/s_fleming_thesis.pdf. [Accessed on: 6 January 2019].

[233] R. C. o. Physicians, *National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party.* London: RCP, 2017.

[234] S. Sendelbach and M. Funk, "Alarm fatigue: A patient safety concern," *AACN Advanced Critical Care*, vol. 24, no. 4, pp. 378–386, 2013.

[235] D. R. Prytherch, G. B. Smith, P. E. Schmidt, and P. I. Featherstone, "ViEWS—Towards a national early warning score for detecting adult inpatient deterioration," *Resuscitation*, vol. 81, no. 8, pp. 932–937, 2010.

[236] P. J. Watkinson, M. A. F. Pimentel, D. A. Clifton, and L. Tarassenko, "Manual centile-based early warning scores derived from statistical distributions of observational vital-sign data," *Resuscitation*, vol. 129, pp. 55–60, 2018.

[237] U. Nwulu, D. Westwood, D. Edwards, F. Kelliher, and J. J. Coleman, "Adoption of an electronic observation chart with an integrated early warning scoring system on pilot wards: A descriptive study," *CIN: Computers, Informatics, Nursing*, vol. 30, no. 7, pp. 371–379, 2012.

[238] C. Downey, W. Tahir, R. Randell, J. M. Brown, and D. G. Jayne, "Strengths and limitations of early warning scores: A systematic review and narrative synthesis," *International Journal of Nursing Studies*, vol. 76, pp. 106–119, 2017.

[239] L. Tarassenko, D. A. Clifton, M. R. Pinsky, M. T. Hravnak, J. R. Woods, and P. J. Watkinson, "Centile-based early warning scores derived from statistical distributions of vital signs," *Resuscitation*, vol. 82, no. 8, pp. 1013–1018, 2011.

[240] T. Bonnici, L. Tarassenko, D. A. Clifton, and P. Watkinson, "The digital patient," *Clinical Medicine*, vol. 13, no. 3, pp. 252–257, 2013.

[241] M. A. F. Pimentel, "Modelling of vital-sign data from post-operative patients," DPhil thesis, University of Oxford, 2015. [Online]. Available: http://www.robots.ox.ac.uk/~davidc/pubs/thesisMP.pdf. [Accessed on: 6 January 2019].

[242] A. Kapoor, H. Ahn, and P. R. W, "Mixture of gaussian processes for combining multiple modalities," in *the Proceedings of the International Workshop on Multiple Classifier Systems (MCS)*, pp. 86–96, 2005.

[243] H. Xiao, H. Xiao, and C. Eckert, "Learning from multiple observers with unknown expertise," in *the Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pp. 595–606, 2013.

[244] D. Dias and J. Cunha, "Wearable health devices-vital sign monitoring, systems and technologies," *Sensors*, vol. 18, no. 8, pp. 2414–2432, 2018.

[245] F. Seoane, I. Mohino-Herranz, J. Ferreira, L. Alvarez, R. Buendia, D. Ayllon, C. Llerena, and R. Gil-Pita, "Wearable biomedical measurement systems for assessment of mental status of combatants in real time," *Sensors*, vol. 14, no. 4, pp. 7120–7141, 2014.

[246] T. Yilmaz, R. Foster, and Y. Hao, "Detecting vital signs with wearable wireless sensors," *Sensors*, vol. 10, no. 12, pp. 10837–10862, 2010.

[247] M. Chan, D. Esteve, J. Fourniols, C. Escriba, and E. Campo, "Smart wearable systems: Current status and future challenges," *Artificial Intelligence in Medicine*, vol. 56, no. 3, pp. 137–156, 2012.

[248] S. P. Lee, G. K. Ha, D. Wright, Y. Ma, E. Sen-Gupta, N. R. Haubrich, P. C. Branche, W. Li, G. L. Huppert, M. Johnson, H. B. Mutlu, K. Li, N. Sheth, J. A. Wright, Y. Huang, M. Mansour, J. A. Rogers, and R. Ghaffari, "Highly flexible, wearable, and disposable cardiac biosensors for remote and ambulatory monitoring," *NPJ Digital Medicine*, vol. 1, pp. 1–8, 2017.

[249] I. Kim, P. H. Lai, R. Lobo, and B. Gluckman, "Challenges in wearable personal health monitoring systems," in *the Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5264–5267, 2014.

[250] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable photoplethysmographic sensors-past and present," *Electronics*, vol. 3, no. 2, pp. 282–302, 2014.

[251] S. M. A. Salehizadeh, "Motion and noise artifact detection and vital sign reconstruction in ECG/PPG based wearable devices," PhD dissertation, University of Connecticut. [Online]. Available: https://opencommons.uconn.edu/cgi/viewcontent.cgi?article=7207&context=dissertations. [Accessed on: 6 January 2019].