Preprint of invited book chapter
Published in: Wu & Sabuncu, eds. Machine learning in medical imaging. Elsevier, 2016

CHAPTER 10 - COVER PAGE

Learning pathological deviations from a normal pattern of myocardial motion: added-value for CRT studies?

Nicolas Duchateau[1]*, Gemma Piella[2], Alejandro Frangi[3], Mathieu De Craene[4].

[1] Asclepios research project, Inria Sophia Antipolis, France

[2] Universitat Pompeu Fabra, Barcelona, Spain

[3] CISTIB, University of Sheffield, UK

[4] Philips Medisys, Suresnes, France

*** Corresponding author:**

Address: Inria Asclepios, 2004 route des Lucioles BP 93 06902 Sophia Antipolis Cedex, France

Tel: +33.49238.5024

Fax +33.49238.7669

Email: nicolas.duchateau@inria.fr

Preprint of invited book chapter
Published in: Wu & Sabuncu, eds. Machine learning in medical imaging. Elsevier, 2016

NON-PRINT ITEMS

Abstract

Strong links exist between mechanical dyssynchrony and the response to cardiac resynchronization therapy (CRT). Recent publications recommend identifying correctable dyssynchrony patterns with a specific motion and deformation signature. The learning of these patterns is visual and highly subjective. We take advantage of statistical atlas and dimensionality reduction tools to learn a representation of these patterns. We hypothesize that myocardial motion patterns belong or lie close to a non-linear manifold, and model them as a pathological deviation from normality. Furthermore, we propose distances to compare new subjects with those patterns and with normality. We evaluate the value of this approach on 2D echocardiographic sequences from CRT candidates at baseline, with pacing on, and at one-year follow-up. We demonstrate that relating pattern changes with patient response is valuable, and paves the ground towards better therapy planning.

Key Words: Manifold learning; pattern analysis; cardiac atlas; myocardial motion; cardiac resynchronization therapy.

Preprint of invited book chapter
Published in: Wu & Sabuncu, eds. Machine learning in medical imaging. Elsevier, 2016

CHAPTER STARTS HERE

1) Introduction

1.1) Cardiac resynchronization therapy

In the last decade, cardiac resynchronization therapy (CRT) raised as the recommended procedure to treat heart failure in patients with asynchronous contraction of the cardiac chambers (Yu et al. 2008). A biventricular pacing device optimizes the left/right ventricular delay, and eventually the atrio-ventricular delay. When deficiencies in the conduction system are compensated, synchronous contraction is expected to improve the heart efficiency and patient survival. Established international guidelines recommend pacing patients with symptomatic heart failure, electrical abnormalities and decreased left ventricular function (Auricchio et al. 2013). Nonetheless, these selection criteria are sub-optimal. The therapy fails to "improve enough" patient condition in approximately 30% of the cases (clinical response), and reverse remodeling in 50% of the cases (volume response) (Bleeker et al. 2006). This issue is of primary concern given the prevalence of the symptoms (between 25 and 50% of heart failure patients, at least 15 million people in Europe (van Veldhuisen et al. 2009)), and the associated costs (better implanting 1% of the devices in the United States between 2003 and 2007 (Laskey et al. 2012) already means an impact of 105M€ ---15K€ per device, 7000 devices).

Currently, no consensus exists on better selection criteria. Publications abound on new predictive indexes better than others. Paradoxically, they over-focus on the blind assessment of mechanical dyssychrony (mainly through peak or time-to-event measurements), instead of better understanding the mechanisms of therapy response. A change towards a more comprehensive strategy was more recently claimed (Fornwalt 2011). Some studies highlighted the relevance of distinguishing between specific types of dyssynchrony, each one associated to a pattern correctable by CRT (Parsai et al. 2009 ; Doltra et al. 2014). They pave the ground towards interpreting those mechanical patterns in light of precise electrical and structural abnormalities (Vernooy et al. 2014).

According to these findings, improving the patient selection should therefore move towards the recognition of these different types of dyssynchrony in new candidates.

1.2) Patterns of motion / deformation

Myocardial motion and deformation can be extracted from cardiac sequences by combining segmentation and temporal registration/tracking. Local speckles attached to the myocardium allow tracking along 2D or 3D echocardiographic sequences and computing these parameters locally (Duchateau et al. 2013a). However, once temporal and anatomical variability are removed, the comparison of spatiotemporal patterns of motion and deformation is challenging.

Quantifying the distance from a patient to a normal motion pattern is an intuitive way of encoding pathological patterns. Clinicians actually do so when diagnosing a new patient: given the representation of a normal contraction (previously learnt), where and how much does the observed patient differ from normality? Then, once categorized as abnormal, can a similar process be extended to quantify the distance to a known (ab)normal pattern?

This philosophy guided our work on dyssynchrony patterns amenable to CRT response. In a first instance, an atlas of normal motion was built from healthy subjects (Duchateau et al. 2011). Average and covariance of myocardial velocities were encoded at each point of the myocardium. The motion of CRT candidates was locally compared to the atlas by a statistical distance to normality (p-value associated to the Mahalanobis distance). Motion abnormalities were coherent across subgroups of dyssynchrony patterns (Duchateau et al. 2012b). Notably, a specific pattern of intra-ventricular dyssynchrony (also called *septal flash*) predominated among the responders to the therapy, and mostly disappeared at mid-term follow-up. Qualitative observations previously reported that this pattern is associated to high response rates, if corrected (Parsai et al. 2009 ; Doltra et al. 2014).

In a second instance, we therefore intended to represent this specific pattern as a progressive deviation from normality (Duchateau et al. 2012a). We hypothesized that pattern variations were encoded in a non-linear

space of lower dimensionality, which could be modeled as a manifold. Then, we proposed to quantify the distance between a new patient and this specific pattern, in complement to its distance to normality. We expected that such distances could (before CRT) reflect the probability of response, and (after CRT) link the abnormality reduction with the actual response (Duchateau et al. 2013b).

1.3) Summary of the challenges

Due to the choice of a machine learning methodology for such a clinical application, we believe that three challenges should be highlighted:

- First, determining the *amount of pre-processing required* to study motion and deformation patterns. Features quality strongly impacts the performance of machine learning algorithms, although exaggerated efforts are often dedicated to their extraction. The literature abounds of methods to estimate myocardial motion via cardiac segmentation and tracking. Specific concrete physiological knowledge could refine the features extraction in our application: local anatomical coordinates (radial, circumferential and longitudinal), spatiotemporal a-priori (phases of the cycle and anatomy should match), quasi-incompressibility of the cardiac motion (DeCraene et al. 2012), etc. In addition to these specificities, we decided to pre-compute distances to a reference for normality to imitate the process of learning of clinicians, and enhance specific features of the motion patterns (Duchateau et al. 2012a).

- Then, knowing how to *"learn" a representation of a given pathology*. We believe that different grades of the same disease, for a patient or within a population, can be modeled as progressive impairments of a normal condition. We hypothesized that motion patterns belong or lie close to a non-linear manifold that can be learnt from data by means of non-linear dimensionality reduction. Specific issues consist in choosing relevant learning techniques, depending on which feature characteristics to highlight, and comparing new subjects to this learnt representation.

- Finally, evaluating whether this is representation *useful* for the clinical application. As indicated in (Parsai et al. 2009 ; Doltra et al. 2014), relating pattern changes with patient response is valuable for the therapy

planning. However, their assessment is currently qualitative or based on simple quantitative indexes. We therefore focused on demonstrating the value of our method to go beyond these limitations and quantify pattern-specific changes, to investigate if the pattern was fully or partially corrected after treatment.

2) Features extraction: statistical distance from normal motion

2.1) Construction of abnormality maps

We chose to use myocardial velocities as local descriptor for cardiac motion. They correspond to small transformations that do not require elaborated computational anatomy techniques (Duchateau et al. 2011). Furthermore, the signature of the septal flash pattern is more contrasted on velocities than on displacements or strain patterns (Duchateau et al. 2014). These velocities are extracted from echocardiographic sequences by spatiotemporal registration (DeCraene et al. 2012) or commercial speckle-tracking. These velocities are normalized both spatially and temporally to take into account differences in the anatomy and in the timing of physiological events (Duchateau et al. 2011).

We decided to post-process these data by highlighting values significantly different from normality (Fig 10.1). This consists of three steps. At each location of the myocardium, average and covariance of velocities over a set of healthy volunteers encode a representation of local normal motion (the atlas in (Duchateau et al. 2011)). Then, at the same locations, the velocity of each individual is compared to the distribution of velocity vectors for this reference population, using the Mahalanobis distance. Finally, the p-value associated to this distribution (assumed as Gaussian) encodes abnormality at each location: a low p-value indicates high abnormality, and a p-value close to 1 stands for normal motion.

In our application, contrast in the abnormality maps is enhanced by using the logarithm of this p-value. On top of this, the maps are multiplied by the sign of the radial velocity, so that the characteristic pattern of septal flash is highlighted (the inward and outward motion of the septum during early-diastole, Fig 10.1).

2.2) Which statistics for motion patterns?

Voxel-based statistics are often retained to analyze group-wise and inter-group differences after alignment to a common reference. Each voxel is considered independently from the others, which may affect the statistical power of the results and bias conclusions. In our application, the link between neighboring regions cannot be discarded due to the data nature: mechanical tissue properties, regional noise patterns such as speckles, smoothness and temporal consistency of the extracted motion, etc.

Global statistics are also often retained. Methods search for an optimal space to compare subjects, according to specific criteria (e.g. explaining the data variance, or discriminating groups of subjects). However, they often neglect the constraints of imaging data, in the sense that generalizations of the model should remain physiologically coherent.

When the structure of the data space is known (e.g. the manifold of diffusion tensors or diffeomorphic transformations), mathematical operators exist to define statistics compliant with this space. In contrast, in our application, the structure of such a manifold should be learnt from data. We chose a specific non-linear dimensionality reduction algorithm that is suitable when the data distribution is not clustered (Isomap (Tenenbaum et al. 2000)). However, our framework is flexible and could easily be adapted to other embedding algorithms, which share similar principles (Yan et al. 2007).

3) Manifold learning: characterizing pathological deviations from normality

Our dataset consists of 109 maps of abnormality as the one in Fig 10.1, where rows and columns respectively correspond to the position along the septum and the time along the cycle. They were extracted from 2D echocardiographic sequences in a 4-chamber view. Focus was kept on the isovolumic contraction (where septal flash appears) and the systole. One map, completely synthetic, was filled with 0 values and served to encode true normality. The *learning set* was composed by this synthetic map and 50 real maps

corresponding to a septal flash (Sec.1). The *testing set* was made of the remaining 58 real maps, corresponding to 6 cases with septal flash, 31 CRT candidates without septal flash, and 21 healthy volunteers.

The abnormality maps were considered as high-dimensional objects (20 dimensions for time x 31 dimensions for space), handled as column vectors by our method.

[Insert Figure 10.1 here]

3.1) Learning part: pathological deviations from normality (manifold learning)

From high-dimensional motion patterns to low-dimensional coordinates (training set).

The Isomap algorithm (Tenenbaum et al. 2000) is used to reduce the dimensionality of the data while preserving specific data arrangement. It maps each (high-dimensional) input sample to a space of (low-dimensional) coordinates, where the Euclidean distance approximates the geodesic distance between two samples. One hypothesis is required: there is a low-dimensional manifold that can explain the main variations in the data.

The algorithm consists of three steps. First, input samples are connected via a nearest-neighbors search, using the Euclidean distance as metric. Then, the shortest path between each pair of samples is taken as surrogate for the geodesic distance, and stored into a matrix. Finally, this matrix is centered and diagonalized. Due to the matrix diagonalization and our initial hypotheses, dimensions of lower significance can be removed. This leads to a space of low-dimensional coordinates into which the geodesic distance is approximated by the Euclidean distance (Fig 10.2). The top N eigenvectors from the matrix diagonalization define the coordinates in this new N-dimensional space.

[Insert Figure 10.2 here]

Tuning parameters.

The algorithm starts building a graph of the training samples, by connecting nearest neighbors according to a given metric (in our case, the Euclidean distance). In this procedure, the number of neighbors needs to be determined. Synthetic experiments on the Swiss roll (Fig 10.3) can show that the accuracy of the manifold estimation increases with the number of neighbors until a short-circuit happens, which breaks the data arrangement. The ratio between the geodesic and the Euclidean distances (the *Isomap error*), computed in the output space, highlights the apparition of such a short-circuit (Fig 10.3b). However, this measure may not be optimal on real data, in particular when the number of training samples is low.

We complemented this measure by the *node flow on a graph* (Choi & Choi 2007), which reflects the number of shortest paths passing through each node of the graph. This measurement is directly performed on the graph, before any dimensionality reduction, and is therefore independent of the choice of dimensionality. A uniform arrangement of the training samples is reflected by the uniformity of the node flow distribution. In contrast, much higher node flow is observed at the nodes where a short-circuit happens, as most of the shortest paths pass by this point (Fig 10.3c).

[Insert Figure 10.3 here]

However, in this synthetic example, the dimensionality of the output space is known---the Swiss roll is a 2D plane embedded in a 3D space. This is not the case on real data, and the number of dimensions of the output space should be estimated together with the number of nearest neighbors. Thus, parameter estimation should start by checking the presence of a short-circuit by measuring the node flow, and then jointly evaluating the Isomap error against the dimensionality of the output space and the number of nearest neighbors. On our data, an increase in the number of neighbors did not point out the apparition of a short-circuit (Fig 10.4a). According to the Isomap error (Fig 10.4b), the dimensionality was set to 4. Few error changes were observed when using less than 30 neighbors, and this number was set to 5 to minimize the computational time.

The Isomap algorithm may be highly affected by an heterogeneous density in the samples distribution, and in particular in the presence of “holes” in the distribution. This is partially the case on our data, as visible in

Fig 10.4c. On this aspect of the learning, our method could be improved by algorithms that specifically target this robustness, such as the diffusion maps (Coifman & Lafon 2006). Note that similar concerns arise for the mapping part of our method, addressed in the following section.

[Insert Figure 10.4 here]

Visualization: data spread and main directions.

The first necessary checking consists in visualizing the low-dimensional embedding of the data. Two-dimensional representations such as in Fig 10.5a are easily readable. The depicted patterns correspond to the abnormality maps of the training set. We can qualitatively check that patterns are arranged in a coherent way, with more abnormal patterns on the border zone of the graph.

This information is complemented by Fig 10.5b, where synthetic patterns were generated from coordinates evolving from normality along the first two dimensions (the dashed lines in Fig 10.5a). Notably, the characteristic inward/outward events of the septal flash are preserved, while this is not the case with linear dimensionality reduction (Duchateau et al. 2012a). Note that this figure requires being able to reconstruct motion patterns from the low-dimensional coordinates, using the techniques described in Sec.3.2.

[Insert Figure 10.5 here]

3.2) Testing part: distances to the modeled pathology and to normality

From high-dimensional motion patterns to low-dimensional coordinates (testing set).

The low-dimensional coordinates associated to the motion pattern of a new subject can be estimated by means of non-linear regression (Fig 10.6a). We formulated it as an inexact matching problem using kernels---also referred to as ridge regression. The optimization looks for a "smooth enough" interpolating function (regularization term) that gets "close enough" to the training samples (similarity term). The kernel formulation both constrains the search into a given space of smooth functions (fixed by the kernel bandwidth), and

provides an analytical solution to the interpolation problem. A scalar weight balances the contribution of the regularization and similarity terms.

To improve the quality of the mapping of new subjects, we adapted the methods on two aspects.

First, our dataset includes a synthetic zero-valued abnormality map that encodes true normality (the bold map in Fig 10.5). The interpolation problem is therefore adapted to an exact matching for this specific sample and an inexact matching for the remaining ones (Fig 10.7a). This is simply addressed by modifying one matrix constraint in the analytical solution of the kernel regression.

[Insert Figure 10.6 here]

Then, similarly to the learning algorithm of Sec.3.1, this subpart of the algorithm is also sensitive to the non-uniform distribution of samples. Artifacts may arise from the use of a single-scale kernel, generally set to the average density of the samples. Two methods were therefore tested:

- A *locally-adjustable kernel* (Duchateau et al. 2012a), whose bandwidth is adapted to the local density of the samples, defined as the local neighborhood size. Although this approach provides robustness to local density variations, it deviates from the original formulation where a given kernel determines the space of functions where the problem is solved.

- A *multiscale strategy* (Duchateau et al. 2013c), where the interpolation process is iterated across scales, by dividing the kernel bandwidth by two at each iteration and interpolating the residual (Fig 10.7b). This approach also has the advantage of removing one parameter of the method, as the kernel bandwidth is automatically determined, starting from the data spread to the average density of the samples.

[Insert Figure 10.7 here]

From low-dimensional motion coordinates to high-dimensional motion patterns (testing set).

The second part of the algorithm solves the reverse problem: it estimates the motion pattern that corresponds to the low-dimensional coordinates determined by the first step (Fig 10.6b). The problem is formulated similarly, by switching the role of the motion patterns and the coordinates in the previous formulation.

Projection to the manifold and distances computation.

As coordinates belong to the low-dimensional space associated to the manifold of motion patterns, all the patterns generated from these coordinates belong to the manifold. In the case of our application, the manifold is learnt from a population with a specific abnormal pattern, and all reconstructed data are expected to show out this pattern (Fig 10.5b). This means that any motion abnormality pattern can be "projected" to the manifold by the combination of these two interpolations. In other words, the method estimates the element of the manifold that shares coordinates with a tested individual, possibly out of this manifold.

Two distances are defined (Fig 10.6c):

- A distance between any abnormality map and the manifold (*distance to the manifold*), which quantifies how far a given subject is from the modeled abnormal pattern. It corresponds to the total error between a given abnormality map and its reconstruction.

- A *distance to normality*. It corresponds to the Euclidean distance between the coordinates of the new subject and the origin---the synthetic map defining true normality.

Tuning parameters.

As for the learning part of the algorithm, the method parameters need to be estimated. Each interpolation depends on a kernel bandwidth and a regularization weight. Kernel bandwidths are automatically determined (Sec.3.2). The regularization weights for the two consecutive interpolations are jointly estimated by heuristic tests. Their optimal values are determined based on the generalization ability over the training set---the

reconstruction error when using a leave-one-out approach (Davies et al. 2010). In our application, they corresponded to 10^{1} and $10^{0.5}$, respectively (Duchateau et al. 2012a).

Naturally, the performance of the methods depends on the size of the training population. The two estimated distances (to the manifold and to normality) were computed from randomly generated datasets of smaller size. Convergence was assessed by the number of subjects above which the results stabilize to their final value ±5%: above 45 subjects for the distance to the manifold, and above 41 subjects for the distance to normality (Duchateau et al. 2012a). Both indicate that results on our population of 50 real subjects can be trusted.

A summary of the estimation of the method parameters is given in Table 10.1.

[Insert Table 10.1 here]

4) Back to the clinical application: understanding CRT-induced changes

The motion patterns of our population were analyzed with respect to the two proposed distances: the distance to the manifold, which models the abnormal pattern of septal flash, and the distance to normality along the manifold. These distances define a 2D space where subgroup differences can easily be visualized (Fig 10.8), as discussed below.

Analysis per population

We first checked that the two distances provided a meaningful organization of all the data at baseline.

The training set, made of septal flash cases, has a low distance to the modeled pattern. This actually reflects the reconstruction error of the interpolations, formulated as an inexact matching. Such cases spread all along the horizontal axis from low to high abnormal patterns, which also indicates that the whole support of the manifold defines the distance to normality. Few of them are close to the origin for normality, which may

arise from the accuracy of the septal flash diagnosis and the accuracy of the abnormality maps to detect low-grade abnormalities.

Among the testing set, patients with septal flash are mapped close to the training cases, according to their respective abnormality. Volunteers lie close from the origin for normality. Two subjects are outliers (black crosses out of the dashed ellipse). They correspond to cases with high abnormalities due high velocities despite a normal pattern, lying out of the range of the other healthy subjects. Their pattern of abnormality cannot be reconstructed by the training set of septal flash cases, which leads to a large distance to the modeled pattern. Similarly, patients with an abnormal pattern different from septal flash are mapped far from the manifold, and out of the normality range.

Link with CRT response

The manifold is exclusively learnt from septal flash patterns. When looking at individual evolutions with CRT, our method only evaluates if a studied pattern (not necessarily septal flash) got closer to septal flash or normality. It the pattern is different from septal flash, it does not tell if this pattern is still present or not. Thus, we focused the outcome analysis on the subjects diagnosed with septal flash at baseline (Fig 10.8b and 8c).

[Insert Figure 10.8 here]

Patients were separated according to the volume response to CRT, measured by a clinical expert as a reduction ≥15% in left-ventricular end-systolic volume, without heart transplantation. Three subgroups were identified: responders who recovered normal motion at follow-up (who went within the normality range defined by the healthy subjects), responders for which abnormal motion is still present, and non-responders.

These subgroups cannot be distinguished at baseline: the presence of septal flash at baseline is not a sufficient condition for CRT response. This coincides with the philosophy of (Parsai et al. 2009 ; Doltra et al. 2014), which interpreted response at follow-up in light of the correction of the identified abnormal mechanisms. In our population, significant changes are visible at follow-up. Patients within the range of normal motion are mostly responders. The other patients have higher distance to the manifold, suggesting that septal

flash may have disappeared. However, their abnormalities are still high, which may explain their lack of response. Beyond these qualitative interpretations, the amount of changes can be quantified with our method and could further be related with quantitative indexes of therapy outcome.

Further results and discussion can be found in (Duchateau et al. 2013b). They notably include the evolution of the whole population, and the outcome immediately after pacing. Complementary interpretations are also given in (Duchateau et al. 2012a), which points out the limits of a linear analysis (PCA) and the meaning of the proposed distances against the total abnormality in each map.

5) Discussion / Future work

We presented a method to model pathological myocardial motion patterns as a smooth deviation from normality, along a non-linear manifold structure, learnt from data. This representation serves for the comparison of new samples to the modeled pattern and to normality, and provides quantitative insights into the mechanisms of CRT response.

Pattern-based comparisons.

On technical aspects, our main contribution resides in the methodology to compare myocardial motion patterns. Previous work used atlas tools to align data in time and space, and to enhance the contrast in the features of interest by a statistical comparison to normality at each voxel (Duchateau et al. 2011). These tools already facilitated the quantitative analysis of CRT response, by highlighting the prevalence and the correction of specific abnormal patterns in groups of responders and non-responders. In this chapter, we illustrated the potential of the pattern-wise comparisons proposed in (Duchateau et al. 2012a). Pattern variations now model the spatiotemporal interdependences in the data. Furthermore, the proposed distances provide physiologically-coherent support to the interpretation of the patterns evolution with therapy.

The methodology is based on standard manifold learning and non-linear regression. Improvements could be made on several parts of the pipeline, depending on the application targeted: getting more power out of the

metrics, adding robustness to the density of the samples during the learning, combining features of different types (Sanchez-Martinez et al. 2015), modeling several abnormal patterns at once, etc. The pipeline is easily transposable to other applications, imaging modalities, and features of interest (not necessarily in cardiac imaging). However, the main decision should be taken as early as possible, to decide whether using a similar but complex framework (statistical atlas and learning a manifold) can add value to the clinical application. In our case, there was a clear claim for the quantitative analysis of dyssynchrony patterns amenable to CRT response (Fornwalt 2011), and a convergence of clinical studies towards mechanistic approaches (Parsai et al. 2009), which has been confirmed afterwards (Doltra et al. 2014). 2D echocardiographic data were retained due to the wide implantation of the modality in clinical routine and CRT monitoring. The high temporal resolution and the easy checking of 2D echocardiographic data outputs also facilitated this decision, at a moment when 3D data was not straightforward to handle.

Going beyond Parsai's paper?

The results in (Parsai et al. 2009) were already a great improvement to CRT, as the decision algorithm improved patient selection by removing ambiguity on 3 out of 5 groups, which always or never respond. Simple criteria defined the inclusion to a given group and facilitated the reproducibility of the results. Our methodology is less easy to implement in clinical routine, and simpler but comprehensive indexes that characterize dyssynchrony patterns may be more relevant in this context. However, our approach allows to quantify pattern evolution with therapy, which would not be straightforward with the indexes in (Parsai et al. 2009), in particular if the pattern is highly modified.

The results in (Parsai et al. 2009) also offered a broader view on the way to treat dyssynchrony. Intra-ventricular dyssynchrony is purely electrical, meaning that resynchronization should lead to response in subjects with septal flash---therefore defining a hyper-responders population. In contrast, structural disease on top of the electrical abnormalities worsens the outcome: although electrical dyssynchrony is corrected, mechanical abnormalities (myocardial infarct) still exist and the probability of response is highly decreased. Going beyond these interpretations, confirmed by our results, should consider additional variety in these mechanisms. External factors could affect the clinical condition, such as the presence of atrial fibrillation, the lack of contractile reserve, or the lead position, and should be considered in further studies. Nonetheless, from

a learning perspective, the user should always keep in mind the amount of effort required to extract features of interest for these data.

Finally, one of the strongest limitations to CRT resides in the definition of response. Paradoxically, no agreement exists (Fornwalt et al. 2010). We hope that our approach, based on the quantification of abnormality, could open the way towards less binary decision of therapy success. Further studies should evaluate the added-value of a continuous spectrum of responses, which may include measures of the abnormality evolution with the therapy.

Acknowledgements

This work was done between 2008 and 2013 under the partial support of the Spanish Industrial and Technological Development Center (CDTeam and cvREMOD CEN-20091044), the Spanish Ministry of Science and Innovation, Plan E and ERDF (STIMATH TIN2009-14536-C02-01), and the European Commission's 7th Framework Program (euHeart FP7-ICT-224495). The authors acknowledge the contribution of their co-authors on related publications, in particular the ones from Bart Bijnens (ICREA, Universitat Pompeu Fabra, Barcelona, Spain) and Marta Sitges (Hospital Clínic, Barcelona, Spain) on the clinical application, and Vicent Caselles (Universitat Pompeu Fabra, Barcelona, Spain) on density invariance aspects.

References

(Auricchio et al. 2013) Brignole M, Auricchio A, Baron-Esquivias G, Bordachar P, Boriani G, Breithardt OA, Cleland J, Deharo JC, Delgado V, Elliott PM, Gorenek B, Israel CW, Leclercq C, Linde C, Mont L, Padeletti L, Sutton R, Vardas PE; ESC Committee for Practice Guidelines (CPG), Zamorano JL, Achenbach S, Baumgartner H, Bax JJ, Bueno H, Dean V, Deaton C, Erol C, Fagard R, Ferrari R, Hasdai D, Hoes AW, Kirchhof P, Knuuti J, Kolh P, Lancellotti P, Linhart A, Nihoyannopoulos P, Piepoli MF, Ponikowski P, Sirnes PA, Tamargo JL, Tendera M, Torbicki A, Wijns W, Windecker S; Document Reviewers, Kirchhof P, Blomstrom-Lundqvist C, Badano LP, Aliyev F, Bänsch D, Baumgartner H, Bsata W, Buser P, Charron P, Daubert JC, Dobreanu D, Faerestrand S, Hasdai D, Hoes AW, Le Heuzey JY, Mavrakis H, McDonagh T, Merino JL, Nawar MM, Nielsen JC, Pieske B, Poposka L, Ruschitzka F, Tendera M, Van Gelder IC & Wilson CM 2013, '2013 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy', *European Heart Journal*, vol. 34, no.29, pp. 2281-2329.

(Bleeker et al. 2006) Bleeker GB, Bax JJ, Fung JW, van der Wall EE, Zhang Q, Schalij MJ, Chan JY & Yu CM 2006, 'Clinical versus echocardiographic parameters to assess response to cardiac resynchronization therapy'. *American Journal of Cardiology*, vol. 97, no.2, pp. 260-263.

(Choi & Choi 2007) Choi H & Choi S 2007, 'Robust kernel isomap', Pattern Recognition, vol. 40, no. 3, pp. 853-862.

(Coifman & Lafon 2006) Coifman R & Lafon S 2006, 'Diffusion maps', *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5-30.

(Davies et al. 2010) Davies RH, Twining CJ, Cootes TF & Taylor CJ 2010, 'Building 3-d statistical shape models by direct optimization', *IEEE Transactions on Medical Imaging*, vol. 29, no. 4, pp. 961-981.

(De Craene et al. 2012) De Craene M, Piella G, Camara O, Duchateau N, Silva E, Doltra A, D'hooge J, Brugada J, Sitges M & Frangi AF 2012, 'Temporal diffeomorphic free-form deformation: application to motion and strain estimation from 3D echocardiography', *Medical Image Analysis*, vol. 16, no. 2, pp. 427-450.

(Doltra et al. 2014) Doltra A, Bijnens B, Tolosana JM, Borràs R, Khatib M, Penela D, De Caralt TM, Castel MÁ, Berruezo A, Brugada J, Mont L & Sitges M 2014, 'Mechanical abnormalities detected with conventional echocardiography are associated with response and midterm survival in CRT', *JACC Cardiovascular Imaging*, vol. 7, no. 10, pp. 969-979.

(Duchateau et al. 2011) Duchateau N, De Craene M, Piella G, Silva E, Doltra A, Sitges M, Bijnens BH & Frangi AF 2011, 'A spatiotemporal statistical atlas of motion for the quantification of abnormalities in myocardial tissue velocities', *Medical Image Analysis*, vol. 15, no. 3, pp. 316-328.

(Duchateau et al. 2012a) Duchateau N, De Craene M, Piella G & Frangi AF 2012, 'Constrained manifold learning for the characterization of pathological deviations from normality', *Medical Image Analysis*, vol. 16, no. 8, pp. 1532-1549.

(Duchateau et al. 2012b) Duchateau N, Doltra A, Silva E, De Craene M, Piella G, Castel MÁ, Mont L, Brugada J, Frangi AF & Sitges M 2012, 'Atlas-based quantification of myocardial motion abnormalities: added-value for understanding the effect of cardiac resynchronization therapy', *Ultrasound in Medicine and Biology*, vol. 38, no. 12, pp. 2186-2197.

(Duchateau et al. 2013a) Duchateau N, Bijnens BH, D'hooge J & Sitges M 2013, 'Three-dimensional assessment of cardiac motion and deformation' in *3D echocardiography, 2nd ed*, ed T Shiota, CRC Press, pp. 201-213.

(Duchateau et al. 2013b) Duchateau N, Piella G, Doltra A, Mont L, Bijnens BH, Sitges M & De Craene M 2013, 'Manifold learning characterization of abnormal myocardial motion patterns: application to CRT-induced changes' in *Proc. Functional Imaging and Modeling of the Heart, LNCS 7945*, eds S Ourselin, D Rueckert & N Smith, Springer-Verlag, pp. 450-457.

(Duchateau et al. 2013c) Duchateau N, De Craene M, Sitges M & Caselles V 2013, 'Adaptation of multiscale function extension to inexact matching. Application to the mapping of individuals to a learnt manifold' in *Proc. SEE International Conference on Geometric Science of Information, LNCS 8085*, eds F Nielsen & F Barbaresco, Springer-Verlag, pp. 578-586.

(Duchateau et al. 2014) Duchateau N, Sitges M, Doltra A, Fernández-Armenta J, Solanes N, Rigol M, Gabrielli L, Silva E, Barceló A, Berruezo A, Mont L, Brugada J & Bijnens B 2014, 'Myocardial motion and deformation patterns in an experimental swine model of acute LBBB/CRT and chronic infarct', *International Journal of Cardiovascular Imaging*, vol. 30, no. 5, pp. 875-887.

(Fornwalt et al. 2010) Fornwalt BK, Sprague WW, BeDell P, Suever JD, Gerritse B, Merlino JD, Fyfe DA, León AR & Oshinski JN 2010, 'Agreement is poor among current criteria used to define response to cardiac resynchronization therapy', *Circulation*, vol. 121, no. 18, pp. 1985-1991.

(Fornwalt 2011) Fornwalt BK 2011, 'The dyssynchrony in predicting response to cardiac resynchronization therapy: A call for change', *Journal of the American Society of Echocardiography*, vol. 24, no. 2, pp. 180-184.

(Laskey et al. 2012) Laskey W, Awad K, Lum J, Skodacek K, Zimmerman B, Selzman K & Zuckerman B 2012, 'An analysis of implantable cardiac device reliability. The case for improved postmarketing risk assessment and surveillance', *American Journal of Therapeutics*, vol. 19, no. 4, pp. 248-254.

(Parsai et al. 2009) Parsai C, Bijnens B, Sutherland GR, Baltabaeva A, Claus P, Marciniak M, Paul V, Scheffer M, Donal E, Derumeaux G & Anderson L 2009, 'Toward understanding response to cardiac resynchronization therapy: left ventricular dyssynchrony is only one of multiple mechanisms', *European Heart Journal*, vol. 30, no. 8, pp. 940-949.

(Sanchez-Martinez et al. 2015) Sanchez-Martinez S, Duchateau N, Bijnens B, Erdei T, Fraser A & Piella G 2015, 'Characterization of myocardial motion by multiple kernel learning: application to heart failure with preserved ejection fraction' in *Proc. Functional Imaging and Modeling of the Heart, LNCS 9126*, eds H van Assen, P Bovendeerd & T Delhaas, pp. 65-73.

(Tenenbaum et al. 2000) Tenenbaum JB, de Silva V & Langford JC 2000, 'A global geometric framework for nonlinear dimensionality reduction', *Science*, vol. 290, no. 5500, pp. 2319-2323.

(van Veldhuisen et al. 2009) van Veldhuisen DJ, Maass AH, Priori SG, Stolt P, van Gelder IC, Dickstein K & Swedberg K 2009, 'Implementation of device therapy (cardiac resynchronization therapy and implantable cardioverter defibrillator) for patients with heart failure in Europe: changes from 2004 to 2008', *European Journal of Heart Failure*, vol. 11, no. 12, pp. 1143-1151.

(Vernooy et al. 2014) Vernooy K, van Deursen CJ, Strik M & Prinzen FW 2014, 'Strategies to improve cardiac resynchronization therapy', *Nature Reviews Cardiology*, vol. 11, no. 8, pp. 481-493.

(Yan et al. 2007) Yan S, Xu D, Zhang B, Zhang HJ, Yang Q & Lin S 2007, 'Graph embedding and extensions: a general framework for dimensionality reduction', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40-51.

(Yu et al. 2006) Yu CM, Hayes DL & Auricchio A 2006, *Cardiac resynchronization therapy*, Wiley-Blackwell.

Preprint of invited book chapter
Published in: Wu & Sabuncu, eds. Machine learning in medical imaging. Elsevier, 2016

Tables

Tab 10.1: Determination of the main parameters for our method.

Parameter	Estimation method	Concerned set
Number of nearest neighbors	Node flow + Isomap error	Training
Dimensionality of the output space	Node flow + Isomap error	Training
Kernel bandwidth	Distribution of neighborhood sizes + adaptive kernel or multiscale formulation	Testing
Interpolation weight	Generalization ability	Testing
Sample size	Convergence of the results	Training + testing

Preprint of invited book chapter
Published in: Wu & Sabuncu, eds. Machine learning in medical imaging. Elsevier, 2016

Figures legend

Fig 10.1: Overview of the extraction of motion features, leading to the computation of spatiotemporal abnormality maps.

Fig 10.2: Non-linear dimensionality reduction via the Isomap algorithm. Adapted from (Duchateau et al. 2012a).

Fig 10.3: (a) 3D Swiss roll and its associated 2D parameterization. Influence of the number of nearest neighbors K (b) against the Isomap error (c) and the node flow (d), highlighting the apparition of a short-circuit (black arrows). Adapted from (Duchateau et al. 2012a).

Fig 10.4: (a) Node flow distribution against the number of nearest neighbors. Errorbars indicate the median and $1^{st}/3^{rd}$ quartiles over the training set. (b) Isomap error against the dimensionality of the output space and the number of nearest neighbors. White crosses indicate the minimum value of each column. (c) Training set distribution of the neighborhood sizes (average distance between the 5 nearest neighbors). Adapted from (Duchateau et al. 2012a).

Fig 10.5: (a) 2D visualization of the estimated embedding. The bold-framed map corresponds to the perfectly normal pattern used to constrain the manifold. (b) Progressive deviations from normality along the two first principal directions. The abnormal inward/outward motion pattern is preserved. Adapted from (Duchateau et al. 2012a).

Fig 10.6: Steps for comparing a new subject to the modeled pattern and to normality. Adapted from (Duchateau et al. 2012a).

Fig 10.7: (a) Single scale interpolation without/with forcing the curve to pass by a given point. (b) Robustness to local density variations: single- vs. multi-scale interpolation. Adapted from (Duchateau et al. 2013c). With permission of *Springer*.

Fig 10.8: Subjects distribution according to the distances to the modeled pattern and to normality. Dashed lines indicate the normality range defined by the healthy population. Adapted from (Duchateau et al. 2013b). With permission of *Springer*.

Fig 10.1

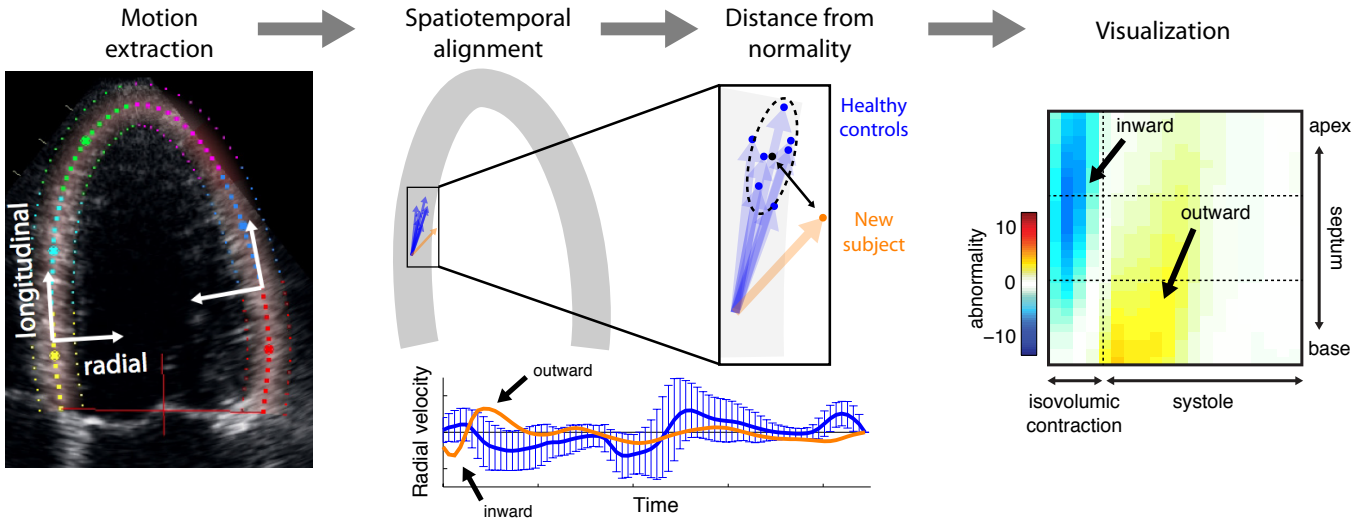

Motion extraction
Spatiotemporal alignment
Distance from normality
Visualization
longitudinal
radial
Healthy controls
New subject
outward
inward
Radial velocity
Time
inward
outward
apex
septum
base
abnormality
10
0
−10
isovolumic contraction
systole

Fig 10.2

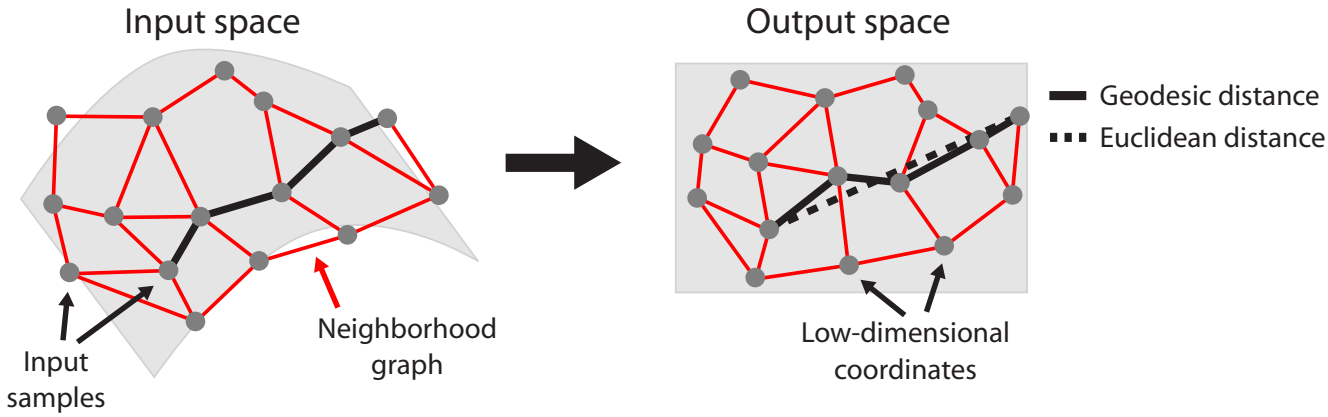

Input space
Output space
Geodesic distance
Euclidean distance
Input
samples
Neighborhood
graph
Low-dimensional
coordinates

Fig 10.3

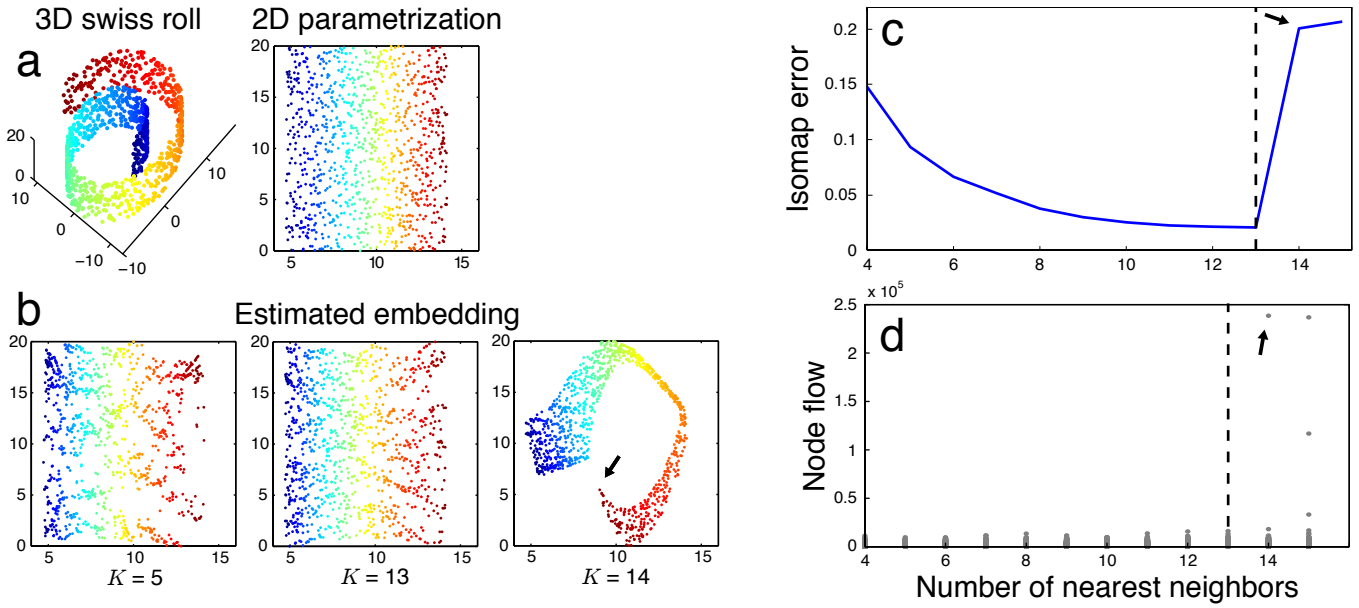

3D swiss roll
2D parametrization
a
b
Estimated embedding
K = 5
K = 13
K = 14
c
Isomap error
d
Node flow
x 10^5
Number of nearest neighbors

Fig 10.4

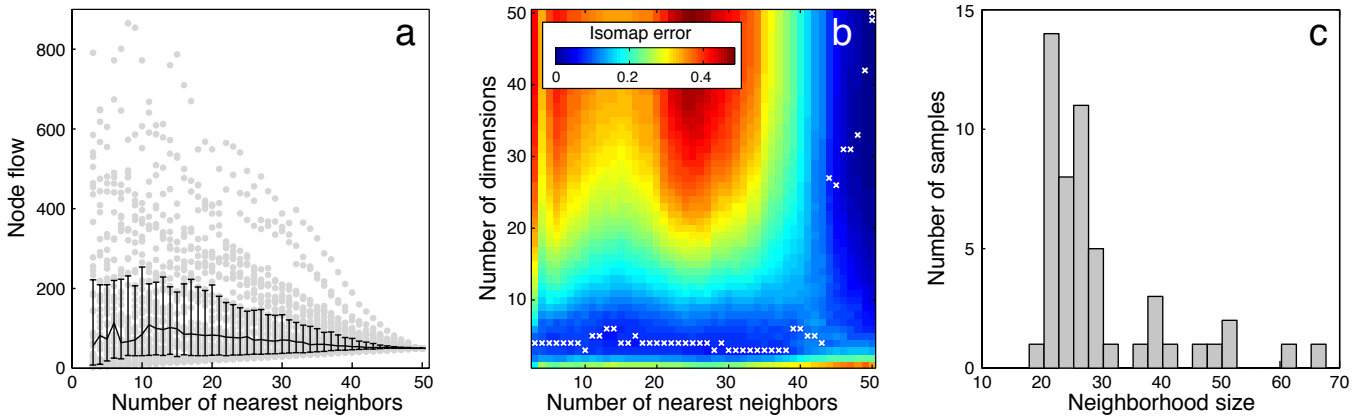

a
Node flow
Number of nearest neighbors
b
Isomap error
0
0.2
0.4
Number of dimensions
Number of nearest neighbors
c
Number of samples
Neighborhood size

Fig 10.5

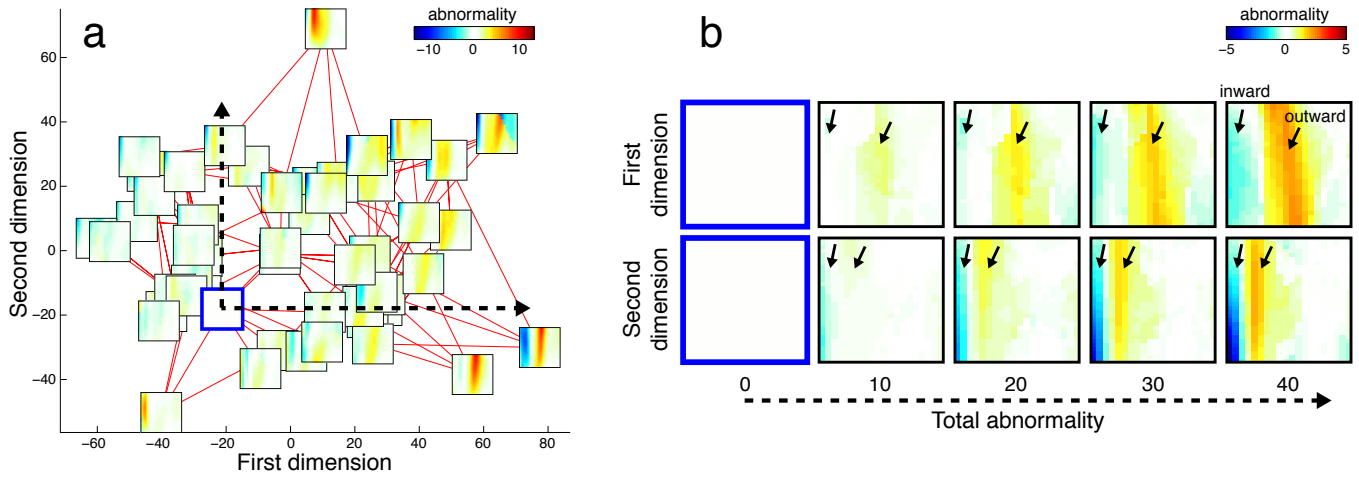

a
abnormality
−10
0
10
Second dimension
First dimension
60
40
20
0
−20
−40
−60
−40
−20
0
20
40
60
80
b
abnormality
−5
0
5
inward
outward
First dimension
Second dimension
0
10
20
30
40
Total abnormality

Fig 10.6

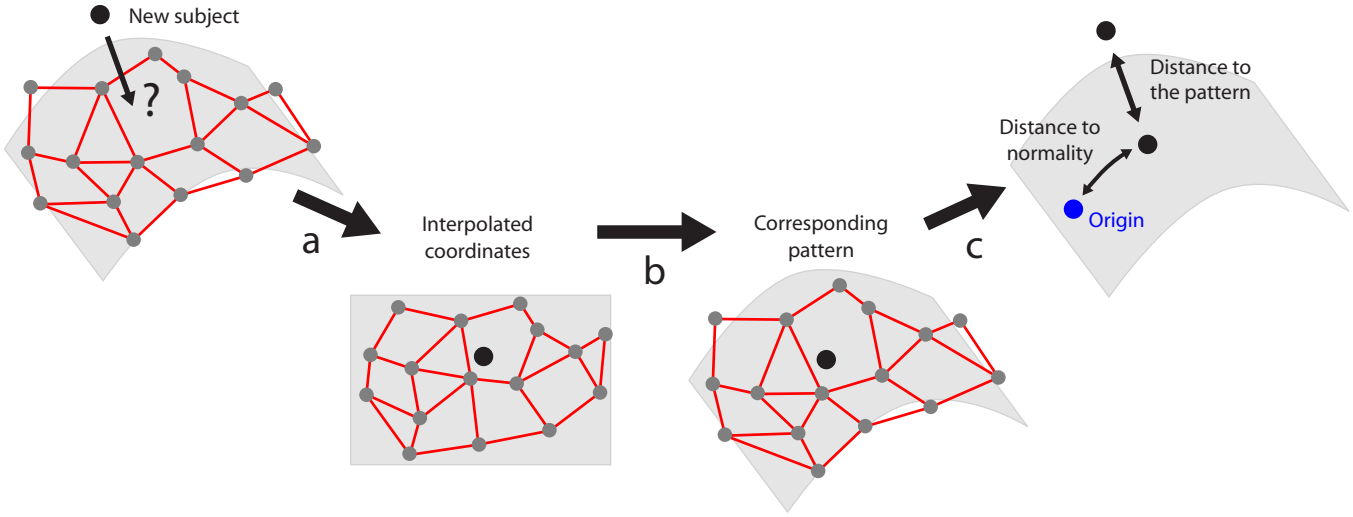

New subject
?
a
Interpolated coordinates
b
Corresponding pattern
c
Distance to the pattern
Distance to normality
Origin

Fig 10.7

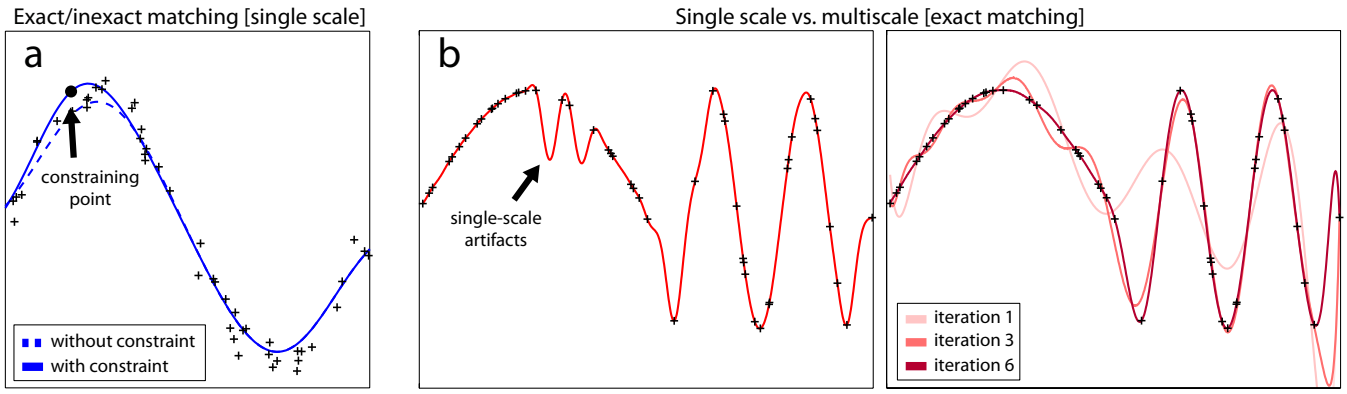
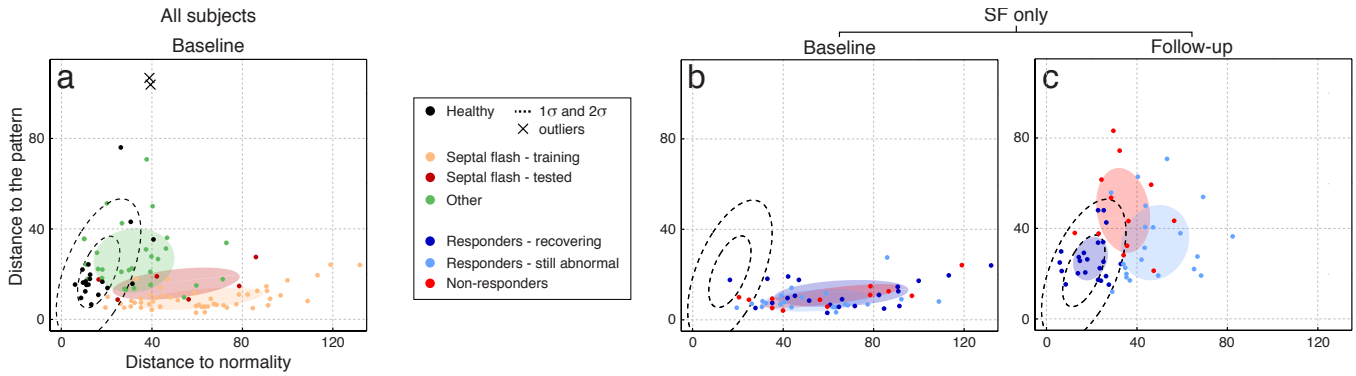

Exact/inexact matching [single scale]
a
constraining point
without constraint
with constraint
Single scale vs. multiscale [exact matching]
b
single-scale artifacts
iteration 1
iteration 3
iteration 6

Fig 10.8


All subjects
Baseline
a
Distance to the pattern
Distance to normality
0
40
80
120
Healthy
1σ and 2σ
outliers
Septal flash - training
Septal flash - tested
Other
Responders - recovering
Responders - still abnormal
Non-responders
SF only
Baseline
Follow-up
b
c