Multivariate regression approaches for surrogate-based diffeomorphic estimation of respiratory motion in radiation therapy

M Wilms[1,4], R Werner[2,4], J Ehrhardt[1], A Schmidt-Richberg[1,5], H-P Schlemmer[3], and H Handels[1]

[1]Institute of Medical Informatics, University of Lübeck, 23538 Lübeck, Germany
[2]Department of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany
[3]Department of Radiology, German Cancer Research Center, 69120 Heidelberg, Germany

E-mail: `wilms@imi.uni-luebeck.de`

Abstract. Breathing-induced location uncertainties of internal structures are still a relevant issue in radiation therapy of thoracic and abdominal tumours. Motion compensation approaches like gating or tumour tracking are usually driven by low-dimensional breathing signals, which are acquired in real-time during the treatment. These signals are only surrogates of the internal motion of target structures and organs at risk, and, consequently, appropriate models are needed to establish correspondence between the acquired signals and the sought internal motion patterns. In this work, we present a diffeomorphic framework for correspondence modelling based on the Log-Euclidean framework and multivariate regression. Within the framework, we systematically compare standard and subspace regression approaches (principal component regression, partial least squares, canonical correlation analysis) for different types of common breathing signals (1D: spirometry, abdominal belt, diaphragm tracking; multi-dimensional: skin surface tracking). Experiments are based on 4D CT and 4D MRI data sets and cover intra- and inter-cycle as well as intra- and inter-session motion variations. Only small differences in internal motion estimation accuracy are observed between the 1D surrogates. Increasing the surrogate dimensionality, however, improved the accuracy significantly; this is shown for both 2D signals, which consist of a common 1D signal and its time derivative, and high-dimensional signals containing the motion of many skin surface points. Eventually, comparing the standard and subspace regression variants when applied to the high-dimensional breathing signals, only small differences in terms of motion estimation accuracy are found.

Keywords: Respiratory motion, Motion estimation, Correspondence modelling, Breathing surrogates, Regression, Image registration, Radiotherapy

1. Introduction

Breathing-induced organ and tumour motion introduces a major source of uncertainty during radiotherapy treatment of thoracic and abdominal tumours (Keall *et al* 2006).

[4] These authors contributed equally to the work.
[5] A. Schmidt-Richberg is now with the Biomedical Image Analysis Group, Imperial College London, London, SW7 2AZ, UK.

Full citation: M. Wilms, R. Werner, J. Ehrhardt, A. Schmidt-Richberg, H.-P. Schlemmer and H. Handels, "Multivariate regression apporaches for surrogate-based diffeomorphic estimation of respiratory motion in radiation therapy", Phys. Med. Biol. (59) 1147-1164, 2014

Motion compensation techniques like respiratory gating (Kubo & Hill 1996) or tumour tracking (Schweikard *et al* 2000) require precise information about the location and motion of internal structures. Direct and continuous monitoring of these structures is, however, hardly feasible in clinical practise. As a consequence, gating and tracking approaches currently rely on (mainly externally acquired) breathing signals of so-called surrogates of the internal motion, which are eventually used to infer the position of the internal structures of interest by patient-specific correspondence models (McClelland *et al* 2013).

The correspondence models can be roughly characterized by three components: a surrogate, an internal motion representation, and a mathematical description of their assumed relationship. Concerning the first aspect and the given application context, breathing surrogates are, for example, abdominal belts and spirometry devices (Keall *et al* 2006). These surrogates provide only one-dimensional breathing signals. Bearing in mind that the signal dimensionality determines the degrees of freedom of the correspondence model, the resulting models can be considered as being over-restrictive in the given context. This motivates the use of multi-dimensional breathing signals, like range images of the moving skin surface (Schaller *et al* 2008). Internal motion representations, the second component of a correspondence model, range from displacements of only selected points of interest (Schweikard *et al* 2000, Cerviño *et al* 2010) over parameters of parametric transformations (Martin *et al* 2013) to dense displacement fields (Zhang *et al* 2007, Li *et al* 2011). The probably most frequent representation are dense displacement fields, which are regarded as a convenient way to describe the potentially complex non-linear motion/deformation of the tumour and organs at risk and can be derived from treatment planning 4D CT or MR image sequences by non-linear registration. Addressing the correspondence modelling itself, the last component of a correspondence model, again several approaches can be found in the literature (see, e.g., the overview by McClelland *et al* (2013)). The probably most common approach in the context of radiation therapy is to assume a linear relationship between the breathing signal and the internal motion; cf. (Schweikard *et al* 2000, Cerviño *et al* 2010, Martin *et al* 2013, Klinder *et al* 2010) as examples.

In this work, we address the following aspects: If a correspondence model is built on dense displacement fields as internal motion representation, there is usually no guarantee that transformations derived by the model are diffeomorphic (globally one-to-one, differentiable, invertible with differentiable inverse). This holds for both interpolation (signal measurement inside range of the training surrogate data), and especially for extrapolation scenarios. However, restricting the transformations to diffeomorphisms is a natural choice in the given context to ensure the topology of the internal structures to be preserved (Beg *et al* 2005). With this goal in mind, we introduce a diffeomorphic framework for surrogate-based motion estimation, which exploits the Log-Euclidean framework (Arsigny *et al* 2006) to perform statistics on diffeomorphisms, here for multivariate regression between the surrogate signals and dense diffeomorphic transformations representing the internal motion (Werner *et al* 2012a). To tackle the problem arising from correlated dimensions of high-dimensional surrogates, we further incorporate subspace regression approaches (principal component regression, partial least squares, canonical correlation analysis) in addition to a standard ordinary least squares regression (Wilms *et al* 2013). Besides these methodical aspects, what has been missing so far is a systematic comparison of the different multivariate regression approaches and common surrogates. Therefore, this work aims also to provide an extensive evaluation of the performance of these

methods and surrogates with the focus lying on lung motion estimation. Such an evaluation requires an appropriate amount of ground truth patient (4D) image data and corresponding breathing signal measurements, which unfortunately are rarely acquired in current clinical practise. Given 4D CT and 4D MRI data sets, our idea was to generate and use image-based simulations of breathing signals. For evaluation purposes, standard measures known from registration evaluation are applied to obtain a fair comparison between different models and scenarios of intra- and inter-cycle as well as intra- and inter-session motion variations.

2. A diffeomorphic framework for surrogate-based motion estimation

In the following, we briefly describe the theoretical background of our diffeomorphic surrogate-based motion estimation framework introduced in (Werner *et al* 2012a, Wilms *et al* 2013). For model training, we assume a 4D image data set $(I_j)_{j\in\{1,\ldots,n_{ph}\}}$ consisting of n_{ph} 3D images $I_j : \Omega \to \mathbb{R}$ $(\Omega \subset \mathbb{R}^3)$ representing the patient's anatomy at breathing phases j is given. Moreover, let $(\zeta_j)_{j\in\{1,\ldots,n_{ph}\}}$ denote a set of corresponding surrogate measurements $\zeta_j \in \mathbb{R}^{n_{sur}}$ with dimension n_{sur}. The first step in training a correspondence model is to determine the internal motion represented by non-linear transformations $\varphi_j : \Omega \to \Omega$ between an arbitrary reference breathing phase (here: I_1) and each I_j. As we want to define a diffeomorphic framework, the φ_j are computed using a diffeomorphic registration scheme.

2.1. Diffeomorphic image registration

In general, a diffeomorphic transformation $\varphi : \Omega \to \Omega$ is a differentiable bijective mapping with a differentiable inverse φ^{-1} (Ehrhardt *et al* 2011). Given the transport equation

$$\frac{\partial}{\partial t}\phi_t(x) = v(\phi_t(x), t) \quad \text{with} \quad \phi_0(x) = x \text{ and } \quad t \in [0,1] \ , \tag{1}$$

and a time-dependent, sufficiently smooth velocity field $v : \Omega \times [0,1] \to \mathbb{R}^3$, which parameterises the flow $\phi : \Omega \times [0,1] \to \Omega$, the sought diffeomorphic transformation φ is the solution of (1) at time $t = 1$ (Dupuis *et al* 1998, Beg *et al* 2005). However, registration algorithms based on time-varying velocity fields typically suffer from high computational and memory cost (Beg *et al* 2005). Therefore, Arsigny *et al* (2006) proposed using a stationary velocity field instead, resulting in a solution of (1) given by $\varphi(x) = \phi_1(x) = \exp(v(x))$. The group exponential map $\exp(v(x))$ needed here can be computed by the efficient scaling and squaring algorithm (Arsigny *et al* 2006). This approach restricts φ to a subgroup of diffeomorphisms (Arsigny *et al* 2006, Ehrhardt *et al* 2011). Nevertheless, previous studies comparing non-stationary and stationary parameterisations of diffeomorphisms in the context of medical image registration revealed that this limitation does not significantly influence the registration accuracy (Ashburner 2007, Hernandez *et al* 2009, Vercauteren *et al* 2009).

Relying on this theoretical framework, we are interested in finding a diffeomorphic transformation $\varphi_j = \exp(v_j)$ between the reference image I_1 and an arbitrary target image I_j at breathing phase $j \in \{1, \ldots, n_{ph}\}$ that minimizes the energy functional

$$\mathcal{J}[v_j] = \mathcal{D}[I_1, I_j \circ \varphi_j] + \alpha\mathcal{S}[v_j] \ . \tag{2}$$

Here, $\mathcal{D}$ denotes a dissimilarity measure and $\mathcal{S}$ is a regularisation term that guarantees the smoothness of the velocity field. In our diffeomorphic registration scheme, a

diffusive regulariser and so-called normalized SSD forces are employed. Additional details are provided in (Schmidt-Richberg *et al* 2010).

2.2. Diffeomorphic correspondence modelling

Given diffeomorphic transformations $(\varphi_j)_{j\in\{1,\dots,n_{ph}\}}$, computed with the registration approach described in section 2.1, and corresponding surrogate measurements $\zeta_j \in \mathbb{R}^{n_{sur}}$ as training data, our goal is to learn a diffeomorphic correspondence model that describes their relationship using multivariate linear regression. The standard approach would be to represent the internal motion by the displacement fields $(u_j)_{j\in\{1,\dots,n_{ph}\}}$, with $\varphi_j = id + u_j : \Omega \to \Omega$, see, e.g., (Klinder *et al* 2010). However, transformations generated by performing classical linear statistics on displacement fields cannot be guaranteed to be diffeomorphic because the space of diffeomorphisms is not linear (Arsigny *et al* 2006, Ehrhardt *et al* 2011). Therefore, our framework uses a subgroup of diffeomorphisms parametrized by stationary velocity fields (cf. section 2.1). These velocity fields form a linear space (the tangent space at identity), which allows us to perform vectorial statistics on diffeomorphisms (Arsigny *et al* 2006). No additional calculations are required, as these velocity fields are computed as part of our registration scheme.

2.2.1. Diffeomorphic multivariate regression by ordinary least squares (OLS) From here on, the velocity fields v_j and surrogate measurements ζ_j are interpreted as random variables $\mathbf{V}_j \in \mathbb{R}^{3m}$ and $\mathbf{Z}_j \in \mathbb{R}^{n_{sur}}$, where m denotes the number of voxels of image I_j. Then, the basis form of our diffeomorphic correspondence model is defined by

$$\hat{\mathbf{V}} = \overline{\mathbf{V}} + \mathbf{B}(\hat{\mathbf{Z}} - \overline{\mathbf{Z}}) \ , \tag{3}$$

which generates a velocity field $\hat{\mathbf{V}}$ for a given breathing signal observation $\hat{\mathbf{Z}}$ with $\overline{\mathbf{V}} = 1/n_{ph}\sum_{j=1}^{n_{ph}} \mathbf{V}_j$ and $\overline{\mathbf{Z}} = 1/n_{ph}\sum_{j=1}^{n_{ph}} \mathbf{Z}_j$. After mean centering the random variables ($\mathbf{V}_j^c = \mathbf{V}_j - \overline{\mathbf{V}}$ and $\mathbf{Z}_j^c = \mathbf{Z}_j - \overline{\mathbf{Z}}$), matrices $\mathbf{V} := (\mathbf{V}_1^c, \dots, \mathbf{V}_{n_ph}^c)$ and $\mathbf{Z} := (\mathbf{Z}_1^c, \dots, \mathbf{Z}_{n_{ph}}^c)$ are formed to estimate the coefficient matrix $\mathbf{B} \in \mathbb{R}^{3m} \times \mathbb{R}^{n_{sur}}$ required in (3) in an ordinary least squares (OLS) sense by performing a multivariate linear regression (Hastie *et al* 2009):

$$\mathbf{B}^{OLS} = \arg\min_{\mathbf{B}'} \operatorname{tr}\Big[(\mathbf{V} - \mathbf{B}'\mathbf{Z})(\mathbf{V} - \mathbf{B}'\mathbf{Z})^T\Big] = \mathbf{V}\mathbf{Z}^+ \ . \tag{4}$$

Here, matrix $\mathbf{Z}^+$ represents the Moore-Penrose pseudoinverse of the $n_{sur} \times n_{ph}$ matrix $\mathbf{Z}$. Following Albert (1972), two possible ways of calculating $\mathbf{Z}^+$ are

$$\mathbf{Z}^+ = \mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T)^{-1} \text{ and} \tag{5}$$

$$\mathbf{Z}^+ = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T \ . \tag{6}$$

In (5) it is assumed that $\mathbf{Z}$ has full row rank, whereas in (6) the columns of $\mathbf{Z}$ have to be linearly independent. However, in our context, at least the assumption of (5) is frequently violated by high-dimensional surrogates (e.g., range images), which often contain perfectly correlated dimensions. Therefore, in our implementation, the non-invertibility of covariance matrix $\mathbf{\Sigma_{ZZ}} = \mathbf{Z}\mathbf{Z}^T$ and $\mathbf{Z}^T\mathbf{Z}$ is avoided by approximating them with $\mathbf{\Sigma_{ZZ}} + \gamma\mathbf{I}$ and $\mathbf{Z}^T\mathbf{Z} + \gamma\mathbf{I}$, ($\gamma > 0$), respectively. This is known as ridge regression or Tikhonov regularisation (Hastie *et al* 2009). For computational efficiency, we choose (5) if $n_{ph} \geq n_{sur}$ and (6) if $n_{sur} > n_{ph}$.

For a ridge regression, the choice of a suitable regularization parameter γ is rather heuristic, and several authors proposed using subspace approaches instead. These methods aim at revealing the low-dimensional hidden structure of the high-dimensional surrogate data via dimensionality reduction, cf. (Klinder *et al* 2010, Gao *et al* 2008, Liu *et al* 2010) as examples of use in the given context.

2.2.2. Principal component regression (PCR) Principal component regression (PCR) is standard linear regression with a dimensionality reduction based on principal components analysis (PCA) performed on the surrogate measurements first (Jolliffe 2002, Klinder *et al* 2010). Therefore, singular value decomposition $\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{W}^T$ is applied to determine a unitary matrix $\mathbf{U}$ of the left-singular vectors of $\mathbf{Z}$ and a diagonal matrix of corresponding singular values $\mathbf{D}$. Now, the idea of PCR is to use only the first n_c left-singular vectors $\mathbf{U}_{n_c}$ with positive singular values to approximate $\mathbf{\Sigma_{ZZ}}$ by $\mathbf{U}_{n_c}(\mathbf{D}_{n_c}^T\mathbf{D}_{n_c})\mathbf{U}_{n_c}^T$. This leads to a coefficient matrix

$$\mathbf{B}^{PCR} = \mathbf{\Sigma_{VZ}}\mathbf{U}_{n_c}(\mathbf{D}_{n_c}^T\mathbf{D}_{n_c})^{-1}\mathbf{U}_{n_c}^T \ , \tag{7}$$

where matrix $\mathbf{\Sigma_{VZ}} = \mathbf{V}\mathbf{Z}^T$ represents the cross-covariance between $\mathbf{V}$ and $\mathbf{Z}$ and only a diagonal matrix needs to be inverted.

2.2.3. Partial least squares (PLS) In PCR, $\mathbf{U}_{n_c}$ defines a subspace of maximum variation of the surrogate data, without taking the structure of the internal motion data into account. It is therefore not guaranteed that the dimensionality reduced surrogate data contains information that is useful for the estimation of the internal motion. Partial least squares (PLS) tries to circumvent this problem by searching for orthonormal bases $\mathbf{U}_{n_c} := (\mathbf{u}_1, \ldots, \mathbf{u}_{n_c})$ and $\mathbf{P}_{n_c} := (\mathbf{p}_1, \ldots, \mathbf{p}_{n_c})$, consisting of pairs $\mathbf{u}_i$ and $\mathbf{p}_i$ that maximize the cross-covariance $\rho_i = \mathbf{u}_i^T\mathbf{\Sigma_{ZV}}\mathbf{p}_i$. As detailed in (Borga *et al* 1997), the basis vectors sought are solutions of a generalized eigenvalue problem, which can be efficiently solved by the non-linear iterative partial least squares algorithm (NIPALS). Having determined a new basis $\mathbf{U}_{n_c}$ this way, a new coefficient matrix is defined as

$$\mathbf{B}^{PLS} = \mathbf{\Sigma_{VZ}}\mathbf{U}_{n_c}(\mathbf{U}_{n_c}^T\mathbf{\Sigma_{ZZ}}\mathbf{U}_{n_c})^{-1}\mathbf{U}_{n_c}^T \ . \tag{8}$$

2.2.4. Canonical correlation analysis (CCA) The canonical correlation analysis (CCA) has much in common with PLS, but instead of searching for new basis vectors $\mathbf{u}_i$ and $\mathbf{p}_i$ that maximize the cross-covariance between the projected data sets, CCA maximizes their correlation $\rho_i = \frac{\mathbf{u}_i^T\mathbf{\Sigma_{ZV}}\mathbf{p}_i}{\sqrt{\mathbf{u}_i^T\mathbf{\Sigma_{ZZ}}\mathbf{u}_i\mathbf{p}_i^T\mathbf{\Sigma_{VV}}\mathbf{p}_i}}$. Again, solving a generalized eigenvalue problem leads to the sought vectors $\mathbf{u}_i$ and $\mathbf{p}_i$ (Borga *et al* 1997). However, this time the inverse of covariance matrix $\mathbf{\Sigma_{ZZ}}$ (and $\mathbf{\Sigma_{VV}}$) is required. We therefore perform a PCA on $\mathbf{V}$ and $\mathbf{Z}$ first, to reduce their dimensionality as suggested by Gao *et al* (2008). Given the new basis $\mathbf{U}_{n_c}$, the coefficient matrix is calculated in the same fashion as $\mathbf{B}^{PLS}$:

$$\mathbf{B}^{CCA} = \mathbf{\Sigma_{VZ}}\mathbf{U}_{n_c}(\mathbf{U}_{n_c}^T\mathbf{\Sigma_{ZZ}}\mathbf{U}_{n_c})^{-1}\mathbf{U}_{n_c}^T \ . \tag{9}$$

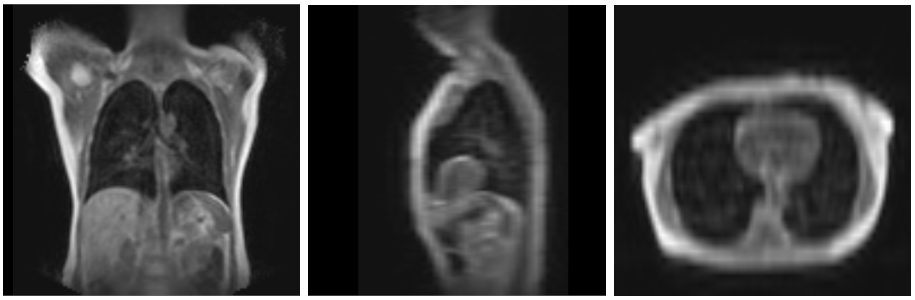


Figure 1. Sample 3D MRI image of a 4D MRI data set for illustration of the highly anisotropic spatial resolution of $3.91 \times 10 \times 3.91$ mm (RL, AP, HF): (a) coronal slice, (b) sagittal slice, (c) axial slice.

3. Study Design

3.1. Image Data

For our evaluation, 34 publicly available and proprietary 4D image data sets from different modalities (CT & MRI) are used:

- 12 4D CT data sets of lung cancer patients taken from our in-house database with an average spatial resolution of $1 \times 1 \times 1.5$ mm (right-left (RL), anterior-posterior (AP), head-foot (HF)). They were acquired during free breathing and each 4D data set consists of 3D images reconstructed at 10 to 14 breathing phases as described in (Ehrhardt *et al* 2007).
- 10 4D CT data sets with 10 breathing phases and a spatial resolution of $0.97 - 1.16 \times 0.97 - 1.16 \times 2.5$ mm (RL, AP, HF) publicly available from the DIR-Lab, The University of Texas M. D. Anderson Cancer Center, USA (Castillo *et al* 2009).
- 6 4D CT data sets with 10 breathing phases and a spatial resolution of $0.879 - 1.172 \times 0.879 - 1.172 \times 2$ mm (RL, AP, HF) publicly available from the Léon Bérard Cancer Center & CREATIS lab, Lyon, France (Vandemeulebroucke *et al* 2011).
- 3 4D CT and 4D MRI data sets of lung cancer patients acquired at the German Cancer Research Center (DKFZ), Heidelberg, Germany. For each of the 3 patients, a 4D CT and a 4D MRI data set were acquired. Each 4D CT data set consists of 7 3D images with a spatial resolution of $0.977 \times 0.977 \times 1.5 - 3$ mm (RL, AP, HF). The 4D MRI data sets consist of 157 3D images acquired during free breathing with a temporal resolution of 0.5 s and a spatial resolution of $3.91 \times 10 \times 3.91$ mm (RL, AP, HF). Figure 1 shows a sample MRI image.

3.2. Simulated types of surrogates

Using the 4D images described above, we simulated the subsequently described breathing signals.

3.2.1. Spirometry Spirometry results in a common 1D (pure signal) and 2D (signal + time derivative) breathing signal, respectively. For the 4D CT data sets, tidal volume

measurements are simulated by image-based analysis of the air content inside the lungs and interpreted as spirometry measurements. This approach follows (Lu *et al* 2005), in which a linear relationship between air content values inside the lungs, derived by a 4D CT-based voxel-wise analysis of the Hounsfield units, and tidal volume has been theoretically derived and experimentally verified. Time derivatives are approximated by finite differences. Relying on the Hounsfield units, air content cannot be determined for the 4D MRI data sets. For them, we approximated the entire lung volumes by lung segmentations, which are generated by an automatic approach (Wilms *et al* 2012).

3.2.2. Abdominal belt Belts measure the expansion of the chest or abdomen during respiration and also represent common 1D (only signal) and 2D (signal + time derivative) breathing signals. We simulate a belt signal by measuring the body volume over a small number of slices (≈ 1 cm of the skin surface) in the abdominal region. The corresponding body volume is calculated by counting body voxels in the selected slices, using a body segmentation generated by volume growing and subsequent hole-filling. An optimal belt location is determined by selecting a position with a clearly visible movement of the body during respiration. We experimentally verified the simulation approach by using 3 patients of our in-house data base for which cine CT volumes of the abdomen and corresponding belt measurements were available. Strong correlations between real and simulated signals (correlation coefficient $c \geq 0.98$) were observed.

3.2.3. Diaphragm translation The motion of the diaphragm is a direct indicator of respiratory motion and can be tracked in real-time in several ways (e.g., fluoroscopy, MRI navigators) (Cerviño *et al* 2010). For this study, 1D (only signal) and 2D (signal + time derivative) breathing signals are generated consisting of the average translation of the dome of the left and right hemi-diaphragm in cranio-caudal direction. The positions of the domes over time in a selected coronal slice of the 4D data sets are automatically determined with sub-voxel accuracy by edge detection.

3.2.4. Range imaging (RI) As a typical multi-dimensional surrogate signal, we simulate a range image-based tracking of the raising/lifting of the chest wall and the abdominal skin. Therefore, we position a virtual range imaging sensor above the patient's body. Rays that start at n_{sur} points of the assumed sensor plane are followed until they intersect with the body in anterior-posterior direction. The intersection is determined with sub-voxel accuracy using a grey value threshold and linear interpolation. Due to the low spatial resolution, the 4D MRI data sets are first resampled to the spatial resolution of the corresponding 4D CT data sets to allow an improved detection of the skin surface. Six different spatial sampling patterns of a rectangular region of interest are simulated: $n_{sur} = 1$ point located over the xiphoid process, $n_{sur} = 1$ point centrally located over the position of the belt, and $n_{sur} = 10, 100, 1000, 10000$ equally spread points.

3.3. Experiments

Considering the dimensionality of the different breathing signals, we evaluated the following regression approach/surrogate combinations: OLS (with Tikhonov regularisation) with all breathing signals. PCR, PLS, and CCA, each combined with the simulated multi-dimensional range images with $n_{sur} = 10, 100, 1000, 10000$ points. The evaluation procedures are described subsequently.

3.3.1. Evaluation based on 4D CT data sets For the 4D CT data sets a leave-out evaluation strategy is applied. Selecting the phase at end-inspiration (EI) as a reference for intra-patient registration and correspondence-modelling, three different models are built for each data set and regression/surrogate combination. To evaluate the extrapolation power of the approaches (signal measurement out of the range of the training data), one model is trained on all phases, but the phases at EE-1, EE (end expiration), and EE+1, and is used to estimate the motion $\hat{\varphi}_{EE}$ between EI and EE. The two interpolation models are evaluated by leaving out the phases at mid-expiration (ME) and mid-inspiration (MI) during training and estimating the corresponding transformations $\hat{\varphi}_{ME}$ and $\hat{\varphi}_{MI}$. The accuracy of the estimated transformations is evaluated by computing the mean target registration error (TRE) based on manually defined corresponding inner-lung landmarks. For our own data sets, on average 70 equally distributed landmarks per patient and 4 selected breathing phases (EE, EI, MI and ME) are available. The DIR-Lab and CREATIS data sets are provided with 300 (DIR-Lab) and 100 (CREATIS) landmarks for the images at EI and EE. Therefore, the extrapolation models are built for all 28 4D CT data sets available, whereas the interpolation capabilities are only evaluated on our own data sets.

3.3.2. Evaluation based on 4D MRI data sets As each of the three 4D MRI data sets consists of 157 3D MRI images and 20-30 complete breathing cycles, separate subsets of the data are used for training and evaluation. The training data for each patient consists of two selected average respiratory cycles, allowing for inter- and extrapolation cases during the evaluation using all remaining data. A landmark-based evaluation is not possible for the 4D MRI data due to the low spatial resolution of the images. Instead, we use the distance between available lung segmentations to assess the accuracy of the motion estimation between the reference phase and all remaining phases in the evaluation subset. For this purpose, the lung segmentations of the phases in the evaluation subset are warped to the reference phase by applying the estimated transformations. Subsequently, the mean symmetric distances between points on the surface of the reference segmentation and points on the surface of the warped segmentations are computed. Then, a patient-specific mean value is calculated by averaging the mean surface distances of the phases in the evaluation subset.

3.3.3. Evaluation based on 4D CT & 4D MRI data sets The 4D CT and 4D MRI data pairs are used to simulate and evaluate a possible clinical workflow. This workflow comprises learning a correspondence model on a 4D CT data set acquired for treatment planning, which is then used for surrogate-based motion estimation during treatment. Here, the 4D MRI data sets only serve as ground truth data for evaluation, meaning that the different correspondence models trained on the 4D CT data are used to estimate the motion in the corresponding 4D MRI data set. As the CT and MRI data sets were acquired during different sessions with different scanners, an alignment is needed to account for differences in the position of the lungs in CT and MRI space. For this alignment, a reference phase in the 4D MRI data representing a breathing state similar to that of the reference phase in the 4D CT data (EI) is chosen. Similarity is determined using the lung volume available from the lung segmentations. Next, an affine transformation φ_{aff} between the EI CT and MRI lung segmentations is estimated by a surface-based ICP registration. Using this affine transformation, displacement fields estimated by a correspondence model can be warped from the

model space (CT) to the MRI space. For evaluation, the same methods and measures are applied as described for the 4D MRI-based experiments.

3.3.4. Comparison of standard and diffeomorphic correspondence modelling To illustrate the advantages of our diffeomorphic modelling approach, additional standard correspondence models (i.e., the regression is directly performed between displacement fields and surrogate data; cf. section 2.2) are built for our twelve proprietary 4D CT data sets, the three 4D CT and 4D MRI data pairs and all surrogates. The models are compared with respect to the motion estimation accuracy (measures as before) and the number of singularities in the motion fields.

3.3.5. Parameters and implementation details Depending on the regression approach, different parameters have to be determined during correspondence modelling: For Tikhonov regularisation of (4) the parameter γ is required to be chosen, and all subspace approaches depend on the number of components n_c used for dimensionality reduction. These parameters are determined on a patient-/regression approach-/experiment-specific level by a leave-one-out cross validation on the training data, considering the Euclidean distance between simulated and left out velocity fields as quality measure. For all experiments, lung segmentations are used to restrict the registration and regression-based modelling approaches to the lung regions to save memory/computation time and to prevent potential sliding motion between the visceral and parietal pleura from affecting the lung motion estimation.

4. Results

The results of the experiments described in section 3.3 are listed in tables 1 and 2 and are detailed in the following subsections. The *overall mean values* ($\pm$ standard deviation) reported in table 1 were calculated by averaging the patient-specific mean target registration errors (4D CT data sets) and the patient-specific mean surface distances (4D MRI data sets) for each experiment. Paired t-tests with a signficance level of 5% ($p < 0.05$) were performed by pairing the patient-specific mean values in order to assess the statistical significance of differences in the overall mean values between different surrogates, regression approaches, and scenarios. Table 2 additionally lists the corresponding *patient-specific mean values* of the 4D MRI and 4D CT & 4D MRI experiments and the three patients used in these experiments, respectively. To determine statistical significane of the differences on a patient-specific level, t-tests were this time performed by pairing the mean surface distances of the $\approx$130 phases of each patient's evaluation subset.

4.1. Motion estimation accuracy using 1D surrogate signals

Comparing the motion estimation accuracy achieved in all three experiments (4D CT, 4D MRI, 4D CT & 4D MRI; scenarios analysed independently) with standard OLS regression as reported in table 1, almost no statistically significant differences ($p > 0.05$ for the landmark-based TRE values/mean surface distances) are observed between the use of the simulated 1D breathing signals. Two exceptions exist: Considering only the 4D CT scenarios, tracking the skin surface point located over the xiphoid process (RI 1 point xiphoid) leads to a significantly decreased motion estimation accuracy in comparison to all other surrogates ($p < 0.05$). Furthermore, the difference

Table 1. Landmark-based mean target registration errors (4D CT) and mean lung surface distances (4D MRI and 4D CT & 4D MRI), obtained for the surrogate-based estimation of lung motion for the different image data sets, regression approaches, and surrogates. The mean observer error for the landmarks provided with the proprietary and DIR-Lab 4D CT data sets is 0.92 ± 0.34 mm. This (intra-)observer error refers to landmark position differences after repeated manual (i.e., performed by a human observer) landmark transferring from EE to EI. The final goal would be to end up with TRE values in the order of or below the observer errors. All listed values are mean values ($\pm$ standard deviation) over the patients' mean values obtained in the individual experiments. 4D-CT: Interpolation performed for 12 patients and 2 phases (mean values are calculated by averaging the MI & ME results of all 12 patients), Extrapolation performed for 28 patients (only EE phases are used). 4D MRI and 4D CT & 4D MRI: Results are averaged over 3 patients and ≈ 130 test images per patient (cf. section 3.3.2 and table 2).

	TRE [mm]	**Mean Surface Distance [mm]**	
Motion Estimation	**4D CT Inter-/Extrapolation**	**4D MRI**	**4D CT & 4D MRI**
No motion estimation	$3.59 \pm 1.46/7.59 \pm 2.72$	2.81 ± 1.54	
Intra-Patient-Registration	$1.52 \pm 0.15/1.40 \pm 0.49$	1.12 ± 0.42	
Surrogate-based motion estimation using $\mathbf{B}^{OLS}$ *and 1D surrogate = ...*			
Spirometry	$1.87 \pm 0.31/2.05 \pm 0.85$	1.31 ± 0.51	1.94 ± 1.16
Belt	$1.94 \pm 0.32/2.03 \pm 0.82$	1.40 ± 0.54	2.00 ± 1.19
Diaphragm	$1.90 \pm 0.32/2.01 \pm 0.88$	1.27 ± 0.50	2.13 ± 1.25
RI 1 point belt pos.	$1.92 \pm 0.27/2.08 \pm 0.91$	1.46 ± 0.62	2.31 ± 1.31
RI 1 point xiphoid pos.	$2.18 \pm 0.66/3.76 \pm 2.02$	1.58 ± 0.54	2.17 ± 0.75
Surrogate-based motion estimation using $\mathbf{B}^{OLS}$ *and 2D (1D+ time derivative) surrogate = ...*			
Spirometry	$1.73 \pm 0.19/1.96 \pm 0.81$	1.28 ± 0.50	1.88 ± 1.08
Belt	$1.72 \pm 0.19/1.98 \pm 0.73$	1.37 ± 0.52	1.94 ± 1.08
Diaphragm	$1.74 \pm 0.19/2.04 \pm 0.86$	1.24 ± 0.48	2.23 ± 1.46
RI 1 point belt pos.	$1.72 \pm 0.17/2.04 \pm 0.84$	1.38 ± 0.58	2.31 ± 1.27
RI 1 point xiphoid pos.	$2.25 \pm 0.71/3.12 \pm 1.91$	1.52 ± 0.59	2.56 ± 1.20
Surrogate-based motion estimation using $\mathbf{B}^{OLS}$ *and RI surrogate with ...*			
10 points	$1.74 \pm 0.21/2.12 \pm 0.86$	1.44 ± 0.59	1.99 ± 0.89
100 points	$1.71 \pm 0.22/1.86 \pm 0.63$	1.41 ± 0.57	2.01 ± 0.89
1000 points	$1.71 \pm 0.20/1.88 \pm 0.59$	1.40 ± 0.56	1.94 ± 0.80
10000 points	$1.80 \pm 0.35/2.02 \pm 0.68$	1.40 ± 0.56	1.95 ± 0.81
Surrogate-based motion estimation using $\mathbf{B}^{PCR}$ *and RI surrogate with ...*			
10 points	$1.74 \pm 0.20/2.36 \pm 1.49$	1.48 ± 0.62	1.99 ± 0.88
100 points	$1.71 \pm 0.23/1.87 \pm 0.61$	1.41 ± 0.58	1.97 ± 0.82
1000 points	$1.68 \pm 0.19/1.87 \pm 0.61$	1.40 ± 0.57	1.96 ± 0.81
10000 points	$1.69 \pm 0.20/1.82 \pm 0.62$	1.40 ± 0.56	1.96 ± 0.82
Surrogate-based motion estimation using $\mathbf{B}^{PLS}$ *and RI surrogate with ...*			
10 points	$1.73 \pm 0.20/2.25 \pm 1.16$	1.48 ± 0.61	2.00 ± 0.89
100 points	$1.70 \pm 0.24/1.86 \pm 0.61$	1.41 ± 0.57	1.96 ± 0.82
1000 points	$1.68 \pm 0.19/1.79 \pm 0.59$	1.40 ± 0.56	1.96 ± 0.81
10000 points	$1.69 \pm 0.20/1.79 \pm 0.59$	1.39 ± 0.56	1.96 ± 0.82
Surrogate-based motion estimation using $\mathbf{B}^{CCA}$ *and RI surrogate with ...*			
10 points	$1.75 \pm 0.20/2.22 \pm 0.86$	1.45 ± 0.66	2.21 ± 1.15
100 points	$1.77 \pm 0.23/2.04 \pm 0.96$	1.34 ± 0.52	2.10 ± 1.01
1000 points	$1.72 \pm 0.19/2.02 \pm 0.91$	1.32 ± 0.51	1.98 ± 0.84
10000 points	$1.72 \pm 0.19/1.88 \pm 0.67$	1.48 ± 0.42	1.97 ± 0.83

Table 2. Patient-specific mean lung surface distances obtained for the surrogate-based estimation of lung motion for the 4D MRI and 4D CT & 4D MRI experiments. Given are the mean values ($\pm$ standard deviation) of the patient-specific mean surface distances averaged over ≈ 130 phases per patient (cf. section 3.3.2 and table 1).

	Mean Surface Distance [mm]	
Motion Est.	**4D MRI** Patient 1 / Patient 2 / Patient 3	**4D CT & 4D MRI** Patient 1 / Patient 2 / Patient 3
No motion est.	1.47 ± 0.47 / 2.46 ± 0.60 / 4.49 ± 1.19	
Intra-Patient-Reg.	0.70 ± 0.12 / 1.13 ± 0.20 / 1.54 ± 0.16	
Surrogate-based motion estimation using $\mathbf{B}^{OLS}$ *and 1D surrogate = ...*		
Spirometry	0.77 ± 0.13 / 1.38 ± 0.26 / 1.77 ± 0.17	1.00 ± 0.26 / 1.57 ± 0.36 / 3.24 ± 0.46
Belt	0.78 ± 0.12 / 1.64 ± 0.33 / 1.77 ± 0.15	1.00 ± 0.25 / 1.68 ± 0.31 / 3.32 ± 0.62
Diaphragm	0.77 ± 0.13 / 1.26 ± 0.26 / 1.76 ± 0.16	1.01 ± 0.28 / 1.89 ± 0.20 / 3.48 ± 0.70
RI 1 pt. belt pos.	0.77 ± 0.11 / 1.65 ± 0.39 / 1.97 ± 0.38	1.12 ± 0.23 / 2.11 ± 0.47 / 3.71 ± 0.87
RI 1 pt. xiph. pos.	1.07 ± 0.32 / 1.55 ± 0.33 / 2.14 ± 0.52	1.72 ± 0.75 / 1.74 ± 0.34 / 3.03 ± 0.34
Surrogate-based motion estimation using $\mathbf{B}^{OLS}$ *and 2D (1D+ time derivative) surrogate = ...*		
Spirometry	0.75 ± 0.13 / 1.33 ± 0.26 / 1.74 ± 0.15	1.01 ± 0.27 / 1.55 ± 0.32 / 3.09 ± 0.43
Belt	0.77 ± 0.13 / 1.59 ± 0.32 / 1.74 ± 0.12	1.01 ± 0.25 / 1.67 ± 0.30 / 3.13 ± 0.52
Diaphragm	0.76 ± 0.13 / 1.26 ± 0.26 / 1.71 ± 0.13	1.27 ± 0.23 / 1.51 ± 0.28 / 3.91 ± 0.90
RI 1 pt. belt pos.	0.76 ± 0.12 / 1.48 ± 0.30 / 1.90 ± 0.34	1.15 ± 0.20 / 2.10 ± 0.46 / 3.66 ± 0.82
RI 1 pt. xiph. pos.	0.98 ± 0.29 / 1.44 ± 0.28 / 2.14 ± 0.55	1.95 ± 0.84 / 1.78 ± 0.32 / 3.95 ± 0.84
Surrogate-based motion estimation using $\mathbf{B}^{OLS}$ *and RI surrogate with ...*		
10 points	0.76 ± 0.12 / 1.75 ± 0.39 / 1.82 ± 0.23	1.20 ± 0.27 / 1.81 ± 0.34 / 2.95 ± 0.45
100 points	0.75 ± 0.13 / 1.74 ± 0.40 / 1.74 ± 0.15	1.20 ± 0.29 / 1.86 ± 0.48 / 2.96 ± 0.42
1000 points	0.75 ± 0.13 / 1.71 ± 0.39 / 1.74 ± 0.15	1.22 ± 0.30 / 1.79 ± 0.41 / 2.80 ± 0.38
10000 points	0.75 ± 0.13 / 1.70 ± 0.39 / 1.74 ± 0.15	1.22 ± 0.30 / 1.80 ± 0.43 / 2.82 ± 0.38
Surrogate-based motion estimation using $\mathbf{B}^{PCR}$ *and RI surrogate with ...*		
10 points	0.77 ± 0.12 / 1.85 ± 0.42 / 1.82 ± 0.23	1.20 ± 0.28 / 1.83 ± 0.40 / 2.93 ± 0.41
100 points	0.75 ± 0.13 / 1.75 ± 0.41 / 1.74 ± 0.13	1.21 ± 0.30 / 1.86 ± 0.49 / 2.83 ± 0.39
1000 points	0.75 ± 0.13 / 1.71 ± 0.39 / 1.74 ± 0.15	1.22 ± 0.30 / 1.81 ± 0.44 / 2.83 ± 0.38
10000 points	0.75 ± 0.13 / 1.71 ± 0.39 / 1.74 ± 0.15	1.23 ± 0.30 / 1.80 ± 0.43 / 2.85 ± 0.38
Surrogate-based motion estimation using $\mathbf{B}^{PLS}$ *and RI surrogate with ...*		
10 points	0.77 ± 0.12 / 1.85 ± 0.42 / 1.81 ± 0.23	1.20 ± 0.28 / 1.83 ± 0.40 / 2.97 ± 0.42
100 points	0.75 ± 0.13 / 1.74 ± 0.40 / 1.74 ± 0.14	1.21 ± 0.30 / 1.85 ± 0.44 / 2.83 ± 0.39
1000 points	0.75 ± 0.13 / 1.71 ± 0.39 / 1.74 ± 0.14	1.22 ± 0.30 / 1.82 ± 0.45 / 2.83 ± 0.38
10000 points	0.75 ± 0.13 / 1.70 ± 0.39 / 1.74 ± 0.16	1.23 ± 0.30 / 1.80 ± 0.36 / 2.85 ± 0.38
Surrogate-based motion estimation using $\mathbf{B}^{CCA}$ *and RI surrogate with ...*		
10 points	0.76 ± 0.12 / 1.51 ± 0.30 / 2.09 ± 0.43	1.22 ± 0.27 / 1.94 ± 0.42 / 3.48 ± 0.68
100 points	0.77 ± 0.14 / 1.48 ± 0.34 / 1.78 ± 0.17	1.21 ± 0.29 / 1.89 ± 0.49 / 3.20 ± 0.46
1000 points	0.75 ± 0.13 / 1.48 ± 0.33 / 1.73 ± 0.14	1.22 ± 0.30 / 1.84 ± 0.45 / 2.88 ± 0.42
10000 points	0.99 ± 0.23 / 1.71 ± 0.39 / 1.73 ± 0.15	1.23 ± 0.30 / 1.81 ± 0.42 / 2.86 ± 0.42

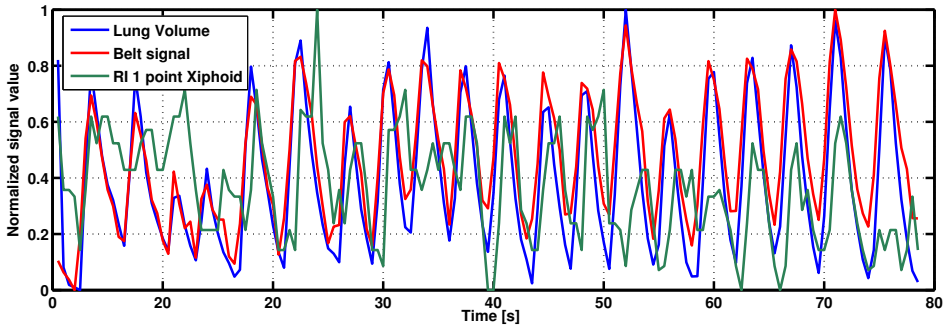


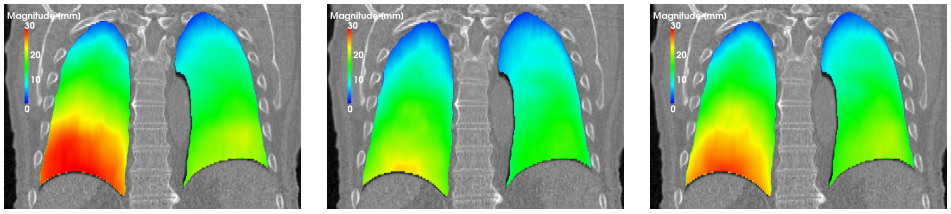
Figure 2. Three different simulated breathing signals (Lung volume, Belt signal, RI 1 point at the xiphoid position) for one patient with a length of 78.5 s. Simulations are based on a 4D MRI data set consisting of 157 3D images. All signals are normalized to the range $[0, 1]$. The lung volume and belt signal are highly correlated (correlation coefficient $c = 0.91$), while the correlation between lung volume/belt signal and the translation of the skin surface at the xiphoid position is weak ($c = 0.25$ / $c = 0.11$).

between tracking the point located over the belt and evaluating diaphragm motion was statistically significant for the combined 4D CT & 4D MRI experiment ($p = 0.046$).

Thus, in general the results suggest that only slight differences in motion estimation accuracy exist between the signals, as also illustrated in figure 2. However, there exist differences between the patients included in our study. As an example, for case 02 of the CREATIS 4D CT data and using tidal volume as surrogate, an underestimation of the motion amplitude can be seen when compared to the displacement field estimated via intra-patient registration (cf. figure 3). In this case, using the diaphragm translation as surrogate visibly reduces the underestimation. Further patient-specific differences are also revealed by the detailed results of the 4D MRI and 4D CT & 4D MRI experiments listed in table 2. For patient 01, all 1D surrogate signals lead to nearly equivalent results (except for the RI 1 point xiphoid) for the 4D MRI experiment, while in the case of patient 02 the accuracy obtained by the diaphragm tracking is significantly better than the results of all other 1D signals ($p < 0.05$). Interestingly, the corresponding results are different for the combined 4D CT & 4D MRI experiments. Here, the spirometry surrogate allows for the most accurate motion estimation for patient 02, and even the accuracy of the RI 1 xipoid surrogate is significantly higher than the diaphragm result. For patient 03, the RI 1 xiphoid surrogate, leading to the worst accuracy in the 4D MRI experiment, is the best 1D surrogate in the 4D CT & 4D MRI experiment, with the differences between all 1D surrogates being significant ($p < 0.05$). These patient-specific differences between 4D MRI and combined 4D CT & 4D MRI experiments can be interpreted as suggesting that the models built using our framework are only able to compensate for some of the occuring inter-session motion variability. However, please note that this specific part of our analysis is based on only a small number of patients (=3) and related conclusions have to be verified for a larger data set (cf. section 5).

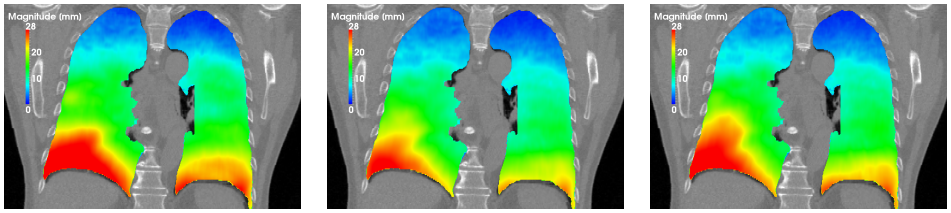
4.2. Motion estimation accuracy using 2D surrogate signals

As expected, in most instances the use of a 2D surrogate signal (1D signal + time derivative) and standard OLS regression leads to an improved estimation accuracy



(a) Intra-patient registration TRE: 1.36 mm (b) OLS Spirometry TRE: 4.17 mm (c) OLS Diaphragm translation TRE: 2.02 mm

Figure 3. Visualization of the lung motion between EI and EE for case 02 of the CREATIS 4D CT data sets estimated by intra-patient registration (a) and surrogate-based motion estimation relying on spirometer-determined tidal volume (b) and diaphragm translation (c).



(a) Intra-patient registration TRE: 1.40 mm (b) OLS RI 1 point belt pos. TRE: 2.53 mm (c) OLS RI 1000 points TRE: 2.00 mm

Figure 4. Visualization of the lung motion between EI and EE for patient 05 of our proprietary 4D CT data sets estimated by intra-patient registration (a) and surrogate-based motion estimation driven by a 2D signal consisting of the translation of 1 skin surface point located at the belt position and its time derivative (b, OLS regression) and 1000 equally spread points (c, OLS regression).

compared to the 1D results. This can be seen especially in case of the 4D CT interpolation scenarios (cf. table 1), for which, in addition, the reported differences are statistically significant ($p < 0.05$, except for the RI 1 point xiphoid signal with $p = 0.56$), indicating that models relying on multi-dimensional surrogate signals are able to account for intra-cycle motion variations like hysteresis. Furthermore and similar to the 1D signal findings, the related differences between the individual 2D signals are not statistically significant ($p > 0.05$) again with the exception of the results obtained for the RI 1 point xiphoid surrogate during the 4D CT experiments ($p < 0.01$).

4.3. Motion estimation accuracy using high-dimensional RI surrogate signals

The results of the standard OLS approach reported in table 1 show the potential of using a high-dimensional surrogate signal, which contains the motion of 10 to 10000 equally spread skin surface points instead of only one point (w/o the time derivative), here located at the xiphoid or belt position. However, when comparing the results of the RI 1 point belt signal in table 1 to the OLS results obtained using 10 to 10000 points, only the differences reported for the 4D CT interpolation scenarios are statistically significant ($p < 0.05$). But even these significant differences diminish or vanish when using the 1D signal with its time derivative (=2D signal) in most cases. Moreover, the differences between the best combinations of the OLS approach with a

high-dimensional RI surrogate signal for each experiment (e.g., 4D CT extrapol.: 100 points) and the results of the 2D signals (except for the RI 1 point xiphoid signal) in table 1 are not statistically significant ($p > 0.05$). But there are again patients for which larger differences occur (cf. figure 4 and table 2). For example, the results achieved for patient 02 in the 4D MRI experiments by using the 2D surrogate signals are all significantly better than any result obtained by combining the OLS approach with a high-dimensional RI surrogate.

Our results also show that for the OLS method and both 4D CT scenarios at some point increasing the number of surface points significantly decreases the estimation accuracy, presumably because of overfitting (e.g., 4D CT interpol.: 1000 points vs. 10000 points, 1.88 ± 0.59 vs. 2.02 ± 0.68, $p = 0.02$). In this case, the subspace approaches show their potential. However, most of the differences between the OLS results shown in table 1 and the results of the subspace-based models are not statistically significant ($p > 0.05$). In addition, the differences between the subspace approaches themselves are generally not significant with some exceptions involving the CCA method and again patient-specific differences. For example, the CCA method combined with 100 or 1000 points turns out to be the best subspace approach for patient 2 when looking at the results for the 4D MRI experiments.

To sum up, our results suggest that (1) using subspace approaches might not be necessary in general and (2) using more sophisticated methods like PLS and CCA instead of PCR might also not be required. The latter suggestion is also supported by the nearly similar number of components (< 5) used by each approach on average.

4.4. Standard vs. diffeomorphic correspondence modelling

The results of the comparison of standard and diffeomorphic correspondence modelling are summarized in table 3. The differences in accuracy between both approaches are generally very small and only the differences obtained for the 4D CT interpolation scenarios with $\mathbf{B}^{OLS}$ and spirometry as surrogate are statistically significant (Interpolation: $p < 0.05$; Extrapolation: $p = 0.34$). In contrast to the standard models, the diffeomorphic correspondence models successfully avoid singularities in the estimated displacement fields even in the case of extrapolation for the 4D CT experiments. For some of the 4D MRI and 4D CT & 4D MRI experiments, a very small number of singularities can also be found in displacement fields estimated by the diffeomorphic modelling approach. These singularities are caused by very large and unusual extrapolation factors (>4, usually factors < 3), which are a consequence of a short period of coughing (4D MRI experiments) and, in case of the combined 4D CT & 4D MRI experiments, also due to the limited number of training samples (one 4D CT data set) and the difficult compensation of baseline differences between the surrogate signals simulated for both modalities.

Formally speaking, the large extrapolation factors eventually lead to velocity fields, which are – at least for the applied configuration of the scaling and squaring algorithm (fixed number of 6 scaling and squaring operations) – non-sufficiently smooth and violate the premises of the underlying theory. For a more detailed discussion on this specific issue please refer to (Werner *et al* 2012b). Despite the occasional presence of singularities in the displacement fields resulting from the diffeomorphic modelling approach, it has to be noted that their number is much smaller than for the corresponding standard models. This, still, illustrates the advantage of the proposed modelling approach.

Table 3. TRE values/mean surface distances and number of singularities (voxels x with $\det \nabla \hat{\varphi}(x) < 0$) for the displacement fields resulting from either diffeomorphic or standard surrogate-based estimation of lung motion. Values are given for the 4D CT interpolation and extrapolation scenarios (evaluated using our twelve proprietary 4D CT data sets), the 4D MRI experiments, the combined 4D CT & 4D MRI scenarios (cf. section 3.3.2) and different surrogate signals (cf. table 1); Listed are the averaged values of the patient mean values and the corresponding standard deviations. Presence of singularities is only evaluated for inner-lung voxels (cf. section 3.3.5; 4D CT: on avg. $\approx$ 300000 inner-lung voxels; 4D MRI: on avg. $\approx$ 30000 inner-lung voxels).

	4D CT			
	Interpolation		**Extrapolation**	
Motion estimation approach	TRE [mm]	Singularities	TRE [mm]	Singularities
Diffeomorphic framework:				
$\mathbf{B}^{OLS}$ & Spirometry	1.87 ± 0.31	0	1.84 ± 0.45	0
$\mathbf{B}^{OLS}$ & Belt	1.94 ± 0.32	0	1.79 ± 0.34	0
$\mathbf{B}^{OLS}$ & RI 1000 points	1.71 ± 0.20	0	1.67 ± 0.29	0
Standard framework:				
$\mathbf{B}^{OLS}$ & Spirometry	1.86 ± 0.30	0	1.83 ± 0.44	0.33 ± 1.15
$\mathbf{B}^{OLS}$ & Belt	1.93 ± 0.32	0	1.80 ± 0.34	109.58 ± 357.53
$\mathbf{B}^{OLS}$ & RI 1000 points.	1.70 ± 0.20	0	1.69 ± 0.28	29.75 ± 71.90
	4D MRI		**4D CT & 4D MRI**	
Motion estimation approach	Mean Surf. Dist. [mm]	Singularities	Mean Surf. Dist. [mm]	Singularities
Diffeomorphic framework:				
$\mathbf{B}^{OLS}$ & Spirometry	1.31 ± 0.51	0	1.93 ± 1.16	0.13 ± 0.22
$\mathbf{B}^{OLS}$ & RI 1 point xiphoid pos.	1.58 ± 0.54	0.03 ± 0.05	2.17 ± 0.75	0.30 ± 0.36
$\mathbf{B}^{CCA}$ & RI 10000 points	1.48 ± 0.42	0	1.97 ± 0.83	1.81 ± 3.14
Standard framework:				
$\mathbf{B}^{OLS}$ & Spirometry	1.29 ± 0.49	0.61 ± 1.06	1.90 ± 1.13	257.93 ± 442.30
$\mathbf{B}^{OLS}$ & RI 1 point xiphoid pos.	1.57 ± 0.52	15.46 ± 26.77	2.15 ± 0.72	245.27 ± 252.48
$\mathbf{B}^{CCA}$ & RI 10000 points	1.46 ± 0.40	3.97 ± 6.65	1.95 ± 0.81	472.04 ± 644.89

4.5. Computation times

All experiments presented here were performed on a standard PC equipped with an Intel W3520 2.67GHz CPU and 24GB RAM. Given a set of velocity fields and a sampled breathing signal, the computation time required for model training depends on several factors: regression approach, dimensionality of the surrogate signal, size of the velocity fields, and the number of phases used, which determines the number of cross-validations performed (cf. section 3.3.5). Average computation times for the current implementation of the algorithms to build, e.g., the 4D CT-based models range from $\approx$ 1 minute (OLS, 1D spirometry, 7 training phases) to $\approx$ 60 minutes (PLS, RI 10000 points, 13 training phases). The time for generation of a new velocity field for a given signal measurement is $<$ 1 second. The subsequent application of the scaling and squaring algorithm to compute the corresponding displacement field takes approx. 10

seconds. Current implementations are not runtime-optimized. However, at least the time needed for model building should be of minor importance for most applications in radiotherapy as this would usually be done offline during treatment planning.

5. Discussion and Conclusion

Respiratory motion compensation methods in radiotherapy of thoracic and abdominal tumours are usually guided by breathing signals of so-called surrogates of the internal motion. In this work, we presented a diffeomorphic framework for surrogate-based motion estimation, which relies on the Log-Euclidean framework and multivariate regression. Besides standard ordinary least squares regression, we also investigated the use of different subspace regression variants to train diffeomorphic correspondence models. The regression approaches were compared for different types of common (simulated) breathing signals within the proposed framework using evaluation measures known from registration evaluation and ground truth patient 4D image data sets covering intra- and inter-cycle as well as intra- and inter-session motion variations.

The results of our study focusing on lung motion estimation showed only small differences in estimation accuracy between the different 1D surrogates. The only exception were breathing signals generated by tracking a skin surface point located over the xiphoid process, which turned out to be not suitable for reliable surrogate-based motion estimation. Furthermore, we found out that the estimation accuracy can be significantly increased by increasing the surrogate dimensionality. This was shown for 2D signals consisting of a common 1D signal and its time derivative, and high-dimensional signals containing the motion of many skin surface points. Most differences between results obtained by the standard OLS and subspace regression approaches for the high-dimensional signals were not significant. Therefore, using subspace approaches instead of the standard OLS regression might not be necessary in general but can help to avoid or reduce overfitting in case of very high-dimensional signals (e.g., 10000 skin surface points). It was striking that most of the significant differences between surrogates and/or regression approaches were only observable for experiments involving the 4D CT data sets. This is most likely a result of the limited number of 4D MRI data sets available (only 3) and their low spatial resolution, which prevented a more detailed evaluation of inner-lung motion estimation accuracy.

We additionally compared diffeomorphic correspondence models trained using the proposed framework with standard (non-diffeomorphic) correspondence models, which rely on displacement fields instead of velocity fields. It was shown that our diffeomorphic approach avoids singularities that are present in the displacement fields computed within the non-diffeomorphic setting, while it still maintains the accuracy of the standard modelling approach.

Although our results indicate that on average most of the surrogates considered here are equally suitable for motion estimation purposes, patient-specific differences exist. Therefore, surrogate simulations and the proposed framework could be used to determine optimal patient-specific surrogates (+ positions) during treatment planning.

A limitation of our study is that the results are solely based on simulated surrogate signals. Therefore, the accuracy of the motion estimation also depends on the accuracy of the simulations, which is highly influenced by the quality and spatial resolution of the images used. Furthermore, with this approach, we completely ignore any device-specific characteristics (e.g., noise) one would have to account for in clinical practise.

Our future work will therefore focus on the following aspects: First, we plan to

verify the simulations of the range-imaging-based surrogate signals by comparing them to real measurements. This will include the incorporation of device-specific noise into the simulation framework and an analysis of the influence of noise on the estimation accuracy. Furthermore, the evaluation of the estimation accuracy in the presence of intra- and inter-session breathing variations needs to be extended by including more patients and an increased number of 4D MRI image data sets with ideally a higher spatial resolution. We are further thinking about extending our studies by incorporation of additional image information sources like 4D cone-beam CT or kV fluoroscopy acquired during radiotherapy treatment of lung tumour patients. Such information could help to overcome some limitations of especially the considered 4D CT scenarios: 4D CT data sets are actually reconstructed from data acquired over several breathing cycles. This leads to the problem that we are (potentially) modelling relationships between information coming from different cycles in our experiments. Moreover, no direct inter-cycle variability analyses are possible on the basis of the reconstructed volumes; they would require repeated 4D CT imaging, which is hardly feasible due to additional dose exposure to the patient. The other information sources mentioned above, however, have their own specific disadvantages (4D cone-beam CT: currently low image quality; kV fluoroscopy: only 2D+time information) and are at the moment rarely acquired in clinical routine. Acquisition and incorporation of such information would, nevertheless, be one further step towards clinical practise. This would finally be in line with our next goal, which is is to incorporate surrogate-based information about breathing motion variations into 4D dose calculation (Werner *et al* 2012a, Werner *et al* 2012c).

Acknowledgments

This work is supported by the German Research Foundation DFG (HA 2355/9-2).

References

Albert A 1972 *Regression and the Moore-Penrose Pseudoinverse* Academic Press.

Arsigny V, Commowick O, Pennec X & Ayache N 2006 A log-Euclidean framework for statistics on diffeomorphisms *Proc. MICCAI 2006* **4190**, 924–931.

Ashburner J 2007 A fast diffeomorphic image registration algorithm. *Neuroimage* **38**(1), 95–113.

Beg M F, Miller M I, Trouvé A & Younes L 2005 Computing large deformation metric mappings via geodesic flows of diffeomorphisms *International Journal of Computer Vision* **61**(2), 139–157.

Borga M, Landelius T & Knutsson H 1997 A Unified Approach to PCA, PLS, MLR and CCA Technical Report LiTH-ISY-R-1992 Linköping University, The Institute of Technology.

Castillo R, Castillo E, Guerra R, Johnson V E, McPhail T, Garg A K & Guerrero T 2009 A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets *Phys Med Biol* **54**, 1849–1870.

Cerviño L I, Jiang Y, Sandhu A & Jiang S B 2010 Tumor motion prediction with the diaphragm as a surrogate: a feasibility study *Physics in Medicine and Biology* **55**(9), N221.

Dupuis P, Grenander U & Miller M 1998 Variational problems on flows of diffeomorphisms for image matching *Quarterly of Applied Mathematics* **LVI**(4), 587–600.

Ehrhardt J, Werner R, S. D *et al* 2007 An optical flow based method for improved reconstruction of 4D CT data sets acquired during free breathing. *Medical Physics* **34**(2), 711–721.

Ehrhardt J, Werner R, Schmidt-Richberg A & Handels H 2011 Statistical Modeling of 4D Respiratory Lung Motion Using Diffeomorphic Image Registration. *IEEE Trans Med Imag* **30**(2), 251–65.

Gao G, McClelland J, Tarte S, Blackall J & Hawkes D 2008 Modelling the respiratory motion of the internal organs by using Canonical Correlation Analysis and dynamic MRI *Proc. First International Workshop on Pulmonary Image Analysis, MICCAI 2008* pp. 145–154.

Hastie T, Tibshirani R & Friedman J 2009 *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* 2nd edn Springer.

Hernandez M, Bossa M N & Olmos S 2009 Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows *Int J Comput Vision* **85**(3), 291–306.

Jolliffe I 2002 *Principal Component Analysis* 2nd edn Springer.

Keall P J *et al* 2006 The management of respiratory motion in radiation oncology report of AAPM Task Group 76 *Medical Physics* **33**(10), 3874–3900.

Klinder T, Lorenz C & Ostermann J 2010 Prediction framework for statistical respiratory motion modeling. *Proc. MICCAI 2010* **6363**, 327–334.

Kubo H D & Hill B C 1996 Respiration gated radiotherapy treatment: a technical study *Physics in Medicine and Biology* **41**(1), 83.

Li R, Lewis J H, Jia X, Zhao T, Liu W, Wuenschel S, Lamb J, Yang D, Low D A & Jiang S B 2011 On a PCA-based lung motion model *Physics in Medicine and Biology* **56**(18), 6009.

Liu X, Oguz I, Pizer S M & Mageras G S 2010 Shape-correlated deformation statistics for respiratory motion prediction in 4D lung *Proc. SPIE Medical Imaging 2010: Visualization, Image-Guided Procedures* **7625**.

Lu W, Parikh P J, Naqa I M E, Nystrom M M, Hubenschmidt J P, Wahab S H, Mutic S, Singh A K, Christensen G E, Bradley J D & Low D A 2005 Quantitation of the reconstruction quality of a four-dimensional computed tomography process for lung cancer patients *Medical Physics* **32**(4), 890–901.

Martin J, McClelland J, Yip C, Thomas C, Hartill C, Ahmad S, OBrien R, Meir I, Landau D & Hawkes D 2013 Building motion models of lung tumours from cone-beam CT for radiotherapy applications *Physics in Medicine and Biology* **58**(6), 1809.

McClelland J, Hawkes D, Schaeffter T & King A 2013 Respiratory motion models: A review *Medical Image Analysis* **17**(1), 19 – 42.

Schaller C, Penne J & Hornegger J 2008 Time-of-flight sensor for respiratory motion gating. *Medical Physics* **35**(7), 3090–3093.

Schmidt-Richberg A, Ehrhardt J, Werner R & Handels H 2010 Diffeomorphic diffusion registration of lung CT images *Proc. Workshop Medical Image Analysis for the Clinic: A Grand Challenge - MICCAI 2010* pp. 55–62.

Schweikard A, Glosser G, Bodduluri M, Murphy M J & Adler J R 2000 Robotic motion compensation for respiratory movement during radiosurgery *Computer Aided Surgery* **5**(4), 263–277.

Vandemeulebroucke J, Rit S, Kybic J, Clarysse P & Sarrut D 2011 Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs *Medical Physics* **38**(1), 166–178.

Vercauteren T, Pennec X, Perchant A & Ayache N 2009 Diffeomorphic demons: Efficient non-parametric image registration *NeuroImage* **45**(1, Supplement 1), S61 – S72.

Werner R, Ehrhardt J, Schmidt-Richberg A, Albers D, Frenzel T, Petersen C, Cremers F & Handels H 2012c Towards accurate dose accumulation for step-&-shoot IMRT: Impact of weighting schemes and temporal image resolution on the estimation of dosimetric motion effects *Z Med Phys* **22**, 109–122.

Werner R, Ehrhardt J, Schmidt-Richberg A, Wilms M, Blendowski M & Handels H 2012b A Diffeomorphic MLR Framework for Surrogate-based Motion Estimation in Radiation Therapy: Concept and First Evaluation *Proc. Informatik 2012* **P-208**, 1774–1785.

Werner R, Wilms M, Ehrhardt J, Schmidt-Richberg A, Blendowski M & Handels H 2012a A Diffeomorphic MLR Framework for Surrogate-based Motion Estimation and Situation-adapted Dose Accumulation *Proc. Image-Guidance and Multimodal Dose Planning in Radiation Therapy: a MICCAI Workshop - MICCAI 2012* pp. 42–49.

Wilms M, Ehrhardt J & Handels H 2012 A 4D Statistical Shape Model for Automated Segmentation of Lungs with Large Tumors *Proc. MICCAI 2012* **7511**, 347–354.

Wilms M, Werner R, Ehrhardt J, Schmidt-Richberg A, Blendowski M & Handels H 2013 Surrogate-based Diffeomorphic Motion Estimation for Radiation Therapy: Comparison of Multivariate Regression Approaches *Proc. SPIE Medical Imaging 2013: Image Processing* **8669**, 866915.

Zhang Q, Pevsner A, Hertanto A, Hu Y C, Rosenzweig K E, Ling C C & Mageras G S 2007 A patient-specific respiratory model of anatomical motion for radiation treatment planning *Medical Physics* **34**(12), 4772–4781.