



UNIVERSITÄT ZU LÜBECK  
INSTITUT FÜR ROBOTIK  
UND KOGNITIVE SYSTEME

# From Univariate to Multivariate Respiratory Motion Compensation

A Bayesian Way to Increase Treatment Accuracy in Robotic  
Radiotherapy

Robert Dürichen

Dissertation

Universität zu Lübeck  
Institut für Robotik und Kognitive Systeme



From the Institute of Robotics and Cognitive Systems  
of the University of Lübeck  
Director: Prof. Dr.-Ing. Achim Schweikard

# From Univariate to Multivariate Respiratory Motion Compensation

A Bayesian Way to Increase Treatment Accuracy in Robotic  
Radiotherapy

Dissertation  
for Fulfillment of  
Requirements  
for the Doctoral Degree  
of the University of Lübeck

from the Department of Computer Sciences

Submitted by

Robert Dürichen  
from Meißen (Germany)

Lübeck 2014





Robert Dürichen – PhD student  
Institute for Robotics and Cognitive Systems  
Universität zu Lübeck  
Ratzeburger Allee 160  
23538 Lübeck, Germany

Board of examiners

First referee: Prof. Dr.-Ing. Achim Schweikard  
Second referee: Prof. Dr.-Ing. Erhardt Barth

Date of oral examination: 28.11.2014

Approved for printing. Lübeck, 07.07.2015



# Zusammenfassung

Krebserkrankungen gehören nach wie vor zu den häufigsten Todesursachen weltweit. Trotz intensiver Forschung und verbesserter diagnostischer und therapeutischer Methoden liegen die durchschnittlich Fünf-Jahres-Überlebensraten für Krebsarten wie dem Lungen-, Leber- oder Bauchspeicheldrüsenkarzinom noch immer unter 25 %.

Neben chirurgischen und chemotherapeutischen Behandlungsansätzen stellt die Strahlentherapie eine der wichtigsten Behandlungsmethoden dar. Bei Tumoren in der Brust- und Bauchregion kommt dabei insbesondere die robotergestützte Strahlentherapie zum Einsatz. Sie bietet den Vorteil, dass Tumorbewegungen, die aufgrund der Atembeziehungsweise Herzaktivitäten des Patienten entstehen, kompensiert werden können. Klinisch genutzte Systeme sind unter anderem das CyberKnife<sup>®</sup> oder Systeme mit Multilamellenkollimatoren. Um eine präzise und patientenschonende adaptive Bewegungskompensation durchführen zu können, ist ein Prädiktions- und ein Korrelationsmodell notwendig. Ersteres kompensiert Bestrahlungsungenauigkeiten aufgrund von Latenzzeiten. Diese entstehen durch kinematische Limitierungen des jeweiligen Systems sowie durch Datenaufnahme und -verarbeitung. Das zweite Modell reduziert eine zusätzliche Strahlenbelastung des Patienten, die durch die Lokalisierung des Tumors mittels stereoskopische Röntgenbildgebung entsteht. Aufgrund relativ langer Behandlungszeiten würde eine kontinuierliche Bewegungsverfolgung des Tumors zu einer nicht zu vernachlässigbaren Strahlungsdosis führen. Um diese zu vermeiden, wird ein Korrelationsmodell trainiert, das die Tumorposition basierend auf Daten von externen Surrogaten berechnet. Klinisch genutzte Prädiktions- und Korrelationsmodelle basieren ausschließlich auf externen optischen Markern.

Die Messung von Atmungsaktivitäten ist für viele medizinische Bereiche relevant. Folglich haben sich verschiedene Messmethoden etabliert, welche teilweise auf unterschiedlichen Modalitäten und Messpositionen basieren. Dabei erfassen Sensoren lediglich einen Ausschnitt des ihnen zugrundeliegenden komplexen Atmungsprozesses, der teilweise durch andere Störeinflüsse überlagert wird. Motiviert durch diese Beobachtung ist das primäre Ziel dieser Arbeit die Entwicklung von multimodalen Prädiktions- und Korrelationsmodellen. Das Ziel ist es, relevante Informationen von verschiedenen Senso-

ren zu kombinieren, um die Behandlungsgenauigkeit zu erhöhen.

Dabei werden insbesondere wahrscheinlichkeitsbasierte Algorithmen untersucht. Aufgrund der immer weiter steigenden Leistungsfähigkeit von modernen Computern sind diese berechnungsintensiveren Algorithmen in den letzten Jahren auch für Echtzeitanwendungen nutzbar geworden. Im Vergleich zu alternativen Ansätzen besitzen diese Modelle einen für medizinische Anwendungen relevanten Vorteil. Das Ergebnis entspricht einer Wahrscheinlichkeitsverteilung mit Mittelwert und Varianz. Neben dem Mittelwert, der z.B. der berechneten Position des Tumors entsprechen kann, enthält die Varianz Informationen über die "Zuverlässigkeit" des Modells in dem aktuell berechneten Wert. Durch eine Auswertung der Varianz können Rückschlüsse über die aktuelle Behandlungsgenauigkeit getroffen werden, was den zweiten Schwerpunkt dieser Arbeit bildet.

Eine detaillierte Literaturrecherche zeigte, dass auch im wissenschaftlichen Umfeld primär nur optische, externe Surrogate untersucht wurden. Des Weiteren wurde bisher nur ein einziger wahrscheinlichkeitsbasierter Algorithmus für das Problem der Atemkompensation präsentiert mit teilweise widersprüchlichen Ergebnissen. Folglich wurden in dieser Arbeit zunächst zwei neue wahrscheinlichkeitsbasierte Bewegungskompensationsalgorithmen entwickelt, welche auf Gaußprozess (GP)-Modellen und Relevanzvektor-Maschinen (RVMs) basieren. Bevor diese Modelle in einem multivariaten Szenario untersucht wurden, fand eine intensive Evaluierung statt. Für die Kompensation von Latenzzeiten (Prädiktionsalgorithmus) wurde der Einfluss von modelspezifischen Parametern anhand von repräsentativ ausgewählten Atmungssignalen untersucht. Dies beinhaltet unter anderem den Einfluss von Anzahl von Trainingspaaren (im Fall von RVM Modellen) oder Kovarianzfunktionen (im Fall von GP-Modellen). Abschließend wurde eine Evaluation auf einem umfassenden Datensatz durchgeführt, der insgesamt 304 Atmungsbewegungen beinhaltet. In einer früheren Vergleichsstudie wurde dieser Datensatz genutzt, um die Genauigkeit von sechs verschiedenen Prädiktionsalgorithmen (unter anderem Modelle basierend auf Support Vektor Regression (SVR), Kalman Filtern, und einer Wavelet-basierten Methode der kleinsten Fehlerquadrate (wLMS)) für verschiedene Latenzzeiten zu untersuchen. Die Ergebnisse zeigen, dass die durchschnittliche Prädiktionsgenauigkeit eines RVM-Modells mit linearer Basisfunktion alle bisherigen Modelle übertrifft, unabhängig von der zu prädizierenden Latenz. Die Ergebnisse des besten GP-Modells zeigen eine erhöhte mittlere Prädiktionsgenauigkeit für kurze Latenzzeiten ( $h = \{77, 115, 154\}$  ms). Im direkten Vergleich mit dem wLMS-Algorithmus, dem Prädiktionsalgorithmus mit der bisher höchsten Genauigkeit, kann der GP- 78.62 %



und der **RVM**-Algorithmus 92.43 % der Daten mit einer höheren oder zumindestens gleichwertigen Genauigkeit vorhersagen (für eine Latenzzeit von  $h = 115$  ms).

Zur Untersuchung der Fehlerkontrolle in Echtzeit mit Hilfe der Varianz wurden die bei der vorherigen Untersuchung gleichzeitig berechneten Varianzdaten genutzt. Am Beispiel des **RVM**-Modells mit einer linearen Basisfunktion wurde ein simples, varianzbasiertes Schwellwertverfahren angewendet. Bei Überschreitung des Varianzschwellwertes wurde die Behandlung unterbrochen. Die Ergebnisse zeigen, dass in Abhängigkeit des gewählten Schwellwertes der Prädiktionsfehler der verbleibenden Daten kontrolliert werden kann. Weitere Untersuchungen ergaben, dass die Varianz auch als Kriterium zur Konstruktion von Hybridalgorithmen genutzt werden kann. Ein für die Praxis besonders relevanter Ansatz stellt die Kombination von mehreren **RVM**-Modellen mit verschiedenen Basisfunktionen dar, da im Vorfeld kein spezifischer Varianzgrenzwert festgelegt werden muss. Durch Nutzung des Hybridalgorithmus  $\text{HYB}_{\text{RVM}}$  ist eine weitere Reduktion des durchschnittlichen Prädiktionsfehlers möglich. Im Vergleich zum **wLMS**-Algorithmus prädiziert  $\text{HYB}_{\text{RVM}}$  94.74 % der Bewegungsdaten mit einer höheren oder zumindestens gleichwertigen Genauigkeit. Im Gegensatz zu alternativen Maßen, die in Echtzeit ausgewertet werden können, wie dem aktuellen Prädiktionsfehler, basiert die Varianz einzig auf den aktuellen Eingangsdaten  $x$  und den vorhandenen Trainingsdaten des Modells. Dies ist ein Vorteil vor allem bei Anwendungen, in denen der Zielwert nur selten gemessen werden kann, wie zum Beispiel bei Korrelationsalgorithmen (Vermeidung von zusätzlichen Röntgenaufnahmen). Die Berechnung des aktuellen Prädiktionsfehlers, der auf dem gemessenen Zielwert  $y$  basiert, ist in diesem Fall nur sehr selten möglich.

Um multivariate Bewegungskompensationsansätze zu untersuchen, wurde eine Studie mit 18 Probanden durchgeführt. Die Studie beinhaltete zwei Messphasen, in denen die Probanden entweder normal oder bewusst unregelmäßig (z.B. Husten oder Gähnen) atmeten. Während dieser Phasen wurden Daten von sechs verschiedenen externen Modalitäten aufgenommen (unter anderem Beschleunigung, Dehnung und Oberflächenmuskellaktivität (**sEMG**)). Gleichzeitig wurde die innere Bewegung der Leber über 4D-Ultraschall (**US**) aufgezeichnet. Eine gut sichtbare Gefäßbifurkation diente als simuliertes Bestrahlungsziel. Eine initiale Untersuchung der Korrelation zwischen den externen Surrogaten untereinander und zwischen internen und externen Daten bestätigt eine hohe Korrelation mit  $0.54 \leq r \leq 0.93$  zwischen den nicht-optischen und den optischen Surrogaten, beziehungsweise der internen Bewegung.

Ausgehend von diesen Ergebnissen wurden im folgenden Schritt multivariate Prädiktionsalgorithmen untersucht. Dabei wurden die Eingangsdaten von drei nicht wahr-scheinlichkeitsbasierten Algorithmen (**nLMS**, **wLMS**, **SVR**) und des **RVM**-Modells durch multivariate Daten erweitert. Um ein mögliches Übertrainieren der Modelle zu vermeiden, wurde ein sequentielles Vorwärtsselektionsverfahren (**SFS**) genutzt, um die Marker mit den relevantesten und am wenigsten redundanten Daten auszuwählen. Durch Nutzung von multivariaten im Vergleich zu univariaten Daten konnte der mittlere quadratische Gesamtfehler (**RMSE**) des **RVM**-Algorithmus (der Algorithmus mit der höchsten Prädiktionsgenauigkeit) um 20 % im Fall von normaler Atmung und um 12 % im Fall von unregelmäßiger Atmung reduziert werden.

Ein alternativer Ansatz wurde gewählt, um **GP**-Modelle auf multivariate Daten zu erweitern. Im Gegensatz zur Erweiterung der Eingangsdaten des Modells wurde ein Multi-Output Modell analysiert. Dieses sogenannte Mehrfach-Aufgaben Gaussprozess (**MTGP**)-Modell kann zusätzlich zur Korrelation zwischen Datenpunkten eines Markers (entspricht dem Fall eines normalen **GP**-Modells) die Korrelation zu anderen Markern lernen. Des Weiteren hat die Nutzung von **MTGP**-Modellen folgende Vorteile:

- Prädiktion von beliebigen Latenzzeiten unabhängig von der Abtastfrequenz,
- Einbeziehung von im Vorfeld bekannten Signalcharakteristiken und
- Nutzung von markerspezifischen Trainingsdaten.

Die letzte Eigenschaft ist besonders für Bewegungskompensationsalgorithmen relevant, da sie eine Möglichkeit darstellt, das Prädiktions- und Korrelationsproblem innerhalb eines Modells zu beschreiben. Die bisherige Aufteilung in zwei getrennte Modelle ist durch die unterschiedliche Anzahl von vorhandenen externen und internen Trainingsdaten begründet. **MTGP**-Modelle bieten die Möglichkeit, in einem einzigen Modell alle vorhanden Daten (sowohl externe als auch interne) zu integrieren, unabhängig von deren Anzahl, Markerzugehörigkeit und dem Zeitpunkt der Datenaufnahme. Die Eigenschaften dieses neuen Ansatzes wurden an verschiedenen synthetischen Beispielen studiert und führten zu der Entwicklung einer frei zugänglichen Toolbox<sup>1</sup>.

Im Kontext der Bewegungskompensation wurde die Genauigkeit von uni- und multivariaten **MTGP**-Modellen, sowohl als reine Korrelationsmodelle als auch als kombinierte Prädiktions- und Korrelationsmodelle, untersucht. Im Vergleich zu alternativen Korrelationsmodellen, die unter anderem auf polynomiellen Modellen oder **SVR** basieren, belegen die Untersuchungen, dass **MTGP** Modelle im Mittel eine höhere Genauigkeit

---

<sup>1</sup><http://www.robots.ox.ac.uk/~davidc/>

erzielen. Durch die Nutzung von multivariaten Daten konnte der mittlere Korrelationsmodellfehler (RMSE) um 0.2 mm gesenkt werden ( $\approx 10\%$ ). Ähnliche Ergebnisse konnten auch für ein kombiniertes Prädiktions- und Korrelationsmodell erzielt werden.

Durch ihre hohe Genauigkeit und große Flexibilität (Kombination von Prädiktion und Korrelation, Einbindung multivariater Daten) sind MTGP-Modelle sehr vielfältig einsetzbar. Aktuelle Forschungstrends in der robotergestützten Strahlentherapie können gut mit MTGP-basierten Bewegungskompensationsmodellen kombiniert werden. Dabei ist insbesondere die verstärkte Nutzung von US-Bildgebung zur Tumorverfolgung zu erwähnen. Aufgrund von hohem Bildrauschen beziehungsweise einer Abschattung des Tumors durch umliegende Strukturen (US-Schatten) ist eine kontinuierliche Verfolgung des Tumors nicht immer garantiert. Ein multivariates MTGP-Modell kann leicht durch die Positionsinformationen der US-Bildgebung erweitert werden. Somit kann zum Beispiel ein erhöhtes Bildrauschen während der Behandlung automatisch detektiert und durch Informationen von anderen externen Surrogaten kompensiert werden.

Die Ergebnisse dieser Arbeit zeigen, dass wahrscheinlichkeitsbasierte Prädiktionsalgorithmen im Vergleich zu bisher veröffentlichten Ergebnissen über 95 % der Datensätze genauer vorhersagen können. Die Gesamtgenauigkeit von adaptiven Bewegungsansätzen kann durch Nutzung multivariater Daten durchschnittlich um 0.2 mm verbessert werden. Bei den verwendeten Sensoren handelt es sich teilweise um bereits in anderen klinischen Bereichen verwendeten Geräte, wodurch eine schnelle und kostengünstige Überführung in strahlentherapeutische Behandlungssysteme möglich ist.



# Abstract

Cancer is one of the most prevalent causes of death worldwide. Even though intense research effort has been dedicated to advance diagnostic and treatment techniques, statistics indicate that the five year survival rates especially for lung, liver and pancreas tumours are still very low.

Radiotherapy is a relevant treatment option beside surgical and chemotherapeutic approaches. One treatment option for tumours located in the breast and abdominal regions is robotic radiotherapy. The advantage is that tumour motion caused by respiratory or cardiac motion of patients can be compensated which increases the treatment accuracy. Clinically used systems are the CyberKnife<sup>®</sup> or **multileaf collimators (MLCs)**. To perform an adaptive **motion compensation (MC)**, two problems have to be addressed. First, time latencies of the specific treatment system (due to data processing and kinematic limitations) have to be compensated. Second, real-time acquisition of the internal tumour position is not possible without exposing the patient to an additional radiation dose. To overcome these problems, prediction and correlation models are used, which both depend on data of univariate external optical surrogates.

The problem of respiratory motion compensation is relevant for various medical applications. In general, a diverse number of techniques has been developed to measure respiratory activity including different sensor modalities and measurement positions. However, data of a single sensor covers only information of a specific aspect of the more complex underlying process of respiration. Motivated by this observation, the first and main focus of this work is to develop multivariate extensions of current prediction and correlation models. The aim is to efficiently combine information of multiple sensors to further increase treatment accuracy.

In particular, probabilistic algorithms will be applied for this purpose. Due to the increased power of modern computers, these computationally more demanding models have become applicable for real-time applications. In contrast to non-probabilistic models, the output of these algorithms is a probability distribution with predicted mean and variance. The variance contains relevant information about the “certainty” of the algorithm in the current output. Such information might be useful to control the prediction

error during the treatment, which is the second main focus of this work.

An extensive literature review confirms the initial assumptions that primarily optical surrogates were investigated in the literature. Furthermore, only one probabilistic algorithm was evaluated for the purpose of respiratory motion compensation - with controversial results. As a consequence, two new **MC** algorithms were developed, namely **Gaussian process (GP)** models and **relevance vector machines (RVMs)**. Before applying these in a multivariate setting, the performance of these algorithms was examined for the purpose of standard univariate prediction models. The influence of model specific parameters such as the number of training pairs (in case of **RVM**) and covariance function (in case of **GP**) was studied on selected motion traces. Furthermore, a comprehensive evaluation was performed on a dataset consisting of 304 motion traces. They were compared to six previously published algorithms including **support vector regression (SVR)**, Kalman Filter, and wavelet-based least mean square (**wLMS**) methods for four different prediction latencies. On average, the **RVM** model with a linear basis function had a superior accuracy and outperformed all previous algorithms independently of the prediction latencies. The result of the best **GP** model indicates a superior accuracy for shorter prediction horizons ( $h = \{77, 115, 154\}$  ms) compared to the previously published algorithms. For  $h = 115$  ms, the **GP** approach predicts 78.62% and the **RVM** model 92.43% of the data more accurately or equally good as the **wLMS** method (which had been the best algorithm so far).

The benefit of monitoring the variance was investigated on the same dataset for the **RVM** model, exemplarily. Applying a simple variance threshold, to interrupt the treatment if exceeded, confirmed that the prediction error can be controlled by the variance. Further experiments showed that the variance can be exploited as a criterion to design hybrid algorithms. A practically relevant approach (as it does not require further parameters) is the combination of three **RVM** models with different basis functions (**HYB<sub>RVM</sub>**), which could even predict 94.74% of the data more accurately or equally good as the **wLMS** method. In contrast to alternative real-time measures such as the current prediction error, the variance is only based on the input features  $x$  and not on the output  $y$ . This allows an error control for applications where the true output cannot be constantly observed as in the case of correlation algorithms.

To investigate multivariate motion compensation techniques, a human study with 18 subjects was performed. Data of six external modalities such as acceleration, respiratory flow, or **surface electromyography (sEMG)** were acquired for two measurement

phases; focussing on either normal or irregular breathing. Internal data were recorded by 4D-ultrasound (US) of the liver. An initial analysis of the correlation between the data of external, and external and internal signals revealed high correlation coefficients ( $0.54 \leq r \leq 0.93$ ) between non-optical and optical surrogates or the internal motion, respectively.

In the following, multivariate prediction algorithms were investigated. The input data of three non-probabilistic algorithms (nLMS, wLMS, SVR) and the RVM model were extended to incorporate multivariate data. To prevent overfitting, a sequential forward selection (SFS) method was used to select the most relevant and least redundant sensor combination. Using a multivariate instead of a univariate setting, the mean root mean square error (RMSE) of RVM (which outperformed all other approaches) decreased by 20 % in the case of normal breathing and by 12 % in the case of irregular breathing.

In order to extend GP models to multivariate data, an alternative approach was investigated. Instead of extending the input data, a multi-output model was examined. We refer to this approach as multi-task Gaussian Process (MTGP) models. These models are capable of learning the correlation between and within sensors and offer several distinct advantages such as

- prediction of arbitrary latencies independent of the sampling frequency,
- incorporation of prior knowledge of the signal characteristics, and
- use of signal-specific training observations.

The latter is particularly relevant for motion compensation algorithms. With the MTGP approach, unified prediction and correlation models can be constructed which make use of all available training data. The properties of this new approach was intensively studied on synthetic examples and led to the development of an open-access toolbox<sup>2</sup>. Further, the approach was evaluated as uni- and multivariate correlation as well as a unified prediction and correlation algorithm. The comparison to alternative correlation approaches based on polynomial models or SVR revealed a superior performance of MTGPs. By using multivariate data, the mean RMSE of an MTGP based correlation algorithm could be decreased by 0.2 mm ( $\approx 10\%$ ). Similar results were observed for a unified MTGP prediction and correlation model for different prediction latencies.

The results of this work show that probabilistic prediction algorithms can predict over 95 % of the data more accurately compared to previous results. Furthermore, the overall accuracy of adaptive motion compensation can be increased by on average 0.2 mm when

---

<sup>2</sup><http://www.robots.ox.ac.uk/~davidc/>

## *Abstract*

---

using multivariate data. Most of the applied sensors are already used in clinical practise which allows a fast and cost-efficient integration into current radiotherapy systems.



# Contents

<b>Mathematical Notation and Indices</b>	<b>xxi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Medical Background . . . . .	1
1.1.1 Characterization of Respiratory Motion . . . . .	1
1.1.2 Measuring Respiration . . . . .	4
1.1.3 The Problem of Respiratory Motion Compensation . . . . .	8
1.2 Motion Compensation in Radiotherapy . . . . .	12
1.2.1 Radiotherapy . . . . .	12
1.2.2 Adaptive Motion Compensation and Clinical Systems . . . . .	17
1.3 Purpose of this Work . . . . .	21
1.4 Organisation . . . . .	23
<b>2 Literature Review</b>	<b>25</b>
2.1 Basic Notation . . . . .	25
2.2 Evaluation Measures . . . . .	28
2.2.1 Prediction Accuracy Measures . . . . .	28
2.2.2 Auxiliary Measures . . . . .	30
2.2.3 Example of Evaluation Measures . . . . .	31
2.3 Motion Compensation Algorithms . . . . .	33
2.3.1 Prediction Algorithms . . . . .	34
2.3.2 Correlation Algorithms . . . . .	46
2.4 Conclusion . . . . .	52
<b>3 Univariate Probabilistic Respiratory Motion Prediction</b>	<b>55</b>
3.1 Dataset . . . . .	55
3.2 Probabilistic Machine Learning . . . . .	57
3.2.1 Relevance Vector Machines . . . . .	60
3.2.2 Gaussian Process Models . . . . .	68

3.3	Comparative Evaluation . . . . .	79
3.3.1	Evaluation Procedure . . . . .	79
3.3.2	Results . . . . .	80
3.3.3	Discussion . . . . .	85
3.4	Exploiting Probabilistic Uncertainties . . . . .	88
3.4.1	Experimental Evaluation of the Variance . . . . .	89
3.4.2	Hybrid Algorithms . . . . .	93
3.5	Conclusion . . . . .	99
<b>4</b>	<b>Multivariate Respiratory Motion Prediction</b>	<b>103</b>
4.1	Multivariate Dataset . . . . .	103
4.1.1	Measurement Phases and Subjects . . . . .	104
4.1.2	Measurement Setup . . . . .	104
4.1.3	Data Pre-processing . . . . .	109
4.2	Mathematical Background . . . . .	110
4.2.1	Multivariate Prediction Algorithms . . . . .	110
4.2.2	Feature Selection . . . . .	112
4.3	Evaluation Procedure and Proposed Methods . . . . .	115
4.4	Experimental Evaluation . . . . .	118
4.4.1	Correlation Analysis . . . . .	118
4.4.2	Multivariate Prediction . . . . .	121
4.4.3	Discussion . . . . .	125
4.5	Further Respiratory Motion Features . . . . .	130
4.5.1	EEG and EOG Features . . . . .	131
4.5.2	sEMG Features . . . . .	132
4.6	Conclusion . . . . .	135
<b>5</b>	<b>Multi-Task Gaussian Process Models</b>	<b>139</b>
5.1	Datasets . . . . .	139
5.2	Mathematical Background . . . . .	141
5.2.1	From GPs to Multi-Task GPs . . . . .	142
5.2.2	Correlation Matrix and Normalisation . . . . .	148
5.2.3	Time Shift Estimation . . . . .	150
5.2.4	MTGP Tracking Algorithm . . . . .	152
5.2.5	Alternative Algorithms . . . . .	154
5.3	Evaluation Procedure . . . . .	155

5.4	Experimental Evaluation . . . . .	158
5.4.1	Correlation Analysis . . . . .	158
5.4.2	MTGP - Number of Internal Training Points . . . . .	159
5.4.3	MTGP - Multivariate Prediction and Correlation . . . . .	163
5.4.4	Discussion . . . . .	163
5.5	Further MTGP Extensions . . . . .	167
5.5.1	Linear Model of Coregularization . . . . .	167
5.5.2	Convolution of Kernels . . . . .	170
5.6	Conclusion . . . . .	172
<b>6</b>	<b>Conclusion and Future Work</b>	<b>175</b>
6.1	Probabilistic Motion Compensation Algorithms . . . . .	176
6.2	Univariate versus Multivariate . . . . .	177
6.3	Real-Time Feedback . . . . .	179
6.4	MTGP Models in Radiotherapy . . . . .	181
6.5	MTGP Models in General . . . . .	185
	<b>Bibliography</b>	<b>187</b>
	<b>List of Figures</b>	<b>207</b>
	<b>List of Tables</b>	<b>211</b>
	<b>Glossary and List of Abbreviations</b>	<b>213</b>



# Mathematical Notation and Indices

Symbol	Description
$b$	linear offset used, e.g., in interpolation and linear regression algorithm, adaptive filters, and <b>SVR</b>
$c$	approximation coefficients (wavelet decomposition; <b>wLMS</b> )
$d$	dimension of input features
$e$	error
$f$	frequency in [Hz]
$h$	time latency in [s] ( $h = \xi / f_s$ )
$i$	time index (if not stated differently)
$k(\cdot)$	covariance function (see <b>GP</b> )
$l$	label indicating the belonging of $y$ to sensor $l$ (see <b>MTGP</b> )
$m$	number of training pairs
$m(\cdot)$	mean function (see <b>GP</b> )
$o$	number of available sensors
$p(x)$	probability of $x$
$p(x y)$	probability of $x$ given $y$
$t$	time in [s]
$u$	updating factor for hyperparameters (see <b>GP</b> )
$w$	weight
$x$	input feature of a prediction or correlation algorithm
$y$	acquired data of a sensor or observation (unit depend on sensor modality)
$\delta(k, j)$	Kronecker-Delta ( $\delta(k, j) = 1$ for $k = j$ , $\delta(k, j) = 0$ for $k \neq j$ )
$\gamma$	kernel parameter of a <b>RBF</b>
$\varepsilon$	parameter to model measurement noise (often $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ ) (parameter also used in $\varepsilon$ - <b>SVR</b> - defines width of $\varepsilon$ -tube)
$\mu$	learning rate ( <b>LMS</b> algorithm) or mean of Gaussian distribution (see <b>RVM</b> )
$\varphi(x)$	basis function to transform input feature $x$ (see <b>RVM</b> )
$\phi(x)$	phase shift in [°]

$\sigma$	standard deviation (often in [mm])
$\sigma^2$	variance (often in [mm <sup>2</sup> ])
$\theta$	hyperparameter of a GP model
$\xi$	time latency index
$C$	normalization factor in cost function for SVR and KDE
$F$	complete feature set (all sensors are used)
$J(y)$	jitter of signal $y$ [77]
$J$	maximum number of wavelet decompositions (LMS)
$N$	number of data points
$S$	selected features set
$T_p$	breathing period in [s]
$W$	detail coefficients (wavelet decomposition; wLMS)
$\Phi$	basis function applied to input feature matrix $\mathbf{X}$ (see RVM)
$\Sigma$	covariance matrix (see RVM)
$\mathcal{N}$	normal distribution
$\mathcal{U}$	uniform distribution
$\mathbf{I}$	identity matrix
$\mathbf{K}$	covariance matrix (see GP models)
$\mathbf{K}_c$	correlation covariance matrix (Eq. 5.3; see MTGP models)
$\mathbf{K}_t$	temporal covariance matrix (Eq. 5.3; see MTGP models)
$\mathbf{L}$	lower triangular matrix (Cholesky decomposition)
$\mathbf{T}$	training set used for an MC algorithm ( $\mathbf{T} = \{\mathbf{X}, \mathbf{y}\}$ )
$\mathbf{X}$	matrix of input features

*Superscript indices*

$y^*$	computed output of a prediction or correlation algorithm
$x^{int} / y^{int}$	feature / data of internal sensor (not inside the body)
$x^{ext} / y^{ext}$	feature / data of external sensor (inside the body)
$x^{OM1} / y^{OM1}$	feature / data of optical marker 1
$x^{OM2} / y^{OM2}$	feature / data of optical marker 2
$x^{OM3} / y^{OM3}$	feature / data of optical marker 3
$x^{Flow} / y^{Flow}$	feature / data of flow sensor
$x^{Strain} / y^{Strain}$	feature / data of strain sensor
$x^{ACC} / y^{ACC}$	feature / data of acceleration sensor
$x^j / y^j$	feature / data of sensor $j$

$\sigma^{2*}$	variance predicted by a probabilistic algorithm (RVM or GP)
MTGP <sup>NLML</sup>	multivariate MTGP model optimised based on NLML value
MTGP <sup>best</sup>	multivariate MTGP model optimised retrospectively on the RMSE

*Subscript indices*

$t_i$	time in [s] at time index $i$ ( $t_i = i/f_s$ )
$y_i$	data at time index $i$
$d_{th}$	error threshold (in general in [mm])
$e_i$	error at time index $i$ ( $e_i = y_i - y_i^*$ )
$e_{th}$	error threshold (in general in [mm])
$f_s$	sampling frequency in [Hz]
$f_c$	frequency to which all sensors are synchronized in [Hz]
$f_{global}$	update factor for SVR models [62]
$f_\varepsilon$	update factor for SVR models [62]
$n_{ext}$	number of external training points (for a correlation model)
$n_{int}$	number of internal training points (for a correlation model)
$n_{iter}$	number of iterations
$n_{feat}$	number of features
$n_{imp-FS}$	number of improved features sets
$p_{th}$	probability threshold
$r$	Euclidean distance between two features
$r_{x,y}$	Pearson's correlation coefficient between $x$ and $y$
$ r_{x,y} $	absolute Pearson's correlation coefficient between $x$ and $y$
$\Delta r_{x,y} $	absolute Pearson's correlation coefficient difference (depending on two time segments)
$t_{comp}$	computation time in [s] (depend on used hardware)
$k_{SE}$	squared-exponential covariance function (Eq. 3.37)
$k_M$	Matérn covariance function (Eq. 3.38)
$k_{RQ}$	rational quadratic covariance function (Eq. 3.39)
$k_P$	periodic covariance function (Eq. 3.40)
$k_N$	noise covariance function (Eq. 3.41)
$k_{QP}$	quasi-periodic covariance function (Eq. 3.42)
$\sigma_{th}^2$	threshold variance
$\theta_L$	time scaling hyperparameter (see GP models)
$\theta_S$	$y$ -scaling hyperparameter (see GP models)

$\theta_P$	period hyperparameter (see <b>GP</b> models)
$\theta_D$	hyperparameter defining the degree of a covariance function
$\text{GP}_{QRSN}$	<b>GP</b> model with a summed covariance function of $k_{QP}$ , $k_{RQ}$ , $k_{SE}$ , and $k_N$
$\text{GP}_{QRM}$	<b>GP</b> model with a summed covariance function of $k_{QP}$ , $k_{RQ}$ , and $k_M$
$\text{HYB}_{wR-lin}$	hybrid model consisting of a <b>wLMS</b> and linear <b>RVM</b> model
$\text{HYB}_{wR-quad}$	hybrid model consisting of a <b>wLMS</b> and quadratic <b>RVM</b> model
$\text{HYB}_{wR-cub}$	hybrid model consisting of a <b>wLMS</b> and cubic <b>RVM</b> model
$\text{HYB}_{RVM}$	hybrid model consisting of a three <b>RVM</b> model
$\text{MTGP}_P$	<b>MTGP</b> model with periodic covariance function
$\text{MTGP}_{SE}$	<b>MTGP</b> model with squared-exponential covariance function
$\text{MTGP}_{QP}$	<b>MTGP</b> model with quasi-periodic covariance function
$N_{pos}$	number of data points below a certain threshold
$\text{RVM}_{lin}$	<b>RVM</b> model with a linear basis function
$\text{RVM}_{quad}$	<b>RVM</b> model with a quadratic basis function
$\text{RVM}_{cub}$	<b>RVM</b> model with a cubic basis function
$\text{wLMS}_{J0}$	<b>wLMS</b> model with $J = 0$
$\text{wLMS}_{J3}$	<b>wLMS</b> model with $J = 3$



# 1 Introduction

This chapter gives a general introduction to the problem of respiratory motion compensation. The physiology of respiration and the relevance for different medical applications will be discussed in sec. 1.1. Section 1.2, addresses the problem of motion compensation within radiotherapy which is one of the most relevant applications. There, different compensation techniques as well as clinical systems will be presented. In the last two sections, the purpose (sec. 1.3) and the structure (sec. 1.4) of this work are outlined.

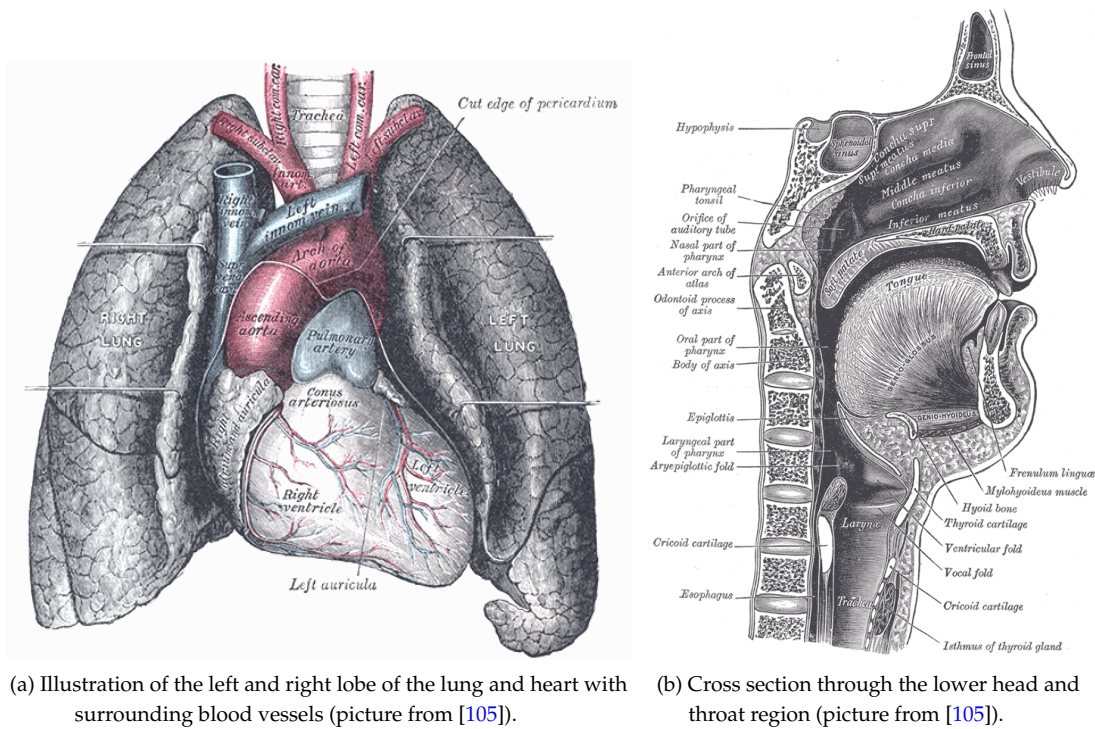
## 1.1 Medical Background

Within sec. 1.1.1, a brief introduction to the physiology of respiration and an overview of prospective motion amplitudes for different organs is provided. Afterwards in sec. 1.1.2, measurement modalities and locations to acquire respiratory activity are summarized. Please note, we refer to respiratory activity as all measurable signals which contain information about respiratory motion starting from the electro-physiological signals in the brain and the nerves to the respiratory flow. In sec. 1.1.3, the most relevant biomedical applications are listed which can benefit from any kind of respiratory motion compensation.

### 1.1.1 Characterization of Respiratory Motion

The term respiration refers to the process of regulating the gas exchange (intake of oxygen and removal of carbon dioxide) between the surrounding environment and the human body [119, 131]. It is also referred to as physiological respiration. Note the difference to cellular respiration which specifies the metabolic process within a cell to generate energy by the reaction of oxygen and glucose. The focus of this work lies solely on physiological respiration.

The gas exchange takes place in the lung. However, several additional structures are required for gas transport such as the nose, throat, and trachea. Respiratory movements can be divided into inspiration and expiration. The lung is an elastic organ which is not able to perform movements by itself. Respiratory motion is caused by the respiratory



**Figure 1.1:** Anatomical details of the respiratory system

muscles. The primary respiratory muscle is the diaphragm which is spanned in the transversal plane and separates the chest and the abdominal cavity. During inspiration, the diaphragm contracts which results in a displacement towards the abdomen. The organs and other structures within the abdomen will be displaced anteriorly and inferiorly. The second respiratory muscles are the intercostal muscles. During inspiration, the contraction of the external intercostal muscles leads to a superior and anterior displacement of the ribs. This results in an increased diameter of the thorax. It can be summarized that the contraction of the diaphragm results in an increased **superior-inferior (SI)** dimension of the chest cavity and of the external intercostal muscles in an increased **anterior-posterior (AP)** dimension. During quiet breathing, exhalation takes place passively without interaction of muscles. In case of deep breathing, the exhalation is an active movement as additional respiratory muscles are used such as the internal intercostal muscles. Figure 1.1 and 1.3 illustrate the anatomical details of the respiratory system.

The normal breathing volume varies in a range of about 500 ml [110, 119] up to 800 ml [20]. In extreme cases, a breathing volume of 1549 ml has been reported [19]. These values depend on the height, gender, and age of the subjects. For deep inspiration and expira-

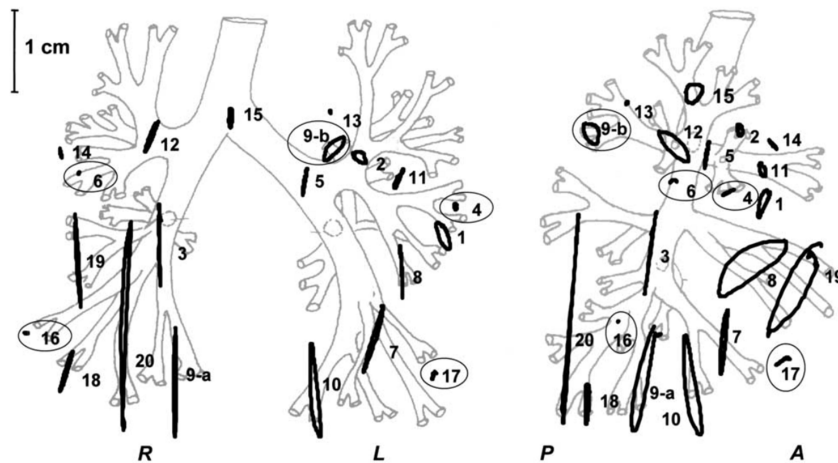
tion, the volumes increase up to 3500 – 5600 ml [110, 119]. The respiratory frequency for normal breathing of an adult is approximately 0.25 Hz [110, 119]. However, frequency variations between 0.1 – 0.5 Hz have been reported [19].

Respiratory motion can be classified as involuntary motion [131]. In general, patients breathe unconsciously. However, within certain limits, it is possible to control the breathing pattern such as the frequency and the volume. Furthermore, Wade *et al.* [270] could not find evidence that it is possible to directly control the diaphragm. Consequently, a wide variation of breathing patterns can be expected [22, 125, 131, 205, 279].

Various studies investigated the amplitude and the direction of internal respiratory motion. In the following, some of the results shall be presented depending on specific organs or structures to give an overview of normal motion amplitudes.

**Diaphragm:** One of the earlier studies was presented by Wade *et al.* [270]. The authors investigated the diaphragm motion, circumference of the torso and tidal value of ten subjects. For normal breathing, they reported a movement of 15 mm of the diaphragm and an increased circumference of the torso of 7 mm in supine position. In case of deep breathing, the diaphragm motion can increase up to 70 – 130 mm and the chest circumference increases by 50 – 110 mm. In a more recent study [101], an amplitude of  $34.3 \pm 20.4$  mm was reported for deep breathing. Further studies reported similar results [52, 97].

**Lung:** The motion range of the lung has been investigated thoroughly, e.g., in [56, 163, 236, 244, 245]. Shirato *et al.* [245] scrutinised the movement of fiducials in the lung of 21 patients. On average, they reported a motion of 8.2 mm in **right-left (RL)**, 10.7 mm in **SI**, and 8.8 mm in **AP** direction. In extreme cases, the motion was  $\geq 28$  mm for **SI** and **AP**. In [236], the trajectory of 21 lung tumours for 20 patients was investigated. A hysteresis was observed for ten tumours meaning that the trajectory was different for inspiration and expiration. The averaged trajectories of these tumours are illustrated in Fig. 1.2. Furthermore, the authors in [244] reported that the respiratory pattern and amplitude cannot be predicted by either gender, weight, age or height. Dietrich and Suh [56] differentiated between fiducials in the upper, middle, and lower lungs. Their results suggest a large variance of amplitudes for all lung segments. They reported that the highest variance was found in the lower lung in all three spatial directions with amplitude ranges between 1.5 – 14.0 mm for **SI**, 0.6 – 7.0 mm for **AP**, and 0.1 – 12.6 mm for **RL** (26 motion traces investigated).



**Figure 1.2:** Averaged trajectories of 21 lung tumours in the frontal plane (left) and the sagittal plane (right). The trajectories were acquired via fluoroscopy. Tumours attached to a bony structures are circled (picture from [236]).

**Liver:** Liver motion due to respiration was investigated in [107, 137, 257, 286]. In [137], Kitamura *et al.* investigated the motion of the 20 liver tumours with fluoroscopy. They reported an averaged amplitude of 4 mm (range 1 – 12 mm) in **RL**, 9 mm (range 2 – 19 mm) in **SI**, and 5 mm (range 2 – 12 mm) in **AP** directions. Analysing the 3D trajectory revealed a hysteresis for four tumours.

**Pancreas:** In [56], Dietrich and Suh investigated the motion of tumours in the pancreas for nine subjects. In total, 28 fractions were investigated. They reported a range from 0.2 – 9.4 mm in **SI**, 0.3 – 4.8 mm in **RL**, and 0.9 – 4.8 mm in **AP** directions. Similar results were reported by [99], where a motion range varying from 4.4 – 9.9 mm for seven patients was found.

### 1.1.2 Measuring Respiration

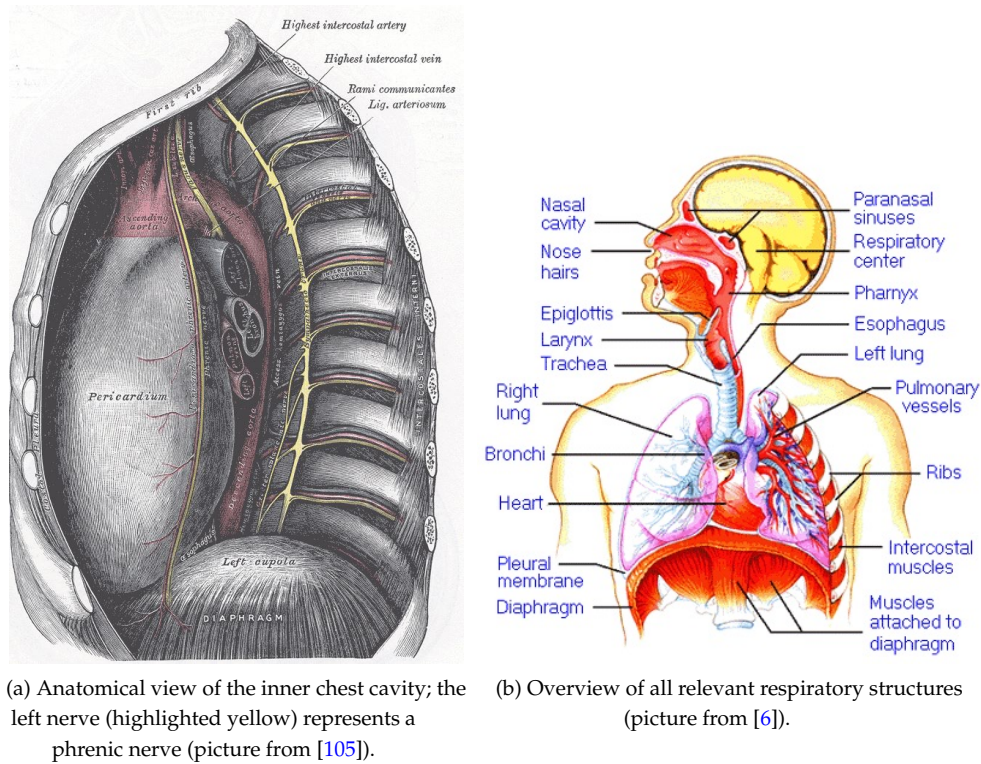
Measuring the respiratory activity of humans is essential for many medical diagnostic and therapeutic applications such as pulmonary function test for athletes, monitoring of patients in an intensive care units, investigation of obstructive sleep apnea or other chronic obstructive pulmonary diseases (e.g. bronchitis, emphysema). Consequently, various measuring techniques have been investigated for different purposes. They differ in, e.g., sensor placement, measuring principle / modality, **signal-to-noise ratio (SNR)**,

or invasiveness. In the following, a brief overview is provided on potential measuring locations for the purpose of respiratory motion compensation. The relevant structures are illustrated in Fig. 1.3.b. At each location, potentially relevant signals and measurement techniques are discussed. Thereby, the focus lies on signals which might be used to identify the current or prospective position in the respiratory cycle (e.g. change from inspiration to expiration) and/or contain information about the amplitude or pattern of the respiratory motion. The measurement techniques are discussed with respect to the SNR, the invasiveness, and if measurements are possible in real-time. Note that in this context, we define that real-time requirements are fulfilled if the breathing activity can be detected without averaging over multiple breathing cycles.

**Central nervous system:** In principle, it would be optimal to measure the initial impulse causing the respiratory motion at its source in the brain. The generator of the unconscious respiratory rhythm lies in the medulla oblongata, which is the lower part of the brain stem [107]. The precise function of this oscillator and its influencing parameters are still content of ongoing research. In [93, 225], the authors present evidence of the existence of two distinct respiratory oscillators controlling either the inspiratory or expiratory activity. However, a non-invasive, real-time measurement of the brain stem is not possible. The relevant regions are located too deep for standard surface **electroencephalography (EEG)** measurements and alternative techniques, such as **functional magnetic resonance imaging (fMRI)**, cannot be used for real-time measurements. However, it shall be mentioned that in [168] a significant difference in the **fMRI** activity was reported for volunteers who breathed either spontaneously or consciously.

In case of conscious respiration, evidence of an increased activity within the supplementary motor area and the primary motor cortex was reported in **fMRI** studies [87, 168]. Furthermore, Raux *et al.* [209] reported **EEG** potentials of up to 1.43 s before the onset of a new respiratory cycle. However, the **EEG** potentials were identified by averaging over 120 respiratory cycles. Consequently, no real-time measurements are possible.

**Neck:** The respiratory information is transmitted from the central nervous system to the diaphragm via the phrenic nerves. They originate from the spinal cord and exit the vertebral column between vertebral body C3 and C5. The phrenic nerves can be divided in a left and right nerve. So far, no study focusing on a direct measurement of this nerve could be found. However, several studies focused on the stimulation of the phrenic nerves via electric or magnetic fields [159, 169, 177] to investigate the transition time between the nerves and the contraction of the diaphragm which is between 5 – 7 ms.



**Figure 1.3:** Anatomical details of the respiratory system.

**Thorax and Abdomen:** The respiratory impulse of the nerves results in the contraction of the diaphragm and, if necessary, other supporting respiratory muscles. The contraction of the diaphragm can be measured via **electromyography (EMG)**. A high **SNR** ratio can be achieved by direct puncture of the diaphragm using **EMG** needle electrodes [29]. However, this invasive procedure can lead to a damage of the pleural space, which might result in a pneumothorax. To lower this risk, the use of ultrasound guidance was recently proposed [31]. A non-invasive measurement might be possible through **surface electromyography (sEMG)** electrodes. In [60, 148, 161, 201] different **sEMG** measurement techniques have been discussed and evaluated. It has to be considered that these recordings are affected by the activity of surrounding surface muscles, and the contribution of the deeper located diaphragm to the measured signal remains unclear. An alternative measurement position through a natural orifice is offered by the oesophagus [159, 246]. A literature review of **EMG** measurement techniques of the diaphragm is presented in [121].

The resulting respiratory motion can be measured internally and externally by various modalities. In case of external measurements, one frequently used technique is to

measure the movement of the skin. This can be done by optical tracking of individually placed markers or of the complete surface. One example for marker-based tracking is the **real-time position management (RPM)** system of Varian [75, 103, 126, 166] which is frequently used in clinical practise. Surface measurements can be performed either with time of flight cameras [182, 198, 230] or laser triangulation based methods [18, 229]. Alternatively, the expansion of the torso and abdomen can be monitored by using acceleration [17, 53, 120] or strain sensors [33, 90]. The latter are frequently used within sleep laboratories.

Acquiring information of the internal motion of specific structures and organs is a very important and wide research field. Different techniques have been used for this such as fluoroscopy [210, 266], 4D **computer tomography (CT)** scans [273, 275], **magnetic resonance imaging (MRI)** [24, 153] and **ultrasound (US)** [36, 85, 256]. Further studies and details will be discussed in sec. 2.3.

**Mouth and Nose:** The effect of respiration motion can be measured at the nose or mouth via different sensors. First, the respiratory pressure can be acquired by, e.g., differential pressure transducers [89, 124]. Second, the respiratory volume and flow can be measured by spirometry [176]. In general, these measuring techniques are non-invasive, have a fast response time and provide accurate measurement results. However in both cases, the patient has to breathe into a mouth piece, which decreases the patient comfort. Alternatively, the temperature difference in front of the nose can be measured via thermistors [90]. The electrical characteristics of these sensors change based on the surrounding temperature. In general, these sensors are small, robust, have high patient comfort and are relatively inexpensive. However, they often have a slow dynamic response and the electrical signal and airflow is correlated nonlinearly [90]. Consequently, they cannot be used for qualitative measurements of the magnitude and waveform of the respiratory flow.

Recently, thermal imaging has been used to evaluate the spatio-temporal waveforms of nostril breathing [92]. This technique offers a contact free measurement possibility.

**Further Locations:** The expansion of the lungs also affects the cardiac system. The time interval between successive R-R peaks of the heart rhythm as well as the mean systolic and diastolic pressure correlate with the respiratory rate [194]. In recent years, there has been an increased effort to non-invasively estimate the respiratory rate from pulse oximetry data in real-time [10, 130, 172]. However, the **SNR** is relatively low com-

pared to direct measurements of, e.g., the respiratory flow, as blood pressure and cardiac pattern fluctuations can be caused by various reasons. The measurements are commonly performed at the finger tips.

### 1.1.3 The Problem of Respiratory Motion Compensation

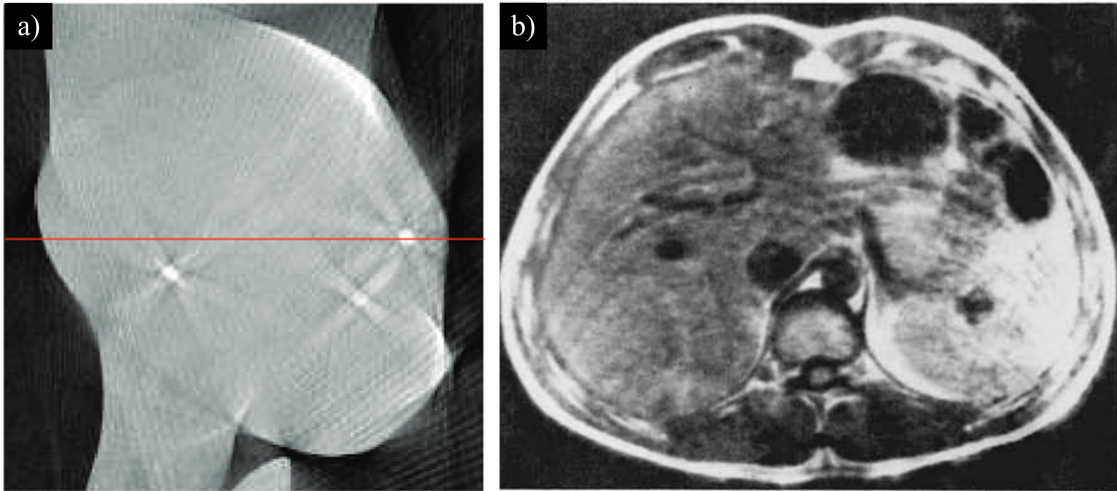
Investigating the problem from a broader perspective, respiratory motion can be classified as one source of involuntary motion. Most of these motions are controlled by the **vegetative nervous system (VNS)**. The **VNS** is part of the peripheral nervous system and is essential for maintaining homeostasis. It controls functions such as heart rate, respiration, blood pressure, and digestion. Further involuntary motions are caused by, e.g., the heart [212, 285] and the digestive system [260]. These movements can influence the outcome of diagnostic and therapeutic applications. For example, the effect of digestive motion was discussed in [260]. The authors investigated the motion and dosimetric consequences on prostate tumours based on the filling of the rectum and bladder. Another involuntary motion (which is not controlled by the **VNS**) is eye motion. Involuntary eye motion can influence the accuracy of, e.g., photorefractive laser surgery [180].

So far only involuntary motions of the patients have been discussed. In case of microsurgery, where a high accuracy of the surgeon is required, involuntary motions of the surgeon have to be considered [102, 217, 218, 247]. As discussed in [10], the physiological tremor motion of the hand of the surgeon results in a decreased positioning accuracy of the instruments and limits further miniaturisation.

All these motions differ in their signal characteristics such as amplitude of the motion and occurrence. While tremor motion tends to have a high frequency ( $f > 5$  Hz) [218], respiratory and cardiac movements occur at a lower frequency ( $0.1 < f < 5$  Hz). In general, digestive motions are often related with slow movements (on a day-to-day basis) depending on the filling of the gastrointestinal system [146].

In the past, there has been a strong research effort to compensate the effect of these motions for specific applications. In case of respiratory motion, one advantage is that it can be controlled up to a certain extent. This allows simple respiratory motion techniques such as breath holding or smoothing of the motion pattern by providing the patient with visual feedback [136]. However, these methods might not be suitable for all applications. Technically more advanced methods are based on adaptive motion compensation or even artificial respiration. An overview of respiratory motion models which can be used for adaptive motion compensation is given in [167]. Note, it could be argued that respiratory motion compensation is not required in case of artificial respiration and is consequently the preferred technique. However, artificial ventilation techniques only





**Figure 1.4:** a) **CT** motion artefacts of a moving heart phantom acquired with a C-arm scanner (phantom has three circular lesions (white dots); picture from [28]); b) **MRI** motion artefacts (picture from [16]). Both images were acquired without further motion compensation.

stabilize the breathing pattern and makes it more predictable. The motion is still present and compensation techniques such as gating will still be required. Furthermore, it has to be considered that artificial respiration requires anaesthesia which is an invasive procedure which can cause serious complications and requires further trained employees and equipment which increases the costs and the treatment time. The preferred compensation technique strongly depends on the specific application.

In the following, a list of medical applications will be presented whose outcome can be improved by some kind of respiratory motion compensation.

**Medical Imaging:** Internal structures in the thorax and/or abdomen can be visualized by different imaging technologies such as **CT**, **MRI**, **positron emission tomography (PET)**, and **US**. Depending on the acquisition time, all modalities are more or less affected by respiratory motion. The resulting artefacts can be broadly divided into motion during image acquisition and between images [94]. The first category leads to a decreased quality of the individual images such as blurring effects. The second group does not affect the image quality but increases the requirements on postprocessing such as the image registration.

Motion during the acquisition of **CT** images causes artefacts such as blurring, streaks, or discontinuities. However, as the acquisition time of one image is relatively low, respiratory motion artefacts can be compensated by simple breath holding [157]. If the or-

gan motion under normal breathing is to be investigated or the patient is not capable of breath-holding, different methods have been examined such as minimizing the acquisition time [216], respiratory gating [163, 178, 196], or using correction methods [157, 275].

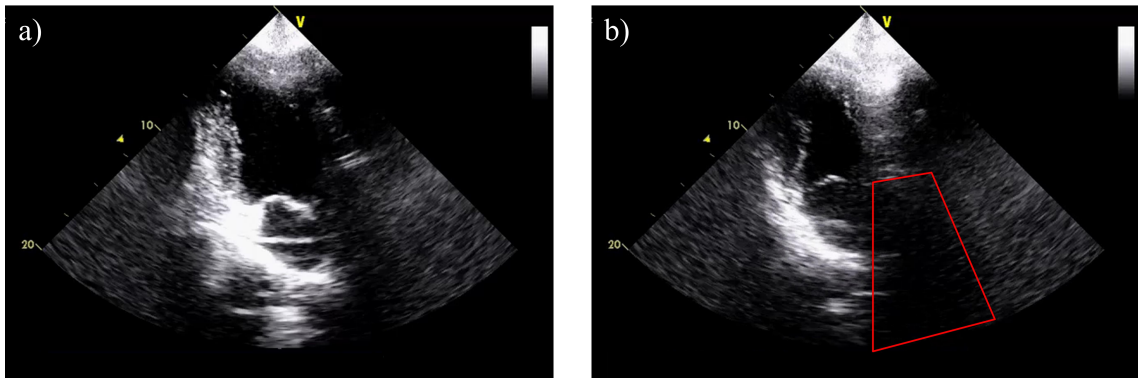
The problem of **CT** motion compensation becomes more challenging when changing from standard **CT** scanners used for diagnostic purposes to mobile C-arm scanners which are often used in image guided-interventions [28]. The acquisition time can be up to several seconds. The effect of uncompensated respiratory motion artefacts is shown in Fig. 1.4.a on a heart phantom with three lesions. Blurring and streaks are visible around the lesions.

In general, the acquisition time of **magnetic resonance (MR)** images is higher compared to **CT** images. Thus, **MR** images are more susceptible to motion artefacts. Furthermore, **MR** images, such as contrast-enhanced **MR** images, have acquisition times which makes the use of breath-holding techniques impossible [94]. Motion artefacts occur in the form of blurring effects or as ghost images of the moving structure [16, 162]. One example of a blurred **MR** image is shown in Fig. 1.4.b. Similar to **CT** imaging, the artefacts can be compensated by gating approaches or motion models. A detailed review of respiratory motion models can be found in [167].

In case of **PET**, respiratory motion leads to two effects [191]. First, the volume of the structure which is highlighted by a radioactive marker such as  $^{18}\text{F}$  fluorodioxycogluose is increased. Second, due to the increased volume, the **standard uptake value (SUV)** is reduced. In [191], a respiratory gating approach was investigated based on an external surrogate. The authors reported a 28 % decreased volume of lung tumours and a 56.5 % increase of the **SUV**. With the appearance of clinical **PET-CT** scanners, the motion-affected **PET** scans can be further corrected using **CT**-based attenuation maps [197].

In general, the acquisition time of ultrasound is fast enough that the images do not suffer from blurring effects. However, the appearance of motion might decrease the contact pressure between the skin surface and the **US** transducer. This can lead to decreased image quality. Furthermore, due to motion bones might appear in the field of view of the **US** transducer, which creates an **US** shadow for the regions beyond [193]. An example is shown in Fig. 1.5. Currently, robot-controlled approaches are investigated which enable an automatic tracking of the **US** transducer [145, 188].

**Oncology:** Another application area where respiratory motion compensation is essential is the precise treatment of tumours in the thorax or abdomen. Tumours which are located, e.g., in the lung or liver move due to respiratory and cardiac activity. To avoid damaging healthy tissue, respiratory motion compensation is required. Within this cat-



**Figure 1.5:** Example of the appearance of an US shadow within the field of view due to respiratory motion (image of the heart). Images a) and b) are acquired at the same position with a temporal difference of several seconds. Shadow is highlighted in red.

egory, the most important application is radiotherapy. Thereby, an external radiation beam is focused on the moving tumour. The aim is to damage the cell division process by a high radiation dose [272]. An overview of respiratory motion techniques for radiotherapy is provided in [131] and will be discussed in detail in sec. 1.2.

Alternative approaches are **radiofrequency (RF)** ablation techniques and **high-intensity focused ultrasound (HIFU)**. In case of HIFU applications, a lesion is generated by thermal ablation via focused ultrasound. Similar to radiotherapy, the technique is non-invasive. As the generation of a thermal lesion takes several seconds, respiratory motion has to be taken into account [50]. If it is not considered, two effects can occur. First, the volume of the lesion might be increased. Second, the temperature increase within the lesion is lower due to the increased surface. Consequently, not all tumour cells might be ablated. Approaches based on adaptive motion compensation have been investigated in [14, 108, 189, 215].

In RF ablation, a needle is inserted into the target area. At the tip of the needle, a lesion is generated by thermal ablation. In contrast to radiotherapy or HIFU, no motion compensation is required during the ablation process. However, the organ motion has to be considered during needle insertion. In general, the steering process is monitored by X-ray fluoroscopy or US [23]. Nonetheless, due to organ motion and deformation during the treatment, it might be difficult to find the preoperatively planned target position. Respiratory models are investigated to take these motions and deformations into account [167].

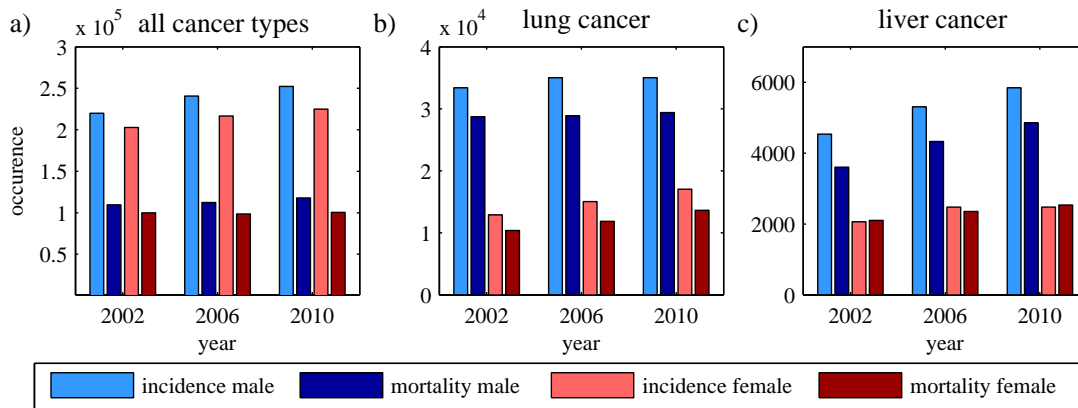
**Minimally Invasive Interventions:** The problem of precise needle placement mentioned above is only one example within the field of minimally invasive or image-guided interventions. In general, minimally invasive instruments such as ablation or biopsy needles are placed under image guidance (fluoroscopy, CT, or MRI), or by using optical or electromagnetic tracking systems. In all applications, respiratory motion has to be compensated to precisely reach the target area. Other applications are bronchoscopic interventions [98], atrial fibrillation procedures [34, 35], or biopsies [32, 164].

## 1.2 Motion Compensation in Radiotherapy

As this work focusses on motion compensation in radiotherapy, further details shall be presented about radiotherapy (sec. 1.2.1), adaptive motion compensation techniques and some of the clinically used systems (sec. 1.2.2). Nonetheless, it shall be emphasized that results of this work can be adapted to other medical applications.

### 1.2.1 Radiotherapy

According to the data of the Robert Koch-Institut (Berlin, Germany), there were approximately 477 300 people diagnosed with cancer in the year 2010 in Germany [4]. In the same year, approximately 218 258 people died due to cancer. This makes cancer the second most widely cause of death after cardiovascular diseases within Germany [7]. Figure 1.6 gives an overview of the absolute number of newly diagnosed cancer patients (light blue and red colours) and mortality (dark blue and red colours) for the years 2002 to 2010 in Germany. Figure 1.6.a points out that the number of newly diagnosed cancer patients seems to be gender unspecific and slowly increases. However, due to the intensive research effort within the field of cancer diagnostics and treatment, the number of people dying from cancer is approximately half the newly diagnosed patients. The mortality rate remains almost constant (for female) or increases only slowly (for male). Nonetheless, the treatment success strongly depends on the type of cancer. The absolute incidence and mortality numbers of lung (Fig. 1.6.b) and liver (Fig. 1.6.c) cancer indicate a low survival rate. This observation is confirmed by Table 1.1 which shows the gender specific prevalence and five year survival rate for 2010 of different cancer types. The prevalence rate is defined with regard to 100 000 people. It can be observed that the survival rates are low for liver, pancreas, and lung cancer compared to other cancer types such as prostate, uterus, or kidney. This is especially critical in case of lung cancer which has a relatively high prevalence rate. Consequently, there is an urgent need for improved treatments for these cancer types.



**Figure 1.6:** Gender specific absolute number of incidence and mortality for the year 2002, 2006, and 2010 in Germany for (a) all types of cancer, (b) lung, and (c) liver cancer (data of the Robert Koch-Institut, Berlin, Germany [4]).

Radiotherapy is one oncological treatment option among others such as chemotherapy or surgery [272]. The principle idea is to damage the cell division process by radiation. This will lead to a shrinkage of the tumour. However, this also affects healthy cells. Consequently, precise delivery of the radiation dose is essential for the treatment outcome. The radiation can be delivered using either external radiation beams or by internally implanted seeds or containers. The latter is referred to as brachytherapy and has the disadvantage that the seeds or containers have to be placed in or adjacent to the tumour, which is - in most cases - an invasive procedure. In contrast, using external beams allows a non-invasive treatment of tumours. In the following, we refer to external beam radiotherapy only as radiotherapy.

The radiation beam is generated by a **linear accelerator (LINAC)**. Electrons are accelerated and directed onto a target plate (often tungsten). The collision results in photons which are used as treatment beam. The size and shape of the beam can be controlled by a collimator. The radiation beams which enter the body linearly penetrate the tissue. By targeting the tumour from multiple angles, a hot spot can be generated within the tumour while, at the same time, the surrounding healthy tissue is spared. This is illustrated in Fig. 1.7. The total dose is often divided into fractions to which the patient is exposed in multiple sessions over days or weeks. The total dose and number of fractions depends on the type of cancer and size of the tumour. As the procedure is non-invasive, patients can be treated in special radiotherapy centres and do not have to stay in a hospital, which decreases treatment costs and increases the comfort of the patients.

Basically, radiotherapy can be applied to all kind of tumours anywhere in the body.

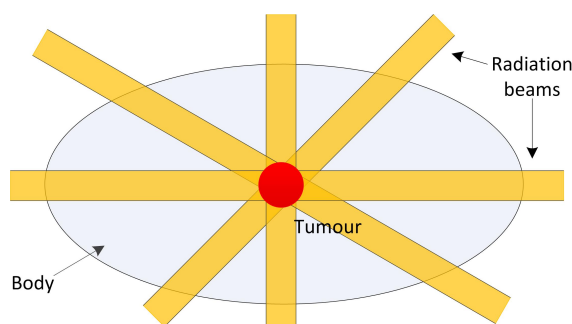
**Table 1.1:** Gender specific prevalence (per 100 000 people) and five year survival rate for 2010 of different cancer types in Germany. Data based on the Robert Koch-Institute [4]. (Very low survival rate are highlighted bold.)

Type of cancer	Prevalence rate		Survival rate [%]	
	Male	Female	Male	Female
Stomach	48.6	33.1	32	33
Colon	289.8	235.5	64	64
Liver	19	6.7	<b>15</b>	<b>12</b>
pancreas	18.1	17.7	<b>7</b>	<b>8</b>
Lung	122.2	65.2	<b>16</b>	<b>21</b>
Kidney	83.7	50.6	75	77
Non-Hodgkin lymphomas	70.4	60.8	66	67
Uterus	-	112.5	-	81
Prostate	695.6	-	93	-
All cancer types	1919.6	1808.9	61	67

The term stereotactic radiosurgery is used if external beam radiation therapy is applied to tumours in the brain and spine. For tumours within the rest of the body, the term **stereotactic body radiation therapy (SBRT)** is used. As discussed in sec. 1.1.3, respiratory motion has to be considered in these cases. Motion compensation techniques can be broadly divided into five categories [131]:

- increase of the target area,
- respiratory gating,
- breath-hold techniques,
- forced shallow-breathing, and
- adaptive motion compensation.

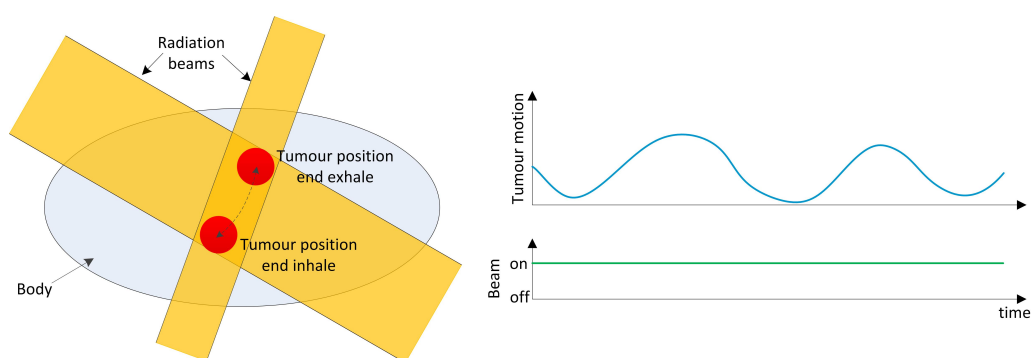
The disadvantages of the first four techniques are either irradiation of healthy tissue, increase of the treatment time, decrease of the patient comfort, or a combination of all three. Their principle ideas, advantages and disadvantages will be briefly discussed in the following. However, this work focuses on the technically more complex adaptive motion compensation which will be discussed separately in the next section (sec. 1.2.2).



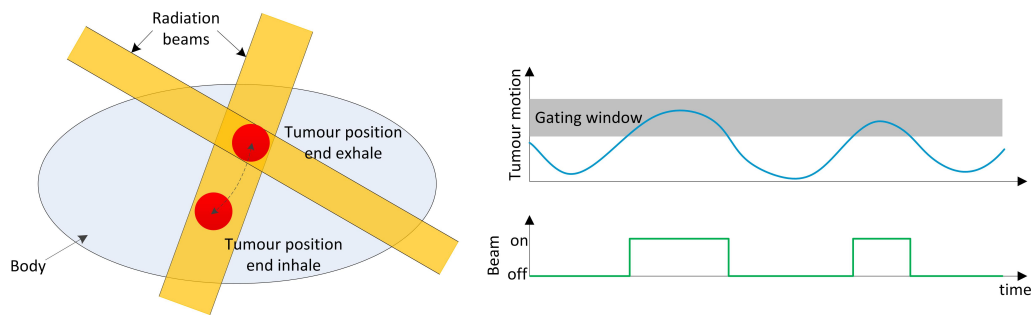
**Figure 1.7:** Irradiation of a tumour from multiple angles to generate a hot spot in the tumour, while surrounding healthy tissue is spared.

**Increasing the target area:** Increasing the target area is the simplest motion compensation technique and is illustrated in Fig. 1.8. Thereby, the target volume is increased to cover the complete motion range of the tumour. Note that the increase can be achieved either by increasing the size of the beam (as illustrated in Fig. 1.8) or using multiple smaller beams. This method does not require special hardware and allows the patient to breathe normally. The obvious drawback is that depending on the motion range a large amount of healthy tissue will be irradiated. This increases the chance of second primary cancer in the future [104, 179].

**Respiratory gating:** This technique is based on the observation that within a specific window (“gate”) during the respiratory cycle the tumour position is almost constant [243]. Therefore, the tumour motion has to be analysed in combination with an external surrogate such as spirometry [144] or external optical marker [100, 280] before treatment. As illustrated in Fig. 1.9, the radiation beam is only activated if the value of the external



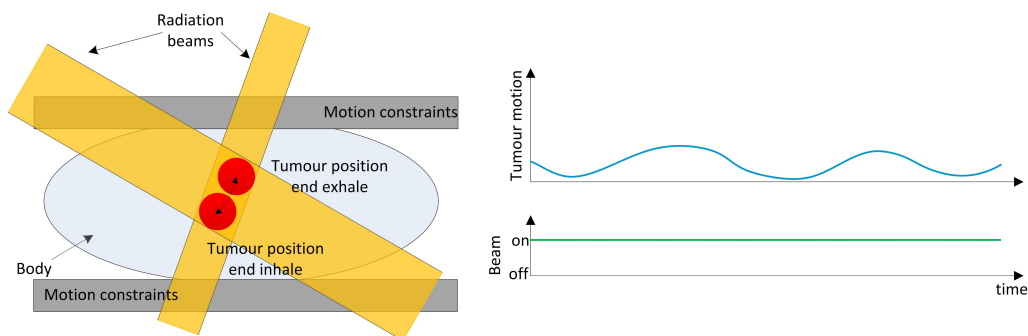
**Figure 1.8:** Illustration of motion compensation by increasing the target area; the target area is increased to cover the complete motion range of the tumour.



**Figure 1.9:** Illustration of respiratory gating; the beam is only activated if tumour position is within a predefined gating window.

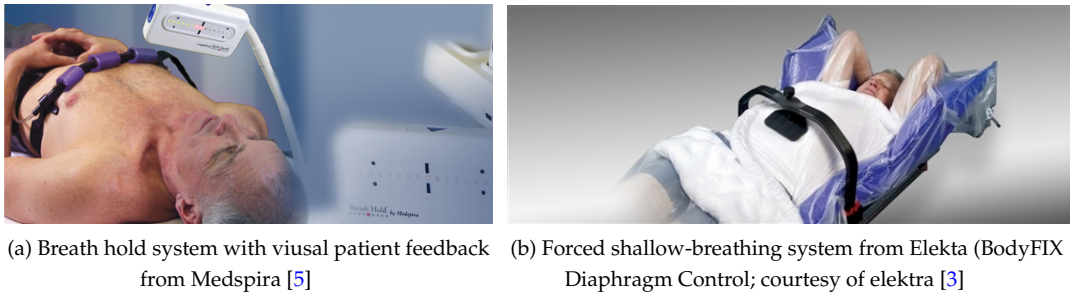
signal is within a predefined window. As a result, the treatment time will be increased. For each patient, the size of the gating window and residual tumour motion within this window has to be optimized to balance between treatment time and accuracy [264]. A commercially available product which provides additional visual feedback for the patient is shown in Fig. 1.11.a.

**Breath-hold techniques:** Similar to respiratory gating methods, breath-holding techniques also divide the treatment time in “beam on” and “beam off” times. However, the patient receives an acoustic or visual feedback which indicates that he has to hold his breath at a certain position [42, 135, 141]. The beam is only activated at these positions. Consequently, the treatment outcome depends on the active cooperation of the patient. Ideally, patient coaching has to be performed before treatment. This results in further increase of the treatment time.



**Figure 1.10:** Illustration of forced shallow-breathing; the potential motion range is limited due to motion constraints such as an abdominal press.





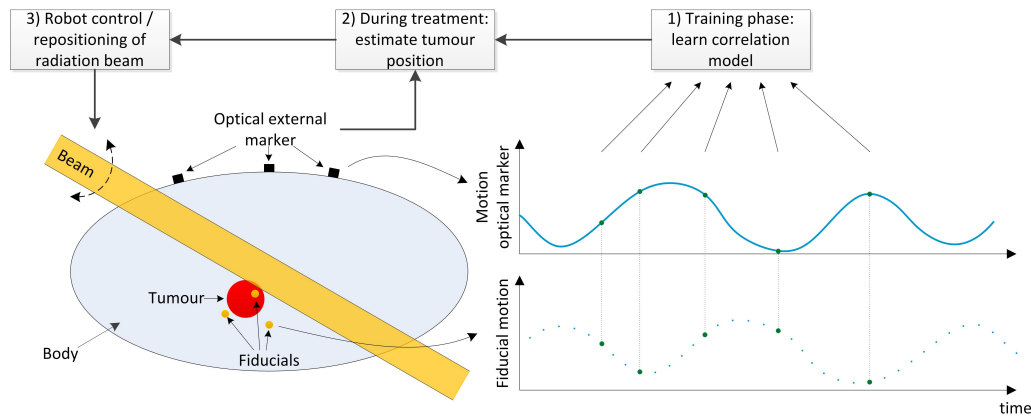
**Figure 1.11:** Clinically used motion compensation systems.

**Forced shallow-breathing:** An alternative approach is forced immobilization of the patient or at least relevant parts such as the abdomen [154, 190]. Negoro *et al.* [190] demonstrated that an abdominal press can be efficiently used to reduce the motion range of lung tumours from 8 – 20 mm to 2 – 11 mm on average. The idea is illustrated in Fig. 1.10 and a clinically used system is shown in Fig. 1.11.b. The drawback of such systems is a decreased patient comfort and that possibly not the complete tumour motion can be suppressed. Compared to gating and breath-hold methods, here, a balance between patient comfort and treatment accuracy has to be found.

### 1.2.2 Adaptive Motion Compensation and Clinical Systems

All previously discussed motion compensation techniques have disadvantages with respect to either treatment time, accuracy, or patient comfort. Adaptive motion compensation offers the possibility to overcome all these drawbacks. The basic idea is to track the tumour position and accordingly adapt the beam position in real-time.

In clinical practise the tumour position is tracked indirectly via multiple gold fiducials. These markers are placed before treatment in or adjacent to the tumour. The positions of the fiducials can be detected automatically by stereoscopic *X-ray* imaging during treatment and are further used to compute the position of the tumour. Ideally, the fiducial positions would be acquired in real-time. However, this would lead to a substantial additional radiation dose for the patient. Shirato *et al.* [243] reported that the diagnostic *X-ray* imaging system used in their phantom experiments caused  $0.208 - 21.48 \cdot 10^{-3}$  Gy for an exposure time of 2 min. An alternative is the estimation of the internal marker position based on the information of external surrogates, such as optical markers, which are placed on the torso or abdomen [231, 232]. The external surrogates can be sampled at a high frequency without additional harm to the patient. At the start of each treatment session, multiple internal and external samples are acquired simultaneously such that a



**Figure 1.12:** Adaptive motion compensation based on external surrogates: 1) Learning of a correlation between data of internal fiducials and external optical markers; 2) Estimation of the internal tumour position based on data of the external markers during treatment; 3) Estimated tumour position is used to reposition the beam position.

complete respiratory cycle is covered. Often 5 – 12 points are sampled in clinical practise [227] which are further used to train a correlation model. During treatment, only external data is acquired and the internal marker position is estimated using the correlation model. The complete process of learning the correlation between internal and external data and the repositioning of the radiation beam are illustrated in Fig. 1.12. Frequently used correlation models assume a linear or quadratic relationship between the data of internal and external markers. As the correlation might change with varying breathing patterns, further internal points are acquired during the treatment to update the correlation model. A detailed overview of studies focusing on the correlation between external and internal markers and investigated correlation models is presented in sec. 2.3.2.

If the current tumour position is known, the beam position has to be corrected accordingly. Different technical solutions have been proposed for this. They can be broadly divided into moving the patient and shaping or realignment of the radiation beam. The first can be achieved by using a Hexapod based robotic patient couch (Fig. 1.13.a) [39, 58, 59, 109, 276]. Instead of realignment of the beam, the patient is repositioned. Note some patients might get motion sick from this compensation technique [258].

Representatives of the second category are **multileaf collimators (MLCs)** [132, 143, 170, 192, 200], the CyberKnife<sup>®</sup> (Accuray Inc., Sunnyvale, CA, USA) [134, 227], and the Vero System (BrainLAB AG, Feldkirchen, Germany, and Mitsubishi Heavy Industries, Tokyo, Japan) [54, 55, 181]. In case of MLCs, a collimator consisting of multiple leaves is placed in front of the LINAC (Fig. 1.13.b). The individual leaves can be repositioned in real-

time to shift the beam as well as to modify its shape. In case of the CyberKnife<sup>®</sup> system, a light-weight LINAC is mounted at the end of an effector of an industrial robotic arm (Fig. 1.13.c). Based on the current tumour position, the robot can be controlled to reposition the LINAC. Furthermore, the LINAC can be moved on a sphere around the patient. This increases the possible angles to irradiate a tumour compared to conventional gantry-based systems. The beam of the VERO system can be repositioned by a combination of an MLC and two orthogonal gimbals which allow the LINAC and MLC to be tilted (Fig. 1.13.d).

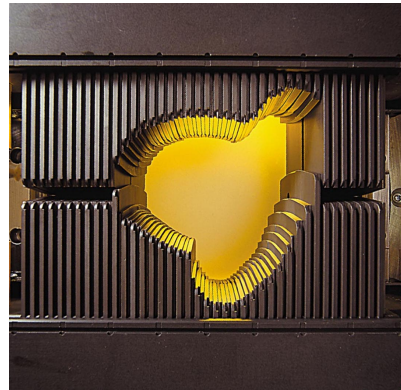
Recently, first results were published using an MR guided radiation therapy system [117, 187], the ViewRay<sup>™</sup> system (ViewRay, Oakwood Village, Ohio, USA). The MRI scanner and three static cobalt radiation sources share the same isocenter. The movement of the tumour can be monitored in real-time and radiation beams are shaped using three individual MLCs (Fig. 1.13.e). An alternative combination consisting of CT imaging and radiation therapy systems offers the TomoTherapy<sup>®</sup> (Accuray Inc., Sunnyvale, CA, USA) (Fig. 1.13.f) [156, 158]. Here, motion compensation is performed by binary MLCs.

All repositioning systems have in common that they suffer from time latencies. As investigated in [213], the delays are caused by data acquisition and processing, and kinematic limitations of the specific system. The time latencies vary from 50 ms for the VERO system [55], 115 ms for the CyberKnife<sup>®</sup> [134] up to several hundred milliseconds for robotic patient couches [39, 58, 276] and MLCs [200]. For example, the MLC based TomoTherapy<sup>®</sup> has a time latency of < 300 ms (including the imaging time) [158]. These time latencies can be compensated by predicting either the internal or the external marker position. Therefore, a second model is required to which we refer to as prediction model or prediction algorithm. A literature review on prediction algorithms for the purpose of radiotherapy is presented in sec. 2.3.1.

Adaptive motion compensation is the technically most demanding method compared to previously presented techniques (sec. 1.2.1). It has to be considered that the output of a new prediction and correlation algorithm has to be computed in real-time. In case of the latest CyberKnife<sup>®</sup> version, the sampling frequency of the external surrogates is  $f_s = 26$  Hz. This means that the outputs have to be computed within  $t = 38.5$  ms. Furthermore, several aspects are still content of ongoing research, such as the number, modality and optimal placement of the external surrogates or robustness of the method in case of unexpected irregular breathing patterns.



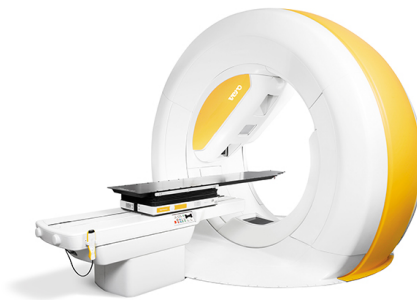
(a) Six degrees of freedom robotic patient couch; HexaPOD™ evo RT system of Elekta (courtesy of elektra [3])



(b) Millennium™ MLC with 120 leaves of Varian (Image courtesy of Varian Medical Systems, Inc. All rights reserved. [8])



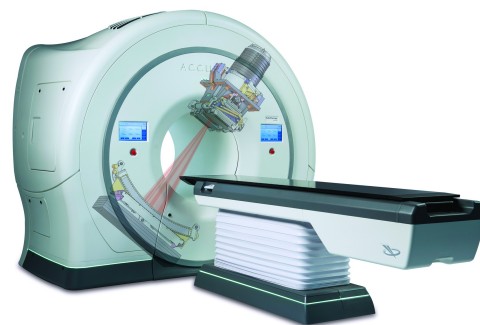
(c) CyberKnife® of Accuray Inc. [1]



(d) VERO SBRT system of BrainLAB AG [2]



(e) Combination of MR imaging and radiotherapy; ViewRay™ system [9]



(f) Combination of CT imaging and radiotherapy; TomoTherapy® H™ [1]

**Figure 1.13:** Clinically used adaptive motion compensation systems

## 1.3 Purpose of this Work

The focus of this work lies on the improvement of the accuracy and robustness of adaptive motion compensation techniques. Thereby, two particular aspects will be addressed which we refer to as question one (Q.1) and question two (Q.2). First (and primarily), the potential of multivariate respiratory motion compensation techniques is investigated (Q.1). As it was illustrated in sec. 1.1.2, respiratory activity can be measured at various locations by different sensor modalities. The data of these sensors contain different information of the same underlying biological process. An efficient integration of different sensor information might result in more accurate motion compensation algorithms.

Second, real-time feedback systems are explored which indicate the current accuracy or the “certainty” of the algorithm in its predictions (Q.2). Respiratory motion patterns have a high variability as discussed in sec. 1.1.1. Slow changes of the respiration frequency or amplitude as well as unexpected breathing irregularities such as coughing or sneezing are to be expected. One of the simplest feedback systems is to monitor the current prediction or correlation error. In general, such systems can be used to increase the robustness of motion compensation techniques throughout the treatment such as to either stop the treatment or to adapt the motion compensation technique. Here, the focus will be on alternative feedback systems which do not rely on the current error.

At first sight, these two aspects seem relatively independent from each other. However, within the methodical sections of this work, probabilistic machine prediction and correlation algorithms will be investigated. This group of algorithms makes the investigation of alternative real-time feedback systems in addition to the first aspect obvious.

In this work, all investigations are performed on the example of adaptive respiratory motion compensation techniques within the field of robotic radiotherapy. In particular, the CyberKnife<sup>®</sup> will be often used as an example. The choice was motivated as radiotherapy is a major application area which can benefit most from increased respiratory motion techniques (sec. 1.1.3). Furthermore, the Institute for Robotics and Cognitive System has gathered a lot of expertise with the CyberKnife<sup>®</sup> system in recent history. However, it has to be highlighted that the results are not restricted to this application (or to the CyberKnife<sup>®</sup>) and might be beneficial for other applications such as medical imaging.

To increase the structure of this work, we pose multiple questions which will be addressed in the following chapters. The questions are divided into two groups representing the two aspects investigated in this work:

**Table 1.2:** Main questions of this work and overview in which chapter they will be addressed.

<b>Q.1</b> Is it possible to improve respiratory motion compensation by using multivariate external surrogates?	
<b>Q.1.1</b> Which algorithms can be used for modelling multivariate data? Which general properties do algorithms need to have for multivariate prediction and correlation?	chapters 4, 6
<b>Q.1.2</b> How to select the most relevant and least redundant markers?	chapters 2, 4 - 6
<b>Q.1.3</b> Can the accuracy of respiratory motion prediction be increased using a multivariate external sensor setup?	chapters 2, 4 - 6
<b>Q.1.4</b> Can the accuracy of respiratory motion correlation be increased using a multivariate external sensor setup?	chapters 2, 4, 6
<b>Q.1.5</b> What are the most relevant sensors or sensor combinations?	chapters 2, 4 - 6
<b>Q.2</b> Is it possible to increase treatment accuracy through a real-time feedback about current prediction errors?	
<b>Q.2.1</b> What criteria could be used to evaluate current prediction accuracy?	chapters 2, 3, 6
<b>Q.2.2</b> What limitations do these criteria have?	chapters 3, 6
<b>Q.2.3</b> How can these criteria be used to control the current prediction error?	chapters 3, 6

Within the questions, the terms multivariate and real-time play a central role. Therefore the terms are further specified.

We refer to multivariate surrogates as the use of multiple sensors. This means that the measuring setup can consist of either multiple univariate sensors (all sensors measuring the same modality), multiple sensors acquiring data of different modalities, or a combination of both.

The term real-time in **Q.2** depends on the specific treatment system and its sampling frequency. In the case of the CyberKnife<sup>®</sup>, external data are acquired at a sampling frequency of  $f_s = 26$  Hz. The computations have to be performed within  $t = 1/f_s \approx 38.5$  ms. A real-time feedback for the CyberKnife<sup>®</sup> will give a feedback every  $t \approx 38,5$  ms.

## 1.4 Organisation

In the following the structural outline of this work is briefly presented. Note, at the end of each chapter, the main results of the particular chapter will be discussed with respect to the questions proposed in Table 1.2. The second column of Table 1.2 provides an overview which questions will be discussed in which chapters.

**Chapter 2:** This chapter provides the basic mathematical notation which is used throughout this work to investigate motion compensation algorithms. Furthermore, a detailed literature review is presented of frequently used evaluation measures (sec. 2.2) and prediction and correlation algorithms (sec. 2.3). Concluding, the algorithms are discussed with respect to multivariate extensions and probabilistic approaches.

**Chapter 3:** Two probabilistic algorithms are introduced which are based on **relevance vector machine (RVM)** and **Gaussian process (GP)** models. In this chapter, both algorithms are investigated for the purpose of respiratory motion prediction using univariate data to allow a comparison to already existing prediction algorithms. First, the mathematical background of both algorithms is presented and algorithm-specific parameters are evaluated on three representatively chosen motion traces (sec. 3.2). Second, the **RVM** and **GP** algorithms are evaluated on a dataset consisting of 304 motion traces for various prediction latencies in sec. 3.3. The results are compared to the **wavelet-based LMS (wLMS)**, the **multi-step linear method (MULIN)**, and the **support vector regression (SVR)** algorithms. The output of probabilistic prediction algorithms is a probability distribution with a predicted mean and variance. In sec. 3.4, the predicted variance is evaluated for the purpose of compensating prediction errors during the treatment. The investigation includes the analysis of characteristic variance patterns (sec. 3.4.1) as well as the evaluation of variance-based hybrid algorithms (sec. 3.4.2).

**Chapter 4:** This chapter focuses on the question if the accuracy of prediction algorithms can be increased by using multivariate external data. For this purpose, a multivariate measurement was performed which included modalities such as respiratory flow, acceleration and strain (sec. 4.1). Mathematical details to extended motion prediction algorithms to multivariate input data are discussed in sec. 4.2. Furthermore, algorithms are introduced to select the most relevant and least redundant data. Two experiments were performed focusing on the correlation between sensors and on the accuracy of multivariate extended prediction algorithms. In the latter experiment, the **RVM** algorithm and four non-probabilistic algorithms were investigated. The experiments are described

in sec. 4.3 and their results are presented and discussed in sec. 4.4. The chapter concludes with an outlook on further potentially relevant respiratory motion features based on EEG, electrooculography (EOG), and sEMG data (sec. 4.5).

**Chapter 5:** The extension of the GP algorithm from univariate to multivariate data is more challenging and is the focus this chapter. One possibility is the use of multi-task Gaussian Process (MTGP) models. After the presentation of the datasets used in the chapter (sec. 5.1), the mathematical background of MTGP is provided (sec. 5.2). The MTGP framework is very flexible and different properties are illustrated on synthetic examples. As it turns out, the MTGP approach makes an efficient combination of the prediction and correlation problem possible. An MTGP racking algorithm which combines prediction and correlation is presented in sec. 5.2.4 and alternative approaches based on separate prediction and correlation algorithms are discussed in sec. 5.2.5. In sec. 5.3, three experiments are outlined which focus on the correlation between internal and external data, on the number of internal training points, and on multivariate extensions. The results and discussion of the three experiments are presented in sec. 5.4. Even though the MTGP approach shows on average a superior performance compared to the alternative approaches, the results point out situations where MTGP does not work optimally. Further extensions to overcome these limits are presented in sec. 5.5.

**Chapter 6:** The main results are summarized in chapter 6. Section 6.1 presents the main findings of the two probabilistic algorithms investigated for the purpose of motion prediction and correlation. The following two sections discuss the results with respect to the main questions Q.1 (sec. 6.2) and Q.2 (sec. 6.3). Furthermore, the advantages of the MTGP approach are summarized and discussed with respect to radiotherapy (sec. 6.4) and in general for alternative applications (sec. 6.5).



## 2 Literature Review

This chapter provides a general overview of the existing literature within the field of respiratory motion prediction and correlation. Even though this is a relevant topic in different research areas such as medical imaging or microsurgery, the focus of this review lies on adaptive radiotherapy. In general, this research field is relatively new. Some of the first publications are by Adler *et al.* [11] and Schweikard *et al.* [231], which propose a new method to compensate for tumour motion based on external surrogates.

In sec. 2.1, general definitions and notations are introduced to describe respiratory motion. Afterwards, sec. 2.2 gives an overview of frequently used error measures and discusses their limitations. The main focus of this chapter is presented in sec. 2.3, a review of prediction and correlation algorithms. Closing remarks and consequences for this work are addressed in sec. 2.4.

### 2.1 Basic Notation

Let us assume that the data of a sensor is equidistantly sampled at a sampling frequency  $f_s$ . As an example, the measured signal could be the position of an optical marker or the voltage between two electrodes. Let  $t$  be the time, given in seconds, and  $y$  be the value of a sensor. The unit of  $y$  depends on the specific sensor. Further, we assume that  $t_i \in \mathbb{R}$  denotes the time at the time index  $i$ , i. e.  $t_i = i/f_s$ . The signal value at  $t_i$  is denoted by  $y_i \in \mathbb{R}$ . Different sensors are indicated by a superscript index or label. Therefore, the internal position of a fiducial could be expressed by  $y^{int}$  and the external position of an optical marker by  $y^{ext}$ .

In general, we refer to algorithms which could be used to perform adaptive respiratory motion compensation as **motion compensation (MC)** algorithms. The predicted value of an **MC** algorithm is denoted by  $y^*$ . These algorithms consist of two sub-algorithms, which we specify as *prediction* and *correlation* algorithms. The algorithms are used in sequence, meaning that the output of one algorithm is used as input for the other algorithm. Thereby, the order is arbitrary. In case of respiratory prediction algorithms the aim is to predict a future position  $y_{i+\xi}$  based on previous observations. Here,  $\xi$  is the latency index.

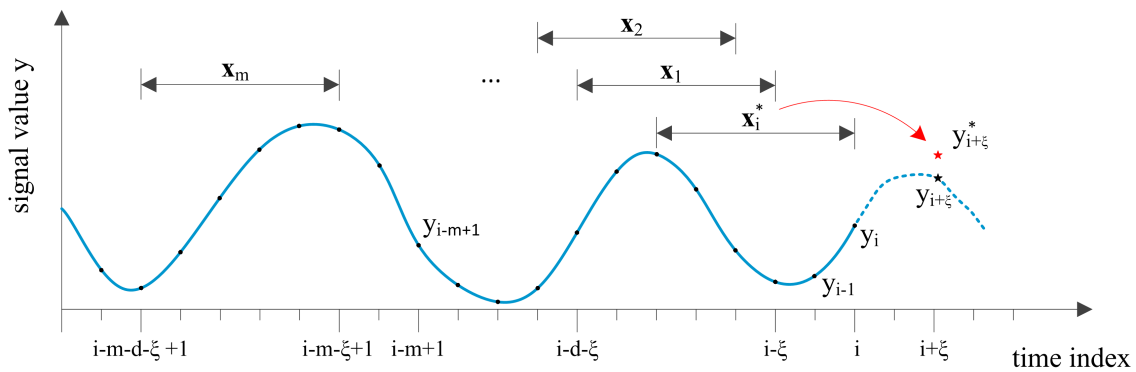
The prediction latency expressed in seconds is defined as  $h = \xi/f_s$ . Within the literature, the prediction latency is also referred to as response time [266], lag time [123], look ahead time [219] or prediction latency [63]. The output of a prediction algorithm at index  $i$  is denoted by  $y_{i+\xi}^*$ . In contrast to prediction algorithms, the aim of correlation algorithms is to estimate an internal position, such as the position of a tumour or a fiducial, based on the data of external surrogates at the same time step. Consequently, the output of a correlation algorithm at index  $i$  is  $y_i^*$ .

In general, prediction and correlation algorithms are based on a function  $f$

$$y = f(\mathbf{x}) \tag{2.1}$$

which allows the calculation of an output  $y \in \mathbb{R}$  based on an input feature  $\mathbf{x} \in \mathbb{R}^{d \times 1}$  where  $d$  is the dimension of the feature space. As it is common within the field of machine learning, we may also refer to the outputs  $y$  as label. The input features  $\mathbf{x}$  can be previous observations  $y$  or time points  $t$ , depending on the algorithm. To compute an output of a MC algorithm a training and test phase is required. Within the first phase, a training dataset  $\mathbf{T} = \{\mathbf{X}, \mathbf{y}\}$  is constructed with  $\mathbf{X} = [\mathbf{x}_1^\top, \dots, \mathbf{x}_m^\top]^\top \in \mathbb{R}^{m \times d}$  and  $\mathbf{y} = [y_{q_1}, \dots, y_{q_m}]^\top \in \mathbb{R}^{m \times 1}$ . Here,  $m$  denotes the number of training pairs,  $\mathbf{x}_j = [x_{j,1}, \dots, x_{j,d}]^\top \in \mathbb{R}^{d \times 1}$ , and  $q_1$  to  $q_m$  refer to time indices of the training labels  $\mathbf{y}$ . If the function  $f$  is learned, a prediction can be computed in the test phase for an unknown label  $y$ . The outcome of a MC algorithm is  $y^*$  based on the test feature  $\mathbf{x}^*$ . Note, to differentiate a training and test feature, the test feature is label by an asterisk.

Figure 2.1 illustrates the notations on an example for a prediction algorithm. The blue



**Figure 2.1:** Basic notation illustrated using the example of a prediction algorithm. Looking at the time index  $i$  with  $y_i$  being the most recent observation, the aim is to predict  $y_{i+\xi}$  based on the test features  $\mathbf{x}_i^*$ . The output of the prediction algorithm is  $y_{i+\xi}^*$  based on the training feature  $\mathbf{X} = [\mathbf{x}_1^\top, \dots, \mathbf{x}_m^\top]^\top$  and the training labels  $\mathbf{y} = [y_i, \dots, y_{i-m+1}]^\top$ .

solid line represents the signal which was sampled at time points indicated by the black dots. It is assumed that the algorithm uses previously observed signal values as input features and that the training set consists of the  $m$  most recent training pairs. The resulting training set would be:

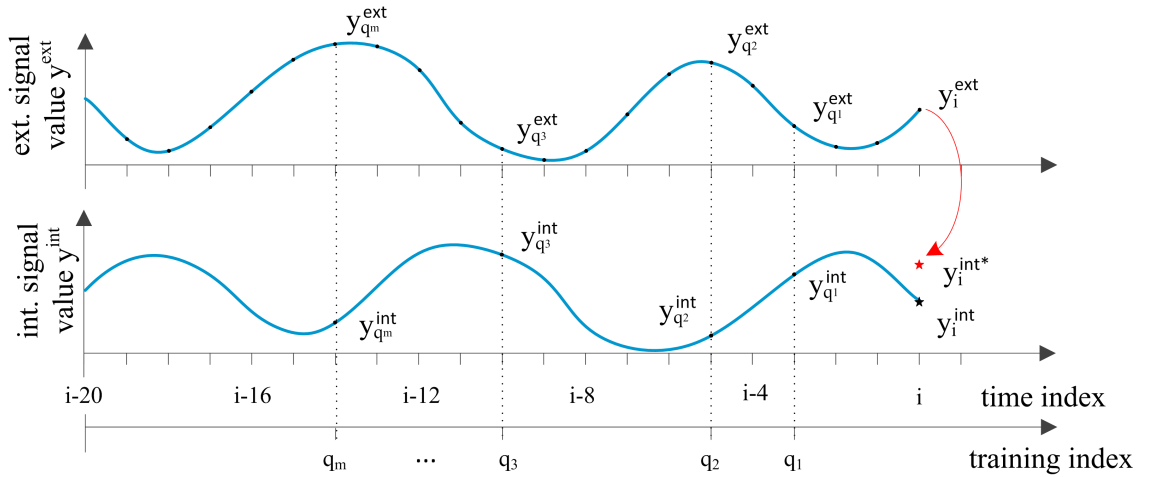
$$\mathbf{X} = \begin{bmatrix} y_{i-\xi} & y_{i-\xi-1} & \dots & y_{i-\xi-d} \\ y_{i-\xi-1} & y_{i-\xi-2} & \dots & y_{i-\xi-d-1} \\ \vdots & & \ddots & \vdots \\ y_{i-\xi-m+1} & y_{i-\xi-m} & \dots & y_{i-\xi-d-m+1} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_i \\ y_{i-1} \\ \dots \\ y_{i-m+1} \end{bmatrix} \quad (2.2)$$

In the illustration, the latency index is set to  $\xi = 2$  and the dimension of the feature space to  $d = 5$ . The aim is to compute  $y_{i+\xi}$ . The output of the prediction algorithm is  $y_{i+\xi}^*$  based on  $\mathbf{x}_i^*$ .

Alternatively, prediction algorithm might use the time  $t$  as input feature. In this case, the feature dimension would be  $d = 1$  and the training set would be:

$$\mathbf{X} = \begin{bmatrix} t_i & t_{i-1} & \dots & t_{i-m+1} \end{bmatrix}^\top, \quad \mathbf{y} = \begin{bmatrix} y_i & y_{i-1} & \dots & y_{i-m+1} \end{bmatrix}^\top. \quad (2.3)$$

An illustration of a typical scenario for a correlation algorithm is shown in Fig. 2.2. The two blue solid lines represent the external and internal signal. The external signal is equidistantly sampled. However, data of the internal signal is acquired only at a few



**Figure 2.2:** Basic notation illustrated at the example of a correlation algorithm. Looking at the time index  $i$  with  $y_i^{ext}$  being the most recent external observation, the algorithm computes  $y_i^{int*}$  based on the test feature  $\mathbf{x}_i^* = [y_i^{ext}]$ . The prediction algorithm is determined using the training feature  $\mathbf{X} = [x_1, \dots, x_m]$  with  $x_j = [y_{q_j}^{ext}]$  and the training labels  $\mathbf{y} = [y_{q_1}^{int}, \dots, y_{q_m}^{int}]^\top$ .

time instances which are marked in the example by  $q_1$  to  $q_m$ . It is assumed that the feature space dimension is  $d = 1$  and the training features are specified as  $x_j = y_{q_j}^{ext}$  for  $j = 1, \dots, m$ . The training set  $\mathbf{T}$  for a correlation algorithm would be:

$$\mathbf{X} = \begin{bmatrix} y_{q_1}^{ext} & y_{q_2}^{ext} & \dots & y_{q_m}^{ext} \end{bmatrix}^\top, \quad \mathbf{y} = \begin{bmatrix} y_{q_1}^{int} & y_{q_2}^{int} & \dots & y_{q_m}^{int} \end{bmatrix}^\top. \quad (2.4)$$

At time instance  $i$ , the objective is to predict  $y_i^{int}$ . The output of the correlation algorithm is  $y_i^{int*}$  based on the test feature  $\mathbf{x}_i^* = y_i^{ext}$ .

## 2.2 Evaluation Measures

To evaluate the outcome of an **MC** algorithm, the predicted signal  $y^*$  can be retrospectively compared to the true signal  $y$ . Depending on the evaluation measure used, different characteristics of the predicted signal can be analysed, such as prediction accuracy or smoothness of the signal [76]. Reviewing the literature reveals several different evaluation measures which will be briefly described within the following section. As the majority of the criteria focuses on prediction accuracy, they are divided into two categories - one focusing on accuracy measures and the other on the auxiliary measures. The latter might be the smoothness of the computed signal or the computation time.

### 2.2.1 Prediction Accuracy Measures

To compare the results of **MC** algorithms with respect to the measured data at time index  $i$ , we define the error  $e_i = y_i - y_i^*$ . In case of prediction algorithms, this is called the prediction error and in case of correlation algorithms the correlation error.

The simplest evaluation measure considering the complete signal is the **mean error (ME)**, which is defined as the mean of the error  $e_i$

$$\text{ME} = \frac{1}{N} \sum_{i=1}^N e_i = \frac{1}{N} \sum_{i=1}^N (y_i - y_i^*), \quad (2.5)$$

where  $N$  is the number of investigated points. Overall, this evaluation measure is rarely used [38, 242] as it has the drawback that positive and negative errors can compensate each other and lead to a small **ME**. This effect can be compensated by using the **mean absolute error (MAE)**

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - y_i^*|, \quad (2.6)$$

which has been widely used [114, 122, 152, 203, 222, 242, 278]. Each error contributes linearly to the resulting MAE value. However in practical applications, large prediction errors, even if they appear only for a short time instance, often have larger impact on the treatment than small prediction errors. One possibility to express large prediction errors is the maximum absolute error (mAE).

$$\text{mAE} = \max_i (|y_i - y_i^*|), \quad i = 1, \dots, N, \quad (2.7)$$

which was used in [118, 206, 242]. The obvious disadvantage of this measure is that the prediction or correlation accuracy is only evaluated at one time instance and all residual time points are neglected. Consequently, this criterion should only be evaluated in combination with another accuracy criterion, which considers the complete signal.

An alternative possibility to weight large errors stronger is the root mean square error (RMSE), which also considers the complete error signal. It is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_i^*)^2}. \quad (2.8)$$

This measure is the most commonly used accuracy measure [39, 48, 81, 86, 103, 118, 122, 126, 128, 142, 160, 203, 206, 211, 214, 221–223, 226, 241, 249, 266].

All evaluation measures presented so far are criteria which can be used to analyse data of one motion trace. If multiple motion traces, such as data from different subjects, should be analysed, the average of the individual MAE or RMSE values could be computed. In general, this cannot be recommended as each individual measure is affected by the amplitude of the motion. To allow for this, the MAE and RMSE can be normalized. In case of RMSE, mainly two normalisations have been used:

1. Murphy *et al.* [184] proposed a normalisation coefficient based on the standard deviation  $\sigma$  of the true signal. The normalised root mean square error (nRMSE) is defined as

$$\text{nRMSE} = \frac{\text{RMSE}}{\sigma} = \frac{\text{RMSE}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}}, \quad (2.9)$$

with  $\bar{y}$  being the mean of the true signal. This normalization was also used in [115, 123, 126, 183, 185, 186, 219]. The advantage of this normalization is that it can be used for prediction and correlation algorithms.

2. An easily interpretable normalisation for prediction algorithms is the approach presented by Ernst *et al.* [81], to which they referred to as relative root mean square

error ( $RMSE_{rel}$ )

$$RMSE_{rel} = \frac{RMSE}{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_{i-\xi})^2}}. \quad (2.10)$$

The  $RMSE$  is normalised by the  $RMSE$  for the case of no prediction, which adds a simple interpretation of the results. A  $RMSE_{rel} = 100\%$ , indicates that the prediction accuracy over all samples  $N$  is the same compared to “no prediction”. A  $RMSE_{rel} < 100\%$  means that the prediction accuracy is improved compared to no prediction. This normalization has been frequently used [63, 68, 77–79, 82, 84, 86, 142]. However, it is limited to prediction algorithms since  $\xi$  has to be larger than zero.

Less frequently used normalization coefficients are based on the range [118] and the average amplitude [166] of the true signal.

Besides these accuracy measures, the temporal distribution of the errors or its histogram can be of interest. This can give insight if the error has for example a certain periodic behaviour. The latter is especially relevant within respiratory motion compensation, as increased error is frequently observed around the maximum inhalation or exhalation. Besides simple distribution measures based on the standard deviation [86, 114, 160, 203, 210, 242] or confidence intervals [203] of the error, Ernst *et al.* [76] presented a more detailed analysis based on the frequency spectrum of the error.

## 2.2.2 Auxiliary Measures

Several publications investigate the potential use of  $MC$  algorithms for gated radiotherapy treatments. During a gated procedure, the radiation beam is only activated when the patient is within a certain breathing state such as end of exhalation or inhalation. The reduced treatment time can be evaluated by the **duty cycle (DC)** which is defined as

$$DC = \frac{N_{on}}{N}, \quad (2.11)$$

where  $N_{on}$  is the number of sample points which are within a predefined breathing state [128, 241]. This concept can be adopted to investigate the error distribution for adaptive radiation therapy using an error threshold  $e_{th}$ . In this case,  $N_{on}$  is the number of points with  $|e_i| \leq e_{th}$  for  $i = 1, \dots, N$ . This evaluation measure was used for example in [249] for  $e_{th} = \{0.2, 0.5, 1, 2\}$  mm and in [242] for  $e_{th} = \{1, 2, 4, 5, 10\}$  mm.

Ernst *et al.* [76] proposed a measure to investigate the smoothness of a prediction signal to which they referred to as jitter  $J$ . The main motivation for this measure are

the kinematic limitations of the specific system. To compensate for respiratory motion, actuators have to be controlled to move a specific leaflet of a collimator or the complete radiation unit as in case of the CyberKnife<sup>®</sup>. The wear of these actuators can be reduced by a smooth control signal. For completeness, it has to be mentioned that often in practise the predicted signal  $y^*$  is smoothed by a low pass filter [227] before it is transferred to the control unit of the robotic system. However, these filters add an additional latency. Therefore, an already smooth predicted signal is desirable. The jitter is defined as

$$J(y^*) = \frac{f_s}{(N-2)} \sum_{i=1}^{N-1} \|y_i^* - y_{i+1}^*\|. \quad (2.12)$$

Similar to the normalization coefficient presented in sec. 2.2.1, the relative jitter  $J_{rel}$  [86] can be defined by

$$J_{rel} = \frac{J(y^*)}{J(y)}. \quad (2.13)$$

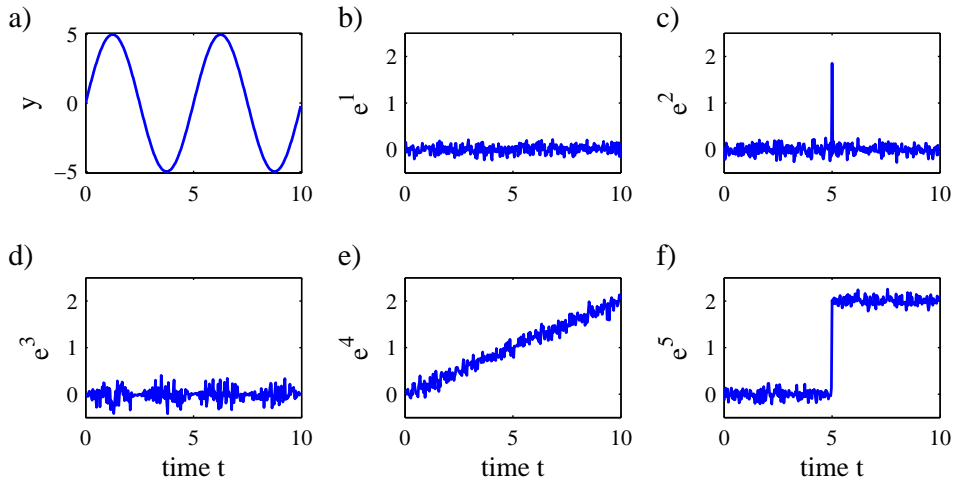
A  $J_{rel}$  close to 100 % is desirable, as the predicted and the true signal would be on average equivalent in their smoothness.

Another auxiliary evaluation is the computation speed [63, 80] which is essential to perform real-time treatment. Assuming a sampling rate of  $f_s = 26$  Hz as in the case of the CyberKnife<sup>®</sup>, the output of an MC algorithm has to be computed within 38.5 ms. In contrast to other measures, this criterion depends on several parameters such as the actual implementation of the algorithm, the programming language, and the hardware. An independent measure simply counts the number of floating point operations [263]. Of minor importance is the computational storage which is discussed in [263].

### 2.2.3 Example of Evaluation Measures

Several of the previously discussed evaluation measures will now be applied to different synthetic error signals to highlight the differences between them. We assume that the true signal  $y$  is given by

$$y = a_1 \cdot \sin(2\pi t/T_p), \quad a_1 = 5, \quad t = [0, 1, \dots, 1000]/f_s, \quad f_s = 26 \text{ Hz},$$



**Figure 2.3:** a) True signal  $y$ ; b-f) Prediction errors  $e^1$  to  $e^5$  for the signals  $y^{1*}$  to  $y^{5*}$ .

where  $T_p = 5$  s represents the breathing period. The output of the MC algorithms is assumed to be one of the following five signals

$$\begin{aligned}
 y^{1*} &= a_1 \cdot \sin(2\pi t/T_p) + \varepsilon \\
 y^{2*} &= \begin{cases} a_1 \cdot \sin(2\pi t/T_p) + \varepsilon & t \neq 5 \text{ s} \\ a_1 \cdot \sin(2\pi t/T_p) + \varepsilon + a_2 & t = 5 \text{ s} \end{cases} \\
 y^{3*} &= (a_1 + a_2 \cdot \varepsilon) \cdot \sin(2\pi t/T_p) \\
 y^{4*} &= a_1 \cdot \sin(2\pi t/T_p) + \varepsilon - a_2 \cdot t/t_{max} \\
 y^{5*} &= \begin{cases} a_1 \cdot \sin(2\pi t/T_p) + \varepsilon & t < 5 \text{ s} \\ a_1 \cdot \sin(2\pi t/T_p) + \varepsilon - a_2 & t \geq 5 \text{ s} \end{cases}
 \end{aligned}$$

where  $\varepsilon \sim \mathcal{N}(0, 0.01)$ ,  $a_2 = 2$ ,  $t_{max} = 1000/f_s$ , and  $t \in [0, 10]$  s. The signal  $y$  and the resulting errors  $e^k = y - y^{k*}$  with  $k = 1, \dots, 5$  are shown in Fig. 2.3. The error signals represent different scenarios such as a constant (Fig. 2.3.b), periodic (Fig. 2.3.d) or linearly increasing (Fig. 2.3.e) prediction error. Table 2.1 lists the ME, MAE, mAE, RMSE, nRMSE,  $RMSE_{rel}$  assuming a latency index of  $\xi = 3$  and duty cycle for a threshold of  $e_{th} = 0.1$  and  $e_{th} = 0.5$ .

As expected, the different kinds of errors cannot be differentiated by the ME, as the MEs for  $e^1$  to  $e^3$  are close to zero or for  $e^4$  and  $e^5$ , close to one. Comparing the MAE and the mAE confirms that the mAE alone can be misleading as in case of  $e^2$ . This signal represents the scenario of a constant small error with one outlier. Due to the outlier,



**Table 2.1:** Result of different evaluation measure

	ME	MAE	mAE	RMSE	nRMSE	RMSE <sub>rel</sub> ( $\xi = 3$ )	DC ( $e_{th} = 0.1$ )	DC ( $e_{th} = 0.5$ )
$e^1$	0.004	0.084	0.355	0.105	0.030	35.5 %	64.7 %	64.7 %
$e^2$	0.007	0.080	2.006	0.148	0.042	50.2 %	73.3 %	99.7 %
$e^3$	0.013	0.094	0.475	0.133	0.038	45.0 %	64.7 %	100 %
$e^4$	1.000	1.003	2.134	1.158	0.327	392.2 %	4.7 %	25 %
$e^5$	1.001	1.042	2.282	1.421	0.401	481.6 %	33.3 %	50 %

the mAE of  $e^2$  is high even though the MAE is slightly smaller than the MAE of signal  $e^1$ . Signal  $e^3$  represents the case of an increased periodic error at maximum inhale or exhale which is often observed in practise. The difference between  $e^2$  and  $e^3$  can also be observed in the MAE and RMSE. The MAE of  $e^3$  is higher than that of  $e^2$ . In contrast, the RMSE of  $e^3$  is lower than that of  $e^2$ , as higher prediction errors are penalized stronger. The differences between the two signals can also be observed in the DCs. While a high percentage of the data of  $e^2$  has an error below  $e_{th} = 0.1$  (73.3 %), a small percentage also has an error above  $e_{th} = 0.5$  (0.3 %). In contrast,  $e^3$  has a DC = 100 % for  $e_{th} = 0.5$ .

Column six of table 2.1 shows the RMSE<sub>rel</sub> assuming that the signals  $y^{1*}$  to  $y^{5*}$  are the results of a prediction algorithm with a latency index of  $\xi = 3$ . As the RMSE<sub>rel</sub> of  $e^1$  to  $e^3$  is < 100 %,  $y^{1*}$  to  $y^{3*}$  would improve the prediction accuracy compared to doing no prediction. In contrast,  $y^{4*}$  and  $y^{5*}$  would worsen the prediction accuracy compared to doing no prediction.

## 2.3 Motion Compensation Algorithms

The following two subsections give an overview of algorithms for respiratory motion prediction and correlation. As stated previously, a wider range of publications can be found for respiratory motion prediction. The reasons are twofold. First, motion prediction algorithms can be evaluated on either external or internal motion data. Consequently, the measurement requirements are lower, if for example only external data is used. It can be observed that several authors used the RPM system to acquire optical external data [103, 115, 126, 166, 203, 219, 222, 223]. Second, the problem can be classified as a standard time-series prediction problem. This increases the variety of algorithms as methods from different research fields such as stock market or weather prediction can be applied without large modifications.

The opposite is the case for correlation algorithms. An increased measurement effort is necessary since internal and external motion data has to be acquired. Additionally, the internal position is often measured via fluoroscopy. To limit the amount of additional radiation dose to the patient due to imaging, only a small number of images is acquired. This makes the evaluation of more complex correlation algorithms challenging.

One recent literature review covering this topic was presented by Vedam *et al.* [268]. Even though the majority of the algorithms were discussed, several - partly recent - publications are missing. Furthermore, the survey does not clearly differentiate between prediction and correlation algorithms. However, this is important as not all presented algorithms can be used as correlation algorithms.

### 2.3.1 Prediction Algorithms

Several authors have presented different approaches for respiratory motion prediction algorithms. Besides the mathematical model, these publications differ in several other aspects, which shall be discussed briefly. Table 2.2 provides the resulting summary.

**Data acquisition** Different measurement modalities have been used to measure respiratory motion. They can be broadly divided into the measurement of internal tumour and external surrogates motion. The latter was measured exclusively with optical tracking systems. There, **optical markers (OMs)** are attached to the chest and/or abdomen of the subject. The position of these markers can be acquired by a camera at a high sampling frequency  $f_s$ . The measurement is non-invasive and the patient is not exposed to additional radiation. Consequently, long measurements with high resolution are easily possible. Data based on external **OMs** were most frequently acquired either by the **RPM** system of Varian ( $f_s = 30$  Hz) or the optical tracking system integrated in the CyberKnife<sup>®</sup> ( $f_s = 26$  Hz).

Internal tumour motion was most frequently acquired by fluoroscopy. The 3D position of one or more fiducials can be automatically tracked via two fluoroscopic imaging systems. In general, the fiducials are gold seeds with a diameter between 1 – 2 mm which are placed near the tumour. These measurements are invasive and often relatively short ( $t < 300$  s), due to the additional radiation. The sampling frequency in the publications varies between 10 Hz [266] and 33 Hz [210]. Ma *et al.* [160] presented an alternative approach to image the tumour motion without fiducials directly using a mega-voltage imaging system. However, the authors report a low sampling frequency of  $f_s = 3.5$  Hz. Another approach without fiducials was presented in [266] where only the diaphragm motion was measured.

**Amount of validation data** Respiratory motion is a quasi-periodic motion with high variability. Signal characteristics such as amplitude, period, and regularity differ significantly between patients but also between treatment sessions of the same patient as well as within a treatment session [131]. Optimally, prediction algorithms should be evaluated on a large number of motion traces with a long duration. The literature shows a high variance in the amount and duration of validation data used. The number of motion traces varies between one or ten [80, 166, 222] up to over 300 [86, 226]. Similarly, the duration of these motion traces are within the range of several seconds [266] up to more than an hour [86]. However, these variations can be partly explained by the data acquisition of the motion traces. In general, the number and duration of evaluated motion traces is lower if internal data is used.

**Patient coaching** Patient coaching can be used to stabilize the breathing pattern. This is typically carried out using a visual or acoustic feedback to the patient. As coaching techniques increase the treatment time (adaptation phase of the patient), this approach was only investigated by a few authors [186, 266]. The majority of the validation datasets consist of normal, free breathing.

**Evaluation measure** As discussed the sec. 2.2, several different evaluation measure were used.

**Latency** As pointed out in sec. 1.2.2, different latencies have to be considered. Relevant latencies are between  $h = 50$  ms for the Vero system up to several hundred for MLC tracking and moving patient couches. In the literature, prediction latencies were investigated between 0.033 s [128, 241, 278] up to  $h = 1.8$  s [265].

However, the main difference between these publications are the mathematical models used to compute a prediction value. These models will be described within the following subsections.

### 2.3.1.1 Extrapolation Techniques

One intuitive approach to predict respiratory motion is the use of linear extrapolation as presented in [241, 243, 259]. By assuming a constant velocity, the predicted amplitude is estimated by

$$y_{i+\xi} = y_i + \Delta y = y_i + (t_{i+\xi} - t_i) \frac{y_i - y_{i-\xi}}{t_i - t_{i-\xi}}. \quad (2.14)$$

However, this approach is not capable of predicting complex motion patterns as it considers only two previous observations. Further, the assumption of constant velocity is not valid for high latencies [241]. In [214], the idea was extended by using spline extrapolation.

An extension of this idea was presented in [78] to which the authors refer to as **MULIN**. **MULIN** takes more observations into account by considering higher order differences of  $\Delta y$ . Additionally, an exponential smoothing parameter was introduced to render the algorithm less sensitive to abrupt signal changes.

### 2.3.1.2 Shape Models

Within this subsection, intuitive approaches utilising the quasi-periodic shape of respiratory motion pattern are presented. Besides the approach in [222], most of these methods are not accurate enough for real motion traces. Nonetheless, they are mentioned for completeness.

Sayeh *et al.* [227] proposed the idea of using a simple-dictionary based prediction algorithm. At the current time step  $i$ , a motion fragment is defined containing the  $d$  most recent observations. Within the dictionary, the most similar pattern is searched. The predicted point is the already observed point which is  $h$  seconds ahead of the most similar pattern.

Ruan *et al.* [222] extended this idea and proposed an algorithm using the  $r$  nearest neighbours. After finding the most relevant neighbours, the regression weights were computed using a kernel.

Instead of using extrapolation techniques, a sinusoidal model can be fitted to previously observed data as suggested in [266]. The model assumes that

$$y_{i+\xi} = a_1 \sin(a_2 x_i + a_3) + a_4, \quad (2.15)$$

where  $x_i = t_{i+\xi}$  and  $a_1$  to  $a_4$  are parameters to model the amplitude, frequency, phase shift, and linear offset, respectively. The parameters are fitted based on a signal history.

This idea can be further extended by computing an average trajectory as proposed by Neicu *et al.* [192]. This model investigates the standard deviation of the amplitude, positions, and periods of multiple breathing cycles. This information is used to construct an average model with a mean breathing period.

### 2.3.1.3 Adaptive Filters

Adaptive filters are a frequently used prediction technique, which are based on the **autoregressive (AR)** properties of the signal. The so called **AR** models assume in their simplest

implementation that the output can be computed based on a linear combination of previously observed values [241].

$$y_{i+\xi} = \mathbf{w}_i^\top \mathbf{x}_i + b + \varepsilon_{i+\xi} = w_1 y_i + w_2 y_{i-1} + \dots + w_d y_{i-d+1} + b + \varepsilon_{i+\xi}, \quad (2.16)$$

where  $\mathbf{w}_i = [w_1, \dots, w_d]^\top$  is a weight vector,  $\mathbf{x}_i$  is the input vector,  $b$  a linear offset, and  $\varepsilon_{i+\xi}$  is noise at time  $t_{i+\xi}$ . Here,  $\mathbf{x}_i$  is defined as  $\mathbf{x}_i = [y_i, y_{i-1}, \dots, y_{i-d+1}]^\top$ , but also alternative implementations with a temporal scaling factor  $l$  are possible such that  $\mathbf{x}_i = [y_i, y_{i-l}, \dots, y_{i-l(d-1)}]^\top$ . A profound investigation of the optimal dimension  $d$ , also referred to as model order, was done in [68] using information criteria. The weights can be optimised based on a **least mean squares (LMS)** criterion. If the signal is assumed to be stationary, it is sufficient to perform an **LMS** optimization once on a single training dataset. However, this assumption is not valid for respiratory motion as Isaksson *et al.* [123] demonstrated. For non-stationary signals, the **LMS** algorithm can be used

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mu(y_i^* - y_i)\mathbf{x}_i, \quad (2.17)$$

with  $\mu \in [0, 1]$  being the learning parameter. This algorithm has been widely used as in [77, 78, 81, 86, 123, 142, 184, 185, 206, 211, 222, 266]. The stability of updated weights can be increased by simultaneously considering multiple training pairs  $m$  at each time step [81] or by normalizing the input vector  $\mathbf{x}_i$  in Eq. 2.17 [77, 118]. The latter is referred to as the **normalised LMS (nLMS)** algorithm. Alternatively, the weight vector can be updated by the **recursive least squares (RLS)** algorithm [79].

Several variations of this model have been investigated. In [226], an adaptive bilinear filter was evaluated which assumes different weighting factors for inspiration and expiration. A hybrid approach of an **LMS** filter, a simple fuzzy logic and a zero error predictor can be found in [242]. This is also the prediction algorithms currently implemented in the CyberKnife<sup>®</sup>. Huang *et al.* [118] proposed an acceleration enhanced **nLMS** filter which combines two separate filters - one for the position and one for the acceleration. The idea is, that the second filter predicts a correction term which enhances the predicted accuracy of the first filter. Ernst *et al.* [81], presented an approach using the á trous wavelet decomposition which decomposes the observed signal into  $J + 1$  scales

$$y_i = c_{J,i} + \sum_{k=1}^J W_{j,i}, \quad (2.18)$$

where the approximation coefficients  $c_J$  represent a smoothed version of the observed signal and the detail coefficients  $W_j$  the high frequency parts. The **LMS** algorithm can

then be applied to each scale. The predicted value will be

$$y_{i+\xi} = \mathbf{w}_{J+1,i}^\top \mathbf{c}_{J,i} + \sum_{j=1}^J \mathbf{w}_{j,i}^\top \mathbf{W}_{j,i}, \quad (2.19)$$

$$\mathbf{W}_{j,t} = (W_{j,i}, \dots, W_{j,i-d+1}), \quad \mathbf{c}_{J,t} = (c_{J,i}, \dots, c_{J,i-d+1}).$$

The **wLMS** algorithm has been extensively evaluated on several datasets, e.g. in [25, 86].

Another commonly used extension of the **AR** model is the **autoregressive moving average (ARMA)** model. In contrast to Eq. 2.16, an **ARMA** model is extended by an averaging component which considers the noise components of previous time steps by

$$y_{i+\xi} = \mathbf{w}_i^\top \mathbf{u}_i + \mathbf{v}_i^\top \boldsymbol{\varepsilon}_i + c + \varepsilon_{i+\xi}, \quad (2.20)$$

with  $\mathbf{v}_i \in \mathbb{R}^{p \times 1}$  being a weight vector and  $\boldsymbol{\varepsilon}_i = [\varepsilon_i, \varepsilon_{i-1}, \dots, \varepsilon_{i-p+1}]^\top \in \mathbb{R}^{p \times 1}$  a vector of previous noise terms. Here,  $p$  is the order of the moving average model. **ARMA** models for respiratory motion prediction have been used in [152, 210].

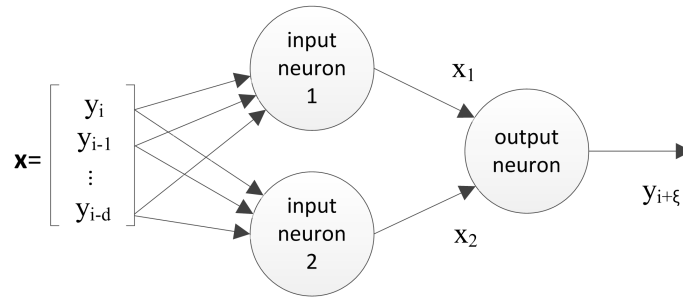
McCall *et al.* [166] proposed a dual-component model which is based on an alternative signal decomposition. First, the mean period within the last 60 s is estimated. This information is used to generate a periodic signal which is subtracted from the true signal. The resulting difference signal is used as input for an **ARMA** filter. The prediction  $y_{i+\xi}^*$  is the sum of the periodic signal at  $t_{i+\xi}$  and the prediction result of the **ARMA** filter.

Researchers of Tohoku University, Japan, investigated the influence of the temporal scaling factor  $l$  in Eq. 2.16 [114, 122]. In [122], Ichiji *et al.* presented a time-varying seasonal **AR** model which incorporates the periodic nature of respiratory motion by defining  $l = T_p f_s$  with  $T_p$  being the breathing period. They investigate different possibilities to estimate  $T_p$  using a correlation analysis.

One of the most recent publications investigated **kernel adaptive filter (KAF)** methods, which combine kernel methods with adaptive filters [263]. In general, kernel methods are widely used in combination with support vector regression (sec. 2.3.1.6). The idea is to transform the input vector  $\mathbf{x}_i$  into a higher dimensional kernel space, in which predictions are performed. Even though this is a very promising approach, the article focuses only on the comparison between different **KAF** methods and therefore lacks the comparison to other methods.

### 2.3.1.4 Artificial Neuronal Networks

**Artificial neuronal networks (ANNs)** are a modelling technique which is inspired by the biological behaviour of neurons in the brain. Using **ANNs** for respiratory motion compensation was first proposed by Murphy *et al.* [184] in 2002. The most commonly used



**Figure 2.4:** Illustration of an artificial neuronal network with two neurons at the input layer and one neuron at the output layer.

**ANN** consists of an input and an output layer [103, 123, 183–186, 241, 249]. The input layer consists of  $n$  neurons which receive as input the  $d$  most recent observations. The output of these neurons can be computed via an activation function, which is often a sigmoid function. Then, the output can be defined as

$$x_k = \frac{1}{1 + \exp(-\sum_{j=1}^d v_{k,j} y_{i-j+1})}, \quad (2.21)$$

where  $v_{k,j}$  is the weight for the input  $j$  at neuron  $k$ . In general, the output layer consists of one neuron with a linear activation function. The output is then defined as

$$y_{i+\xi} = \sum_{k=1}^n w_k x_k, \quad (2.22)$$

where  $w_k$  is the weight of the output at neuron  $k$ . This structure is illustrated in Fig. 2.4 with  $n = 2$ . Murphy *et al.* [185] showed that **ANNs** with a nonlinear activation function have a superior performance compared to linear **ANNs**, especially for highly irregular breathing patterns. In [48], the authors compared the performance of two and three layer **ANNs** and reported slightly improved prediction accuracy for three layer **ANNs**.

Goodband *et al.* [103] focused on the training of the weights  $v$  and  $w$ . They compared three training algorithms: conjugate-gradient, Levenberg-Marquardt and a hybrid algorithm based on conjugate-gradient and Bayesian regularization. They conclude that the highest prediction accuracy can be achieved by using the latter algorithm.

Alternative networks using an augmented input space were discussed in [37] (quadratic neuronal unit) and [103] (generalized regression neural networks).

### 2.3.1.5 Kalman Filters

The **Kalman filter (KF)** is a widely used technique to model linear dynamic systems, which is described in detail in [274]. The model assumes a time dependent internal state vector  $\mathbf{x}_i$  which cannot be observed directly via measurements. However, the state vector  $\mathbf{x}_{i+1}$  can be predicted recursively by

$$\mathbf{x}_{i+1} = \mathbf{A} \mathbf{x}_i + \boldsymbol{\varepsilon}_i, \quad (2.23)$$

where  $\mathbf{A}$  is the transition matrix and  $\boldsymbol{\varepsilon}_i$  a normally distributed noise vector. Knowing the state  $\mathbf{x}_i$ , a measurement can be predicted by

$$y_{i+1} = \mathbf{B} \mathbf{x}_{i+1} + \nu_{i+1}, \quad (2.24)$$

with  $\mathbf{B}$  being the measurement matrix and  $\nu_i$  normally distributed noise. At each time index  $i$ , the Kalman filter performs a “prediction” a and “correction” step. First, based on the previous state, a state  $\mathbf{x}_{i+1}$  and its covariance  $\mathbf{P}_{i+1}$  is predicted. The predicted state is used to predict a measurement  $y_{i+1}^*$ . After acquiring a new observation  $y_{i+1}$ , the predicted state and its covariance is corrected by computing the so-called Kalman gain  $\mathbf{K}_{i+1}$ .

For respiratory motion prediction, different state models were investigated. In [203, 226, 249] a **constant velocity (CV)** and a **constant acceleration (CA)** Kalman filter implementation were proposed. The **CV** model assumes a two-dimensional state vector consisting of position and velocity. This state model is extended in the **CA** approach by also considering the acceleration. The state update equation for the **CV** and **CA** models can be written as

$$\mathbf{x}_{i+1}^{CV} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \mathbf{x}_i^{CV} + \begin{bmatrix} \Delta t^2/2 \\ 1 \end{bmatrix} \boldsymbol{\varepsilon}_i, \quad \mathbf{x}_{i+1}^{CA} = \begin{bmatrix} 1 & \Delta t & \Delta t^2/2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_i^{CA} + \begin{bmatrix} \Delta t^2/2 \\ \Delta t \\ 1 \end{bmatrix} \boldsymbol{\varepsilon}_i, \quad (2.25)$$

with  $\Delta t$  being the time interval between two sample points. The measurement matrix is set to  $\mathbf{B} = [1 \ 0]$  for the **CA** model and to  $\mathbf{B} = [1 \ 0 \ 0]$  for the **CV** model. However, as model names indicate, they assume constant velocity or acceleration. As this is not valid for respiratory motion, the authors conclude that these models can only be used for short latencies. A hybrid approach of the **CV** and **CA** model is presented in [203].

To model nonlinear functions, the **extended Kalman filter (EKF)** or **unscented Kalman filter (UKF)** can be used [274]. So far, two approaches using the **EKF** have been investigated. Ramrath *et al.* [206] proposed an **EKF** model based on a sum of multiple sinusoidal



models. Each sinusoidal model was represented by three parameters in the state vector, namely the amplitude, frequency and phase shift. Further, a parameter is added to represent linear offsets. Another **EKF** approach is based on the mean estimator by Ruan *et al.* [223]. An augmented elliptic pattern can be observed by considering multiple current and delayed observations. The local circular model [118, 126] defines a state vector in this augmented space based on the current position, velocity and angular velocity.

### 2.3.1.6 Support Vector Regression

Support vector machines have become a frequently used tool for classification and regression problems. A practically very relevant approach is  $\varepsilon$ -**SVR** which is described in detail in [250]. Within respiratory motion compensation, **SVR** was first investigated by Ernst *et al.* [80]. It is assumed that a training dataset  $\mathbf{T} = \{\mathbf{X}, \mathbf{y}\}$  with  $\mathbf{X} = [\mathbf{x}_1^\top, \dots, \mathbf{x}_m^\top]$  is given (see sec. 2.1). This training dataset can be interpreted as a dictionary of size  $m$  with previous input-output pairs. The **SVR** algorithm then tries to optimize a function  $f(\mathbf{x})$  such that the function values have at most  $\varepsilon$  deviation from the training labels. The function can be written as

$$y_{i+\xi} = f(\mathbf{x}_i) = \langle \mathbf{w}, \mathbf{x}_i \rangle + b, \quad (2.26)$$

where  $\langle \cdot, \cdot \rangle$  denotes the dot product,  $\mathbf{w}$  is a weight vector, and  $b$  a linear offset. To solve for  $f(\cdot)$ , a convex optimization problem can be formulated in which  $\mathbf{w}$  is penalized by the  $L_2$ -norm. As a result of this optimization, a subset  $\mathbf{M} \subset \mathbf{T}$  of the training dataset is specified, which is also referred to as support vectors. These are the training pairs with  $|f(\mathbf{x}_j) - y_j| \geq \varepsilon$ . The trained function  $f(\cdot)$  can be used to predict an unobserved point  $y_{i+\xi}^*$  based on  $\mathbf{x}_i^*$ . Furthermore, **SVR** can be extended to nonlinear functions by using kernel functions. A frequent choice is the Gaussian kernel or **radial basis function (RBF)**, which was investigated in [62, 63, 80, 86, 142, 211]. Choi *et al.* [48] compared a linear and an **RBF** kernel and reported an increased prediction accuracy for the **RBF** kernel.

One major limitation are the high computational requirements of **SVR**, which are extremely problematic in real-time applications. Optimally, a large dataset  $\mathbf{T}$  should be used to decrease the prediction error, but this also increases the computational load. In [48, 80], an adaptive **SVR** implementation was evaluated and in [62] different update rules of the training set were investigated.

A hybrid approach was presented in [63], combining **SVR** and the wavelet decomposition presented in [81]. However, this approach further increases the computational requirements as multiple **SVRs** are used.

### 2.3.1.7 Probabilistic Approaches

So far the only probabilistic approach was investigated for respiratory motion prediction by Ruan *et al.* [219] which is based on **kernel density estimation (KDE)**. Given a training set  $\mathbf{T}$ , the algorithm computes a joint probability distribution of the training features and labels using a Gaussian kernel. Knowing the covariance matrix of the features  $\Sigma_u$  and the variance of the labels  $\sigma_y$ , the conditional probability  $p(y^*|\mathbf{x}^*)$  of an unobserved label  $y^*$  given the feature  $\mathbf{x}^*$  can be computed by

$$p(y_{i+\xi}^*|\mathbf{x}_i^*) = \frac{1}{C} \sum_{i=1}^m w_i \exp(-\|y_{i+\xi}^* - y_i\|/\sigma_y^2), \quad (2.27)$$

$$w_i = \exp(-(\mathbf{x}_i^* - \mathbf{x})^\top \Sigma_u^{-1} (\mathbf{x}_i^* - \mathbf{x})), \quad (2.28)$$

where  $C$  is a normalization parameter. The predicted value  $y_{i+\xi}^*$  is then defined as the mean of the conditional distribution

$$y_{i+\xi}^* = \frac{1}{\sum_{j=1}^m w_j} \sum_{i=1}^m w_i y_i. \quad (2.29)$$

Alternative solutions could be either the median or the maximum *a posteriori* probability of the conditional distribution.

An interesting advantage of this method was discussed in [220]. The authors proposed that by considering the second-order statistics of the conditional distribution, i.e. the variance, the ‘‘certainty’’ of the algorithm in the prediction can be expressed. This information could be used to detect potential large prediction errors.

### 2.3.1.8 Finite State Models

Finite state models assume that the respiratory motion pattern can be divided into a sequence of discrete states. Such a classification was proposed by Wu *et al.* in [277]. This model is applied to motion prediction in [278] by defining an online subsequence matching and similarity algorithm and was further extended using hidden Markov models [128]. The authors assume that a normal breathing cycle can be divided into three piecewise linear states: exhale, end-of-exhale, and inhale. The states were classified by their velocity. Additionally, an irregular breathing state was defined. Analysing a database of motion patterns, the authors computed the probabilities of being in one state and their transition probabilities. Within [128], the authors especially focused on the prediction of an optimal gating signal.

The appealing advantage of these models is the analysis of the patient-specific breathing pattern. This allows for simple characterisation and classification of different breathing pattern as shown in [279, 281].

### 2.3.1.9 Comparative Studies

As shown in the previous subsections, several different models have been proposed, which were partly evaluated on small datasets and using different evaluation measures. This makes a direct comparison difficult. Here, the most comprehensive studies will be highlighted, which were selected based on a high number of motion traces and prediction algorithms.

In case of predicting internal tumour motion acquired by fluoroscopy, the studies of Sharp *et al.* [241] and Krauss *et al.* [142] have to be mentioned. In [241], 14 lung cancer patients were selected which did show breathing amplitudes greater than 8 mm. The signal duration was between 48 s and 342 s. The authors compared the performance of linear extrapolation, a linear filter, two ANNs (one predicting the position and the other one the velocity), and one KF against no prediction for a prediction latency of  $h \in \{0.033, 0.2, 1\}$  s. However, the extrapolation method and the velocity ANN were only evaluated on  $h = 0.033$  s as they assume a constant velocity. The linear filter was equivalent to an adaptive AR model, whose weights were only optimized once during the training phase. Sharp *et al.* conclude that the highest prediction accuracy over all latencies can be achieved by the linear filter followed by the ANN which predicts the position. However, for short latencies and a high sampling frequency, no prediction led to superior results compared to the ANN.

In the study of Krauss *et al.* [142], the first 83 s of twelve motion traces of six lung cancer patients were evaluated. The authors compared an ANN, an SVR, a KDE, and a linear filter. The authors investigated different parameter update strategies (optimizing the parameter only once at the beginning or at regular intervals), prediction latencies ( $h \in \{0.2, 0.4, 0.6\}$  s), sampling frequencies ( $f_s \in \{7.5, 15, 30\}$  Hz), and different representations of the input signals (normal input signal and using a principle component analysis on the input signals). They reported only minor differences in case of using different input representations and parameter updating strategies. The latter could be strongly influenced by the short duration of the motion traces. The best performance on average was achieved by using the ANN approach followed by the SVR, linear regression and KDE methods.

Three other comprehensive studies are based on externally acquired motion traces

**Table 2.2:** Overview of the published prediction algorithms in the years 2000 to 2014. The publications are categorized based on the mathematical model used, investigated latency, acquired data, and if the subjects did have breathing coaching.

	publication year			
	2000 – 2004	2005 – 2007	2008 – 2010	2011 – 2014
<b>model</b>				
extrapolation	[214, 241, 243]		[78, 259]	
shape models	[192, 265]	[222, 227]		
adaptive filter	[152, 184, 214, 241, 266]	[81, 123, 160, 166, 206, 210, 222, 226, 242]	[77, 79, 82, 114, 118, 211]	[39, 86, 122, 142, 263]
neural networks	[184, 241]	[123, 222, 249]	[37, 103, 183, 186]	[48, 142]
Kalman filter	[241]	[202, 206, 222, 226, 249]	[77, 203]	[86, 115, 126]
support vector regression		[278]	[80, 211]	[48, 63, 86, 142]
finite state models			[128]	
probabilistic models			[219, 221]	[142]
comparative study	[241]	[226]		[48, 86, 142]
<b>latency</b>				
$h \leq 100$ ms	[48, 86, 122, 128, 166, 185, 186, 203, 210, 241, 278]			
100 ms < $h \leq 200$ ms	[184, 214, 241, 266]	[81, 123, 166, 185, 206, 210, 222, 242, 249, 278]	[77–80, 82, 103, 128, 186, 203, 219, 219]	[39, 48, 62, 63, 68, 86, 115, 122, 142, 263]
200 ms < $h \leq 300$ ms	[152]	[166, 185, 226, 249, 278]	[118, 128, 186, 203]	[39, 48, 115, 122]
300 ms < $h \leq 500$ ms	[184, 266]	[123, 166, 185, 210, 222, 226, 249]	[37, 103, 128, 183, 186, 203, 211, 259]	[48, 86, 115, 122, 142]
500 ms < $h$	[184, 241, 266]	[123, 166, 185, 222]	[114, 128, 203, 211, 219, 219]	[48, 115, 122, 126, 142]
<b>data</b>				
internal	[184, 192, 241, 243, 266]	[123, 160, 210]	[37, 114, 128, 183, 211]	[122, 142]
external	[152, 184, 214]	[81, 123, 160, 166, 185, 206, 222, 226, 242, 249, 278]	[78–80, 82, 103, 118, 183, 186, 203, 219, 221, 223]	[39, 62, 63, 68, 86, 115, 126, 263]
<b>coaching</b>	[266]		[186]	

[48, 86, 226]. Due to the different data acquisition, the amount and duration of the validation data was increased. A wider spectrum of motion patterns as well as a higher variability within the motion traces has to be expected. This has to be considered when comparing results of these publications with the first two studies [142, 241]. Sahih *et al.* [226] used 331 motion traces with a duration of 4 min each. The validation set of Choi *et al.* [48] did consist of 87 motion traces with a duration between 3.6 – 8 min. The most comprehensive dataset was evaluated by Ernst *et al.* [86] with 304 motion traces and a mean duration of 71 min.

Sahih *et al.* investigated two KFs (CV and CA), a multi modal approach based on multiple KF, an adaptive bilinear filter and two ANN structures for  $h \in \{0.2, 0.4\}$  s. The ANNs were optimized with four different training algorithms. They report that the highest prediction accuracy could be achieved by using a KF based on a constant velocity model followed by the adaptive bilinear filter. These results are surprising, considering the high prediction latencies. Sharp *et al.* argued in [241] that the assumption of a constant velocity seems unrealistic and excluded the linear extrapolation model and the velocity based ANN in their investigations for high prediction latencies.

The study presented in [48] compared the performance between three ANNs and a linear and a nonlinear SVR (using an RBF kernel). Five prediction latencies were investigated,  $h \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$  s. The authors reported that the highest accuracy could be achieved by the nonlinear SVR followed by the linear SVR. These findings are in contrast to the results reported in [142]. They argued that the improved performance of the SVR is due the use of an adaptive SVR, like in [80, 86].

In the paper of Ernst *et al.* [86], six algorithms were compared, namely: LMS, nLMS, RLS, wLMS [81], MULIN [78], EKF (based on [206]), and an adaptive SVR using a Gaussian kernel. Three prediction latencies were investigated  $h = \{0.077, 0.154, 0.308\}$  s. On average the highest prediction accuracy was achieved using the wLMS approach followed by SVR and MULIN. The authors confirmed the results of Krauss *et al.* [142] that SVR can outperform standard adaptive filters (LMS, nLMS, RLS). However, using a wavelet-based LMS filter resulted in superior prediction accuracy compared to other filters and SVR. In contrast to a standard adaptive filter, the wLMS approach considers at each time index multiple training pairs ( $m > 1$ ), which might result in more robust predictions. It has to be highlighted that the parameters of the wLMS algorithm such as the number of wavelet scales were fixed and not optimized for different patients.

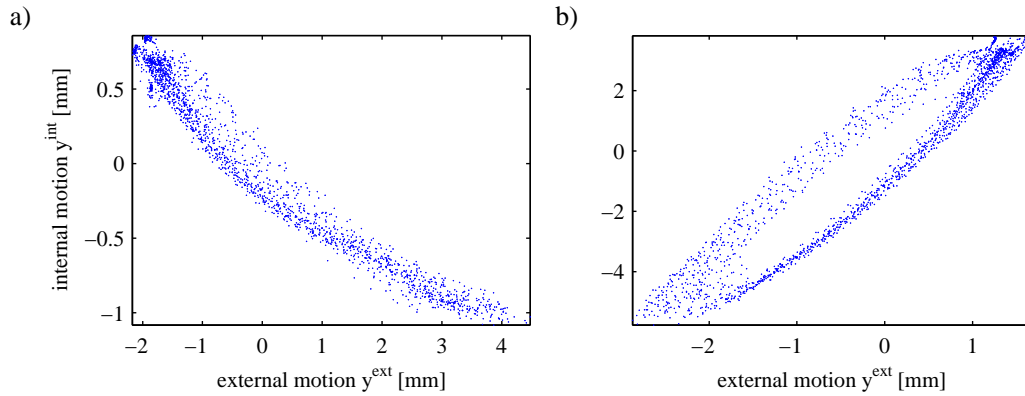
Summarising the results of these studies indicate that linear filters (especially the wLMS algorithm and the adaptive bilinear filter) and the SVR algorithm are able to predict respiratory motion very accurately for various latencies. Due to strongly varying

results, no clear conclusion can be drawn for prediction algorithms based on ANN and KF.

### 2.3.2 Correlation Algorithms

The problem of describing internal motion based on surrogate data was and still is investigated by many researchers. A very profound literature review across different medical applications can be found in [167]. McClelland *et al.* pointed out that different external surrogates have been utilised such as spirometry (airvolume and airflow [155, 284]), respiratory bellows [94], optical tracking of single markers [123, 231], and tracking of the skin surface using time of flight cameras [91]. Internal surrogates have also been investigated such as the motion of the diaphragm [41, 43]. This surrogate information was used in different models to predict, for example, the motion of an internal point, a surface or a deformation field. To validate the internal motion, different imaging modalities have been used, such as fluoroscopy, MRI [133], CT, and US. Within this section, an overview of correlation models which use external surrogates to predict the motion of one internal point representing a tumour or a fiducial is given. Publications investigating the internal surrogates, such as the motion of the diaphragm, as replacement of the external surrogates were ignored. Even though not exclusively, the focus is placed on imaging modalities, which could be used for real-time treatment such as fluoroscopy and US. The interested reader is referred to [167] to obtain a complete overview of respiratory motion models.

Since the proposal of using image guidance for adaptive motion compensation [231], there has been an increased interest in the correlation between internal and external motion. Many authors exploited the correlation of external optical markers and internal motion [12, 112, 133, 140, 231, 283]. Ahn *et al.* [12] investigated the correlation of 34 locations in the lung and 14 in the diaphragm to external skin markers and discovered a strong correlation ( $0.77 \pm 0.12$ ). Koch *et al.* [140] reported similar results. However, their results additionally indicated a strong dependency of the correlation on the direction of the internal movement. They reported a strong correlation in SI direction ( $0.89 \pm 0.09$  for volunteers and  $0.87 \pm 0.23$  for patients) and a weaker correlation in AP direction ( $0.72 \pm 0.23$  for volunteers and  $0.44 \pm 0.27$  for patients). They concluded that a combination of multiple external optical markers and other physiological signals, such as lung volume and air flow, would result in a higher correlation accuracy. In [283], Yan *et al.* reported a strong variation of the correlation coefficients with respect to the breathing pattern and marker placement. They investigated the correlation error of a linear regression model by using one or multiple external optical markers. The results indicated that a decreased correlation error was possible by using multiple external markers, which is in agreement with



**Figure 2.5:** Illustration of a quadratic (a) and a hysteresis (b) correlation between external optical marker and internal fiducial based on the data of [83]

[140]. In [112, 113, 155], the authors investigated the potential of spirometric measures. Hoisak *et al.* [112] showed that the respiratory volume had a higher and more reproducible correlation to the movement of lung tumours than abdominal markers.

Several authors also reported a temporal shift between the internal and external motion. Hoisak *et al.* [112] described shifts between  $-0.65$  s and  $0.5$  s. These temporal shifts led to a hysteresis effect which was discussed in [83, 85, 116, 224, 227, 237]. The hysteresis between internal and external data is illustrated in Fig. 2.5.b based on data of a porcine study [83]. This effect can be influenced by the breathing pattern of the subject. While inhalation is often performed actively by tension of the muscles, exhalation is done passively under normal breathing conditions. Depending on the tumour location, this can result in different trajectories.

Several studies analysed the variability of the correlation and temporal shifts over time [112, 116, 165]. Malinowski *et al.* [165] published a retrospective analysis of 55 lung and 29 pancreas fractions treated with the CyberKnife<sup>®</sup>. They divided each fraction into blocks of 10 min and investigated mean shifts, tumour position distributions, and changes in the spatial relationship between internal and external motion. A mean shift of  $> 5$  mm between the first and the third block was observed in 13% of lung and 7% of pancreas cases, respectively. Furthermore, a significant change in the correlation between tumour and external surrogates was observed in 63% of all fractions over 30 min. Interfractional variations were investigated by Hoisak *et al.* [112]. The correlation of the abdominal motion and the respiratory volume to an internal tumour motion was investigated for five patients before, in between and after treatment. The correlation was only consistent for one patient over multiple days. A result of these investigations is that cor-

relation models have to be constantly updated by acquiring new internal observations. In [129], different update strategies have been compared.

To model internal motion based on external surrogates, different mathematical methods have been proposed, which will be discussed in the following subsections. Many models are based on the same methods as presented in sec. 2.3.1 and will not be explained in detail here. Table 2.3 summarizes the used literature with respect to the mathematical model and the used imaging modality to track the internal position.

### 2.3.2.1 Linear and Quadratic Regression Models

The majority of publications proposed to use linear models [12, 113, 129, 133, 140, 153, 155, 175, 184, 227, 231–234, 237, 265, 283]. It is assumed that the internal position can be predicted by

$$y_i^{int} = a y_i^{ext} + b, \quad (2.30)$$

with  $a$  being the slope and  $b$  being the offset of the model. In general, this model is trained by minimising the least mean squares error of  $m$  observations. The advantage of this simple model is that, in theory, only  $m = 2$  internal images have to be acquired and consequently the additional radiation dose due to imaging can be kept low. However, this model is not capable of compensating the hysteresis effect. Several extensions have been presented to overcome this drawback. Vedam *et al.* [265] proposed a linear model which incorporates the time shift  $\Delta t$  between internal and external motion by

$$y_i^{int} = a y_{i-\gamma}^{ext} + b, \quad (2.31)$$

where  $\gamma = \Delta t \cdot f_s^{ext} \in \mathbb{Z}$  and  $f_s^{ext}$  is the sampling frequency of the external modality. By using the temporally shifted external signal value as input for the correlation model, the hysteresis can be efficiently compensated. This approach was further used in [113]. To train this model, sufficient internal observations to determine the time shift have to be available. Assuming two periodical signals with the same period, this could be achieved by a minimum of  $m = 3$ . An obvious alternative is the use of two models - one for inspiration and one for expiration. We refer to this model as a dual-linear model. Additionally, the accuracy of all models presented so far can be extended by the use of higher order polynomials, such as a quadratic function. We refer to a model consisting of two quadratic functions as a dual-quadratic model. The linear, quadratic and dual-quadratic models are currently implemented in the CyberKnife<sup>®</sup> [227] and were used for evaluation in [83, 85, 237, 262]. A quadratic model requires at least  $m = 3$  and



**Table 2.3:** Overview of the published correlation models in the years 2000 to 2012. The publications are categorized based on the mathematical model used and imaging modality used to acquire the evaluation data.

	publication year			
	2000 – 2004	2005 – 2007	2008 – 2010	2011 – 2014
<b>model</b>				
linear and quadratic regression	[12, 133, 140, 153, 184, 231–233, 265]	[113, 129, 155, 175, 227, 234, 237, 283]	[45, 46, 83, 262, 287]	[85]
adaptive filter	[184]	[123, 160]	[47, 183, 223, 239]	
neuronal networks	[184]	[123, 282]	[183, 262]	[238–240]
support vector regression			[57, 83]	[85]
kernel methods				[151]
<b>imaging modality</b>				
fluoroscopy	[12, 112, 184, 231–233, 265]	[227]	[45, 57, 83, 116, 183, 224, 113, 123, 129, 234, 237, 282, 283]	[46, 47, 151, 165, 238, 240]
MRI	[133, 140, 153]			
CT		[155]	[287]	
megavolt imaging		[160, 175]		[47]
US				[85]

a dual-quadratic model at least  $m = 6$  training points. Additionally, dual-linear and dual-quadratic models require classification of the training points into inhale and exhale phases.

Combinations of a linear and a dual-quadratic model have also been discussed [83, 85, 227]. Higher order polynomials can have the drawback that they lead to high correlation errors at the end of inhalation and exhalation phases. This is especially true in situations where the subject performs a deep inhalation and exhalation. In these cases, a fall-back to a linear model can decrease the correlation error.

### 2.3.2.2 Adaptive Filters

Equivalent to sec. 2.3.1.3, adaptive filters can be applied to predict the internal position by

$$y_i^{int} = \mathbf{w}_i^\top \mathbf{x}_i + b + \varepsilon_{i+\xi} = w_1 y_i^{ext} + w_2 y_{i-1}^{ext} + \dots + w_d y_{i-d+1}^{ext} + b. \quad (2.32)$$

The minimum number of training points depends on the selected model order  $m = d + 1$ . These models were successfully applied in [123, 160, 184]. By considering previous external observations in the model, the model can also compensate the hysteresis effect [224]. Ruan *et al.* [224] proposed a special case of an adaptive filter by assuming a constant order of  $d = 2$  and adding a temporal offset  $k$  between the current and the delayed external observations. The model can be formulated as

$$y_i^{int} = w_1 y_i^{ext} + w_2 y_{i-k}^{ext} + b, \quad (2.33)$$

with  $k \geq 0$ . The authors suggested to use a temporal offset  $k$  within the range of 0 s and about half the average breathing period. Similar to the quadratic model, Ruan *et al.* extended their approach to a nonlinear model by

$$y_i^{int} = w_1 y_i^{ext} + w_2 y_{i-k}^{ext} + w_3 (y_i^{ext})^2 + w_4 (y_{i-k}^{ext})^2 + w_5 y_i^{ext} y_{i-k}^{ext} + b. \quad (2.34)$$

The authors reported a superior correlation accuracy of their approaches compared to the linear and the quadratic models [224]. This approach was further used in [45, 239].

### 2.3.2.3 Artificial Neuronal Networks

Murphy *et al.* [184] proposed the use of an artificial neuronal network with one input and one output layer. The units of the input layer have a nonlinear activation function. However, the results indicated a lower accuracy compared to adaptive filters. In [123], the comparison was repeated and a superior performance of ANNs compared to an adaptive LMS filter was reported. However, the results were affected by the small number of motion traces (only one in case of [184] and three for [123]). To train the two layer ANN, the authors used 20 – 25 initial training points. In [282], a linear ANN was proposed which is equivalent to an adaptive filter.

Recently, Seregni *et al.* [238] published results on a nonlinear ANN which is based on the position of the external optical markers and their velocity. Considering the velocity, instead of multiple previous observations, is an alternative approach to compensate the hysteresis and was investigated in [147]. Within their experiments, Seregni *et al.* [238] varied the number of hidden layers and number of units per layer. The results indicated that one hidden layer and a small number of units ( $< 5$ ) is sufficient. To train the network,

$m = 100$  training points were used. The same research group published experimental results on the use of correlation models for particle therapy [239, 240]. They compared the linear approach of Ruan *et al.* [224] (however with  $d = 3$ ) and the velocity-based ANN. They reported a lower correlation error for the adaptive filter approach of Ruan *et al.* and a significant decrease of the dosimetric error compared to no motion compensation.

#### 2.3.2.4 Support Vector Regression

A more complex correlation model can be designed by the use of the  $\varepsilon$ -SVR algorithm. Preliminary results of a linear  $\varepsilon$ -SVR were presented in [57]. In [83, 85], Ernst *et al.* investigated the performance of a nonlinear  $\varepsilon$ -SVR which used an RBF kernel. The input features were a vector consisting of the position, a binary flag indicating inhalation or exhalation, the velocity, and/or the acceleration. Additionally, the feature set could be extended by multiple external markers. The authors compared the performance of their SVR approach to different (dual-)linear and (dual-)quadratic regression models. In [83], a porcine study was performed and the internal motion data was acquired by fluoroscopy. The results indicated a superior performance of the SVR approach. In [85], a validation on seven human subjects using US was performed, resulting in comparable correlation accuracies for SVR, and linear and quadratic regression models. However, the expressiveness of the second validation dataset could be limited due to the lower internal resolution of the US imaging modality.

Ernst *et al.* used  $m = 20$  data points to train their SVR approach. However, in all three studies (D'Souza *et al.* [57] and Ernst *et al.* [83, 85]), parameters of the SVR, such as  $\varepsilon$ , were fixed to different pre-defined values. In a real application, these parameters have to be learned, resulting in the need for additional observations.

#### 2.3.2.5 Kernel Regression

In [151], Li *et al.* proposed an alternative approach to which they refer as memory-based learning. The principal idea is similar to the idea of shape models (sec. 2.3.1.2). The correlation result  $y^*$  is a weighted sum of previously observed training labels. The weights are determined based on the distance between the test feature  $\mathbf{x}^*$  and the training features  $\mathbf{x}$  in a kernel space. The resulting correlation equation is equivalent to the mean estimation of the probabilistic KDE approach presented in [219]. The authors also proposed an extension of this approach, to which they refer as locally weighted regression. Even though the results are very promising and have been evaluated on 171 motion traces, the prediction results strongly depend on the size of the training data. The authors investigated

different training windows, based on the most recent 5 – 30 s. Considering the sampling frequency of  $f_s = 30$  Hz, this results in training dataset of  $m = 150 - 900$ . In contrast, the discussed ANN and SVR approaches were evaluated with  $m \leq 25$  training points.

Besides these approaches, a physiological model was investigated by Li *et al.* [150] which can be used to predict the motion of the diaphragm. Furthermore, a model based on Takens theorem was presented in [160] and one based on Fuzzy logic in [262].

In contrast to the prediction algorithms, there are no large-scale comparative studies.

## 2.4 Conclusion

In general, this literature review shows that more publications have been presented for prediction algorithms. This is mainly due to easier data acquisition, as several studies only used external respiratory motion traces. 45 publications focusing on prediction algorithms have been analysed. 30 of these used external data, 11 used internal data, and only 4 used internal and external data. Additionally, these data are sampled at a higher frequency. These two facts lead to more diverse mathematical approaches for prediction algorithms. In case of correlation algorithms, mainly linear and quadratic regression models have been applied. In general, more complex correlation algorithms such as the SVR approach of Ernst *et al.* [85] or the kernel approach of Li *et al.* [151] require an increased amount of training data compared to simpler models. Because usually fluoroscopy is used to image the internal target position, such models would increase the radiation dose for the patients.

One interesting observation is that a unified prediction and correlation approach was only investigated in [123, 184]. In these two publications, an LMS filter and an ANN have been evaluated for combined prediction and correlation depending on different prediction latencies and update rates. On average, the results indicate a superior accuracy of the ANN approach. However, the publications are lacking a direct comparison between two LMS/ANN filters (one for prediction and one for correlation) and one LMS/ANN filter (for prediction and correlation).

Focusing on the main questions of this work (sec. 1.3) reveals that the majority of the questions was not addressed. The following results can be summarized for some of the questions:

- **Q.1.2:** How to select the most relevant and least redundant markers?

This question was not addressed in the publications discussed. However, the rel-

evance of this question is illustrated in [83]. In this study, Ernst *et al.* had two evaluation datasets. Using the first dataset, an increased prediction accuracy was reported if all six available markers were used simultaneously. In contrast, an decreased correlation error was reported by using two optical markers in the second experiment. This indicates that the optimal number of markers is most likely to be patient specific. Consequently, if a multivariate sensor setting is used, approaches have to be investigated how to find the most relevant and least redundant sensor combination.

- **Q.1.3:** Can the accuracy of respiratory motion prediction be increased by using a multivariate external sensor setup?

All publications presented focused on the prediction of external optical markers. None of the papers used a different external surrogate such as respiratory volume or considered a sensor setup consisting of multiple different external surrogates. In [211], the authors investigated whether the accuracy of adaptive filters for one dimension can be increased if the filter is extended to also consider other dimensions. They reported an increased prediction accuracy of their multiple-input single-output filter compared to their single-input single-output filter for a prediction latency of 1 s.

- **Q.1.4:** Can the accuracy of respiratory motion correlation be increased using a multivariate external sensor setup?

In case of correlation models, different external surrogates have been investigated as in [112, 113]. In [112], a higher and more reliable correlation between respiratory volume and the motion of lung tumours was reported in contrast to abdominal external motion. This was confirmed by a partly decreased correlation error in [113]. Unfortunately, the correlation accuracy of a linear model based on the respiratory volume and abdominal displacement was not investigated. In [83], Ernst *et al.* showed that the correlation accuracy could be increased by using the information of multiple optical markers.

- **Q.2.1:** What criteria could be used to evaluate current prediction accuracy?

In [220], Ruan *et al.* discussed the potential of utilizing the predicted variance of the KDE algorithm to decrease prediction errors. The variance could be interpreted as the “confidence” of the algorithm in the predicted value and can be computed

at each time index  $i$ . However, this approach is limited to probabilistic algorithms. Currently, the prediction algorithm of Ruan *et al.* is the only evaluated respiratory prediction or correlation algorithm. Nonetheless, it is difficult to use this algorithm as a correlation algorithm in practise due to the high amount of training data (Li *et al.* [151] used  $m = 150 - 900$ ). Additionally in case of prediction algorithms, the comparative study of Krauss *et al.* [142] revealed that the KDE approach had the lowest prediction accuracy compared to an ANN, an SVR and an LMS model.

To conclude, the reviewed literature indicates that alternative external sensor modalities might be beneficial to increase the correlation accuracy. Furthermore, first results have been presented which point out that multiple external optical markers result in an increased prediction accuracy. However, no studies have been published so far, which combine both, multiple external sensors which acquire data from different modalities. Additionally, the predicted variance of a probabilistic algorithm could be used to control the prediction and correlation accuracy. Up to now, only one probabilistic approach has been evaluated for respiratory motion compensation, the KDE method presented in [219]. In the comparative study of Krauss *et al.* [142], the KDE method did have the highest prediction errors. Consequently, if the predicted variance should be utilized, alternative probabilistic approaches have to be investigated.

## 3 Univariate Probabilistic Respiratory Motion Prediction

In this chapter, alternative probabilistic algorithms will be investigated and evaluated for respiratory motion compensation. Two approaches are discussed here, namely **RVM** and **GP** models. The approaches are evaluated on an univariate dataset. This enables a comparison to alternative prediction algorithms which have been previously evaluated on this dataset. Furthermore, the predicted variance is investigated as criterion to control the prediction error in real-time.

The validation dataset and previously published prediction results are presented in sec. 3.1. Further, the **RVM**, the **GP** models and their essential parameters are introduced and discussed in sec. 3.2. In sec. 3.3, the algorithms are evaluated on the complete dataset for different prediction latencies. The potential of using the predicted variance as real-time feedback criterion is investigated sec. 3.4.

### 3.1 Dataset<sup>1</sup>

The validation dataset<sup>2</sup> used in this chapter has been presented previously in Ernst *et al.* [86]. It consists of 304 motion traces which have been recorded during CyberKnife<sup>®</sup> treatment sessions at the Georgetown University Hospital, Washington, DC. The traces are from 31 patients which were treated with up to seven fractions, leading to 102 fractions in total. As reported in [227], each patient wore a tight vest. Three optical markers are placed on the chest and abdomen of the patient. The markers are tracked by the dubbed Synchrony Respiratory Motion Tracking System (Accuracy, Inc.). The data were acquired at a sampling frequency of approximately 26 Hz. Two of the three markers were defect in one fraction. The motion traces of the original dataset have a duration between 80 min and 150 min. As the dataset consists of real treatments, several motion artefacts due to patient re-alignment can be observed. These motion artefacts have been excluded. After motion correction, the traces have a duration between 6.5 and 132 min (average duration

<sup>1</sup>Parts of this section have been published in [86]

<sup>2</sup>The dataset is available online at <http://signals.rob.uni-luebeck.de>

**Table 3.1:** Mean **RMSE** and **RMSE<sub>rel</sub>** of prediction algorithms which have been evaluated on the dataset of Ernst *et al.* [86]

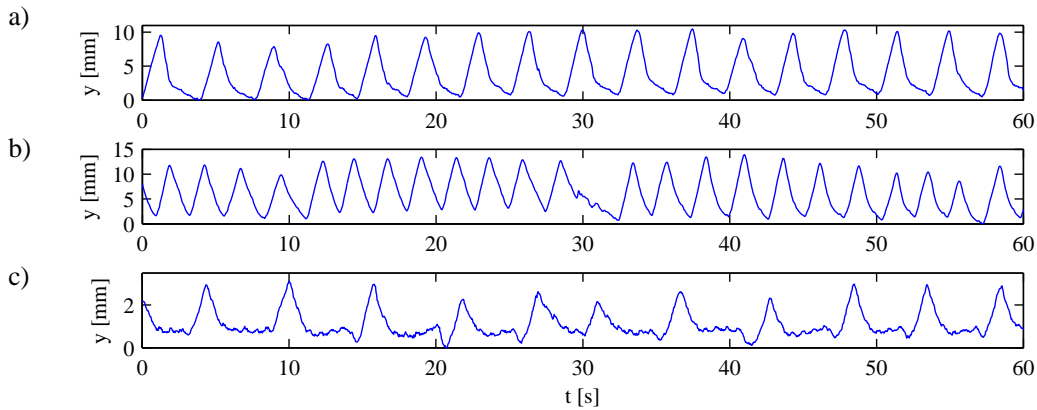
	no pred	MULIN	nLMS	RLS	wLMS	EKF	SVR
<i>h</i> = 77 ms							
RMSE [mm]	0.257	0.143	0.176	0.162	0.135	0.187	0.193
RMSE <sub>rel</sub> [%]	-	59.62	69.45	67.22	56.51	73.91	76.17
<i>h</i> = 154 ms							
RMSE [mm]	0.484	0.277	0.391	0.327	0.253	0.368	0.288
RMSE <sub>rel</sub> [%]	-	60.73	80.90	70.85	55.95	76.69	62.79
<i>h</i> = 308 ms							
RMSE [mm]	0.905	0.612	0.850	0.782	0.517	0.761	0.508
RMSE <sub>rel</sub> [%]	-	69.46	93.84	87.7	60.09	83.62	59.32

of 71 min). Further, a principle component analysis was applied, similar to [142, 221], as respiratory motion occurs predominantly in one direction. Further details about the data pre-processing can be found in [86].

In [86], Ernst *et al.* compared the performance of six prediction algorithms, namely **nLMS** [77, 81], **wLMS** [81], **RLS** [79], **MULIN** [78], Kalman filter [206], and **SVR** [80]. The Kalman filter approach is based on multiple overlaid sinusoidal models. The **SVR** model used an **RBF** kernel. The parameters of the **wLMS** algorithm were fixed for all motion traces and prediction latencies. The parameters of the residual algorithms were learned on an initial training set consisting of the first 2000 points. The optimization was performed by either grid search or a bound constrained optimization method [86]. The latencies  $h = \{77, 154, 308\}$  ms have been investigated, which corresponds to a latency index of  $\xi = \{2, 4, 8\}$ . Table 3.1 shows the mean **RMSE** and **RMSE<sub>rel</sub>** for all algorithms and prediction latencies of this study. On average, all prediction algorithm led to an improvement compared to no prediction. The **wLMS** algorithm had the lowest mean **RMSE** and **RMSE<sub>rel</sub>** for a prediction latency of 77 ms and 154 ms followed by the **MULIN** and the **SVR** algorithm. For the prediction latency of 308 ms, the **SVR** algorithm showed a superior performance followed by the **wLMS** and **MULIN** algorithm.

Three motion fragments have been selected from this dataset to study the influence of the different parameters on the prediction accuracy. The fragments are illustrated in Fig. 3.1 and represent regular (Fig. 3.1.a), irregular (Fig. 3.1.b), and noisy-irregular (Fig. 3.1.c) breathing. Each breathing fragment has a duration of 1 min.





**Figure 3.1:** Three selected motion fragments representing (a) regular, (b) irregular, and (c) noisy-irregular breathing

## 3.2 Probabilistic Machine Learning

This section provides the mathematical background for the **RVM** and **GP** methods. Furthermore, results of initial experiments are shown whose purpose was to investigate essential parameters such as the number of training pairs  $m$  and the feature dimension  $d$ . The experiments were performed on the motion fragments shown in Fig. 3.1. Before doing so, common probabilistic terms and equations shall be provided. The explanations follow the examples of [21, 208] which contain further details.

The basic probability terms shall be illustrated on a simple example. We assume that a green, red, and yellow box are given. The set of boxes is specified by  $Y = \{\text{green, red, yellow}\}$  with  $y_j \in Y$  and  $j \in \{1, 2, 3\}$ . All three boxes contain two different kind of objects (cubes and balls) which define the set  $X = \{\text{cube, ball}\}$  with  $x_i \in X$  and  $i \in \{1, 2\}$ . The setting is illustrated in Fig. 3.2.

A probability  $p(x)$  can be interpreted as the likeliness of a particular event  $x$  with  $p \in [0, 1]$ . In this discrete example the probability  $p$  can be computed as a fraction. For example, the probability  $p$  of taking out one specific object  $x_i$  from a specific box  $y_j$  is defined as

$$p(X = x_i, Y = y_j) = p(x_i, y_j) = \frac{n_{ij}}{N} \quad (3.1)$$

with  $n_{ij}$  being the number of objects  $x_i$  in box  $y_j$  and  $N$  the total number of objects in all boxes. The  $p(X = x_i, Y = y_j)$  is known as the joint probability. The joint probabilities of the example are listed in the cells of Table 3.2.a highlighted in blue.

The probability of taking out one specific object  $x_i$  irrespective of the boxes is known as the marginal probability  $p(x_i)$  which can be computed as

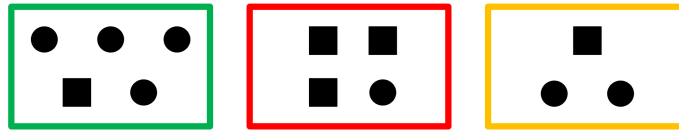


Figure 3.2: Illustration of box probability example

$$p(X = x_i) = p(x_i) = \frac{n_i}{N} \quad (3.2)$$

with  $n_i$  being the number of objects  $x_i$  in all boxes. Table 3.2.a shows  $p(x_i)$  highlighted in red and  $p(y_i)$  highlighted in green. In general, the marginal probability of  $p(x_i)$  can be computed by applying the “sum rule” of probabilities which means that the influence of, i.e., a variable  $Y$  can be integrated out by

$$p(x_i) = \sum_{j=1}^3 p(x_i, y_j). \quad (3.3)$$

If the box is known before ( $Y = y_j$ ), the conditional probability  $p(x_i|y_i)$  can be computed as

$$p(x_i|y_j) = \frac{n_{ij}}{n_j}. \quad (3.4)$$

Table 3.2.a contains the conditional probability  $p(x_i|y_i)$ . From Eq. 3.1, 3.2, and 3.4 the “product rule” of probabilities can be derived as

$$p(x_i, y_j) = p(y_j|x_i)p(x_i). \quad (3.5)$$

The previous concepts for discrete variables can be extended to the case of a continuous variables. In general, a probability density function  $p(x)$  has to be specified over a continuous variable  $x$ . The probability of  $x$  can be computed as the integral over the interval  $(x, x + dx)$  for  $dx \rightarrow 0$

$$p(x) = \int_x^{x+\delta x} p(x)dx. \quad (3.6)$$

Table 3.2: a) Joint and marginal probabilities  $p(x_i, y_i)$  (blue),  $p(x_i)$  (red), and  $p(y_i)$  (green); b) Conditional probability  $p(x_i|y_i)$

		Y			
		green	red	yellow	
X	cube	1/12	1/4	1/12	5/12
	ball	1/3	1/12	1/6	7/12
		5/12	1/3	1/4	

		Y		
		green	red	yellow
X	cube	1/5	3/4	1/3
	ball	4/5	1/4	2/3

By definition the probability density function  $p(x)$  must fulfil two conditions

$$p(x) \geq 0, \quad \int_{-\infty}^{+\infty} p(x)dx = 1. \quad (3.7)$$

For continuous variables, the sum and product of Eq. 3.3 and 3.5 can be generalized to

$$p(x) = \int p(x, y)dy, \quad p(x, y) = p(y|x)p(x). \quad (3.8)$$

The product rule allows the derivation of the Bayes theorem which is defined as

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}, \quad (3.9)$$

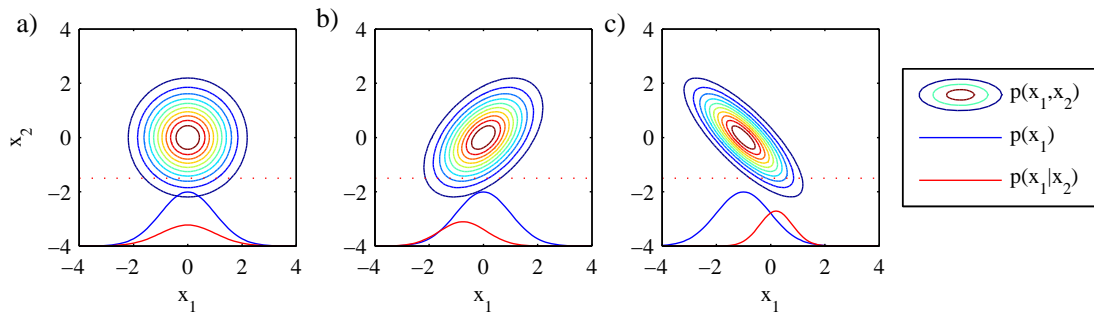
where  $p(x)$  is called the prior probability as it does specify the initial belief without observing any data. The Bayes theorem can be used to compute the posterior probability  $p(x|y)$  which expresses the probability of  $x$  after observing  $y$ . The term  $p(y|x)$  is referred to as the likelihood and specifies the probability of observing  $y$  given  $x$ . Note  $p(y|x)$  is not a probability distribution over  $y$  as the integral  $y$  does not need to be equal one. This might be also observed in Table 3.2.b, where  $\sum_{j=1}^3 p(x = \text{cube}|y_j) = 1.283$ . As a consequence, a normalization constant is required to ensure that the posterior distribution is a valid probability distribution. The normalization is realized by the marginal likelihood  $p(y)$ .

Within the literature, there are various probability distributions. An overview of relevant distributions for machine learning applications can be found in chapter two of [21]. Here, only the Gaussian distribution shall be discussed, as it is relevant for the subsequent sections. The Gaussian distribution, also known as normal distribution, is defined for a one-dimensional variable as

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right), \quad (3.10)$$

with  $\mu$  being the mean of the distribution and  $\sigma^2$  the variance. The square root of  $\sigma^2$  is known as standard deviation  $\sigma$ . In case of the multivariate variable  $\mathbf{x}$ , the Gaussian distribution is fully specified by its mean vector  $\boldsymbol{\mu} \in \mathbb{R}^{d \times 1}$  and its covariance matrix  $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$  with  $d$  being the dimension of  $\mathbf{x}$ . A multi-dimensional Gaussian distribution is defined as

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}^2) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad (3.11)$$



**Figure 3.3:** Examples of two-dimensional Gaussian distributions and their respective marginal probability  $p(x_1)$  (blue) and conditional probability  $p(x_1|x_2)$  (red) for  $x_2 = -1.5$  (red dotted line) for different mean vector  $\mu$  and covariance matrix  $\Sigma$

with  $|\Sigma|$  being the determinant of  $\Sigma$ . Figure 3.3 illustrates three examples of two-dimensional Gaussian distributions with parameters

$$\begin{aligned} \mu_a &= \begin{bmatrix} 0 & 0 \end{bmatrix}^\top, & \mu_b &= \begin{bmatrix} 0 & 0 \end{bmatrix}^\top, & \mu_c &= \begin{bmatrix} 0 & 0 \end{bmatrix}^\top, \\ \Sigma_a &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, & \Sigma_b &= \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, & \Sigma_c &= \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}, \end{aligned}$$

with the subscript label  $a$  to  $c$  referring to Fig. 3.3.a to 3.3.c, respectively. Each figure shows also the marginal distribution  $p(x_1)$  (blue solid line) and the conditional distribution  $p(x_1|x_2)$  (red solid line) for  $x_2 = -1.5$  (red dotted line). It can be observed that the off-diagonal elements of  $\Sigma$  can be interpreted as the correlation between  $x_1$  and  $x_2$ .

### 3.2.1 Relevance Vector Machines<sup>3</sup>

**RVMs** are a general Bayesian technique to solve regression and classification tasks and were first proposed by Tipping [261] and further illustrated in [96]. This framework assumes a specific prior distribution over the weights which leads to a sparse solutions of the input features. As a result, most of the weights are zero. Consequently, the number of training features is reduced to only a few relevant features to which Tipping refers as the “relevance vectors”. This sparse training set is of high practical relevance as it decreases computational requirements and prevents overfitting. Even though the concrete optimisation is different, similarities can be seen to the frequently used **SVR**. In case of **SVR**, the training set is also reduced to a sparse set, the so called “support vectors” [250]. However, Tipping illustrates in [261] that the size of the sparse training set and the prediction error

<sup>3</sup>Parts of this section have been published in [67]

is lower for the RVMs compared to the SVRs.

### 3.2.1.1 Mathematical background

We assume that the test label  $y_{i+\xi}^*$  can be computed by a linearly weighted sum of the test feature  $\mathbf{x}_i^*$

$$y_{i+\xi}^* = \mathbf{w}^\top \mathbf{x}_i^* + \varepsilon_{i+\xi}, \quad (3.12)$$

where  $\mathbf{x}_i^* = [y_i, y_{i-1}, \dots, y_{i-d+1}]^\top \in \mathbb{R}^{d \times 1}$  is a vector of the test feature at the time index  $i$  containing the  $d$  most recent observations  $y$ . The vector  $\mathbf{w} \in \mathbb{R}^{d \times 1}$  represents the weight vector and  $\varepsilon_{i+\xi}$  the measurement noise at time index  $i + \xi$ . It is assumed that the noise is normally distributed with variance  $\sigma^2$  ( $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ ). To consider nonlinear relationships between the features and the labels, the model can be extended by using a set of  $k$  basis functions  $\varphi_k(\mathbf{x}_i^*)$

$$y_{i+\xi}^* = \mathbf{w}^\top \boldsymbol{\varphi}(\mathbf{x}_i^*) + \varepsilon_{i+\xi}. \quad (3.13)$$

For the simplest case, we define a set of  $d$  linear basis functions according to

$$\varphi_k(\mathbf{x}_i^*) = \sum_{j=1}^d \delta(k, j) \mathbf{x}_i^*[j], \quad \text{with } \delta(k, j) = \begin{cases} 1, & \text{if } k = j, \\ 0, & \text{if } k \neq j. \end{cases} \quad (3.14)$$

where  $\mathbf{x}_i^*[k]$  is the  $k$ -th element of the vector  $\mathbf{x}_i^*$ . Beside a linear relationship, two non-linear functions will be investigated, to incorporate quadratic and cubic relationships between  $\mathbf{x}_i$  and  $y_{i+\xi}$ . The vectors of the resulting linear and nonlinear basis functions are defined as

$$\begin{aligned} \varphi_{lin}(\mathbf{x}_i^*) &= \mathbf{x}_i^*, & \varphi_{quad}(\mathbf{x}_i^*) &= [\mathbf{x}_i^*, \mathbf{x}_i^*[k]_{k=1, \dots, m}^2]^\top, \\ \varphi_{cub}(\mathbf{x}_i^*) &= [\mathbf{x}_i, \mathbf{x}_i^*[k]_{k=1, \dots, m}^2, \mathbf{x}_i^*[k]_{k=1, \dots, m}^3]^\top. \end{aligned} \quad (3.15)$$

Here,  $\mathbf{x}_i^*[k]_{k=1, \dots, m}^2$  and  $\mathbf{x}_i^*[k]_{k=1, \dots, m}^3$  are two vectors with the second and third power of the individual vector elements  $\mathbf{x}_i[k]$ , respectively. We refer to these algorithms as  $\text{RVM}_{lin}$ ,  $\text{RVM}_{quad}$ , and  $\text{RVM}_{cub}$ , respectively.

Equivalent to other algorithms, the aim is to learn the weight vector  $\mathbf{w}$ . Therefore, a training set  $\mathbf{T}$  can be defined. Considering the training phase, Eq. 3.12 can be formulated as

$$\mathbf{y}_i = \boldsymbol{\Phi}_i \mathbf{w} + \boldsymbol{\varepsilon}_i, \quad (3.16)$$

with  $\boldsymbol{\Phi}_i = [\boldsymbol{\varphi}(\mathbf{x}_{i-\xi})^\top, \dots, \boldsymbol{\varphi}(\mathbf{x}_{i-\xi-m+1})^\top]$  the design matrix,  $\mathbf{y}_i = [y_i, \dots, y_{i-m+1}]^\top$  the vector of the training labels, and  $\boldsymbol{\varepsilon}_t$  being the noise vector. As the noise vector is normally distributed, the probability  $p(\mathbf{y}_i | \boldsymbol{\Phi}_i, \mathbf{w}, \sigma^2)$  of the vector  $\mathbf{y}_i$  given  $\mathbf{w}$ ,  $\sigma^2$  and the trainings

data matrix  $\Phi_i$  is also a normal distribution and can be expressed according to Eq. 3.10 as

$$p(\mathbf{y}_i|\mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{-m/2} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}_i - \Phi_i\mathbf{w}\|^2\right). \quad (3.17)$$

To generate a sparse feature set, the algorithm can be constrained to smooth and less complex functions. This can be achieved by specifying a prior probability distribution on  $\mathbf{w}$ . By choosing a zero-mean Gaussian prior

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{j=1}^d \mathcal{N}(0, \alpha_j^{-1}), \quad (3.18)$$

the algorithm will prefer weights which are zero or close to it and will penalize large weights within the vector. Note that each weight  $w_j$  has an individual parameter  $\alpha_j$ , which describes the inverse variance of  $w_j$ . The parameters  $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_d]^T$  and  $\sigma^2$  are called hyperparameters. For simplicity, it was assumed that  $p(\boldsymbol{\alpha})$  and  $p(\sigma^2)$  followed a uniform distribution. In [261], the complex general cases are derived if e.g.  $p(\boldsymbol{\alpha})$  and  $p(\sigma^2)$  are drawn from a Gamma distribution.

All unknown parameters can be optimized by maximising the posterior probability  $p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{y}_i)$ . Using Bayes theorem, this probability can be expressed by

$$p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{y}_i) = \frac{p(\mathbf{y}_i|\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)}{p(\mathbf{y}_i)}. \quad (3.19)$$

However, as the probability  $p(\mathbf{y}_i)$  cannot be computed analytically, an approximation has to be found. As proposed in [261], this can be achieved by decomposing the posterior probability to

$$p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{y}_i) = p(\mathbf{w}|\mathbf{y}_i, \boldsymbol{\alpha}, \sigma^2)p(\boldsymbol{\alpha}, \sigma^2|\mathbf{y}_i). \quad (3.20)$$

The first part yields:

$$p(\mathbf{w}|\mathbf{y}_i, \boldsymbol{\alpha}, \sigma^2) = \frac{p(\mathbf{y}_i|\mathbf{w}, \sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})}{p(\mathbf{y}_i|\boldsymbol{\alpha}, \sigma^2)} = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \quad (3.21)$$

The numerators are defined by Eq. 3.17 and 3.18 and the denominator is given by the integral  $p(\mathbf{y}_i|\boldsymbol{\alpha}, \sigma^2) = \int p(\mathbf{y}_i|\mathbf{w}, \sigma^2)p(\mathbf{w}|\boldsymbol{\alpha}) d\mathbf{w}$ . The mean and covariance of the resulting normal distribution can be expressed by

$$\boldsymbol{\mu} = \sigma^{-2}\boldsymbol{\Sigma}\Phi_i^\top \mathbf{y}_i, \quad \boldsymbol{\Sigma} = \left(\mathbf{A} + \sigma^{-2}\Phi_i^\top \Phi_i\right)^{-1} \text{ with } \mathbf{A} = \text{diag}(\boldsymbol{\alpha}). \quad (3.22)$$

To evaluate the mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$ , the hyperparameters  $\boldsymbol{\alpha}$  and  $\sigma^2$  has to be found. This can be achieved by maximising  $p(\boldsymbol{\alpha}, \sigma^2|\mathbf{y}_i)$ , the second part of Eq. 3.20. The hyperparameter probability can be approximated by a delta function at the most probable values of  $\boldsymbol{\alpha}$  and  $\sigma^2$  ( $p(\boldsymbol{\alpha}, \sigma^2|\mathbf{y}_i) \approx \delta(\boldsymbol{\alpha})\delta(\sigma^2)$ ). Doing so, the problem reduces to maximising

of hyperparameter posterior mode with respect to  $\alpha$  and  $\sigma^2$ . The probability  $p(\alpha, \sigma^2 | \mathbf{y}_i)$  can be decomposed into

$$p(\alpha, \sigma^2 | \mathbf{y}_i) \propto p(\mathbf{y}_i | \alpha, \sigma^2) p(\alpha) p(\sigma^2). \quad (3.23)$$

Assuming unified hyperpriors ( $p(\alpha)$  and  $p(\sigma^2)$  can be ignored), the probability is given by

$$p(\mathbf{y}_i | \alpha, \sigma^2) = (2\pi)^{-N/2} |\sigma^2 \mathbf{I} + \Phi_i \mathbf{A}^{-1} \Phi_i^\top|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{y}_i^\top (\sigma^2 \mathbf{I} + \Phi_i \mathbf{A}^{-1} \Phi_i^\top)^{-1} \mathbf{y}_i \right\}. \quad (3.24)$$

This probability is also referred to as the marginal likelihood. Maximising this probability is known as type-II maximum likelihood. The hyperparameters can be estimated by setting the derivatives of  $p(\mathbf{y}_i | \alpha, \sigma^2)$  with respect to the hyperparameters zero, which leads to

$$\alpha_j^{new} = \frac{\gamma_j}{\mu_j^2}, \quad (\sigma^2)^{new} = \frac{\|\mathbf{y}_i - \Phi_i \boldsymbol{\mu}\|^2}{m - \sum_{j=1}^d \gamma_j}, \quad \text{with } \gamma_j = 1 - \alpha_j \Sigma_{jj} \quad (3.25)$$

where  $\mu_j$  is the  $j$ -th element of  $\boldsymbol{\mu}$  and  $\Sigma_{jj}$  is the  $j$ -th diagonal element of  $\boldsymbol{\Sigma}$ . Setting the derivative of Eq. 3.24 to zero, does hence not lead to a closed form solution of the hyperparameters. In contrast, it is an iterative processes. Starting with some initial values for  $\alpha$  and  $\sigma^2$ , the mean  $\boldsymbol{\mu}$  and covariance  $\boldsymbol{\Sigma}$  can be computed by Eq. 3.22. The updated  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  can be used to compute  $\alpha_j^{new}$  and  $(\sigma^2)^{new}$  (Eq. 3.25). This process can be repeated until  $p(\mathbf{y}_i | \alpha, \sigma^2)^{new} - p(\mathbf{y}_i | \alpha, \sigma^2)^{old} < p_{th}$ , with  $p_{th}$  being a predefined probability difference threshold, or a certain number of iterations  $n_{iter}$  was performed.

The optimized hyperparameters  $\alpha$  and  $\sigma^2$  can be used to make a predictions according to Eq. 3.13. The predicted mean and variance are

$$y_{i+\xi} = \boldsymbol{\mu}^\top \boldsymbol{\varphi}(\mathbf{x}_i^*), \quad \sigma_{i+\xi}^{2*} = \sigma^2 + \boldsymbol{\varphi}(\mathbf{x}_i^*)^\top \boldsymbol{\Sigma} \boldsymbol{\varphi}(\mathbf{x}_i^*). \quad (3.26)$$

The predicted variance  $\sigma_{i+\xi}^{2*}$  is the sum of the variances caused by the measurement noise and the uncertainty in the prediction of  $\mathbf{w}_t$ .

The RVM algorithm can be easily used as an adaptive tracking algorithm. An overview of the required steps for an RVM tracking algorithm based on a fixed number of iterations  $n_{iter}$  is shown in Table 3.3.

Note that the basis functions presented in Eq. 3.14 will result in a simplified version of the original RVM. Meaning that  $y_{i+\xi}^*$  is the sum of weighted previous observations  $\mathbf{x}_i^*$ . The weight vector will be  $\mathbf{w} \in \mathbb{R}^d$  and the relevance vectors indicate the position of the relevant entries in  $\mathbf{x}_i^*$ . This approach can be interpreted as an adaptive filter with a cost

**Table 3.3:** Outline of an **RVM** tracking algorithm with a maximum number of iterations  $n_{iter}$

---

```

specify basis function  $\varphi(\cdot)$ 
initialize  $m, d$ , and  $n_{iter}$ 
for  $i = 1$  to  $N$  do
  update design matrix  $\Phi_i = [\varphi(\mathbf{x}_{i-\xi}^\top), \dots, \varphi(\mathbf{x}_{i-\xi-m+1}^\top)]$ 
  update training labels  $\mathbf{y}_i = [y_i, \dots, y_{i-m+1}]^\top$ 
  update test feature  $\mathbf{x}_i^* = [y_i, y_{i-1}, \dots, y_{i-d+1}]^\top$ 
  initialize  $\alpha$ , and  $\sigma^2$ 
  for  $k = 1$  to  $n_{iter}$  do
    update mean and covariance of weight vector  $\mathbf{w}$  (Eq. 3.22)
    compute new hyperparameter  $\alpha$  and  $\sigma^2$  (Eq. 3.25)
  end
  compute  $y_{i+\xi}^*$  and  $\sigma_{i+\xi}^{2*}$  (Eq. 3.26)
end

```

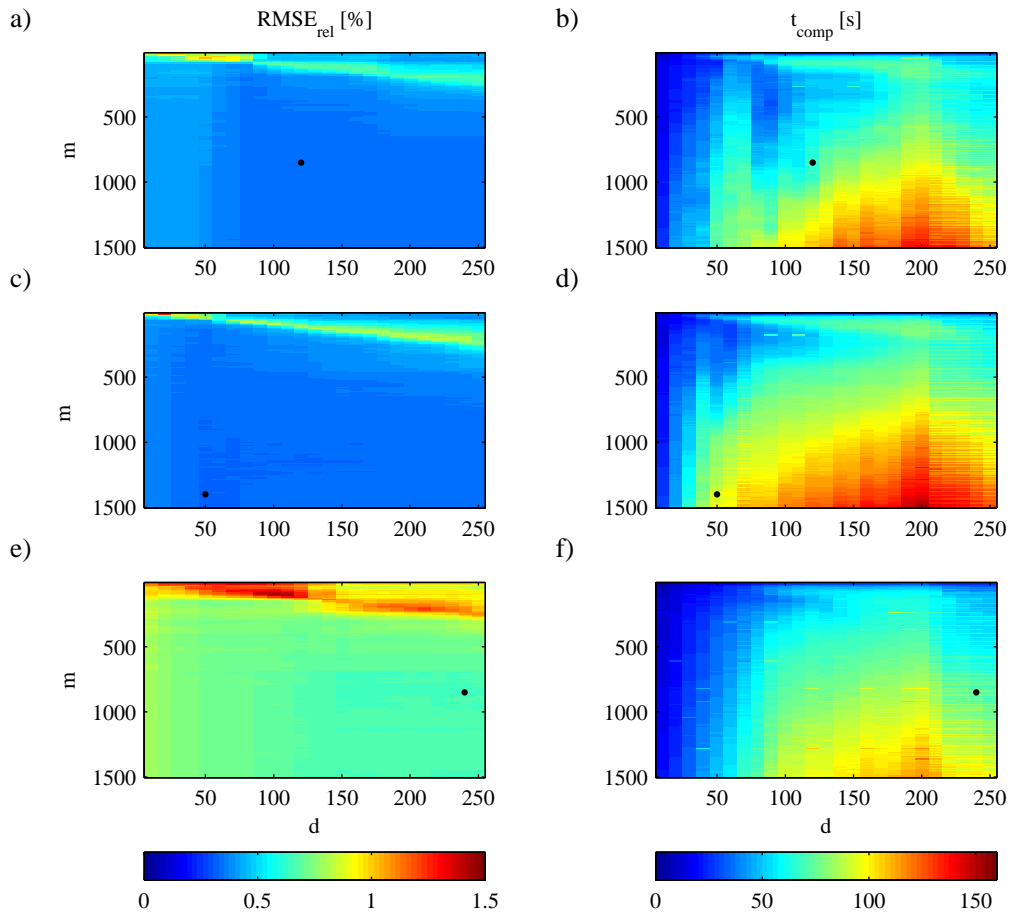
---

function, which tries to find a sparse weight vector. In contrast, Tipping proposed to use an **RBF** kernel [261] for Eq. 3.14. The prediction result is the sum of weighted kernel evaluations  $k(\mathbf{x}', \mathbf{x}_i^*)$  with  $\mathbf{x}' \in \mathbf{X}$ . In this case, the relevance vectors will indicate the relevant training features of  $\mathbf{X}$  with  $\mathbf{w} \in \mathbb{R}^m$  which would be comparable to the **SVR** algorithm. However, this approach would required further optimization of the kernel parameter, which could be done by grid search. The simpler approach was chosen as adaptive filter approaches have already been proven to be a valid approach for respiratory motion prediction.

#### 3.2.1.2 Parameter selection

For a given latency, the **RVM** tracking algorithm has three main parameters which have to be optimized. These are the dimension of the feature vector  $d$ , the number of training pairs  $m$ , and either the number of iterations  $n_{iter}$  or the probability difference threshold  $p_{th}$ . The advantage of a fixed number of iterations is that the maximum computation time can be better controlled, which is crucial in real-time applications. On the other side, a drawback is that the marginal likelihood might not have converged to a steady state, which can result in an increased prediction error. In general, a fixed threshold  $p_{th}$  could prevent this. However, the relationship between the marginal probability and the

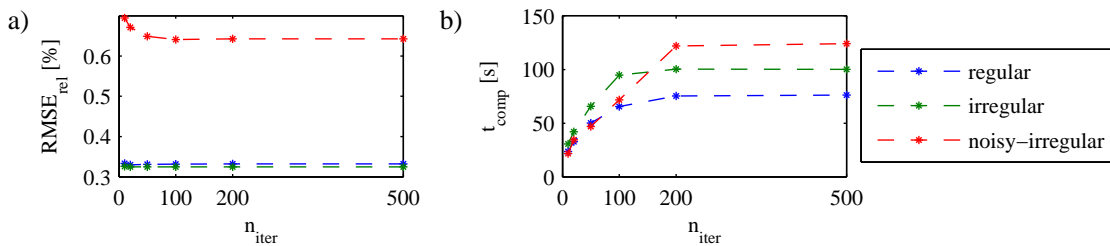




**Figure 3.4:** Relative  $\text{RMSE}_{rel}$  and computation time  $t_{comp}$  of a linear RVM algorithm for the regular (a-b), irregular (c-d) and noisy-irregular (e-f) breathing fragments. Results are shown depending on number of training pairs  $m$  and number of feature dimensions  $d$  for  $n_{iter} = 100$  iterations. (black dots indicate lowest  $\text{RMSE}_{rel}$  result)

prediction error or other evaluation measures is difficult or even impossible to express. This approach was not further investigated and a fixed number of  $n_{iter}$  was used.

**Experimental Setup** An initial experiment was performed to investigate the influence of these parameters based on the three breathing fragments presented in sec. 3.1. A linear basis function (Eq. 3.15) was selected. The feature dimension was varied between  $d \in \{10, 20, \dots, 250\}$  and the number of training pairs between  $m \in \{10, 20, \dots, 1500\}$ . The prediction horizon was set to  $\xi = 3$ , which is equivalent to the CyberKnife<sup>®</sup> Synchrony. The results were evaluated with respect to the  $\text{RMSE}_{rel}$  and the computation time  $t_{comp}$



**Figure 3.5:** Optimal  $\text{RMSE}_{rel}$  (a) of a linear RVM algorithm and respectively required computation time  $t_{comp}$  (b) depending on the number iterations  $n_{iter}$

of the complete 60 s fragment for  $n_{iter} \in \{10, 20, 50, 100, 200, 500\}$ . The algorithm was implemented in Matlab using the SBS toolbox on an office computer (i7@3.4GHz, 16GB RAM).

**Results and Discussion** Figure 3.4 shows the  $\text{RMSE}_{rel}$  and  $t_{comp}$  of the three motion fragments depending on  $m$  and  $d$  for  $n_{iter} = 100$ . The lowest  $\text{RMSE}_{rel}$  are highlighted by a black dot for each signal respectively. Comparing the  $\text{RMSE}_{rel}$  indicates that the error is comparable for the regular and irregular breathing fragment and higher for the noisy-irregular motion trace. Differences between the regular and irregular breathing fragment might be observed for the computation time (Fig. 3.4.b and 3.4.d), which seems to be increased for the irregular motion fragment.

Figure 3.5 illustrates the  $\text{RMSE}_{rel}$  and  $t_{comp}$  based on the number of iterations. For each  $n_{iter}$  the lowest  $\text{RMSE}_{rel}$  has been selected independently on  $m$  and  $d$ , retrospectively. It can be observed that the influence of  $n_{iter}$  on the  $\text{RMSE}_{rel}$  for the regular and irregular motion fragment is relatively small. In case of the noisy-irregular breathing fragment, the  $\text{RMSE}_{rel}$  decreases until  $n_{iter} = 100$  and remains constant for  $n_{iter} = \{200, 500\}$ . The results indicate that  $t_{comp}$  increases until  $n_{iter} = 200$  for all motion fragments. To be able to use the algorithm in real-time,  $t_{comp}$  should be  $\leq 60$  s. Based on this experiment, the number of iterations should be limited to approximately  $n_{iter} = 50$ . However, it has to be considered that these results represent the average computation time over 60 s. A computation time below  $1/f_s$  at each time index is not guaranteed. The reported computation times should be interpreted more as approximate values, as the algorithm was implemented in Matlab and without a special focus on the computational load. Further decrease of the computation time should be easily possible by using different programming language or hardware.

Table 3.4 shows the minimal  $\text{RMSE}_{rel}$  depending on  $n_{iter} = \{20, 50, 100\}$ . Addition-

ally,  $t_{comp}$ ,  $d$  and  $m$  are listed for the minimum  $RMSE_{rel}$ . A strong variation of  $d$  and  $m$  can be observed for the motion fragments. Intuitively, it may be expected that the number of feature dimensions increases with a more complex motion pattern. This cannot be observed in Table 3.4. However, the **RVM** algorithm performs a prediction based on a sparse feature set, consequently the number of relevance vectors is has to be investigated. The mean number of relevance vectors are  $\bar{n}_{RV} = 10.6$  for the regular,  $\bar{n}_{RV} = 13.3$  for irregular and  $\bar{n}_{RV} = 17.8$  for noisy-irregular motion fragment. This is in agreement with the above mentioned intuition.

Additionally, Table 3.4 contains the results of the **wLMS** algorithm for comparison. The parameters of the **wLMS** algorithm were specified according to [81] with three wavelet decompositions ( $J = 3$ ), a learning parameter  $\mu = 0.0204$ , and a feature dimension of  $d = 199$ . Comparing the  $RMSE_{rel}$  indicates that the **RVM** algorithm with a linear basis function can in principle outperform the **wLMS** algorithm, also for a low number of iterations. However, the **wLMS** algorithm is less computational intensive.

Further improvements might be achieved by using a different design of the feature vector  $\mathbf{x}$ . Instead of using the  $d$  most recent observations, the vector could contain observations which have a linearly or quadratically increasing space between them [80]. Additionally, a combination of the number of iterations  $n_{iter}$  and the probability difference threshold  $p_{th}$  is possible. The optimization could be stopped if the threshold is reached

**Table 3.4:** Minimum  $RMSE_{rel}$  depending on the number of iteration  $n_{iter}$  as well as the computation time  $t_{comp}$ , number of features dimension  $d$ , and number of training pairs  $m$  of linear **RVM** for a regular, irregular and noisy-irregular signal. Result of the **wLMS** algorithm are shown for comparison.

signal	<b>RVM</b>					<b>wLMS</b>	
	$n_{iter}$	$RMSE_{rel}$	$t_{comp}$	$d$	$m$	$RMSE_{rel}$	$t_{comp}$
regular	20	33.02 %	32.89 s	150	850	39.67 %	6.86 s
	50	33.12 %	50.05 s	120	850		
	100	33.18 %	65.44 s	120	850		
irregular	20	32.49 %	41.89 s	70	1290	36.49 %	6.88 s
	50	32.48 %	65.66 s	50	1450		
	100	32.45 %	94.79 s	50	1400		
noisy-irregular	20	67.11 %	34.58 s	150	990	69.46 %	6.86 s
	50	64.87 %	46.89 s	240	820		
	100	64.12 %	71.7 s	240	850		

even though the maximum number of iterations is not. This could further decrease the computation time.

#### 3.2.2 Gaussian Process Models<sup>4</sup>

The previously presented **RVM** algorithm can be seen as a special case of a **GP** model [40]. The **RVM** assumes a specific prior distribution of the weight vector  $\mathbf{w}$  which is linearly combined with the basis functions. Both, the prior distribution and the linear combination, can be interpreted as specific assumption over the model. A more general way is offered by **GP** models, which assume a prior distribution over the functions itself. A very solid introduction to **GPs** and how this model technique can be used for regression and classification tasks can be found in [208]. Similar to sec. 3.2.1, the focus of the following introduction lies on regression problems.

##### 3.2.2.1 Mathematical background

It is assumed that an observation  $y_i$  at time  $t_i$  is drawn from a Gaussian distribution, which is specified by

$$y_i = f(t_i) \sim \mathcal{N}(\mu, \sigma^2). \quad (3.27)$$

Considering  $m$  observations simultaneously, a joint Gaussian distribution can be defined according to

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (3.28)$$

with  $\boldsymbol{\mu} \in \mathbb{R}^m$  being the mean vector and  $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$  the covariance matrix. In principle infinite number of points  $m$  can be considered. A Gaussian process defines a probability distribution over functions

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (3.29)$$

where  $m(\mathbf{x})$  is the mean function and  $k(\mathbf{x}, \mathbf{x}')$  be the covariance function. The covariance function describes the coupling between two values of  $\mathbf{x}$ . However, as it is not possible to perform computations with infinite dimensions, the process is discretised with respect to  $\mathbf{x}$ .

Considering our general regression problem, a joint probability distribution over the training features and test feature can be specified. Note that in contrast to the **RVM**, we consider the feature dimension to be  $d = 1$  with  $x_i = t_i$  being the time at index  $i$  instead of an observation  $y_i$ . Considering  $m$  training pairs, the training features are defined as

---

<sup>4</sup>Parts of this section have been published in [69]

$\mathbf{x}_i = [x_i, \dots, x_{i-m+1}]^\top \in \mathbb{R}^{m \times 1}$ , the test labels as  $\mathbf{y}_i = [y_i, \dots, y_{i-m+1}]^\top \in \mathbb{R}^{m \times 1}$ , and the test feature and label as  $x_{i+\xi}^*, y_{i+\xi}^* \in \mathbb{R}$ . The joint probability distribution over the features is defined as

$$\begin{bmatrix} \mathbf{f}(\mathbf{x}_i) \\ f(x_{i+\xi}^*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{m}(\mathbf{x}_i) \\ m(x_{i+\xi}^*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{x}_i, \mathbf{x}_i) & \mathbf{k}(\mathbf{x}_i, x_{i+\xi}^*) \\ \mathbf{k}(x_{i+\xi}^*, \mathbf{x}_i) & k(x_{i+\xi}^*, x_{i+\xi}^*) \end{bmatrix}\right). \quad (3.30)$$

Here,  $\mathbf{K}(\cdot, \cdot)$  refers to a covariance matrix and  $\mathbf{k}(\cdot, \cdot)$  refers to a covariance vector with  $\mathbf{k}(\mathbf{x}_i, x_{i+\xi}^*) = \mathbf{k}(x_{i+\xi}^*, \mathbf{x}_i)^\top$  and  $\mathbf{k}(\mathbf{x}_i, x_{i+\xi}^*) \in \mathbb{R}^{m \times 1}$ . In general,  $\mathbf{K}(\cdot, \cdot)$  is defined as

$$\mathbf{K}(\mathbf{v}, \mathbf{w}) = \begin{bmatrix} k(\mathbf{v}[1], \mathbf{w}[1]) & k(\mathbf{v}[1], \mathbf{w}[2]) & \dots & k(\mathbf{v}[1], \mathbf{w}[r]) \\ k(\mathbf{v}[2], \mathbf{w}[1]) & k(\mathbf{v}[2], \mathbf{w}[2]) & \dots & k(\mathbf{v}[2], \mathbf{w}[r]) \\ \vdots & & \ddots & \vdots \\ k(\mathbf{v}[q], \mathbf{w}[1]) & k(\mathbf{v}[q], \mathbf{w}[2]) & \dots & k(\mathbf{v}[q], \mathbf{w}[r]) \end{bmatrix},$$

where  $\mathbf{v} \in \mathbb{R}^{q \times 1}$  and  $\mathbf{w} \in \mathbb{R}^{r \times 1}$ .

As the training labels  $\mathbf{y}_i$  are known, the conditional probability  $p(y_{i+\xi}^* | x_{i+\xi}^*, \mathbf{x}_i, \mathbf{y}_i)$  of the test label  $y_{i+\xi}^*$  given  $x_{i+\xi}^*, \mathbf{x}_i$ , and  $\mathbf{y}_i$  can be defined. The result will also be a Gaussian distribution

$$p(y_{i+\xi}^* | x_{i+\xi}^*, \mathbf{x}_i, \mathbf{y}_i) \sim \mathcal{N}(\bar{y}^*, \text{var}[y^*]). \quad (3.31)$$

where the mean and variance are given by

$$\bar{y}^* = m(x_{i+\xi}^*) - \mathbf{k}(x_{i+\xi}^*, \mathbf{x}_i) \mathbf{K}(\mathbf{x}_i, \mathbf{x}_i)^{-1} (\mathbf{y}_i - \mathbf{m}(\mathbf{x}_i)), \quad (3.32)$$

$$\text{var}[y^*] = k(x_{i+\xi}^*, x_{i+\xi}^*) - \mathbf{k}(x_{i+\xi}^*, \mathbf{x}_i) \mathbf{K}(\mathbf{x}_i, \mathbf{x}_i)^{-1} \mathbf{k}(\mathbf{x}_i, x_{i+\xi}^*). \quad (3.33)$$

Without loss of generality the mean function can be set to zero, as every mean function can be expressed within the covariance functions [208]. So far the model does not consider observation noise. If a regression model such as  $y = f(x) + \varepsilon$  with  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$  and  $\sigma^2$  being the noise variance is assumed, the GP model can be extended to

$$\begin{bmatrix} \mathbf{y}_i \\ y_{i+\xi}^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}(\mathbf{x}_i, \mathbf{x}_i) + \sigma^2 \mathbf{I} & \mathbf{k}(\mathbf{x}_i, x_{i+\xi}^*) \\ \mathbf{k}(x_{i+\xi}^*, \mathbf{x}_i) & k(x_{i+\xi}^*, x_{i+\xi}^*) \end{bmatrix}\right), \quad (3.34)$$

where  $\mathbf{0}$  is a zero vector of length  $m$  and  $\mathbf{I} \in \mathbb{R}^{m \times m}$  is the identity matrix. The mean and variance of  $p(y_{i+\xi}^* | x_{i+\xi}^*, \mathbf{x}_i, \mathbf{y}_i)$  is specified by

$$\bar{y}^* = \mathbf{k}(x_{i+\xi}^*, \mathbf{x}_i) [\mathbf{K}(\mathbf{x}_i, \mathbf{x}_i) + \sigma^2 \mathbf{I}]^{-1} \mathbf{y}_i, \quad (3.35)$$

$$\text{var}[y^*] = k(x_{i+\xi}^*, x_{i+\xi}^*) - \mathbf{k}(x_{i+\xi}^*, \mathbf{x}_i) [\mathbf{K}(\mathbf{x}_i, \mathbf{x}_i) + \sigma^2 \mathbf{I}]^{-1} \mathbf{k}(\mathbf{x}_i, x_{i+\xi}^*). \quad (3.36)$$

It is assumed that the test label  $y^* = \bar{y}^*$ . It can be observed that  $y^*$  (Eq. 3.35) and its variance (Eq. 3.36) only depend on the covariance function and the variance of the observation noise. Similar to kernel functions in case of the SVR, covariance functions have to fulfil the Mercer theorem [208]. In general, covariance functions encode our prior believe of the functional behaviour which shall be modelled. In case of time-series modelling such prior knowledge can be for example with respect to the smoothness or periodicity of the signal. Rasmussen and Williams gave an overview over various covariance functions in [208]. Within this work five covariance functions are considered, namely the squared-exponential (SE), rational quadratic (RQ), Matérn (M), periodic (P) and white noise (N) covariance functions, which are defined as

$$k_{SE}(r) = \theta_S^2 \exp \left\{ -\frac{r^2}{2\theta_L^2} \right\}, \quad (3.37)$$

$$k_M(r) = \theta_S^2 \left( 1 + \frac{\sqrt{3}r}{\theta_L} \right) \exp \left\{ -\frac{\sqrt{3}r}{\theta_L} \right\}, \quad (3.38)$$

$$k_{RQ}(r) = \theta_S^2 \left( 1 + \frac{r^2}{2\theta_D\theta_L} \right)^{-\theta_D}, \quad (3.39)$$

$$k_P(r) = \theta_S^2 \exp \left\{ -2 \sin^2[(\pi/\theta_P)r] \right\}, \quad (3.40)$$

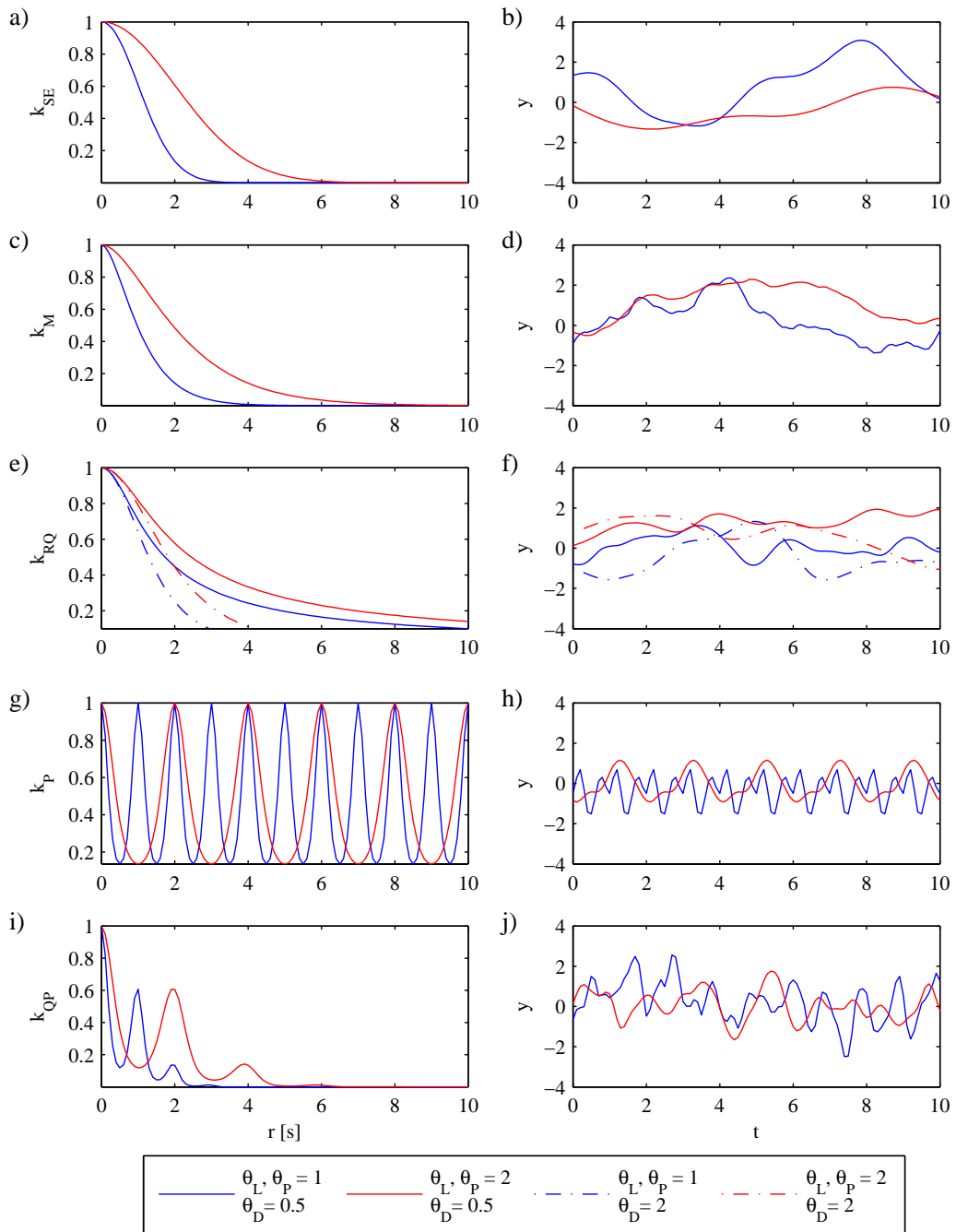
$$k_N(\mathbf{x}[p], \mathbf{x}[q]) = \theta_S^2 \delta(p, q), \quad (3.41)$$

where  $\theta_S, \theta_L, \theta_P \in \mathbb{R}$  and  $\theta_D \in \mathbb{Z}$  are hyperparameters which model the  $y$ -scaling, the  $t$ -scaling, the period, and the degree of the covariance functions, respectively. Except for the noise covariance function  $k_N$ , all functions depend on  $r = \|x - x'\|_2$ , which denotes the Euclidean distance between two time points. The noise covariance function, depends on the Kronecker delta, which is  $\delta(q, p) = 1$  if  $q = p$  and  $\delta(q, p) = 0$  otherwise. Note that the Matérn function (Eq. 3.38) is just one frequently used spacial case of a group of Matérn functions.

These covariance functions can be considered as “basic” covariance functions. More “complex” covariance functions can be designed by summation, multiplication or convolution of two or more basic covariance functions. A frequently used example is the quasi-periodic covariance function  $k_{QP}$  which is a multiplication of a squared-exponential and periodic function

$$k_{QP}(r) = \theta_S^2 \exp \left\{ -\frac{r^2}{2\theta_L^2} \right\} \times \exp \left\{ -\frac{\sin^2[(2\pi/\theta_P)r]}{2} \right\}. \quad (3.42)$$

In the following section 3.2.2.2 various basic and complex covariance functions will be presented and evaluated on the three breathing fragments of sec. 3.1. Here, the basic



**Figure 3.6:** Output of covariance functions depending on the feature distance  $r$  (left column) and examples functions drawn from the GP models respectively (right column) for (a-b) a squared exponential  $k_{SE}$ , (c-d) a Matérn  $k_M$ , (e-f) a rational quadratic  $k_{RQ}$ , (g-h) a periodic  $k_P$ , and (i-j) a quasi-periodic  $k_{QP}$  covariance function for different hyperparameters  $\theta$ .

properties of the covariance functions of Eq. 3.37-3.40 and 3.42 shall be illustrated. Figure 3.6 illustrates the output of the covariance functions depending on  $r$  in the left column. The hyperparameters have been exemplarily chosen as  $\theta_L, \theta_P = \{1, 2\}$ ,  $\theta_D = \{0.5, 2\}$ , and  $\theta_S = 1$ . The figures in the right column show drawn example functions from these GPs, respectively. The function value of  $k_{SE}$ ,  $k_M$ , and  $k_{RQ}$  is 1 for  $r = 0$  (Fig. 3.6.a, 3.6.c, 3.6.e). With increasing distance  $r$  the output of these covariance functions decreases towards zero. How fast the function values decrease is influenced by the hyperparameters  $\theta_L$  and  $\theta_D$ . These covariance functions can be used to model the smoothness of a signal. For example in case of  $k_{SE}$ , for  $\theta_L = 2$  (red line), the correlation between two data points with a distance  $r$  is higher than for  $\theta_L = 1$  (blue line) (Fig. 3.6.a). Consequently, drawn examples from this process with  $\theta_L = 2$  are more likely to be smoother as for  $\theta_L = 1$ , which can be observed in Fig. 3.6.b. Which covariance functions is most suited for a specific applications, depends in general only on the data. However, Stein [255] argued that the smoothness assumptions of a SE covariance functions is to often too strong for real applications and suggested to use functions of the Matérn class.

Fig. 3.6.e shows the output of the periodic covariance functions depending on  $r$  which has a periodic pattern with a period of  $\theta_P$ . Consequently, example functions drawn from such a process will have a constant periodic pattern as well (Fig. 3.6.f). For practical applications, this covariance function will most likely result in high prediction errors as the breathing pattern changes over time. A more flexible function is the quasi-periodic covariance function shown in Fig. 3.6.g which combines the smoothing properties of a  $k_{SE}$  and periodic behaviour of a  $k_P$  function. Samples drawn from this process are shown in Fig. 3.6.h. Further quasi-periodic covariance functions can be generated by multiplication of  $k_M$  and  $k_p$  with  $k_p$ .

The remaining challenge is to find the optimal hyperparameters. This can be achieved in a similar manner as in sec. 3.2.1 by maximising the marginal likelihood  $p(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta})$  with  $\boldsymbol{\theta}$  a vector containing all hyperparameters. Assuming a mean of zero and observation noise, the marginal likelihood is defined according Eq. 3.34 as  $p(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta}) \sim \mathcal{N}(0, \mathbf{K}(\mathbf{x}_i, \mathbf{x}_i) + \sigma^2\mathbf{I})$ . In practise, the **negative logarithmic marginal likelihood (NLML)** is minimized

$$-\log p(\mathbf{y}_i|\mathbf{x}_i, \boldsymbol{\theta}) = \frac{1}{2}\mathbf{y}_i^\top (\mathbf{K}(\mathbf{x}_i, \mathbf{x}_i) + \sigma^2\mathbf{I})^{-1} \mathbf{y}_i + \frac{1}{2} \log |\mathbf{K}(\mathbf{x}_i, \mathbf{x}_i) + \sigma^2\mathbf{I}| + \frac{m}{2} \log 2\pi. \quad (3.43)$$

The **NLML** consist of three summands. The latter is a constant which is based on the number of training pairs  $m$ . Only the first two term are influenced by varying the hyperparameters. The first terms can be interpreted as an error term, as it is based on the training features  $\mathbf{x}_i$  and labels  $\mathbf{y}_i$ . It decreases if the training error is small. The second



**Table 3.5:** Outline of a GP tracking algorithm with fixed number of iteration  $n_{iter}$ 


---

```

specify covariance function  $k(\cdot, \cdot)$ 
initialize  $m, n_{iter}$ , and  $\theta$ 
for  $i = 1$  to  $N$  do
  update training features  $\mathbf{x}_i = [t_i, \dots, t_{i-m+1}]^\top$ 
  update training labels  $\mathbf{y}_i = [y_i, \dots, y_{i-m+1}]^\top$ 
  update test feature  $x_{i+\xi}^* = t_{i+\xi}$ 
  for  $l = 1$  to  $n_{iter}$  do
    update hyperparameters  $\theta$ 
  end
  compute  $y_{i+\xi}^*$  and  $\sigma_{i+\xi}^{2*}$ 
end

```

---

term can be interpreted as a model complexity term as it depends only on the training features. The term increases if the correlation between the training features decreases, which is equivalent with an higher model complexity. A result of these two terms is that by minimizing the NLML a bias-variance trade off is performed.

The NLML can be optimized with respect to the hyperparameters by using a gradient descent method as it is also implemented in the GPML toolbox [207]. Similar to the RVM tracking algorithm in sec. 3.2.1, the optimization step could be either performed based on a number of maximum iterations steps  $n_{iter}$  or a probability threshold  $p_{th}$ . The essential steps for a GP tracking algorithm are presented in Table 3.5.

To summarize, it shall be highlighted that the GP algorithm as outlined in this section uses time points  $t$  as input features and not observations  $y$ . This is different compared to the RVM algorithm and most of the prediction algorithms presented in sec. 2.3.1. Advantages of this approach are:

- *Reduced feature dimension:* The dimension of a feature  $x$  is  $d = 1$ . Other algorithms, such as RVM and SVR, have a feature dimension of  $d > 1$  as they use observations  $y$  as features. No additional optimization is required which leads to decreased computational and storage requirements (although it is possible to extent it to  $d > 1$  if necessary).
- *Compensation of missing data:* As a result of the first point, the GP models can compensate missing training data. In case one training observation  $y_o$  is missing at time

index  $o$ , only the training pair  $\{y_o, x_o\}$  has to be removed from the training set. In case of algorithms with  $d > 1$ ,  $y_o$  might be present within multiple training pairs. Consequently, all training pairs containing  $y_o$  would have to be removed. Furthermore, if  $y_o \in \mathbf{x}^*$ , no prediction could be performed.

- *Arbitrary prediction latencies:* Most of the algorithms of sec. 3.2.1.2 can only perform multi-step ahead predictions, meaning that  $h$  depends on the sampling frequency  $f_s$  and possible latencies are  $h = \xi/f_s$  with  $\xi \in \mathbb{Z}$ . The GP algorithm can predict arbitrary prediction latencies independently of the sampling frequency.
- *Multiple prediction latencies:* The previous points can be further extended to the prediction of multiple latencies with the same model and without adaptation of the training pairs. This could be for example relevant for systems where multiple motion compensation systems are used such as a combination of a MLC and a moving patient couch.
- *Reduction of computation requirements:* Note that all discussed covariance functions are stationary covariance functions, meaning that the functions are invariant to translations in the input space [208]. This property could be of practical relevance to decrease computational requirements. Let us assume that  $t$  is equidistantly sampled. The covariance matrices  $\mathbf{K}$  in Eq. 3.35 and 3.36 remains constant at each time step, as the differences between feature values  $r$ , which is equivalent to the difference between sampled time steps, remains constant. The matrices and also the results of the inverse of  $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_i)$  can be stored to reduce computational requirements. However, this assumes that the hyperparameter remains constant, which is not valid for respiratory motion. It has to be further investigated how frequently the hyperparameters have to be updated.

#### 3.2.2.2 Parameter Selection

As presented in the previous section, the most relevant parameters of the GP algorithm are the selected covariance function, the number of training pairs  $m$ , and the number of iterations  $n_{iter}$ . Within this section these parameters will be evaluated in two experiments on the motion fragments presented in sec. 3.1.

A similar and more detailed evaluation of the GP parameters for the purpose of respiratory motion prediction was performed in [88].

### Experiment 1: Covariance selection

The first experiment focuses on the model selection - the choice of the optimal covariance function or combination of functions. One possibility to investigate this is the evaluation of the **NLML**. The marginal likelihood represents the likelihood that the observed training data can be modelled given a certain covariance function. The advantage of using the **NLML** as model selection criterion, instead of the  $\text{RMSE}_{rel}$ , is that the model selection can be investigated independently from the prediction horizon.

**Experimental Setup** Beside the six covariance functions presented in sec. 3.2.2.1, all possible additive combinations were considered. The combinations of a periodic and a quasi-periodic covariance functions were ignored as it can be assumed that only one periodic component is present in the signal. We refer to each of the combinations by a label which consists of the first letter of the individual covariance functions. For example, a combination of a periodic, squared-exponential and Matérn covariance function and a combination of quasi-periodic, Matérn and rational quadratic covariance function are defined as

$$k_{PSM}(r) = k_P(r) + k_{SE}(r) + k_M(r), \quad k_{QMR}(r) = k_{QP}(r) + k_M(r) + k_{RQ}(r). \quad (3.44)$$

The combinations are labelled as the PSM and the QMR covariance function.

For each motion fragment, the complete dataset was considered as known training data ( $m = 1560$ ). The motion traces were scaled to a range of  $y \in [0, 1]$ . Further, the data were corrected by their mean value to enable the assumption of a zero mean function. The hyperparameter of the period was randomly initialized by  $\theta_P \sim \mathcal{U}(0, 5)$  and the residual hyperparameters by  $\theta \sim \mathcal{U}(0, 1)$ . The number of gradient descent steps to optimize the hyperparameters was fixed to  $n_{iter} = 200$ . To avoid local minima, the hyperparameters optimization was repeated ten times for each covariance combination.

**Results and Discussion** Table 3.6 shows the lowest **NLML** for the three breathing segments and the average **NLML** over all three segments. The table is divided into five sections depending on the number of combined covariance functions. The lowest **NLML** of each section is highlighted bold. The lowest **NLML** value over all covariance functions is marked by an asterisk.

On average, it might be observed that the **NLML** is lower for the regular and slightly irregular breathing fragments as for the noisy-irregular breathing fragment. The best individual covariance function on average (first section of Table 3.6) is the Matérn covariance function. This is in agreement with the argument of Stein [255]. However, using

**Table 3.6:** Optimized NLM values depending on the covariance functions for a regular, a irregular, and a noisy irregular motion fragment. Considered covariance functions are the squared-exponential (S), Matérn (M), rational quadratic (R), periodic (P), and noise (N) function and all their additive combinations. (best result per number of covariance functions and motion trace highlighted in bold; the best overall result per motion trace is marked by asterisk)

covariance	NLM				average	function	covariance	NLM				average	function		
	regular	irregular	noisy-irregular	noisy-average				regular	irregular	noisy-irregular	noisy-average				
P	-1258	-5023	-711	-2331	P	PRN	-5034	-5636	-4057	-4909	PRN	-5034	-5636	-4057	-4909
Q	-4846	-5604	-4062	-4838	Q	PMN	-5163	-5786	-4105	-5018	Q	-5163	-5786	-4105	-5018
S	-4677	-5023	-3916	-4539	S	QSR	-5150	<b>-5837</b>	<b>-4187</b>	<b>-5058</b>	S	-5150	<b>-5837</b>	<b>-4187</b>	<b>-5058</b>
R	<b>-4865</b>	-5557	-4043	-4822	R	QSM	-5157	-5816	-4142	-5038	R	<b>-4865</b>	-5557	-4043	-4822
M	-5071	<b>-5734</b>	<b>-4095</b>	<b>-4967</b>	M	QSN	-5139	-5831	-4119	-5030	M	-5071	<b>-5734</b>	<b>-4095</b>	<b>-4967</b>
N	231	73	-273	10	N	QRM	-5134	-5826	-4173	-5044	N	231	73	-273	10
PS	-4835	-5082	-3936	-4618	PS	QRN	-5143	-5826	-4142	-5037	PS	-4835	-5082	-3936	-4618
PR	-5032	-5646	-4062	-4913	PR	QMN	-5152	-5824	-4176	-5051	PR	-5032	-5646	-4062	-4913
PM	-5137	-5787	-4127	-5017	PM	SRM	-5147	-5790	-4123	-5020	PM	-5137	-5787	-4127	-5017
PN	-1255	-436	-3156	-1616	PN	SRN	-5119	-5793	-4103	-5005	PN	-1255	-436	-3156	-1616
QS	-5139	-5792	-4119	-5017	QS	SMN	-5116	-5784	-4114	-5005	QS	-5139	-5792	-4119	-5017
QR	<b>-5145</b>	<b>-5811</b>	-4131	<b>-5029</b>	QR	RMN	-5130	-5784	-4108	-5007	QR	<b>-5145</b>	<b>-5811</b>	-4131	<b>-5029</b>
QM	-5072	-5746	<b>-4179</b>	-4999	QM	PSRM	-5171	-5815	-4127	-5038	QM	-5072	-5746	<b>-4179</b>	-4999
QN	-4884	-5493	-4070	-4816	QN	PSRN	<b>-5207*</b>	-5818	-4102	-5043	QN	-4884	-5493	-4070	-4816
SR	-5126	-5788	-4103	-5006	SR	PSMN	-5156	-5798	-4107	-5020	SR	-5126	-5788	-4103	-5006
SM	-5116	-5784	-4107	-5002	SM	PRMN	-5158	-5791	-4099	-5016	SM	-5116	-5784	-4107	-5002
SN	-4677	-5032	-3916	-4542	SN	QSRM	-5154	-5844	-4171	-5056	SN	-4677	-5032	-3916	-4542
RM	-5130	-5784	-4108	-5007	RM	QSRN	-5169	<b>-5851*</b>	-4171	<b>-5064*</b>	RM	-5130	-5784	-4108	-5007
RN	-4865	-5557	-4043	-4822	RN	QSMN	-5123	-5792	<b>-4204*</b>	-5040	RN	-4865	-5557	-4043	-4822
MN	-5071	-5734	-4095	-4967	MN	QRMN	-5150	-5823	-4192	-5055	MN	-5071	-5734	-4095	-4967
PSR	<b>-5203</b>	-5819	-4114	-5045	PSR	SRMN	-5147	-5807	-4123	-5026	PSR	<b>-5203</b>	-5819	-4114	-5045
PSM	-5149	-5786	-4122	-5019	PSM	PSRMN	<b>-5170</b>	-5786	-4118	-5025	PSM	-5149	-5786	-4122	-5019
PSN	-4846	-5151	-3935	-4644	PSN	QSRMN	-5162	<b>-5823</b>	<b>-4137</b>	<b>-5041</b>	PSN	-4846	-5151	-3935	-4644
PRM	-5163	-5799	-4120	-5027	PRM						PRM	-5163	-5799	-4120	-5027

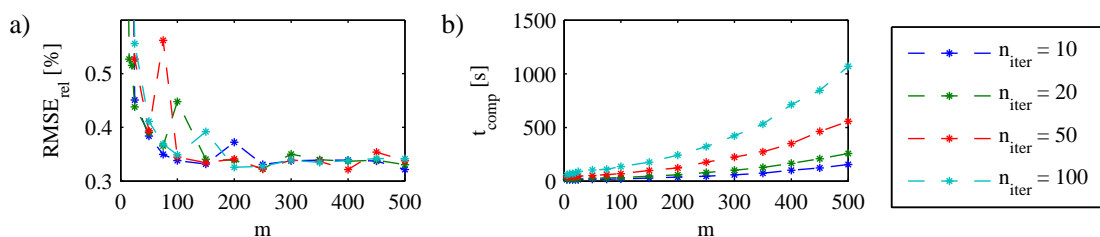
combinations of covariance functions (section 3 to 4 of Table 3.6), the SE function seems to be superior as it is part of the covariance functions combination resulting in the lowest **NLML** for these sections. Comparing the performance of the periodic and the quasi-periodic covariance function, indicates that respiratory motion can be better modelled with a quasi-periodic covariance function as it does result in lower **NLML** values.

Comparing the best results of all covariance functions reveals that the QSRN covariance function has the lowest **NLML** value on average followed by the QSR covariance function. As the optimization was only performed on the training data, it might be expected that the most complex covariance functions combination (QSRMN or PSRMN) would result in the lowest **NLML** value. However, this is not the case due to the separate error and model complexity term in Eq. 3.43. By adding an additional individual covariance function, the increase of the model complexity term is stronger compared to the decrease of the error term. This result demonstrates the bias-variance trade of the **NLML** optimization.

A comparable evaluation on different motion fragments was performed in [69, 88]. There, the lowest **NLML** value was achieved by the QRM covariance function. Here, this covariance function has only the fifth lowest **NLML** value on average. The differences can be explained by the different investigated motion fragments as well as the different hyperparameter optimization process. In [69, 88], a fixed set of initial hyperparameters was used. The optimization is more susceptible to local minima as the multiple random hyperparameter initialisation used in this experiment.

### Experiment 2: Number training pairs $m$ and iterations $n_{iter}$

**Experimental Setup** The aim of the second experiment was to investigate the influence of basic GP model parameters. Based on the results of the previous experiment, GP models were initialised using a QSRN covariance function. The number of training pairs



**Figure 3.7:** (a)  $RMSE_{rel}$  and (b) computation time  $t_{comp}$  of a GP algorithm depending on the number of training pairs  $m$  and iterations  $n_{iter}$  for the irregular motion fragment.

was varied between  $m \in \{5, 10, 15, 20, 25, 50, 75, 100, 150, 200, \dots, 500\}$  and the number of iterations between  $n_{iter} \in \{10, 20, 50, 100\}$ . The hyperparameters were 40 times randomly initialised with  $\theta_P \sim \mathcal{U}(0, 5)$  and  $\theta \sim \mathcal{U}(0, 1)$  for the residual hyperparameters. The set of hyperparameters resulting in the lowest **NLML** was selected as initial set. To reduce the computational load, an update of the hyperparameters was only performed every  $u = 20$  time steps. The results of the three motion fragments were evaluated with respect to the **RMSE<sub>rel</sub>** and the computation time  $t_{comp}$ . The algorithm was implemented in Matlab on an office computer (i7@3.4GHz, 16GB RAM).

**Results and Discussion** Figure 3.7 shows the **RMSE<sub>rel</sub>** and  $t_{comp}$  of the **GP** algorithm for the irregular motion fragment depending on  $m$  and  $n_{iter}$ . Increasing the number of training pairs leads to a decrease of the **RMSE<sub>rel</sub>** and increase of  $t_{comp}$ . Variations of the **RMSE<sub>rel</sub>** are visible for  $m < 300$ . For  $m \geq 300$ , the influence of the number of iterations  $n_{iter}$  seems to be small on the **RMSE<sub>rel</sub>**.

Table 3.7 shows the minimal **RMSE<sub>rel</sub>** depending on  $n_{iter} = \{10, 20, 50\}$ . Additionally,  $t_{comp}$  and  $m$  are listed for the minimum **RMSE<sub>rel</sub>**. Similar to the evaluation for the **RVM** algorithm (see Table 3.4), the results are compared to the **wLMS** algorithm (parameters were selected according to [81]). Comparing the **RMSE<sub>rel</sub>** indicates that the accuracy of the **GP** algorithm is comparable (for the regular motion fragment) or superior (for the irregular and noisy-irregular motion fragment) compared to the **wLMS** algorithm. How-

**Table 3.7:** Minimum **RMSE<sub>rel</sub>** depending on the number of iteration  $n_{iter}$  as well as the computation time  $t_{comp}$  and number of training pairs  $m$  of a **GP** algorithm for a regular, irregular and noisy-irregular signal. Result of the **wLMS** algorithm are shown for comparison.

signal	GP				wLMS	
	$n_{iter}$	RMSE <sub>rel</sub>	$t_{comp}$	$m$	RMSE <sub>rel</sub>	$t_{comp}$
regular	10	40.26 %	54.40 s	300	39.67 %	6.86 s
	20	39.15 %	113.21 s	350		
	50	40.27 %	324.81 s	400		
irregular	10	32.17 %	155.51 s	500	36.49 %	6.88 s
	20	32.30 %	80.60 s	250		
	50	32.14 %	349.48 s	400		
noisy-irregular	10	62.27 %	63.41 s	300	69.46 %	6.86 s
	20	62.04 %	103.12 s	300		
	50	61.49 %	383.24 s	400		

ever, the GP algorithm is computationally more intensive. Further investigations also about the influence of the update factor  $u$  on the  $\text{RMSE}_{rel}$  and  $t_{comp}$  can be found in [88].

### 3.3 Comparative Evaluation<sup>5</sup>

In sec. 3.2, two new probabilistic prediction algorithms, the RVM and the GP model, have been presented. Their main parameters were evaluated exemplarily on three short motion fragments. Within this section, the two algorithms are evaluated on the complete dataset of Ernst *et al.* [86] (sec. 3.1) to make a comparison to other previously investigated algorithms possible.

#### 3.3.1 Evaluation Procedure

The first minute of each motion trace was considered as training set. The remaining data were the test set. The training set was scaled to  $[0, 1]$ . The scaling parameters were used further to scale the data of the test set.

As in [86], the prediction latency was set to  $h = \{77, 154, 308\}$  ms which corresponds to a  $\xi = \{2, 4, 8\}$  step ahead prediction. Additionally, a prediction horizon of  $h = 115$  ms ( $\xi = 3$ ) was evaluated as this correspond to the current prediction latency of the CyberKnife<sup>®</sup> [134, 227].

The RVM algorithm was initialized with  $d = 100$ ,  $m = 1000$ , and  $n_{iter} = 100$ . The linear, quadratic, and cubic basis function were used in this study according to Eq. 3.15 to which we refer to as  $\text{RVM}_{lin}$ ,  $\text{RVM}_{quad}$ , and  $\text{RVM}_{cub}$ . In case of the GP algorithm, the QRSN covariance function was selected for all motion traces according to the results of sec. 3.2.2.2. Furthermore, the QRM covariance function was evaluated, which is the sum of a quasi-periodic, rational quadratic, and Matérn covariance function. We refer to the two GP models as  $\text{GP}_{QRSN}$  and  $\text{GP}_{QRM}$ . The QRM covariance function was the optimal covariance function in previous investigations on respiratory motion fragments [88]. According to [88], the residual GP parameters were specified by  $m = 500$  and  $n_{iter} = 20$ . Further, to reduce the computational load, an update of the hyperparameters was only performed every  $u = 20$  time steps. To prevent the influence of local minimal during the hyperparameter optimization, the initial hyperparameter were ten times randomly initialised and trained on the training set. The optimal initial hyperparameters were selected based on the lowest NLML value.

<sup>5</sup>Parts of this section have been published in [67, 69, 86]

The prediction outcome was compared to the results presented in [86]. To avoid clutter, only the **wLMS**, **MULIN**, and **SVR** algorithm were considered for comparison, as they outperformed the other algorithms. All algorithms were evaluated with respect to the **RMSE** and the  $\text{RMSE}_{rel}$ . Furthermore, the relative RMSE difference  $\Delta\text{RMSE}_{rel}$  is defined as

$$\Delta\text{RMSE}_{rel} = \text{RMSE}_{rel} - \text{RMSE}_{rel}(\text{wLMS}), \quad (3.45)$$

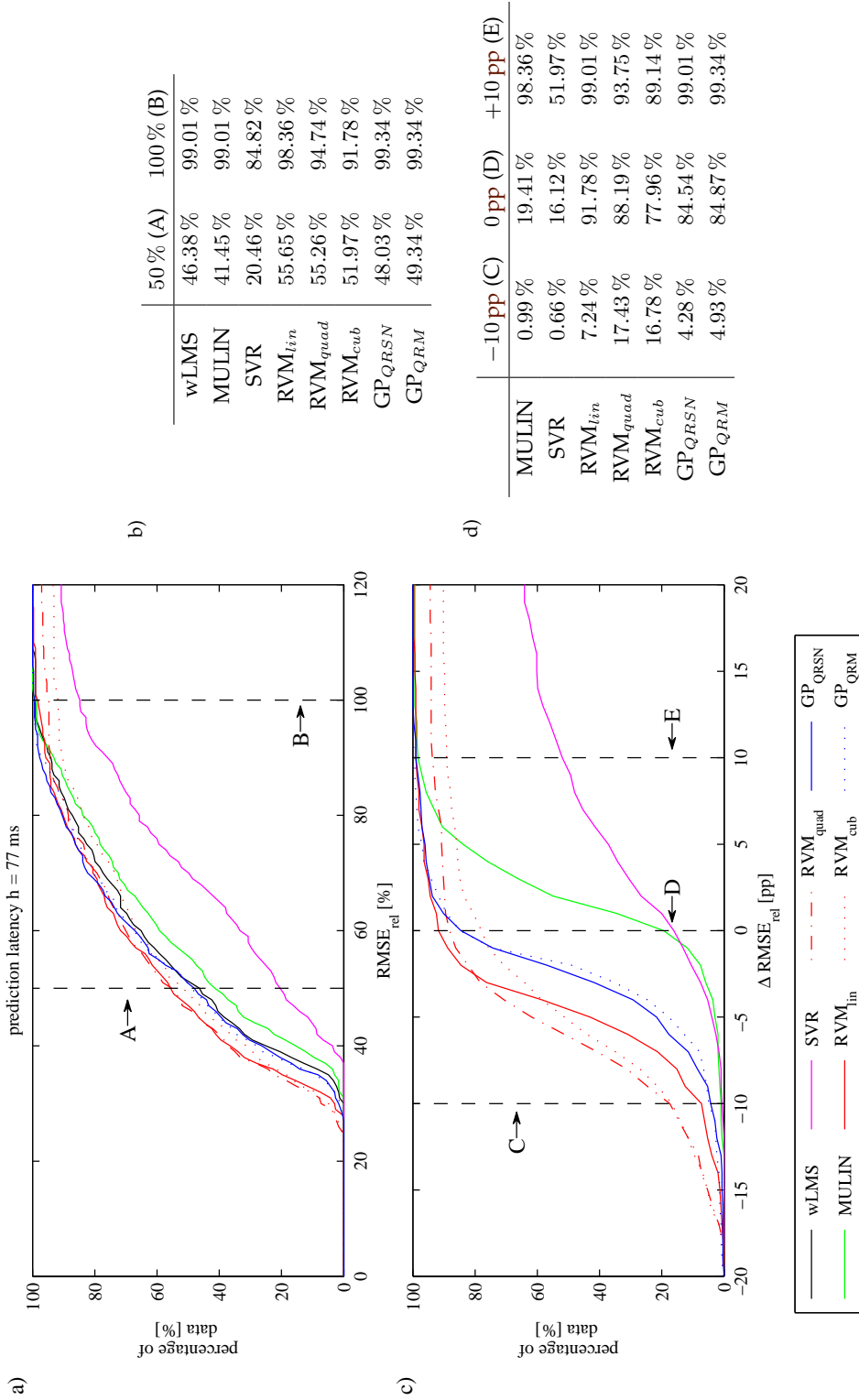
which represents the  $\text{RMSE}_{rel}$  difference with respect to the **wLMS** algorithm and is presented in **percentage points (pp)**. The **wLMS** algorithm was selected as reference algorithm, as it did results in the lowest prediction error on average. Note that in [86] only  $h = \{77, 154, 308\}$  ms have been evaluated. Consequently, the prediction results of the **wLMS**, **MULIN**, and **SVR** algorithm were not available for  $h = 115$  ms. The results of the **wLMS** algorithm were additionally computed for  $h = 115$  ms to enable the computation of  $\Delta\text{RMSE}_{rel}$ . The **wLMS** parameters were selected according to [86].

#### 3.3.2 Results

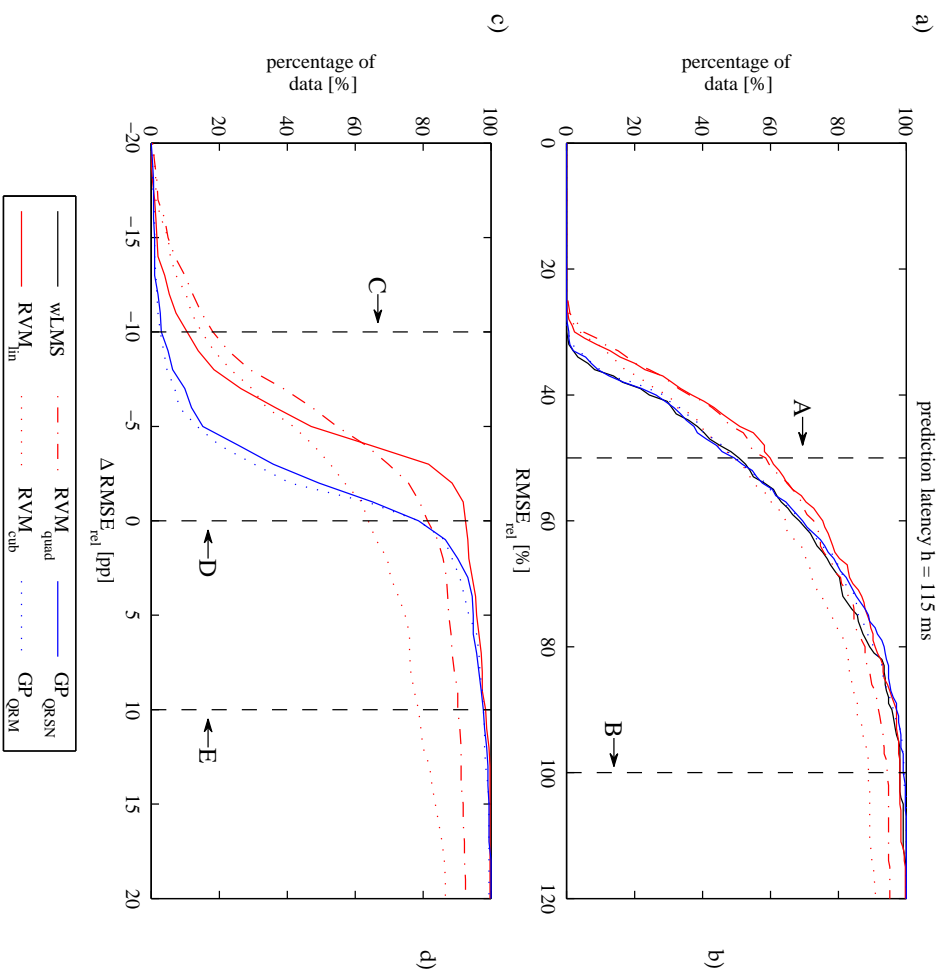
The figures 3.8 to 3.11 show the prediction outcome of the **RVM**, **GP**, and the previously published algorithms for a prediction latency  $h$  of 77 ms (Fig. 3.8), 115 ms (Fig. 3.9), 154 ms (Fig. 3.10), and 308 ms (Fig. 3.11). The upper diagram of each figure presents the  $\text{RMSE}_{rel}$  values as a cumulative histogram for the entire dataset. The histogram displays the percentage of data for which an algorithm could achieve an  $\text{RMSE}_{rel}$  equal or smaller than a certain value. The percentage of data along the two dashed lines A and B are listed in the upper right table of each figure. This way of presenting the results gives insight into the distribution of the  $\text{RMSE}_{rel}$ . Figure 3.8.a shall be considered as an example. The percentage of motion traces which could be predicted with an error of  $\text{RMSE}_{rel} \leq 50\%$  (marked by the left vertical dashed line) is: 20.46% for **SVR**, 41.45% for **MULIN**, 46.38% for **wLMS**, 48.03% for  $\text{GP}_{QRSN}$ , 49.34% for  $\text{GP}_{QRM}$ , 51.97% for  $\text{RVM}_{cub}$ , 55.26% for  $\text{RVM}_{quad}$ , and 55.59% for  $\text{RVM}_{lin}$ . The results indicate a superior performance of the three **RVM** approaches. However, investigating the percentage of motion traces which have an  $\text{RMSE}_{rel} \leq 100\%$  (marked by the right vertical dashed line) reveals a different ranking. The highest percentage could be achieved by the two **GP** approaches with 99.34% followed by the **wLMS** and **MULIN** algorithm with 99.01%. The **RVM** algorithm could predict only 98.36% ( $\text{RVM}_{lin}$ ), 94.74% ( $\text{RVM}_{quad}$ ), and 91.78% ( $\text{RVM}_{cub}$ ) of the data with an  $\text{RMSE}_{rel} \leq 100\%$ . The worst performance was achieved by the **SVR** algorithm with only 84.82%.

The bottom diagram of each figure shows the cumulative histogram of the relative





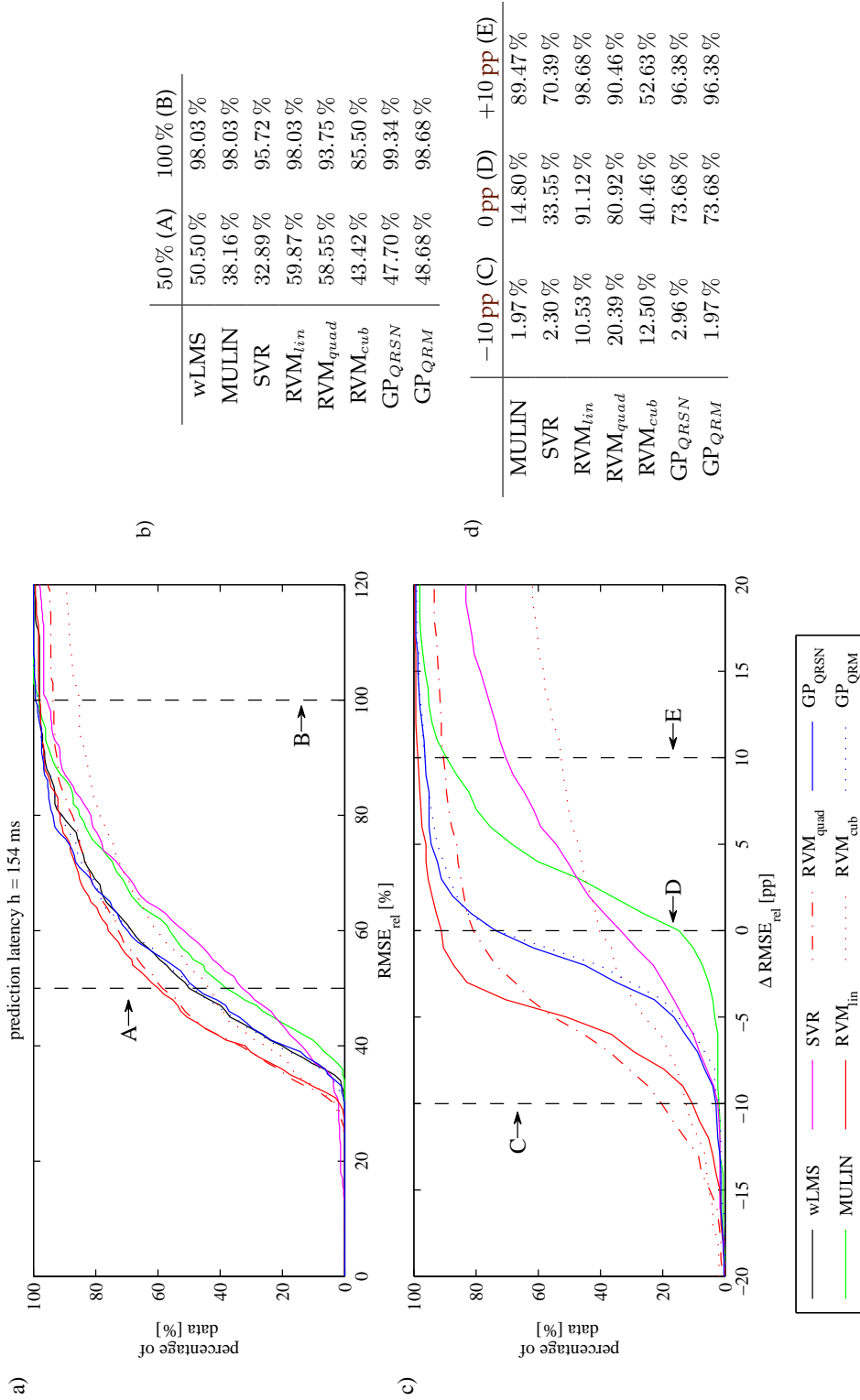
**Figure 3.8:** Prediction outcome for a latency of  $h = 77$  ms; a) Cumulative  $RMSE_{rel}$  histogram evaluated with wLMS, MULIN, SVR, three RVM, and two GP algorithms; values along line A and B are listed in b); c) Cumulative  $\Delta RMSE_{rel}$  histogram of all algorithms with respect to the wLMS algorithm; values along line C, D, and E are listed in d).



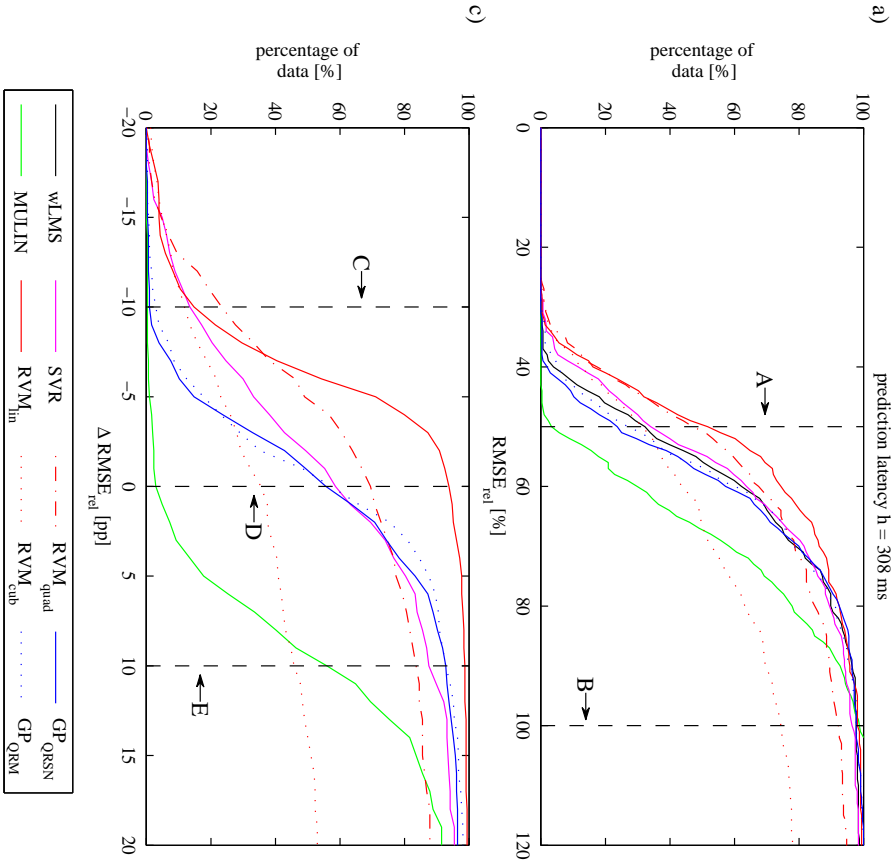
	50% (A)	100% (B)
wLMS	50.66%	98.03%
RVM <sub>lin</sub>	59.87%	98.03%
RVM <sub>quad</sub>	58.55%	94.41%
RVM <sub>cutb</sub>	49.34%	88.82%
GP <sub>QRSN</sub>	49.34%	99.01%
GP <sub>QRM</sub>	49.67%	99.34%

	-10pp (C)	0pp (D)	+10pp (E)
RVM <sub>lin</sub>	10.53%	92.43%	98.36%
RVM <sub>quad</sub>	18.09%	81.25%	90.13%
RVM <sub>cutb</sub>	14.80%	63.82%	78.62%
GP <sub>QRSN</sub>	2.96%	78.62%	97.70%
GP <sub>QRM</sub>	2.63%	78.62%	97.70%

**Figure 3.9:** Prediction outcome for a latency of  $h = 115$  ms; a) Cumulative  $\text{RMSE}_{rel}$  histogram evaluated with wLMS, three RVM, and two GP algorithms; values along line A and B are listed in b); c) Cumulative  $\Delta\text{RMSE}_{rel}$  histogram of all algorithms with respect to the wLMS algorithm; values along line C, D, and E are listed in d).



**Figure 3.10:** Prediction outcome for a latency of  $h = 154$  ms; a) Cumulative  $RMSE_{rel}$  histogram evaluated with wLMS, MULIN, SVR, three RVM, and two GP algorithms; values along line A and B are listed in b); c) Cumulative  $\Delta RMSE_{rel}$  histogram of all algorithms with respect to the wLMS algorithm; values along line C, D, and E are listed in d).



b)

	50% (A)	100% (B)
wLMS	32.24%	97.70%
MULIN	3.31%	98.68%
SVR	33.99%	96.37%
RVM <sub>in</sub>	51.64%	98.07%
RVM <sub>quad</sub>	47.04%	92.11%
RVM <sub>cub</sub>	31.58%	74.34%
GP <sub>QRSN</sub>	23.68%	97.70%
GP <sub>QRM</sub>	26.97%	97.70%

d)

	-10 pp (C)	0 pp (D)	+10 pp (E)
MULIN	0.32%	2.961%	55.92%
SVR	13.49%	58.55%	87.50%
RVM <sub>in</sub>	14.80%	93.75%	98.68%
RVM <sub>quad</sub>	23.36%	69.41%	83.55%
RVM <sub>cub</sub>	11.84%	35.53%	45.72%
GP <sub>QRSN</sub>	0.99%	55.59%	92.76%
GP <sub>QRM</sub>	2.96%	57.24%	92.76%

Figure 3.11: Prediction outcome for a latency of  $h = 308$  ms; a) Cumulative  $RMSE_{rel}$  histogram evaluated with wLMS, MULIN, SVR, three RVM, and two GP algorithms; values along line A and B are listed in b); c) Cumulative  $\Delta RMSE_{rel}$  histogram of all algorithms with respect to the wLMS algorithm; values along line C, D, and E are listed in d)

RMSE difference  $\Delta\text{RMSE}_{rel}$  with respect to the **wLMS** algorithm. A value of  $\Delta\text{RMSE}_{rel} < 0$  pp represents a decreased  $\text{RMSE}_{rel}$  compared to the **wLMS** algorithm. The vertical dashed line D at  $\Delta\text{RMSE}_{rel} = 0$  pp represents the percentage of data which could be predicted with an  $\text{RMSE}_{rel}$  lower or equally good compared to the **wLMS** algorithm. In case of  $h = 77$  ms (Fig. 3.8.c), the percentage of data with an  $\Delta\text{RMSE}_{rel} \leq 0$  pp is: 16.12 % for **SVR**, 19.41 % for **MULIN**, 77.96 % for  $\text{RVM}_{cub}$ , 84.54 % for  $\text{GP}_{QRSN}$ , 84.87 % for  $\text{GP}_{QRM}$ , 88.19 % for  $\text{RVM}_{quad}$ , and 91.78 % for  $\text{RVM}_{lin}$ . The left and right vertical dashed line C and E at  $\Delta\text{RMSE}_{rel} = -10$  pp and  $\Delta\text{RMSE}_{rel} = 10$  pp represent extreme cases for which the  $\text{RMSE}_{rel}$  could be either decreased or increased by at least 10 pp. For example in Fig. 3.8.c, the percentage of data with  $\Delta\text{RMSE}_{rel} \leq -10$  % is with 17.43 % for  $\text{RVM}_{quad}$  and 16.78 % for  $\text{RVM}_{cub}$ . Even though a high percentage of the data could be improved by at least 10 pp, the right vertical line indicates that also a high percentage of the data has a  $\text{RMSE}_{rel}$  which is more than 10 pp higher compared to the **wLMS** algorithm. This is the case for 6.25 % of the data for  $\text{RVM}_{quad}$  and for 10.86 % for  $\text{RVM}_{cub}$ . In contrast, the **GP** approaches lead only for 4.93 % of the data for  $\text{GP}_{QRM}$  and for 4.28 % for  $\text{GP}_{QRSN}$  to an improvement  $\text{RMSE}_{rel}$  of at least 10 pp. However, only 0.66 % of the data have an increased  $\text{RMSE}_{rel}$  of at least 10 pp for  $\text{GP}_{QRM}$  and for 0.99 % of the data for  $\text{GP}_{QRSN}$ .

Differences can be observed for the investigated prediction latencies such as a strong decrease of the prediction accuracy for the  $\text{RVM}_{cub}$  algorithm for increasing  $h$ . While the algorithm predicts 51.97 % of the data with an error of  $\text{RMSE}_{rel} \leq 50$  % for  $h = 77$  ms, only 31.58 % of the data can be predicted for  $h = 308$  ms. A similar strong decrease can be observed for the two **GP** algorithms. The percent of data with  $\text{RMSE}_{rel} \leq 50$  % decreases by 24.35 pp for  $\text{GP}_{QRSN}$  and by 22.37 pp for  $\text{GP}_{QRM}$ . It can be observed that the only increase in the prediction accuracy is visible for the **SVR** algorithm, which was already reported in [86].

The main results are summarized in Table 3.8 for all prediction horizons. The Table lists the mean  $\text{RMSE}$  and  $\text{RMSE}_{rel}$  and their standard deviation for all investigated algorithms. Additionally the  $\text{RMSE}$  results are shown if no prediction algorithm is used. In case of no prediction the mean  $\text{RMSE}$  value increases from 0.257 mm for  $h = 77$  ms to 0.905 mm for  $h = 308$  ms on average. This error can be reduced by half, if the  $\text{RVM}_{lin}$  algorithm is used. The mean  $\text{RMSE}$  of  $\text{RVM}_{lin}$  is 0.257 mm to 0.126 mm for  $h = 77$  ms and 0.469 mm for  $h = 308$  ms.

### 3.3.3 Discussion

The results presented in [86] indicated that the lowest  $\text{RMSE}$  value could be achieved on average by the **wLMS** algorithm for  $h = 77$  ms with 0.135 mm and for  $h = 154$  ms with

**Table 3.8:** Mean and standard deviation of the **RMSE** and the **RMSE<sub>rel</sub>** for the **RVM** and **GP** algorithms. The results of the **wLMS**, **MULIN**, and **SVR** algorithm are according to [86]. Four prediction latencies have been investigated,  $h = \{77, 115, 154, 308\}$  ms.

measure	no pred	wLMS	MULIN	SVR	RVM <sub>lin</sub>	RVM <sub>quad</sub>	RVM <sub>cube</sub>	GP <sub>GRNN</sub>	GP <sub>GRM</sub>
$h = 77$ ms									
RMSE									
mean [mm]	0.257	0.135	0.143	0.193	<b>0.126</b>	0.133	0.179	0.130	0.130
std [mm]	0.153	0.089	0.096	0.140	0.085	0.101	0.271	0.078	<b>0.077</b>
RMSE <sub>rel</sub>									
mean [%]	100.00	56.51	59.62	76.17	<b>52.96</b>	55.57	77.59	54.76	54.99
std [%]	0.00	18.10	17.95	28.72	18.87	28.97	140.28	<b>16.79</b>	17.46
$h = 115$ ms									
RMSE									
mean [mm]	0.372	0.192	-	-	<b>0.178</b>	0.200	0.335	0.188	0.187
std [mm]	0.222	0.126	-	-	0.121	0.169	0.765	0.116	<b>0.113</b>
RMSE <sub>rel</sub>									
mean [%]	100.00	55.51	-	-	<b>51.55</b>	58.35	112.07	54.51	54.66
std [%]	0.00	17.58	-	-	18.57	43.41	408.61	<b>16.03</b>	16.83
$h = 154$ ms									
RMSE									
mean [mm]	0.484	0.253	0.277	0.288	<b>0.234</b>	0.260	0.505	0.251	0.249
std [mm]	0.291	0.165	0.188	0.191	0.156	0.209	1.446	0.156	<b>0.151</b>
RMSE <sub>rel</sub>									
mean [%]	100.00	55.95	60.73	62.79	<b>51.82</b>	59.14	149.49	55.30	55.32
std [%]	0.00	17.32	16.97	22.41	18.36	53.43	765.35	<b>15.46</b>	16.35
$h = 308$ ms									
RMSE									
mean [mm]	0.905	0.517	0.612	0.508	<b>0.469</b>	0.536	1.258	0.532	0.521
std [mm]	0.541	0.316	0.382	0.317	<b>0.292</b>	0.409	4.513	0.329	0.314
RMSE <sub>rel</sub>									
mean [%]	100.00	60.09	69.46	59.32	<b>54.47</b>	65.18	238.92	61.11	60.45
std [%]	0.00	16.22	14.39	18.33	16.04	66.44	1760.14	<b>14.02</b>	15.00

0.253 mm. For  $h = 308$ , ms, the **SVR** algorithm had a slightly improved performance compared to the **wLMS** algorithm with 0.508 mm. The superior **SVR** result for  $h = 308$  ms has to be highlighted as the mean **RMSE** of the **SVR** algorithm is greater than for the **wLMS** and **MULIN** algorithm for  $h = \{77, 154\}$  ms. Nonetheless, the results indicate that the error increases slower for increasing prediction latency [86], concluding that the **SVR** method is superior for high prediction latencies. Incorporating the prediction outcome of the **RVM** and **GP** algorithms reveals that the highest prediction accuracy can be achieved by the  $\text{RVM}_{lin}$  algorithm for all prediction horizons (Table 3.8). Performing a  $t$ -test with  $p = 0.001$  reveals a significant improvement of the  $\text{RMSE}_{rel}$  compared to the **wLMS** algorithm.

Focusing only on the results of the **RVM** methods reveals that on average the performance of  $\text{RVM}_{lin}$  is superior compared to  $\text{RVM}_{quad}$  and  $\text{RVM}_{cub}$ . This becomes especially clear in case of  $\text{RVM}_{cub}$  for  $h = \{115, 154, 308\}$  ms where the mean  $\text{RMSE}_{rel}$  is over 100%. Meaning that on average the performance is worse compared to no prediction.

Investigating the  $\Delta\text{RMSE}_{rel}$  histograms (Fig. 3.8.b - 3.11.b ) reveals that  $\text{RVM}_{quad}$  and  $\text{RVM}_{cub}$  are susceptible to outliers. The **RMSE** of both algorithms is decreased for a high percentage of data by at least 10 pp compared to **wLMS** (marked by the vertical line C in the bottom diagrams). For  $\text{RVM}_{quad}$ , this is true for 17.43 %, 18.09 %, 20.39 %, and 23.36 % of the data for a latency of  $h = \{77, 115, 154, 308\}$  ms, respectively. On the other hand, the error is increased extremely for some of the motion traces. This might be best observed in case of  $\text{RVM}_{cub}$  which leads to an increased  $\text{RMSE}_{rel}$  of more than 10 pp for 54.28 % of the data for  $h = 308$  ms. The results point out that including the second and third power of each feature element (see Eq. 3.15) might lead to a high improvement for a subset of the data. However, it makes the **RVM** algorithm susceptible to strong outliers. The highest  $\text{RMSE}_{rel}$  for one motion trace is 961.64 % for  $\text{RVM}_{quad}$  and  $2.77 \cdot 10^4$  % for  $\text{RVM}_{cub}$  for  $h = 308$  ms. In contrast, the  $\text{RVM}_{lin}$  seems to be more robust over the entire dataset with the highest  $\text{RMSE}_{rel}$  for one motion trace of 1.27 % for  $h = 308$  ms (the highest  $\text{RMSE}_{rel}$  of **wLMS** is 1.39 %).

Table 3.8 indicates that both **GP** methods have on average an increased **RMSE** and  $\text{RMSE}_{rel}$  relative to  $\text{RVM}_{lin}$ . However, only minor differences are visible for short latencies ( $h = \{77, 115\}$  ms). These differences increase for increasing latency. Compared to the **wLMS**, the **MULIN**, and the **SVR** algorithm, the mean **RMSE** and  $\text{RMSE}_{rel}$  are lower for  $h = \{77, 115, 154\}$  ms. For  $h = 308$  ms, the prediction accuracy is only slightly increased compared to the **wLMS** and **SVR** algorithm. Comparing both **GP** methods with each other suggest an almost equal performance. A slightly superior performance of  $\text{GP}_{QRM}$  is visible for  $h = 308$  ms. On average, the **GP** approach seems to be very robust

with respect to the motion traces as it is indicated by the relatively lower standard deviation for all prediction latencies.

Finally, it should be emphasized that the number of training pairs  $m$ , the features dimension  $d$  and the number of optimization steps  $n_{iter}$  were equal for each motion trace. Further improvements might be possible by optimizing these parameters on the initial training set. In contrast, the results of the **MULIN** and **SVR** algorithm were optimized by grid search or a bounded optimisation routine [86].

## 3.4 Exploiting Probabilistic Uncertainties<sup>6</sup>

The aim of adaptive motion compensation is to deliver ablative radiation to a moving tumour as precisely as possible. As the patients are breathing freely without any coaching, changes in the breathing pattern are to be expected. In the optimal case, the MC algorithm is able to adapt to these changes. As the variety of published algorithms and diverse results indicate, this cannot be guaranteed. One obvious solution is to monitor the current prediction error  $e_i$  throughout the treatment. In case of an increased error, the treatment could be interrupted and unnecessary harm to the patient could be prevented. Such monitoring systems are especially prone to errors caused by slow changes of, e.g., the breathing frequency or the amplitude. However, their effectiveness is limited in case of unexpected fast changes, such as coughing or sneezing, as they suffer from time latencies. Let us assume to be at time index  $i$ . A prediction algorithm computes the point  $y_{i+\xi}^*$ . We are interested in the prediction error  $e_{i+\xi}$  at the predicted time index  $i + \xi$ . However, this error can only be computed retrospectively delayed by the latency  $h = \xi/f_s$ , which could be up to several hundreds of milliseconds. In an online setting, only the current prediction error  $e_i$  can be observed, which might be less correlated to the error  $e_{i+\xi}$ .

Probabilistic MC algorithms offer an alternative approach to estimate the prediction error  $e_{i+\xi}$ . The result of these algorithms is a probability distribution. So far, the first order statistics, the mean of the distribution, was used as predicted value  $y_{i+\xi}^*$ . However, also the second order statistics, the variance of the predicted distribution  $\sigma_{i+\xi}^{2*}$ , contain valuable information. The variance can be interpreted as the “certainty” of the algorithm in the prediction. A high variance indicates the possibility of a high prediction error. In contrast to the prediction error  $e_{i+\xi}$ , the variance  $\sigma_{i+\xi}^{2*}$  is available at the current time index  $i$ . This advantage is distinct for probabilistic approaches. Ruan [220] discussed this idea in case of the **KDE** algorithm [219]. The author proposed the possibility to control large prediction errors in real-time by monitoring the variance.

---

<sup>6</sup>Parts of this section have been published in [67, 72, 73]



It has to be mentioned that there are alternative possibilities to estimate the prediction error  $e_{i+\xi}$  with non-probabilistic approaches. For example the training error could be evaluated. However, as pointed out in [220], it can be quite difficult to learn this relationship, resulting in the fact that no studies covering this problem have been published so far. In contrast to probabilistic algorithms, where the predicted variance is a spin-off of the actual prediction.

Within this section the hypothesis of Ruan [220] shall be further investigated. First, the predicted variance of the RVM algorithm is evaluated to control the prediction error  $e_{i+\xi}$  in real-time (sec. 3.4.1). Afterwards, this idea is further extended to construct hybrid prediction algorithms based on the variance (sec. 3.4.2).

### 3.4.1 Experimental Evaluation of the Variance

To investigate the hypothesis that the prediction error can be controlled by the predicted variance an experimental evaluation is performed. In a realistic scenario the variance could be monitored and constantly compared to a threshold  $\sigma_{th}^2$ . If the predicted variance  $\sigma_{i+\xi}^{2*} > \sigma_{th}^2$ , the treatment could be interrupted for the time index  $i + \xi$ . Depending on the selected  $\sigma_{th}^2$ , the effective treatment time will be reduced. This can be evaluated by the duty cycle DC (Eq. 2.11) which expresses the remaining treatment relative to the complete time

$$DC = \frac{N_{pos}}{N} \quad (3.46)$$

where  $N_{pos}$  is the number of points with  $\sigma_{i+\xi}^{2*} \leq \sigma_{th}^2$ . To investigate if limiting the variance limits the prediction error, an error threshold  $e_{th}$  has to be defined. Further, the precision PR is defined as

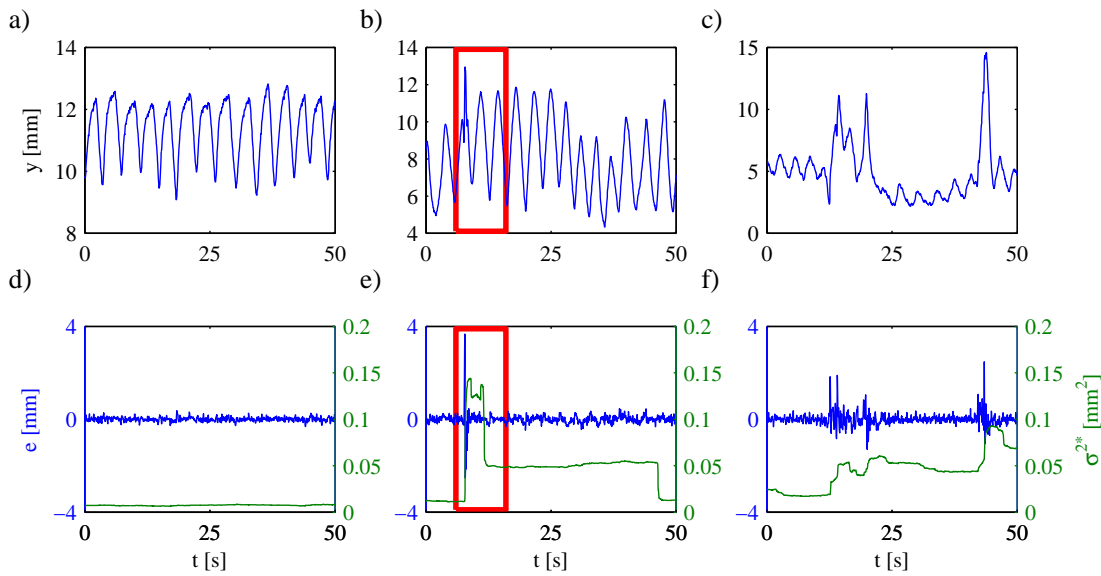
$$PR = \frac{N'_{pos}}{N_{pos}} \quad (3.47)$$

where  $N'_{pos}$  is the number of points of  $N_{pos}$  with  $|e_i| \leq e_{th}$  with  $i = 1, \dots, N_{pos}$ .

**Experimental Setup** In the experiment, a linear RVM algorithm was evaluated on all 304 motion traces. The latency index was set to  $\xi = 3$ , which coincides with the CyberKnife latency of  $h = 115$  ms. The parameters of the RVM algorithm were fixed to  $m = 1000$  and  $d = 100$ , which is equivalent to sec. 3.3.1.

The duty cycle and the precision were evaluated for  $e_{th} = \{0.1, 0.2, 0.5, 1\}$  mm and  $\sigma_{th}^2 = \{0.001, 0.0012, \dots, 10\}$  mm<sup>2</sup>.

For comparison, an equivalent evaluation was performed using the current prediction error  $e_i$ . If  $e_i < d_{th}$ , the treatment was interrupted at the time step  $t_{i+\xi}$ . Here,  $d_{th}$  is an



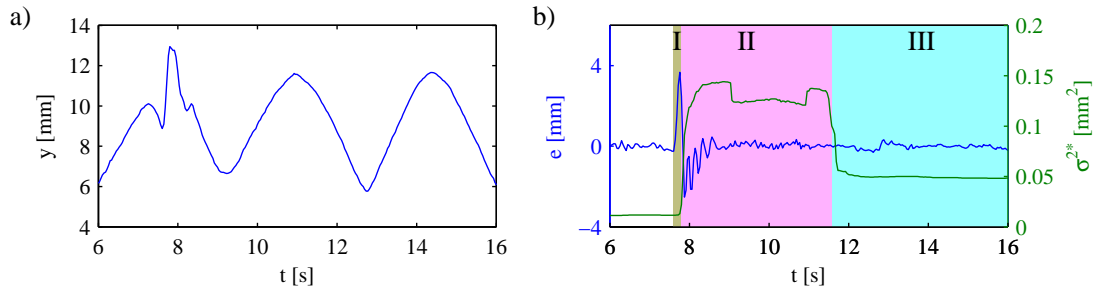
**Figure 3.12:** Measured position  $y$  of three motion fragments which are examples for regular (a), regular breathing with one artefact (b) and irregular breathing with multiple artefacts (c); d-f) Prediction error  $e$  and predicted variance  $\sigma^{2*}$  of a linear RVM algorithm for each motion fragment, respectively. Areas marked by an red rectangle are shown enlarged in Fig. 3.13.

error threshold which was set to  $d_{th} = \{0.001, 0.0012, \dots, 10\}$  mm.

Note that within this experiment  $N$  refers to the number of investigated points over all motion traces.

**Results and Discussion** Figure 3.12 shows three example motion traces, the error and predicted variance. The motion fragments have a duration of 50 s and can be classified as regular (a), regular breathing with one motion artefacts at  $t \approx 7.8$  s (b), and irregular breathing with multiple artefacts (c). The bottom row (Fig. 3.12.d-f) shows the prediction error (blue) and the variance (green), respectively. For the regular breathing fragment, the prediction error and the variance are small. With the appearance of one or multiple breathing artefacts, the prediction error as well as the variance increases. Note, the prediction error can only be evaluated retrospectively delayed by the latency. In contrast the variance is available at the current time index.

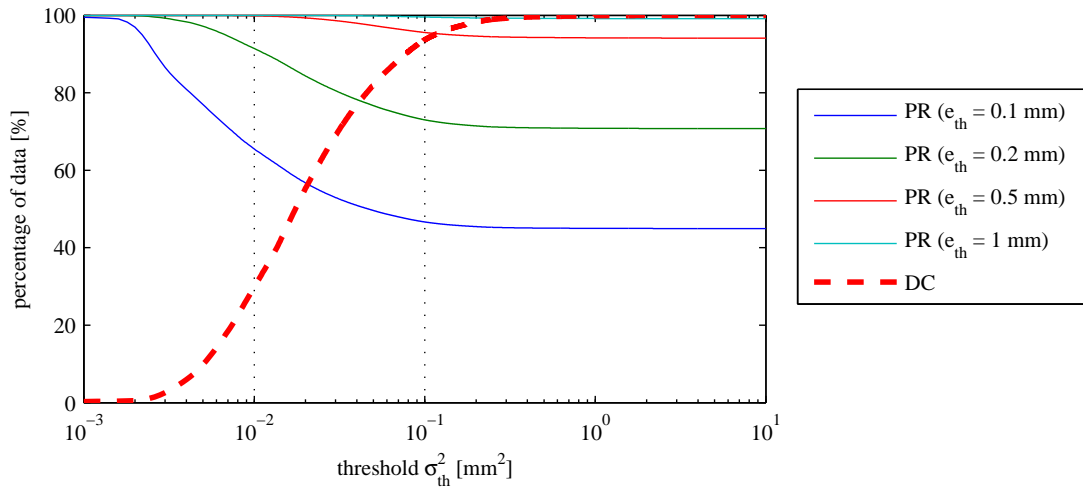
The motion fragment in 3.12.b reveals a characteristic variance pattern, which has been observed frequently in the data. The areas marked by a red rectangle in Fig. 3.12.b and 3.12.e are shown enlarged in Fig. 3.13. To investigate this pattern, it is divided into



**Figure 3.13:** (a) Measured position  $y$ , and (b) prediction error  $e$  and predicted variance  $\sigma^{2*}$  of the linear RVM algorithm. The error and the variance are divided into phase I to III. (enlarged illustration of Fig. 3.12.b and 3.12.e.)

phase I to III.

- Phase I:* During this phase, the irregular motion pattern occurs. The irregular motion pattern has not yet been observed. The algorithm performs a “normal” prediction based on the training dataset. This results in a high prediction error and a low variance. The duration of this phase is very short and depends mainly on the latency  $h$ . It can be assumed that  $t_I \approx h$ . The results indicate that the prediction error in this phase cannot be controlled by the variance.
- Phase II:* The second phase is characterized by a high variance and partly high prediction errors. Within this phase, the high prediction error of phase I decreases depending on the duration of the irregular motion pattern. The variance remains high throughout this phase, as the irregular motion pattern is still part of the test features  $\mathbf{x}_{i+\xi}^*$ . The duration of this phase depends on the dimension  $d$  of the features vectors. Assuming an irregular motion fragment which occurs at just one time instance, the duration can be specified by  $t_{II} \approx d/f_s = 3.83$  s.
- Phase III:* The last phase can be distinguished by a small prediction error and an increased variance. The irregular motion pattern is no longer present within the  $\mathbf{x}_{i+\xi}^*$  but is part of the training set. This results in an increased variance compared to phase I. Compared to phase II, the variance is decreased as the irregular motion fragment is not present in all training features. Considering again a motion artefact which occurs at one time instance, the minimal duration of this phase can be estimated by  $t_{III} \approx (m - d)/f_s = 34.61$  s. This is in agreement with the results of Fig.3.12.e.



**Figure 3.14:** Precision for the error threshold  $e_{th} = \{0.1, 0.2, 0.5, 1\}$  mm and duty cycle depending on the variance threshold  $\sigma_{th}^2$ .

This variance pattern could be observed in case of the appearance of one irregular motion fragment. However, if multiple motion artefacts occur, multiple such patterns occur and influence each other (Fig. 3.12.f). Several conclusions can be drawn from this analysis:

1. Spontaneous irregular motion patterns and their resulting prediction error (phase I) cannot be controlled by the variance. The occurrence of an irregular motion pattern has to have indications in the test feature  $\mathbf{x}_i^*$ .
2. The variance can be used to control the prediction errors of phase II and III.
3. The duration of phase II and III can be influenced by the feature dimension  $d$  and the size of the training set  $m$ . The increased variance of phase III can be interpreted as the increased probability of the appearance of an irregular motion pattern in case of an recently observed irregular pattern. A more conservative treatment is possible if, e.g.,  $m$  is increased. However, it has to be considered that varying  $d$  and  $m$  will effect the prediction accuracy (see. 3.2.1.2).

The precision (solid lines) and the duty cycle (red dashed line) over all motion traces depending on the variance threshold  $\sigma_{th}^2$  are shown in Fig. 3.14. The duty cycle decreases with decreasing  $\sigma_{th}^2$ , while the precision increases. Assuming a constant variance threshold of  $\sigma_{th}^2 = 0.1$  mm (right vertical dotted line), the precision is 46.64% for  $e_{th} = 0.1$  mm and 95.57% for  $e_{th} = 0.5$  mm with a duty cycle of DC= 93.66%. Decreasing the variance threshold to  $\sigma_{th}^2 = 0.001$  mm<sup>2</sup> (left vertical dotted line), results in

a decrease of the duty cycle of 63.74 pp. In contrast the precision increases by 18.58 pp for  $e_{th} = 0.1$  mm and 6.21 pp for  $e_{th} = 0.5$  mm. These results are in agreement with the proposed variance pattern. By decreasing  $\sigma_{th}^2$ , the prediction error of phase II and phase III are controlled. Nonetheless, the errors of phase I cannot be bounded. This becomes obvious as the highest precision of  $e_{th} = 0.1$  mm is only PR= 99.51 %.

In general, it might be observed that the precision with  $e_{th} = 1$  mm is close to 100 % independently of  $\sigma_{th}^2$ . This confirms the high accuracy of the linear RVM algorithm.

Figure 3.15 shows the precision (solid lines) and the duty cycle (red dashed line) by using the current prediction error  $e_i$  as criterion instead of the predicted variance. By varying the error threshold  $d_{th}$  a similar decrease and increase of the duty cycle can be observed compared to Fig. 3.14. However by decreasing  $d_{th}$ , the precision increases until it reaches an maximum at  $d_{th} \approx 10^{-2} - 10^{-3}$  mm depending on  $e_{th}$ . This maximum is below 90 % for  $e_{th} = \{0.1, 0.2\}$  mm. A further decrease of  $d_{th}$  only decreases the duty cycle. The difference is caused due to the fact that the current prediction error only depends on the test label  $y_i^*$  and not on the test  $\mathbf{x}_i^*$  and training features  $\mathbf{X}$ .

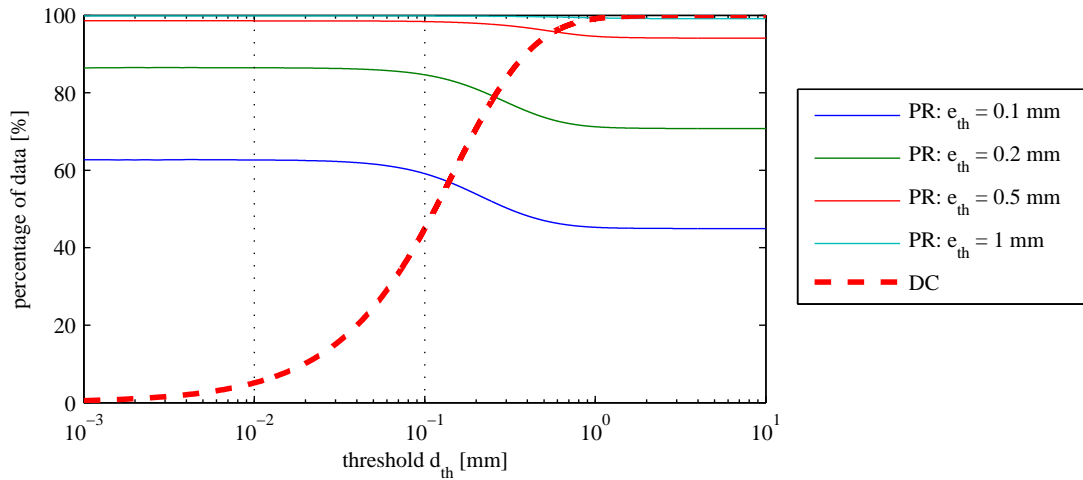
The results indicate that a superior control of the prediction error  $e_{i+\xi}$  is possible by monitoring the variance  $\sigma_{i+\xi}^{2*}$  instead of the current prediction error  $e_i$ . A similar experiment can be performed for the GP algorithm. However, only the RVM algorithm was evaluated as similar results can be expected.

For practical implementations, the variance threshold can be seen as another algorithm parameter, which can be determined on an initial training set. Further, an adaptation of  $\sigma_{th}^2$  is easily possible in real-time as this parameter does not influence the prediction algorithm itself.

### 3.4.2 Hybrid Algorithms

Based on the results of the previous section, the variance could be used to construct hybrid algorithms. Instead of interrupting the treatment if  $\sigma_{i+\xi}^{2*} > \sigma_{th}^2$ , the results of an alternative prediction algorithm could be used.

In principle, there exist various possibilities to construct hybrid algorithms. In [203], a hybrid algorithm was proposed based on two Kalman filters assuming either a constant velocity or a constant acceleration. As each assumption is only valid for certain states of the breathing phase, the authors designed a transition matrix containing the probability going from one state to the next state. The current CyberKnife<sup>®</sup> prediction algorithm is also a hybrid algorithm [227]. It consists of three individual algorithms. In contrast to [203], the results of all algorithms are computed in parallel. The final result depends on the two prediction algorithms, whose prediction results have the smallest absolute dif-



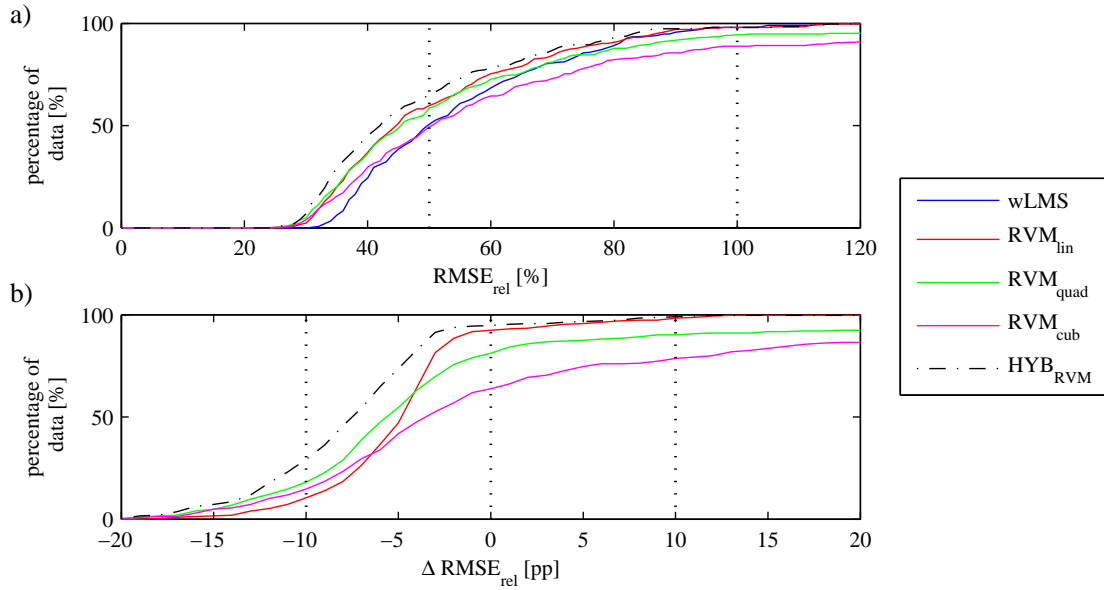
**Figure 3.15:** Precision for the error threshold  $e_{th} = \{0.1, 0.2, 0.5, 1\}$  mm and duty cycle depending on the error threshold  $d_{th}$ .

ference to each other.

Here, two approaches are investigated to construct hybrid algorithms based on the variance. First, we consider a hybrid approach consisting of multiple probabilistic algorithms. As each algorithm computes an individual variance, the variance across the algorithms can be compared. The prediction result  $y_{i+\xi}^*$  is the result of the algorithm with the lowest variance. This approach is evaluated exemplarily with three **RVM** algorithms using the covariance functions defined in Eq. 3.15. The hybrid algorithm is referred to as  $HYB_{RVM}$ . The second approach combines a non-probabilistic and a probabilistic algorithm. The variance is compared to a threshold  $\sigma_{th}^2$ . The non-probabilistic algorithm is only selected if  $\sigma_{i+\xi}^{2*} > \sigma_{th}^2$ . The **wLMS** algorithm was chosen as non-probabilistic algorithm based on the superior performance in [86]. The linear, quadratic, and cubic **RVM** algorithm were used as probabilistic algorithms. We refer to the hybrid algorithms as  $HYB_{wR-lin}$ ,  $HYB_{wR-quad}$ , and  $HYB_{wR-cub}$ , respectively.

**Experimental Setup** The four hybrid algorithms were evaluated on all 304 motion traces. Equivalent to sec. 3.4.1, the latency index was set to  $\xi = 3$  and the parameters of the **RVM** algorithms to  $m = 1000$  and  $d = 100$ . The parameters of the **wLMS** algorithm were initialised according to [80] with  $d = 193$ , a learning rate of  $\mu = 0.0204$  and  $J = 3$  number of wavelet decomposition levels. The variance threshold of the second hybrid approach was specified by  $\sigma_{th}^2 \in \{0.001, 0.0025, 0.005, 0.0075, 0.01, 0.025, \dots, 1\}$  mm<sup>2</sup>.

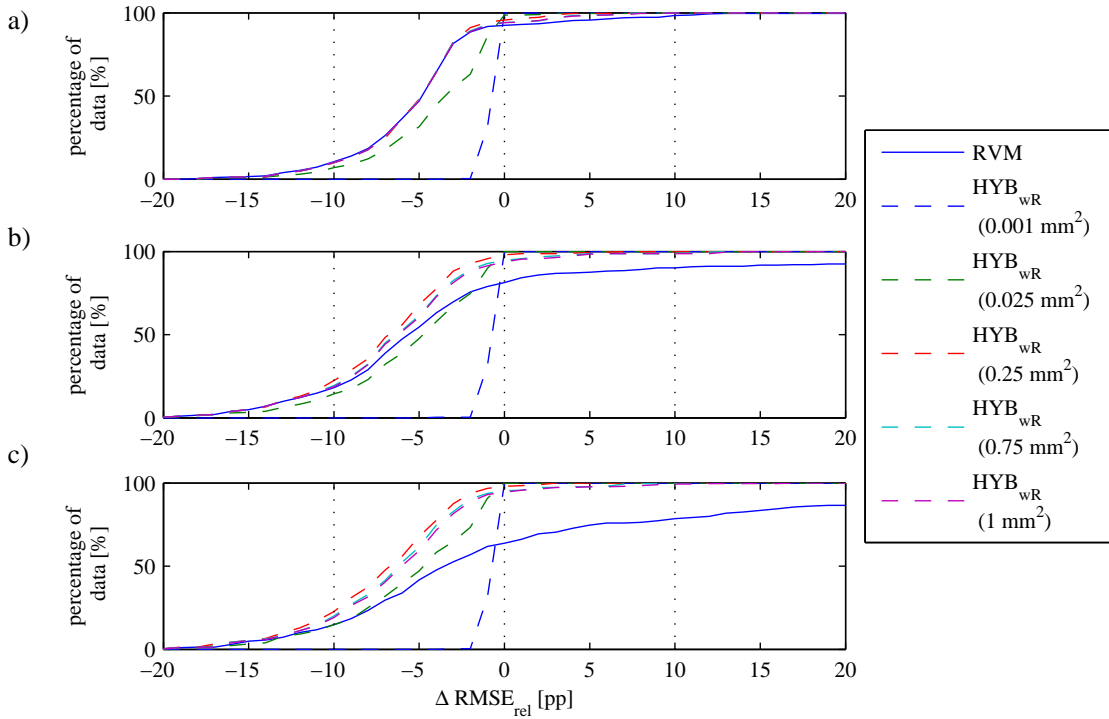
The results were evaluated with respect to the **RMSE** and **RMSE<sub>rel</sub>**. Additionally, the



**Figure 3.16:** a) Cumulative histogram of the  $RMSE_{rel}$  for  $wLMS$ ,  $RVM_{lin}$ ,  $RVM_{quad}$ ,  $RVM_{cub}$ , and  $HYB_{RVM}$ ; b) Cumulative histogram of the  $RMSE_{rel}$  difference between  $wLMS$  and  $RVM_{lin}$ ,  $RVM_{quad}$ ,  $RVM_{cub}$ , and  $HYB_{RVM}$ .

$\Delta RMSE_{rel}$  was computed which was defined in Eq. 3.45 and represents the  $RMSE_{rel}$  difference with respect to the  $wLMS$  algorithm. Furthermore the maximum absolute error mAE was investigated (Eq. 2.7).

**Results and Discussion** Figure 3.16.a shows a cumulative histogram of the  $RMSE_{rel}$  for the first hybrid algorithm,  $HYB_{RVM}$ . The results of the individual  $RVM$  and the  $wLMS$  algorithm are shown for comparison. It might be observed that the  $HYB_{RVM}$  algorithm has a superior performance compared to the residual algorithms. The percentage of data which has a  $RMSE_{rel} \leq 50\%$  (left vertical dotted line) is 49.34% for  $RVM_{cub}$ , 50.66% for  $wLMS$ , 58.55% for  $RVM_{quad}$ , 59.87% for  $RVM_{lin}$  and 64.47% for  $HYB_{RVM}$ . Further, Fig. 3.16.b shows the relative RMSE differences  $\Delta RMSE_{rel}$  with respect to the  $wLMS$  algorithm. Again, also the results of the individual  $RVM$  algorithms are plotted. The  $HYB_{RVM}$  approach outperforms the individual algorithms on average and also for extreme values of  $\Delta RMSE_{rel}$ . Using the  $HYB_{RVM}$  instead of the  $wLMS$  approach, leads to a decrease of the  $RMSE_{rel}$  of 10 pp or more for 29.28% of the data (marked by the left vertical dotted line). For comparison, this improvement could be achieved only for 18.09% of the data by using the  $RVM_{quad}$ , 14.8% by the  $RVM_{cub}$ , and 10.53% by the  $RVM_{lin}$ . On



**Figure 3.17:** Cumulative histogram of the  $\text{RMSE}_{rel}$  difference between  $w\text{LMS}$  and  $\text{HYB}_{wR-lin}$  and  $\text{RVM}$  for a (a) linear, (b) quadratic, (c) and cubic basis function.  $\text{HYB}_{wR-lin}$  results are shown for  $\sigma_{th}^2 = \{0.001, 0.025, 0.25, 0.75, 1\} \text{ mm}^2$ .

the other hand, the  $\text{RVM}_{quad}$  and  $\text{RVM}_{cub}$  algorithm leads to a relatively high percentage of data with an error which is increased by  $\text{RMSE}_{rel} \geq 10 \text{ pp}$  compared to the  $w\text{LMS}$  algorithm (right vertical line). The  $\text{HYB}_{\text{RVM}}$  algorithm instead increases the  $\text{RMSE}_{rel}$  of only 0.99% of the data by 10 pp or more compared to the  $w\text{LMS}$  algorithm. This indicates that the variance can be used as feature to compensate the partly high prediction errors of  $\text{RVM}_{quad}$  and  $\text{RVM}_{cub}$ .

The cumulative histogram of the  $\Delta \text{RMSE}_{rel}$  for the  $\text{HYB}_{wR-lin}$ ,  $\text{HYB}_{wR-quad}$  and  $\text{HYB}_{wR-cub}$  for different  $\sigma_{th}^2$  is shown in Fig. 3.17.a-c, respectively. The solid blue line is the result of the individual  $\text{RVM}$  algorithm. As  $\Delta \text{RMSE}_{rel}$  is computed with respect to the  $w\text{LMS}$  algorithm, a vertical line at  $\Delta \text{RMSE}_{rel} = 0 \text{ pp}$  represents the results of the  $w\text{LMS}$  algorithm. For illustration purposes, not all evaluated thresholds are shown. The threshold  $\sigma_{th}^2$  controls how frequently the results of the  $w\text{LMS}$  or the  $\text{RVM}$  algorithm were selected. In case of a small threshold such as  $\sigma_{th}^2 = 0.001 \text{ mm}^2$ , the results are comparable to the  $w\text{LMS}$  algorithm and only slight improvements can be observed. Increasing the threshold, the  $\text{RVM}$  results will be more frequently selected. This leads to an decreased

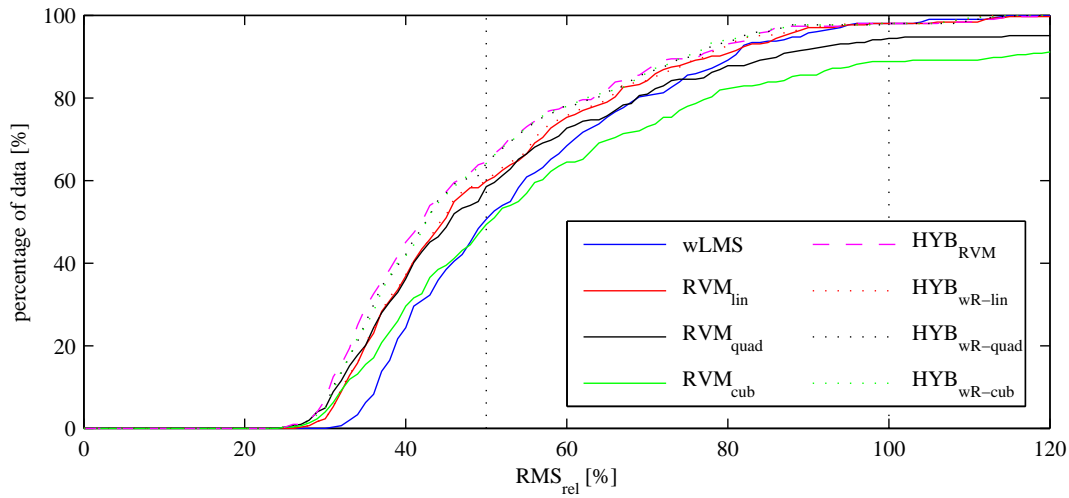


**Table 3.9:** Mean **RMSE**, **RMSE<sub>rel</sub>** and **mAE** over all 304 motion traces of **wLMS**, **RVM**, **HYB<sub>RVM</sub>**, and best results of **HYB<sub>wR</sub>** ( $\sigma_{th}^2 = 25 \text{ mm}^2$ ). The values in parenthesis are the standard deviation. (best results highlighted bold)

algorithm	RMSE [mm]	RMSE <sub>rel</sub> [%]	mAE [mm]
wLMS	0.1922 (0.126)	55.51 (17.58)	<b>3.885</b> <b>(3.252)</b>
RVM <sub>lin</sub>	0.1781 (0.121)	51.55 (18.57)	4.19 (4.032)
RVM <sub>quad</sub>	0.1997 (0.167)	58.35 (43.41)	9.19 (18.411)
RVM <sub>cub</sub>	0.3345 (0.765)	112.07 (408.61)	39.884 (193.505)
HYB <sub>wR-lin</sub>	0.1770 (0.123)	51.09 (17.89)	3.928 (3.295)
HYB <sub>wR-quad</sub>	0.1694 (0.118)	49.35 (18.03)	4.157 (3.19)
HYB <sub>wR-cub</sub>	0.1691 (0.118)	49.26 (17.77)	4.101 (3.135)
HYB <sub>RVM</sub>	<b>0.1679</b> <b>(0.115)</b>	<b>48.92</b> <b>(18.40)</b>	4.177 (4.021)

$\Delta\text{RMSE}_{rel}$  on average. The highest mean  $\Delta\text{RMSE}_{rel}$  can be achieved by  $\sigma_{th}^2 = 0.025 \text{ mm}^2$  for all three hybrid algorithms. Interestingly, the results of best **HYB<sub>wR</sub>** is not an averaged **RMSE** between the **wLMS** and the **RVM** algorithm. The results are rather superior compared to the individual algorithms. This is especially visible in case of **HYB<sub>wR-quad</sub>** and **HYB<sub>wR-cub</sub>**. Less percent of the data have a  $\Delta\text{RMSE}_{rel} > 0 \text{ pp}$ . This indicates that especially the dataset with a high prediction error of **RVM<sub>quad</sub>** and **RVM<sub>lin</sub>** can be improved, which is similar to the results of the first hybrid approach.

The main average results of both hybrid approaches and their standard deviation are listed for comparison in Table 3.9. The best averaged **RMSE**, **RMSE<sub>rel</sub>** and maximum absolute error **mAE** (Eq. 2.7) are highlighted in bold. Further, Fig. 3.18 shows a cumulative histogram of the **RMSE<sub>rel</sub>** for all approaches. On average, all hybrid approaches lead to a decreased **RMSE** and **RMSE<sub>rel</sub>** compared to their individual algorithms. The best result is achieved by the **HYB<sub>RVM</sub>** algorithm, closely followed by **HYB<sub>wR-cub</sub>** and **HYB<sub>wR-quad</sub>**.



**Figure 3.18:** Cumulative histogram of the  $RMSE_{rel}$  for  $wLMS$ ,  $RVM$ ,  $HYB_{RVM}$  and best results of  $HYB_{wR}$  with linear, quadratic, and cubic basis function.

Further advantages of the  $HYB_{RVM}$  algorithm are that it does not need an additional parameter ( $\sigma_{th}^2$ ) and that the algorithm can be easily extended to incorporate multiple probabilistic algorithms such as other  $RVM$  algorithms with different basis functions or the  $GP$  algorithms. This makes the  $HYB_{RVM}$  algorithm very simple to transfer into clinical practise. However, it is restricted to probabilistic algorithm. The second approach offers a simple solution if a non-probabilistic algorithm should be integrated. Similar to the variance threshold  $\sigma_{th}^2$  of sec. 3.4.1, the parameter can be estimated either on an initial training set or adapted throughout the treatment. Again, note that the choice of evaluating a  $wLMS$ - $RVM$  hybrid algorithm was motivated by the superior performance of the  $wLMS$  algorithm in [86]. Further improvements might be achieved by using alternative algorithm such as  $SVR$  or  $ANNs$ .

The hybrid algorithms will result in increased computational requirements. However, each individual algorithm can be computed in parallel. Consequently, each algorithm can be used for the discussed hybrid approaches as long as the algorithm itself fulfils real-time requirements.

### 3.5 Conclusion

The chapter focused on two parts. First, two novel probabilistic algorithms, the **RVM** and the **GP** algorithm have been introduced for respiratory motion prediction. The influence of essential parameters such as the feature dimension and number of training pairs has been evaluated on three exemplary motion traces. Further, both algorithms were evaluated with different basis or covariance functions on the dataset of [86] for four prediction latencies  $h$  between 77 ms and 308 ms. The results indicate a superior performance of the **RVM** algorithm with a linear basis function compared to previously evaluated algorithms for all investigated latencies. In case of the **GP** method, a slightly decreased prediction accuracy compared to  $\text{RVM}_{lin}$  could be found for short prediction latencies ( $h \leq 154$  ms) which decreases further for longer prediction latencies ( $h = 308$  ms). Nonetheless, the **GP** algorithm was able to outperform on average all previously published algorithms for  $h \leq 154$  ms.

Second, the predicted variance  $\sigma_{i+\xi}^{2*}$  was investigated as a criterion to control the prediction error  $e_{i+\xi}$ . This is a distinct advantage of probabilistic algorithms. Only the **RVM** algorithms were evaluated in experiments based on the promising results of  $\text{RVM}_{lin}$ . However, similar results can be expected for the **GP** algorithm. The investigation of the variance revealed a characteristic variance pattern, which gave further insight into which prediction errors might be compensated by the variance. Furthermore, two hybrid algorithms were proposed which are based on the variance. The evaluation on all 304 motion traces of [86] did indicate that a **RVM** hybrid algorithm can further increase the prediction accuracy. The hybrid algorithm consists of three individual **RVM** algorithms with a linear, quadratic and cubic basis function. The results indicate that the variance can be used effectively to select the optimal **RVM** algorithm at each time instance .

Referring back to the questions proposed in sec. 1.3, the results of this chapter can be summarized as followed:

- **Q.2.1:** What criteria could be used to evaluate current prediction accuracy?

This chapter investigates the possibility to compensate the current prediction error by monitoring the predicted variance of probabilistic prediction algorithms. The variance is a spin-off result, as the investigated probabilistic prediction algorithms predict at each time step a Gaussian distribution, whose mean is assumed to be the prediction target. Alternative criteria could be the current prediction error  $e_i$  or the training error.

- **Q.2.2:** What limitations do these criteria have?

The investigation revealed a characteristic variance pattern which is shown in Fig. 3.13.b. As discussed in sec. 3.4.1, the results indicate that prediction errors of phase I, cannot be controlled by the variance. The width of phase I is approximately the prediction horizon  $h$ . Consequently, an increase of the variance is delayed by  $h$ . The hypothesis to use the variance to predict prospective errors could not be confirmed. However in a direct comparison with alternative measures such as the current prediction error  $e_i$ , the results show that prediction error  $e_{i+\xi}$  can be better controlled by the predicted variance  $\sigma_{i+\xi}^{2*}$  (compared Fig. 3.14 and 3.15). Three main differences between monitoring the current prediction error  $e_i$  and the predicted variance  $\sigma_{i+\xi}^2$  can be observed. First,  $e_i$  depends only on the predicted test labels  $y_i^*$  and the current observation  $y_i$ . On the other hand, the variance depends on the current test features  $\mathbf{x}_i^*$  and the training features  $\mathbf{x}$ . In principle, it might be the case that the algorithm computed a  $y_i^*$ , which led to a small  $e_i$ . However, using  $y_i$  as input to construct the next test feature vector  $\mathbf{x}_i^*$ , might lead to an increased variance as no comparable training features  $\mathbf{x}$  might be present within the training set  $\mathbf{T}$ . In this case, monitoring the variance would be an advantage.

Second, assuming the occurrence of a motion artefact at one time instance, the prediction error is increased for that time instances and decreases quickly as shown in Fig. 3.13.b. However, the variance remains increased as the artefact is still present within the test features and/or the training dataset. A practical interpretation of this results is that the variance remains increased even though the current prediction error might be low, as an artefact occurred in the “recent” history. The width of this history can be influenced by the feature dimension and number of training pairs.

Third, a feedback criterion which is based on the features  $\mathbf{x}_i$  instead of the labels  $y_i$  has the advantage that it can be evaluated even though the labels are not observed. This is an advantage especially for correlation algorithms where the internal position is only measured at a few time instances (to reduce the additional imaging dose). The monitoring of the current prediction error would not be possible.

- **Q.2.3:** How can these criteria be used to control the current prediction error?

An increased variance can be either used as an indicator to stop the treatment or

to use the prediction output of an alternative algorithm. Both approaches have been investigated. If the variance exceeds a threshold  $\sigma_{th}^2$ , the treatment could be interrupted. Depending on the selected threshold, this results in an decreased duty cycle. However, as the results in sec. 3.4.1 indicate, not all prediction errors can be controlled as the variance is delayed by the prediction latency  $h$ .

In sec. 3.4.2, two kind of hybrid algorithms have been investigated. The first one is based on a combination of one probabilistic algorithm and one non-probabilistic algorithm. If the variance of the probabilistic algorithm increases, the prediction results of the non-probabilistic algorithm could be used. The results indicate that potential hybrid algorithms between the **wLMS** and the **RVM** algorithm resulted in a superior performance compared to the results of the individual algorithms. However, this approach requires an additional parameter - the threshold variance. The second hybrid approach combines multiple probabilistic algorithms. The variance of each algorithm can be computed in parallel. The prediction output of the algorithm with the lowest variance is selected. A combination of three **RVM** algorithms was evaluated as an example and led to the lowest mean **RMSE** and **RMSE<sub>rel</sub>** values over the complete dataset for  $h = 115$  ms.



## 4 Multivariate Respiratory Motion Prediction<sup>1</sup>

In this chapter the potential of multivariate external data for respiratory motion prediction is explored. Several aspects will be considered such as the extension of univariate to multivariate prediction algorithms and the selection of the most relevant and least redundant sensors. These aspects refer to the questions **Q.1.1** and **Q.1.2**. The main focus, however, lies on question **Q.1.3** - can the prediction accuracy be increased by using multivariate external data? These questions can be addressed independently from question **Q.2**. Both, probabilistic and non-probabilistic prediction algorithms could benefit from multivariate data. Consequently, this chapter discusses prediction algorithms in general. Experimental results are shown for probabilistic as well as non-probabilistic prediction algorithms.

In sec. 4.1, a multivariate measurement is described which will be used later for evaluation purposes. The acquired datasets contain the information of multiple external sensors such as flow and acceleration. Section 4.2 provides the mathematical background on multivariate prediction and feature selection algorithms. The evaluation procedure including the used prediction and feature selection algorithms is described in sec. 4.3. The experimental results are shown and discussed in sec. 4.4. The chapter closes with an outlook on other potentially relevant external sensors such as EEG and sEMG (sec. 4.5). However, only preliminary results are shown as either the correlation to the respiratory motion is very low or the used pre-processing steps are not possible in real-time. Section 4.6 summarizes the main results with respect to the questions **Q.1.1** to **Q.1.3** and draws initial conclusions about the most relevant sensor combination (**Q.1.5**).

### 4.1 Multivariate Dataset

To investigate multivariate MC algorithms, a study was performed with 18 human subjects<sup>2</sup>. The measurement phases and a summary of the subject population are described

<sup>1</sup>Parts of this section have been published in [64–66, 74]

<sup>2</sup>The study was approved by the ethics committee of the University of Lübeck - file number 13-055

in sec. 4.1.1. The measurement hardware and sensor placement are presented in sec. 4.1.2. Further data pre-processing steps are discussed in sec. 4.1.3.

### 4.1.1 Measurement Phases and Subjects

The measurement did consist of two phases which we will refer to as M1 and M2. The objective of M1 was to simulate a normal treatment condition. The subjects were asked to breathe normally without any restrictions. Similar to a real treatment, the subjects were left alone in the measurement room. The complete duration of M1 was 20 min including time intervals to start and terminate all sensors. Slow occurring variations in the breathing pattern may be occur such as a decrease of the breathing frequency due to relaxation of the subjects or changes between abdominal and chest breathing. The second measurement phase M2 focused on unexpected spontaneous variations - breathing artefacts. The duration of M2 was 5 min. Throughout the first two minutes the subjects were asked to breathe normally. Afterwards, the subjects heard an acoustic signals every 15s. In total, ten acoustic signals were generated. Hearing a signal, the subjects were asked to simulate a breathing artefacts such as coughing, sneezing, harrumphing, a short breath hold or speaking. The kind and order of the breathing artefacts was not predetermined.

The measurement was performed with 18 subjects (11 male and 7 female). Details on the subject population are summarized in Table 4.1.

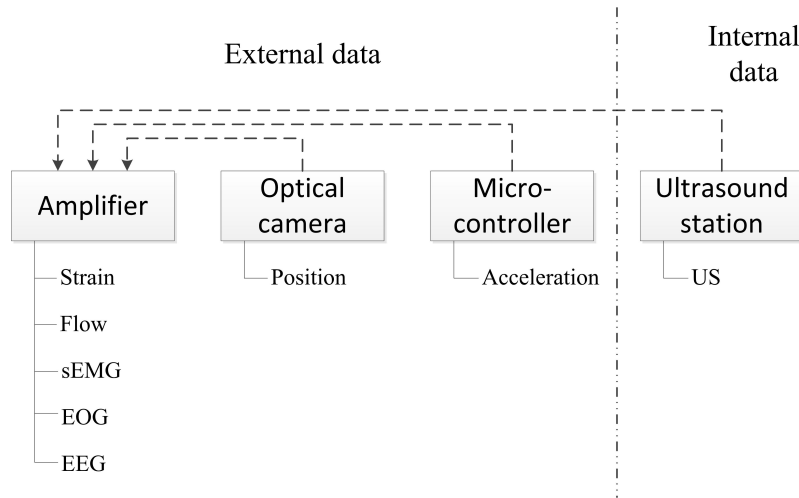
### 4.1.2 Measurement Setup

The measurement setup incorporated one internal and eight external sensor modalities. Fig. 4.1 shows a schematic representation of the measurement setup. It consisted of four independent measurement units (measurement computers, mircocontroller or compar-

**Table 4.1:** Summary of subject population divided into female and male

gender	age	body height [m]	body weight [kg]
female			
mean	29.3	1.65	60.5
range	22 - 34	1.53 - 1.8	50 - 78
male			
mean	24.6	1.85	77.6
range	21 - 29	1.75 - 1.95	65 - 93





**Figure 4.1:** Schematic representation of the four measurement units and their connections. The dashed arrows indicate transmitted strobe values.

able hardware which have an own system clock). The data of the strain, respiratory flow, sEMG, EEG and EOG sensors were acquired directly with an amplifier unit. The data of the other modalities (position, acceleration and ultrasound) were recorded by three separate measurement units. Temporal synchronization between the four units was achieved via strobe values. The tracking server (controlling the optical camera), the microcontroller (controlling the acceleration sensor) and the ultrasound station transmitted a specific binary pattern of strobe values which were additionally recorded by the amplifier. Details about the synchronization are presented in [51].

The amplifier unit consists of two g.tec USB amplifiers (g.tec medical engineering GmbH, Austria) which were operated in a master-slave mode. Each amplifier had 16 measurement channels with a resolution of 30 nV. The data of the five directly connected sensor modalities and the three strobe signals were recorded at a sampling frequency of  $f_s = 1200$  Hz.

The sensor placement is shown in Fig. 4.2.a. In the following each sensor modality will be described independently:

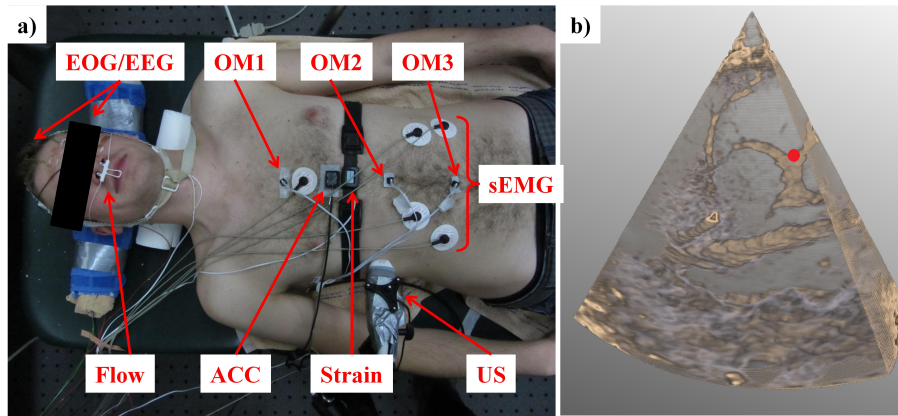
- *Position:* Equivalent to the currently used clinical setup of the CyberKnife<sup>®</sup> Synchrony, three active optical markers (OM) were used to acquire the position of the abdomen and the chest. The optical markers were attached along the median line of the chest and the abdomen of each subject. The first optical marker (OM1) was placed at the height of the areolas mamme. Data recorded by this marker represents chest breathing. The second optical marker (OM2) was placed below the sternum.

Changes of the position of OM2 were caused by movements of the chest and/or abdomen. The position of the abdomen was measured with the third optical marker (OM3), which was placed above the navel. The information of all three OMs were acquired by an accuTrack 250 system (Atracsys LLC, Switzerland). The camera was controlled via a tracking server. In [76], a 3D RMSE error of 0.082 mm was estimated for one moving OM. The OMs were sampled at an average frequency of  $f_s = 47$  Hz. Latencies due to data acquisition were corrected [51]. The remaining latencies were caused by network delays and were below 10 ms. Further details and a practical evaluation of the optical tracking system can be found in [76].

- *Acceleration (ACC)*: To measure the acceleration of the chest, one custom made ACC sensor was placed on the median line between the OM1 and the OM2. The sensor (STMicroelectronics, LIS3LV02DQ) acquired the acceleration for all three directions in a range of  $\pm 2g$  at a sampling frequency of  $f_s = 160$  Hz. The resolution was 12 bit. The sensor was interfaced by an Atmel AT90USB162 microcontroller. The microcontroller transmitted the data via a USB interface to an independent measurement computer. The acceleration sensor and the microcontroller were covered in a plastic housing. The microcontroller transmitted strobe values with a frequency of 10 Hz.
- *Strain*: The extension of a cross section of the torso was measured by a respiratory belt which consists of a piezo crystal sensor (SleepSense<sup>®</sup>, Double Buckle Piezo Crystal Effort Sensor) and an elastic belt. The belt was placed between the ACC sensor and the OM2. Depending on the strain of the belt, a small voltage could be measured at the piezo crystal. The sensitivity of the sensor varied between 100 – 250  $\mu V/mm$  depending on the tightness of the elastic belt. An increased tightness does result in an increased sensitivity. However, it can also limit the normal breathing of the subjects. As the tightness of the respiratory belt was not measured, the acquired data cannot be compared between patients.
- *Flow*: The respiratory flow was measured indirectly via thermistor elements which were placed in front of the nose and the mouth of the subjects. The sensor (SleepSense<sup>®</sup> Airflow Thermistor - Embla Compatible) is commonly used in sleep laboratories. In contrast to spirometry, the subjects did not need to breath into a mouth piece. This resulted in an increased comfort for the subjects and in a more normal breathing pattern. However, the sensor measured the respiratory flow based on the temperature difference in front of the nose and the mouth and not the actual air flow. This could result in a decreased measurement accuracy. Furthermore, addi-

tional dead times between the air flow and the measured voltage can be assumed as time is required to increase or decrease the temperature of the thermistor elements.

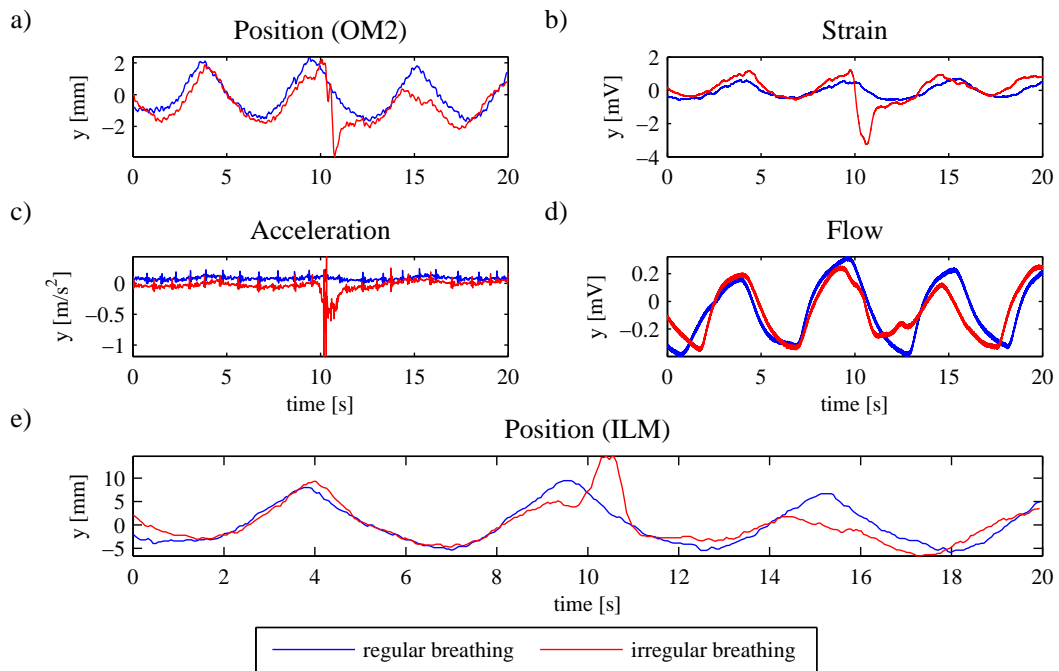
- *Surface electromyography (sEMG)*: To measure the activity of the respiratory muscles, the electrode placement proposed by Maarsingh *et al.* [161] was used. In contrast to the original electrode placement, two further measurement electrodes were added, which resulted in four measurement and one ground and reference electrode, respectively. The measurement electrodes were placed bilateral at the costal margin between the seventh and eighth rib and between the tenth and eleventh rib. The reference electrode was placed at the sternum between the OM1 and the ACC sensor and the ground electrode at the ankle of the right foot. The electrodes were standard self-adhesive **electrocardiography (ECG)** electrodes (ASF40C, Asmuth GmbH, Germany).
- *Electroencephalography (EEG) and electrooculography (EOG)*: The measurement setup included six further electrodes to measure the **EEG** and **EOG** of each subject. Two electrodes were placed above and below the left or right eye. The **EEG** was recorded at the C3 and C4 location of the 10-20 measuring system [127, 138]. The electrodes were standard Ag-AgCl ring-electrodes which were placed on the skin via a electrode cream (EC2 Genuine Grass Electrode Cream, Grass Technologies, USA). The ground electrode was placed central at the forehead and the reference electrode behind the right ear. The information from these electrodes could give indications about the alertness of each subject by counting for example the number of blinks per minutes or evaluation of the alpha frequency band of the **EEG** [139]. These additional features can potentially be used to classify the current respiratory states.
- *Ultrasound (US)*: To enable the possibility of investigating multivariate correlation algorithms, the movement of an internal target was monitored. Here, the movement of the liver was recorded by 4D **US**. The choice of **US** was motivated by the relatively high sampling frequency (compared to **MRI** and **CT**) and the lack of additional radiation as in case of fluoroscopy. An unambiguous point in the liver such as a vessel bifurcation was selected as target for each subject. It was assumed that this point represents the movement of a tumour. We refer to this target as **internal landmark (ILM)**. An example of an **US** image is shown in Fig. 4.2.b. A transfer function is used for illustration purposes to highlight the blood vessels. The target is marked by a red dot. The **US** transducer was placed lateral between two ribs. Before the measurement, the movement of the target was observed to verify that the target was



**Figure 4.2:** a) Placement of the EEG, EOG and sEMG electrodes, the three optical markers (OM1-3), the ultrasound transducer (US), the respiratory flow sensor (Flow), the acceleration sensor (ACC) and the respiratory belt (Strain); b) Example of an acquired US image and the internal target (ILM), which is highlighted by a red dot.

visible within the field of view of the transducer throughout a complete respiratory cycle. The transducer was attached via a tripod to the patient couch. The US images were acquired with a modified GE Vivid 7 Dimension ultrasound station with a 3V 3D/4D transducer. The frame rate was on average 17 Hz with a spatial resolution of 0.33 mm. Each recorded frame triggered one strobe value which was measured by the g.tec amplifier. The maximum expectable system latency is 28 ms according to [36]. The movement of the target was tracked via template matching using the minimum of the sum of squared distances. A template size of 15 and 25 mm<sup>3</sup> was investigated. The maximum distance in one direction between the templates of two consecutive frames was varied between 3 and 8 pixels. Additionally, a linear penalty factor was used, which penalises larger movements between two frames. The best signal was chosen by visual inspection. Further details about the used US framework and tracking algorithm can be found in Bruder *et al.* [36].

Note the focus of this chapter lies on multivariate prediction algorithms with external data. The recorded US data are mentioned here for completeness. In chapter 5, the US data will be used to investigate multivariate correlation algorithms.



**Figure 4.3:** Measured data of the external position of (a) OM2, (b) strain, (c) acceleration, (d) flow and (e) internal position of **ILM** for one subject for a regular and an irregular breathing segment.

### 4.1.3 Data Pre-processing

Some of the acquired data were further pre-processed before they were used as features in an **MC** algorithm. A 50 Hz notch filter was applied to the flow and strain data to remove noise most likely caused by the supply voltage. Further, the dimensionality of the **OMs**, the **ACC** sensor and the **ILM** were reduced to their first principle component as breathing motion does occur predominantly in one direction [142, 221]. This dimensionality reduction was performed to simplify further evaluations. Depending on the following experiments, the data were downsampled and interpolated to a common frequency  $f_c$ . In this chapter, only the external data are considered. The common frequency was set to  $f_c = 26$  Hz, which is equivalent to the sampling frequency of the CyberKnife<sup>®</sup> system. In chapter 5, the common frequency was set to the sample frequency of the **US** data to investigate motion correlation algorithms. The data of **sEMG**, **EEG**, and **EOG** sensor does require more complex pre-processing steps to extract relevant features. They will be discussed in detail in sec. 4.5.<sup>3</sup>

An example of a pre-processed recording for a regular and irregular breathing seg-

<sup>3</sup>All datasets can be downloaded at <http://signals.rob.uni-luebeck.de>.

ment of one subject is shown in Fig. 4.3. The motion segments were temporally aligned to highlight the effect of an irregular motion pattern. It can be observed that the same motion artefact does appear differently depending on the signal modality. The shape of the strain signal is comparable to the signal measured by the OM2. However, the OM2 signal seems to be noisier, as it represents only the movement of one point on the surface instead of the extension of a complete cross-section of the torso. In case of the flow sensor, the irregular motion pattern does not lead to an increased amplitude of the signal. This signal seems to be very smooth compared to the residual signal, which could be explained by the additional dead times due to the indirect measurement and human physiology. Investigating the acceleration signal indicates that the normal inhale and exhale movements are almost not visible. In contrast minor regular peaks are visible which correlate with the heart activity. Comparing the time point of the maximum absolute amplitude of the OM2 and the ACC signal indicates that the OM2 is slightly delayed with respect to the ACC sensor. Fig. 4.3.e shows the movement of the ILM. Even though the signal of OM2 and ILM seem to be strongly correlated, differences are clearly visible.

## 4.2 Mathematical Background

As presented in chapter 2.3.1, various prediction algorithms have been proposed. However, the majority of them was only evaluated by using the information of one sensor. In sec. 4.2.1, it will be investigated which of these algorithms can be used for multivariate datasets. The relevance of feature selection algorithms is discussed in sec. 4.2.2.

Note, following the notation for MC algorithms (defined in sec. 2.1), the input features of a prediction or correlation algorithm are  $\mathbf{X} = \{\mathbf{x}_j^\top | j = 1, \dots, m\}$ . In case of multivariate inputs, we extend these input features by adding additional features of other sensors. We define multivariate inputs to be  $\mathbf{X} = \{[\mathbf{x}_j^k]^\top | j = 1, \dots, m; k = 1, \dots, o\}$  with  $o$  being the number of sensors. Further the term **feature set (FS)** shall be introduced to simplify the notations and group the features of one or multiple sensors. Exemplarily, the FS  $S = \{\text{OM2}\}$  means that only features of the marker OM2 are used and that the input features are  $\mathbf{X} = \{[\mathbf{x}_j^{\text{OM2}}]^\top | j = 1, \dots, m\}$ . An FS  $S = \{\text{OM2}, \text{ACC}\}$  indicates that features of the OM2 and the ACC sensor are used and that  $\mathbf{X} = \{[\mathbf{x}_j^{\text{OM2}}, \mathbf{x}_j^{\text{ACC}}]^\top | j = 1, \dots, m\}$ . The cardinal number of a FS is denoted by  $s$ , which would be  $s = 2$  in the latter case.

### 4.2.1 Multivariate Prediction Algorithms

In [76], prediction algorithms were classified into model-based and model-free methods. The first class of algorithms makes use of prior knowledge about the functional beha-

viour. In case of modelling respiratory motion, one obvious assumption is the quasi-periodic nature of the signal. This knowledge can be used to construct prediction algorithms, as it was done in case of the sinusoidal models presented in [192, 266] and the extended Kalman filter approach of [206]. The GP model as it was introduced in sec. 3.2.2 is also a model-based algorithm. Note that the GP framework is very general and a model-free GP prediction algorithm could easily be designed. According to Ernst [76], the second class of algorithms does not require prior knowledge about the signal. The majority of the published prediction algorithms belong to the second group, such as the ANN approaches of [185, 241], the adaptive filter algorithms of [77, 81] and SVR algorithms of [80, 142]. The RVM algorithm of sec. 3.2.1 is another representative of this class. These algorithms have in common that they are based on the general assumption of an ARMA model (defined in Eq. 2.20), meaning that a prediction can be computed by the weighted sum of previous observations.

Strictly speaking, the term “model-free” may be misleading as the algorithms assume an ARMA model. Nonetheless the term is used in this context to indicate that these models do not use assumptions which are specific for respiratory motion signals and to be consistent with [76].

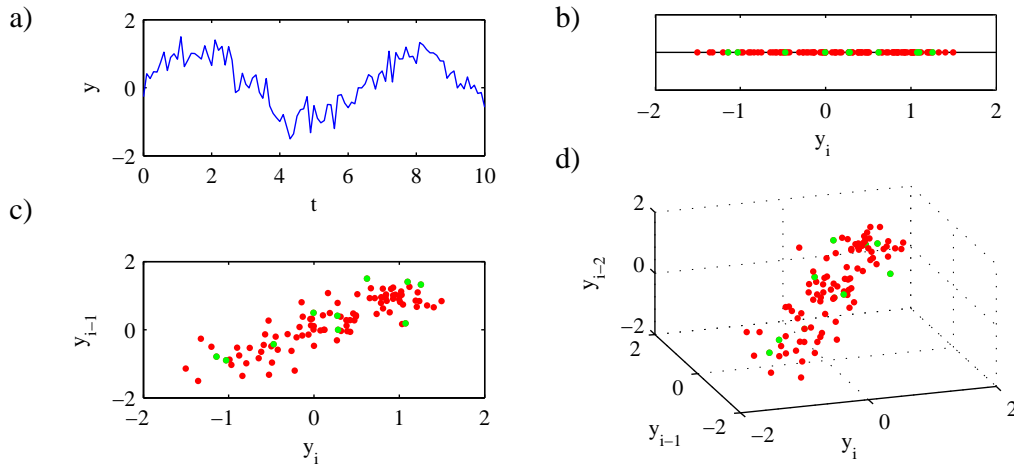
In principle, model-based and model-free algorithms can be extended to multivariate inputs/features. However, it is more challenging to extend model-based algorithms as the assumptions are often depending on the underlying sensor modality. In case of multivariate inputs, these assumptions and their interaction have to be known for all modalities. To illustrate this, the sinusoidal model proposed by Vedam *et al.* [266] shall be considered as an example. It is defined as Eq. 2.15

$$y_{i+\xi} = a_1 \sin(a_2 x_i + a_3) + a_4 \quad (4.1)$$

with  $x_i = t_{i+\xi}$ . The model parameters  $a_1$  to  $a_4$  are estimated based on a training dataset consisting of features of one sensor. This model could be extended to multivariate features by assuming an individual sinusoidal model for each sensor. The prediction results  $y_{i+\xi}^*$  could be the weighted sum of the individual models such as

$$y_{i+\xi} = \sum_{k=1}^o a_1^k \sin(a_2^k x_i + a_3^k) + a_4^k \quad (4.2)$$

with  $o$  being the number of sensors. However, considering the example recordings shown in Fig. 4.3, it becomes obvious that the sinusoidal model is not appropriate for data of all sensors. Fig. 4.3.c shows the data of the ACC sensor where only minor periodic movements are visible. As a result, the model assumptions have to be adapted for each



**Figure 4.4:** (a) Simulated periodic signal; Feature space assuming (b)  $\mathbf{x}$  to be  $x_i = y_i$ , (c)  $\mathbf{x}_i = [y_i, y_{i-1}]^\top$ , and (d)  $\mathbf{x}_i = [y_i, y_{i-1}, y_{i-2}]^\top$ . The feature space of the simulated signal is indicated by red dots and the reduced feature space by the green dots.

sensor combination.

In contrast, model-free algorithms are more flexible. In the simplest case these models assume only

$$y_{i+\xi} = \mathbf{w}_i^\top \mathbf{x}_i \quad (4.3)$$

with  $\mathbf{u}_i$  being a feature vector of the  $m$  last observations (similar to Eq. 2.16). This algorithm can be extended to multivariate features by simply extending the feature vector  $\mathbf{x}_i$  to

$$\tilde{\mathbf{x}}_i = [\mathbf{x}_i^1, \dots, \mathbf{x}_i^o]^\top \quad (4.4)$$

with  $\tilde{\mathbf{x}}_i \in \mathbb{R}^{mo \times 1}$  assuming that each individual feature vector considers the last  $m$  observations. A prediction can be computed by

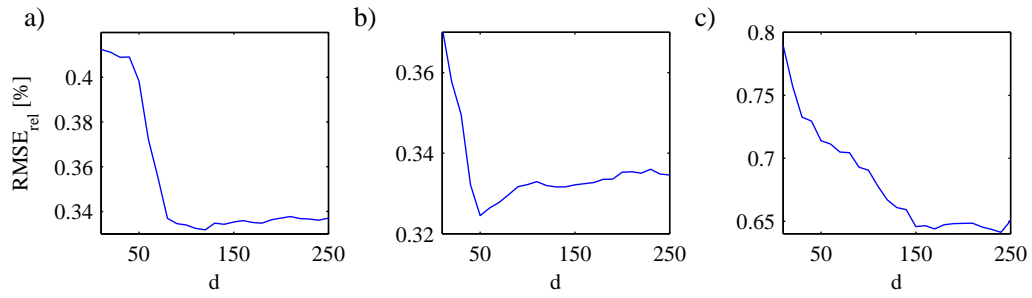
$$y_{i+\xi} = \tilde{\mathbf{w}}_i^\top \tilde{\mathbf{x}}_i \quad (4.5)$$

where  $\tilde{\mathbf{w}}_i = [\mathbf{w}_i^1, \dots, \mathbf{w}_i^o]^\top \in \mathbb{R}^{mo \times 1}$ . No sensor specific prior knowledge is used and arbitrary sensor combinations can be applied without adaptation of the algorithm.

## 4.2.2 Feature Selection

The motivation for extending univariate to multivariate feature sets is that the features of different sensors can contain additional relevant information which can be used to





**Figure 4.5:**  $\text{RMSE}_{rel}$  of the RVM algorithm depending on  $d$  for (a) the regular, (b) the irregular and (c) the noisy-irregular breathing fragments with number of training data  $m$  fixed according to Fig. 3.4.

increase the prediction accuracy. The consequence is an increased dimension  $d$  of the feature vector  $\mathbf{x}$ . Obvious drawbacks are the increase of storage as well as computational requirements. The latter is especially critical for real-time applications. Another problem is the so called “curse of dimensionality” [106, 267]. Increasing the number of features, a model requires more training data to be able to learn the feature space or at least the relevant part of it. The effect is illustrated in Fig. 4.4. We assume that the aim is to predict the synthetically generated periodic signal shown in Fig. 4.4.a. Fig. 4.4.b illustrates the feature space if only the most recent observation is used as feature ( $x_i = y_i$ ). The red dots represent the feature space if all points are considered as training features. A reduced training set, where only every tenth point is considered, is indicated by the green dots. It can be observed that the reduced training set covers the complete relevant feature space. However, it seems unrealistic that future signal values can be predicted based on only the most recent observation. If the feature dimension is increased to the two and three most recent observations (Fig. 4.4.c-d), the reduced training set (green dots) covers less percent of the relevant feature space (red dots). A learned prediction model will not be valid for the complete feature space, which can lead to an increased prediction error. Verleysen *et al.* [267] argued that the amount of training data should increase exponentially with the dimension. If 10 training features are considered to be optimal for a model with a one-dimensional feature space, 100 training features should be used for a two-dimensional feature space.

This effect could be already observed in Fig. 3.4 while investigating the parameters of the RVM algorithm (sec. 3.2.1). Fig. 4.5 shows the  $\text{RMSE}_{rel}$  for the three motion fragments if the number of training pairs  $m$  is fixed to the optimal  $\text{RMSE}_{rel}$  result. If  $d$  is small, the  $\text{RMSE}_{rel}$  is high as the feature dimension is too low to predict the complex res-

piratory motion pattern. Depending on the complexity of the motion fragment, a feature dimension between  $50 \leq d \leq 245$  does lead to the lowest  $\text{RMSE}_{rel}$ . Increasing  $d$  further (without increasing the number of training data) results in an increased prediction error as the algorithm succumbs to the “curse of dimensionality”.

One possibility to reduce this effect offer feature selection methods. The general idea is to generate a subset  $S$  of all possible features  $F$ . This can be, e.g., a forward selection algorithm. Assuming that an initial FS  $S_0$  is given, a new feature should only be added to the set if it has a high relevance with respect to the prediction labels and a low redundancy with respect to features which are already present in  $S_0$ .

A profound overview of different feature selection algorithms is given in [106]. Roughly, these methods can be divided into two classes - filter and wrapper methods. Filter methods use general information criteria such as the Pearson’s correlation coefficient or mutual information to rank the relevance of features. The Pearson’s correlation coefficients for two signals  $x$  and  $y$  is defined as

$$r_{x,y}(x, y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}, \quad (4.6)$$

where  $\bar{x}$  and  $\bar{y}$  are the mean of  $x$  and  $y$ . The mutual information is defined for discrete variables as

$$I_{x,y}(x, y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}, \quad (4.7)$$

with  $p(x)$  and  $p(y)$  being the marginal probability distribution of  $X$  and  $Y$  respectively and  $p(x, y)$  being the joint probability distribution of  $X$  and  $Y$ .

As discussed in sec. 2.3.2, the Pearson’s correlation coefficient was frequently used in the literature to investigate if data from external sensors can be used to predict internal tumour. In general the ranking is carried out independent of the specific prediction algorithm and the learning objective such as decreasing the  $\text{RMSE}_{rel}$ . This may make these methods unspecific. However, they are often appealing through their low computational requirements. Further information criteria are the Akaike and Bayesian information criterion. These criteria were evaluated in [68] for the LMS and wLMS prediction algorithms on the dataset presented in sec. 3.1. The results indicate that the Akaike information criterion is a valid method to estimate a patient specific feature dimension  $d$ .

Wrapper methods are specific with respect to the selected prediction algorithm and the learning objective. Often wrapper methods use the prediction algorithm as a black box and simply evaluate the outcome of this box (e.g. the  $\text{RMSE}_{rel}$ ) depending on its input (the selected features). Unsurprisingly, these methods tend to find better FSs than

filter methods for specific applications. Depending on the selected prediction algorithm, the computational requirements can be very high. As pointed out in [106], efficient search strategies through the potential feature combinations are required. The most trivial strategies are the forward selection and backward elimination. In case of forward selection methods, the method starts with none or only one feature and adds further features successively until the learning objective cannot be further decreased. In contrast, backward elimination algorithms are initialised with all possible features and aim to eliminate features successively.

### 4.3 Evaluation Procedure and Proposed Methods

Two experiments were performed to investigate the potential of multivariate features. First, the Pearson's correlation coefficient between all external sensors was computed for the different measurement phases. This enabled a principle investigation of the correlation independent of a specific prediction algorithm. A high Pearson's correlation coefficient can indicate that data from another sensor such as flow or ACC are relevant to predict future values of, e.g., an optical marker. The second experiment focused on the increase of the prediction accuracy of four selected prediction algorithms if multivariate instead of univariate features were used. The features sets were estimated by a sequential forward selection method. In the following, further details about the two experiments are presented.

**Experiment 1: Correlation Analysis** The Pearson's correlation coefficient  $r$  was computed for all external sensors with respect to the three optical markers OM1 - 3. The correlation coefficient was defined in Eq. 4.6. In case of the measurement phase M1, the complete measurement was divided into motion fragments with a duration of 1 min. The first and last minute of each recording was discarded to avoid the influence of unnatural breathing patterns due to the initialization or termination of the measurement. This resulted in 18 motion fragments per subject. The mean absolute correlation coefficient  $|r|$  and its standard deviation were computed over all motion fragments and all subjects. Furthermore, the variation of the correlation coefficients between the 1<sup>st</sup> and the 6<sup>th</sup>, the 12<sup>th</sup>, and the 18<sup>th</sup> minute was investigated. We refer to  $v_i^k$  as the observations  $y^k$  of marker  $k$  within the time interval  $t \in ((i-1) \cdot 60, i \cdot 60]$  s with  $i = 1, \dots, 18$ . The absolute correlation coefficient difference  $|\Delta r|$  was defined as

$$|\Delta r_{x,y}^{1,i}| = |r_{x,y}(v_1^x, v_1^y) - r_{x,y}(v_i^x, v_i^y)| \quad (4.8)$$

with  $i \in \{6, 12, 18\}$ ,  $x \in \{OM1, OM2, OM3\}$ , and  $y \in \{OM1, OM2, OM3, Strain, Flow, ACC\}$ . A high  $|\Delta r_{x,y}^{1,j}|$  indicates a stronger variation of the correlation coefficient between between the first minute and minute  $i$ .

In case of the measurement phase M2, two time windows were investigated. The first window represented regular breathing and was defined as  $t \in [55, 115]$  s. The second window represented irregular breathing within  $t \in (115, 260]$  s. Again, the residual measurement data of M2 was ignored to avoid the influence of the initialization and termination of the measurement. The mean absolute correlation coefficient and standard deviation was computed over all subjects.

The Fisher transform was used to compute the mean and standard deviation of the correlation coefficients and their differences [95, 254].

**Experiment 2: Multivariate Prediction** In this experiment, four model-free prediction algorithms were evaluated on multivariate input features. The least complex algorithm was the **nLMS** algorithm. Furthermore, the **wLMS** and **SVR** algorithm were selected based on the results of Ernst *et al.* [86] as well as the **RVM** algorithm presented in sec. 3.2.1. The first and the last minute of M1 and the first and the last minute 45 s of M2 were removed of each recording to eliminate irregular motion artefacts. This lead to a duration of the motion traces of 18 min for M1 and 210 s for M2.

The first minute of the remaining motion traces was selected as the training dataset for each algorithm. The residual data was designated as the test set. The **SVR** and **RVM** algorithm might be sensitive to the scaling of the signals. Therefore each training set was scaled to  $[0, 1]$ . The scaling parameters were also used to pre-process the unknown test data.

The algorithms were optimized by grid search over the entire parameter space. To accelerate the training process a sequential grid search was used, which optimized the parameters independently of each other. In the first step, ten equidistantly sampled parameters values were selected across the parameter space for all parameters. The parameter values leading to the lowest prediction error were used in a second step to define a narrowed parameter space. This new parameter space was used to generate ten new equidistantly sampled parameters values. This optimization was six times repeated. The **SVR** approach had the highest number of parameters. To accelerate the **SVR** training process further, a constant update factor  $f_{global} = 10$  was used [62]. Throughout the test phase, an  $\varepsilon$ -tube dependent update factor of  $f_{\varepsilon} = 2$  was used as discussed in [62]. In case of the **wLMS** algorithm, the optimal number of wavelet decompositions for univariate optical features is  $J = 3$  according to [81]. Within this experiment,  $J = 0$  was also

**Table 4.2:** Overview of all algorithm specific parameters

algorithm	number of parameters	parameters opt. by grid search	fixed parameters
nLMS	2	$d, \mu$	-
wLMS <sub>J0</sub>	4	$d, \mu$	$J = 0, m = 199$ [81]
wLMS <sub>J3</sub>	4	$d, \mu$	$J = 3, m = 199$ [81]
SVR	5	$d, C, \gamma, \epsilon$	$m = 1000$ [80]
RVM	2	$d$	$m = 1000$ [67]

evaluated to which we refer to as wLMS<sub>J0</sub> and wLMS<sub>J3</sub> (for  $J = 3$ ). The evaluation of wLMS<sub>J3</sub> was simplified by assuming a constant feature dimension  $d$  for each wavelet scale. The number of training pairs  $m$  was selected according to previous publications. An overview of the parameters of all algorithms is shown in Table 4.2.

A **sequential forward selection (SFS)** method was chosen to find the optimal sensor combination based on the  $\text{RMSE}_{rel}$ . The complete feature set was defined as  $F = \{\text{OM1}, \text{OM2}, \text{OM3}, \text{ACC}, \text{Strain}, \text{Flow}\}$ . The second optical marker OM2 was selected as prediction target for all algorithms. Note OM2 was selected as an example, the evaluation could have been performed also with OM1 or OM3. The initial feature set was defined as  $S_0 = \{\text{OM2}\}$ . With  $S_0$ , a prediction error  $\text{RMSE}_{rel}^{S_0}$  could be achieved. In the next stage of the forward selection method,  $S_0$  was extended to  $S_1$  by one of the remaining features  $F \setminus S_0$ . The feature combination leading to the lowest  $\text{RMSE}_{rel}^{S_1}$  was selected to be the optimal feature set. The procedure was repeated up to a feature set size of  $s = 3$  or until no further features could be found that decreased the  $\text{RMSE}_{rel}$ . In the optimal case, three improved feature sets could be found to which we refer as  $S_1, S_2$ , and  $S_3$ .

Two evaluation scenarios were considered for both measurement phases M1 and M2:

- **Test set unknown (TSUK):** This scenario represented the practical relevant case, that the test dataset was unknown. The feature sets  $S_1$  to  $S_3$  were computed based on the  $\text{RMSE}_{rel}$  of the training dataset.
- **Test set known (TSK):** This scenario assumed that all data was given and the optimal feature combination could be estimated based on the test dataset.

Even though **TSK** was not practically relevant, the results of **TSK** could be interpreted as the best possible results. Thus, the differences between **TSK** and **TSUK** gave insight into how good the optimal feature set for the test set could be selected based on the training

set.

The prediction latency was set to  $h = 115$  ms ( $\xi = 3$ ). The algorithms were evaluated with respect to the **RMSE**, the **RMSE<sub>rel</sub>**, and the **ΔRMSE<sub>rel</sub>**. In contrast to the last chapter (Eq. 3.45), here **ΔRMSE<sub>rel</sub>** is defined as the difference **RMSE** value with respect to the initial feature set  $S_0$

$$\Delta\text{RMSE}_{rel} = \text{RMSE}_{rel}(W) - \text{RMSE}_{rel}(S_0), \quad (4.9)$$

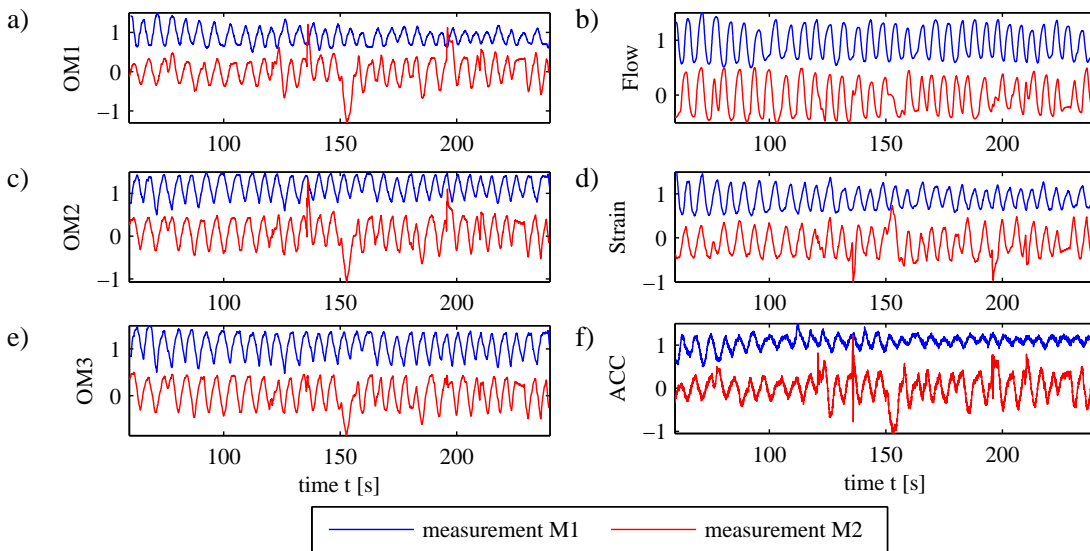
with  $\text{RMSE}_{rel}(S_0)$  being the **RMSE<sub>rel</sub>** of the initial **FS**  $S_0$  and  $\text{RMSE}_{rel}(W)$  the **RMSE<sub>rel</sub>** of the investigated **FS**  $W \in \{S_0, S_1, S_2, S_3, F\}$ .

## 4.4 Experimental Evaluation

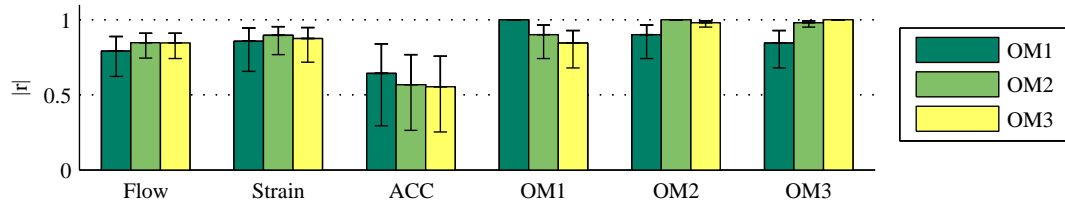
The results of the correlation analysis are presented in sec. 4.4.1. Section 4.4.2 contains the multivariate prediction results of the four investigated algorithms and the two evaluation scenarios. Finally, the results of both experiments are discussed in sec. 4.4.3.

### 4.4.1 Correlation Analysis

Figure 4.6 illustrates exemplarily of the recordings for all six external sensors of one subject. Each separate figure shows a scaled motion fragment of measurement phase M1 and



**Figure 4.6:** Scaled example recordings of the six external markers for measurement M1 (blue) and M2 (red). The signals are shifted by  $\pm 0.5$  for visualization purposes.

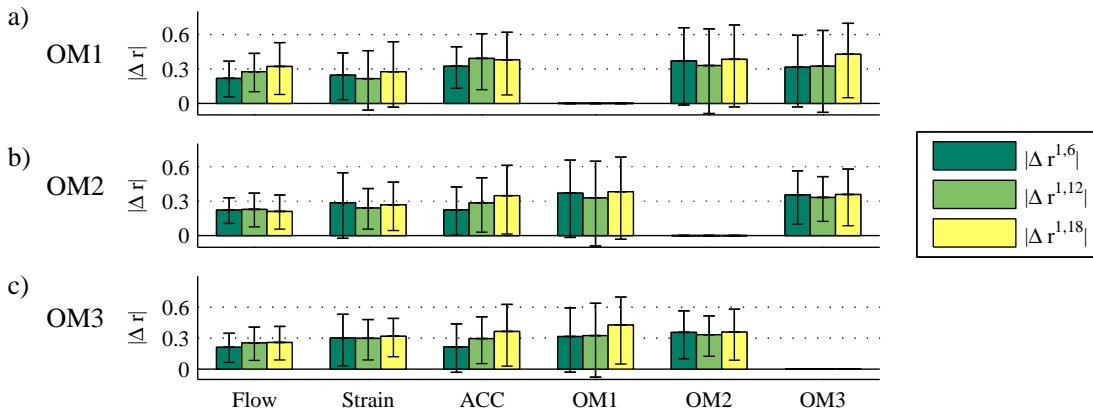


**Figure 4.7:** Mean absolute correlation coefficient  $|r|$  for the external markers and OM1, OM2, and OM3 of the measurement phase M1.

M2. The two signals are shifted by an offset ( $\pm 0.5$ ) along the  $y$ -axis for visualization purpose. The lower signal represents M2 and contains several motion artefacts for  $t > 115$  s. Similar to Fig. 4.3, the artefacts are manifested differently for different sensor modalities. Furthermore, differences can also be observed within the same sensor modality by comparing OM1 - 3.

The mean absolute correlation coefficients of measurement phase M1 over all subjects and one-minute fragments are shown in Fig. 4.7. The mean correlation of an optical marker with itself is  $|r_{OM1,OM1}| = |r_{OM2,OM2}| = |r_{OM3,OM3}| = 1$ . In general, the correlation values are high for all investigated markers as the mean correlation is  $|r| > 0.5$ . The highest mean correlation and lowest standard deviation is achieved for OM2 and OM3 with  $|r_{OM2,OM3}| = 0.98$ . On average, the correlation of OM1 with respect to OM2 and OM3 decreases as the spatial distance between the markers increases ( $|r_{OM1,OM2}| = 0.9$ ,  $|r_{OM1,OM3}| = 0.84$ ). These results indicate a dependency on the marker position. The influence of the marker placement may be reduced by using alternative modalities. In case of the strain marker, only a small variation of the mean correlation is observed. The lowest mean correlation is  $|r_{Strain,OM1}| = 0.86$  and the highest  $|r_{Strain,OM2}| = 0.9$ . Note the strain belt is placed between OM1 and OM2. Interestingly, the mean correlation of the strain marker and OM3 ( $|r_{Strain,OM3}| = 0.88$ ) is higher compared to OM1. Similar results can be observed for the data of the flow sensor. The ACC marker has the lowest mean correlation and highest standard deviation with respect to all optical markers. This could be expected as the data of the ACC marker is the second derivative of the data of the optical marker (spatial position) and the Pearson's correlation coefficient considers only linear correlation. Using a nonlinear transformation of the ACC data might lead to higher correlation results.

The absolute mean correlation differences between the correlation coefficients of the 1<sup>st</sup> and the 6<sup>th</sup>, the 12<sup>th</sup>, and the 18<sup>th</sup> minute are shown in Fig. 4.8. Considering only the optical marker, the correlation between the 1<sup>st</sup> and the 6<sup>th</sup> minute did vary between  $|\Delta r_{OM1,OM3}^{1,6}| = 0.31$  for OM1 and OM3 up to  $|\Delta r_{OM1,OM2}^{1,6}| = 0.37$  for OM1 and OM2.

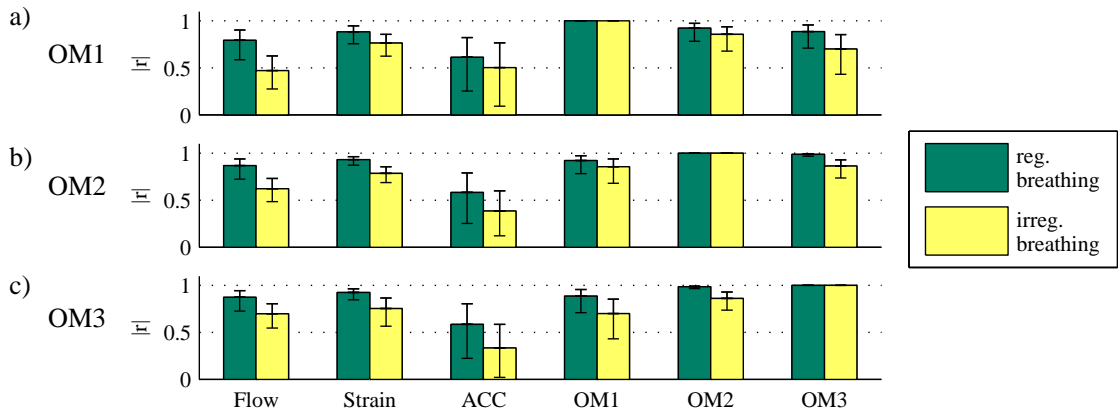


**Figure 4.8:** Mean absolute correlation coefficient difference  $|\Delta r|$  of the external marker with respect to (a) OM1, (b) OM2, (c) and OM3 between the correlation coefficients of the 1<sup>st</sup> and the 6<sup>th</sup>, the 12<sup>th</sup>, and the 18<sup>th</sup> minute of M1.

Increasing the time difference did effect the correlation variation only slightly. The maximum correlation difference could be observed for OM1 and OM3 for the difference between the 1<sup>st</sup> and the 18<sup>th</sup> minute with  $|\Delta r_{OM1,OM3}^{1,18}| = 0.43$ . The investigation of the correlation difference for alternative modalities revealed smaller mean values and standard deviations of  $|\Delta r|$  especially for data of the flow and the strain sensor. The lowest mean correlation difference between the 1<sup>st</sup> and the 6<sup>th</sup> minute was  $|\Delta r_{OM3,Flow}^{1,6}| = 0.21$ , between the 1<sup>st</sup> and the 12<sup>th</sup> minute  $|\Delta r_{OM1,Strain}^{1,12}| = 0.22$ , and between the 1<sup>st</sup> and the 18<sup>th</sup> minute  $|\Delta r_{OM1,Flow}^{1,18}| = 0.21$ .

The mean absolute correlation coefficients  $|r|$  of the second measurement phase M2 with respect to the OM1 to OM3 are shown in Fig. 4.9.a-c. The results are divided into regular (green) and irregular (yellow) breathing. Considering only the correlation results for regular breathing, similar results as in case of M1 can be observed. The optical markers have the highest mean correlation followed by the strain, flow, and ACC marker. Furthermore, a dependency on the position of the optical marker can be observed ( $|r_{OM1,OM2}| = 0.92$ ,  $|r_{OM1,OM3}| = 0.87$ ). Comparing the results between regular and irregular breathing reveals an decrease of the mean correlation coefficients and an increase of the standard deviation for all sensors. As an example, the mean correlation for regular breathing is  $r_{OM2,OM3} = 0.99$  between OM2 and OM3 and decrease to  $|r_{OM2,OM3}| = 0.86$  in case of irregular breathing. The highest difference of  $|r|$  between regular and irregular breathing can be observed in case of the flow marker with respect to OM1 (regular:  $|r_{OM1,Flow}| = 0.79$ ; irregular:  $|r_{OM1,Flow}| = 0.47$ ) and the ACC to OM3 (regular  $|r_{OM3,ACC}| = 0.59$ ; irregular  $|r_{OM3,ACC}| = 0.33$ ).





**Figure 4.9:** Mean absolute correlation coefficient  $|r|$  between (a) OM1, (b) OM2, (c) OM3 and the external markers for the regular and irregular breathing segments of measurement M2.

#### 4.4.2 Multivariate Prediction

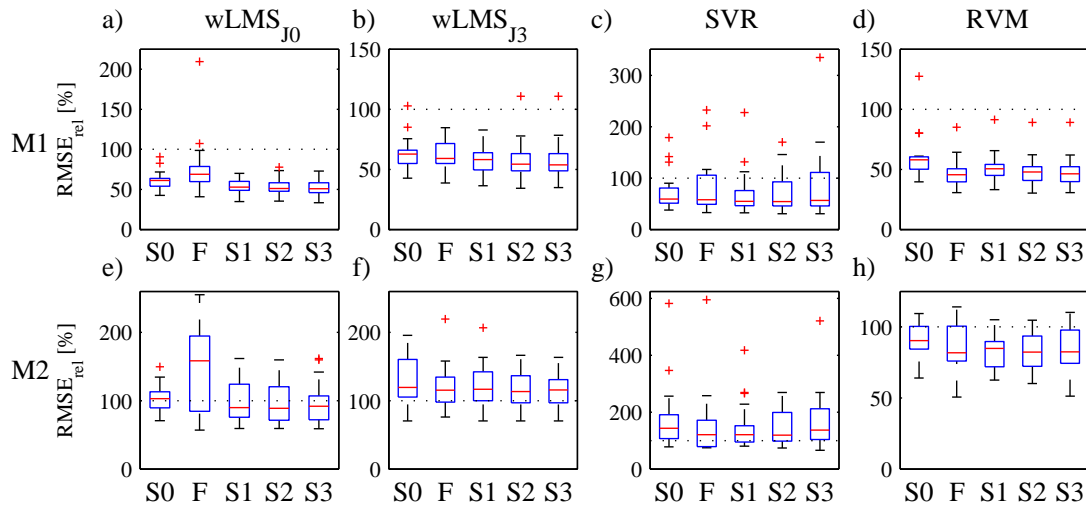
A detailed overview of the results of the second experiment are displayed in Table 4.3. The results of M1 are contained in Table 4.3.a and those of M2 in Table 4.3.b. The mean  $RMSE$ ,  $RMSE_{rel}$ , and  $\Delta RMSE_{rel}$  over the 18 subjects are presented for the initial feature set  $S_0$ , the complete set  $F$  and the optimized sets  $S_1$  to  $S_3$  for the scenario of an unknown (TSUK) and a known (TSK) test set. The best results for TSUK and TSK are highlighted in bold for each algorithm. Comparing the results of  $S_0$  with the residual FSs indicate that except of the nLMS algorithm, all algorithms can benefit by using the sets  $S_1$  to  $S_3$  or  $F$ . The nLMS algorithm has partly high averaged prediction errors due to outliers. To present meaningful results, the outliers were removed. The values marked by † indicate outlier corrected results with the number of removed outliers in parentheses. Values marker by \* indicate a significant different  $RMSE$  or  $RMSE_{rel}$  value compared to the initial set  $S_0$  (t-test,  $p = 0.05$ ).

To further highlight the increase of the prediction accuracy, the  $RMSE_{rel}$  results are shown in box plot diagrams in Fig. 4.10 and the  $\Delta RMSE_{rel}$  in Fig. 4.11 for TSUK. The first row of each figure displays the results of measurement phase M1 and the second row of phase M2. The results of the nLMS algorithm are ignored for illustration purposes. Note the scaling of some figures had to be adapted due to outliers especially for the SVR and the  $wLMS_{J_0}$  algorithm.

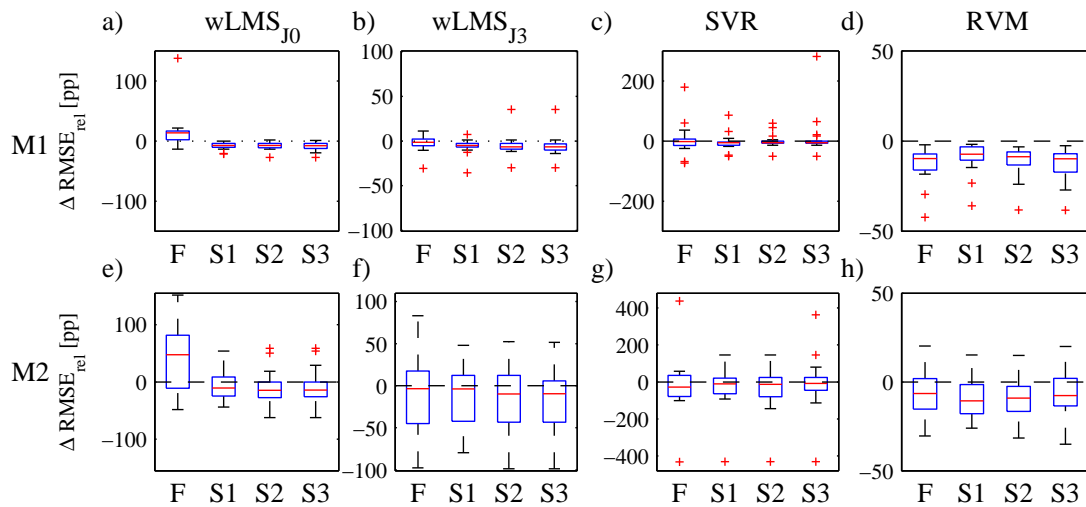
Table 4.4 lists how often a feature occurred within one of the extended feature sets  $S_1$  to  $S_3$  for M1. The values without parentheses represent the occurrence for the evaluation scenario TSUK and those in parentheses for TSK. Table 4.5 contains the results for meas-

**Table 4.3:** Mean **RMSE**, **RMSE<sub>rel</sub>** and **ΔRMSE<sub>rel</sub>** for all 18 subjects for (a) M1 and (b) M2 of the FS  $S_0$ ,  $F$  and  $S_1$  to  $S_3$  of the SFS method for the scenario of unknown (TSUK) and known (TSK) test sets. (Best results for TSUK and TSK highlighted bold; \* significant difference compared to the initial set  $S_0$  (t-test,  $p = 0.05$ ); † outlier corrected result (number of outliers shown in parentheses))

measure	a) Measurement phase M1 - normal breathing																	
	Test set unknown						Test set known											
	$S_0$	$F$	$S_1$	$S_2$	$S_3$	$S_1$	$S_2$	$S_3$										
nLMS	RMSE [mm]	0.187	0.174	0.166† (5)	<b>0.157† (4)</b>	<b>0.157† (4)</b>	0.163	<b>0.161</b>	<b>0.161</b>	RMSE <sub>rel</sub> [%]	74.91	70.50	68.42† (5)	66.15† (4)	<b>65.97† (4)</b>	65.86	65.13	<b>65.04</b>
	ΔRMSE <sub>rel</sub> [pp]	0.0	-4.41	-6.49† (5)	-8.76† (4)	<b>-8.94† (4)</b>	-9.05	-9.78	<b>-9.87</b>		0.0	0.0	-4.41	-6.49† (5)	<b>-8.94† (4)</b>	-9.05	-9.78	<b>-9.87</b>
wLMS <sub>J0</sub>	RMSE [mm]	0.146	0.186	0.126	0.127	<b>0.124</b>	0.125	0.122	<b>0.121</b>	RMSE <sub>rel</sub> [%]	61.00	76.86	52.74*	52.75*	<b>51.72*</b>	52.56*	50.90*	<b>50.47*</b>
	ΔRMSE <sub>rel</sub> [pp]	0.0	15.86	-8.27	-8.26	<b>-9.28</b>	-8.44	-10.10	<b>-10.53</b>		0.0	0.0	-8.27	-8.26	<b>-9.28</b>	-8.44	-10.10	<b>-10.53</b>
wLMS <sub>J3</sub>	RMSE [mm]	0.151	0.145	<b>0.137</b>	0.140	0.139	0.134	0.130	<b>0.129</b>	RMSE <sub>rel</sub> [%]	63.28	60.36	<b>57.13</b>	58.26	57.92	55.87	54.07*	<b>53.80*</b>
	ΔRMSE <sub>rel</sub> [pp]	0.0	-2.93	<b>-6.16</b>	-5.03	-5.36	-7.41	-9.21	<b>-9.48</b>		0.0	0.0	-6.16	-5.03	-5.36	-7.41	-9.21	<b>-9.48</b>
SVR	RMSE [mm]	0.181	0.191	<b>0.176</b>	0.176	0.202	0.146	0.136	<b>0.133</b>	RMSE <sub>rel</sub> [%]	75.26	79.92	<b>70.86</b>	74.84	89.32	60.84	56.56	<b>55.31</b>
	ΔRMSE <sub>rel</sub> [pp]	0.0	4.66	<b>-4.40</b>	-0.42	14.06	-14.42	-18.70	<b>-19.95</b>		0.0	0.0	-4.40	-0.42	14.06	-14.42	-18.70	<b>-19.95</b>
RVM	RMSE [mm]	0.146	0.116	0.124	0.118	<b>0.116</b>	0.124	0.117	<b>0.114</b>	RMSE <sub>rel</sub> [%]	60.51	47.64*	51.33	48.97*	<b>47.90*</b>	51.25	48.37*	<b>47.07*</b>
	ΔRMSE <sub>rel</sub> [pp]	0.0	-12.87	-9.18	-11.54	<b>-12.61</b>	-9.26	-12.14	<b>-13.44</b>		0.0	-12.87	-9.18	-11.54	<b>-12.61</b>	-9.26	-12.14	<b>-13.44</b>
b) Measurement phase M2 - breathing artifacts																		
nLMS	RMSE [mm]	0.657	0.621	0.871	0.845	<b>0.747</b>	0.585	0.579	<b>0.578</b>	RMSE <sub>rel</sub> [%]	113.76	107.20	135.01	134.12	<b>123.88</b>	100.82*	99.89*	<b>99.60*</b>
	ΔRMSE <sub>rel</sub> [pp]	0.0	-6.56	21.25	20.36	<b>10.12</b>	-12.94	-13.87	<b>-14.16</b>		0.0	-6.56	21.25	20.36	<b>10.12</b>	-12.94	-13.87	<b>-14.16</b>
wLMS <sub>J0</sub>	RMSE [mm]	0.585	0.921*	0.546	0.516	0.523	0.473	0.447*	<b>0.444*</b>	RMSE <sub>rel</sub> [%]	105.03	152.02*	99.45	95.66	96.67	85.27*	80.32*	<b>79.83*</b>
	ΔRMSE <sub>rel</sub> [pp]	0.0	46.98	-5.58	-9.38	-8.37	-19.76	-24.72	<b>-25.20</b>		0.0	46.98	-5.58	-9.38	-8.37	-19.76	-24.72	<b>-25.20</b>
wLMS <sub>J3</sub>	RMSE [mm]	0.763	0.676	0.679	0.654	<b>0.640</b>	0.539	0.516*	<b>0.507*</b>	RMSE <sub>rel</sub> [%]	131.48	119.83	119.55	117.06	<b>115.61</b>	96.20*	92.21*	<b>90.60*</b>
	ΔRMSE <sub>rel</sub> [pp]	0.0	-11.65	-11.93	-14.41	<b>-15.87</b>	-35.27	-39.27	<b>-40.87</b>		0.0	-11.65	-11.93	-14.41	<b>-15.87</b>	-35.27	-39.27	<b>-40.87</b>
SVR	RMSE [mm]	1.028	0.940	0.902	0.856	1.023	0.630	0.550*	<b>0.524*</b>	RMSE <sub>rel</sub> [%]	179.72	158.44	150.87	<b>144.76</b>	170.48	106.00*	95.51*	<b>92.06*</b>
	ΔRMSE <sub>rel</sub> [pp]	0.0	-21.28	-28.86	<b>-34.96</b>	-9.24	-73.72	-84.21	<b>-87.66</b>		0.0	-21.28	-28.86	<b>-34.96</b>	-9.24	-73.72	-84.21	<b>-87.66</b>
RVM	RMSE [mm]	0.522	0.475	0.463	<b>0.461</b>	0.466	0.439	0.429	<b>0.422</b>	RMSE <sub>rel</sub> [%]	91.44	84.78	82.17*	<b>81.88*</b>	83.43	77.62*	75.64*	<b>74.87*</b>
	ΔRMSE <sub>rel</sub> [pp]	0.0	-6.66	-9.27	<b>-9.56</b>	-8.01	-13.82	-15.80	<b>-16.56</b>		0.0	-6.66	-9.27	<b>-9.56</b>	-8.01	-13.82	-15.80	<b>-16.56</b>



**Figure 4.10:** Distribution of the  $RMSE_{rel}$  values for  $wLMS_{J_0}$ ,  $wLMS_{J_3}$ , SVR and RVM for the scenario of an unknown test set (TSUK). Each box indicates the 25<sup>th</sup> and 75<sup>th</sup> percentile of the data and the whiskers represent the 1.5 interquartile range. The residual points are classified as outliers and are marked by red crosses.



**Figure 4.11:** Distribution of the  $\Delta RMSE_{rel}$  values for  $wLMS_{J_0}$ ,  $wLMS_{J_3}$ , SVR and RVM for the scenario of a unknown test set (TSUK). Each box indicates the 25<sup>th</sup> and 75<sup>th</sup> percentile of the data and the whiskers represent the 1.5 interquartile range. The residual points are classified as outliers and are marked by red crosses.

**Table 4.4:** Occurrence of a feature within one of the feature sets  $S_1$  to  $S_3$  for M1. Results of **TSUK** are shown without parentheses and of **TSK** in parentheses. The sum of features along a row is  $n_{imp-FS}$  and  $n_{feat}$  along a column.

predictor	para	OM1	OM3	FLOW	STRAIN	ACC	$n_{imp-FS}$
nLMS	$S_1$	1 (3)	2 (2)	1 (2)	8 (11)	6 (0)	18 (18)
	$S_2$	5 (2)	3 (3)	4 (6)	0 (2)	1 (1)	13 (14)
	$S_3$	3 (0)	1 (0)	0 (1)	0 (2)	1 (2)	5 (5)
wLMS $_{J_0}$	$S_1$	0 (1)	0 (1)	0 (0)	13 (11)	4 (5)	17 (18)
	$S_2$	1 (1)	8 (7)	0 (0)	1 (3)	2 (2)	12 (13)
	$S_3$	0 (2)	0 (2)	0 (1)	2 (1)	6 (1)	8 (7)
wLMS $_{J_3}$	$S_1$	0 (1)	3 (3)	0 (0)	10 (10)	4 (3)	17 (17)
	$S_2$	0 (1)	7 (5)	0 (0)	3 (2)	1 (5)	11 (13)
	$S_3$	1 (0)	0 (1)	0 (0)	1 (0)	5 (3)	7 (4)
SVR	$S_1$	0 (1)	2 (2)	0 (2)	12 (7)	4 (6)	18 (18)
	$S_2$	0 (0)	12 (2)	0 (1)	0 (3)	3 (2)	15 (8)
	$S_3$	2 (0)	0 (1)	0 (2)	2 (0)	4 (0)	8 (3)
RVM	$S_1$	0 (1)	3 (2)	0 (0)	12 (11)	3 (4)	18 (18)
	$S_2$	1 (1)	10 (10)	1 (1)	2 (2)	4 (4)	18 (18)
	$S_3$	4 (2)	2 (2)	1 (2)	1 (2)	8 (7)	16 (15)
$n_{feat}$		18 (16)	53 (43)	7 (18)	67 (67)	56 (45)	

urement phase M2. The last column shows how often an improved feature set  $n_{imp-FS}$  could be found. The maximum of  $n_{imp-FS}$  is 18, which means that for all subjects an improved feature set could be found. It can be observed that  $n_{imp-FS}$  is close to the maximum for all algorithms in case of M1 and for feature set  $S_1$ . In case of feature set  $S_2$ , only the **RVM** algorithm has 18 feature sets (Table 4.4). Furthermore, the **RVM** algorithm has the highest number of improved features set  $S_3$  (16 for **TSUK** and 15 for **TSK**). Similar results can be observed for measurement phase M2. The last row in both tables represents the occurrence of a feature within one of the feature sets across all algorithms. Features of the OM3, strain, and **ACC** sensor were used most frequently.

Figure 4.12 illustrates, as an example, the prediction results of one motion fragment of measurement phase M2. The fragment contains three motion artefacts at about  $t = 213$  s,  $t = 227$  s, and  $t = 243$  s. The figure shows the prediction results of wLMS $_{J_0}$ , wLMS $_{J_0}$ , **SVR**, and **RVM** for  $S_0$  and  $S_3$ . Minor differences between  $S_0$  and  $S_3$  can be observed during the first and the last artefact. However, an improved prediction performance is

visible for all algorithms around and after the second artefact for  $S_3$ .

### 4.4.3 Discussion

In the first experiment the correlation between the external marker and the optical markers were investigated. The optical markers were selected as reference since they are currently used in clinical practise and can be interpreted as the “gold-standard”. The results indicate that the optical markers have, on average, the highest correlation followed by the strain, the flow, and the acceleration sensor. The high correlation results of the **OMs** could be expected. However, the results of M1 and M2 indicate a dependency on the marker position. The correlation between OM1 and OM3 is significantly lower (t-test,  $p < 0.01$ ) compared to the correlation between OM1 and OM2. This dependency on the marker position might be reduced by using either a respiratory belt or a flow sensor or by considering multiple external markers. These results are in agreement with the study of Ahn *et al.* [12]. The authors did report a general high correlation between internal and external motion of 0.77. Furthermore, Yan *et al.* reported a strong dependency on the marker placement and breathing pattern in [283].

The results for the correlation difference, shown in Fig. 4.8, reveal correlation variations of  $|\Delta r| > 0.2$  on average for all modalities even between minutes one and six. These differences are most likely due to altered breathing characteristics such as variations of the frequency or changes between abdominal and chest breathing. Further correlation differences can be observed for the comparison between regular and irregular breathing in Fig. 4.9. As a consequence, the relevance of one sensor to predict one of the optical markers might change over time. In an optimal setting a multivariate prediction algorithm should be adaptable to these changes over time. One practically relevant solution can be feature selection algorithms based on filter methods as discussed in sec. 4.2.2. Due to their low computational requirements, these method could be used in a real-time setting to find the currently most relevant sensors.

Even though not analysed in this experiment, the results motivate further research for multivariate correlation algorithms. Due to the large differences of the correlation coefficients between the external sensors, it can be assumed that also high variations can be found for the correlation between external and internal motion data.

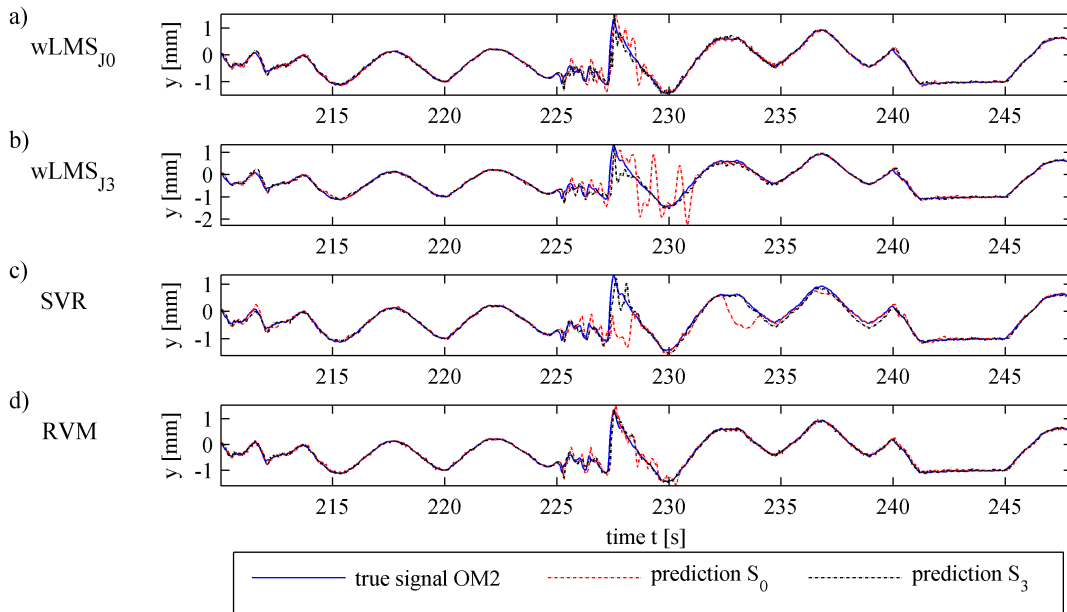
In the second experiment, the outcome of four motion prediction algorithms was evaluated for univariate and multivariate features. The results are shown in Table 4.3 and suggest that the **RMSE** can be decreased by using multivariate input features. An exception is the **nLMS** algorithm which seems to be prone to outliers (with  $\text{RMSE} > 10^4$  mm for some motion traces). The improvements are highlighted in Fig. 4.10 and Fig. 4.11,

**Table 4.5:** Occurrence of a feature within one of the feature sets  $S_1$  to  $S_3$  for M2. Results of **TSUK** are shown without parentheses and of **TSK** in parentheses. The sum of features along a row is  $n_{imp-FS}$  and  $n_{feat}$  along a column.

predictor	para	OM1	OM3	FLOW	STRAIN	ACC	$n_{imp-FS}$
nLMS	$S_1$	1 (2)	0 (2)	1 (5)	12 (6)	4 (3)	18 (18)
	$S_2$	2 (0)	2 (5)	1 (2)	1 (2)	0 (2)	6 (11)
	$S_3$	1 (0)	2 (2)	0 (2)	0 (2)	0 (1)	3 (7)
wLMS $_{J_0}$	$S_1$	0 (2)	2 (1)	0 (1)	12 (3)	3 (8)	17 (15)
	$S_2$	1 (0)	3 (1)	1 (4)	2 (5)	7 (1)	14 (11)
	$S_3$	1 (1)	5 (2)	0 (1)	1 (0)	3 (1)	10 (5)
wLMS $_{J_3}$	$S_1$	0 (1)	2 (1)	1 (3)	11 (6)	3 (6)	17 (17)
	$S_2$	2 (1)	6 (1)	0 (2)	2 (3)	4 (4)	14 (11)
	$S_3$	1 (0)	2 (0)	0 (1)	0 (0)	4 (2)	7 (3)
SVR	$S_1$	0 (0)	2 (3)	0 (6)	9 (7)	7 (1)	18 (17)
	$S_2$	0 (0)	10 (1)	1 (3)	3 (5)	2 (4)	16 (13)
	$S_3$	4 (1)	1 (1)	0 (2)	0 (1)	5 (2)	10 (7)
RVM	$S_1$	0 (0)	3 (0)	0 (0)	10 (12)	5 (5)	18 (17)
	$S_2$	1 (2)	7 (2)	1 (2)	4 (2)	5 (4)	18 (12)
	$S_3$	2 (2)	6 (2)	0 (1)	2 (0)	7 (2)	17 (7)
$n_{feat}$		16 (12)	53 (24)	6 (35)	69 (54)	59 (46)	

where it can be observed that the median  $RMSE_{rel}$  decreases and the mean  $\Delta RMSE_{rel}$  is negative by using one of the features sets  $S_1$  to  $S_3$ . Comparing the results between the complete feature set  $F$  and the improved feature sets, indicate that for the most prediction algorithms a feature selection method is required. In general, the mean  $RMSE_{rel}$  of  $F$  is higher compared to the  $RMSE_{rel}$  of  $S_1$  independently on the measurement phase and evaluation scenario. One exception can be observed for the **RVM** algorithm in measurement phase M1. The mean  $RMSE_{rel}$  of  $F$  is lower compared to  $S_1$  indicating that the prior assumption that the weights  $w$  are normally distributed with a zero mean (discussed in sec. 3.2.1) might prevents overfitting in case of normal breathing. However, these results cannot be confirmed in case of irregular breathing (M2).

Comparing the prediction accuracy of the algorithms with each other for all **FSs** reveals that the **RVM** algorithm has the lowest  $RMSE_{rel}$  followed by the **wLMS**, **SVR**, and **nLMS** algorithm. This order is in agreement with results of sec. 3.3 in case of univariate



**Figure 4.12:** Prediction results for one motion fragment of M2 with three breathing artefacts. The results of the (a)  $wLMS_{J_0}$ , (b)  $wLMS_{J_3}$ , (c)  $SVR$ , and (d)  $RVM$  algorithm are shown for  $S_0$  and  $S_3$ .

features. Including the outliers, the highest  $RMSE$  and  $RMSE_{rel}$  results were achieved by the  $nLMS$  algorithm for both measurement phases. The unstable behaviour of  $nLMS$  is most likely caused by using a common learning parameter  $\mu$  and by normalizing over the complete input vector  $\mathbf{x}_i^*$ . Consequently, the algorithm cannot be used without further modifications.

Comparing the results of the two  $wLMS$  algorithms points out that the mean  $RMSE_{rel}$  of  $wLMS_{J_0}$  is smaller than of  $wLMS_{J_3}$  for M1 and M2. These results indicate that a wavelet decomposition of  $J = 3$  might not be optimal for multivariate features. One explanation for this is that the dimension  $d$  of the features was fixed for each scale and modality. In [81], an adaptable feature dimension for each scale was proposed based on the energy of the decomposed signals. However, this approach was so far only evaluated on optical univariate data. Improvements might be achieved by using a feature dimension  $d_j^k$ , which depends on the sensor  $k$  and the wavelet scale  $j$ .

The  $SVR$  and the  $RVM$  algorithms are based on constraints which do result in either a sparse training or a sparse feature set. In case of  $SVR$ , the sparse set of training features are known as the support vectors. The  $RVM$  as proposed in sec. 3.2.1 leads to a set of sparse features which are known as the relevance vectors. This might enable

both algorithms to be more robust in terms of extended multivariate feature sets. As a consequence, the algorithms have high values of  $n_{imp-FS}$  for TSUK (Table 4.4 and 4.5). One practically relevant difference between both algorithms is the number of parameters. While the SVR algorithm depends on five parameters, the RVM approach has only two parameters (the residual parameters are hyperparameters which will be optimized automatically throughout the training process). Previous studies, such as [63, 80], have shown a high sensitivity to the initial choice of the penalty factor  $C$ , the tube width  $\varepsilon$  and the RBF kernel parameter  $\gamma$  in case of the SVR algorithm. As the signal properties might change over time (M1) or breathing artefacts occur (M2), the selected parameters might change. This can result in the appearance of large outliers as visible in Fig. 4.10 and Fig. 4.11. Furthermore, this might be one reason for the increase of the  $RMSE_{rel}$  results for TSUK from  $S_1$  to  $S_3$ . In case of M1 and TSUK, all parameters are learned on the first minute of each signal. The SVR algorithm finds a high number of FSs  $S_1$  to  $S_3$  (Table 4.4), meaning the  $RMSE_{rel}$  of the training set can be decreased. The mean  $RMSE_{rel}$  of the training set decreases from 40.4 % for  $S_1$  down to 37.71 % for  $S_3$  (data not shown). However, the parameters are not further adapted through the following 17 test minutes. Leading to the result that the mean  $RMSE_{rel}$  of the test set increases from 70.86 % for  $S_1$ , to 74.84 % for  $S_2$ , up to 89.61 % for  $S_3$ . In contrast the  $RMSE_{rel}$  decreases from 60.84 % for  $S_1$  down to 55.31 % for  $S_3$  in case of TSK. One solution would be time- and marker-dependent parameters, but this would further increase the computation time. As the SVR algorithm already has the longest training time compared to the other algorithms, a further increase might be unacceptable for real-time applications.

The RVM approach has the lowest  $RMSE$  and  $RMSE_{rel}$  and the highest  $n_{imp-FS}$  values for both measurement phases. Considering M1, the  $RMSE_{rel}$  is 60.51 % for the initial feature set  $S_0$ , which can be decreased to 47.9 % for  $S_3$ . Due to the prior distribution over  $w$ , the  $RMSE_{rel}$  can be decreased further by  $F$  to 47.64 %. These results are in agreement with [261]. There, the author showed that a higher sparsity can be achieved with the RVM compared to an the SVR algorithm. Furthermore, it has to be emphasized that the RVM is the only algorithm with a mean  $RMSE_{rel}$  below 100 % for  $S_0$  in case of M2. This means that in case of strong artefacts, the other prediction algorithms perform worse compared to no prediction. The advantage of the RVM is that most of the parameters are hyperparameters, which are optimized at each time index  $i$ . Consequently, a real time adaptation is performed automatically.

Some of the above drawn conclusions can be observed in the prediction results shown in Fig. 4.12. In case of the wLMS algorithm, the prediction oscillates around true observations  $y$  after the appearance of the second artefact for  $S_0$ . As the learning parameter  $\mu$  was



optimized on a regular breathing fragment, the value is not optimal for strong irregular motion patterns. The effect is compensated if multivariate features are considered. In case of the **SVR** and the **RVM** algorithm, the prediction accuracy depends on the current training set  $X$ . Using the feature set  $S_3$  instead of  $S_0$ , leads to an larger feature vector. This results in an increased prediction accuracy, as motion artefacts can be represented better within the feature space. The results indicate that a diverse representations of the same artefact, might improve the prediction accuracy.

The results were computed by using an **SFS** feature selection method. Beside the **SVR** algorithm, the resulting **FSs** have a lower mean **RMSE** and mean **RMSE<sub>rel</sub>** compared to the initial feature set  $S_0$  and the complete feature set  $F$ . The strongest increase in the prediction accuracy can be observed at the transition from  $S_0$  to  $S_1$ . Further minor improvements can be observed for  $S_2$  and  $S_3$ . Comparing the results between **TSUK** and **TSK** indicates that a further increase of the prediction accuracy is possible. The difference is especially visible for the artefact measurement M2. Assuming the complete test set is known (**TSK**), the lowest **RMSE<sub>rel</sub>** of the **RVM** algorithm is 74.87%. In contrast, assuming the realistic scenario that only the training data is known, the lowest **RMSE<sub>rel</sub>** is 81.88%. Improvements might be achieved by using filter methods such as the Akaike and Bayesian information criteria, or mutual information. In general, these methods could be used in a real-time setting to estimate the currently best feature set, as they are computational less expensive. As a consequence, the robustness of the algorithms could be increased in case of e.g. changes of the breathing pattern or increase of the measurement noise of one sensor.

Table 4.4 and Table 4.5 give the possibility to investigate how often each feature was used to extend the **FS**  $S_0$ . The last rows of both tables show the  $n_{feat}$ , which represent the occurrence of a particular feature in one of the **FSs** across all algorithms. The highest  $n_{feat}$  could be achieved by OM3, Strain, and **ACC** for **TSUK** and **TSK**. Intuitively, the results seem to be in conflict with the results of the first experiment, where OM1 and the flow sensor did have a high and the **ACC** sensor did have the lowest correlation. This might be caused due to redundant information of some of the features. The preference of extending  $S_0$  by OM3 instead of OM1 might be due to the higher correlation coefficients. However, after including the information of two optical markers, the remaining marker OM1 potentially contains very redundant information. As a consequence alternative features are added. For example, both markers, OM1 and OM3, were only used for three subjects in case of  $S_3$  for the **RVM** algorithm (M1 and **TSUK**). In contrast, only one of the markers was used for 14 subjects. Even though the data of the flow sensor has a high correlation with OM2, this feature was used the least. This might be due to the

used measurement hardware and the human physiology, which both cause dead times between the movement of OM2 and the measured flow. These dead times can be reduced by using a spirometer instead of a thermistor, which only measures the temperature difference. However, the patient would have to breath into a mouth piece, which would lead to a decreased patient comfort.

Comparing the features used for **TSUK** and **TSK** reveals that the relevance of the flow sensor is underestimated. In case of M2, the sensor is only used six times for **TSUK**, while it is used 35 times for **TSK**. Similar results can be observed for M2. In contrast, the use of the OM3 and the **ACC** sensor is overestimated. The biggest difference can be observed for OM3 in case of M2. If only the training set is known, OM3 is used 54 times. However in case of **TSK**, OM3 is only used 24 times. These differences and the resulting increased prediction error might be compensated by using a real-time feature selection method.

Finally, the high **RMSE** values of measurement phase M2 have to be highlighted. Even though M2 illustrates an extreme case with several motion artefacts, the mean **RMSE** for  $S_0$  is 1.03 mm for the **SVR**, 0.76 mm for the **wLMS<sub>J3</sub>**, and 0.52 mm for the **RVM** algorithm. These high errors are only caused by the prediction algorithms and can be further increased if for example the prediction result is used as input for a correlation algorithm. Using multivariate features, the errors can be decreased down to 0.52 mm for the **SVR**, 0.51 mm for the **wLMS<sub>J3</sub>**, and 0.42 mm for the **RVM** algorithm in case of **TSK**.

### 4.5 Further Respiratory Motion Features

As indicated in sec. 4.1.2, additional sensors have been used to measure the **EEG**, **EOG** and the **sEMG**. These sensors are normally not used in clinical practise for the purpose of respiratory motion estimation. The data of these sensors has in common that they require more complex pre-processing steps to extract relevant features, which might not be possible in real-time. The following investigations have been separated from the previous section as it can be interpreted as more basic research. In the current state, it seems unlikely that these features will play a role in clinical practise.

In this section, the basic ideas of **EEG**, **EOG**, and **sEMG** features are discussed. Furthermore, first results of correlation analyses and prediction results are presented. Further details about the pre-processing steps and initial results for all features can be found in [173].

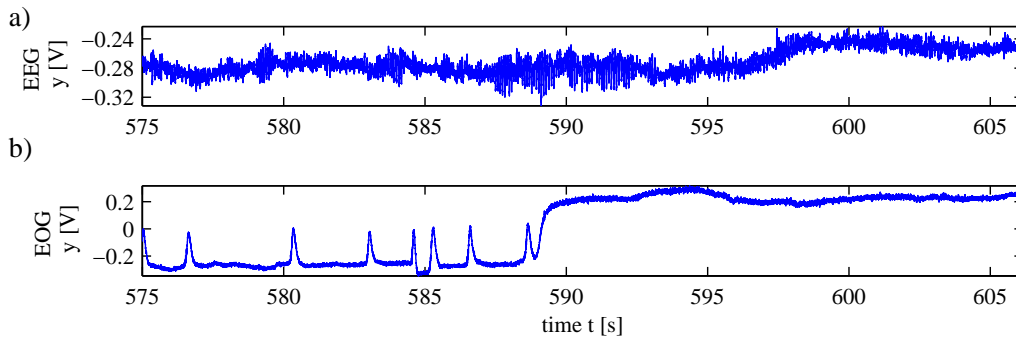


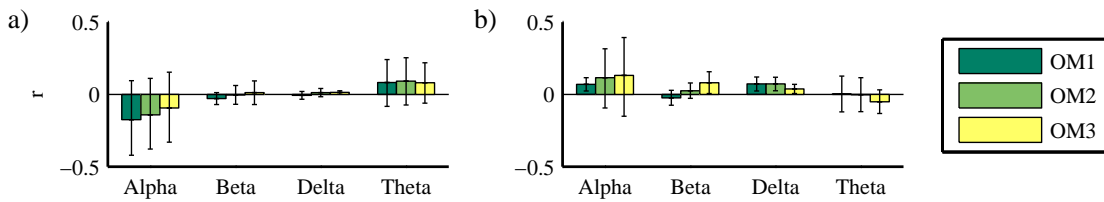
Figure 4.13: Recording of (a) a EOG and (b) a EEG signal.

### 4.5.1 EEG and EOG Features

The idea of measuring the EEG and the EOG of a subject is to extract features which indicate the level of alertness of the patients. In contrast to previous features, these features can be interpreted as “high-level” features. They do not contain information about the current position in the breathing cycle. In practise, patients will change their level of alertness throughout a treatment session. Most likely, the patients will be excited or even nervous at the start of a treatment session when they are left alone in a treatment room. During the treatment, a relaxation up till falling asleep can often be observed. These different stages will effect the breathing pattern. A feature indicating the transition from one specific level of alertness to another could potentially be useful to adapt the MC algorithm.

Figure 4.13.a shows an example of an EEG signal, which is computed based on the difference of the data recorded at position C3 and C4. The EEG signal is relatively noisy and further pre-processing steps are required. One possibility is the analysis of the EEG frequency spectrum. Within neuroscience, a separation into alpha ( $f \approx 8 - 13$  Hz), beta ( $f \approx 13 - 38$  Hz), gamma ( $f \approx 38 - 70$  Hz), delta ( $f \approx 0.5 - 4$  Hz), and theta ( $f \approx 4 - 8$  Hz) waves is typically performed [139]. A possible EEG feature could be the energy of the alpha wave. It is assumed that the energy of the alpha band correlates with the wake-sleep activity of a subject [139].

Figure 4.13.b shows the raw data of a typical EOG recoding. The plotted EOG signal is the difference between data of the electrode above the eye and data of the electrode below the eye. Before  $t < 589$  s, the eyes are open. Several eye blinks are visible, which are characterised by an increase of the amplitude by  $y \approx 300$  mV and a short duration of  $t \approx 100 - 400$  ms. For  $t > 589$  s, the eyes are closed, as the increased amplitude and missing blinks indicate. Furthermore, a slow oscillation of the base line can be often ob-



**Figure 4.14:** Mean correlation coefficient  $r$  between the power of different frequency bands of the EEG data and the mean frequency (a) and amplitude (b) per minute for OM1, OM2, and OM3 of measurement phase M1.

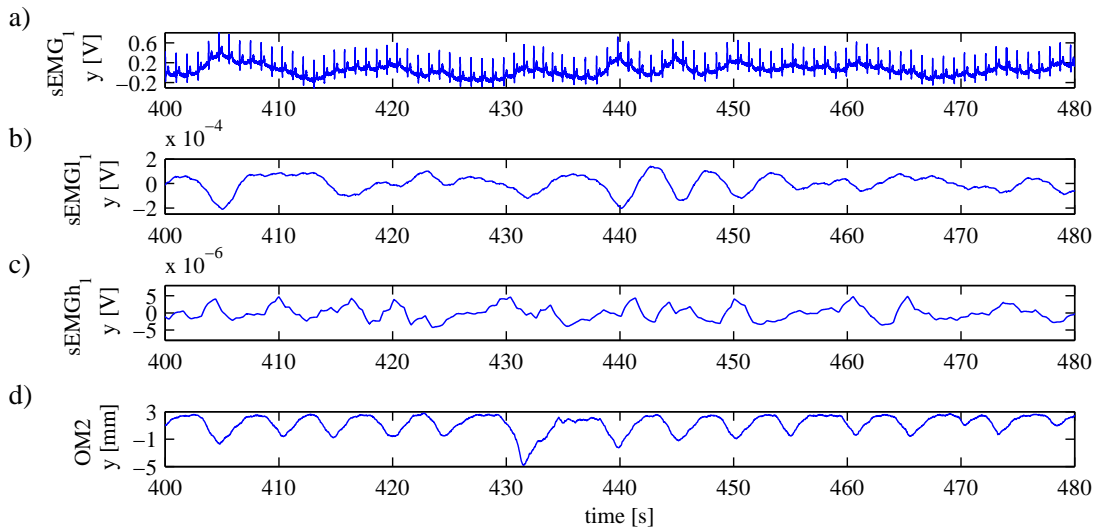
served, which represents normal small movements of the eye ball. Potential “high-level” EOG features could be based on the number of blinks per minute or if the eyes are open or closed.

To investigate the relevance of the EEG features, the correlation coefficients between the power of different frequency bands of the EEG data and the mean frequency and amplitude per minute of OM1 to OM3 were computed for measurement phase M1, respectively. The results, shown in Fig. 4.14, reveal on average a low correlation and a high standard deviation for all combinations. Consequently, these features were not further investigated. Similar results were found for the EOG features. Further details about the pre-processing steps and alternative evaluations, such as a correlation analysis differentiating between open and closed eyes, can be found in [173].

#### 4.5.2 sEMG Features<sup>4</sup>

As shown in Fig. 4.2 four electrodes were used to measure the activity of the surrounding surface and the diaphragm muscles. Ideally, only the muscle activity of the diaphragm should be measured as this is the main muscle for respiratory motion. However, a direct measurement is only possible with invasive needle electrodes. Additionally, it has to be considered that the diaphragm is next to the pleural gap, which could lead to a pneumothorax if damaged. To ensure a high patient safety, a non-invasive measurement of the EMG activity should be preferred. Maarsingh *et al.* [161] proposed a non-invasive sEMG technique to measure the diaphragm activity of children. This measurement technique was adapted for the purpose of respiratory motion prediction. A drawback of this technique is that the contribution of the diaphragm and the surface muscles to the measured activity is unclear. In the worst case only the activity of the surface muscles is measured. In contrast to the EEG and EOG features, the sEMG features could be used directly as

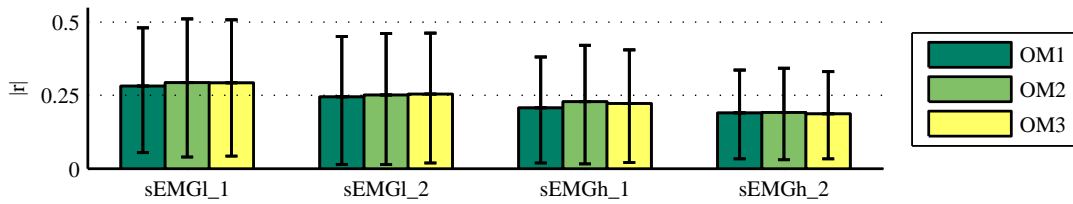
<sup>4</sup>Parts of this section have been published in [174]



**Figure 4.15:** Example of a **sEMG** recording for (a) the  $sEMG_1$  and (b) the computed features  $sEMG1_1$  and (c) the  $sEMGh_1$ ; (d) measured respiratory position of marker OM2 for comparison.

input of an **MC** algorithm similar to the features of the flow or acceleration sensor.

**Methods** Figure 4.15.a shows the measured raw data of one **sEMG** signal. We refer to this signal as  $sEMG_1$  as it is the difference between the data of the two electrodes which were placed bilaterally between the seventh and eighth rib.  $sEMG_2$  denotes the difference signal of the data of the two electrodes bilateral between the tenth and eleventh rib. The **sEMG** signals had to be further processed as they were influenced by the electrical potential of the heart. The sharp peaks in Fig. 4.15.a represent the QRS complex of the heart. Two features were considered to which we refer to as  $sEMG1$  and  $sEMGh$ .  $sEMG1$  was based on the low frequency components of the raw **sEMG** data and was constructed by applying an infinite impulse response filter with a passband of  $0 - 0.3$  Hz. The feature represents the slow baseline oscillations and was caused by the physical expansion of the chest and abdomen during inspiration. An example of the feature  $sEMG1_1$  is shown in Fig. 4.15.b. The second feature,  $sEMGh$ , covers the high frequency components of the **sEMG** raw data. The influence of QRS complex on the raw **sEMG** data was compensated retrospectively by removing the QRS complex ( $0.01$  s before the Q-spike and  $0.025$  s after the S-spike) [161]. The gap was replaced by the average of the data before and after the QRS complex. The low frequency of the **EMG** signal was eliminated and the resulting feature is the upper envelope of the absolute processed signal. Fig. 4.15.c shows an ex-



**Figure 4.16:** Mean absolute correlation coefficient  $|r|$  between the **sEMG** features and the OM1, OM2, and OM3 of measurement phase M1.

ample the **sEMGh<sub>1</sub>** feature. For comparison the signal of OM2 for the same time interval is shown in Fig. 4.15.d. Further details about the pre-processing steps and the results of an initial correlation analysis can be found in [173].

Note in the literature, the muscle activity is measured by investigating the frequency components around  $f \approx 80 - 500$  Hz. Consequently, only the **sEMGh** feature represents real muscle activity. The **sEMGI** feature is not a real **EMG** feature. However, it is named **sEMGI** as the feature is computed from low frequency components of the raw **sEMG** data.

**Experimental Setup** Only the data of the measurement phase M1 were considered for this initial investigation. Similar to sec. 4.3, the correlation coefficients of the **sEMG** features were computed with respect to the three optical marker OM1 to OM3. The first and last minute of each measurement was discarded to avoid irregular motion patterns due to initialisation and termination of the measurements. The signal was divided into 18 one minute segment. The correlation coefficient of each segment was computed. The mean correlation coefficients were computed over all subjects and segments.

The second experiment investigated if the  $RMSE_{rel}$  could be decreased by extending an initial feature set  $S_0$  by one of the **sEMG** features. The aim was to predict the second optical marker OM2. The initial feature set was defined as  $S_0 = \{OM2\}$  and extended by either **sEMGI<sub>1</sub>**, **sEMGI<sub>2</sub>**, **sEMGh<sub>1</sub>**, or **sEMGh<sub>2</sub>**. The first minute was defined as training set. Only the **RVM** algorithm was evaluated based on the results of sec. 4.4.2. The **RVM** parameters were initialised according to sec. 4.3.

**Results and Discussion** Fig. 4.16 shows the mean correlation coefficient  $|r|$  and standard deviation for all subjects and one-minute motion fragments. The highest mean correlation has the **sEMGI<sub>1</sub>** feature with respect to OM2 and OM3 with  $|r_{sEMGI_1, OM2}| = |r_{sEMGI_1, OM3}| = 0.29$ . On average, the results of the **sEMGh** features are lower compared to the **sEMGI** features. The electrodes placed next to rip seven and eight, which were

**Table 4.6:** Number of improvements  $n_{imp}$  and  $RMSE_{rel}$  of RVM for OM2 alone and in combination with one of the sEMG features

feature	OM2	OM2+ <i>sEMGl</i> <sub>1</sub>	OM2+ <i>sEMGl</i> <sub>2</sub>	OM2+ <i>sEMGh</i> <sub>1</sub>	OM2+ <i>sEMGh</i> <sub>2</sub>
$n_{imp}$	-	14	14	9	8
$RMSE_{rel}$					
<i>mean</i> [%]	60.5	58.2	58.8	60.3	60.2
<i>std</i> [%]	19.8	19.8	20	20.1	18.4

used to compute  $sEMGl_1$  and  $sEMGh_1$ , led to a higher correlation as the features of the electrodes next to the rip eleven and twelve. Only a small dependency on the optical marker can be observed. However, the sEMG features have a lower mean correlation and an increased standard deviation compared to the previously investigated external markers (Fig. 4.7).

The results of the second experiment are listed in Table 4.6. It can be observed that the mean  $RMSE_{rel}$  is smaller for all extended feature sets than for the feature OM2 alone. The  $RMSE_{rel}$  improvement is higher for the  $sEMGl$  features, which is in agreement with the increased correlation coefficients in Fig. 4.15. The lowest mean  $RMSE_{rel}$  could be achieved by the  $sEMGl_1$  feature with 58.2%. This feature led to an improved performance for  $n_{imp} = 14$  subjects.

These initial results indicate that sEMG features could potentially be used to increase the prediction accuracy. Even though the correlation is relatively low, the features seem to be less dependent on the position of the optical marker. However, it has to be considered that the features were computed retrospectively. Consequently, these features could not be used in a real-time setting. Further research has to be done to enable an on-line compensation of the QRS complex and to enhance the quality of the sEMG features. A starting point might be the pre-processing steps summarized in the EMG literature review of Hutten *et al.* [121].

## 4.6 Conclusion

This chapter focused on multivariate prediction algorithms. At first a measurement with human subjects was described, which was later on used to evaluate the proposed multivariate prediction algorithms. The measurement included, beside the optical marker,

also a respiration belt, a flow sensor, a acceleration sensor, and electrodes to measure the EEG, EOG, and sEMG. In sec. 4.2, the mathematical background was discussed to extend prediction algorithms to multivariate features as well as the relevance of feature selection methods.

To evaluate our approach the nLMS, wLMS, SVR, and RVM prediction algorithms were exemplarily selected. The aim was to predict OM2 depending on different feature sets. The features set corresponded to data of different sensors. Beside the initial FS  $S_0 = \{\text{OM2}\}$ , the complete FS, and six extended FSs, which were selected based on a features selection method. The results indicate that all algorithms can benefit from additional multivariate features. However, a feature selection method is required. On average, the prediction accuracy of  $F$  was lower than the accuracy of the FSs estimated by the feature selection method. The best performance could be achieved by the RVM algorithm. The results indicate that this algorithm is less effected by the “curse of dimensionality” due to its sparsity assumptions.

Finally, features were discussed which are not normally used to measure respiratory activity. Beside the EEG and EOG features which could be used to indicate different breathing states, first results of sEMG features were presented. Even though the correlation to the optical markers was relatively small, these features led to an decrease of the  $\text{RMSE}_{rel}$  for the RVM algorithm.

Focusing on the questions proposed in sec. 1.3, this chapter focused on the first main question Q.1. The following results can be summarized:

- **Q.1.1:** Which algorithms can be used to model multivariate data? Which general properties do algorithms need to have for multivariate prediction and correlation?

In principle all algorithms can be extended to use multivariate input features. However, algorithm belonging to the group of model-free algorithms are particularly interesting. These algorithm can incorporate additional multivariate features without knowing the exact physical relationship of the new features to already existing features. The majority of the discussed algorithms of chapter 2 can be categorized as model-free algorithms such as the approaches based on ANN and adaptive filters. Consequently, many additional algorithms could benefit from the results presented in this section.

- **Q.1.2:** How to select the most relevant and least redundant markers?



As discussed in sec. 4.2.2, feature selection methods can be utilized to find the most relevant and least redundant features. In the experiments described in sec. 4.4, an **SFS** method is proposed. The **FSs** computed by the **SFS** lead to a lower  $\text{RMSE}_{rel}$  compared to  $S_0$  and  $F$ . However, the **SFS** method is known to be computationally expensive and was only applied once to the training data. In an optimal setting the best **FS** is estimated in real-time, as the relevance and redundancy of certain features might change over time. Filter methods such as information criteria (e.g. mutual information) could be used for this. Referring back to the results of sec. 3.4, probabilistic algorithms offer another possibility. Either the predicted variance or the optimized **NLML** value could be used as a criterion to select the optimal **FS**. Similar to the hybrid approach, discussed in sec. 3.4.2, multiple **RVM** algorithms with different **FS** could be started in parallel. The algorithm with the lowest predicted variance (or **NLML** value) could be selected.

- **Q.1.3:** Can the accuracy of respiratory motion prediction be increased using a multivariate external sensor setup?

The results presented in sec. 4.4.2 point out that the prediction accuracy of probabilistic and non-probabilistic algorithms can be increased by using multivariate features. In case of M1, the mean **RMSE** of the **RVM** algorithm can be decreased by 20 % (from 0.15 mm for  $S_0$  to 0.12 mm for  $S_3$ ; **TSUK**). In case of the artefact measurement M2, the mean **RMSE** can be decreased by 12 % (from 0.52 mm for  $S_0$  to 0.42 mm for  $S_3$ ; **TSUK**). Even though M2 represents an extreme case, these high prediction errors emphasize the relevance of this research and motivate further investigations.

- **Q.1.5:** What are the most relevant sensors or sensor combinations?

Table 4.4 and 4.5 summarize how often data of one sensor was used as feature for a **FS**. Depending on the algorithm, different sensors were more relevant. On average the OM3, strain and **ACC** sensor were utilized most. The preference of the OM3, which was placed next to the navel, over the OM1, which was placed on the chest, most likely depends on the prediction target (which was the OM2 in the experiments). A respiratory belt to measure the strain as well as an acceleration sensor could be easily integrated into the current clinical setup. For example, a tight vest

is used to attach the optical markers in case of the CyberKnife<sup>®</sup>. An integration of these sensors into the vest is easily possible. However, the measurement electronics could potentially be in the radiation beam. It has to be further investigated if this would affect the measurement accuracy or the dose distribution received by the tumour. The advantage of a classical respiratory belt is that the measurement electronics can be placed anywhere along the cross-section of the torso, also at the back of the patient where normally less radiation beams are placed.

The least relevant sensor seems to be the flow sensor, which might be strongly effected by dead times due to the human physiology and the indirect measurement technique. Removing this sensor from the measurement setup would lead to an increased patient comfort, as the patient could breathe more normally.

# 5 Multi-Task Gaussian Process Models<sup>1</sup>

In chapter 3, two probabilistic algorithms, the **RVM** and the **GP** model, have been presented and evaluated as respiratory prediction algorithms on a univariate dataset. In chapter 4, only the first algorithm was evaluated for multivariate respiratory motion prediction, among other non-probabilistic approaches. It is more challenging to extend the second algorithm, the **GP** algorithm, to multivariate inputs, as this algorithm can be categorized as a model-based algorithm according to [76] (see sec. 4.2.1). In this chapter, one possible solution is discussed to which we refer as **MTGP** models. As it turns out, **MTGPs** might be an efficient possibility to combine prediction and correlation algorithms within one model. Consequently, this chapter focuses on using **MTGPs** for respiratory motion prediction and correlation. The **MTGP** framework was implemented as a Matlab open access toolbox<sup>2</sup>.

Note in this chapter, we refer to data of a signal also as a “task” as it is commonly done within the literature of **MTGP** models [30, 269].

The datasets used in this chapter are described in sec. 5.1. In sec. 5.2, the mathematical details to derive a **MTGP** model are presented. As this technique is relatively unknown in the field of biomedical engineering, several extensions of this framework are discussed and illustrated on synthetic examples in this section. The experiments performed on respiratory motion compensation are described in sec. 5.3. The results of the experiments are presented and discussed in sec. 5.4. In sec. 5.5, two extensions of the **MTGP** framework are discussed, which might be relevant for further research. A summary of the main results and conclusions of this chapter is presented in sec. 5.6.

## 5.1 Datasets

The experiments of this chapter are based on the data of two studies. The first study was presented in [83]<sup>3</sup>. We refer to this dataset as dataset I. It consists of data from a porcine study. Four gold fiducials were implanted in the liver of the subject. The movement

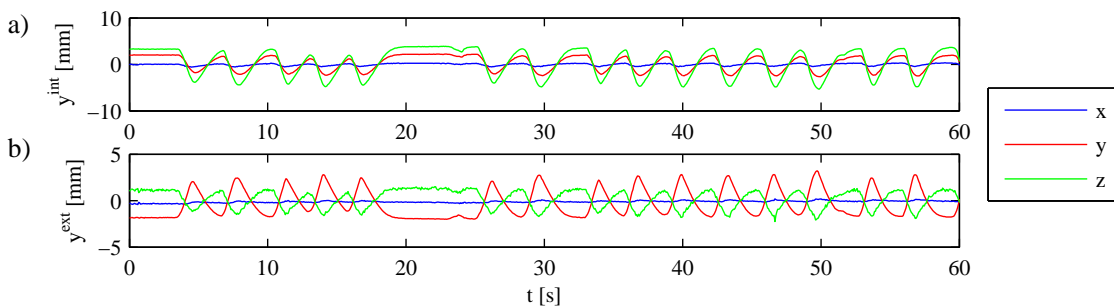
<sup>1</sup>Parts of this section have been published in [61, 70, 71]

<sup>2</sup>The **MTGP** toolbox and further examples are available online at <http://www.robots.ox.ac.uk/~davidc/>

<sup>3</sup>The dataset is available online at <http://signals.rob.uni-luebeck.de>.

of the fiducials was tracked via a two-plane **X-ray** imaging device (Philips Allura Xper FD20/10). The internal motion data was acquired at a sampling frequency of  $f_s = 15$  Hz. The mean amplitude of the four fiducials and the three spatial dimensions was 5.98 mm (min: 1.12 mm, max: 10.78 mm). The surface motion of the subject was recorded via six **OMs**. The data was acquired with an accuTrack 250 system (Atracsys LLC, Switzerland) equivalent to the set used in sec. 4.1. The sampling frequency was  $f_s = 216$  Hz. However, the data was downsampled to match the sampling frequency of the **X-ray** cameras. The recorded respiratory motion traces are very regular as the subject was ventilated manually. Here, only data of the first measurement was used which has a duration of  $t = 120$  s. An example recording for one internal fiducial and one external **OM** is shown in Fig. 5.1. Ernst *et al.* [83] evaluated the performance of various linear and quadratic correlation models as well as an approach based on **SVR**. The proposed **SVR** is based on the position, velocity, and acceleration information of one or multiple external **OMs**. The author reported a superior correlation accuracy of the **SVR** approach. Further details can be found in [76, 83].

The second study was already presented in detail in chapter 4 to which we refer to as dataset II in this chapter. In contrast to the previous chapter, we now also consider the internal motion which was acquired using **4D US**. Details about the data acquisition and pre-processing can be found in sec. 4.1. In contrast to dataset I, unambiguous landmarks were tracked. We refer to these internal targets as **ILM**. The continuous tracking of one unambiguous point within the **US** volumes is difficult. There are several sources for errors which might affect the tracking accuracy. First, **US** images have a high noise level compared to **X-ray** images. Second, the acquired volumes are susceptible to movement errors. This might be due to the **US** transducer losing contact to the skin surface or of rib movements in the field of view of the **US** transducer. Third, the selected **ILM** might



**Figure 5.1:** Example of a recorded motion fragment of dataset I for (a) fiducial one and (b) optical marker three for all spatial directions.

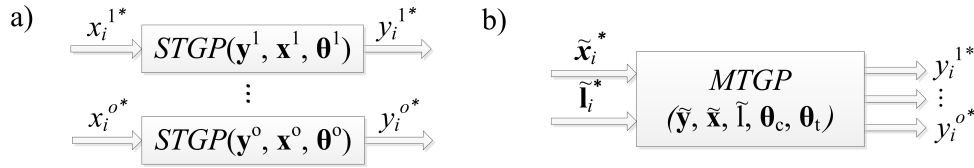
underlie deformations during the respiratory cycle. Fourth, template matching based on the sum of squared distances was used to track the **ILM**. The tracking accuracy depends on several parameters such as the template size or the search distance. As a result, visual inspection of the tracked motion traces revealed that only a subset of the acquired internal motion traces could be used for further experiments. The data of this study was reduced to nine motion traces of measurement phase M1. The data of measurement phase M2 was completely ignored. The artificial motion artefacts of M2 did cause several discontinuities in the tracked internal motion pattern, which indicate that the tracking algorithm jumped between different internal points. The remaining nine motion traces (five females, four males) had an average duration of 17.69 min (min: 15.10 min, max: 19.12 min). The data of the external sensor was downsampled to the sampling rate of the **US**, which was  $f_s = 17$  Hz. The data of the external sensor was preprocessed as described in sec. 4.1.3. The mean amplitude of the internal landmarks was 24.07 mm (min: 13.7 mm, max: 30.47 mm).

## 5.2 Mathematical Background

Within the literature, **GP** models are often used for regression problems such as to model a single time-series. We refer to these models as **single-task Gaussian process (STGP)** models. Examples from the field of biomedical engineering are the vital-sign “early warning system” of Clifton *et al.* [49] or the regression of noisy heart rate data of Stegle *et al.* [253].

In this section, the mathematical background of **MTGP** models is presented. **MTGPs** are multi-output models, which can be used to model multiple time-series (multiple tasks) simultaneously within one model. Thereby, the model learns the correlation between and within tasks. By learning the correlation, the aim is to improve the overall prediction accuracy compared to multiple individual **STGP** models. One advantage of **MTGPs** is that unevenly-sampled time-series can be incorporated in the same model. The method does not require further downsampling or interpolation which might be sources of errors.

In sec. 5.2.1, the mathematical background is presented to extend normal **GP** or **STGP** model (as presented in sec. 3.2.2) to a **MTGP** model. Section 5.2.2 presents a new transformation to normalize the **MTGP** correlation coefficients which makes a comparison to the Pearson’s correlation coefficient possible. In sec. 5.2.3, **MTGP** models are extended to incorporate time latencies between time-series. Section 5.2.4 outlines three **MTGP** tracking algorithms which can be used as combined approaches for correlation and prediction.



**Figure 5.2:** Schematic diagram of (a) multiple single-task Gaussian Processes (STGP) models and (b) one multi-task Gaussian process (MTGP) model to learn  $o$  tasks.

In sec. 5.2.5, alternative algorithms are presented which could be used for prediction and correlation.

The sections 5.2.1 to 5.2.3 contain several examples on synthetic datasets to illustrate the basic properties of the MTGP method. The examples were evaluated with respect to the RMSE (Eq. 2.8) and nRMSE (Eq. 2.9). Furthermore, a probabilistic measure was considered. The prediction outcome at  $t_i$  of a GP model is a normal distribution  $\mathcal{N}(y_i^*, \sigma_i^*)$  with a mean  $y_i^*$  and a variance  $\sigma_i^{2*}$  according to Eq. 3.35 and 3.36. As a consequence, the probability of the true observations  $y_i$  under the model can be evaluated. In [208], the mean standardized log loss (MSLL) is defined as

$$MSLL = \frac{1}{N} \sum_{i=1}^N \left( -\log p(y_i | f(\mathbf{x}_i), x_i^*) + \log p(y_i | \bar{\mathbf{y}}_i, \text{var}(\mathbf{y}_i), x_i^*) \right). \quad (5.1)$$

The first term represents the probability of  $y_i$  given our model  $f(\mathbf{x}_i)$  and the test label  $x_i^*$ . The second term is a normalization term, which represent the probability of  $y_i$  under a trivial model. Here, the trivial model is assumed to be a Gaussian distribution with mean  $\bar{\mathbf{y}}_i$  and variance  $\text{var}(\mathbf{y}_i)$  of the training labels  $\mathbf{y}_i$ . In case of more complex functions with low prediction error, the MSLL will be negative. For simple functions which are close to the mean value of the training labels, the MSLL value will be close to zero.

### 5.2.1 From GPs to Multi-Task GPs

MTGP models are motivated by the problem of modelling  $o$  tasks simultaneously. One example might be the respiratory motion acquired by multiple optical markers. A naïve approach might be to learn  $o$  individual STGP models, which is illustrated in Fig. 5.2.a. However, these models would ignore potential correlation between the tasks.

If the tasks share a common set of input features such as the time  $t$ , all tasks can be simulated within one MTGP model. The GP algorithm presented in sec. 3.2.2 is based on one-dimensional input features  $x_i = t_i$ . The vector of all training features was defined in

sec. 3.2.2 as  $\mathbf{x}_i = [x_i, \dots, x_{i-m+1}]^\top$ . In case of **MTGPs** with  $o$  tasks, we want to generalize this to  $\tilde{\mathbf{x}}_i = [(\mathbf{x}_i^1)^\top, \dots, (\mathbf{x}_i^o)^\top]^\top$ . The superscript labels indicate the affiliation of feature vector  $\mathbf{x}_i^j$  to task  $j$ . Furthermore, it is assumed that each task  $j$  has a task-specific number of training pairs  $m_j$  which are sampled at task-specific time indices  $q_i^j$ . This leads to  $\mathbf{x}_i^j = [x_1^j, \dots, x_{m_j}^j]^\top$  with  $x_i^j = t_{q_i^j}$ . The training labels are defined as  $\tilde{\mathbf{y}}_i = [(\mathbf{y}_i^1)^\top, \dots, (\mathbf{y}_i^o)^\top]^\top$  with  $\mathbf{y}_i^j = [y_{q_1^j}^j, \dots, y_{q_{m_j}^j}^j]^\top$ . Additionally, a label vector  $l_i^j$  is required to specify the affiliation of  $x_i^j$  and  $y_{q_i^j}^j$  to task  $j$ . The label is defined as  $l_i^j = j$ . The vector of training labels is defined as  $\tilde{\mathbf{l}}_i = [(\mathbf{l}_i^1)^\top, \dots, (\mathbf{l}_i^o)^\top]^\top$  with  $\mathbf{l}_i^j = [l_1^j, \dots, l_{m_j}^j]^\top$ . The vectors  $\tilde{\mathbf{x}}_i$ ,  $\tilde{\mathbf{y}}_i$ , and  $\tilde{\mathbf{l}}_i$  have a size of  $\tilde{m} \times 1$  with  $\tilde{m} = \sum_{j=1}^o m_j$ . A **MTGP** model is illustrated in Fig. 5.2.b.

Note, the difference to the multivariate algorithms discussed in chapter 4, such as the **RVM** algorithm, is that data from additional sensors was incorporated by increasing the dimension of the feature space  $\mathbf{x}$  (Eq. 4.4). In case of **MTGP**, the dimension of the feature space remains the same compared to an **STGP** ( $d = 1$ ). However, the number of training pairs is increased to  $\tilde{m}$ . By doing so, **MTGP** models gain flexibility with respect to input features. Data of different sensors do not need to be acquired at a common sampling frequency. Further, if data of one sensor is missing, the data of the residual sensors can still be used.

The extension from an **STGP** to a **MTGP** model takes place within the specification of the covariance functions. The residual assumptions and equations such as for the predicted mean (Eq. 3.35) and variance (Eq. 3.36) remain unchanged. As discussed in sec. 3.2.2.1, complex covariance functions can be designed by summation, multiplication, and convolution of individual covariance functions. It might be assumed that two covariance function are defined as

$$k_{\text{MTGP}}(x, x', l, l') = k_c(l, l') \times k_t(x, x') \quad (5.2)$$

where  $k_c$  is a covariance function representing the correlation between tasks and  $k_t$  is a function representing the correlation within tasks. We refer to the latter as temporal covariance function. Examples of  $k_t$  are the covariance functions discussed in sec. 3.2.2.1, such as the squared-exponential function (Eq. 3.37). Note that  $k_t$  only depends on the indices  $x$  and  $k_c$  only on the labels  $l$ . This means that all tasks are modelled with the same set of temporal hyperparameters  $\boldsymbol{\theta}_t$ .

To simplify the appearance of the next equations, we assume that  $m^j = m$  and  $q_i^j = q_i$  for  $j = 1, \dots, o$  and  $\tilde{m} = om$ . However, the **MTGP** framework is not restricted to this. The covariance matrix of the training features can be written as

$$\mathbf{K}_{\text{MTGP}}(\tilde{\mathbf{x}}, \tilde{\mathbf{l}}, \boldsymbol{\theta}_c, \boldsymbol{\theta}_t) = \mathbf{K}_c(\tilde{\mathbf{l}}, \boldsymbol{\theta}_c) \otimes \mathbf{K}_t(\tilde{\mathbf{x}}, \boldsymbol{\theta}_t) \quad (5.3)$$

where  $\otimes$  is the Kronecker product, and  $\theta_c$  and  $\theta_t$  are vectors containing hyperparameters for  $\mathbf{K}_c$  and  $\mathbf{K}_t$ , respectively. The matrix  $\mathbf{K}_c$  is the correlation matrix between all tasks and has a size of  $o \times o$ . The matrix element  $\mathbf{K}_c[a, b]$  can be interpreted as the correlation between the task  $a$  and  $b$ . Consequently, the diagonal elements represents the correlation of the tasks with themselves and the off-diagonal elements correspond to the correlation between tasks with  $\mathbf{K}_c[a, b] = \mathbf{K}_c[b, a]$ . Assuming  $\mathbf{K}_c = \mathbf{I}$  being the identity matrix, the tasks would be modelled without considering the correlation between them. The matrix  $\mathbf{K}_t$  describes the correlation between features and has a size of  $m \times m$ . The Kronecker product is defined as

$$A \otimes B = C \tag{5.4}$$

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \otimes \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} a_1 b_1 & a_1 b_2 & a_2 b_1 & a_2 b_2 \\ a_1 b_3 & a_1 b_4 & a_2 b_3 & a_2 b_4 \\ a_3 b_1 & a_3 b_2 & a_4 b_1 & a_4 b_2 \\ a_3 b_3 & a_3 b_4 & a_4 b_3 & a_4 b_4 \end{bmatrix}$$

Each element of matrix  $A$  is multiplied with the complete matrix  $B$ . The resulting matrix  $\mathbf{K}_{\text{MTGP}}$  has the size  $\tilde{m} \times \tilde{m}$ .

The remaining problem is the parametrization of matrix  $\mathbf{K}_c$ . In order to make  $\mathbf{K}_{\text{MTGP}}$  a valid covariance matrix, it has to be guaranteed that the covariance matrix  $\mathbf{K}_c$  is positive semi-definite (Mercers's theorem). Different parametrizations of  $\mathbf{K}_c$  have been proposed in the literature [30, 195, 248]. Bonilla *et al.* [30] presented a so-called "free-form" parametrisation, as it allows arbitrary correlations between the tasks. It is based on the Cholesky decomposition. The hyperparameters  $\theta_c$  specify the elements of the lower triangular matrix  $\mathbf{L}$  as

$$\mathbf{K}_c = \mathbf{L}\mathbf{L}^\top, \quad \mathbf{L} = \begin{bmatrix} \theta_{c,1} & 0 & \dots & 0 \\ \theta_{c,2} & \theta_{c,3} & & 0 \\ \vdots & & \ddots & \vdots \\ \theta_{c,k-o+1} & \theta_{c,k-o+2} & \dots & \theta_{c,k} \end{bmatrix} \tag{5.5}$$

where the number of correlation hyperparameters  $\theta_c$  is  $k = o(o+1)/2$ . By multiplication of  $\mathbf{L}\mathbf{L}^\top$ , the matrix  $\mathbf{K}_c$  is guaranteed to be positive semi-definite. It can be observed that the elements of  $\mathbf{K}_c$  are not restricted to take values between  $[-1, 1]$  as, for example, in case of Pearson's correlation coefficients. As a consequence, the diagonal elements of  $\mathbf{K}_c$  can be used as individual  $y$ -scaling parameters for each task. If the diagonal element of  $\mathbf{K}_c$  are one, the data of all tasks is restricted to share the  $y$ -scaling due to the common temporal hyperparameters  $\theta_t$  in  $\mathbf{K}_t$ . One drawback of the "free-form" parametrisation is



that the interpretability of the correlation values decreases. A direct comparison to, e.g., Pearson’s correlation coefficient is not possible as the values are affected by the  $y$ -scaling of each task.

Osborne *et al.* [195] proposed a parametrization based on a spherical decomposition. The parametrization compensates the drawback of the “free-form” parametrization. The correlation values are forced to be within  $[-1, 1]$ . However, it requires equally scaled observations for each task or additional scaling hyperparameters.

The output of a test label  $y_i^{j*}$  for task  $j$  can be computed based on the test features  $\{x_i^{j*}, l_i^{j*}\}$  with  $x_i^{j*} = t_i$  and  $l_i^{j*} = j$ . Equivalently to sec. 3.2.2.1, the hyperparameters  $\theta_c$  and  $\theta_t$  can be optimized by minimizing the NLML.

The MTGP framework has several useful properties such as:

- utilization of task-specific training indices  $m^j$  and time indices  $q_i^j$  (e.g., each task could be sampled at a different sampling frequency);
- similar to STGP, arbitrary prediction latencies can be specified, independent from the sampling frequency;
- the correlation between tasks is learned automatically as it is another hyperparameter which can be optimized; and
- it is assumed that the tasks share similar temporal characteristics and hyperparameters  $\theta_t$ .

Note, the last property might be useful in case of respiratory motion prediction. However, this assumption might be too strong for many other applications. In sec. 5.5.2, an approach to overcome this restriction is presented.

One drawback of the MTGP approach are the increased computational requirements. In terms of computation time, the most costly operation is the inversion of the matrix  $[\mathbf{K}(\mathbf{x}_i, \mathbf{x}_i) + \sigma^2 \mathbf{I}]$  (Eq. 3.35). The computational requirement is  $o \times \mathcal{O}(m^3)$  in case of  $o$  individual STGP models. For an MTGP model, it increases to  $\mathcal{O}(o^3 m^3)$ . Furthermore, the increased number of hyperparameters makes a gradient descent optimization more susceptible to local minima. Alternative approaches to optimize the hyperparameters or sparse Gaussian process models [199, 204, 271] might be required for an increased number of tasks and/or training data.

**MTGP Example I.a** The first example illustrates the scenario of modelling multiple tasks ( $o = 4$ ) with MTGPs and how learning the correlation between tasks can improve the prediction outcome. A 70 s long motion fragment of dataset II was considered. The

**Table 5.1:** Time interval  $d^j$  enclosing the training data for the  $j^{\text{th}}$  task, sampling frequency  $f_s$ , and Pearson's correlation coefficient  $r$  with respect to OM1

	OM1	OM2	OM3	Strain
$d^j$ [s]	(0, 20)	(10, 30)	(25, 40)	(0, 60)
$f_s$ [Hz]	2.6	2.6	2.6	0.52
$r$	1	-0.96	-0.9	0.89

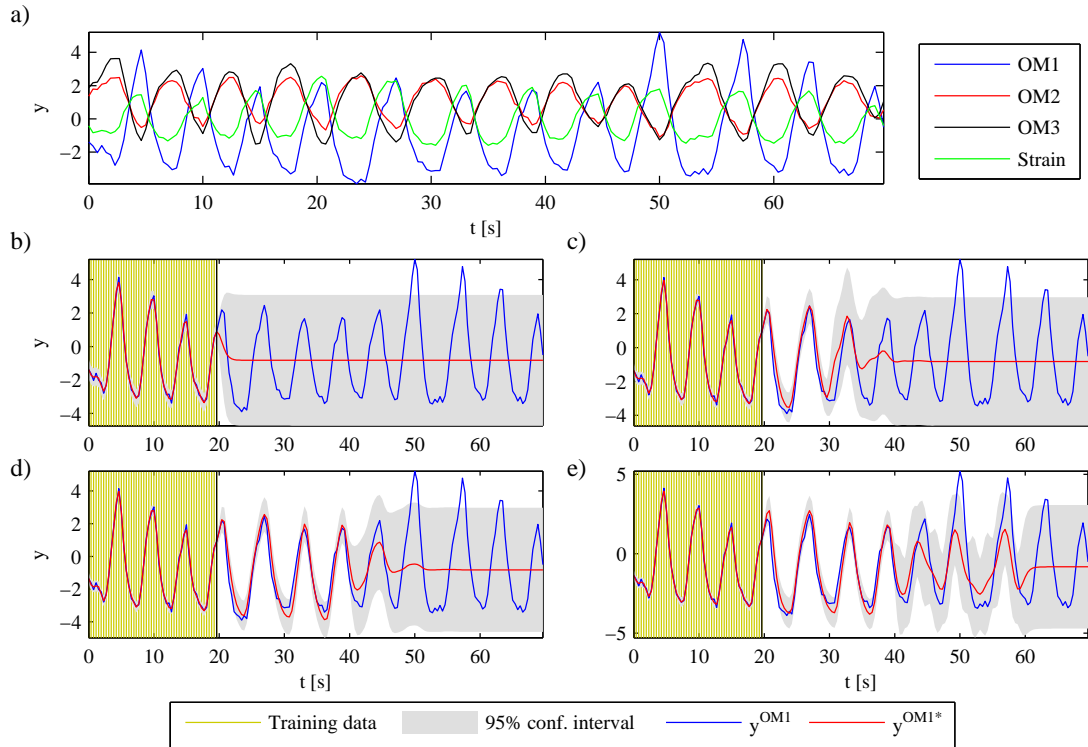
data of the three **OMs** (OM1 - 3) and the strain belt were adapted for this example. We assumed that the data of the **OMs** and strain sensor were acquired at a different sampling frequency and that different training intervals  $d^j$  were known. Table 5.1 lists acquisition parameters for each task. It can be observed that the sampling frequency of the **OMs** was five times higher than that of the strain sensor. OM1 and OM3 had no overlapping training interval. Furthermore, Pearson's correlation coefficients with respect to OM1 are listed (based on the complete signals). They were computed on the complete motion fragments and indicate a high positive or negative correlation. The four motion fragments are shown in Fig. 5.3.a.

The aim of this experiment was to predict  $y^{OM1*}$  for task OM1 with the test range being  $x^{OM1*} \in (20\text{ s}, 70\text{ s}]$ . Four evaluation scenarios were considered. The first scenario (S1) assumed that only data of OM1 was known. This case is equivalent to an **STGP**. The training data of the OM2, the OM3, and the strain sensor were added successively into the **MTGP** model for scenarios S2 to S4. All tasks were considered in S4. A quasi-periodic covariance function (Eq. 3.42) was selected as temporal covariance function. The correlation hyperparameters were initialised assuming independent tasks.

The prediction results for  $y^{OM1*}$  and the 95% confidence intervals for S1 to S4 are shown in Fig. 5.3.b-e, respectively. The 95% confidence intervals is defined as two times

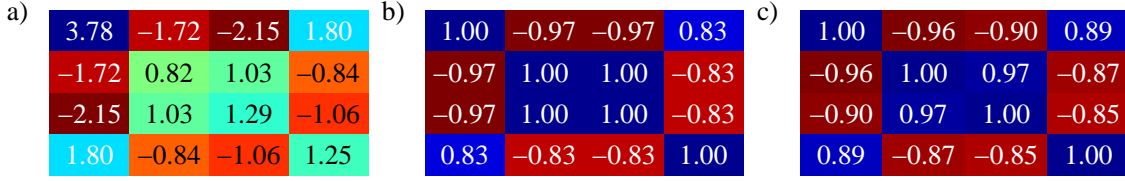
**Table 5.2:** **RMSE**, **nRMSE** and **MSLL** of  $y^{OM1*}$  for prediction scenarios S1 to S4 of **MTGP** experiment I.a

	S1	S2	S3	S4
<b>RMSE</b>	2.244	2.005	1.805	1.474
<b>nRMSE</b>	0.987	0.881	0.793	0.648
<b>MSLL</b>	0.002	-0.278	-0.527	-0.669



**Figure 5.3:** (a) Motion fragments of three optical markers (OM1 - 3) and one respiration belt (Strain); (b) - (e) Predicted position  $y^{OM1*}$  and confidence interval for OM1 of scenarios S1 to S4, respectively.

the square root of the predicted variance  $\sigma^{2*}$ . The training data are highlighted by vertical yellow lines. In case of S1, the difference between  $y^{OM1*}$  and  $y^{OM1}$  is small within the training region ( $t < 20$  s) as these observations were part of the training set. For  $t > 20$  s,  $y^{OM1*}$  moves towards the mean function, which is the mean of the training observations. Consequently, the variance and the prediction error increase. As the distance  $r$  between the test label and the training labels increases, the correlation  $k_t$  decreases and the model predicts the prior distribution. If we consider additional tasks (S2 to S4), it can be observed that the prediction accuracy increases. This observation is confirmed by the **RMSE** and **nRMSE** in Table 5.2. The **RMSE** decreases from 2.244 for S1 to 1.474 for S4. Figure 5.3.d and the **RMSE** value for S3 indicate that data of OM3 can be used to improve the prediction accuracy compared to S2. This shows that the **MTGP** model is able to learn the correlation between OM1 and OM3 even though they do not share an overlapping training region. The correlation is learned via the training data of OM2. Scenario S4 illustrates that also data acquired at a different sampling frequency can be used in



**Figure 5.4:** (a) Correlation matrix  $\mathbf{K}_c$  without normalization; (b) normalized matrix  $\mathbf{K}'_c$  and (c) Pearson's correlation coefficients for scenario S4 of *MTGP* example I.

the *MTGP* model to increase the prediction accuracy. These observations are confirmed by the *MSLL* values. In case of S1, the *MSLL* is 0.002 indicating that the probability to predict the observations  $y^{OM1}$  of the *MTGP* is comparable to the probability of assuming a normal distribution with mean and variance based on the training data for OM1. As the prediction error decreases for scenarios S2 to S4, the probability of predicting  $y^{OM1}$  increases, resulting in decreased *MSLL* values.

## 5.2.2 Correlation Matrix and Normalisation

One disadvantage of the “free-form” parametrisation is that the correlation coefficients are influenced by the task-specific scaling of the observations  $y^j$  [30]. This influence can be compensated by performing a normalization which allows for a direct comparison of the *MTGP* correlation values to Pearson's correlation coefficient. We refer to the normalized correlation hyperparameters as  $\theta'_c$ . By substitution of  $\theta_c$  through  $\theta'_c$ , the normalized correlation matrix  $\mathbf{K}'_c$  can be computed based on Eq. 5.5.

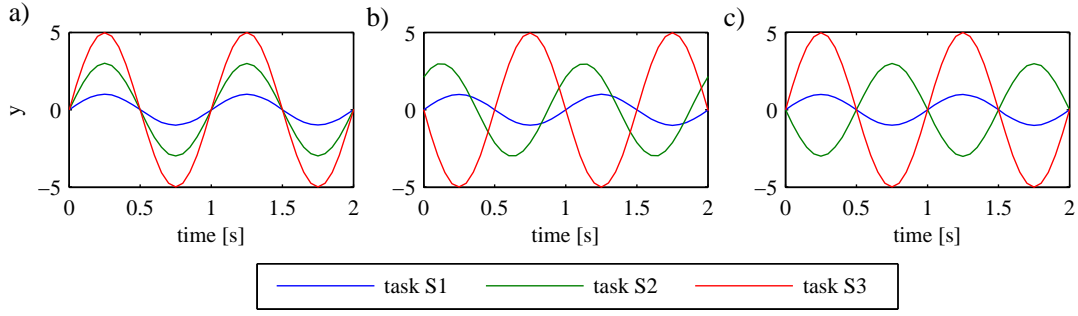
The proposed transformation is based on two assumptions. First, the influence of the task-specific scaling is eliminated through the constraint

$$\text{diag}(\mathbf{K}'_c)[l] = 1, \quad (5.6)$$

with  $\text{diag}(\mathbf{K}'_c)[l]$  being the diagonal elements of  $\mathbf{K}'_c$  for  $l = 1, \dots, o$ . Second, it is assumed that the contribution of the hyperparameters  $\theta_c$  to the diagonal elements of  $\mathbf{K}_c$  is equivalent to that of  $\theta'_c$  to  $\mathbf{K}'_c$ . These assumptions lead to the following normalization of  $\theta'_c$

$$\theta'_c[l] = \text{sgn}(\theta_c[l]) \sqrt{\frac{\theta_c[l]^2}{\sum_{k=\delta_1+1}^{\delta_2} \theta_c[k]^2}} \quad (5.7)$$

with  $\delta_1 = j(j-1)/2$ ,  $\delta_2 = j(j+1)/2$ ,  $\delta_1 < l \leq \delta_2$  and  $j \in \{1, \dots, o\}$  with  $o$  being the number of tasks.



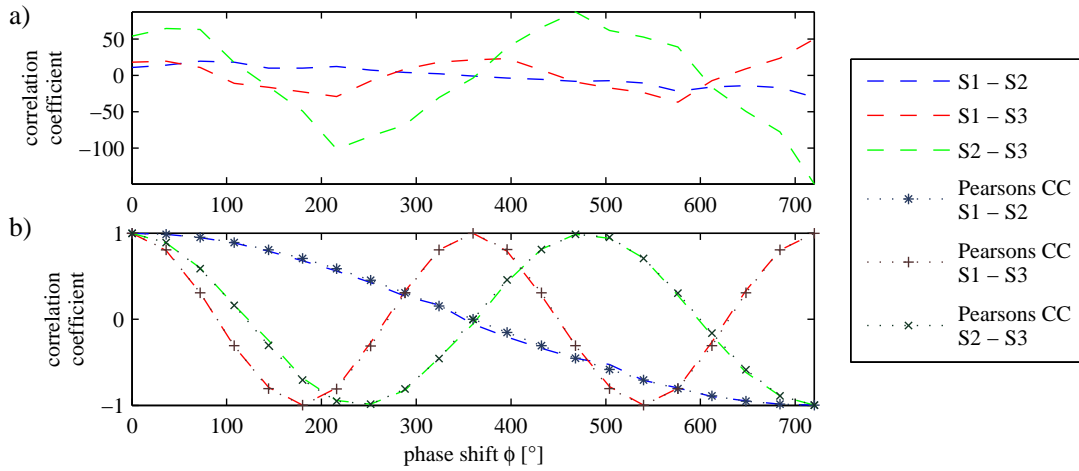
**Figure 5.5:** Illustration of the three tasks for (a)  $\phi = 0$ , (b)  $\phi = \pi$ , and (c)  $\phi = 4\pi$ .

**MTGP Example I.b** The example of the previous section will be considered to highlight the effect of the normalization. Applying Eq. 5.7 leads to the following normalized hyperparameters

$$\begin{aligned} \theta'_c[1] &= \text{sgn}(\theta_c[1]) \sqrt{\frac{\theta_c[1]^2}{\sum_{k=1}^1 \theta_c[k]^2}} = 1, & \theta'_c[2] &= \text{sgn}(\theta_c[2]) \sqrt{\frac{\theta_c[2]^2}{\sum_{k=2}^3 \theta_c[k]^2}}, \\ \theta'_c[3] &= \text{sgn}(\theta_c[3]) \sqrt{\frac{\theta_c[3]^2}{\sum_{k=2}^3 \theta_c[k]^2}}, & \theta'_c[4] &= \text{sgn}(\theta_c[4]) \sqrt{\frac{\theta_c[4]^2}{\sum_{k=4}^6 \theta_c[k]^2}}, \\ & \vdots & & \vdots \\ \theta'_c[9] &= \text{sgn}(\theta_c[9]) \sqrt{\frac{\theta_c[9]^2}{\sum_{k=7}^{10} \theta_c[k]^2}}, & \theta'_c[10] &= \text{sgn}(\theta_c[10]) \sqrt{\frac{\theta_c[10]^2}{\sum_{k=7}^{10} \theta_c[k]^2}}. \end{aligned}$$

Figure 5.4.a shows the correlation matrix  $\mathbf{K}_c$  of scenario S4 after optimization. The correlation values are within  $-2.15$  and  $3.78$ . Applying normalization, the correlation values are within  $-1$  and  $1$  as indicated in Fig. 5.4.b. Figure 5.4.c displays the Pearson's correlation coefficient in the same manner as the MTGP correlation coefficients. Minor differences between the normalized MTGP and Pearson's correlation coefficients can be observed. The highest difference is  $0.07$  for the correlation between OM1 and OM3. These differences were expected as the MTGP correlation coefficients were computed based on the training data and the Pearson's correlation coefficient on the complete motion fragments. Note the computation of the Pearson's correlation coefficient between OM1 and OM3 only on the training data would not be possible as OM1 and OM3 have no overlapping training region.

**MTGP Example II** Three sinusoidal signals were generated to further validate the proposed normalization. The sinusoidal signals had a common period of  $1$  s and differed in their amplitudes. We refer to the signals as tasks S1 to S3. To investigate the correlation,



**Figure 5.6:** (a) **MTGP** correlation coefficients without normalization; (b) normalized **MTGP** and Pearson's correlation coefficients depending on the phase shift  $\phi$ .

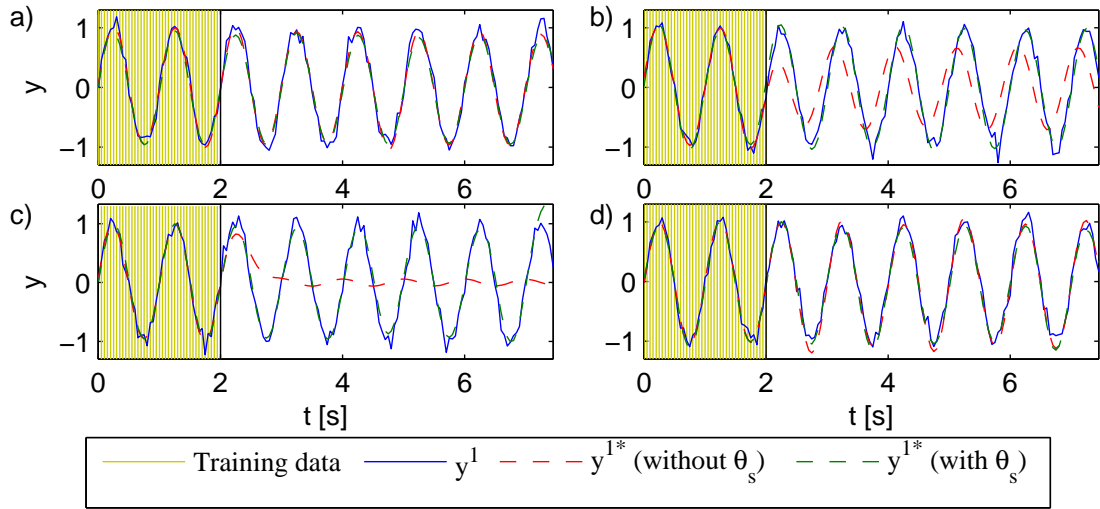
tasks S2 and S3 were shifted by  $\phi/4$  and  $\phi$ , respectively, with  $\phi \in \{0, \pi/2, \pi, \dots, 4\pi\}$ . Figure 5.5 illustrates the three tasks for different phase shifts. An **MTGP** model was trained for each  $\phi$  with a squared-exponential covariance function. It was assumed that all points were given as training points for the **MTGP** models. All models were initialized assuming no correlation between the tasks.

The **MTGP** correlation coefficients between the three tasks without normalization are shown in Fig. 5.6.a. The correlation coefficients are within the range of  $-149.4$  and  $87.44$ . Applying normalization leads to the normalized correlation coefficients which are shown in Fig. 5.6.b. The Pearson's correlation coefficients are shown for comparison and demonstrate good agreement with the normalized **MTGP** coefficients. The absolute difference between the normalized **MTGP** and Pearson's correlation coefficients is on average  $0.015$ .

### 5.2.3 Time Shift Estimation

The **MTGP** example II illustrated that the correlation between two signals could be influenced if one task is temporally shifted relative to another task. In extreme cases, the correlation may become zero, meaning that the prediction outcome of a **MTGP** model would be comparable to two **STGP** models with the same temporal hyperparameters  $\theta_t$ . Within respiratory motion compensation, temporal shifts between different sensors are to be expected. This might result in a decreased prediction accuracy.

To overcome this problem, the **MTGP** model can be extended by additional hyperparameters  $\theta_s$  which represent the temporal shift between tasks. We assume that the first



**Figure 5.7:** Predicted  $y^{1*}$  of an **MTGP** model with and without a shift hyperparameter  $\theta_s$  for a phase shift of (a)  $\phi = 0$ , (b)  $\phi = \pi/4$ , (c)  $\phi = \pi/2$ , and (d)  $\phi = \pi$ .

task is selected to be the reference task. If the feature space of all residual tasks is shifted by an individual hyperparameter,  $o - 1$  additional hyperparameters are needed. The feature space of one task can be shifted according to

$$x^{j'} = x^j + \theta_{s,j-1} \quad (5.8)$$

with  $j \in \{2, \dots, o\}$ ,  $\theta_s = [\theta_{s,1}, \dots, \theta_{s,o-1}]^\top$ , and  $o$  the number of tasks. The number of shift hyperparameters can be reduced if prior knowledge is available such as that two tasks are temporally aligned or have a constant shift. Equivalent to the residual hyperparameter,  $\theta_s$  can be learned by minimizing the **NLML**.

**MTGP example III** Two tasks were considered in this example. The tasks represented two sinusoidal signals with a period of 1 s. A noise component with a distribution of  $\mathcal{N}(0, 0.01)$  was added to both tasks. The training data was  $x^1 \in [0, 2]$  s for task one and  $x^2 \in [0, 7.5]$  s for task two. Additionally, a phase shift of  $\phi \in \{0, \pi/4, \pi/2, \pi\}$  was applied to task two. The aim was to predict  $y^{1*}$  of task one depending on  $\phi$  for the test region  $x^{1*} \in (2, 7.5]$  s. The **MTGP** models were initialized with and without a shift hyperparameter  $\theta_s$ .

Figure 5.7 shows the prediction results of  $y^{1*}$  depending on the phase shift  $\phi$ . The outcome of the **MTGP** models without  $\theta_s$  is illustrated by a red dashed line. It can be observed that in Fig. 5.7.a and 5.7.d,  $y^{1*}$  and  $y^1$  are in good agreement for the test range. Table 5.3 indicates that the Pearson's correlation coefficients between the tasks are close

to either 1 or  $-1$  for these cases. As a consequence, accurate prediction can be computed for  $y^{1*}$  based on the training data of  $y^2$ .

For  $\phi = \pi/2$ , the Pearson's correlation coefficient is 0 (Table 5.3). This leads to the effect that the available training data of  $y^2$  cannot be used to predict  $y^{1*}$ . The prediction outcome of  $y^{1*}$  is comparable to an **STGP** model (Fig. 5.7.c). As the distance between test and training features increases, the prediction of  $y^{1*}$  is equivalent to the prior distribution (mean of the training labels). Compared to Fig. 5.3.b, small oscillations around zero are visible. Table 5.3 reveals that the **MTGP** model learned a low correlation of  $\mathbf{K}'_c[1, 2] = 0.054$ , which is most likely due to the added noise component. The **MSLL** value is with 1.048 very high and indicates that the models perform worse compared to a trivial model.

Figure 5.7.b illustrates the scenario if the correlation is between 0 and 1. As the correlation is decreased, the predicted signal  $y^{1*}$  has a decreased amplitude. Furthermore, as the distance between  $y^{1*}$  and the training labels  $y^1$  increases, the phase shift of  $y^2$  is projected to the prediction results of  $y^{1*}$ .

The green dashed lines represent the prediction output of **MTGP** models with  $\theta_s$ . It can be observed that the prediction outcome seems to be independent of the phase shift  $\phi$ . This is confirmed by the **RMSE**, **nRMSE** and **MSLL** values shown in Table 5.3. The normalized correlation values  $\mathbf{K}'_c[1, 2]$  of the **MTGP** models with  $\theta_s$  are consistently close to 1 or  $-1$ . The last row of Table 5.3 shows the estimated shift hyperparameter  $\theta_s$  in seconds. The shift is  $\theta_s = 0.123$  s and  $\theta_s = -0.247$  s for  $\phi = \pi/4$  and  $\phi = \pi/2$ , respectively. This correlates with the true phase shift values of 0.125 s and  $-0.25$  s. Note that multiple phase shift solutions are possible due to the periodicity of the signals. The presented results depend on the initially chosen hyperparameter, which was  $\theta_s = 0$  s.

Adding additional hyperparameters increases the model complexity. However, in cases where this is not necessary, this might lead to a decreased prediction accuracy as more hyperparameters have to be optimized. This is illustrated in Table 5.3 for the case of  $\phi = 0$  and  $\phi = \pi$ . The **RMSE** and **nRMSE** values of the **MTGP** models with  $\theta_s$  is slightly increased compared to the **MTGP** models without  $\theta_s$ .

### 5.2.4 MTGP Tracking Algorithm

The **MTGP** approach described in the previous sections can be used to efficiently solve the prediction and correlation problem within one model. Compared to alternative approaches, the model can incorporate the information from multiple motion traces even though they might be sampled with different frequencies. This is also true for extreme cases, if, e.g., one motion trace is constantly sampled at a high frequency and one motion



**Table 5.3:** Estimated Pearson’s correlation coefficient  $r$  and normalised **MTGP** correlation coefficient  $k'_c(1, 2)$  between task one and two as well as **RMSE**, **nRMSE**, and **MSLL** for an **MTGP** model without and with additional shift hyperparameter  $\theta_s$

$\phi$	0	$\pi/4$	$\pi/2$	$\pi$
$r$	0.982	0.684	0.001	-0.983
MTGP without $\theta_s$				
$k'_c(1, 2)$	0.999	0.7	0.054	-1
RMSE	0.114	0.504	0.709	0.124
nRMSE	0.158	0.718	0.975	0.173
MSLL	-1.231	-0.082	1.048	-1.186
MTGP with $\theta_s$				
$k'_c(1, 2)$	1	0.999	-1	-1
RMSE	0.129	0.116	0.187	0.128
nRMSE	0.180	0.164	0.261	0.178
MSLL	-1.186	-1.149	-1.109	-1.208
$\theta_s[s]$	0.004	0.123	-0.247	-0.001

trace is only sampled at a few discrete time points.

In the following, the outline of the **MTGP** tracking algorithms used will be described. The tracking algorithm considered the correlation between multiple tasks by using the “free-form” parametrisation (sec. 5.2.1) as well as temporal delays between tasks (sec. 5.2.3). This resulted in additional hyperparameters  $\theta_c$  and  $\theta_s$  compared to a normal **STGP** model (sec. 3.2.2). To limit the total number of hyperparameters, only basic temporal covariance functions were considered (see sec. 3.2.2). Three tracking algorithms were designed using either a squared-exponential (Eq. 3.37), periodic (Eq. 3.40), or quasi-periodic (Eq. 3.42) covariance function. We refer to these **MTGP** algorithms as **MTGP<sub>SE</sub>**, **MTGP<sub>P</sub>**, and **MTGP<sub>QP</sub>**, respectively. In general, we considered  $o \geq 2$  number of tasks (one internal and one or multiple external motion traces). All internal and external data were acquired at the same sampling frequency  $f_s$ . During the training phase of the algorithms, we assumed that  $n_{int}$  internal and  $n_{ext}$  external training points were given. In case of multiple external motion traces, we assumed that the same number of external training points  $n_{ext}$  was given for each motion trace. In general, the amount of internal data is smaller than external data  $n_{int} < n_{ext}$ . During the test phase, no additional internal points were acquired. The outcome of the **MTGP** approach was  $y_{i+\xi}^{int*}$  based on

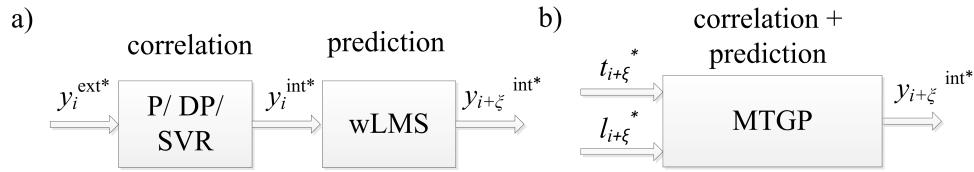
$x_i^{int*} = t_{i+\xi}$ . In case of  $\xi = 0$ , the **MTGP** tracking algorithm was used as a correlation model. For  $\xi > 0$ , the **MTGP** approach was used as a combined prediction and correlation model. Note, the proposed **MTGP** tracking algorithm was applied to datasets whose data was acquired at the same sampling frequency  $f_s$ . This simplification was done to enable a comparison to alternative prediction and correlation models, which will be presented in the next section. The **MTGP** model is not restricted to these simplifications.

During the training phase, the hyperparameters of the **MTGP** algorithm were optimized based on all available internal and external training data. In case of **MTGP<sub>P</sub>** and **MTGP<sub>QP</sub>**, the hyperparameter  $\theta_P$  was initialised based on the main breathing frequency estimated by the external training data. The shift hyperparameters  $\theta_s$  were initially set to the delay which maximises the cross-correlation between internal and external training data. The internal task was selected as reference task for the shift hyperparameters. The residual hyperparameters were randomly initialised and optimized using 200 gradient descent steps. The training phase was repeated 30 times to avoid the effect of local minima. The set of hyperparameters resulting in the lowest **NLML** value was selected for the test phase.

During the test phase, new external points were acquired as time progressed. At each new time step  $t_i$ , the hyperparameters of the temporal covariance function were further optimized by  $n = 10$  gradient descent steps. The optimization was based on a sliding window approach with a fixed window size of  $n_w = 100$  to reduce the computational requirements. The shift and correlation hyperparameters  $\theta_s$  and  $\theta_c$  were fixed as no new internal data points were acquired.

### 5.2.5 Alternative Algorithms

The outcome of the **MTGP** models was compared to the prediction and correlation results of previously presented algorithms. We used the output of a correlation model as input for a prediction model, as it was for example the case for the CyberKnife<sup>®</sup> [227]. Three correlation models were considered: a second-order **polynomial (P)**, a **dual-polynomial (DP)**, and an **SVR** model. The **DP** model was based on two quadratic functions, one for inspiration and one for expiration. The **SVR** correlation model was presented in [83] and evaluated on dataset I. In [83], various input features were investigated such as position, velocity, and acceleration of one or multiple dimensions. The best performance on this dataset could be achieved by using the position and velocity information of the three spatial dimensions of one marker. The parameter of the **SVR** model were set to  $\varepsilon = 0.2$  and  $C = 1$  while using an **RBF** kernel [83]. Furthermore, the **DP** and **SVR** model re-



**Figure 5.8:** Illustration of correlation and prediction with (a) the alternative algorithms and (b) the **MTGP** approach. Note during the test phase, the input of the alternative approaches is the observation  $y_i^{ext*}$  of one external sensor (or multiple as in case of **SVR**). The **MTGP** approach requires the time point  $t_{i+\xi}^*$  of the prediction and the label  $l_i^*$  specifying the internal motion task.

quire an additional binary feature to differentiate between inspiration and expiration. To highlight the difference between the alternative models and the **MTGP** approach, Fig. 5.8 illustrate both approaches.

The **wLMS** approach [81] was selected as prediction algorithm due to its superior performance in [86]. The **wLMS** parameters were selected as  $\mu = 0.0204$ ,  $J = 3$ , and  $m = 199$  according to [81].

We refer to the combined prediction and correlation algorithms as **P-wLMS**, **DP-wLMS**, and **SVR-wLMS**, respectively.

### 5.3 Evaluation Procedure

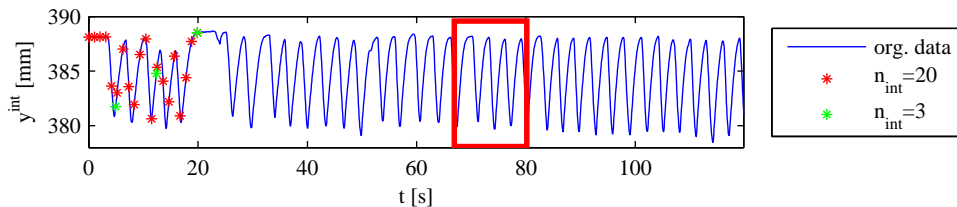
Three experiments were performed. First, Pearson’s correlation coefficient was computed to investigate the correlation between external and internal motion using data of dataset II (sec. 5.1). This experiment was comparable to the first investigation described in sec. 4.3. It can be interpreted as an initial investigation of the correlation and its variation between internal and external motion traces. The following two experiments focus on using the **MTGP** model for combined prediction and correlation. The outcome of the **MTGP** models was compared to alternative approaches described in sec. 5.2.4. In the second experiment, the influence of the number of internal training points was evaluated on the data of the porcine study (dataset I, sec. 5.1). The third experiment used data of the human study and investigated multivariate extension of the **MTGP** method.

**Experiment 1 (E1): Correlation Analysis** Pearson’s correlation coefficient  $r$  was computed between the internal landmark and all external markers of the second dataset. In contrast to the correlation analysis in sec. 4.3, only data of measurement phase M1 (see sec. 4.1) from nine subjects was considered, since only for these subjects a continuous

internal motion trace could be acquired (see sec. 5.1). The first minute of each recording was discarded to avoid the influence of unnatural breathing patterns (equivalent to sec. 4.3). The residual data was divided into motion fragments of 1 min. As the recorded internal motion traces had a varying duration, the last motion fragment could have a duration of below 1 min. These shortened fragments were discarded. This resulted in 151 motion fragments over all nine subjects. The mean absolute correlation coefficient  $|r|$  and its standard deviation were computed over all motion fragments. Furthermore, the absolute correlation coefficient difference  $|\Delta r|$  between the 1<sup>st</sup> and the 6<sup>th</sup> and between the 1<sup>st</sup> and the 12<sup>th</sup> motion fragment were investigated. A computation of the variance difference to the 18<sup>th</sup> motion fragment as in the first experiment of sec. 4.3 was not possible due to the shortened length of the internal motion traces. The mean and the standard deviation of the correlation coefficients and their differences are computed using the Fisher transformation [95, 254].

**Experiment 2 (E2): MTGP - Number of Internal Training Points** The aim of this experiment was to investigate the influence of the number of internal data points  $n_{int}$ . The experiment was performed on the data of the porcine study (dataset I). The first 20 s were assumed to be the training set. Within the training set, all external points were given ( $n_{ext} = 300$ ). The number of internal training points was  $n_{int} \in \{3, 5, 7, 10, 13, 15, 17, 20\}$ . For  $n_{int} \geq 10$ , the internal training points were equidistantly sampled across the training set. To ensure that a complete breathing cycle was represented, the training points were manually selected for  $n_{int} < 10$ . The test set was the following 100 s. Figure 5.9 illustrates the selected training points for  $n_{int} = 3$  and  $n_{int} = 20$  for one internal motion trace. In total, twelve internal motion traces were evaluated (four fiducials with three spatial dimensions each).

The prediction accuracy was evaluated using the **RMSE**. The latency index was selected as  $\xi \in \{0, 1, \dots, 4\}$ , which corresponded to  $h \in \{0, 66.7, 133.3, 200, 266.7\}$  ms. The outcome of the **MTGP** models was compared to the alternative algorithms. The **MTGP**



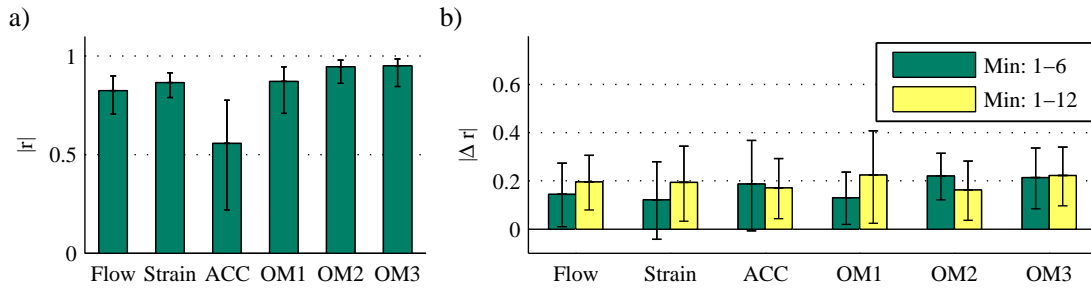
**Figure 5.9:** Example recording of the position of one fiducial with the training data for  $n_{int} = 3$  and  $n_{int} = 20$ . The motion fragment highlighted in red will be further analysed in sec. 5.4.2.

approaches, the **P-wLMS**, and the **DP-wLMS** algorithm computed  $y^{int*}$  based on one spatial dimension of one external marker ( $o = 2$ ). In contrast, the **SVR** algorithm was evaluated equivalently to [83], where the **SVR** algorithm used the position and the velocity of three spatial dimensions of one external marker.

One remaining challenge was the selection of the optimal external marker which results in the lowest **RMSE**. Due to the small number of internal training points  $n_{int}$ , a marker selection criterion might be strongly influenced by the selected points. To investigate the principle performance of the **MTGP** approach relative to the alternative algorithms, the best marker was selected retrospectively. This is equivalent to the evaluation of [83]. Furthermore, the **NLML** value of the training set was investigated as a potential marker selection criterion. The marker combination resulting in the lowest **NLML** value was selected. We refer to results of this approach as  $MTGP^{NLML}$ . The results of the retrospective analysis are referred to as  $MTGP^{best}$ .

**Experiment 3 (E3): MTGP - Multivariate Prediction and Correlation** The third experiment focused on the multivariate extensions of the **MTGP** approach. The experiment was performed on the data of the human study (dataset II). Similar to the second experiment of sec. 4.3, it was investigated if the prediction accuracy of an **MTGP** model could be increased if multiple external markers were considered simultaneously ( $o > 2$ ). The experiment was evaluated on the  $MTGP_{SE}$  tracking algorithm with either  $o \in \{2, 3, 4\}$  tasks, resulting in one to three external tasks. For comparison, the outcome of the **SVR-wLMS** approach was investigated as well as the **DP-wLMS** approach. Note that only the **SVR-wLMS** approach can be extended for varying numbers of external signals. The first 30 s of each dataset were considered as training set. The number of internal training points was fixed to  $n_{int} = 20$ , which were equidistantly-sampled across the training set. All external points were used, resulting in  $n_{int} = 509$  training points per external signal. The latency index was set to  $\xi \in \{0, \dots, 4\}$ , which resulted in a prediction latency of  $h \in \{0, 58.8, 117.6, 176.5, 235.3\}$  ms.

Equivalent to experiment E2, the best possible marker combinations for  $MTGP_{SE}$ , **SVR-wLMS**, and **DP-wLMS** were selected retrospectively based on the **RMSE** of the test set, as the focus lied on a principle investigation if a multivariate prediction and correlation model could result in an increased prediction accuracy. Furthermore, the **NLML** value of the training set was investigated for the **MTGP** approach as a potential marker selection criterion. We refer to these results as  $MTGP^{NLML}$  and to the results of the retrospective analysis as  $MTGP^{best}$ .



**Figure 5.10:** (a) Mean absolute correlation coefficient  $|r|$  and standard deviation of the external motion traces with respect to the motion trace of the internal landmark for nine human subjects; (b) Mean absolute correlation difference  $|\Delta r|$  between the motion fragment of minutes one and six and one and twelve.

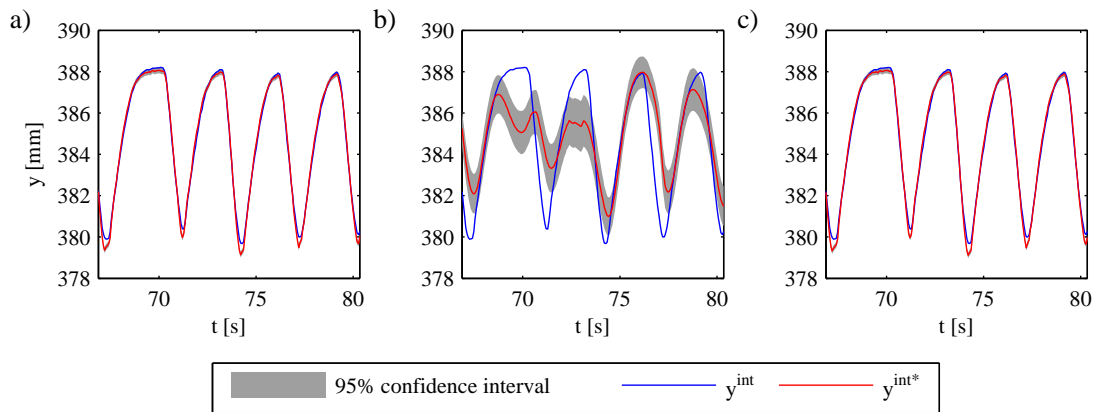
## 5.4 Experimental Evaluation

The results of the correlation analysis between internal and external motion traces are presented in sec. 5.4.1. Section 5.4.2 and 5.4.3 contain the results of the **MTGP** experiments. The experiments and their results are discussed in the context of respiratory motion compensation in sec. 5.4.4.

### 5.4.1 Correlation Analysis

Figure 5.10.a shows the mean absolute correlation coefficients of the external markers with respect to the **ILM**. The highest mean absolute correlation is found between OM2 and OM3 with  $|r_{OM2,ILM}| = |r_{OM3,ILM}| = 0.95$ . Similar to the results of sec. 4.4.1, a marker placement dependency can be observed in the case of OM1. The mean absolute correlation coefficient is  $|r_{OM1,ILM}| = 0.87$ , which is comparable to the correlation of the strain and flow sensors with  $|r_{Strain,ILM}| = 0.87$  and  $|r_{Flow,ILM}| = 0.83$ . The lowest mean correlation (and the highest standard deviation) can be observed for the acceleration sensor with  $|r_{ACC,ILM}| = 0.56$ .

The mean absolute correlation coefficient differences  $|\Delta r|$  between the first and the sixth and the first and the twelfth minute are illustrated in Fig. 5.10.b. The lowest mean correlation difference can be observed between the strain sensor and the **ILM** with  $|\Delta r_{Strain,ILM}| = 0.12$  between minutes one and six and between OM2 and **ILM** with  $|\Delta r_{OM2,ILM}| = 0.16$  between minutes one and twelve. The highest correlation differences on average can be observed between the **OMs** and **ILM** (e.g.,  $|\Delta r_{OM1,ILM}| = |\Delta r_{OM3,ILM}| = 0.22$  between minutes one and twelve). These results indicate a relatively high variation of the mean correlation coefficient and standard deviation over time for



**Figure 5.11:** Outcome of the **MTGP** methods for the motion fragment highlighted in Fig. 5.9 using an **MTGP** model with (a) a squared-exponential, (b) periodic, and (c) quasi-periodic temporal covariance function.

all external markers.

#### 5.4.2 MTGP - Number of Internal Training Points

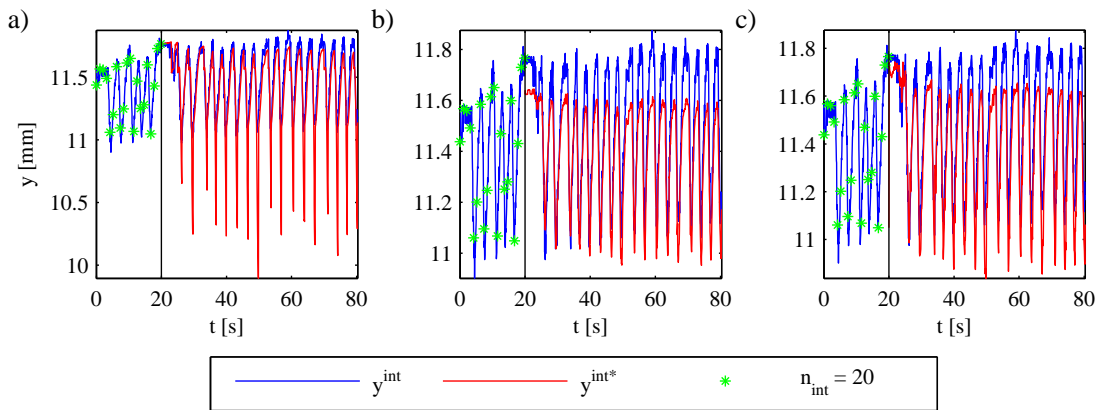
Figures 5.11.a - 5.11.c illustrate the prediction results of the  $\text{MTGP}_{SE}$ ,  $\text{MTGP}_P$ , and  $\text{MTGP}_{QP}$  tracking algorithms, respectively. The combined prediction and correlation results are shown as an example for the retrospective analysis ( $\text{MTGP}^{best}$ ) with a prediction latency of  $h = 133.3$  ms, and with  $n_{int} = 20$  internal training points. The red solid line represents the mean of the predicted Gaussian distribution and the 95% confidence interval is indicated as a gray area. Figure 5.11.b points out that the outcome of the  $\text{MTGP}_P$  algorithm will lead to high prediction errors. Similar to the results of an **STGP** algorithm (see experiment one of sec. 3.2.2.2), modelling respiratory motion with a periodic covariance function seems to be too inflexible. A more accurate outcome can be computed by using either a squared-exponential or a quasi-periodic covariance function which is indicated by Fig. 5.11.a and 5.11.c. The increased accuracy also leads to a smaller confidence interval in case of  $\text{MTGP}_{SE}$  and  $\text{MTGP}_{QP}$ .

The mean **RMSE** and mean **nRMSE** over the twelve internal motion traces are listed in Table 5.4 for  $n_{int} = 20$ . The results of the retrospective analysis are separated from the results using the **NLML** as marker selection criterion by two vertical lines. The best results depending on the prediction latency are highlighted bold. For  $h = 0$  ms, the algorithms are used as correlation algorithms. This means that in the case of **P-wLMS**, **DP-wLMS**, and **SVR-wLMS**, the **wLMS** algorithm is not used. The table reveals that the **DP-wLMS** approach has the lowest **RMSE** value for  $h \leq 66.7$  ms. For  $h \geq 66.7$  ms,

**Table 5.4:** Mean **RMSE** and **nRMSE** for the alternative approaches (P-wLMS, DP-wLMS, SVR-wLMS) and the MTGP models depending on the prediction latency  $h$  (for  $n_{int} = 20$ ). **RMSE** and **nRMSE** values marked by an asterisk represent the outlier corrected results. (best results depending on  $h$  highlighted in bold)

$h$ [ms]	measure	RMSE [mm]												
		P-wLMS			DP-wLMS			SVR-wLMS			MTGP <sup>best</sup>			MTGP <sup>NLML</sup>
		wLMS	wLMS	wLMS	wLMS	wLMS	wLMS	wLMS	SE	PER	QP	SE	PER	QP
0	RMSE [mm]	0.418	<b>0.206</b>	0.245	0.216	0.784	0.215	0.242	1.617	0.243				
	nRMSE	0.436	<b>0.358</b>	0.371	0.375	0.530	0.385	0.386	1.263	0.397				
	RMSE* [mm]	0.439	<b>0.208</b>	0.251	0.218	0.842	0.216	0.247	1.725	0.248				
66.7	nRMSE*	0.168	<b>0.103</b>	0.129	0.112	0.338	0.112	0.122	0.710	0.124				
	RMSE [mm]	0.413	<b>0.215</b>	0.260	0.216	0.865	<b>0.215</b>	0.241	1.645	0.243				
	nRMSE	0.434	<b>0.359</b>	0.377	0.373	0.560	0.384	0.384	1.275	0.397				
133.3	RMSE* [mm]	0.432	0.218	0.268	0.218	0.930	<b>0.216</b>	0.246	1.756	0.247				
	nRMSE*	0.166	<b>0.106</b>	0.135	0.111	0.374	0.112	0.1216	0.721	0.124				
	RMSE [mm]	0.409	0.241	0.297	0.216	0.954	<b>0.215</b>	0.240	1.675	0.244				
200	nRMSE	0.435	<b>0.367</b>	0.393	0.370	0.596	0.384	0.381	1.286	0.397				
	RMSE* [mm]	0.428	0.247	0.308	0.219	1.027	<b>0.217</b>	0.245	1.789	0.249				
	nRMSE*	0.165	0.114	0.149	<b>0.111</b>	0.414	0.112	0.121	0.732	0.124				
266.7	RMSE [mm]	0.413	0.291	0.355	0.227	1.047	<b>0.225</b>	0.249	1.708	0.255				
	nRMSE	0.441	0.384	0.418	<b>0.373</b>	0.639	0.388	0.384	1.297	0.401				
	RMSE* [mm]	0.433	0.301	0.371	0.230	1.128	<b>0.228</b>	0.254	1.825	0.260				
266.7	nRMSE*	0.168	0.131	0.171	<b>0.115</b>	0.455	0.116	0.124	0.745	0.129				
	RMSE [mm]	0.435	0.363	0.431	0.274	1.141	<b>0.261</b>	0.294	1.743	0.291				
	nRMSE	0.456	0.411	0.452	<b>0.395</b>	0.687	0.404	0.408	1.308	0.419				
266.7	RMSE* [mm]	0.456	0.379	0.453	0.282	1.230	<b>0.267</b>	0.303	1.863	0.299				
	nRMSE*	0.179	0.157	0.201	0.135	0.497	<b>0.130</b>	0.144	0.758	0.144				





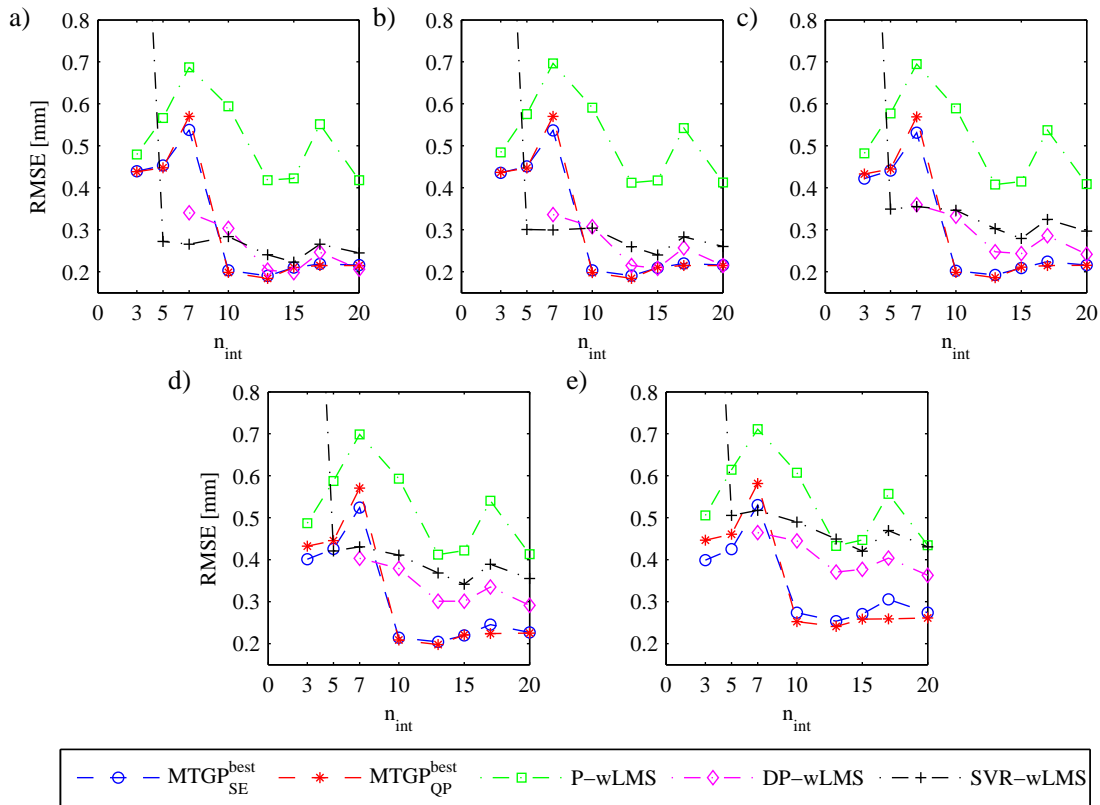
**Figure 5.12:** Outcome of (a) the **DP-wLMS**, (b) the **SVR-wLMS**, and (c) the **MTGP<sub>QP</sub><sup>best</sup>** algorithm for the outlier motion trace for  $n_{int} = 20$  internal training points and  $h = 0$  ms. The vertical black line separates the training and the test data.

the **MTGP<sub>QP</sub><sup>best</sup>** algorithm has the lowest **RMSE** value on average, closely followed by the **MTGP<sub>SE</sub><sup>best</sup>** algorithm. The **DP-wLMS** algorithm has the lowest **nRMSE** for  $h \leq 133.3$  ms and the **MTGP<sub>SE</sub><sup>best</sup>** for  $h > 133.3$  ms. As Fig. 5.11.b already indicated, the mean **RMSE** for **MTGP<sub>P</sub>** is very high.

Comparing the **MTGP** results of the retrospective analysis and of the **NLML** criterion shows only a minor decrease of the prediction accuracy for **MTGP<sub>SE</sub>** and **MTGP<sub>QP</sub>**. Table 5.4 points out a superior performance of the **MTGP** approaches especially for high prediction latencies. As an example, the mean **RMSE** increases for **MTGP<sub>SE</sub><sup>best</sup>** from 0.216 mm for  $h = 0$  ms to 0.274 mm for  $h = 266.7$  ms. In contrast, the **RMSE** of **DP-wLMS** increases from 0.206 mm to 0.363 mm and the **RMSE** of the **SVR-wLMS** method from 0.245 mm to 0.431 mm.

The evaluation of the twelve motion traces reveals the existence of one outlier. No algorithm was able to learn the correlation of the first spatial dimension of the first fiducial, which is the motion trace with the smallest amplitude. The motion trace and the outcome of the **DP-wLMS**, the **SVR-wLMS**, and the **MTGP<sub>QP</sub><sup>best</sup>** are illustrated in Fig. 5.12 for  $h = 0$  ms. The correlation errors might be caused by the small amplitude of the trace, which leads to a stronger influence of the measurement noise. Furthermore, an increasing mean offset is visible in the observed data. The results of the **SVR-wLMS** and the **MTGP<sub>QP</sub><sup>best</sup>** (Fig. 5.12.b and 5.12.c) point out that neither algorithm was able to learn the mean shift.

The values marked by an asterisk in Table 5.4 represent the outlier corrected mean **RMSE** and **nRMSE** values. Due to the small amplitude of the motion trace, the influence



**Figure 5.13:** Mean **RMSE** of the **MTGP** models using either a squared-exponential or quasi-periodic covariance function, the **P-wLMS**, the **DP-wLMS**, and the **SVR-wLMS** approaches depending on the number of internal training points  $n_{int}$  for a prediction latency of (a)  $h = 0$  ms, (b)  $h = 66.7$  ms, (c)  $h = 133.3$  ms, (d)  $h = 200$  ms, and (e)  $h = 266.7$  ms.

of the prediction error is small on the mean **RMSE** over all motion traces. However, the influence of the outlier becomes visible in the mean **nRMSE**. For example, the mean **RMSE** of **DP-wLMS** is 0.206 mm for  $h = 0$  and the mean outlier corrected **RMSE\*** is 0.208 mm. In contrast, the **nRMSE** is 0.358 and the outlier corrected **nRMSE\*** is 0.103.

The influence of the number of internal training points  $n_{int}$  is illustrated for different latencies  $h$  in Fig. 5.13.a - 5.13.e, respectively. The results of the **MTGP<sub>P</sub>** algorithm have been discarded for illustration purposes. Further, no result can be computed for the **DP-wLMS** approach with  $n_{int} \in \{3, 5\}$ , as a correlation model consisting of two quadratic curves requires at least six training points. The **RMSE** of the **MTGP** models indicate only a small influence of the number of internal training points for  $n_{int} \geq 10$ . For  $n_{int} < 10$ , the mean **RMSE** of both **MTGP** models increases. In general, the **SVR-wLMS** algorithm seems to be less affected by  $n_{int}$ . However for  $n_{int} = 3$ , the **RMSE** increases strongly, as it

is more challenging to train an SVR model with only three training points. Similar to the results shown in Table 5.4, a strong increase of the **RMSE** for **DP-wLMS** and **SVR-wLMS** can be observed for increasing prediction latencies.

### 5.4.3 MTGP - Multivariate Prediction and Correlation

The results of the third experiment are summarized in Table 5.5. It contains the mean **RMSE** and mean **nRMSE** over all subjects for the **DP-wLMS**, the **SVR-wLMS**, and the  $MTGP_{SE}$  models depending on the prediction latency  $h$ . The numbers in the headers of the **SVR-wLMS** and the  $MTGP_{SE}$  columns indicate the number of external markers used. The lowest **RMSE** and **nRMSE** are highlighted in bold depending on  $h$ . The highest prediction accuracy could be achieved by the  $MTGP_{SE}^{best}$  models using three external markers for all latencies.

In general, the results reveal a decrease of the prediction error for the **SVR-wLMS** and the  $MTGP_{SE}^{best}$ , while increasing the number of external markers from one to three. The mean **RMSE** decreases in case of  $MTGP_{SE}^{best}$  from 2.054 mm for one external marker to 1.853 mm for three external markers. However, a strong difference between the **RMSE** and the **nRMSE** values of the  $MTGP_{SE}^{best}$  and the  $MTGP_{SE}^{NLMML}$  can be observed. For  $MTGP_{SE}^{NLMML}$ , the prediction error decreases if the information of two external marker are considered instead of one. However, the error increases again in case of three external markers. This results in the highest **RMSE** and **nRMSE** values on average.

Table 5.6 lists the occurrence of one external marker within the optimal marker combinations which leads to the lowest **RMSE**. In case of using only the data of one external marker, such as for the **DP-wLMS**, and the first columns of **SVR-wLMS**,  $MTGP_{SE}^{best}$ , and  $MTGP_{SE}^{NLMML}$ , the sum of occurrences is nine. In case of two and three external markers, the sum of occurrences is 18 or 27, respectively. The used external markers are comparable for the **SVR-wLMS** and the  $MTGP_{SE}^{best}$  methods. However, it can be observed that the  $MTGP_{SE}^{NLMML}$  algorithm prefers the flow marker, as it is present in every marker combination.

### 5.4.4 Discussion

The correlation coefficients of experiment E1 (Fig. 5.10) reveal a strong variation between the internal and external motion traces. This is in agreement with the correlation investigations of Yan *et al.* [283].

Comparing the **RMSE** results of experiment E2 and E3 (Table 5.4 and 5.5) indicates a higher **RMSE** in general for measurement E3. This is mainly caused by the increased

**Table 5.5:** Mean **RMSE** for experiment E3, for **DP-wLMS**, **SVR-wLMS** and **MTGP<sub>SE</sub><sup>NLML</sup>** for  $m_{ext} \in \{1, 2, 3\}$  depending on the prediction horizon  $h$  (best results highlighted bold)

$h$ [ms]	measure	RMSE [mm]									
		DP-wLMS	SVR-wLMS			MTGP <sub>SE</sub> <sup>best</sup>			MTGP <sub>SE</sub> <sup>NLML</sup>		
0	RMSE [mm]	2.148	2.059	1.927	1.854	2.054	1.880	<b>1.853</b>	3.100	2.465	3.911
	nRMSE	0.168	0.158	0.148	<b>0.142</b>	0.157	0.147	<b>0.142</b>	0.223	0.199	0.289
	RMSE [mm]	2.168	2.071	1.929	1.862	2.056	1.890	<b>1.860</b>	3.099	2.465	3.859
58.8	nRMSE	0.168	0.157	0.148	0.143	0.157	0.147	<b>0.142</b>	0.223	0.199	0.286
	RMSE [mm]	2.223	2.100	1.951	1.889	2.063	1.905	<b>1.871</b>	3.097	2.468	3.794
	nRMSE	0.170	0.160	0.150	0.145	0.157	0.148	<b>0.143</b>	0.225	0.199	0.281
176.5	RMSE [mm]	2.305	2.131	1.988	1.936	2.079	1.930	<b>1.895</b>	3.094	2.482	3.721
	nRMSE	0.174	0.162	0.153	0.149	0.158	0.149	<b>0.144</b>	0.222	0.199	0.276
	RMSE [mm]	2.391	2.178	2.040	1.999	2.110	1.973	<b>1.940</b>	3.092	2.510	3.648
235.3	nRMSE	0.179	0.166	0.157	0.154	0.160	0.151	<b>0.146</b>	0.221	0.200	0.270

**Table 5.6:** Occurrence of the external markers in one of optimal marker combinations for **DP-wLMS**, **SVR-wLMS**, **MTGP<sub>SE</sub><sup>best</sup>**, and **MTGP<sub>SE</sub><sup>NLML</sup>**.

Marker	DP-wLMS	RMSE [mm]								
		SVR-wLMS			MTGP <sub>SE</sub> <sup>best</sup>			MTGP <sub>SE</sub> <sup>NLML</sup>		
OM1	2	1	2	3	1	2	3	1	2	3
OM2	1	2	7	8	1	4	5	0	2	2
OM3	5	2	2	2	1	5	5	0	0	1
Strain	0	5	6	6	5	6	7	0	3	7
Flow	0	0	0	3	2	2	5	0	4	8
ACC	1	0	3	4	0	1	4	9	9	9
	0	0	0	4	0	0	1	0	0	0

mean amplitude of the internal landmarks of dataset II compared to the mean amplitude of the fiducials of dataset I (see sec. 5.1). On average the amplitude is 5.98 mm for dataset I and 24.07 mm for dataset II. Computing the **nRMSE** reveals a lower prediction error for E3 on average. Comparing the outlier corrected **nRMSE\*** of E2 and the **nRMSE** of E3, reveals an only slightly decreased prediction accuracy for experiment E3. The remaining differences can be explained due to the increased duration of the test set of E3 and the used internal imaging modalities (**X-ray** vs. **US**). The test phase of E3 is relatively long ( $t > 15$  min). In a real scenario, additional internal training points would have been acquired to update the correlation model. As a result, the models would be more adaptive to variations of the breathing pattern. Furthermore, it has to be highlighted that tracking an internal point with **US** is not clinical standard. In general, tracking of internal points with **US** has a reduced accuracy compared to **X-ray**, due to the increased amount of measurement noise and lower imaging resolution. Additionally, the **US** tracking accuracy strongly depends on the location of the tumour in the body. The use of **US** in dataset II was only possible as the target was one clearly visible point.

Overall, the outcome of the  $\text{MTGP}^{best}$  model indicates a comparable or increased prediction accuracy for experiments E2 and E3 relative to the alternative approaches (Table 5.4 and 5.5). The **RMSE** and **nRMSE** increases more slowly for increasing the prediction latency  $h$  compared to the **DP-wLMS** and **SVR-wLMS** model in both experiments. This results in the highest prediction accuracy (**RMSE** and **nRMSE**) for **MTGP** approaches for  $h \geq 133,3$  ms. Consequently, the approach will be very relevant for systems such as robotic patient couches or **MLCs** with high latencies. Additionally in case of high prediction latencies, the error can be further decreased by using more complex temporal covariance functions such as the quasi-periodic function. For short latencies, the outcome of  $\text{MTGP}_{SE}^{best}$  and  $\text{MTGP}_{QP}^{best}$  is comparable. Using a periodic covariance function seems to be too inflexible to model respiratory motion. These results are comparable with the results of the first experiment in sec. 3.2.2.2.

The outcome of experiment E2 indicates only a small dependency on the number of internal training points of the  $\text{MTGP}^{best}$  models for  $n_{int} > 10$ . For  $n_{int} < 10$ , the **RMSE** of the  $\text{MTGP}^{best}$  models increases (see Fig. 5.13). External and internal signals are often temporally delayed. This leads to a hysteresis between internal and external motion, as it is illustrated in Fig. 2.5.b. To compensate for this effect, models may require an additional input classifying inspiration or expiration. Examples are the **DP-wLMS** and the **SVR-wLMS** models. In case of irregular breathing, it might be difficult to classify the current breathing state. This can lead to additional errors. In contrast, the presented **MTGP** models learn the temporal delay between the different motion traces. As a result,

no additional classification feature is required. Nonetheless, extending the **MTGP** model to learn the temporal delay increases the number of hyperparameters. This could explain the increased errors for  $n_{int} < 10$  as ambiguous solutions are possible.

The focus of the last experiment E3 was the multivariate extension of the **MTGP** approaches. The results reveal that the **SVR-wLMS** as well as the  $\text{MTGP}_{SE}^{best}$  approaches can benefit from data of multiple external markers. For example in the case of  $h = 0$  ms, the mean **RMSE** decreases by 0.205 mm for the **SVR-wLMS** model and by 0.201 mm for the  $\text{MTGP}_{SE}^{best}$  model. The increased accuracy of the **SVR-wLMS** algorithm confirms the results of Ernst *et al.* [83]. The authors reported an increased correlation accuracy of the **SVR** model by using multiple external optical markers. This experiment goes one step further by combining multiple external markers which have different modalities. Table 5.6 reveals that all modalities were used. However, the results indicate an increased relevance of the **OMs**. In contrast, the **ACC** marker was only used once in case of  $\text{MTGP}_{SE}^{best}$  and four times in case of **SVR-wLMS**. The low occurrence of the **ACC** marker in case of  $\text{MTGP}_{SE}^{best}$  can be explained by the shared temporal hyperparameters  $\theta_t$ . The **MTGP** models are based on the assumption that all signals have the same temporal hyperparameters. As the motion traces in Fig. 4.3 indicate, the temporal characteristics of **ACC** data might be different from to the **OMs**, strain, flow, and **ILM** data.

One remaining problem is the selection of the external marker or marker combination resulting in the lowest **RMSE**. As discussed in sec. 4.2.2, the marker selection criteria could be based on filters (such as the correlation between external and internal signals) or wrapper methods (such as leave-one-out methods or evaluation of the training error). However, all criteria will be strongly influenced by the selected internal training points. Furthermore, including multiple external markers might result in an increased prediction error due to the “curse of dimensionality” (see sec. 4.2.2). Within E2 and E3, the **NLML** value of the training data was investigated as a potential marker selection criterion. As discussed in sec. 3.2.2.1, the term consists of a model complexity as well as a data fit term, which balances between under- and overfitting. The outcome of experiment E2 indicates that the **NLML** criterion can be used to select the most relevant marker. The **RMSE** and **nRMSE** values of the  $\text{MTGP}_{SE}^{NLML}$  and the  $\text{MTGP}_{QP}^{NLML}$  are only slightly increased compared to  $\text{MTGP}_{SE}^{best}$  and  $\text{MTGP}_{QP}^{best}$ , respectively. However, the results of experiment E3 reveal that this criterion might be biased towards certain modalities. Table 5.6 points out that the flow marker was used in all  $\text{MTGP}^{best}$  models for all subjects.

One drawback of the **MTGP** approach is the increased computational requirements. The computation time is  $\mathcal{O}(\tilde{m}^3)$  with  $\tilde{m} = \sum_{j=1}^o m_j$  due to the matrix inversion of  $\mathbf{K}_{\text{MTGP}}$  (Eq. 5.3). Furthermore, an increased amount of external markers leads to an increased

number of correlation hyperparameters  $\theta_c$  (Eq. 5.5). The computational requirements can be reduced by using a sparse approximation of  $\mathbf{K}_{\text{MTGP}}^{-1}$  [199, 204, 235, 251, 252, 271] or an alternative parametrisation of the correlation matrix  $\mathbf{K}_c$ . The principle idea of sparse GP models is to find a set of pseudo training data  $\tilde{n}$  with  $\tilde{n} \ll n$ . As shown in [204], the majority of the sparse GP algorithms only differ in the way of finding the reduced set.

## 5.5 Further MTGP Extensions

The results of the previous sections point out two situations where the MTGP tracking algorithms are not working optimally. These are:

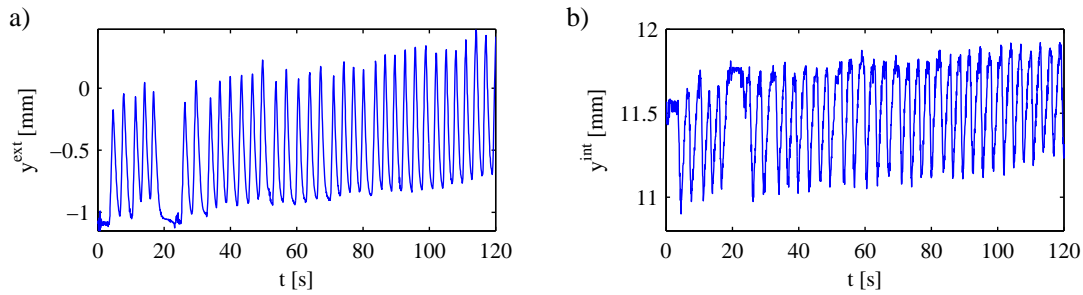
- prediction of motion traces with a baseline shift such as the outlier in experiment E2 (see Fig. 5.12); and
- multivariate extension of MTGP models to external motion traces with different temporal characteristics such as the ACC sensor.

Within this section, extensions of the MTGP model will be discussed which can be used to overcome the limitations of the current MTGP model in these situations. In sec. 5.5.1, the **linear model of coregularization (LMC)** is presented, which can be used to model baseline shifts. Further, the idea of convoluted kernels is discussed in sec. 5.5.2 which can be used to model tasks with different temporal characteristics.

These extensions have not been considered for the experiments in sec. 5.4 as they lead to twice the number of temporal hyperparameters  $\theta_t$  (convoluted kernels) and/or correlation hyperparameters  $\theta_c$  (for LMC). As a consequence, it is more challenging to train the MTGP models. Further research has to be undertaken to learn MTGP models with a large number of hyperparameters.

### 5.5.1 Linear Model of Coregularization

The first MTGP extension is motivated by the outlier in experiment E2. Further investigating the internal motion trace reveals a superposition of a positive and negative correlations. Figure 5.14.a shows, as an example, one external motion trace and Fig. 5.14.b the internal motion trace classified as outlier. It can be observed that periodic breathing motion between external and internal motion trace is negatively correlated, meaning that during a respiratory cycle a maximum of the external motion trace is a minimum of the internal motion trace. However, a second positive correlation can also be observed as both traces have an increasing baseline shift. The MTGP tracking algorithm is only able



**Figure 5.14:** Recorded data of (a) one external motion trace and (b) of the internal motion trace which was classified as outlier in experiment E2.

to learn one correlation which will result in an increased **RMSE**.

The **MTGP** model can be extended to learn multiple correlations. This approach is known as “linear model of coregularization” [44]. The idea is to extend Eq. 5.3 linearly such that

$$\mathbf{K}_{\text{MTGP}}(\tilde{\mathbf{x}}, \tilde{\mathbf{l}}, \boldsymbol{\theta}_c, \boldsymbol{\theta}_t) = \sum_{j=1}^u \mathbf{K}_{c,u}(\tilde{\mathbf{l}}, \boldsymbol{\theta}_{c,u}) \otimes \mathbf{K}_{t,u}(\tilde{\mathbf{x}}, \boldsymbol{\theta}_{t,u}), \quad (5.9)$$

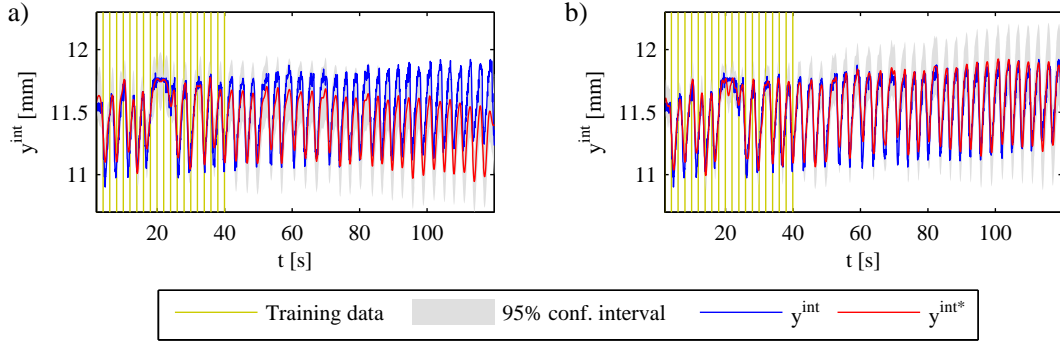
with  $u$  being the number of correlations. Assuming superposition of two correlations  $u = 2$ , the number of hyperparameters  $\boldsymbol{\theta}_c$  and  $\boldsymbol{\theta}_t$  would be doubled. The **LMC** can be further extended to use different covariance functions for different correlations such as using a quasi-periodic function for the periodic motion and a squared-exponential function for the baseline shift.

**MTGP Example IV** The external and internal motion traces shown in Fig. 5.14 were considered as a dataset for this example. The training data of the external task were sampled at  $f_s^{\text{ext}} = 5 \text{ Hz}$  for  $t \in (0, 120] \text{ s}$ . 20 internal training points were given, which were sampled equidistantly within the first 40 s. The residual internal points within  $t \in (0, 120] \text{ s}$  were considered as test set. Two **MTGP** models were trained. The first model

**Table 5.7:** Normalized **MTGP** correlation coefficients, **RMSE**, **nRMSE**, and **MSLL** of the **MTGP** model with one ( $u = 1$ ) and two ( $u = 2$ ) correlation terms.

$u$	$\mathbf{K}_{c,1}[2, 1]$	$\mathbf{K}_{c,2}[2, 1]$	RMSE	nRMSE	MSLL
1	-0.87	-	0.160 mm	0.659	-0.476
2 (MTGP <sub>LMC</sub> )	-0.94	0.99	0.073 mm	0.301	-0.816





**Figure 5.15:** Predicted internal position  $y^{int*}$  and confidence interval for (a) an **MTGP** model with one correlation term ( $u = 1$ ) and (b) an **MTGP** model with two correlation terms ( $u = 2$ ).

assumed only one correlation term ( $u = 1$ ) and a quasi-periodic temporal covariance function. The second **MTGP** model assumed a linear combination of two correlation terms ( $u = 2$ ) with a quasi-periodic (to model the periodic behaviour) and a squared-exponential temporal covariance function (to model the baseline shift), respectively. We refer to this model as **MTGP<sub>LMC</sub>**. It is specified as

$$\mathbf{K}_{\text{MTGP}}(\tilde{\mathbf{x}}, \tilde{\mathbf{l}}, \boldsymbol{\theta}_c, \boldsymbol{\theta}_t) = \mathbf{K}_{c,1}(\tilde{\mathbf{l}}, \boldsymbol{\theta}_{c,1}) \otimes \mathbf{K}_{QP,1}(\tilde{\mathbf{x}}, \boldsymbol{\theta}_{t,1}) + \mathbf{K}_{c,2}(\tilde{\mathbf{l}}, \boldsymbol{\theta}_{c,2}) \otimes \mathbf{K}_{SE,2}(\tilde{\mathbf{x}}, \boldsymbol{\theta}_{t,2}). \quad (5.10)$$

The hyperparameters of both models were initialised randomly multiple times. The best set of hyperparameters was selected based on the lowest **NLML** value.

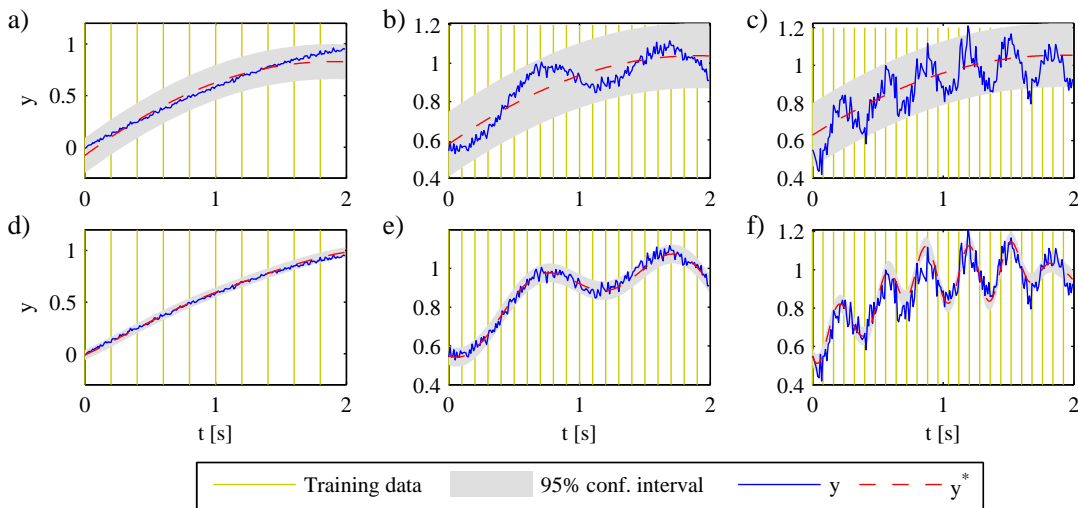
Figure 5.15.a shows the outcome of the **MTGP** model with a single correlation term. The results indicate that the model learns a negative correlation between the external and internal training data representing the periodic motion. However, the model is not able to learn the positive baseline shift. In contrast, the **MTGP<sub>LMC</sub>** model is able to learn both correlations as shown in Fig. 5.15.b. These observations are confirmed by the normalized **MTGP** correlation coefficients which are shown in Table 5.7. For **MTGP<sub>LMC</sub>**,  $\mathbf{K}_{c,1}$  is negative. According to Eq. 5.10,  $\mathbf{K}_{c,1}$  is multiplied by the quasi-periodic temporal covariance function which is suitable for model periodic motion. The second correlation term  $\mathbf{K}_{c,2}$  is positive. The matrix  $\mathbf{K}_{c,2}$  is multiplied by the squared-exponential covariance function and models the linear baseline shift. The **RMSE**, **nRMSE**, and **MSLL** values shows an increased accuracy of the **MTGP<sub>LMC</sub>** model.

### 5.5.2 Convolution of Kernels

The second **MTGP** extension focuses on the use of task-specific temporal covariance functions and hyperparameters. Using the same temporal covariance function and hyperparameters for each task reduces the number of hyperparameters. However, this could lead to the result that data with different temporal characteristics relative to the internal motion trace might not be used. This can be observed in Table 5.6 where the ACC marker was the marker used least frequently. In contrast, the ACC marker was one of the most frequently used markers in case of multivariate prediction using the **RVM** and **SVR** algorithms (see Table 4.4 and 4.5).

When introducing task-specific covariance functions  $k_t^j$  with  $j \in \{1, \dots, o\}$  and  $o$  being the number of tasks, it has to be guaranteed that Mercer's theorem is still fulfilled. This can be achieved by a convolution of two covariance functions as discussed in [111]. The result of two convoluted covariance function is again a valid covariance function. This idea has been investigated in [171]. The authors presented various convoluted covariance functions such as the convolution of two squared-exponential functions

$$k_{SE \times SE}(x, x', l, l') = \sqrt{\frac{2\theta_L(l)\theta_L(l')}{\theta_L(l)^2 + \theta_L(l')^2}} \exp\left(-\frac{r^2}{\theta_L(l)^2 + \theta_L(l')^2}\right) \quad (5.11)$$



**Figure 5.16:** (a-c) Prediction  $y^*$  and 95 % confidence interval of an **MTGP** model with one squared-exponential covariance function for tasks one to three, respectively; (d-f) Prediction  $y^*$  and 95 % confidence interval of an **MTGP** model with three task-specific squared-exponential covariance functions for task one to three, respectively.

**Table 5.8:** Hyperparameter  $\theta_L$ , RMSE, nRMSE, and MSLL of the three task modelled for an MTGP model with one and three squared-exponential covariance functions, respectively.

Task	1	2	3
MTGP with one SE cov. func.			
$\theta_L[s]$		3.38	
RMSE	0.048	0.068	0.098
nRMSE	0.169	0.425	0.424
MSLL	-0.421	0.175	0.2
MTGP with three SE cov. func.			
$\theta_L[s]$	1.182	0.715	0.125
RMSE	0.017	0.023	0.068
nRMSE	0.061	0.144	0.144
MSLL	-1.083	-0.496	-0.414

with  $r = \|x - x'\|_2$  and  $\theta_L(l)$  being the x-scaling hyperparameter of task  $l$ . Note for  $l = l'$ , Eq. 5.11 is equivalent to a “normal” squared-exponential function (Eq. 3.37). The temporal covariance function in Eq. 5.2 will be replaced by Eq. 5.11

$$k_{\text{MTGP}}(x, x', l, l') = k_c(l, l') \times k_{SE \times SE}(x, x', l, l'). \quad (5.12)$$

**MTGP Example V** The use of convoluted kernels will be illustrated in another example using synthetic data. In this example three tasks were considered with different temporal characteristics. The tasks are illustrated in Fig. 5.16.a to 5.16.c, respectively. They share an increasing long-term trend. However, they differ in their short-term behaviour. Tasks two and three were generated by adding a periodic signal and Gaussian noise to task one, respectively. Two MTGP models were trained, one MTGP model with a single squared-exponential covariance function for all tasks and one with three task-specific squared-exponential covariance functions. Furthermore, the tasks were generated to have task-specific sampling frequencies which were  $f_s^1 = 5$  Hz,  $f_s^2 = 10$  Hz and  $f_s^3 = 12.5$  Hz. The training data is illustrated in Fig. 5.16. The test data are the residual points in the time frame  $t \in (0\text{ s}, 2\text{ s}]$ .

The prediction  $y^*$  for tasks one to three using an MTGP model with a single covariance function is shown in Fig. 5.16.a - 5.16.c. The common temporal hyperparameter is

$\theta_L = 3.38$  s as shown in Table 5.8. The algorithm is not able to model the short-term characteristics which results in increased **RMSE**, **nRMSE**, and **MSLL** values. In contrast, using task-specific covariance functions adds two further temporal  $x$ -scaling hyperparameters  $\theta_L$ , which vary between 0.15 s for task three and 1.18 s for task one. The outcome of the **MTGP** model with task-specific covariance functions is shown in Fig. 5.16.d to 5.16.f and indicates an increased prediction accuracy. Table 5.8 confirms a decreased **RMSE**, **nRMSE**, and **MSLL** values for all tasks.

## 5.6 Conclusion

This chapter focused on the multivariate extension of the **GP** model, which was presented in sec. 3.2.2. The proposed **MTGP** approach is an efficient possibility to combine prediction and correlation algorithms and solve both problems within one model. To the author's knowledge, this is the first combined **MC** algorithm which makes use of all available internal and external training data even if they are sampled at different frequencies. Previously presented combined **MC** algorithms [123, 184] can only use external and internal data as training pairs if both were sampled at the same time instance. As a consequence, the majority of the external data sampled at higher frequencies will be ignored. In contrast, **MTGP** models also consider the time  $t$  at which each internal and external observation was acquired. Hence, all data can be used to learn the correlation within and between the tasks.

The proposed **MTGP** tracking algorithm offers several advantages compared to alternative prediction and correlation algorithms:

1. The model can use signal-specific training observations which might be sampled at arbitrary time instances. The low mean **RMSE** and **nRMSE** of experiment E2 indicate that the **MTGP** models can predict the internal motion even though only  $n_{int} = 10$  internal and  $n_{ext} = 300$  external training points are available.
2. Arbitrary prediction latencies can be selected, independently from the sampling frequency. In the experiments E2 and E3, a 0 to 4 sample look-ahead prediction was only selected to make a comparison to the alternative algorithms possible. Furthermore, multiple latencies can be computed without retraining the model. This might be useful in treatment systems with two motion compensation systems such as a **MLC** and a robotic patient couch.
3. The prediction result of the **MTGP** model is a Gaussian distribution. Similar to the experiments in sec. 3.4, the predicted variance can be used as real-time feedback

system during the treatment (see sec. 3.4).

4. The **MTGP** framework is very flexible and can be easily extended to consider multiple external markers as experiment E3 indicates.
5. In case of multivariate settings, the **MTGP** approach is robust against a breakdown of one or multiple sensors. In case of losing the connection to one of the external sensors, the alternative approaches would require retraining of the models. In contrast, an **MTGP** model is able to automatically compensate the loss.
6. **MTGP** models learn the correlation between internal and external motion traces. The results of experiment E1 indicate mean correlation differences between the **OMs** and the **ILM** of up to 0.22 over time. Assuming frequently acquired internal and external data throughout the treatment, variations in the correlation can automatically be detected and compensated by **MTGPs**.

Referring to the questions of sec. 1.3, the following conclusions can be drawn:

- **Q.1.2:** How to select the most relevant and least redundant markers?

This problem is more challenging in case of correlation algorithms than for prediction algorithms due to the low number of internal training points  $n_{int}$ . Each marker selection criterion will be strongly influenced by the selected internal points. In experiments E2 and E3, the **NLML** value of the trained **MTGP** model was investigated as a criterion. The results indicate that the **NLML** value might be used in case of multiple univariate markers such as multiple **OMs**. However, the criterion seems to be biased towards certain modalities. In experiment E3, the criterion selected the flow marker for each subject. It has to be investigated further how this marker dependency can be compensated.

- **Q.1.3 + Q.1.4:** Can the accuracy of respiratory motion prediction and correlation be increased using a multivariate external sensor setup?

As the **MTGP** model is used in a combined setting for correlation and prediction, the questions **Q.1.3** and **Q.1.4** will be discussed together. The mean **RMSE** and **nRMSE** values of experiment E3 point out that the **SVR-wLMS** algorithm and the **MTGP** model can be decreased by using multiple external markers in case of a combined correlation and prediction setting. In case of  $h = 0$  ms, no prediction is performed and both algorithms are used as correlation algorithms. The mean **RMSE**

of the **SVR-wLMS** model decreases from 2.06 mm to 1.85 mm and for the  $MTGP_{SE}$  approach from 2.05 mm to 1.85 mm if three instead of one external marker is considered. These results are in agreement with [83]. For respiratory motion prediction, no independent conclusion can be drawn from experiment E3. For the **MTGP** approach and  $h > 0$  ms, the decreased **RMSE** and **nRMSE** values might be caused by an increased correlation accuracy and not by an increased prediction accuracy. Note, the increased accuracy for  $h > 0$  ms in case of the **SVR-wLMS** algorithm is only caused by an increased correlation accuracy. The **SVR** algorithm, which is used as correlation algorithm, computes a one-dimensional internal motion trace which is used as input for the **wLMS** prediction algorithm. However, it has already been shown in sec. 4.4.2 that the prediction accuracy of the **wLMS** algorithm can be increased by using multiple external markers.

- **Q.1.5:** What are the most relevant sensors or sensor combinations?

The most relevant external sensors are the **OMs** for the **DP-wLMS**, the **SVR-wLMS** and the **MTGP** approaches. This is in agreement with the high correlation coefficients between the **OMs** and the **ILM** shown in experiment E1. The **ACC** marker is the least relevant sensor in case of the **MTGP** approach. This is due to the different temporal characteristics compared to the internal motion trace. A possible solution is the use of task-specific covariance functions as discussed in sec. 5.5.2. Comparing the selected marker in case of using an **SVR** method as correlation algorithm or as prediction algorithm (see sec. 4.4.2) results in different relevance of the strain, flow, and **ACC** marker. The least relevant marker for **SVR** prediction is the flow marker according to Table 4.4. However, in case of the **SVR** correlation algorithm, the strain and **ACC** marker seem to be the least relevant markers. Nonetheless, it has to be considered that only nine motion traces have been evaluated and a strong dependency on the breathing pattern as well as on the position of the **ILM** can be expected.

## 6 Conclusion and Future Work

The primary objective of this work was to increase the accuracy of respiratory adaptive motion compensation techniques. As discussed within sec. 1.1.3, these techniques are relevant for various medical applications such as medical imaging, image-guided interventions, or radiotherapy. Within this work, we focused exclusively on motion compensation in radiotherapy. However, we want to emphasize that the investigated algorithms and presented results can be transferred to other applications. The algorithms are not restricted to respiratory motion. As discussed in sec. 1.1.3, respiratory motion is one source of involuntary motion. The investigated **RVM** or **GP** algorithms can also be applied to compensate cardiac or tremor motion.

To improve **MC** algorithms, two approaches have been investigated

1. extension from uni- to multivariate **MC** algorithms (**Q.1**), and
2. improved accuracy by real-time feedback about the current prediction error (**Q.2**).

These approaches were specified by the questions proposed in sec. 1.3.

The extension to multivariate algorithms was motivated by two observations. First, optical markers are used as external surrogates in clinical systems such as the CyberKnife<sup>®</sup> [134]. However, studies have indicated that the correlation between external optical markers and the internal fiducials depends on the placement of the external markers and the breathing pattern [282]. As respiratory motion patterns have a high variability [205], the initial correlation is most likely to change during the treatment time. Second, the literature review in chapter 2 revealed that optical tracking is used almost exclusively as external surrogate modality. For example in case of prediction algorithms, none of the 45 considered publications investigated the potential of alternative modalities. In contrast, we presented in sec. 1.1.2 that respiratory activity can be measured by different sensor modalities at various body positions. These observations suggest that an algorithm considering either multiple univariate or multivariate external surrogates should reduce the dependency of the marker placement and consequently increase the robustness and the accuracy of **MC** algorithms.

The second approach investigated, the potential of a real-time feedback system, can

be implemented in its simplest form by monitoring the current prediction or correlation error. However, this error can only be computed after acquiring the true value. In case of prediction algorithms, this means that the current prediction error is delayed by the latency. Ruan *et al.* [220] proposed a possibility to overcome these limitations by exploiting the predicted variance of the probabilistic KDE algorithm. However, the results of the comprehensive study of Krauss *et al.* [142] showed that the prediction accuracy of the KDE approach is inferior to alternative algorithms such as SVR or neuronal networks.

To combine both approaches, the investigation of alternative probabilistic MC algorithms was required. In a first step of this work, the performance of these algorithms was evaluated relative to alternative algorithms on a comprehensive dataset (sec. 6.1). In the following, these algorithms were used to address question Q.1 and Q.2 whose main results are summarized in sec. 6.2 and 6.3, respectively. The investigation of multivariate GP models led to MTGP models which are the first combined prediction and correlation model. The differences of a MTGP model compared to the currently used CyberKnife<sup>®</sup> Synchrony model are highlighted in sec. 6.4. This section contains also a discussion about the potential of MTGP models in future adaptive motion compensation techniques (in particular the use of US tracking). An summary and outlook of MTGP models in general is presented in sec. 6.5.

## 6.1 Probabilistic Motion Compensation Algorithms

To evaluate the benefit of the predicted variance, we introduced two new probabilistic algorithms (RVM and GP models) for the problem of respiratory motion compensation. After discussing the mathematical background of the RVM and the GP approach, both algorithms were tested on a large dataset for the purpose of respiratory motion prediction (sec. 3.3). The algorithms were evaluated on a comprehensive dataset and compared to six previously published algorithms for different prediction latencies. On average, the results indicate a superior performance of the linear RVM algorithm for all investigated prediction latencies (Tables 3.8 - 3.11). The two investigated GP models could outperform all previously evaluated algorithms for  $h = \{77, 115, 154\}$  ms. For long prediction horizons ( $h = 308$  ms), the performance is worse compared to the wLMS and SVR algorithms. To highlight these findings, the  $RMSE_{rel}$  differences between the wLMS (the best algorithm so far) and the  $RVM_{lin}$ , and the  $GP_{QRSN}$  are shown in Fig. 6.1 for  $h = 115$  ms. 92.43% of the data for  $RVM_{lin}$  and 78.62% for  $GP_{QRSN}$  can be predicted more accurately or equally well as the wLMS algorithm. These results show that the RVM and GP approach can be used as highly accurate respiratory motion prediction

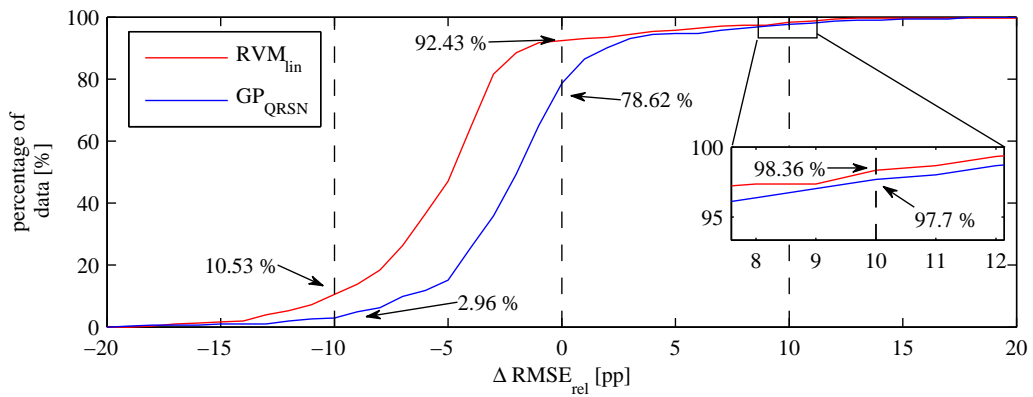


methods.

For the case of correlation algorithms, only the **GP** algorithm and in particular its extension to **MTGPs** has been investigated (whose main contributions will be summarized in sec. 6.4). The **RVM** algorithm can be used in the same manner as the **SVR** correlation algorithm presented by Ernst *et al.* [83]. An **RVM** model can be trained which uses the following input features: the position, the velocity, the acceleration, and a binary flag (indicating inspiration or expiration) of one or multiple external surrogates. However, this approach was not further investigated as the problem of training such an algorithm with a small number of training points remains difficult. Furthermore, algorithms requiring a binary flag indicating inspiration or expiration seem practically less robust, as this might not be classified correctly in real-time. The alternatively investigated solution of using an **MTGP** model (sec. 6.4) seemed to be more promising as it has several advantages since it does not require a binary flag and can be used for combined motion prediction and correlation. It has to be further investigated how a potential **RVM** correlation algorithm (as well as the **SVR** approach of [83]) could be put into practise.

## 6.2 Univariate versus Multivariate

The research questions focusing on multivariate **MC** algorithms were grouped under question **Q.1** whose main results will be discussed here. Within this work, multivariate extensions were considered for prediction and correlation algorithms. The investigation was performed in several steps.



**Figure 6.1:** Comparative performance of a linear **RVM** (red) and **GP** (blue) algorithm with respect to the **wLMS** algorithm. Diagram shows the cumulative histogram of the relative **RMSE** difference over 304 motion traces.

First, Pearson's correlation coefficients were computed between data of different external (sec. 4.4.1), and external and internal sensors (sec. 5.4.1). In both experiments, the highest mean correlation over all subjects was achieved by the data of one of the external optical markers. In case of correlation between external and internal data, the highest mean absolute correlation was  $|r_{OM2,ILM}| = |r_{OM3,ILM}| = 0.95$  (between optical marker two and three, and the internal landmark). The generally high correlation is in agreement with previous publications [12, 112, 140] and confirms the use of optical tracking in current clinical univariate systems. Nonetheless, the results also indicate a correlation dependency on the marker placement (Fig. 4.7 and 5.10.b). Furthermore, data acquired by alternative modalities such as flow or strain also have high correlation values which vary between  $0.54 \leq |r| \leq 0.93$ . This indicates that alternative modalities could be useful for adaptive motion compensation. The highest mean absolute correlation could be achieved by the strain sensor for both measurements (for external-external  $r_{Strain,OM2} = 0.93$ ; for external-internal  $r_{Strain,ILM} = 0.87$ ).

The long measurement duration of M1 (sec. 4.1) allowed the investigation of correlation variations. These variations were higher in the case of data of multiple external sensors (up to a maximum of  $|\Delta r_{ACC,OM1}| = 0.39$ ; between minutes one and twelve) than for data of external and internal sensors (up to a maximum of  $|\Delta r_{OM1,ILM}| = 0.23$ ; between minutes one and twelve). These results and the high standard deviation point out a strong variation throughout the long treatment session. The practical conclusion of these initial findings is that the correlation and the relevance of external sensors is likely to change over time and that MC algorithms have to be adaptive over time.

In a following step, multivariate prediction and correlation algorithms have been investigated. Thereby, probabilistic and non-probabilistic algorithms were applied. The results confirm that the prediction and the correlation accuracy can be increased by using a multivariate setting (Tables 4.3 and 5.5). Applying multivariate prediction algorithms, the averaged RMSE of the RVM algorithm (which outperformed all investigated non-probabilistic approaches) decreased from 0.15 mm to 0.12 mm in the case of normal breathing and from 0.52 mm to 0.46 mm in case of irregular breathing. Similar results were found for multivariate correlation algorithms (decrease of the RMSE by 0.2 mm for the SVR-wLMS and the MTGP algorithms; Table 5.5).

For completeness it has to be mentioned that the dosimetric effect through an increased prediction and correlation accuracy has not been evaluated in this work. This very relevant topic is part of ongoing studies such as [26, 27].

One essential problem for multivariate approaches is the selection of the optimal sensor combination. The results indicate that different sensor combinations should be used

depending on the algorithm and the purpose. For example, in the case of the **SVR** prediction algorithm, the **ACC** sensor was frequently selected while the flow sensor was never used (Table 4.4). The opposite was found for the **SVR**-based correlation algorithm (Table 5.6). This makes a general conclusion about the most relevant alternative modality difficult.

Consequently, sensor selection criteria are required. Here, a sequential forward selection criterion based on the  $RMSE_{rel}$  (for prediction) and on the **NLML** value (for correlation) was investigated. Even though this criterion was useful for these initial studies, the relevance for clinical practise is very limited. First, the partly high computation costs limit the use for real-time systems. Second, using the criterion based on the **NLML** value seems to be biased toward certain modalities. Third, the criterion was only used once during the initial training phase. Correlation variations over time were not considered. Consequently, further improvements are possible if an adaptive criterion would be used (see differences between **TSUK** and **TSK** in Table 4.3).

The results of this work clearly show that multivariate settings can increase the prediction and correlation accuracy in adaptive motion compensation. In principle, additional sensors such as a thermistor (to measure the flow) or a strain belt should be easy to integrate into a current clinical motion compensation system. However, further research about sensor selection methods needs to be done (especially for correlation algorithms) to make such a system usable in clinics and treatment centres.

### 6.3 Real-Time Feedback

This section summarizes the main findings regarding question **Q.2**. The potential benefit of using the predicted variance as real-time feedback was evaluated on the **RVM** algorithm for the purpose of respiratory motion prediction. Equivalent investigations are possible with the **GP** algorithm. The analysis of the predicted variance revealed a characteristic variance pattern (see Fig. 3.13.b). This pattern points out that monitoring the variance  $\sigma_{i+\xi}^{2*}$  suffers from the same problem as monitoring the current prediction error  $e_i$ . In both cases, an increased error or variance would be delayed by the prediction latency (see Fig. 3.13.b)

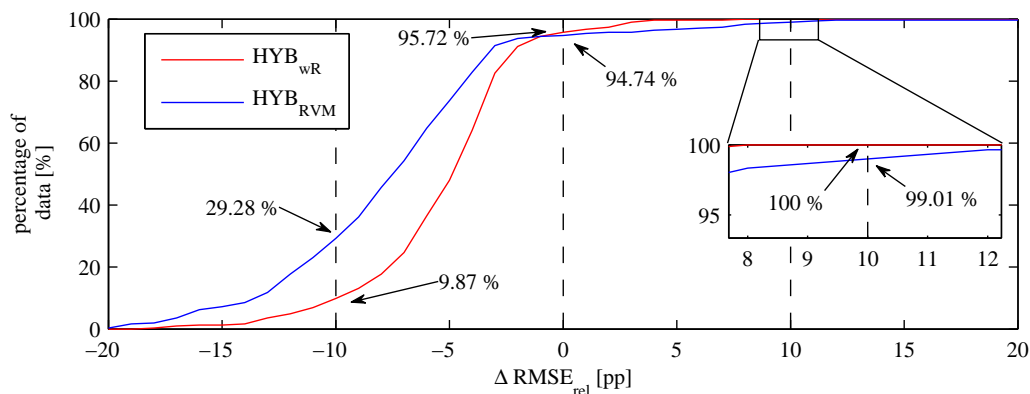
Nonetheless, further investigations showed that by monitoring the variance  $\sigma_{i+\xi}^{2*}$  instead of the current prediction error  $e_i$  results in a superior control of the prediction error  $e_{i+\xi}$ . In a first experiment, the treatment was interrupted when the variance exceeded a specific threshold variance. The results point out that by reducing the variance threshold, the average error on the remaining data decreases (Fig. 3.14). A similar evaluation was

performed by interrupting the treatment when the current prediction error exceeded an error threshold. By reducing the error threshold, the average error on the remaining data reaches a minimum which is higher than the error by using the variance criterion (compared Fig. 3.14 and Fig. 3.15).

In a second experiment the variance was used as a feature for potentially relevant hybrid algorithms. Two kind of hybrid algorithms have been investigated: a combination of non-probabilistic and probabilistic algorithms, and of multiple probabilistic algorithms. Especially the latter is of high practical relevance as it does not require additional parameters and can be extended to an arbitrary number of probabilistic algorithms. In contrast, the first hybrid approach is limited to two algorithms. Figure 6.2 shows the  $\Delta\text{RMSE}_{rel}$  of a hybrid algorithm consisting of the  $w\text{LMS}$  and a linear  $\text{RVM}$  algorithm ( $\text{HYB}_{wR}$ ), and of a hybrid algorithm consisting of three  $\text{RVM}$  algorithms with different basis functions ( $\text{HYB}_{\text{RVM}}$ ) for  $h = 115$  ms. Using one of the hybrid approaches results in a superior performance compared to using the algorithms individually (compare to Fig. 6.1).

The main difference between  $\sigma_{i+\xi}^2$  and  $e_i$  is that the predicted variance  $\sigma_{i+\xi}^2$  depends on the test features  $\mathbf{x}_i^*$  and the features in the training set  $\mathbf{T}$ . In contrast, the current prediction error  $e_i$  is computed based on the predicted label  $y_i^*$  and the observed position  $y_i$ . In the case of the predicted variance, this leads to a "memory effect" (sec. 3.4.1). Depending on the size of the feature vector and/or the training set, the predicted variance remains high if a motion artefact occurred in "recent" history.

This difference between the two measures can be very relevant for the purpose of cor-



**Figure 6.2:** Cumulative histogram of the relative RMSE difference for the best  $\text{HYB}_{wR}$  (with linear  $\text{RVM}$  basis function and a variance threshold of  $\sigma_{th}^2 = 0.025 \text{ mm}^2$ ) and the  $\text{HYB}_{\text{RVM}}$  algorithm over 304 motion traces (difference computed with respect to the  $w\text{LMS}$  algorithm).

relation algorithms. In this scenario, only a few internal fiducial positions are known while the external data are sampled frequently. The current prediction error cannot be monitored as the internal observations  $y_i^{int}$  are not observed frequently. In contrast, the predicted variance can be computed at any time index.

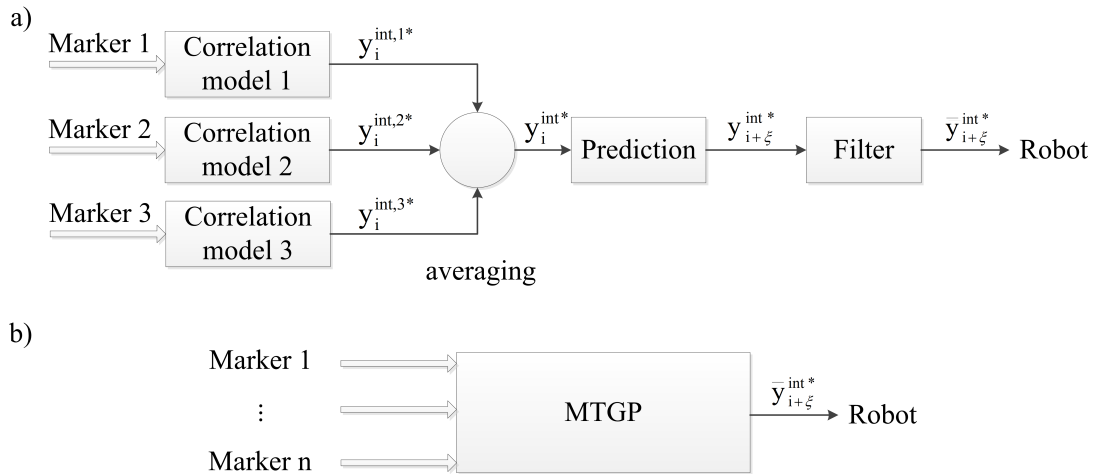
Consequently, the results presented here should be interpreted as a preliminary investigation showing that the variance can be used as a feature for a real-time feedback system. Further investigations have to be performed to evaluate the variance for the purpose of correlation algorithms.

## 6.4 MTGP Models in Radiotherapy

Even though it was not the primary focus of this work, the extension of GP models to MTGP models revealed the first combined prediction and correlation model which makes use of all available training data. Previously published combined algorithms [123, 184] can only use external and internal data as training pairs if both were acquired at the same time instance (or delayed by the system latency). The majority of the external data, which contains relevant information about the breathing characteristics, is ignored.

The aim of using one model for both problems is to increase the accuracy of the complete system. Using two independent models for prediction and correlation requires two training processes, which can be sources of errors. Further, the algorithms are used in a sequence, meaning that the output of the first algorithm is the input of the second algorithm. As both algorithms will not be able to perfectly model the true physiological process (due to measurement errors and unintended movements), errors of the first algorithm can be amplified by the second. Such effects can be avoided by using a single model.

To highlight the difference and the relevance of this approach, a potential MTGP workflow is compared to the workflow of the CyberKnife<sup>®</sup> Synchrony system [134, 227]. The complete CyberKnife<sup>®</sup> Synchrony workflow is illustrated in Fig. 6.3.a. Currently, up to three external optical markers are used. A correlation model between the internal position and the data of each external marker is trained. A linear, quadratic, or dual-quadratic model can be selected as correlation model to compensate for potential phase shifts between internal and external motion. To build these models, Sayeh *et al.* [227] recommend to acquire 5-6 internal data points for a linear model and 10-12 points for a dual quadratic model. The input of the prediction algorithm  $y_i^{int*}$  is the averaged output of all correlation models  $y_i^{int,1*}$  to  $y_i^{int,3*}$ . The result of the prediction algorithm  $y_{i+\xi}^{int*}$  is passed through a filter to smooth the signal before sending it to the robot. An alternat-



**Figure 6.3:** Schematic diagram of adaptive motion compensation for (a) the CyberKnife® Synchrony system (illustration based on [227]) and (b) an alternative approach based on **MTGP** models.

ive **MTGP** workflow which is capable of performing the same task is illustrated in Fig. 6.3.b for  $n$  external markers. Table 6.1 highlights the differences and similarities between the two approaches. The cells of the table are highlighted light green or red to indicate advantage or disadvantage of one of the approaches relative to the other. The cells are highlighted light yellow if the approaches are comparable. The classification is based on the author’s opinion.

**Table 6.1:** Comparison of the adaptive motion compensation techniques using either the CyberKnife® Synchrony or an **MTGP**-based workflow.

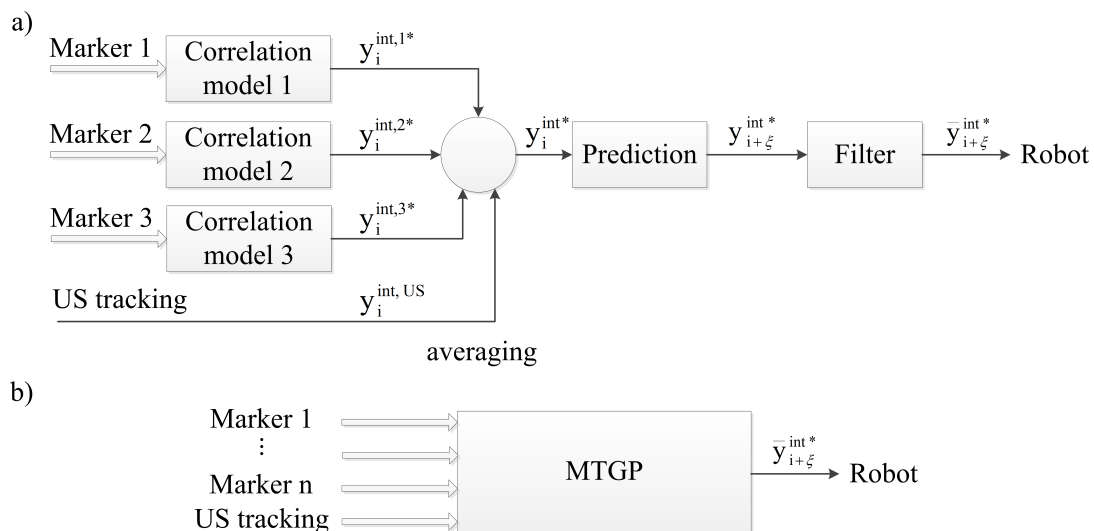
CyberKnife® Synchrony	MTGP
<b>number of models</b>	
two models (prediction + correlation); one smoothing filter	one model
<b>prediction latencies</b>	
one fixed latency (an integer multiple of the sampling frequency)	prediction of multiple numbers of latencies (independent of the sampling frequency)
<b>filtering</b>	

a smoothing filter is used to compensate jitter and unnatural jumps (caused by the prediction or correlation model)	a smoothed prediction result can be achieved by restricting the $x$ -scaling hyperparameter $\theta_L$ (see Eq. 3.37 - 3.39); an additional filter is not required
<b>compensation of phase shift between external and internal data</b>	
use of a dual quadratic correlation model	use of additional time shift hyperparameters (see sec. 5.2.3)
<b>number of internal data points required</b>	
10 – 12 for a dual quadratic correlation model [227]	10 points are required according to Fig. 5.13
<b>use of multiple univariate sensors</b>	
the current model uses up to three external optical markers and can be extended to more (note, all external markers are weighted equally due to averaging, independent of their correlation to the internal target)	the model can use multiple univariate markers (Table 5.4); each external marker is weighted based on its correlation to the internal target
<b>use of a multivariate setting</b>	
in principle possible; however most likely not beneficial due to averaging and high correlation variations between modalities (see sec. 5.4.1)	is possible and beneficial (Table 5.5); each marker is weighted based on its correlation
<b>effect of a short or permanent breakdown of one external sensor</b>	
the averaging will be performed over two instead of three markers; this can lead to jumps within the prediction	no or damped jumps (the MTGP model considers previous data of the affected marker, smoothing effect of temporal covariance function)
<b>common sampling frequency</b>	
it is assumed that internal and external data are acquired simultaneously; however external markers are tracked sequentially; delay is not considered	each marker can be sampled at different time instances; approach not limited to a common sampling frequency
<b>real-time feedback of current prediction accuracy</b>	

feedback is only possible when internal points are acquired (additional imaging dose)

real-time feedback is possible due to the probabilistic nature of **MTGP** (evaluation of the predicted variance; see sec. 3.4.1); further experiments have to be performed investigating variance controlled correlation algorithms

In recent years, several studies focussed on the use of ultrasound for target tracking in the case of prostate radiotherapy [15, 149, 228]. Ongoing research effort is used to evaluate the possibility of using **US** guidance for other targets such as in the liver [36, 83]. The advantage of such systems is that the internal tumour position can be tracked at a high frequency without additional negative side effects for the patient. Within these developments, **MTGP** models provide a useful tool. In the optimal case, the internal target can be tracked through the entire treatment with **US**. However, as stated earlier (sec. 5.1), the internal target might not be visible at each time instance. Reasons can be high image noise, a limited field of view, or **US** shadows (see Fig. 1.5). In these situations, the tumour position has to be predicted based on external markers and previously observed internal positions. Consequently, hybrid systems combining **US** guidance and external motion correlation are required. One simple example of such a system is illustrated in



**Figure 6.4:** Schematic diagram of a hybrid motion compensation system consisting of external markers and **US** data based on (a) an extension of the CyberKnife® Synchrony approach and (b) on an **MTGP** model.



Fig. 6.4.a which is based on the Synchrony system (proposal of author). If the internal target position acquired by US is visible,  $y_i^{int,US}$  is averaged with the position computed by the correlation models. However, this approach assumes that  $y_i^{int,US}$  is sampled at the same time instance as the external markers. This might not be the case due to the acquisition system. Furthermore, the computation of  $y_i^{int,US}$  might suffer from additional time latencies due to image processing times and network latencies. Consequently, a hybrid approach such as in Fig. 6.4.a cannot be used. In contrast, these requirements are no obstacle for an MTGP approach. As illustrated in Fig. 6.4.b, the US data are just another task, which is sampled at its own frequency and is included in the model if visible. The high flexibility makes the MTGP approach an optimal choice for future applications.

## 6.5 MTGP Models in General

As shown in the previous section, MTGP models have several useful properties which might be relevant for alternative applications. These properties can be summarized as followed:

- The MTGP model can use signal-specific training observations.
- The training observations can be sampled at arbitrary time instances (they do not have to be sampled equidistantly).
- The model can simultaneously predict multiple prediction latencies without re-training.
- The prediction latencies can be selected independently of the sampling frequency.
- The correlation between sensors is learned automatically and can be used for further investigations.
- The MTGP framework can incorporate either multiple univariate sensors or be used in a multivariate setting.
- The predicted result of a task and at a specific time instant will be a Gaussian distribution. The predicted variance can be used as feedback about the “confidence” of the algorithm in the predicted value.
- Depending on the application and prior knowledge about the signal characteristics, different temporal covariance functions (or a superposition of them) can be applied.
- The method is robust against short or permanent breakdown of one input signal.

In general, **MTGPs** can be used for all kind of multivariate time-series problems such as the analysis of financial time-series [13], environmental sensor networks [195], or compiler performances [30]. In the field of biomedical engineering, one alternative applications is the modelling of vital-signs data of hospital patients [70]. Often, clinical data such as respiratory rate or heart pressure are acquired by different sensors (with different sampling frequency) or even manually by the clinical staff. **MTGP** can be used to construct robust forecasting models using all available data.

To further stimulate research, a web-page was created with an open-access **MTGP** toolbox<sup>1</sup>. The page contains further examples of **MTGP** applications such as the use of two-dimensional tasks.

---

<sup>1</sup>The **MTGP** toolbox is available online at <http://www.robots.ox.ac.uk/~davidc/>

# Bibliography

- [1] Accuray Inc., Sunnyvale, CA, USA. <http://www accuray.com/>. (visited on 2014-07-22).
- [2] Brainlab, Feldkirchen, Germany. <https://www.brainlab.com/>. (visited on 2014-07-22).
- [3] Elekta, Stockholm, Sweden. <http://www.elekta.com>. (visited on 2014-07-22).
- [4] Zentrum für Krebsregisterdaten, Robert Koch Institute, Berlin, Germany. <http://www.krebsdaten.de>. (visited on 2014-06-23).
- [5] Medspira, Minneapolis, MN, USA. <http://www.medspira.com/company.html>. (visited on 2014-07-22).
- [6] Atlas of Human Body, American Medical Association (AMA), Chicago, IL, USA. [http://www.ama-assn.org/ama/pub/physician-resources/patient-education-materials/atlas-of-human-body.page?](http://www.ama-assn.org/ama/pub/physician-resources/patient-education-materials/atlas-of-human-body.page) (visited on 2014-06-26).
- [7] Staat & Gesellschaft - Todesursachen, Statistisches Bundesamt, Wiesbaden, Germany. <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Gesundheit/Todesursachen/Todesursachen.html>. (visited on 2014-06-23).
- [8] Varian Medical Systems, Palo Alto, CA, USA. <http://www.varian.com/>. (visited on 2014-07-22).
- [9] ViewRay, Oakwood Village, Ohio, USA. <http://www.viewray.com/system>. (visited on 2014-07-22).
- [10] P. S. Addison, J. N. Watson, M. L. Mestek, J. P. Ochs, A. A. Uribe, and S. D. Bergese. Pulse oximetry-derived respiratory rate in general care floor patients. *Journal of Clinical Monitoring and Computing*, pp. 1–8, 2014. DOI: [10.1007/s10877-014-9575-5](https://doi.org/10.1007/s10877-014-9575-5).
- [11] J. R. Adler Jr, M. J. Murphy, S. D. Chang, and S. L. Hancock. Image-guided robotic radiosurgery. *Neurosurgery*, **44**(6):1299–1306, 1999.
- [12] S. Ahn, B. Yi, Y. Suh, J. Kim, S. Lee, S. Shin, S. Shin, and E. Choi. A feasibility study on the prediction of tumour location in the lung from skin motion. *British Journal of Radiology*, **77**(919):588–596, 2004. DOI: [10.1259/bjr/64800801](https://doi.org/10.1259/bjr/64800801).
- [13] M. Alvarez, D. Luengo-Garcia, M. Titsias, and N. Lawrence. Efficient multioutput gaussian processes through variational inducing kernels. In *Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 25–32, 2011.
- [14] P. Arnold, F. Preiswerk, B. Fasel, R. Salomir, K. Scheffler, and P. Cattin. Model-based respiratory motion compensation in MRgHIFU. In *Information Processing in Computer-Assisted Interventions*, no. 7330 in Lecture Notes in Computer Science, pp. 54–63. Springer Berlin Heidelberg, 2012. DOI: [10.1007/978-3-642-30618-1\\_6](https://doi.org/10.1007/978-3-642-30618-1_6).

- [15] X. Artignan, M. H. P. Smitsmans, J. V. Lebesque, D. A. Jaffray, M. van Her, and H. Bartelink. On-line ultrasound image guidance for radiotherapy of prostate cancer: impact of image acquisition on prostate displacement. *International Journal of Radiation Oncology\*Biological\*Physics*, 59(2):595–601, 2004. DOI: [10.1016/j.ijrobp.2004.01.043](https://doi.org/10.1016/j.ijrobp.2004.01.043).
- [16] D. R. Bailes, D. J. Gilderdale, G. M. Bydder, A. G. Collins, and D. N. Firmin. Respiratory ordered phase encoding (ROPE): A method for reducing respiratory motion artefacts in MR imaging. *Journal of Computer Assisted Tomography*, 9(4):835–838, 1985. DOI: [10.1097/00004728-198507010-00039](https://doi.org/10.1097/00004728-198507010-00039).
- [17] A. Bates, M. Ling, J. Mann, and D. K. Arvind. Respiratory rate and flow waveform estimation from tri-axial accelerometer data. In *2010 International Conference on Body Sensor Networks (BSN)*, pp. 144–150, 2010. DOI: [10.1109/BSN.2010.50](https://doi.org/10.1109/BSN.2010.50).
- [18] S. Bauer, B. Berkels, S. Ettl, O. Arold, J. Hornegger, and M. Rumpf. Marker-less reconstruction of dense 4-d surface motion fields using active laser triangulation for respiratory motion management. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS, pp. 414–421. Springer, 2012. DOI: [10.1007/978-3-642-33415-3\\_51](https://doi.org/10.1007/978-3-642-33415-3_51).
- [19] G. Benchetrit. Breathing pattern in humans: diversity and individuality. *Respiration Physiology*, 122(2–3):123–129, 2000. DOI: [10.1016/S0034-5687\(00\)00154-7](https://doi.org/10.1016/S0034-5687(00)00154-7).
- [20] C. D. Bianca, E. Yorke, C.-S. Chui, P. Giraud, K. Rosenzweig, H. Amols, C. Ling, and G. S. Mageras. Comparison of end normal inspiration and expiration for gated intensity modulated radiation therapy (IMRT) of lung cancer. *Radiotherapy and Oncology*, 75(2):149–156, 2005. DOI: [10.1016/j.radonc.2005.01.008](https://doi.org/10.1016/j.radonc.2005.01.008).
- [21] C. M. Bishop. *Pattern recognition and machine learning*, volume 1. Springer New York, 2006. ISBN 0387310738.
- [22] J.-P. Bissonnette, K. N. Franks, T. G. Purdie, D. J. Moseley, J.-J. Sonke, D. A. Jaffray, L. A. Dawson, and A. Bezjak. Quantifying interfraction and intrafraction tumor motion in lung stereotactic body radiotherapy using respiration-correlated cone beam computed tomography. *International Journal of Radiation Oncology\* Biological\* Physics*, 75(3):688–695, 2009.
- [23] J. Blackall, G. Penney, A. King, and D. Hawkes. Alignment of sparse freehand 3-d ultrasound with pre-operative images of the liver using models of respiratory motion and deformation. *IEEE Transactions on Medical Imaging*, 24(11):1405–1416, 2005. DOI: [10.1109/TMI.2005.856751](https://doi.org/10.1109/TMI.2005.856751).
- [24] J. M. Blackall, S. Ahmad, M. E. Miquel, J. R. McClelland, D. B. Landau, and D. J. Hawkes. MRI-based measurements of respiratory motion variability and assessment of imaging strategies for radiotherapy planning. *Physics in Medicine and Biology*, 51(17):4147, 2006. DOI: [10.1088/0031-9155/51/17/003](https://doi.org/10.1088/0031-9155/51/17/003).
- [25] O. Blanck, R. Dürichen, F. Ernst, J. Dunst, D. Rades, G. Hildebrandt, and A. Schweikard. Evaluation of a wavelet-based least mean square motion prediction algorithm for lung and liver patients. In *European Society for Radiotherapy & Oncology (ESTRO) 31, Barcelona Spain*, 2012.
- [26] O. Blanck, J. Krause, R. Dürichen, N. Andratschke, D. Rades, J. Dunst, G. Hildebrandt, A. Schweikard, and A. Schläefer. Retrospective accuracy estimation for motion compensated robotic radiosurgery of the liver. In *American Association of Physicists in Medicine (AAPM) 2012*, Charlotte, USA, 2012.
- [27] O. Blanck, J. Krause, R. Dürichen, N. Andratschke, S. Wurster, A. Kovacs, G. Gaffke, K. Bogun, D. Rades, M. Birth, J. Dunst, G. Hildebrandt, A. Schweikard, and A. Schläefer. Pilotstudie zur analyse der klinischen genauigkeit der robotergestützten radiochirurgie für lebermetastasen. In *Jahrestagung der Deutschen Gesellschaft für Radioonkologie (DEGRO)*, Wiesbaden, Germany, 2012.

- [28] M. Boegel, H. G. Hofmann, J. Hornegger, R. Fahrig, S. Britzen, and A. Maier. Respiratory motion compensation using diaphragm tracking for cone-beam c-arm CT: A simulation and a phantom study. *International Journal of Biomedical Imaging*, 2013. DOI: [10.1155/2013/520540](https://doi.org/10.1155/2013/520540).
- [29] C. F. Bolton, F. Grand'maison, A. Parkes, and M. Shkrum. Needle electromyography of the diaphragm. *Muscle & Nerve*, **15**(6):678–681, 1992. DOI: [10.1002/mus.880150608](https://doi.org/10.1002/mus.880150608).
- [30] E. V. Bonilla, K. M. A. Chai, and C. K. Williams. Multi-task gaussian process prediction. In *Neural Information Processing Systems Foundation (NIPS)*, Vancouver, BC, Canada, 2007.
- [31] A. J. Boon, K. I. Alsharif, C. M. Harper, and J. Smith. Ultrasound-guided needle EMG of the diaphragm: Technique description and case report. *Muscle & Nerve*, **38**(6):1623–1626, 2008. DOI: [10.1002/mus.21187](https://doi.org/10.1002/mus.21187).
- [32] J. Borgert, S. Kruger, H. Timinger, J. Krucker, N. Glossop, A. Durrani, A. Viswanathan, and B. Wood. Respiratory motion compensation with tracked internal and external sensors during CT-guided procedures. *Computer Aided Surgery*, **11**(3):119–125, 2006. DOI: [10.1080/10929080600740871](https://doi.org/10.1080/10929080600740871).
- [33] A. Boudewyns, M. Willems, M. Wagemans, W. De Cock, P. Van de Heyning, and W. De Backer. Assessment of respiratory effort by means of strain gauges and esophageal pressure swings: a comparative study. *Sleep*, **20**(2):168–170, 1997.
- [34] A. Brost, R. Liao, N. Strobel, and J. Hornegger. Respiratory motion compensation by model-based catheter tracking during EP procedures. *Medical Image Analysis*, **14**(5):695–706, 2010. DOI: [10.1016/j.media.2010.05.006](https://doi.org/10.1016/j.media.2010.05.006).
- [35] A. Brost, W. Wu, M. Koch, A. Wimmer, T. Chen, R. Liao, J. Hornegger, and N. Strobel. Combined cardiac and respiratory motion compensation for atrial fibrillation ablation procedures. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, **14**:540–547, 2011. DOI: [10.1007/978-3-642-23623-5\\_68](https://doi.org/10.1007/978-3-642-23623-5_68).
- [36] R. Bruder, F. Ernst, A. Schlaefer, and A. Schweikard. A framework for real-time target tracking in radiosurgery using three-dimensional ultrasound. In *Proceedings of the 25th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS)*, pp. S306–S307, Berlin, Germany, 2011.
- [37] I. Bukovsky, K. Ichiji, N. Homma, M. Yoshizawa, and R. Rodriguez. Testing potentials of dynamic quadratic neural unit for prediction of lung motion during respiration for tracking radiation therapy. In *International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6, 2010. DOI: [10.1109/IJCNN.2010.5596748](https://doi.org/10.1109/IJCNN.2010.5596748).
- [38] I. Buzurovic, T. Podder, K. Huang, and Y. Yu. Tumor motion prediction and tracking in adaptive radiotherapy. In *2010 IEEE International Conference on BioInformatics and BioEngineering (BIBE)*, pp. 273–278, 2010. DOI: [10.1109/BIBE.2010.52](https://doi.org/10.1109/BIBE.2010.52).
- [39] I. Buzurovic, K. Huang, Y. Yu, and T. K. Podder. A robotic approach to 4d real-time tumor tracking for radiotherapy. *Physics in Medicine and Biology*, **56**(5):1299–1318, 2011. DOI: [10.1088/0031-9155/56/5/005](https://doi.org/10.1088/0031-9155/56/5/005).
- [40] J. Q. Candela. *Learning with Uncertainty - Gaussian Processes and Relevance Vector Machines*. PhD thesis, Technical University of Denmark, 2004.
- [41] L. I. Cerviño, A. K. Y. Chao, A. Sandhu, and S. B. Jiang. The diaphragm as an anatomic surrogate for lung tumor motion. *Physics in Medicine and Biology*, **54**(11):3529, 2009. DOI: [10.1088/0031-9155/54/11/017](https://doi.org/10.1088/0031-9155/54/11/017).

- [42] L. I. Cerviño, S. Gupta, M. A. Rose, C. Yashar, and S. B. Jiang. Using surface imaging and visual coaching to improve the reproducibility and stability of deep-inspiration breath hold for left-breast-cancer radiotherapy. *Physics in Medicine and Biology*, **54**(22):6853, 2009. DOI: [10.1088/0031-9155/54/22/007](https://doi.org/10.1088/0031-9155/54/22/007).
- [43] L. I. Cerviño, Y. Jiang, A. Sandhu, and S. B. Jiang. Tumor motion prediction with the diaphragm as a surrogate: a feasibility study. *Physics in Medicine and Biology*, **55**(9):N221, 2010. DOI: [10.1088/0031-9155/55/9/N01](https://doi.org/10.1088/0031-9155/55/9/N01).
- [44] K. M. Chai. *Multi-task learning with Gaussian processes*. PhD thesis, University of Edinburgh, 2010.
- [45] B. Cho, Y. Suh, S. Dieterich, and P. J. Keall. A monoscopic method for real-time tumour tracking using combined occasional x-ray imaging and continuous respiratory monitoring. *Physics in Medicine and Biology*, **53**(11):2837, 2008. DOI: [10.1088/0031-9155/53/11/006](https://doi.org/10.1088/0031-9155/53/11/006).
- [46] B. Cho, P. R. Poulsen, and P. J. Keall. Real-time tumor tracking using sequential kV imaging combined with respiratory monitoring: a general framework applicable to commonly used IGRT systems. *Physics in Medicine and Biology*, **55**(12):3299, 2010. DOI: [10.1088/0031-9155/55/12/003](https://doi.org/10.1088/0031-9155/55/12/003).
- [47] B. Cho, P. R. Poulsen, A. Sawant, D. Ruan, and P. J. Keall. Real-time target position estimation using stereoscopic kilovoltage/megavoltage imaging and external respiratory monitoring for dynamic multileaf collimator tracking. *International Journal of Radiation Oncology\* Biology\* Physics*, **79**(1):269–278, 2011.
- [48] S. Choi, Y. Chang, N. Kim, S. H. Park, S. Y. Song, and H. S. Kang. Performance enhancement of respiratory tumor motion prediction using adaptive support vector regression: Comparison with adaptive neural network method. *International Journal of Imaging Systems and Technology*, **24**(1):8–15, 2014. DOI: [10.1002/ima.22073](https://doi.org/10.1002/ima.22073).
- [49] L. Clifton, D. Clifton, M. Pimentel, P. Watkinson, and L. Tarassenko. Gaussian process regression in vital-sign early warning systems. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6161–6164, San Diego, USA, 2012. DOI: [10.1109/EMBC.2012.6347400](https://doi.org/10.1109/EMBC.2012.6347400).
- [50] S. Dasgupta, J. Wansapura, P. Hariharan, R. Pratt, D. Witte, M. R. Myers, and R. K. Banerjee. HIFU lesion volume as a function of sonication time, as determined by MRI, histology, and computations. *Journal of biomechanical engineering*, **132**(8):081005, 2010.
- [51] L. Davenport. *Synchronisation und Evaluation von multimodalen Sensoren zur Bewegungskompensation in der allgemeinen Strahlentherapie*. Bachelor thesis, Institute for Robotics and Cognitive Systems, Universität zu Lübeck, Lübeck, Germany, 2013.
- [52] S. C. Davies, A. L. Hill, R. B. Holmes, M. Halliwell, and P. C. Jackson. Ultrasound quantitation of respiratory organ motion in the upper abdomen. *Ultrasound*, **67**(803), 1994.
- [53] P. Dehkordi, M. Marzencki, K. Tavakolian, M. Kaminska, and B. Kaminska. Validation of respiratory signal derived from suprasternal notch acceleration for sleep apnea detection. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3824–3827, 2011. DOI: [10.1109/IEMBS.2011.6090950](https://doi.org/10.1109/IEMBS.2011.6090950).
- [54] T. Depuydt, O. Haas, D. Verellen, S. Erbel, M. De Ridder, and G. Storme. Geometric accuracy evaluation of the new VERO stereotactic body radiation therapy system. In *UKACC International Conference on CONTROL*, pp. 259–264, Coventry, UK, 2010. ISBN 978-1-84600-038-6. DOI: [10.1049/ic.2010.0291](https://doi.org/10.1049/ic.2010.0291).

- [55] T. Depuydt, D. Verellen, O. Haas, T. Gevaert, N. Linthout, M. Duchateau, K. Tournel, T. Reynders, K. Leysen, M. Hoogeman, G. Storme, and M. D. Ridder. Geometric accuracy of a novel gimbals based radiation therapy tumor tracking system. *Radiotherapy and Oncology*, **98**(3):365–372, 2011. DOI: [10.1016/j.radonc.2011.01.015](https://doi.org/10.1016/j.radonc.2011.01.015).
- [56] S. Dieterich and Y. Suh. Tumor motion ranges due to respiration and respiratory motion characteristics. In *Treating Tumors that Move with Respiration*, pp. 3–13. Springer Berlin Heidelberg, 2007. DOI: [10.1007/978-3-540-69886-9\\_1](https://doi.org/10.1007/978-3-540-69886-9_1).
- [57] W. D’Souza, K. Malinowski, and H. Zhang. Machine learning for intra-fraction tumor motion modeling with respiratory surrogates. In *International Conference on Machine Learning and Applications (ICMLA)*, pp. 463–467, Miami Beach, USA, 2009. DOI: [10.1109/ICMLA.2009.56](https://doi.org/10.1109/ICMLA.2009.56).
- [58] W. D. D’Souza and T. J. McAvoy. An analysis of the treatment couch and control system dynamics for respiration-induced motion compensation. *Medical Physics*, **33**(12):4701–4709, 2006. DOI: [10.1118/1.2372218](https://doi.org/10.1118/1.2372218).
- [59] W. D. D’Souza, S. A. Naqvi, and C. X. Yu. Real-time intra-fraction-motion tracking using the treatment couch: a feasibility study. *Physics in Medicine and Biology*, **50**(17):4021–4033, 2005. DOI: [10.1088/0031-9155/50/17/007](https://doi.org/10.1088/0031-9155/50/17/007).
- [60] M. L. Duiverman, L. A. van Eykern, P. W. Vennik, G. H. Koeter, E. J. W. Maarsingh, and P. J. Wijkstra. Reproducibility and responsiveness of a noninvasive EMG technique of the respiratory muscles in COPD patients and in healthy subjects. *J Appl Physiol*, **96**(5):1723–1729, 2004. DOI: [10.1152/jappphysiol.00914.2003](https://doi.org/10.1152/jappphysiol.00914.2003).
- [61] R. Dürichen, M. A. Pimentel, L. Clifton, A. Schweikard, and D. A. Clifton. Multi-task gaussian processes for multivariate physiological time-series analysis. *Transactions on Biomedical Engineering*, (in review).
- [62] R. Dürichen, T. Wissel, and A. Schweikard. Efficient SVR model update approaches for respiratory motion prediction. In *Deutsche Gesellschaft für Computer- und Roboterassistierte Chirurgie (CURAC)*, Düsseldorf, Germany, 2012. <https://www.curac.org>.
- [63] R. Dürichen, T. Wissel, and A. Schweikard. Prediction of respiratory motion using wavelet based support vector regression. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Santander, Spain, 2012. DOI: [10.1109/MLSP.2012.6349742](https://doi.org/10.1109/MLSP.2012.6349742).
- [64] R. Dürichen, R. Bruder, L. Davenport, T. Wissel, F. Ernst, and A. Schweikard. Correlation and variation of a multi-modal sensor setup for respiratory motion prediction and correlation. In *55th Annual Meeting & Exhibition of AAPM*, pp. 186, Indianapolis, USA. American Association of Physicists in Medicine (AAPM). Published in *Medical Physics*, **40**:186, 2013. DOI: [10.1118/1.4814365](https://doi.org/10.1118/1.4814365).
- [65] R. Dürichen, L. Davenport, R. Bruder, T. Wissel, A. Schweikard, and F. Ernst. Evaluation of the potential of multi-modal sensors for respiratory motion prediction and correlation. In *IEEE Engineering in Medicine and Biology Society - EMBC 2013*, pp. 5678–5681, Osaka, Japan, 2013. DOI: [10.1109/EMBC.2013.6610839](https://doi.org/10.1109/EMBC.2013.6610839).
- [66] R. Dürichen, T. Wissel, F. Ernst, and A. Schweikard. Multi-modal respiratory motion prediction using sequential forward selection method. In *Deutsche Gesellschaft für Computer- und Roboterassistierte Chirurgie (CURAC)*, pp. 183–187, Innsbruck, Austria, 2013. <https://www.curac.org>.
- [67] R. Dürichen, T. Wissel, F. Ernst, and A. Schweikard. Respiratory motion compensation with relevance vector machines. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8150 of *LNCS*, pp. 108–115. Springer Berlin Heidelberg, Nagoya, Japan, 2013. DOI: [10.1007/978-3-642-40763-5\\_14](https://doi.org/10.1007/978-3-642-40763-5_14).

- [68] R. Dürichen, T. Wissel, and A. Schweikard. Optimized order estimation for autoregressive models to predict respiratory motion. *International journal of computer assisted radiology and surgery*, 8(6): 1037–1042, 2013. DOI: [10.1007/s11548-013-0900-0](https://doi.org/10.1007/s11548-013-0900-0).
- [69] R. Dürichen, X. Fang, T. Wissel, and A. Schweikard. Gaussian process models for respiratory motion compensation. In *Proceedings of the 28th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS'14)*, pp. 286–287, Fukuoka, Japan. Published in. *International Journal of Computer Assisted Radiology and Surgery*, 9:286–287, 2014.
- [70] R. Dürichen, M. A. F. Pimentel, L. Clifton, A. Schweikard, and D. A. Clifton. Multi-task gaussian process models for biomedical applications. In *Proceedings of the International Conference on Biomedical and Health Informatics (IEEE BHI)*, pp. 492–495, Valencia, Spain, 2014. DOI: [10.1109/BHI.2014.6864410](https://doi.org/10.1109/BHI.2014.6864410).
- [71] R. Dürichen, T. Wissel, F. Ernst, M. A. F. Pimentel, D. A. Clifton, and A. Schweikard. A unified approach for respiratory motion prediction and correlation with multi-task gaussian processes. In *IEEE International Workshop on Machine Learning for Signal Processing*, accepted for publication, Reims, France, 2014.
- [72] R. Dürichen, T. Wissel, and A. Schweikard. Controlling motion prediction errors in radiotherapy with relevance vector machines. *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–9, 2014. DOI: [10.1007/s11548-014-1008-x](https://doi.org/10.1007/s11548-014-1008-x).
- [73] R. Dürichen, T. Wissel, and A. Schweikard. Exploiting probabilistic uncertainty measures for respiratory motion prediction. In *Proceedings of the 28th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS'14)*, pp. 59, Fukuoka, Japan. Published in. *International Journal of Computer Assisted Radiology and Surgery*, 9:59, 2014.
- [74] R. Dürichen, T. Wissel, F. Ernst, A. Schlaefer, and A. Schweikard. Multivariate respiratory motion prediction. *Physics in Medicine and Biology*, 2014 (accepted for publication).
- [75] Y. E. Erdi, S. A. Nehmeh, T. Pan, A. Pevsner, K. E. Rosenzweig, G. Mageras, E. D. Yorke, H. Schoder, W. Hsiao, O. D. Squire, P. Vernon, J. B. Ashman, H. Mostafavi, S. M. Larson, and J. L. Humm. The CT motion quantitation of lung lesions and its impact on PET-measured SUVs. *Journal of Nuclear Medicine*, 45(8):1287–1292, 2004.
- [76] F. Ernst. *Motion Compensation in Robotic Radiosurgery*. Springer New York, 2012. DOI: [10.1007/978-1-4614-1912-9\\_2](https://doi.org/10.1007/978-1-4614-1912-9_2).
- [77] F. Ernst and A. Schweikard. A family of linear algorithms for the prediction of respiratory motion in image-guided radiotherapy. In *Proceedings of the 22nd International Conference and Exhibition on Computer Assisted Radiology and Surgery (CARS)*, pp. 31–32, Barcelona, Spain, 2008. [http://www.rob.uni-luebeck.de/publikationen\\_downloads/es\\_08a.pdf-ba3176bee2bf7a80a6d46a2d45fa8dfa.pdf](http://www.rob.uni-luebeck.de/publikationen_downloads/es_08a.pdf-ba3176bee2bf7a80a6d46a2d45fa8dfa.pdf).
- [78] F. Ernst and A. Schweikard. Predicting respiratory motion signals for image-guided radiotherapy using multi-step linear methods (MULIN). *International Journal of Computer Assisted Radiology and Surgery*, 3(1-2):85–90, 2008. DOI: [10.1007/s11548-008-0211-z](https://doi.org/10.1007/s11548-008-0211-z).
- [79] F. Ernst and A. Schweikard. Prediction of respiratory motion using a modified Recursive Least Squares algorithm. In *Deutschen Gesellschaft für Computer- und Roboterassistierte Chirurgie (CURAC)*, pp. 157–160, Leipzig, Germany, 2008. <https://www.curac.org>.
- [80] F. Ernst and A. Schweikard. Forecasting respiratory motion with accurate online support vector regression (SVRpred). *International Journal of Computer Assisted Radiology and Surgery*, 4(5):439–447, 2009. DOI: [10.1007/s11548-009-0355-5](https://doi.org/10.1007/s11548-009-0355-5).



- [81] F. Ernst, A. Schlaefer, and A. Schweikard. Prediction of respiratory motion with wavelet-based multiscale autoregression. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 4792 of *LNCS*, pp. 668–675. Springer, Berlin, Heidelberg, 2007. DOI: [10.1007/978-3-540-75759-7\\_81](https://doi.org/10.1007/978-3-540-75759-7_81).
- [82] F. Ernst, A. Schlaefer, S. Dieterich, and A. Schweikard. A fast lane approach to LMS prediction of respiratory motion signals. *Biomedical Signal Processing and Control*, **3**(4):291–299, 2008. DOI: [10.1016/j.bspc.2008.06.001](https://doi.org/10.1016/j.bspc.2008.06.001).
- [83] F. Ernst, V. Martens, S. Schlichting, A. Beširević, M. Kleemann, C. Koch, D. Petersen, and A. Schweikard. Correlating chest surface motion to motion of the liver using  $\epsilon$ -SVR – a porcine study. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 5762 of *LNCS*, pp. 356–364. Springer Berlin Heidelberg, London, UK, 2009. DOI: [10.1007/978-3-642-04271-3\\_44](https://doi.org/10.1007/978-3-642-04271-3_44).
- [84] F. Ernst, A. Schlaefer, and A. Schweikard. Predicting the outcome of respiratory motion prediction. *Medical Physics*, **38**(10):5569–5581, 2011. DOI: [10.1118/1.3633907](https://doi.org/10.1118/1.3633907).
- [85] F. Ernst, R. Bruder, A. Schlaefer, and A. Schweikard. Correlation between external and internal respiratory motion: a validation study. *International Journal of Computer Assisted Radiology and Surgery*, **7**(3):483–492, 2012. DOI: [10.1007/s11548-011-0653-6](https://doi.org/10.1007/s11548-011-0653-6).
- [86] F. Ernst, R. Dürichen, A. Schlaefer, and A. Schweikard. Evaluating and comparing algorithms for respiratory motion prediction. *Physics in Medicine and Biology*, **58**(11):3911–3929, 2013. DOI: [10.1088/0031-9155/58/11/3911](https://doi.org/10.1088/0031-9155/58/11/3911).
- [87] K. C. Evans, S. A. Shea, and A. J. Saykin. Functional MRI localisation of central nervous system regions associated with volitional inspiration in humans. *The Journal of Physiology*, **520**(2):383–392, 1999. DOI: [10.1111/j.1469-7793.1999.00383.x](https://doi.org/10.1111/j.1469-7793.1999.00383.x).
- [88] X. Fang. *Respiratory motion prediction with Gaussian process models*. Master thesis, Institute for Robotics and Cognitive Systems, Universität zu Lübeck, Lübeck, Germany, 2013.
- [89] R. Farré, R. Peslin, D. Navajas, C. Gallina, and B. Suki. Analysis of the dynamic characteristics of pressure transducers for studying respiratory mechanics at high frequencies. *Medical and Biological Engineering and Computing*, **27**(5):531–537, 1989. DOI: [10.1007/BF02441474](https://doi.org/10.1007/BF02441474).
- [90] R. Farré, J. M. Montserrat, and D. Navajas. Noninvasive monitoring of respiratory mechanics during sleep. *European Respiratory Journal*, **24**(6):1052–1060, 2004. DOI: [10.1183/09031936.04.00072304](https://doi.org/10.1183/09031936.04.00072304).
- [91] H. Fayad, T. Pan, O. Pradier, and D. Visvikis. Patient specific respiratory motion modeling using a 3d patient’s external surface. *Medical Physics*, **39**(6):3386–3395, 2012. DOI: [10.1118/1.4718578](https://doi.org/10.1118/1.4718578).
- [92] J. Fei and I. Pavlidis. Thermistor at a distance: unobtrusive measurement of breathing. *IEEE Transactions on Biomedical Engineering*, **57**(4):988–998, 2010. DOI: [10.1109/TBME.2009.2032415](https://doi.org/10.1109/TBME.2009.2032415).
- [93] J. L. Feldman and C. A. Del Negro. Looking for inspiration: new perspectives on respiratory rhythm. *Nat Rev Neurosci*, **7**(3):232–241, 2006. DOI: [10.1038/nrn1871](https://doi.org/10.1038/nrn1871).
- [94] M. Filipovic, P.-A. Vuissoz, A. Codreanu, M. Claudon, and J. Felblinger. Motion compensated generalized reconstruction for free-breathing dynamic contrast-enhanced MRI. *Magnetic Resonance in Medicine*, **65**(3):812–822, 2011. DOI: [10.1002/mrm.22644](https://doi.org/10.1002/mrm.22644).
- [95] R. A. Fisher and o. undefined. On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron*, **1**:3–32, 1921.
- [96] T. Fletcher. Relevance vector machines explained, technical report, University College London, Oct., 2010. <http://www.tristanfletcher.co.uk/RVM%20Explained.pdf>. (visited on 2014-06-23).

- [97] E. C. Ford, G. S. Mageras, E. Yorke, K. E. Rosenzweig, R. Wagman, and C. C. Ling. Evaluation of respiratory movement during gated radiotherapy using film and electronic portal imaging. *International Journal of Radiation Oncology\* Biology\* Physics*, 52(2):522–531, 2002. DOI: [10.1016/S0360-3016\(01\)02681-5](https://doi.org/10.1016/S0360-3016(01)02681-5).
- [98] I. Gergel, J. Hering, R. Tetzlaff, H.-P. Meinzer, and I. Wegner. An electromagnetic navigation system for transbronchial interventions with a novel approach to respiratory motion compensation. *Medical Physics*, 38(12):6742–6753, 2011. DOI: [10.1118/1.3662871](https://doi.org/10.1118/1.3662871).
- [99] D. P. Gierga, G. T. Chen, J. H. Kung, M. Betke, J. Lombardi, and C. G. Willett. Quantification of respiration-induced abdominal tumor motion and its impact on IMRT dose distributions. *International Journal of Radiation Oncology\* Biology\* Physics*, 58(5):1584–1595, 2004. DOI: [10.1016/j.ijrobp.2003.09.077](https://doi.org/10.1016/j.ijrobp.2003.09.077).
- [100] D. P. Gierga, J. Brewer, G. C. Sharp, M. Betke, C. G. Willett, and G. T. Chen. The correlation between internal and external markers for abdominal tumors: Implications for respiratory gating. *International Journal of Radiation Oncology\* Biology\* Physics*, 61(5):1551–1558, 2005. DOI: [10.1016/j.ijrobp.2004.12.013](https://doi.org/10.1016/j.ijrobp.2004.12.013).
- [101] P. Giraud, Y. De Rycke, B. Dubray, S. Helfre, D. Voican, L. Guo, J.-C. Rosenwald, K. Keraudy, M. Housset, E. Touboul, and J.-M. Cosset. Conformal radiotherapy (CRT) planning for lung cancer: analysis of intrathoracic organ motion during extreme phases of breathing. *International Journal of Radiation Oncology\* Biology\* Physics*, 51(4):1081–1092, 2001. DOI: [10.1016/S0360-3016\(01\)01766-7](https://doi.org/10.1016/S0360-3016(01)01766-7).
- [102] B. Gonenc, E. Feldman, P. Gehlbach, J. Handa, R. H. Taylor, and I. Iordachita. Towards robot-assisted vitreoretinal surgery: Force-sensing micro-forceps integrated with a handheld micromanipulator. In *Proceedings of Optical Fibers and Sensors for Medical Diagnostics and Treatment Applications XII, SPIE 8218*, San Francisco, CA, USA, 2012. <https://cld.pt/dl/download/f9658d95-0c61-4ebd-8d70-3cada6be2c0b/ICRA2014/media/files/1509.pdf>. DOI: [10.1117/12.909602](https://doi.org/10.1117/12.909602).
- [103] J. H. Goodband, O. C. L. Haas, and J. A. Mills. A comparison of neural network approaches for on-line prediction in IGRT. *Medical Physics*, 35(3):1113–1122, 2008. DOI: [10.1118/1.2836416](https://doi.org/10.1118/1.2836416).
- [104] T. Grantzau, M. S. Thomsen, M. Væth, and J. Overgaard. Risk of second primary lung cancer in women after radiotherapy for breast cancer. *Radiotherapy and Oncology*, 111(3):366 – 373, 2014. DOI: [10.1016/j.radonc.2014.05.004](https://doi.org/10.1016/j.radonc.2014.05.004).
- [105] H. Gray and W. H. Lewis. *Anatomy of the human body*, Philadelphia, PA, 1918. <http://www.bartleby.com/107/>. (visited on 2014-06-23).
- [106] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3: 1157–1182, 2003.
- [107] G. Harauz and M. J. Bronskill. Comparison of the liver’s respiratory motion in the supine and upright positions: Concise communication. *J Nucl Med*, 1979.
- [108] A. R. Hareendranathan. *Real-time movement compensation for synchronous robotic hifu surgery*. PhD thesis, Nanyang Technological University (NTU), Singapore, 2011. <http://dr.ntu.edu.sg/>.
- [109] C. Herrmann, K. Schilling, and L. Ma. Modeling a hexapod for tumor motion compensation in robot assisted radiotherapy. In *Robotics (ISR), 41st International Symposium on and 2010 6th German Conference on Robotics (ROBOTIK)*, pp. 1–7, 2010.
- [110] C. Hick, A. Hick, F. Jockenhövel, and R. Merker. *Kurzlehrbuch Physiologie*. Urban & Fischer, München; Jena, 4th edition, 2002. ISBN 9783437418914.
- [111] D. Higdon. Space and space-time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues*, pp. 37–56, 2002. DOI: [10.1007/978-1-4471-0657-9\\_2](https://doi.org/10.1007/978-1-4471-0657-9_2).

- [112] J. D. Hoisak, K. E. Sixel, R. Tirona, P. C. Cheung, and J.-P. Pignol. Correlation of lung tumor motion with external surrogate indicators of respiration. *International Journal of Radiation Oncology\*Biophysics*, **60**(4):1298–1306, 2004. DOI: [10.1016/j.ijrobp.2004.07.681](https://doi.org/10.1016/j.ijrobp.2004.07.681).
- [113] J. D. P. Hoisak, K. E. Sixel, R. Tirona, P. C. F. Cheung, and J.-P. Pignol. Prediction of lung tumour position based on spirometry and on abdominal displacement: Accuracy and reproducibility. *Radiotherapy and Oncology*, **78**(3):339–346, 2006. DOI: [10.1016/j.radonc.2006.01.008](https://doi.org/10.1016/j.radonc.2006.01.008).
- [114] N. Homma, M. Sakai, H. Endo, M. Mitsuya, Y. Takai, and M. Yoshizawa. A new motion management method for lung tumor tracking radiation therapy. *WSEAS Transactions on Systems*, **8**(4):471–480, 2009.
- [115] S.-M. Hong, B.-H. Jung, and D. Ruan. Real-time prediction of respiratory motion based on a local dynamic model in an augmented space. *Physics in Medicine and Biology*, **56**(6):1775, 2011. DOI: [10.1088/0031-9155/56/6/016](https://doi.org/10.1088/0031-9155/56/6/016).
- [116] M. Hoogeman, J.-B. Prévost, J. Nuyttens, J. Pöll, P. Levendag, and B. Heijmen. Clinical accuracy of the respiratory tumor tracking system of the CyberKnife: Assessment by analysis of log files. *International Journal of Radiation Oncology\*Biophysics*, **74**(1):297–303, 2009. DOI: [10.1016/j.ijrobp.2008.12.041](https://doi.org/10.1016/j.ijrobp.2008.12.041).
- [117] Y. Hu, O. P. Green, P. Parikh, J. Olsen, and S. Mutic. TH-e-BRA-07: Initial experience with the ViewRay system - quality assurance testing of the imaging component. *Medical Physics*, **39**(6):4013–4013, 2012. DOI: [10.1118/1.4736368](https://doi.org/10.1118/1.4736368).
- [118] K. Huang, I. Buzurovic, Y. Yu, and T. K. Podder. A comparative study of a novel AE-nLMS filter and two traditional filters in predicting respiration induced motion of the tumor. In *Proceedings of the 2010 IEEE International Conference on Bioinformatics and Bioengineering*, pp. 281–282. IEEE, 2010. DOI: [10.1109/BIBE.2010.53](https://doi.org/10.1109/BIBE.2010.53).
- [119] R. Huch and K. D. Jürgens. *Mensch, Körper, Krankheit*. Urban & Fischer Verlag/Elsevier GmbH, München u.a., 5th edition, 2007. ISBN 9783437267918.
- [120] P. D. Hung, S. Bonnet, R. Guillemaud, E. Castelli, and P. T. N. Yen. Estimation of respiratory waveform using an accelerometer. In *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008*, pp. 1493–1496, 2008. DOI: [10.1109/ISBI.2008.4541291](https://doi.org/10.1109/ISBI.2008.4541291).
- [121] G. Hutten, H. van Thuijl, A. van Bellegem, L. van Eykern, and W. van Aalderen. A literature review of the methodology of EMG recordings of the diaphragm. *Journal of Electromyography and Kinesiology*, **20**(2):185–190, 2010. DOI: [10.1016/j.jelekin.2009.02.008](https://doi.org/10.1016/j.jelekin.2009.02.008).
- [122] K. Ichiji, N. Homma, M. Sakai, Y. Narita, Y. Takai, X. Zhang, M. Abe, N. Sugita, and M. Yoshizawa. A time-varying seasonal autoregressive model-based prediction of respiratory motion for tumor following radiotherapy. *Computational and Mathematical Methods in Medicine*, **2013**, 2013. DOI: [10.1155/2013/390325](https://doi.org/10.1155/2013/390325).
- [123] M. Isaksson, J. Jalden, and M. J. Murphy. On using an adaptive neural network to predict lung tumor motion during respiration for radiotherapy applications. *Medical Physics*, **32**(12):3801–3809, 2005. DOI: [10.1118/1.2134958](https://doi.org/10.1118/1.2134958).
- [124] A. C. Jackson and A. Vinegar. A technique for measuring frequency response of pressure, volume, and flow transducers. *J Appl Physiol*, **47**(2):462–467, 1979.
- [125] T. Juhler Nøttrup, S. S. Korreman, A. N. Pedersen, L. R. Aarup, H. a. Nyström, M. Olsen, and L. Specht. Intra- and interfraction breathing variations during curative radiotherapy for lung cancer. *Radiotherapy and oncology*, **84**(1):40–48, 2007. DOI: [10.1016/j.radonc.2007.05.026](https://doi.org/10.1016/j.radonc.2007.05.026).

- [126] B.-H. Jung, B.-H. Kim, and S.-M. Hong. Respiratory motion prediction with extended kalman filters based on local circular motion model. *International Journal of Bio-Science and Bio-Technology*, 5(1):51–58, 2013.
- [127] V. Jurcak, D. Tsuzuki, and I. Dan. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *NeuroImage*, 34(4):1600–1611, 2007. DOI: [10.1016/j.neuroimage.2006.09.024](https://doi.org/10.1016/j.neuroimage.2006.09.024).
- [128] A. Kalet, G. Sandison, H. Wu, and R. Schmitz. A state-based probabilistic model for tumor respiratory motion prediction. *Physics in medicine and biology*, 55(24):7615–7631, 2010. DOI: [10.1088/0031-9155/55/24/015](https://doi.org/10.1088/0031-9155/55/24/015).
- [129] E. Kanoulas, J. A. Aslam, G. C. Sharp, R. I. Berbeco, S. Nishioka, H. Shirato, and S. B. Jiang. Derivation of the tumor position from external respiratory surrogates with periodical updating of the internal/external correlation. *Physics in Medicine and Biology*, 52(17):5443–5456, 2007. DOI: [10.1088/0031-9155/52/17/023](https://doi.org/10.1088/0031-9155/52/17/023).
- [130] W. Karlen, S. Raman, J. Ansermino, and G. Dumont. Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering*, 60(7):1946–1953, 2013. DOI: [10.1109/TBME.2013.2246160](https://doi.org/10.1109/TBME.2013.2246160).
- [131] P. J. Keall, G. S. Mageras, J. M. Balter, R. S. Emery, K. M. Forster, S. B. Jiang, J. M. Kapatoes, D. A. Low, M. J. Murphy, B. R. Murray, C. R. Ramsey, M. B. V. Herk, S. S. Vedam, J. W. Wong, and E. Yorke. The management of respiratory motion in radiation oncology report of AAPM task group 76a. *Medical Physics*, 33(10):3874–3900, 2006. DOI: [10.1118/1.2349696](https://doi.org/10.1118/1.2349696).
- [132] P. J. Keall, A. Sawant, B. Cho, D. Ruan, J. Wu, P. Poulsen, J. Petersen, L. J. Newell, H. Cattell, and S. Korreman. Electromagnetic-guided dynamic multileaf collimator tracking enables motion management for intensity-modulated arc therapy. *International Journal of Radiation Oncology\*Biography\*Physics*, 79(1):312–320, 2011. DOI: [10.1016/j.ijrobp.2010.03.011](https://doi.org/10.1016/j.ijrobp.2010.03.011).
- [133] A. Khamene, J. K. Warzelhan, S. Vogt, D. Elgort, C. Chefd’Hotel, J. L. Duerk, J. Lewin, F. K. Wacker, and F. Sauer. Characterization of internal organ motion using skin marker positions. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 526–533. Published in LNCS, 3217:526–533, 2004. DOI: [10.1007/978-3-540-30136-3\\_65](https://doi.org/10.1007/978-3-540-30136-3_65).
- [134] W. Kilby, J. R. Dooley, G. Kuduvalli, S. Sayeh, and C. R. Maurer Jr. The CyberKnife robotic radiosurgery system in 2010. *Technology in cancer research & treatment*, 9(5):433–452, 2010.
- [135] T. Kimura, Y. Hirokawa, Y. Murakami, M. Tsujimura, T. Nakashima, Y. Ohno, M. Kenjo, Y. Kaneyasu, K. Wadasaki, and K. Ito. Reproducibility of organ position using voluntary breath-hold method with spirometer for extracranial stereotactic radiotherapy. *International Journal of Radiation Oncology\*Biography\*Physics*, 60(4):1307–1313, 2004. DOI: [10.1016/j.ijrobp.2004.07.718](https://doi.org/10.1016/j.ijrobp.2004.07.718).
- [136] V. R. Kini, S. S. Vedam, P. J. Keall, S. Patil, C. Chen, and R. Mohan. Patient training in respiratory-gated radiotherapy. *Medical Dosimetry*, 28(1):7–11, 2003. DOI: [10.1016/S0958-3947\(02\)00136-X](https://doi.org/10.1016/S0958-3947(02)00136-X).
- [137] K. Kitamura, H. Shirato, Y. Seppenwoolde, T. Shimizu, Y. Kodama, H. Endo, R. Onimaru, M. Oda, K. Fujita, S. Shimizu, and K. Miyasaka. Tumor location, cirrhosis, and surgical history contribute to tumor movement in the liver, as measured during stereotactic irradiation using a real-time tumor-tracking radiotherapy system. *International Journal of Radiation Oncology\*Biography\*Physics*, 56(1):221–228, 2003. DOI: [10.1016/S0360-3016\(03\)00082-8](https://doi.org/10.1016/S0360-3016(03)00082-8).
- [138] G. H. Klem, H. O. Lüders, H. H. Jasper, and C. Elger. The ten-twenty electrode system of the international federation. the international federation of clinical neurophysiology. *Electroencephalography and clinical neurophysiology. Supplement*, 52:3–6, 1999.

- [139] W. Klimesch. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, **29**(2–3):169–195, 1999. DOI: [10.1016/S0165-0173\(98\)00056-3](https://doi.org/10.1016/S0165-0173(98)00056-3).
- [140] N. Koch, H. H. Liu, G. Starkschall, M. Jacobson, K. Forster, Z. Liao, R. Komaki, and C. W. Stevens. Evaluation of internal lung motion for respiratory-gated radiotherapy using MRI: Part I—correlating internal lung motion with skin fiducial motion. *International Journal of Radiation Oncology\*Biography\*Physics*, **60**(5):1459–1472, 2004. DOI: [10.1016/j.ijrobp.2004.05.055](https://doi.org/10.1016/j.ijrobp.2004.05.055).
- [141] S. S. Korreman, A. N. Pedersen, T. J. Nøttrup, L. Specht, and H. Nyström. Breathing adapted radiotherapy for breast cancer: Comparison of free breathing gating with the breath-hold technique. *Radiotherapy and Oncology*, **76**(3):311–318, 2005. DOI: [10.1016/j.radonc.2005.07.009](https://doi.org/10.1016/j.radonc.2005.07.009).
- [142] A. Krauss, S. Nill, and U. Oelfke. The comparative performance of four respiratory motion predictors for real-time tumour tracking. *Physics in Medicine and Biology*, **56**(16):5303–5317, 2011. DOI: [10.1088/0031-9155/56/16/015](https://doi.org/10.1088/0031-9155/56/16/015).
- [143] A. Krauss, S. Nill, M. Tacke, and U. Oelfke. Electromagnetic real-time tumor position monitoring and dynamic multileaf collimator tracking using a siemens 160 MLC: Geometric and dosimetric accuracy of an integrated system. *International Journal of Radiation Oncology\*Biography\*Physics*, **79**(2):579–587, 2011. DOI: [10.1016/j.ijrobp.2010.03.043](https://doi.org/10.1016/j.ijrobp.2010.03.043).
- [144] H. D. Kubo and B. C. Hill. Respiration gated radiotherapy treatment: a technical study. *Physics in Medicine and Biology*, **41**(1):83, 1996. DOI: [10.1088/0031-9155/41/1/007](https://doi.org/10.1088/0031-9155/41/1/007).
- [145] I. Kuhlemann. *Force and Image Adaptive Strategies for Robotised Placement of 4D Ultrasound Probes*. Master thesis, Institute for Robotics and Cognitive Systems, Universität zu Lübeck, Lübeck, Germany, 2013.
- [146] K. M. Langen and D. T. L. Jones. Organ motion and its management. *International Journal of Radiation Oncology\*Biography\*Physics*, **50**(1):265–278, 2001. DOI: [10.1016/S0360-3016\(01\)01453-5](https://doi.org/10.1016/S0360-3016(01)01453-5).
- [147] U. W. Langner and P. J. Keall. Accuracy in the localization of thoracic and abdominal tumors using respiratory displacement, velocity, and phase. *Medical Physics*, **36**(2):386–393, 2009. DOI: [10.1118/1.3049595](https://doi.org/10.1118/1.3049595).
- [148] R. Lansing and J. Savelle. Chest surface recording of diaphragm potentials in man. *Electroencephalography and Clinical Neurophysiology*, **72**(1):59–68, 1989. DOI: [10.1016/0013-4694\(89\)90031-X](https://doi.org/10.1016/0013-4694(89)90031-X).
- [149] J. Lattanzi, S. McNeeley, A. Hanlon, T. E. Schultheiss, and G. E. Hanks. Ultrasound-based stereotactic guidance of precision conformal external beam radiation therapy in clinically localized prostate cancer. *Urology*, **55**(1):73–78, 2000. DOI: [10.1016/S0090-4295\(99\)00389-1](https://doi.org/10.1016/S0090-4295(99)00389-1).
- [150] G. Li, H. Xie, H. Ning, W. Lu, D. Low, D. Citrin, A. Kaushal, L. Zach, K. Camphausen, and R. W. Miller. A novel analytical approach to the prediction of respiratory diaphragm motion based on external torso volume change. *Physics in Medicine and Biology*, **54**(13):4113, 2009. DOI: [10.1088/0031-9155/54/13/010](https://doi.org/10.1088/0031-9155/54/13/010).
- [151] R. Li, J. H. Lewis, R. I. Berbeco, and L. Xing. Real-time tumor motion estimation using respiratory surrogate via memory-based learning. *Physics in Medicine and Biology*, **57**(15):4771, 2012. DOI: [10.1088/0031-9155/57/15/4771](https://doi.org/10.1088/0031-9155/57/15/4771).
- [152] H. Liu, Y. Yu, W. O’Dell, and M. Schell. A prospective tracking algorithm for motion compensation radiotherapy. In *Annual meeting of American Association of Physicists in Medicine (AAPM)*, pp. 425–440, San Diego, CA, USA, 2003.
- [153] H. H. Liu, N. Koch, G. Starkschall, M. Jacobson, K. Forster, Z. Liao, R. Komaki, and C. W. Stevens. Evaluation of internal lung motion for respiratory-gated radiotherapy using MRI: Part II—margin reduction of internal target volume. *International Journal of Radiation Oncology\*Biography\*Physics*, **60**(5):1473–1483, 2004. DOI: [10.1016/j.ijrobp.2004.05.054](https://doi.org/10.1016/j.ijrobp.2004.05.054).

- [154] F. Lohr, J. Debus, C. Frank, K. Herfarth, O. Pastyr, B. Rhein, M. L. Bahner, W. Schlegel, and M. Wannemacher. Noninvasive patient fixation for extracranial stereotactic radiotherapy. *International Journal of Radiation Oncology\*Biophysics*, **45**(2):521–527, 1999. DOI: [10.1016/S0360-3016\(99\)00190-X](https://doi.org/10.1016/S0360-3016(99)00190-X).
- [155] D. A. Low, P. J. Parikh, W. Lu, J. F. Dempsey, S. H. Wahab, J. P. Hubenschmidt, M. M. Nystrom, M. Handoko, and J. D. Bradley. Novel breathing motion model for radiotherapy. *International Journal of Radiation Oncology\*Biophysics*, **63**(3):921–929, 2005. DOI: [10.1016/j.ijrobp.2005.03.070](https://doi.org/10.1016/j.ijrobp.2005.03.070).
- [156] W. Lu. Real-time motion-adaptive delivery (MAD) using binary MLC: I. static beam (topotherapy) delivery. *Physics in Medicine and Biology*, **53**(22):6491, 2008. DOI: [10.1088/0031-9155/53/22/014](https://doi.org/10.1088/0031-9155/53/22/014).
- [157] W. Lu, P. J. Parikh, J. P. Hubenschmidt, D. G. Politte, B. R. Whiting, J. D. Bradley, S. Mutic, and D. A. Low. Reduction of motion blurring artifacts using respiratory gated CT in sinogram space: a quantitative evaluation. *Medical physics*, **32**(11):3295–3304, 2005. DOI: [10.1118/1.2074187](https://doi.org/10.1118/1.2074187).
- [158] W. Lu, M. Chen, K. J. Ruchala, Q. Chen, K. M. Langen, P. A. Kupelian, and G. H. Olivera. Real-time motion-adaptive-optimization (MAO) in TomoTherapy. *Physics in medicine and biology*, **54**(14):4373, 2009. DOI: [10.1088/0031-9155/54/14/003](https://doi.org/10.1088/0031-9155/54/14/003).
- [159] Y. M. Luo, M. I. Polkey, L. C. Johnson, R. A. Lyall, M. L. Harris, M. Green, and J. Moxham. Diaphragm EMG measured by cervical magnetic and electrical phrenic nerve stimulation. *J Appl Physiol*, **85**(6):2089–2099, 1998.
- [160] L. Ma, C. Herrmann, and K. Schilling. Modeling and prediction of lung tumor motion for robotic assisted radiotherapy. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007. IROS 2007*, pp. 189–194. IEEE, 2007. ISBN 978-1-4244-0912-9. DOI: [10.1109/IROS.2007.4398952](https://doi.org/10.1109/IROS.2007.4398952).
- [161] E. J. W. Maarsingh, L. A. van Eykern, A. B. Sprickelman, M. O. Hoekstra, and W. M. C. van Aalderen. Respiratory muscle activity measured with a noninvasive EMG technique: technical aspects and reproducibility. *J Appl Physiol*, **88**(6):1955–1961, 2000.
- [162] C. K. MacGowan and M. L. Wood. Phase-encode reordering to minimize errors caused by motion. *Magnetic Resonance in Medicine*, **35**(3):391–398, 1996. DOI: [10.1002/mrm.1910350318](https://doi.org/10.1002/mrm.1910350318).
- [163] G. S. Mageras, A. Pevsner, E. D. Yorke, K. E. Rosenzweig, E. C. Ford, A. Hertanto, S. M. Larson, D. M. Lovelock, Y. E. Erdi, S. A. Nehmeh, J. L. Humm, and C. C. Ling. Measurement of lung tumor motion using respiration-correlated CT. *International Journal of Radiation Oncology\*Biophysics*, **60**(3):933–941, 2004. DOI: [10.1016/j.ijrobp.2004.06.021](https://doi.org/10.1016/j.ijrobp.2004.06.021).
- [164] L. Maier-Hein, F. Pianka, A. Seitel, S. A. Müller, A. Tekbas, M. Seitel, I. Wolf, B. M. Schmied, and H.-P. Meinzer. Precision targeting of liver lesions with a needle-based soft tissue navigation system. In *Medical Image Computing and Computer-Assisted Intervention – (MICCAI)*, pp. 42–49. Published in LNCS, **4792**:42–49, 2007. DOI: [10.1007/978-3-540-75759-7\\_6](https://doi.org/10.1007/978-3-540-75759-7_6).
- [165] K. Malinowski, T. J. McAvoy, R. George, S. Dietrich, and W. D. D’Souza. Incidence of changes in respiration-induced tumor motion and its relationship with respiratory surrogates during individual treatment fractions. *International Journal of Radiation Oncology\* Biophysics*, **82**(5):1665–1673, 2012. DOI: [10.1016/j.ijrobp.2011.02.048](https://doi.org/10.1016/j.ijrobp.2011.02.048).
- [166] K. C. McCall and R. Jeraj. Dual-component model of respiratory motion based on the periodic autoregressive moving average (periodic ARMA) method. *Physics in Medicine and Biology*, **52**(12):3455, 2007. DOI: [10.1088/0031-9155/52/12/009](https://doi.org/10.1088/0031-9155/52/12/009).
- [167] J. McClelland, D. Hawkes, T. Schaeffter, and A. King. Respiratory motion models: A review. *Medical Image Analysis*, **17**(1):19–42, 2013. DOI: [10.1016/j.media.2012.09.005](https://doi.org/10.1016/j.media.2012.09.005).

- [168] L. C. McKay, K. C. Evans, R. S. J. Frackowiak, and D. R. Corfield. Neural correlates of voluntary breathing in humans. *J Appl Physiol*, **95**(3):1170–1178, 2003. DOI: [10.1152/jappphysiol.00641.2002](https://doi.org/10.1152/jappphysiol.00641.2002).
- [169] D. K. McKenzie and S. C. Gandevia. Phrenic nerve conduction times and twitch pressures of the human diaphragm. *J Appl Physiol*, **58**(5):1496–1504, 1985.
- [170] D. McQuaid and S. Webb. Target-tracking deliveries using conventional multileaf collimators planned with 4d direct-aperture optimization. *Physics in Medicine and Biology*, **53**(15):4013, 2008. DOI: [10.1088/0031-9155/53/15/001](https://doi.org/10.1088/0031-9155/53/15/001).
- [171] A. Melkumyan and F. Ramos. Multi-kernel gaussian processes. In *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Two*, pp. 1408–1413, Barcelona, Catalonia, Spain. AAAI Press. Published in *IJCAI'11*, pp. 1408–1413, 2011. DOI: [10.5591/978-1-57735-516-8/IJCAI11-238](https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-238).
- [172] D. J. Meredith, D. Clifton, P. Charlton, J. Brooks, C. W. Pugh, and L. Tarassenko. Photoplethysmographic derivation of respiratory rate: a review of relevant physiology. *Journal of medical engineering & technology*, **36**(1):1–7, 2012. DOI: [10.3109/03091902.2011.638965](https://doi.org/10.3109/03091902.2011.638965).
- [173] K. Merkel. *EEG, EOG and EMG Feature Extraction for Motion Compensation in Extracranial Stereotactic Radiotherapy*. Bachelor thesis, Insititute for Robotics and Cognitive Systems, University of Lübeck, Lübeck, Germany, 2013.
- [174] K. Merkel, T. Wissel, A. Schweikard, and R. Dürichen. Respiratory motion prediction with surface EMG features. In *Proceedings of the 28th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS)*, pp. 290–291, Fukuoka, Japan, 2014.
- [175] J. Meyer, K. Baier, J. Wilbert, M. Guckenberger, A. Richter, and M. Flentje. Three-dimensional spatial modelling of the correlation between abdominal motion and lung tumour motion with breathing. *Acta Oncologica*, **45**(7):923–934, 2006. DOI: [10.1080/02841860600897926](https://doi.org/10.1080/02841860600897926).
- [176] M. R. Miller, J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, R. Crapo, P. Enright, C. P. Van der Grinten, P. Gustafsson, and others. Standardisation of spirometry. *Eur Respir J*, **26**(2):319–38, 2005. DOI: [10.1183/09031936.05.00034805](https://doi.org/10.1183/09031936.05.00034805).
- [177] C. Morélot-Panzini, E. Fournier, C. Donzel-Raynaud, O. Dubourg, J.-C. Willer, and T. Similowski. Conduction velocity of the human phrenic nerve in the neck. *Journal of Electromyography and Kinesiology*, **19**(1):122–130, 2009. DOI: [10.1016/j.jelekin.2007.06.017](https://doi.org/10.1016/j.jelekin.2007.06.017).
- [178] M. Mori, K. Murata, M. Takahashi, K. Shimoyama, T. Ota, R. Morita, and T. Sakamoto. Accurate contiguous sections without breath-holding on chest CT: value of respiratory gating and ultrafast CT. *American Journal of Roentgenology*, **162**(5):1057–1062, 1994. DOI: [10.2214/ajr.162.5.8165981](https://doi.org/10.2214/ajr.162.5.8165981).
- [179] L. M. Morton, K. Onel, R. E. Curtis, E. A. Hungate, and G. T. Armstrong. The rising incidence of second cancers: Patterns of occurrence and identification of risk factors for children and adults. *American Society of Clinical Oncology Educational Book*, **34**:e57–e67, 2014. DOI: [10.14694/EdBook\\_AM.2014.34.e57](https://doi.org/10.14694/EdBook_AM.2014.34.e57).
- [180] M. Mrochen, M. S. Eldine, M. Kaemmerer, T. Seiler, and W. Hütz. Improvement in photorefractive corneal laser surgery results using an active eye-tracking system. *Journal of Cataract & Refractive Surgery*, **27**(7):1000–1006, 2001. DOI: [10.1016/S0886-3350\(00\)00884-1](https://doi.org/10.1016/S0886-3350(00)00884-1).
- [181] N. Mukumoto, M. Nakamura, A. Sawada, Y. Suzuki, K. Takahashi, Y. Miyabe, S. Kaneko, T. Mizowaki, M. Kokubo, and M. Hiraoka. Accuracy verification of infrared marker-based dynamic tumor-tracking irradiation using the gimbaled x-ray head of the vero4drt (MHI-TM2000). *Medical Physics*, **40**(4):041706, 2013. DOI: [10.1118/1.4794506](https://doi.org/10.1118/1.4794506).

- [182] K. Müller, C. Schaller, J. Penne, and J. Hornegger. Surface-based respiratory motion classification and verification. In *Bildverarbeitung für die Medizin 2009*, pp. 257–261. Springer, 2009. DOI: [doi={10.1007/978-3-540-93860-6\\_52}](https://doi.org/10.1007/978-3-540-93860-6_52).
- [183] M. Murphy. Using neural networks to predict breathing motion. In *Seventh International Conference on Machine Learning and Applications, 2008. ICMLA '08*, pp. 528–532, 2008. DOI: [10.1109/ICMLA.2008.136](https://doi.org/10.1109/ICMLA.2008.136).
- [184] M. Murphy, M. Isaakson, and J. Jalden. Adaptive filtering to predict lung tumor motion during free breathing. In *Computer Assisted Radiology and Surgery (CARS)*, pp. 539–544, 2002. DOI: [10.1007/978-3-642-56168-9\\_90](https://doi.org/10.1007/978-3-642-56168-9_90).
- [185] M. J. Murphy and S. Dieterich. Comparative performance of linear and nonlinear neural networks to predict irregular breathing. *Physics in Medicine and Biology*, **51**(22):5903, 2006. DOI: [10.1088/0031-9155/51/22/012](https://doi.org/10.1088/0031-9155/51/22/012).
- [186] M. J. Murphy and D. Pokhrel. Optimization of an adaptive neural network to predict breathing. *Medical Physics*, **36**(1):40–47, 2009. DOI: [10.1118/1.3026608](https://doi.org/10.1118/1.3026608).
- [187] S. Mutic and J. F. Dempsey. The ViewRay system: Magnetic resonance-guided and controlled radiotherapy. *Seminars in Radiation Oncology*, **24**(3):196–199, 2014. DOI: [10.1016/j.semradonc.2014.02.008](https://doi.org/10.1016/j.semradonc.2014.02.008).
- [188] F. Najafi and N. Sepehri. A robotic wrist for remote ultrasound imaging. *Mechanism and Machine Theory*, **46**(8):1153–1170, 2011. DOI: [10.1016/j.mechmachtheory.2011.03.002](https://doi.org/10.1016/j.mechmachtheory.2011.03.002).
- [189] W. A. N'Djin, N. R. Miller, J. C. Bamber, J. Y. Chapelon, and D. Melodelima. Effects of respiratory motion on in-vivo HIFU treatments: a comparative study in the liver. In *AIP Conference Proceedings*, pp. 203–206. AIP Publishing, 2010. DOI: [10.1063/1.3367142](https://doi.org/10.1063/1.3367142).
- [190] Y. Negoro, Y. Nagata, T. Aoki, T. Mizowaki, N. Araki, K. Takayama, M. Kokubo, S. Yano, S. Koga, K. Sasai, Y. Shibamoto, and M. Hiraoka. The effectiveness of an immobilization device in conformal radiotherapy for lung tumor: reduction of respiratory tumor movement and evaluation of the daily setup accuracy. *International Journal of Radiation Oncology\*Biophysics\*Physics*, **50**(4):889–898, 2001. DOI: [10.1016/S0360-3016\(01\)01516-4](https://doi.org/10.1016/S0360-3016(01)01516-4).
- [191] S. A. Nehmeh, Y. E. Erdi, C. C. Ling, K. E. Rosenzweig, O. D. Squire, L. E. Braban, E. Ford, K. Sidhu, G. S. Mageras, S. M. Larson, and others. Effect of respiratory gating on reducing lung motion artifacts in PET imaging of lung cancer. *Medical physics*, **29**(3):366–371, 2002. DOI: [10.1118/1.1448824](https://doi.org/10.1118/1.1448824).
- [192] T. Neicu, H. Shirato, Y. Seppenwoolde, and S. B. Jiang. Synchronized moving aperture radiation therapy (SMART): average tumour trajectory for lung patients. *Physics in Medicine and Biology*, **48**(5):587, 2003. DOI: [10.1088/0031-9155/48/5/303](https://doi.org/10.1088/0031-9155/48/5/303).
- [193] T. R. Nelson, D. H. Pretorius, A. Hull, M. Riccabona, M. S. Sklansky, and G. James. Sources and impact of artifacts on clinical three-dimensional ultrasound imaging. *Ultrasound in Obstetrics and Gynecology*, **16**(4):374–383, 2000. DOI: [10.1046/j.1469-0705.2000.00180.x](https://doi.org/10.1046/j.1469-0705.2000.00180.x).
- [194] V. Novak, P. Novak, J. de Champlain, A. R. Le Blanc, R. Martin, and R. Nadeau. Influence of respiration on heart rate and blood pressure fluctuations. *Journal of Applied Physiology*, **74**(2):617–626, 1993.
- [195] M. Osborne, S. Roberts, A. Rogers, S. Ramchurn, and N. Jennings. Towards real-time information processing of sensor network data using computationally efficient multi-output gaussian processes. In *International Conference on Information Processing in Sensor Networks, 2008. IPSN '08*, pp. 109–120, 2008. DOI: [10.1109/IPSN.2008.25](https://doi.org/10.1109/IPSN.2008.25).
- [196] T. Pan, T.-Y. Lee, E. Rietzel, and G. T. Chen. 4d-CT imaging of a volume influenced by respiratory motion on multi-slice CT. *Medical physics*, **31**(2):333–340, 2004.



- [197] S.-J. Park, D. Ionascu, J. Killoran, M. Mamede, V. H. Gerbaudo, L. Chin, and R. Berbeco. Evaluation of the combined effects of target size, respiratory motion and background activity on 3d and 4d PET/CT images. *Physics in Medicine and Biology*, **53**(13):3661, 2008. DOI: [10.1088/0031-9155/53/13/018](https://doi.org/10.1088/0031-9155/53/13/018).
- [198] J. Penne, C. Schaller, J. Hornegger, and T. Kuwert. Robust real-time 3d respiratory motion detection using time-of-flight cameras. *International Journal of Computer Assisted Radiology and Surgery*, **3**(5):427–431, 2008. DOI: [10.1007/s11548-008-0245-2f](https://doi.org/10.1007/s11548-008-0245-2f).
- [199] G. Pilonetto, F. Dinuzzo, and G. De Nicolao. Bayesian online multitask learning of gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**(2):193–205, 2010. DOI: [10.1109/TPAMI.2008.297](https://doi.org/10.1109/TPAMI.2008.297).
- [200] P. R. Poulsen, B. Cho, A. Sawant, D. Ruan, and P. J. Keall. Detailed analysis of latencies in image-based dynamic MLC tracking. *Medical Physics*, **37**(9):4998–5005, 2010. DOI: [10.1118/1.3480504](https://doi.org/10.1118/1.3480504).
- [201] H. F. R. Prechtel, L. A. Van Eykern, and M. J. O'Brien. Respiratory muscle EMG in newborns: a non-intrusive method. *Early Human Development*, **1**(3):265–283, 1977. DOI: [10.1016/0378-3782\(77\)90040-8](https://doi.org/10.1016/0378-3782(77)90040-8).
- [202] D. Putra, O. C. L. Haas, J. Mills, and K. Bumham. Prediction of tumour motion using interacting multiple model filter. In *3rd International Conference On Advances in Medical, Signal and Information Processing (MEDSIP)*, pp. 1–4, 2006.
- [203] D. Putra, O. C. L. Haas, J. A. Mills, and K. J. Burnham. A multiple model approach to respiratory motion prediction for real-time IGRT. *Physics in Medicine and Biology*, **53**(6):1651–1663, 2008. DOI: [10.1088/0031-9155/53/6/010](https://doi.org/10.1088/0031-9155/53/6/010).
- [204] J. Quiñonero-Candela and C. E. Rasmussen. A unifying view of sparse approximate gaussian process regression. *J. Mach. Learn. Res.*, **6**:1939–1959, 2005.
- [205] S. Quirk, N. Becker, and W. Smith. External respiratory motion: Shape analysis and custom realistic respiratory trace generation. *Medical Physics*, **39**(8):4999–5003, 2012. DOI: [10.1118/1.4737095](https://doi.org/10.1118/1.4737095).
- [206] L. Ramrath, A. Schlaefel, F. Ernst, S. Dieterich, and A. Schweikard. Prediction of respiratory motion with a multi-frequency based Extended Kalman Filter. In *Proceedings of the 21st International Conference and Exhibition on Computer Assisted Radiology and Surgery (CARS'07)*, pp. 56–58. CARS. Published in *International Journal of CARS*, **2**:56–58, 2007. DOI: [10.1007/s11548-007-0083-7](https://doi.org/10.1007/s11548-007-0083-7).
- [207] C. E. Rasmussen and H. Nickisch. Gaussian processes for machine learning (GPML) toolbox. *J. Mach. Learn. Res.*, **11**:3011–3015, 2010.
- [208] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. MIT Press, 2006. ISBN 0-262-18253-X.
- [209] M. Raux, C. Straus, S. Redolfi, C. Morelot-Panzini, A. Couturier, F. Hug, and T. Similowski. Electroencephalographic evidence for pre-motor cortex activation during inspiratory loading in humans. *The Journal of Physiology*, **578**(2):569–578, 2007. DOI: [10.1113/jphysiol.2006.120246](https://doi.org/10.1113/jphysiol.2006.120246).
- [210] Q. Ren, S. Nishioka, H. Shirato, and R. I. Berbeco. Adaptive prediction of respiratory motion for motion compensation radiotherapy. *Physics in Medicine and Biology*, **52**(22):6651, 2007. DOI: [10.1088/0031-9155/52/22/007](https://doi.org/10.1088/0031-9155/52/22/007).
- [211] N. Riaz, P. Shanker, R. Wiersma, O. Gudmundsson, W. Mao, B. Widrow, and L. Xing. Predicting respiratory tumor motion with multi-dimensional adaptive filters and support vector regression. *Physics in Medicine and Biology*, **54**(19):5735–5748, 2009. DOI: [10.1088/0031-9155/54/19/005](https://doi.org/10.1088/0031-9155/54/19/005).

- [212] R. Richa, P. Pognet, and C. Liu. Efficient 3d tracking for motion compensation in beating heart surgery. In *Medical Image Computing and Computer-Assisted Intervention – (MICCAI)*, pp. 684–691, New York, USA. Published in. *LNCS*, 5242:684–691, 2008. DOI: [10.1007/978-3-540-85990-1\\_82](https://doi.org/10.1007/978-3-540-85990-1_82).
- [213] L. Richter, F. Ernst, V. Martens, L. Matthäus, and A. Schweikard. Client/server framework for robot control in medical assistance systems. In *Proceedings of the 24th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS)*, pp. 306–307, Geneva, Switzerland, 2010.
- [214] S. R. Riesner. *Korrelations-und Prädiktionsverfahren zur Lageverfolgung in der perkutanen Radioonkologie*. PhD thesis, Technische Universität München, Universitätsbibliothek, 2004. <http://tumblr.biblio.tu-muenchen.de/publ/diss/in/2004/riesner.html>.
- [215] E.-J. Rijkhorst, I. Rivens, G. t. Haar, D. Hawkes, and D. Barratt. Effects of respiratory liver motion on heating for gated and model-based motion-compensated high-intensity focused ultrasound ablation. In *Proceedings of the 14th International Conference on Medical Image Computing and Computer-assisted Intervention (MICCAI)*, pp. 605–612, Toronto, Canada. Published in. *LNCS*, pp. 605–612, 2011.
- [216] C. J. Ritchie, J. D. Godwin, C. R. Crawford, W. Stanford, H. Anno, and Y. Kim. Minimum scan speeds for suppression of motion artifacts in CT. *Radiology*, 185(1):37–42, 1992. DOI: [10.1148/radiology.185.1.1523332](https://doi.org/10.1148/radiology.185.1.1523332).
- [217] C. Riviere. Robotic compensation of biological motion to enhance surgical accuracy. *PROCEEDINGS-IEEE*, 94(9):1705–1716, 2006. DOI: [10.1109/JPROC.2006.880722](https://doi.org/10.1109/JPROC.2006.880722).
- [218] C. Riviere, R. Rader, and N. Thakor. Adaptive cancelling of physiological tremor for improved precision in microsurgery. *IEEE Transactions on Biomedical Engineering*, 45(7):839–846, 1998. DOI: [10.1109/10.686791](https://doi.org/10.1109/10.686791).
- [219] D. Ruan. Kernel density estimation-based real-time prediction for respiratory motion. *Physics in Medicine and Biology*, 55(5):1311, 2010. DOI: [10.1088/0031-9155/55/5/004](https://doi.org/10.1088/0031-9155/55/5/004).
- [220] D. Ruan. Prospective detection of large prediction errors: a hypothesis testing approach. *Physics in Medicine and Biology*, 55(13):3885, 2010. DOI: [10.1088/0031-9155/55/13/021](https://doi.org/10.1088/0031-9155/55/13/021).
- [221] D. Ruan and P. Keall. Online prediction of respiratory motion: multidimensional processing with low-dimensional feature learning. *Physics in Medicine and Biology*, 55(11):3011, 2010. DOI: [10.1088/0031-9155/55/11/002](https://doi.org/10.1088/0031-9155/55/11/002).
- [222] D. Ruan, J. A. Fessler, and J. M. Balter. Real-time prediction of respiratory motion based on local regression methods. *Physics in Medicine and Biology*, 52(23):7137–7152, 2007. DOI: [10.1088/0031-9155/52/23/024](https://doi.org/10.1088/0031-9155/52/23/024).
- [223] D. Ruan, J. A. Fessler, and J. M. Balter. Mean position tracking of respiratory motion. *Medical Physics*, 35(2):782–792, 2008. DOI: [10.1118/1.2825616](https://doi.org/10.1118/1.2825616).
- [224] D. Ruan, J. A. Fessler, J. M. Balter, R. I. Berbeco, S. Nishioka, and H. Shirato. Inference of hysteretic respiratory tumor motion from external surrogates: a state augmentation approach. *Physics in Medicine and Biology*, 53(11):2923, 2008. DOI: [10.1088/0031-9155/53/11/011](https://doi.org/10.1088/0031-9155/53/11/011).
- [225] J. E. Rubin, B. J. Bacak, Y. I. Molkov, N. A. Shevtsova, J. C. Smith, and I. A. Rybak. Interacting oscillations in neural control of breathing: modeling and qualitative analysis. *Journal of Computational Neuroscience*, 30(3):607–632, 2010. DOI: [10.1007/s10827-010-0281-0](https://doi.org/10.1007/s10827-010-0281-0).
- [226] A. Sahih, O. C. Haas, J. H. Goodband, D. Putra, J. A. Mills, and K. J. Burnham. Respiratory motion prediction for adaptive radiotherapy. In *IAR Annual meeting*, 2006.

- [227] S. Sayeh, J. Wang, W. T. Main, W. Kilby, and C. R. M. Jr. Respiratory motion tracking for robotic radiosurgery. In *Treating Tumors that Move with Respiration*, pp. 15–29. Springer Berlin Heidelberg, 2007. DOI: [10.1007/978-3-540-69886-9\\_2](https://doi.org/10.1007/978-3-540-69886-9_2).
- [228] T. J. Scarbrough, N. M. Golden, J. Y. Ting, C. D. Fuller, A. Wong, P. A. Kupelian, and C. R. Thomas Jr. Comparison of ultrasound and implanted seed marker prostate localization methods: Implications for image-guided radiotherapy. *International Journal of Radiation Oncology\*Biophysics*, **65**(2):378–387, 2006. DOI: [10.1016/j.ijrobp.2006.01.008](https://doi.org/10.1016/j.ijrobp.2006.01.008).
- [229] J. Schaerer, A. Fassi, M. Riboldi, P. Cerveri, G. Baroni, and D. Sarrut. Multi-dimensional respiratory motion tracking from markerless optical surface imaging based on deformable mesh registration. *Physics in medicine and biology*, **57**(2):357–373, 2012. DOI: [10.1088/0031-9155/57/2/357](https://doi.org/10.1088/0031-9155/57/2/357).
- [230] C. Schaller, J. Penne, and J. Hornegger. Time-of-flight sensor for respiratory motion gating. *Medical Physics*, **35**(7):3090–3093, 2008. DOI: [10.1118/1.2938521](https://doi.org/10.1118/1.2938521).
- [231] A. Schweikard, G. Glosser, M. Bodduluri, M. J. Murphy, and J. R. Adler. Robotic motion compensation for respiratory movement during radiosurgery. *Computer Aided Surgery*, **5**(4):263–277, 2000. DOI: [10.1002/1097-0150\(2000\)5:4<263::AID-IGS5>3.0.CO;2-2](https://doi.org/10.1002/1097-0150(2000)5:4<263::AID-IGS5>3.0.CO;2-2).
- [232] A. Schweikard, H. Shiomi, and J. Adler. Respiration tracking in radiosurgery. *Medical Physics*, **31**(10):2738–2741, 2004. DOI: [10.1118/1.1774132](https://doi.org/10.1118/1.1774132).
- [233] A. Schweikard, H. Shiomi, J. Fisseler, M. Dötter, K. Berlinger, H.-B. Gehl, and J. Adler. Fiducial-less respiration tracking in radiosurgery. In C. Barillot, D. R. Haynor, and P. Hellier, editors, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3217 of *LNCS*, pp. 992–999. Springer Berlin Heidelberg, 2004. DOI: [10.1007/978-3-540-30136-3\\_120](https://doi.org/10.1007/978-3-540-30136-3_120).
- [234] A. Schweikard, H. Shiomi, and J. Adler. Respiration tracking in radiosurgery without fiducials. *The International Journal of Medical Robotics and Computer Assisted Surgery*, **1**(2):19–27, 2005. DOI: [10.1002/rcs.38](https://doi.org/10.1002/rcs.38).
- [235] M. Seeger, C. K. Williams, and N. D. Lawrence. Fast forward selection to speed up sparse gaussian process regression. In *Workshop on Artificial Intelligence and Statistics*, 2003.
- [236] Y. Seppenwoolde, H. Shirato, K. Kitamura, S. Shimizu, M. van Herk, J. V. Lebesque, and K. Miyasaka. Precise and real-time measurement of 3d tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. *International Journal of Radiation Oncology\*Biophysics*, **53**(4):822–834, 2002. DOI: [10.1016/S0360-3016\(02\)02803-1](https://doi.org/10.1016/S0360-3016(02)02803-1).
- [237] Y. Seppenwoolde, R. I. Berbeco, S. Nishioka, H. Shirato, and B. Heijmen. Accuracy of tumor motion compensation algorithm from a robotic respiratory tracking system: A simulation study. *Medical Physics*, **34**(7):2774–2784, 2007. DOI: [10.1118/1.2739811](https://doi.org/10.1118/1.2739811).
- [238] M. Seregini, P. Cerveri, M. Riboldi, A. Pella, and G. Baroni. Robustness of external/internal correlation models for real-time tumor tracking to breathing motion variations. *Physics in Medicine and Biology*, **57**(21):7053, 2012. DOI: [10.1088/0031-9155/57/21/7053](https://doi.org/10.1088/0031-9155/57/21/7053).
- [239] M. Seregini, R. Kaderka, G. Fattori, M. Riboldi, A. Pella, A. Constantinescu, N. Saito, M. Durante, P. Cerveri, C. Bert, and G. Baroni. Tumor tracking based on correlation models in scanned ion beam therapy: an experimental study. *Physics in Medicine and Biology*, **58**(13):4659, 2013. DOI: [10.1088/0031-9155/58/13/4659](https://doi.org/10.1088/0031-9155/58/13/4659).
- [240] M. Seregini, A. Pella, M. Riboldi, R. Orecchia, P. Cerveri, and G. Baroni. Real-time tumor tracking with an artificial neural networks-based method: A feasibility study. *Physica Medica*, **29**(1):48–59, 2013. DOI: [10.1016/j.ejmp.2011.11.005](https://doi.org/10.1016/j.ejmp.2011.11.005).

- [241] G. C. Sharp, S. B. Jiang, S. Shimizu, and H. Shirato. Prediction of respiratory tumour motion for real-time image-guided radiotherapy. *Physics in Medicine and Biology*, **49**(3):425, 2004. DOI: [10.1088/0031-9155/49/3/006](https://doi.org/10.1088/0031-9155/49/3/006).
- [242] Y. Sheng, S. Li, S. Sayeh, J. Wang, and H. Wang. Fuzzy and hybrid prediction of position signal in synchrony&reg; respiratory tracking system. In *International Conference on Signal and Image Processing (IASTED)*, pp. 459–464, Anaheim, CA, USA, 2007.
- [243] H. Shirato, S. Shimizu, T. Kunieda, K. Kitamura, M. van Herk, K. Kagei, T. Nishioka, S. Hashimoto, K. Fujita, H. Aoyama, K. Tsuchiya, K. Kudo, and K. Miyasaka. Physical aspects of a real-time tumor-tracking system for gated radiotherapy. *International Journal of Radiation Oncology\*Biolog\*Physics*, **48**(4):1187–1195, 2000. DOI: [10.1016/S0360-3016\(00\)00748-3](https://doi.org/10.1016/S0360-3016(00)00748-3).
- [244] H. Shirato, Y. Seppenwoolde, K. Kitamura, R. Onimura, and S. Shimizu. Intrafractional tumor motion: lung and liver. *Seminars in Radiation Oncology*, **14**(1):10–18, 2004. DOI: [10.1053/j.semradonc.2003.10.008](https://doi.org/10.1053/j.semradonc.2003.10.008).
- [245] H. Shirato, K. Suzuki, G. C. Sharp, K. Fujita, R. Onimaru, M. Fujino, N. Kato, Y. Osaka, R. Kinoshita, H. Taguchi, S. Onodera, and K. Miyasaka. Speed and amplitude of lung tumor motion precisely detected in four-dimensional setup and in real-time tumor-tracking radiotherapy. *International Journal of Radiation Oncology\*Biolog\*Physics*, **64**(4):1229–1236, 2006. DOI: [10.1016/j.ijrobp.2005.11.016](https://doi.org/10.1016/j.ijrobp.2005.11.016).
- [246] C. A. Sinderby, J. C. Beck, L. H. Lindstrom, and A. E. Grassino. Enhancement of signal quality in esophageal recordings of diaphragm EMG. *J Appl Physiol*, **82**(4):1370–1377, 1997.
- [247] S. Singhy and C. Riviere. Physiological tremor amplitude during retinal microsurgery. In *Proceedings of the IEEE 28th Annual Northeast Bioengineering Conference*, pp. 171–172, 2002. DOI: [10.1109/NEBC.2002.999520](https://doi.org/10.1109/NEBC.2002.999520).
- [248] G. Skolidis and G. Sanguinetti. Bayesian multitask classification with gaussian process priors. *IEEE Transactions on Neural Networks*, **22**(12):2011–2021, 2011. DOI: [10.1109/TNN.2011.2168568](https://doi.org/10.1109/TNN.2011.2168568).
- [249] P. Skworcow, D. Putra, A. Sahih, J. Goodband, O. C. Haas, K. J. Burnham, and J. A. Mills. Predictive tracking for respiratory induced motion compensation in adaptive radiotherapy. *MEASUREMENT AND CONTROL - UKACC Control*, **40**(1):203–210, 2006.
- [250] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, **14**(3):199–222, 2004. DOI: [10.1023/B:STCO.0000035301.49549.88](https://doi.org/10.1023/B:STCO.0000035301.49549.88).
- [251] E. Snelson and Z. Ghahramani. Sparse gaussian processes using pseudo-inputs. In *Advances in neural information processing systems (NIPS)*, Vancouver, Canada, 2006.
- [252] E. Snelson and Z. Ghahramani. Local and global sparse gaussian process approximations. In *International Conference on Artificial Intelligence and Statistics*, pp. 524–531, 2007.
- [253] O. Stegle, S. Fallert, D. J. C. MacKay, and S. Brage. Gaussian process robust regression for noisy heart rate data. *IEEE Transactions on Biomedical Engineering*, **55**(9):2143–2151, 2008. DOI: [10.1109/TBME.2008.923118](https://doi.org/10.1109/TBME.2008.923118).
- [254] J. H. Steiger. Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, **87**(2):245, 1980. DOI: [10.1037/0033-2909.87.2.245](https://doi.org/10.1037/0033-2909.87.2.245).
- [255] M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer, 1999. ISBN 978-0387986296.
- [256] E. M. Summerhill, Y. A. El-Sameed, T. J. Glidden, and F. D. McCool. MOnitoring recovery from diaphragm paralysis with ultrasound\*. *CHEST Journal*, **133**(3):737–743, 2008. DOI: [10.1378/chest.07-2200](https://doi.org/10.1378/chest.07-2200).

- [257] I. Suramo, M. Päivänsalo, and V. Myllylä. Cranio-caudal movements of the liver, pancreas and kidneys in respiration. *Acta radiologica: diagnosis*, **25**(2):129–131, 1983.
- [258] R. A. Sweeney, W. Arnold, E. Steixner, M. Nevinny-Stickel, and P. Lukas. Compensating for tumor motion by a 6-degree-of-freedom treatment couch: Is patient tolerance an issue? *International Journal of Radiation Oncology\*Biolog\*Physics*, **74**(1):168–171, 2009. DOI: [10.1016/j.ijrobp.2008.07.069](https://doi.org/10.1016/j.ijrobp.2008.07.069).
- [259] M. B. Tacke, S. Nill, A. Krauss, and U. Oelfke. Real-time tumor tracking: Automatic compensation of target motion using the siemens 160 MLC. *Medical Physics*, **37**(2):753–761, 2010. DOI: [10.1118/1.3284543](https://doi.org/10.1118/1.3284543).
- [260] R. K. Ten Haken, J. D. Forman, D. K. Heimburger, A. Gerhardsson, D. L. McShan, C. Perez-Tamayo, S. L. Schoepel, and A. S. Lichter. Treatment planning issues related to prostate movement in response to differential filling of the rectum and bladder. *International Journal of Radiation Oncology\*Biolog\*Physics*, **20**(6):1317–1324, 1991. DOI: [10.1016/0360-3016\(91\)90244-X](https://doi.org/10.1016/0360-3016(91)90244-X).
- [261] M. E. Tipping. Sparse bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.*, **1**: 211–244, 2001. DOI: [10.1162/15324430152748236](https://doi.org/10.1162/15324430152748236).
- [262] A. E. Torshabi, A. Pella, M. Riboldi, and G. Baroni. Targeting accuracy in real-time tumor tracking via external surrogates: a comparative study. *Technology in cancer research & treatment*, **9**(6):551–562, 2010.
- [263] S. Van Vaerenbergh and I. Santamaria. A comparative study of kernel adaptive filtering algorithms. In *Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, pp. 181–186. IEEE, 2013.
- [264] S. S. Vedam, P. J. Keall, V. R. Kini, and R. Mohan. Determining parameters for respiration-gated radiotherapy. *Medical Physics*, **28**(10):2139–2146, 2001. DOI: [10.1118/1.1406524](https://doi.org/10.1118/1.1406524).
- [265] S. S. Vedam, V. R. Kini, P. J. Keall, V. Ramakrishnan, H. Mostafavi, and R. Mohan. Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker. *Medical physics*, **30**(4):505–513, 2003. DOI: [10.1118/1.1558675](https://doi.org/10.1118/1.1558675).
- [266] S. S. Vedam, P. J. Keall, A. Docef, D. A. Todor, V. R. Kini, and R. Mohan. Predicting respiratory motion for four-dimensional radiotherapy. *Medical Physics*, **31**(8):2274–2283, 2004. DOI: [10.1118/1.1771931](https://doi.org/10.1118/1.1771931).
- [267] M. Verleysen and D. François. The curse of dimensionality in data mining and time series prediction. In *Computational Intelligence and Bioinspired Systems*, pp. 758–770. Springer, 2005. DOI: [10.1007/11494669\\_93](https://doi.org/10.1007/11494669_93).
- [268] P. S. Verma, H. Wu, M. P. Langer, I. J. Das, and G. Sandison. Survey: Real-time tumor motion prediction for image-guided radiation treatment. *Computing in Science & Engineering*, **13**(5):24–35, 2011. DOI: [10.1109/MCSE.2010.99](https://doi.org/10.1109/MCSE.2010.99).
- [269] H. Wackernagel. *Multivariate Geostatistics*. Springer, 2003. ISBN 9783540441427.
- [270] O. L. Wade. Movements of the thoracic cage and diaphragm in respiration. *The Journal of physiology*, **124**(2):193–212, 1954.
- [271] Y. Wang and R. Khordon. Sparse gaussian processes for multi-task learning. In *Machine Learning and Knowledge Discovery in Databases*, pp. 711–727. Springer, 2012.
- [272] M. Wannemacher, J. Debus, and F. Wenz. *Strahlentherapie*. Springer, Berlin, 2006. ISBN 9783540228127.
- [273] E. Weiss, K. Wijesooriya, S. V. Dill, and P. J. Keall. Tumor and normal tissue motion in the thorax during respiration: Analysis of volumetric and positional variations using 4d CT. *International Journal of Radiation Oncology\*Biolog\*Physics*, **67**(1):296–307, 2007. DOI: [10.1016/j.ijrobp.2006.09.009](https://doi.org/10.1016/j.ijrobp.2006.09.009).

- [274] G. Welch. An introduction to the kalman filter. 2006, 2006.
- [275] R. Werner, J. Ehrhardt, A. Schmidt-Richberg, A. Heiß, and H. Handels. Estimation of motion fields by non-linear registration for local lung motion analysis in 4d CT image data. *International Journal of Computer Assisted Radiology and Surgery*, 5(6):595–605, 2010. DOI: [10.1007/s11548-010-0418-7](https://doi.org/10.1007/s11548-010-0418-7).
- [276] J. Wilbert, J. Meyer, K. Baier, M. Guckenberger, C. Herrmann, R. Hess, C. Janka, L. Ma, T. Mersebach, A. Richter, M. Roth, K. Schilling, and M. Flentje. Tumor tracking and motion compensation with an adaptive tumor tracking system (ATTS): System description and prototype testing. *Medical Physics*, 35(9):3911–3921, 2008. DOI: [10.1118/1.2964090](https://doi.org/10.1118/1.2964090).
- [277] H. Wu, G. C. Sharp, B. Salzberg, D. Kaeli, H. Shirato, and S. B. Jiang. A finite state model for respiratory motion analysis in image guided radiation therapy. *Physics in Medicine and Biology*, 49(23):5357, 2004. DOI: [10.1088/0031-9155/49/23/012](https://doi.org/10.1088/0031-9155/49/23/012).
- [278] H. Wu, B. Salzberg, G. C. Sharp, S. B. Jiang, H. Shirato, and D. Kaeli. Subsequence matching on structured time series data. In *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, pp. 682–693, New York, NY, USA, 2005. DOI: [10.1145/1066157.1066235](https://doi.org/10.1145/1066157.1066235).
- [279] H. Wu, G. C. Sharp, Q. Zhao, H. Shirato, and S. B. Jiang. Statistical analysis and correlation discovery of tumor respiratory motion. *Physics in Medicine and Biology*, 52(16):4761, 2007. DOI: [10.1088/0031-9155/52/16/004](https://doi.org/10.1088/0031-9155/52/16/004).
- [280] H. Wu, Q. Zhao, R. I. Berbeco, S. Nishioka, H. Shirato, and S. B. Jiang. Gating based on internal/external signals with dynamic correlation updates. *Physics in medicine and biology*, 53(24):7137, 2008. DOI: [10.1088/0031-9155/53/24/009](https://doi.org/10.1088/0031-9155/53/24/009).
- [281] H. Wu, Q. Zhao, and L. Zhao. Knowledge discovery from tumor respiratory motion data. In *International Conference on BioMedical Engineering and Informatics, 2008. BMEI 2008*, pp. 297–301, 2008. DOI: [10.1109/BMEI.2008.116](https://doi.org/10.1109/BMEI.2008.116).
- [282] H. Yan, F.-F. Yin, G.-P. Zhu, M. Ajlouni, and J. H. Kim. Adaptive prediction of internal target motion using external marker motion: a technical study. *Physics in Medicine and Biology*, 51(1):31, 2006. DOI: [10.1088/0031-9155/51/1/003](https://doi.org/10.1088/0031-9155/51/1/003).
- [283] H. Yan, F.-F. Yin, G.-P. Zhu, M. Ajlouni, and J. H. Kim. The correlation evaluation of a tumor tracking system using multiple external markers. *Medical Physics*, 33(11):4073–4084, 2006. DOI: [10.1118/1.2358830](https://doi.org/10.1118/1.2358830).
- [284] D. Yang, W. Lu, D. A. Low, J. O. Deasy, A. J. Hope, and I. E. Naqa. 4d-CT motion estimation using deformable image registration and 5d respiratory motion modeling. *Medical Physics*, 35(10):4577–4590, 2008. DOI: [10.1118/1.2977828](https://doi.org/10.1118/1.2977828).
- [285] S. G. Yuen, S. B. Kesner, N. V. Vasilyev, P. J. D. Nido, and R. D. Howe. 3d ultrasound-guided motion compensation system for beating heart mitral valve repair. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 711–719, New York, USA. Published in. *LNCS*, 5241:711–719, 2008. DOI: [10.1007/978-3-540-85988-8\\_85](https://doi.org/10.1007/978-3-540-85988-8_85).
- [286] Y. Zhang, D. Boye, C. Tanner, A. J. Lomax, and A. Knopf. Respiratory liver motion estimation and its effect on scanned proton beam therapy. *Physics in Medicine and Biology*, 57(7):1779, 2012. DOI: [10.1088/0031-9155/57/7/1779](https://doi.org/10.1088/0031-9155/57/7/1779).
- [287] T. Zhao, W. Lu, D. Yang, S. Mutic, C. E. Noel, P. J. Parikh, J. D. Bradley, and D. A. Low. Characterization of free breathing patterns with 5d lung motion model. *Medical Physics*, 36(11):5183–5189, 2009. DOI: [10.1118/1.3246348](https://doi.org/10.1118/1.3246348).

# List of Figures

1.1	Anatomical details of the respiratory system, part 1 . . . . .	2
1.2	Averaged trajectories of 21 lung tumours . . . . .	4
1.3	Anatomical details of the respiratory system, part 2 . . . . .	6
1.4	CT and MRI motion artefacts . . . . .	9
1.5	Illustration of US shadow . . . . .	11
1.6	Incidence and mortality rate of cancer in Germany . . . . .	13
1.7	Principle of external beam radiotherapy . . . . .	15
1.8	Motion compensation by increasing the target area . . . . .	15
1.9	Motion compensation by respiratory gating . . . . .	16
1.10	Motion compensation by forced shallow-breathing . . . . .	16
1.11	Clinically used motion compensation systems . . . . .	17
1.12	Motion compensation by adaptive compensation . . . . .	18
1.13	Clinically used adaptive motion compensation systems . . . . .	20
2.1	Basic notation for prediction algorithms . . . . .	26
2.2	Basic notation for correlation algorithms . . . . .	27
2.3	Example evaluation measure . . . . .	32
2.4	Illustration of an artificial neuronal network . . . . .	39
2.5	Example of correlation data . . . . .	47
3.1	Regular, irregular, and noisy-irregular motion fragments . . . . .	57
3.2	Illustration of box probability example . . . . .	58
3.3	Example of 2D Gaussian distributions . . . . .	60
3.4	Results RVM depending on feature dimension and number of training pairs	65
3.5	Results RVM depending on number of iterations . . . . .	66
3.6	Example functions drawn from different Gaussian processes . . . . .	71
3.7	Results GP depending on number of training pairs . . . . .	77
3.8	Cumulative $RMSE_{rel}$ and $\Delta RMSE_{rel}$ histogram of RVM and GP algorithms for $h = 77$ ms . . . . .	81

3.9	Cumulative $RMSE_{rel}$ and $\Delta RMSE_{rel}$ histogram of RVM and GP algorithms for $h = 115$ ms . . . . .	82
3.10	Cumulative $RMSE_{rel}$ and $\Delta RMSE_{rel}$ histogram of RVM and GP algorithms for $h = 154$ ms . . . . .	83
3.11	Cumulative $RMSE_{rel}$ and $\Delta RMSE_{rel}$ histogram of RVM and GP algorithms for $h = 308$ ms . . . . .	84
3.12	Prediction error and variance for three motion fragments . . . . .	90
3.13	Enlarged prediction error and variance . . . . .	91
3.14	Duty cycle and precision depending on the variance threshold $\sigma_{th}^2$ . . . . .	92
3.15	Duty cycle and precision depending on the error threshold $d_{th}$ . . . . .	94
3.16	Cumulative $RMSE_{rel}$ and $\Delta RMSE_{rel}$ histogram of $HYB_{RVM}$ algorithm . . . . .	95
3.17	Cumulative $\Delta RMSE_{rel}$ histogram of wLMS-RVM hybrid algorithms . . . . .	96
3.18	Cumulative $RMSE_{rel}$ histogram of all hybrid algorithms . . . . .	98
4.1	Multivariate measurement: connection of measurement units . . . . .	105
4.2	Multivariate measurement: sensor placement . . . . .	108
4.3	Multivariate measurement: example of acquired data I . . . . .	109
4.4	Illustration feature space . . . . .	112
4.5	$RMSE_{rel}$ of the RVM algorithm depending on feature dimension $d$ . . . . .	113
4.6	Multivariate measurement: example of acquired data II . . . . .	118
4.7	Mean absolute correlation coefficient $ r $ of measurement phase M1 . . . . .	119
4.8	Correlation coefficient difference $ \Delta r $ of measurement phase M1 . . . . .	120
4.9	Correlation coefficient $ r $ for the regular and irregular breathing segments of measurement phase M2 . . . . .	121
4.10	Distribution of the $RMSE_{rel}$ values for the scenario TSUK . . . . .	123
4.11	Distribution of the $\Delta RMSE_{rel}$ values for the scenario TSUK . . . . .	123
4.12	Example of multivariate prediction results . . . . .	127
4.13	EOG and EEG raw data . . . . .	131
4.14	Correlation coefficient $ r $ between the EEG features and OMs . . . . .	132
4.15	Postprocessing steps of sEMG data . . . . .	133
4.16	Correlation coefficient $ r $ between the sEMG features and OMs . . . . .	134
5.1	Acquired data of porcine study . . . . .	140
5.2	Schematic diagram of STGPs and MTGPs . . . . .	142
5.3	MTGP Example I.a . . . . .	147
5.4	MTGP Example I.b: correlation . . . . .	148
5.5	MTGP Example II: simulated data . . . . .	149



5.6	MTGP example II: correlation results . . . . .	150
5.7	MTGP example III: time shift . . . . .	151
5.8	Illustration of alternative correlation and prediction, and MTGP models . . . . .	155
5.9	Training and test data of porcine study . . . . .	156
5.10	Correlation evaluation between external and internal marker . . . . .	158
5.11	Combined prediction and correlation result of the MTGP models . . . . .	159
5.12	Outcome of DP-wLMS, SVR-wLMS, and $MTGP_{QP}^{best}$ of an outlier . . . . .	161
5.13	RMSE of MTGP and alternative models depending on the prediction horizons and the number of internal training points . . . . .	162
5.14	Internal and external data of outlier . . . . .	168
5.15	MTGP example IV: LMC . . . . .	169
5.16	MTGP Example V: convoluted kernels . . . . .	170
6.1	Cumulative $\Delta RMSE_{rel}$ difference of RVM and GP algorithm . . . . .	177
6.2	Cumulative $\Delta RMSE_{rel}$ difference of hybrid algorithms . . . . .	180
6.3	Comparison of motion compensation with CyberKnife <sup>®</sup> Synchrony system and MTGP based system . . . . .	182
6.4	Outlook: Motion compensation using US with CyberKnife <sup>®</sup> Synchrony and MTGP based system . . . . .	184



# List of Tables

1.1	Prevalence and five-year survival rate of different cancer types . . . . .	14
1.2	Main questions of this work and overview in which chapter they will be addressed. . . . .	21
2.1	Result of different evaluation measure . . . . .	33
2.2	Overview of prediction algorithms . . . . .	44
2.3	Overview of correlation algorithms . . . . .	49
3.1	Summary of main results of comparative study of Ernst <i>et al.</i> [86] . . . . .	56
3.2	Example of joint, marginal, and conditional probability . . . . .	58
3.3	Outline of RVM tracking algorithm . . . . .	64
3.4	$RMSE_{rel}$ of RVM depending on the number of iteration $n_{iter}$ . . . . .	67
3.5	Outline of a GP tracking algorithm . . . . .	73
3.6	NLML values of a GP models depending on the covariance functions . . . . .	76
3.7	$RMSE_{rel}$ of GP models depending on the number of iteration $n_{iter}$ . . . . .	78
3.8	RMSE and $RMSE_{rel}$ for the RVM and GP algorithms . . . . .	86
3.9	RMSE, $RMSE_{rel}$ and mAE for all hybrid algorithms . . . . .	97
4.1	Summary of subject population divided into female and male . . . . .	104
4.2	Overview of algorithm specific parameters . . . . .	117
4.3	Multivariate prediction results for measurement phase M1 and M2 . . . . .	122
4.4	Occurrence of features in the feature sets for M1 . . . . .	124
4.5	Occurrence of features in the feature sets for M2 . . . . .	126
4.6	Prediction results using sEMG features . . . . .	135
5.1	MTGP example I.a: training intervals, sampling frequency $f_s$ , and Pearson's correlation coefficient . . . . .	146
5.2	MTGP example I.a: results . . . . .	146
5.3	MTGP example III: results . . . . .	153
5.4	Prediction and correlation results of alternative and MTGP models for E2 . . . . .	160

5.5	Prediction and correlation results of multivariate alternative and MTGP models for E3 . . . . .	164
5.6	Occurrence of external markers in the optimal feature sets . . . . .	164
5.7	MTGP example IV: results . . . . .	168
5.8	MTGP example V: results . . . . .	171
6.1	Adaptive motion compensation using CyberKnife <sup>®</sup> Synchrony or an MTGP model . . . . .	182





# Glossary and List of Abbreviations

## **ACC**

Acceleration.

## **AP**

Anterior-posterior.

## **Artificial Neuronal Network (ANN)**

Machine learning technique which is inspired by the biological behaviour of neurons in the brain. Multiple neurons are connected within an artificial network.

## **Autoregressive (AR)**

A linear model for a stationary, time-discrete stochastic process consisting of only an autoregressive model (see [ARMA](#)).

## **Autoregressive Moving Average (ARMA)**

A linear model for a stationary, time-discrete stochastic process consisting of an autoregressive and a moving average model.

## **CA**

Constant acceleration (refers here to a Kalman filter assuming constant acceleration).

## **Computer Tomography (CT)**

Anatomical imaging method to compute 3D volumes using multiple [X-ray](#) images.

## **CV**

Constant velocity (refers here to a Kalman filter assuming constant velocity).

## **DC**

Duty cycle (definition sec. [2.2](#)).

## **DP**

Dual-polynomial correlation model consisting of two quadratic polynomial functions - one for inspiration and one for expiration (see sec. 5.2.5).

## **DP-wLMS**

Combination of a two quadratic (dual-)polynomial correlation and a wLMS prediction [81] model for adaptive motion compensation (see sec. 5.2.5).

## **Electrocardiography (ECG)**

Diagnostic technique to measure the electrical activity of the heart.

## **Electroencephalography (EEG)**

Diagnostic technique which measures the electrical potentials of the brain via highly sensitive electrodes placed on the skull.

## **Electromyography (EMG)**

Medical technique to measure the electrical potentials of muscles (often measured invasively via needle electrodes).

## **Electrooculography (EOG)**

Diagnostic technique which measures the electrical voltages between two electrodes to evaluate the movements of the eyes or the changes of the resting potential of the retina.

## **Extended Kalman Filter (EKF)**

Extension of the KF approach to nonlinear systems by using the first order Taylor expansion.

## **Feature Set (FS)**

The term is used to simplify the notations and group features of one or multiple sensors.

## **Functional Magnetic Resonance Imaging (fMRI)**

Functional imaging method which is based on MRI.



### **Gaussian Process (GP)**

A Gaussian process is probability distribution over functions where each subset follows a multivariate Gaussian distribution. The term is used here in the context of GP models which can be used for machine learning problems.

### **High-Intensity Focused Ultrasound (HIFU)**

Medical procedure to ablate tissue by using focused ultrasound.

### **ILM**

Internal landmark.

### **Kalman Filter (KF)**

Technique to model linear dynamic systems which was introduced by R. E. Kálmán.

### **Kernel Adaptive Filtering (KAF)**

Covers a group of algorithms which combines the principle idea of adaptive filters, such as **LMS**, **nLMS**, and **RLS**, and kernel methods.

### **Kernel Density Estimation (KDE)**

Probabilistic machine learning algorithm which was introduced by Ruan *et al.* [219] for respiratory motion prediction.

### **Least Mean Square (LMS)**

Algorithm to approximate a least mean squares problem which is used in digital signal processing as filter method and prediction technique. The algorithm uses a gradient descent method.

### **Linear Accelerator (LINAC)**

Device used in radiotherapy which accelerates electrons on a target to generate  $\gamma$ -radiation.

### **Linear Model of Coregularization (LMC)**

Extension of a **MTGP** model which considers multiple correlations between signals.

### **MAE**

Mean absolute error (definition sec. 2.2).

**mAE**

Maximum absolute error (definition sec. 2.2).

**Magnetic Resonance (MR)**

See [MRI](#).

**Magnetic Resonance Imaging (MRI)**

Anatomical imaging method which is based on principles of nuclear magnetic resonance.

**ME**

Mean error (definition sec. 2.2).

**Motion compensation (MC)**

Term used in the context of MC algorithms which refers to all algorithms which can be used for adaptive motion compensation.

**MSLL**

Mean standardized log loss (definition sec. 5.2).

**Multi-Step Linear Method (MULIN)**

Regression method which is based on a Taylor-expansion of the prediction error.

**Multi-Task Gaussian Process (MTGP)**

Multi-task extension of a STGP which can be used to model multiple time-series simultaneously.

**Multileaf Collimator (MLC)**

Device which is used to shape the beam of a radiotherapy device and consists of multiple moveable leaves.

**NLML**

Negative logarithmic marginal likelihood (definition sec. 3.2.2.1).

**Normalised Least Mean Squares (nLMS)**

Normalised version of a [LMS](#) algorithm which increases the stability of the method.

**nRMSE**

Root mean square error normalized by variance of the true signal (definition sec. 2.2) [184].

**OM**

Optical marker.

**P**

Polynomial correlation model consisting of one quadratic polynomial function for inspiration and expiration (see sec. 5.2.5).

**P-wLMS**

Combination of a quadratic polynomial correlation and a wLMS prediction [81] model for adaptive motion compensation (see sec. 5.2.5).

**Positron Emission Tomography (PET)**

Functional imaging method which is based on nuclear injected tracers.

**pp**

Percentage points.

**PR**

Precision (definition sec. 3.47).

**Radiofrequency (RF)**

Term used in the context of RF ablation which refers to medical procedure to ablate tissue by use of alternating current ( $f_s = 350 - 500$  kHz).

**RBF**

Radial basis function.

**Real-Time Position Management (RPM)**

Term refers to the real-time position management system of Varian which can be used for acquiring data of the external movement of the torso or abdomen.

**Recursive Least Squares (RLS)**

Adaptive algorithm which optimizes the parameter of a least squares problem recursively by considering the complete data sequence.

**Relevance Vector Machine (RVM)**

A supervised probabilistic learning method which was first proposed by Tipping *et al.* [261]. It can be used for classification or regression problems.

**RL**

Right-left.

**RMSE**

Root mean square error (definition sec. 2.2).

**RMSE<sub>rel</sub>**

Relative root mean square error (definition sec. 2.2) [81]. Measure is only valid for prediction algorithms (RMSE<sub>rel</sub> < 100 % - improvement compared to no prediction; RMSE<sub>rel</sub> > 100 % - no improvement compared to no prediction).

**SFS**

Refers to a sequential forward selection method (see sec. 4.3).

**SI**

Superior-inferior.

**Single-Task Gaussian Process (STGP)**

Supervised probabilistic model which is based on GPs model as defined in sec. 3.2.2 which can be used for time-series prediction.

**SNR**

Signal-to-noise ratio.

**Stereotactic Body Radiation Therapy (SBRT)**

External radiation beam therapy with the focus on tumours outside of the central nervous system.

**Support Vector Regression (SVR)**

A supervised learning method which can be used for regression problems. It is based on the support vector machines which are a maximum margin classifier.

**Surface Electromyography (sEMG)**

Non-invasive measurement technique to acquire the electrical activity of muscles (see EMG). Potentials are often measured by electrodes placed on the skin.

**SUV**

Standard uptake value.

**SVR-wLMS**

Combination of a **SVR** correlation [83] and a **wLMS** prediction [81] model for adaptive motion compensation (see sec. 5.2.5).

**TSK**

Test set known (refers to one evaluation scenarios where the test set is known before evaluation; see sec. 4.3).

**TSUK**

Test set unknown (refers to one evaluation scenarios where the test set is unknown before evaluation; see sec. 4.3).

**Ultrasound (US)**

Imaging method based on ultrasound.

**Vegetative Nervous System (VNS)**

The VNS is part of the peripheral nervous system which is essential for maintaining homeostasis and controls functions such as heart rate, respiration, blood pressure, and digestion..

**Wavelet-based LMS (wLMS)**

Extension of a **LMS** algorithm by using a wavelet decomposition on the input signal.

**X-ray**

Electromagnetic radiation which can be used for medical imaging with a wavelength  $\lambda \in [0.01, 10]$  nm.

## Acknowledgements

Mit Fertigstellung dieser Arbeit möchte ich mich insbesondere bei meinen Eltern für ihre jahrelange Unterstützung bedanken. Zeit meines Lebens waren und sind sie eine wichtige Säule, auf die ich mich auch in schwierigen Zeiten uneingeschränkt verlassen kann. Ohne Sie wäre diese Arbeit nicht entstanden.

I would also like to express my gratitude towards Prof. Dr.-Ing. Achim Schweikard and my colleagues of the Institute of Robotic and Cognitive Systems (University zu Lübeck) for their support, and the good and creative working atmosphere. I enjoyed it a lot to work with them over the last four years. In particular, I would like to thank Tobias Wisel, Dr. Floris Ernst, and Prof. Alexander Schlaefter for their constructive input and feedback. Most of this work is a result of uncountable (and sometimes seemingly endless) discussions with them. A special thanks also to Cornelia Rieckhoff, who helped with her engagement again and again to overcome the organizational obstacles in a university. I also want to acknowledge the work of my students Katharina Merkel, Lucas Davenport, and Xiao Fang, who supported me in realisation and evaluation of my experiments.

Further, I would like to thank my colleagues of the Computational Health Informatics Lab (University of Oxford). Even though, I spent only two months in Oxford, it was an extremely motivating and constructive time. I would especially like to mention Dr. David A. Clifton, Dr. Lei Clifton, and Marco A.F. Pimentel for their support and input on the fascinating topic of Multi-Task Gaussian Processes models.

Last but not least, I want to express my gratitude towards my colleagues and friends for proof reading my work and the valuable feedback. Beside already named persons, I want to mention Christoph Metzner, Alena Woeste, and Camillo Dürichen.

## Curriculum Vitae

Robert Dürichen was born on 22.10.1983 in Meißen, Germany. In 2002, he received his high school degree (Abitur) from Franziskaner-Gymnasium Meißen. After completing his compulsory one-year military service, he studied Mechatronics at the Technical University of Dresden (Germany). During his studies, he focused on medical engineering, microsystems technology, design methods, and information processing. He undertook his diploma thesis with Philips Research, Eindhoven, in the Netherlands. He graduated in 2009 as Diplom-Ingenieur (corresponding to Master of Science), after finishing a thesis entitled "Optimization of a diffuse reflectance spectroscopy method for hemoglobin measurements in biological tissue".

Further, he completed a minor study titled "Regional studies of Latin America" at the Technical University of Dresden in 2007 and studied for one semester at the Universidad de Valladolid, Spain.

Since 2010, Robert Dürichen is a member of the Institute for Robotics and Cognitive Systems, Universität zu Lübeck, and a doctoral candidate of the Graduate School for Computing in Medicine and Life Sciences (Lübeck, Germany). He is working on machine learning approaches for motion compensation techniques in robotic radiotherapy. His research interest in probabilistic approaches within the field of medical applications was strengthened through an internship in the Computation Health Informatics Lab, University of Oxford, under supervision of Dr. David Clifton in 2013.

## Erklärung an Eides Statt

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbständig sowie ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ich versichere an Eides Statt, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

---

(Ort, Datum)

---

(Unterschrift)