

Multi-task Gaussian process models for biomedical applications

Robert Dürichen^{1,2}, Marco A. F. Pimentel³, Lei Clifton³, Achim Schweikard¹, and David A. Clifton³

Abstract—Gaussian process (GP) models are a flexible means of performing non-parametric Bayesian regression. However, the majority of existing work using GP models in healthcare data is defined for univariate output time-series, denoted as *single-task GPs* (STGP). Here, we investigate how GPs could be used to model multiple correlated univariate physiological time-series simultaneously. The resulting *multi-task GP* (MTGP) framework can learn the correlation within multiple signals even though they might be sampled at different frequencies and have training sets available for different intervals. We illustrate the basic properties of MTGPs using a synthetic case-study with respiratory motion data. Finally, two real-world biomedical problems are investigated from the field of patient monitoring and motion compensation in radiotherapy. The results are compared to STGPs and other standard methods in the respective fields. In both cases, MTGPs learned the correlation between physiological time-series efficiently, which leads to improved modelling accuracy.

I. INTRODUCTION

The Gaussian process (GP) framework is a useful non-parametric Bayesian regression tool for machine learning problems [1]. Compared to other regression techniques, such as support vector regression, GP models have the advantage that prior knowledge of the functional behaviour (including periodicity or smoothness) can easily be integrated.

GPs have been used for various regression tasks concerning physiological data. In [2], GPs were used to design a vital-sign “early warning system” and in [9] relevance vector machines were used to compensate for respiratory motion in robotic radiotherapy. Additionally, the hyperparameters of a trained GP can be used to estimate physiological parameters such as the respiratory rate [4]. However, GPs are typically used to predict a single output time-series (“task”) based on one or multiple inputs. Here, we explore the use of multi-task GP (MTGP) models for modelling multiple univariate physiological time-series. By modelling multiple correlated tasks simultaneously, the aim is to learn the correlation between and within the tasks, and thereby to improve overall modelling accuracy compared to single task GPs (STGP). In the field of geostatistics, this approach is known as *co-kriging* [5]. In recent years, MTGPs received

more interest within the machine learning community and have been applied to multi-label text categorisation [6], speed-up prediction for programming tasks [7], robot inverse dynamics [8], and environmental sensor networks [10].

In this paper, we present a novel general framework to model physiological data using MTGPs. We evaluate our approach using synthetic data, and real scenarios with different biomedical datasets. To stimulate further biomedical research, we have implemented an open-source MTGP toolbox, which is available on-line¹. The toolbox is an extension to the GPML toolbox [1], which is one of the most frequently-used toolboxes for performing analysis with GPs.

II. METHODS

A. Single-Task Gaussian Process models

In this section, we provide a brief introduction to STGP models; a more detailed description can be found in [1]. Let $\mathbf{x}_n = \{x_i \mid i = 1, \dots, n\}$ be our training data and $\mathbf{y}_n = \{y_i \mid i = 1, \dots, n\}$ the training labels with $x_i, y_i \in \mathbb{R}$. The aim is to learn a regression model $y = f(x) + \varepsilon$, where $f(x)$ represents a latent function and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ a noise term. The function f can be interpreted as being a probability distribution over functions,

$$\mathbf{y}_n = f(\mathbf{x}_n) \sim \mathcal{GP}(m(\mathbf{x}_n), k(\mathbf{x}_n, \mathbf{x}'_n)) \quad (1)$$

where $m(\mathbf{x}_n)$ is the mean function of the process and $k(\mathbf{x}_n, \mathbf{x}'_n)$ is a covariance function which describes the coupling between two values of \mathbf{x}_n . Given \mathbf{x}_n and \mathbf{y}_n , predictions can be made for the unknown “test” data \mathbf{x}_* (Fig. 1(a)) by computing the conditional distribution $p(y_* | \mathbf{x}_*, \mathbf{x}_n, \mathbf{y}_n)$ which will be a Gaussian distribution,

$$p(y_* | \mathbf{x}_*, \mathbf{x}_n, \mathbf{y}_n) \sim \mathcal{N}(\bar{y}_*, \text{var}[y_*]) \quad (2)$$

where the mean and variance are given by, assuming the mean function m to be zero,

$$\bar{y}_* = k(\mathbf{x}_n, \mathbf{x}_*)^\top k(\mathbf{x}_n, \mathbf{x}_n)^{-1} \mathbf{y}_n \quad (3)$$

$$\text{var}[y_*] = k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathbf{x}_n, \mathbf{x}_*)^\top k(\mathbf{x}_n, \mathbf{x}_n)^{-1} k(\mathbf{x}_n, \mathbf{x}_*) \quad (4)$$

The covariance function encodes our prior knowledge concerning the functional behaviour of the time series that we wish to model. As shown in [1], there exists a large class of covariance functions which could be used. Frequently-used examples are the squared-exponential (SE), periodic (PER) and quasi-periodic (QP) covariance functions:

$$k_{SE}(r) = \theta_S^2 \exp\left\{-\frac{r^2}{2\theta_L^2}\right\}, \quad (5)$$

*This work was partially funded by the Graduate School for Computing in Medicine and Life Sciences, German Excellence Initiative [DFG GSC 235/1]. DAC was supported by a Royal Academy of Engineering Research Fellowship.

¹ Institute for Robotics and Cognitive Systems, Department of Computer Science, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany duerichen@rob.uni-luebeck.de

² Graduate School for Computing in Medicine and Life Sciences, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

³ Institute of Biomedical Engineering, Department of Engineering Science, Old Road Campus Research Building, University of Oxford, Roosevelt Drive, Oxford, OX3 7DQ, UK

¹<http://www.robots.ox.ac.uk/~davidc>

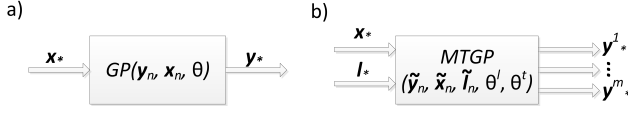


Fig. 1. Box diagram of a single input single output GP (a) and a multi-task GP (b)

$$k_{PER}(r) = \theta_S^2 \exp \left\{ -\frac{\sin^2[(2\pi/\theta_P)r]}{2} \right\}, \quad (6)$$

$$k_{QP}(r) = \theta_S^2 \exp \left\{ -\frac{r^2}{2\theta_L^2} \right\} \times \exp \left\{ -\frac{\sin^2[(2\pi/\theta_P)r]}{2} \right\}, \quad (7)$$

where θ_S , θ_L and θ_P are hyperparameters modelling the y-scaling, x-scaling (or time-scale if the data are time-series), and period of the covariance functions, respectively, and where $r = \|x - x'\|$ denotes the Euclidean distance between two data. The unknown hyperparameters can be optimised by minimising the negative log marginal likelihood NLML = $-\log p(\mathbf{y}_n|\mathbf{x}_n)$. The NLML was optimized using a gradient descent method implemented in the GPML toolbox [1].

B. Multi-task Gaussian Process models

For the MTGP model, we combine the analysis of m tasks within a single GP model. Let $\tilde{\mathbf{x}}_n = \{\mathbf{x}_{n_i}^i | i = 1, \dots, m\}$ and $\tilde{\mathbf{y}}_n = \{\mathbf{y}_{n_i}^i | i = 1, \dots, m\}$ be the vector of training data and labels for the m tasks, where the i^{th} task has n_i data. We hereafter assume that the common independent feature \mathbf{x} corresponds to time for all tasks. To specify that data $\mathbf{x}_{n_i}^i$ and labels $\mathbf{y}_{n_i}^i$ belong to task i , a labels vector $\tilde{\mathbf{l}}_n = \{\mathbf{l}_{n_i}^i | i = 1, \dots, m\}$ with $\mathbf{l}_{n_i}^i = i$ has to be added as additional input as illustrated in Fig. 1(b).

To model the correlation between tasks as well as the temporal behaviour of the tasks within a single GP, two independent covariance functions can be assumed,

$$k_{MTGP}(x, x', l, l') = k^c(l, l') \times k^t(x, x') \quad (8)$$

where k^c and k^t represent the correlation and temporal covariance functions between $\{x, l\}$ and $\{x', l'\}$, respectively. Note that k^t depends only on (x, x') and that k^c depends only on (l, l') . The complete covariance matrix for the training data \mathbf{K}_{MTGP} is

$$\mathbf{K}_{MTGP}(\tilde{\mathbf{x}}_n, \tilde{\mathbf{l}}_n, \theta^c, \theta^t) = \mathbf{K}_n^c(\tilde{\mathbf{l}}_n, \theta^c) \otimes \mathbf{K}_n^t(\tilde{\mathbf{x}}_n, \theta^t) \quad (9)$$

where \otimes is the Kronecker product, \mathbf{K}_n^t has a size of $m \times m$, and where \mathbf{K}_n^c and \mathbf{K}_{MTGP-n} have a size of $\tilde{n} \times \tilde{n}$ with $\tilde{n} = \sum_{i=1}^m n_i$. Here, θ^c and θ^t are two vectors containing all hyperparameters for \mathbf{K}^c and \mathbf{K}^t . This approach is also known as the *intrinsic correlation model* [5].

The remaining challenge is to construct a valid positive semidefinite correlation covariance function \mathbf{K}^c . One solution is to use the Cholesky decomposition and to parametrize the elements of the lower triangular matrix [7],

$$\mathbf{K}^c(l, l') = LL^\top, \quad L = \begin{bmatrix} \theta_1^c & 0 & \dots & 0 \\ \theta_2^c & \theta_3^c & & 0 \\ \vdots & & \ddots & \vdots \\ \theta_{k-m+2}^c & \theta_{k-m+2}^c & \dots & \theta_k^c \end{bmatrix} \quad (10)$$

where $k = \sum_{i=1}^m n_i$ is the number of correlation hyperparameters. Similar to the case with STGPs, the hyperparameters can be optimised by maximising the marginal likelihood, and predictions for test data $\{x_*, l_*\}$ can be made by computing the conditional probability $p(y_* | x_*, l_*, \tilde{\mathbf{x}}_n, \tilde{\mathbf{l}}_n, \theta^t)$.

This framework has several useful properties which we emphasise:

- we may have task-specific training data n_i (i.e., training data may be sampled at different frequencies for different tasks);
- automatic learning of the correlation within tasks occurs by maximising the marginal likelihood;
- we may have task-specific prediction labels y_* ; and
- the framework assumes that the tasks have similar temporal characteristics and hyperparameters θ^t .

III. RESULTS AND DISCUSSION

In this section, we illustrate the potential of MTGPs on a synthetic dataset and two real-world biomedical case studies.

A. Synthetic dataset

We investigate an example with $m = 4$ tasks. The dataset comprises time-series from three optical markers (OM) which are placed along the median line of one subject, at the chest (OM1), at the lower end of the sternum (OM2), and next to the navel (OM3) [9]. Additionally, a respiration belt (RB) was placed around the torso next to OM2. All signals were scaled to $[0, 1]$ and are shown in Fig. 2(a). The subject breathed normally throughout the period of data acquisition. The training data and sample frequency of each task are listed in Tab. I. We note that RB has a sampling frequency that is five times lower than that of OM1, OM2, and OM3. In this example, the objective is to predict y_*^{OM1} within a test range of $x_*^{OM1} = [20, 70)$. We assume that the temporal behaviour of each task can be modelled by a quasi-periodic function, $k^t = k_{QP}$, (7). The table also shows that Pearson's correlation coefficient for all tasks with respect to OM1 indicates a highly positive or negative correlation. The correlation covariance matrix \mathbf{K}^c is initialised assuming independent tasks, by setting it to be the identity matrix. The MTGP model has to learn the correlation during training by optimising the marginal likelihood. We investigate four evaluation scenarios (S1-4). In the first scenario (S1), only the time-series from OM1 is considered. This case is equivalent to a STGP. In the scenarios S2-4, the time-series for OM2, OM3 and RB are

TABLE I
TRAINING DATA n_i AND SAMPLE FREQUENCY f_s FOR OPTICAL MARKERS (OM) 1-3 AND RESPIRATION BELT (RB), ALONG WITH PEARSON'S CORRELATION COEFFICIENT r_{OM1} WITH RESPECT TO OM1

	OM1	OM2	OM3	RB
n_i [s]	(0, 20)	(15, 30)	(25, 40)	(0, 60)
f_s [Hz]	2.6	2.6	2.6	0.52
r_{OM1}	1	-0.96	-0.9	0.88

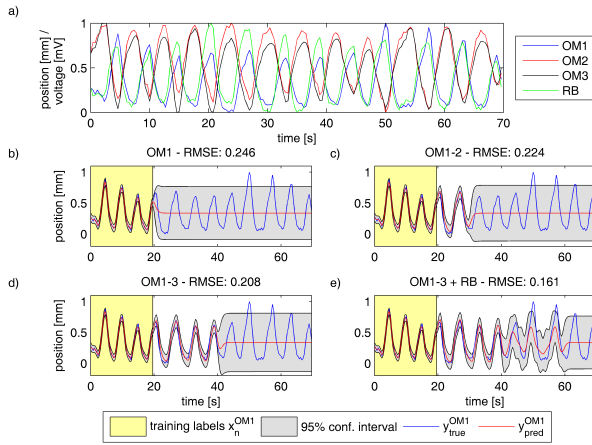


Fig. 2. (a) Signals of three optical markers (OM 1-3) and one respiration belt (RB); (b)-(e) Prediction position of y_{OM1}^{pred} for scenario S1 to S4, respectively

integrated successively into the MTGP. This means that scenario four (S4) will consider all tasks.

Fig. 2(b)-(e) shows the prediction results for the different scenarios including the root mean square error (RMSE) for y_{OM1}^{pred} . It may be seen that the prediction accuracy of y_{OM1}^{pred} improves by taking multiple correlated tasks into account. The RMSE decreases from 0.246 for S1 to 0.161 for S4. We observe that the MTGP is able to learn the correlation between $OM1$ and $OM3$ even though they do not have training data occurring at a shared interval. The correlation is learned via the common intervals of training data for $OM1-OM2$ and $OM2-OM3$. Additionally, the correlation of the RB task, which is the signal with the lowest sampling frequency, can be learned accurately, and which leads to a further decrease of the RMSE.

B. Vital-sign monitoring

In-hospital patients who are recovering from surgery typically have their vital signs (such as heart rate, and respiratory rate) periodically observed by nursing staff in order to monitor the patients' physiological status. One of the limiting factors of existing methods for patient monitoring is deemed to be how they cope with those commonly encountered periods of signal artefact and incomplete or noisy vital-sign data [2]. We investigated the correlation between heart rate and respiratory rate observations from a cohort of patients recovering from cancer surgery in the Oxford University Hospitals NHS Trust. We applied the standard univariate STGP to each physiological time-series, which assumes in dependency between both variables, and compared the results with those obtained by using our MTGP approach (where the same optimisation algorithm was used in both approaches). Fig. 3 shows the results obtained for one example patient.

We can see in Fig. 3 that the MTGP provides a significant improvement in the estimated values for both channels, because the correlation between the two vital signs is taken into account. Particularly, in periods of incomplete data (black arrows in Fig. 3), the predictions are more accurate,

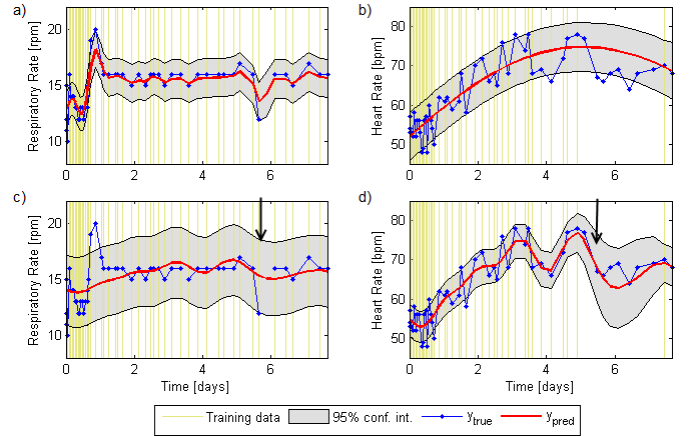


Fig. 3. Application of standard STGP to respiratory rate (a) and heart rate (b). Application of MTGP to the same training data is shown in (c) and (d). Heart rate data from days 5 and 6 were not used for training the GP in both cases.

and can therefore be used to estimate the true value of the data and the distribution over the estimated values.

C. Motion compensation in radiotherapy

In stereotactic body radiotherapy, a particularly challenging task is the precise radiation of moving tumours. Movements are mainly caused by respiration and can be as large as 5cm depending on the position of the tumour. In recent years several technical systems have been developed, such as multileaf collimators and the CyberKnife® (Accuray Inc., Sunnyvale, CA), which enable an active motion compensation. In general, motion compensation is based on a correlation model between external optical markers, measured via tracking cameras, and internal fiducial points, obtained from stereoscopic X-ray imaging [11]. To reduce the additional radiation dose to the patient, due to the X-ray imaging system, the internal fiducial position is only measured at a few selected positions within the breathing cycle.

Fig. 4(a) shows an example of the nonlinear correlation between the external and internal markers. The typical hysteresis is visible, due to the different trajectories for inhalation and exhalation. Current clinical practice is to describe the correlation using a dual polynomial model where one second-order polynomial model describes inhalation and another describes exhalation ("Dual-Poly"). This simple model can only describe the correlation between a single external and internal dimension. Ernst et al. [12] presented a promising alternative approach based on support vector regression (SVR), which uses multiple external dimensions, their first derivatives, and a label to classify inhalation and exhalation as features. However, both methods only make use of external and internal data points which are known at the same time instance during the training phase. The majority of the external data are ignored.

MTGPs can efficiently overcome this problem. To demonstrate this, we use data from a porcine study [12], which

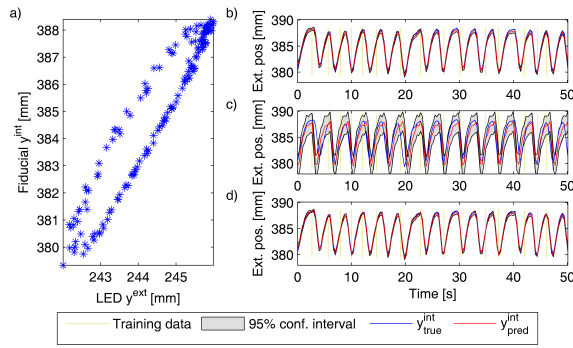


Fig. 4. (a) Nonlinear correlation between internal and external motion; (b)-(d) Respectively, predicted internal fiducial position y^{int}_{pred} for a squared-exponential (SE), periodic (PER) and quasi-periodic (QP) covariance function in the MTGP model ($n_{int} = 15$)

is available online². The subject was manually ventilated. Six external LEDs and four internal fiducial points were constantly measured at $f_s = 14.81$ Hz. The investigated internal signal fragment (150 data points) is shown in Fig. 4(b)-(d). We investigated the RMSE of the predicted internal fiducial position y^{int} depending on the number of internal training points n_{int} for Dual-Poly, the SVR approach of [12], and the MTGP with different $k^l(x, x')$, (5-7). We assumed that all external LED time-series are known ($n_{ext} = 150$). For $n_{int} = \{100, 75, 50\}$, the training data were randomly selected. For $n_{int} = \{20, \dots, 5\}$, the training data were selected by hand to guarantee that a complete breathing cycle is represented. As the number of internal training points n_{int} is equivalent to the number of X-ray images, a lower value of n_{int} is preferable. Note that the results of Dual-Poly and the MTGP models rely only on the data of the third spatial dimension of LED 1. In contrast, the SVR approach uses all three spatial dimensions of LED 1. The MTGP models were extended by a delay hyperparameter $\theta_{\Delta t}$, which transforms x^{ext} to $x^{ext} = x^{ext} + \theta_{\Delta t}$ [10] and is learned automatically.

Investigating the performance of the individual MTGP models reveals that the assumption of a quasi-periodic motion leads to the lowest overall RMSE across all values of n_{int} (Tab. II). In contrast, the RMSE of MTGP(PER) is very

²<http://signals.rob.uni-luebeck.de>

TABLE II
RMSE OF A DUAL POLYNOMIAL MODEL, SVR AND MTGP DEPENDING
ON THE NUMBER OF TRAINING POINTS n_{int} (BEST RESULTS PER n_{int}
HIGHLIGHTED BOLD)

n_{int}	Dual-Poly	SVR	MTGP (SE)	MTGP (PER)	MTGP (QP)
100	0.226	0.172	0.231	1.907	0.154
75	0.233	0.18	0.198	1.885	0.179
50	0.227	0.186	0.318	1.669	0.223
20	0.988	0.207	0.484	1.874	0.307
15	0.542	0.219	0.263	1.993	0.222
10	0.542	0.208	0.393	1.793	0.317
5	-	0.393	1.09	2.245	0.983

high, indicating that the periodic covariance function from eq. (6) is too inflexible for real motion data, as expected. This can be seen in Fig. 4(b)-(d), where the prediction results for all MTGP models for $n_{int} = 15$ are shown. MTGP(PER) cannot adapt to small changes within the breathing cycle, shown in Fig. 4(c), resulting in broader confidence intervals on the prediction. MTGP(SE) can better adapt to these changes, as shown in Fig. 4(a). However, as only few internal training labels are known, the predicted periodic motion is a result of the learned correlation to the external signal. MTGP(QP) assumes a quasi-periodic motion, leading to an improved performance and small confidence intervals. Overall, MTGP(QP) outperforms Dual-Poly and only has a slightly higher RMSE compared to SVR for $n_{int} \leq 50$ and an even a lower error for $n_{int} > 50$. Further improvements can be expected by considering multiple external signals.

IV. CONCLUSION

We have presented a novel, non-parametric method to analyse multivariate physiological time-series based on their correlation. MTGPs are a promising tool for analysing datasets with missing or unevenly-sampled data. Furthermore, prior knowledge of the functional behaviour can be easily integrated and feedback of the current model accuracy can be evaluated using the predicted variance.

REFERENCES

- [1] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, MIT Press, ISBN 0-262-18253 2006.
- [2] L. Clifton, D. A. Clifton, M. A. F. Pimental, P. J. Watkinson, L. Tarassenko, Gaussian Process Regression in Vital-Sign Early Warning Systems, IEEE Engineering in Medicine and Biology Conference, San Diego, USA, pp. 6161 - 6164, 2012.
- [3] R. Dürichen, T. Wissel, F. Ernst, A. Schweikard, Respiratory Motion Compensation with Relevance Vector Machines, in Proc. 16th Int. Conf. MICCAI, vol 8150, pp. 108 - 115, 2013.
- [4] M. A. F. Pimentel, D. A. Clifton, L. Clifton, L. Tarassenko, Probabilistic Estimation of Respiratory Rate using Gaussian Processes, in Proc. IEEE Engineering in Medicine and Biology Conference, Osaka, pp. 2902 - 2905, 2013.
- [5] H. Wackernagel, Multivariate Geostatistics, Springer, ISBN 9783540441427, 2003.
- [6] K. Yu, V. Tresp, A. Schwaighofer, Learning Gaussian Processes from Multiple Tasks, in Proc. 22nd Int. Conf. on Machine Learning, pp. 1012-1019, New York, USA, 2005.
- [7] E. Bonilla, K.M.A. Chai, C.K.I. Williams, Multi-task Gaussian Process Prediction, in Proc. NIPS, 2008.
- [8] K.M.A. Chai, C.K.I. Williams, S. Klanke, S. Vijayakumar Multi-task Gaussian process learning of robot inverse dynamics, in Proc. NIPS, vol. 21, pp. 265-272, 2009.
- [9] R. Dürichen, L. Davenport, R. Bruder, T. Wissel, A. Schweikard, F. Ernst, Evaluation of the potential of multi-modal sensors for respiratory motion prediction and correlation, IEEE Engineering in Medicine and Biology Conference, Japan, pp. 5678-5681, 2013.
- [10] M.A. Osborne, S.J. Roberts, A. Rogers, S. D. Ramchurn, N. R. Jennings, Towards Real-Time Information Processing of Sensor Network Data Using Computationally Efficient Multi-output Gaussian Processes, in Proc. 7th Int. Conf. Information Processing in Sensor Networks, pp. 109-120, St. Louis, 2008.
- [11] A. Schweikard, G. Glosser, M. Bodduluri, M. Murphy, J. R. Adler, Robotic Motion Compensation for Respiratory Movement during Radiosurgery, Computer Aided Surgery, vol. 5, no. 4, pp. 263-277, 2000.
- [12] F. Ernst, V. Martens, S. Schlichting, A. Beširević, M. Kleemann, C. Koch, D. Petersen, A. Schweikard, Correlating Chest Surface Motion to Motion of the Liver Using -SVR A Porcine Study, in Proc. 12th Int. Conf. MICCAI, vol. 5762, pp. 356-364, London, 2009.