On Feature Extraction for Noninvasive Kernel Estimation of Left Ventricular Chamber Function Indices from Echocardiographic Images

Ricardo Santiago-Mozos, MSc, PhD[a,*], José Luis Rojo-Álvarez, MSc, PhD[a,d], J. Carlos Antoranz, PhD[c], Mar Desco, MD, PhD[c], Daniel Rodríguez-Pérez, PhD[c], Raquel Yotti, MD, PhD[b], Javier Bermejo, MD, PhD[b]

[a]*Department of Signal Theory and Communications, Universidad Rey Juan Carlos, Fuenlabrada, Madrid, Spain*
[b]*Department of Cardiology, Hospital General Universitario Gregorio Marañón and Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain.*
[c]*Department of Mathematical Physics and Fluids, Universidad Nacional de Educación a Distancia, Spain.*
[d]*Prometeo, Department of Electric and Electronic Engineering, Universidad de las Fuerzas Armadas ESPE, Ecuador.*

Abstract

Two reference indices used to characterize left ventricular (LV) global chamber function are end-systolic peak elastance (E_{max}) and the time-constant of relaxation rate (τ). However, these two indices are very difficult to obtain in the clinical setting as they require invasive high-fidelity catheterization procedures. We have previously demonstrated that it is possible to approximate these indices noninvasively by digital processing color-Doppler M-mode (CDMM) images. The aim of the present study was twofold: (1) to study which feature extraction from linearly reduced input spaces yields the most useful information for the prediction of the haemodynamic variables from CDMM images; (2) to verify whether the use of nonlinear versions of those linear methods actually improves the estimation. We

*Corresponding author

Email addresses: ricardo.santiago.mozos@urjc.es (Ricardo Santiago-Mozos, MSc, PhD), joseluis.rojo@urjc.es (José Luis Rojo-Álvarez, MSc, PhD), jcantoranz@dfmf.uned.es (J. Carlos Antoranz, PhD), mmdesco@dfmf.uned.es (Mar Desco, MD, PhD), daniel@dfmf.uned.es (Daniel Rodríguez-Pérez, PhD), ryotti@gmail.com (Raquel Yotti, MD, PhD), javbermejo@jet.es (Javier Bermejo, MD, PhD)

studied the performance and interpretation of different linearly transformed input spaces (raw image, discrete cosine transform (DCT) coefficients, partial least squares, and principal components regression), and we compared whether non linear versions of the above methods provided significant improvement in the estimation quality. Our results showed that very few input features are enough for providing a good (medium) quality estimator for E_{max} (for τ), which can be readily interpreted in terms of the measured flows. Additional covariates should be included to improve the prediction accuracy of both reference indices, but especially in τ models. The use of efficient nonlinear kernel algorithms does improve the estimation quality of LV indices from CDMM images when using DCT input spaces that capture almost all energy.

Keywords: Color Doppler, M-mode Image, Kernel Methods, Support Vector Regression, Partial Least Squares, Principal Component Regression, Left Ventricular Function, Elastance, LV relaxation, Cosine Transform.

---

1. Introduction

Characterization of left ventricular (LV) systolic and diastolic chamber function is still a pending issue in the clinical setting. In experimental physiology, peak end-systolic elastance (E_{max}) is well established as the best available index to measure systolic performance of the LV chamber. In turn, the time-constant of LV relaxation (τ) is accepted as the gold standard method accounting for the rate of relaxation of the chamber, one of the main diastolic properties of LV function. Measuring E_{max} requires complex measurements of instantaneous pressure and volume inside the LV chamber as well as preload intervention maneuvers. Measuring τ requires invasive catheterization of the LV using high-fidelity micromanometers. For these reasons, neither E_{max} nor τ are only measured in patients for research purposes.

A number of noninvasive methods have been developed to obtain surrogate indices that correlate with E_{max} and τ. Among them, most research has focused on Doppler-echocardiography, because it is a fully noninvasive, non ionizing, cheap and readily available at the patient's bedside. In a previous work we have shown that τ and E_{max} can be reasonably approximated from CDMM images. Using a fluid-dynamic approach we have shown that E_{max} correlates closely to the peak-ejection pressure difference developed inside the ventricle, which can be computed by solving Euler's equation from the CDMM velocity data (Yotti et al., 2005). Similarly, τ can be approximated by the peak reverse end-ejection

pressure difference with reasonable accuracy (Yotti et al., 2011). Importantly, using a learning from samples approach we have obtained similar approximations without the need of complex fluid-dynamic modeling (Rojo-Alvarez et al., 2006). Hence, an experimental (animal) setup was used to simultaneously measure the catheter-based curves (pressure and flow) and acquire the CDMM images; and a machine-learning model was designed for straightforward estimation of τ and E_{max} parameters from an input space given by the diastolic period of the digitized CDMM image.

In that precedent work, a linear estimator was used for the image input space, which raises several questions. On the one hand, nonlinear relations between CDMM images and indices can be expected, as the haemodynamic variables in the cardiac circulatory system are known to be mostly interrelated by nonlinear fluid dynamic equations. On the other hand, linear kernel estimators are often suggested in the machine learning literature as the most appropriate choice for high-dimensional input spaces, and they also provide with easier to interpret, black-box models than their nonlinear counterparts. Therefore, our aim was to test whether alternative algorithms on the machine learning specifications could improve the prediction of invasive indices E_{max} and τ from CDMM images. First, we wanted to study which feature extraction from linearly reduced input spaces yields the most useful information for the prediction of the haemodynamic variables from CDMM images. Second, we wanted to verify whether the use of nonlinear algorithmic versions of those linear methods actually improves the estimation. Accordingly, we benchmarked the performance of several linear kernel estimators, in terms of linear feature extraction transformations, and in addition we analyzed the physical and clinical meaning of the relevant features in these transformed spaces, when possible. We also benchmarked the nonlinear kernel versions of the above analyzed estimators, hence determining the actual improvement obtained by the consideration of nonlinearity in the estimation kernel machine. For this purpose, we chose several kernel methods, namely, Support Vector Regression (SVR), Principal Component Regression (PCR), and Partial Least Squares (PLS), according to different levels of algorithm complexity in terms of the multidimensional output estimation from the multidimensional input. SVR performs one dimensional output robust estimation, PCR performs dimensionality reduction and multidimensional output estimation, and PLS performs a dimensionality reduction according input-output covariance.

The rationale for the chosen input features was as follows. First, the RAW input space conveys all the image information, hence it represents a necessary benchmarking. Also, a linear machine working on the input space will be easy

to interpret, in terms of the relative temporal and spatial position of the linear weights. Second, DCT input space is a widespread used frequency transform in image problems, and given the smoothness and low-pass frequency content of CDMM images, it can be expected to work well from an image information compression point of view. Third, PCR provides us with features from an intrinsic image decomposition (different from frequency decompositions), with a decoupled regression stage. And finally, PLS provides us with features from an intrinsic image decomposition in which the regression output quality is an embedded optimization criterion.

The scheme of the paper is as follows. In the next section, the fundamental theory of the multidimensional kernel machines is summarized for the SVR, PCR, and PLS algorithms. Then, a detailed set of experiments is presented for benchmarking and interpretation of linear vs nonlinear kernel versions of the estimators. Finally, conclusions are drawn.

2. Multidimensional Kernel Machines

This section first describes the basic equations of SVR, PCR and PLS. These methods allow both linear and nonlinear estimation without explicitly extracting features from the images. Both PCR and PLS implicitly extract features (components or latent vectors) previous to the estimation problem. PCR performs a feature extraction such every new feature captures as much as possible of the remaining variance of the input data, where PLS extracts features that maximize the covariance of the input data and target variables.

2.1. SVR Estimation

SVR is a well-studied technique that allows nonlinear mappings of the input space and works well with high dimensional spaces like images (Smola and Schölkopf, 2004). Given a training set $\{(\mathbf{x}_i, y_i),\ i = 1, \ldots, n\}$ where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, the SVR finds a function f that estimates y_i as

$$\hat{y}_i = f(\mathbf{x}_i) = \boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b = y_i + e_i \tag{1}$$

where $\boldsymbol{\phi} : \mathbb{R}^d \rightarrow \mathcal{H}$ is in general a nonlinear mapping to the feature space $\mathcal{H}$; $(\cdot)^T$ is the matrix transpose operator; $\mathbf{w}$ is the weight column vector in this space; b is a bias term; and e_i is the residual error. Function f is found by minimizing a functional with a regularization term and a loss term, as follows:

$$L = \frac{1}{2}||\mathbf{w}||^2 + \sum_{i=1}^{n} \mathcal{L}(e_i) \tag{2}$$

where $\mathcal{L}$ is in our case the robust ϵ-Huber loss function (Rojo-Alvarez et al., 2006), which increases the flexibility modeling outliers, given by

$$\mathcal{L}(e_i) = \begin{cases} 0 & |e_i| \leq \epsilon \\ \frac{1}{2\delta}(|e_i| - \epsilon)^2 & \epsilon \leq |e_i| \leq \epsilon + \delta C \\ C(|e_i| - \epsilon) - \frac{1}{2}\delta C^2 & \epsilon + \delta C \leq |e_i| \end{cases} \tag{3}$$

where ϵ is the insensitive-zone parameter (no loss for errors lower than ϵ), and δC controls the size of the quadratic zone of the loss function. Finally, we minimize the following convex problem:

$$L = \frac{1}{2}||\mathbf{w}||^2 + \frac{1}{2\delta}\sum_{i \in I_1}(\xi_i^2 + \xi_i^{*2}) + C\sum_{i \in I_2}(\xi_i + \xi_i^*) - \sum_{i \in I_2}\frac{\delta C^2}{2} \tag{4}$$

with respect to $\mathbf{w}, b, \xi_i, \xi_i^*$, taking into account the following convex constraints:

$$y_i - \boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} - b \leq \epsilon + \xi_i \tag{5}$$

$$\boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b - y_i \leq \epsilon + \xi_i^* \tag{6}$$

$$\xi_i, \xi_i^* \geq 0 \tag{7}$$

for $i = 1, \ldots, n$, and where ξ_i, ξ_i^* are positive slack variables to penalize the positive and negative errors, and I_1 and I_2 are respectively the sets of samples that are in the quadratic and linear loss zone.

Following the same procedure that solves standard SVR (Smola and Schölkopf, 2004), we obtain the following solution:

$$\hat{y}_t = f(\mathbf{x}_t) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}_t) + b \tag{8}$$

where $K(\mathbf{x}_i, \mathbf{x}_t) = \boldsymbol{\phi}(\mathbf{x}_i)^T\boldsymbol{\phi}(\mathbf{x}_t)$ is a Mercer kernel function, which is usually constructed without explicitly projecting in $\mathcal{H}$ (i.e., without explicit knowledge of $\boldsymbol{\phi}$) and $C \geq \alpha_i, \alpha_i^* \geq 0$ are the Lagrange multipliers for the restrictions (5), (6). After optimization, some α_i, α_i^* have a non-zero value and their associated sample is named *support vector* (SV), because it influences function f. Three parameters (C, ϵ, δ) need to be tuned for linear SVR (as well as the kernel width σ for the Gaussian case). Linear kernel is defined as $k_L(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^T\mathbf{x}_2$, and Gaussian kernel is defined as $k_G(\mathbf{x}_1, \mathbf{x}_2) = e^{-\frac{||\mathbf{x}_1 - \mathbf{x}_2||^2}{2\sigma^2}}$.

2.2. PCR Estimation

PCR (Massy, 1965) performs a Principal Component Analysis (PCA) on the zero-mean $n \times d$ predictor matrix $\mathbf{X}$, and then applies Ordinary Least Squares (OLS) to the resulting g principal components (scores) $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_g]$ and the zero-mean $n \times k$ dependent variables matrix $\mathbf{Y}$, where k is the number of dependent variables. $\mathbf{X}$ can be factorized as:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T \tag{9}$$

where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_r]$ is the loadings matrix ($\mathbf{P}\mathbf{P}^T = \mathbf{I}$) and r, is the rank of $\mathbf{X}$. Any number $g \leq r$ of components can be selected in order to estimate $\mathbf{Y}$, however, these components should be selected carefully. Massy (1965) proposes using the variance of the components or the dependence (correlation with $\mathbf{Y}$) depending on the purpose of the analysis, but author warns also that low variance components can be useful for predicting $\mathbf{Y}$. Jolliffe (1982) and Hadi and Ling (1998) discourage the direct application of the high-variance rule for selection components.

Predictions for $\mathbf{X}$ and $\mathbf{X}_t$ are done by using the following matrix relations:

$$\mathbf{Y} = \mathbf{T}\mathbf{B} + \mathbf{F} \tag{10}$$

$$\hat{\mathbf{B}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{Y} \tag{11}$$

$$\hat{\mathbf{Y}} = \mathbf{T}\hat{\mathbf{B}} = \mathbf{X}\mathbf{P}\hat{\mathbf{B}} \tag{12}$$

$$\hat{\mathbf{Y}}_t = \mathbf{X}_t\mathbf{P}\hat{\mathbf{B}} \tag{13}$$

where (10) is the regression model based on the principal components; (11) is the OLS estimator for $\hat{\mathbf{B}}$; and (12) and (13) are the predictions.

As described in (Rosipal et al., 2001b; Rosipal and Trejo, 2001), the nonlinear kernel PCR (KPCR) performs kernel PCA (Schölkopf et al., 1999) and then OLS using the principal components in the feature space. The same concerns appear in KPCR as in PCR, low variance components could be useful for prediction, as shown in (Rosipal et al., 2001a). One parameter (g) needs to be tuned for PCR (and the kernel width for Gaussian KPCR).

2.3. PLS Estimation

PLS for regression (see Wold et al. (2001); Rosipal and Krämer (2006) for a description and an overview) are used for modeling relations between blocks of variables (e.g., a block of d explanatory variables and another block of k response

variables), as well as for dimensionality reduction. PLS extracts orthogonal latent vectors (also called score vectors) by maximizing the covariance between blocks of variables; then PLS projects the observed data to its latent structure and use the latent vectors to perform regression of the response variables. PLS techniques assume that the observed data are generated by a process driven by a small number of latent (not directly observed) components.

PLS models the zero-mean $n \times d$ explanatory (predictor) matrix $\mathbf{X}$ and the zero-mean $n \times k$ dependent variables $\mathbf{Y}$ as:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{14}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} = \mathbf{T}\mathbf{D}\mathbf{Q}^T + (\mathbf{H}\mathbf{Q}^T + \mathbf{F}) = \mathbf{T}\mathbf{C}^T + \mathbf{F}^* \tag{15}$$

where $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_g]$ and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_g]$ contain the extracted latent vectors as columns, and g is the number of latent components extracted; $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_g]$ and $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_g]$ are the loading matrices; and $\mathbf{E}$ and $\mathbf{F}$ represent residual matrices. The second equality in (15) comes from the assumption of $\mathbf{T}$ columns being good predictors of $\mathbf{Y}$ and $\mathbf{U} = \mathbf{T}\mathbf{D} + \mathbf{H}$, where $\mathbf{D}$ is a diagonal matrix.

In order to find score vectors $\mathbf{t}_i$ and $\mathbf{u}_i$, a procedure based on NIPALS algorithm (Wold, 1966) finds $\mathbf{w}_i, \mathbf{c}_i$ vectors such that

$$\text{cov}(\mathbf{t}_i, \mathbf{u}_i)^2 = \text{cov}(\mathbf{X}\mathbf{w}_i, \mathbf{Y}\mathbf{c}_i)^2 = \max_{\substack{||\mathbf{r}||=1 \\ ||\mathbf{s}||=1}} \text{cov}(\mathbf{X}\mathbf{r}, \mathbf{Y}\mathbf{s})^2 \tag{16}$$

where $\text{cov}(\mathbf{t}_i, \mathbf{u}_i)$ is the sample covariance between vector $\mathbf{t}_i$ and vector $\mathbf{u}_i$. The procedure starts by a random selection of $\mathbf{u}_i$ vector (it can be initialized as $\mathbf{u}_i = \mathbf{Y}$ if $k = 1$). Then, the following steps are repeated until convergence:

1.- $\mathbf{w}_i = \mathbf{X}^T\mathbf{u}_i / ||\mathbf{X}^T\mathbf{u}_i||$

2.- $\mathbf{t}_i = \mathbf{X}\mathbf{w}_i$

3.- $\mathbf{c}_i = \mathbf{Y}^T\mathbf{t}_i / ||\mathbf{Y}^T\mathbf{t}_i||$

4.- $\mathbf{u}_i = \mathbf{Y}\mathbf{c}_i$

Here, $\mathbf{w}_i, \mathbf{t}_i, \mathbf{c}_i$ and $\mathbf{u}_i$ can also be found by solving eigenvalue problems (Höskuldsson, 1988) or using other approaches as SIMPLS (de Jong, 1993) that directly finds $\mathbf{w}_i$ vectors. The loadings can be computed as

$$\mathbf{p}_i = \mathbf{X}^T\mathbf{t}_i / (\mathbf{t}_i^T\mathbf{t}_i) \qquad \mathbf{q}_i = \mathbf{Y}^T\mathbf{u}_i / (\mathbf{u}_i^T\mathbf{u}_i)$$

Then, $\mathbf{X}$ and $\mathbf{Y}$ are deflated as follows:

$$\mathbf{X} = \mathbf{X} - \mathbf{t}_i\mathbf{p}_i^T \tag{17}$$

$$\mathbf{Y} = \mathbf{Y} - \mathbf{t}_i\mathbf{t}_i^T\mathbf{Y} / (\mathbf{t}_i^T\mathbf{t}_i) = \mathbf{Y} - \mathbf{t}_i\mathbf{c}_i^T \tag{18}$$

and the algorithm is repeated until the required number of components is reached (see Wold et al. (2001) for stopping criteria).

The estimated $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Y}}_t$ for the training data $\mathbf{X}$ and test data $\mathbf{X}_t$ are respectively given by:

$$\mathbf{B} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{C}^T \tag{19}$$

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{B} = \mathbf{T}\mathbf{C}^T \tag{20}$$

$$\hat{\mathbf{Y}}_t = \mathbf{X}_t\mathbf{B} \tag{21}$$

where $\mathbf{W}$, $\mathbf{P}$ and $\mathbf{C}$ respectively are the matrices with $\mathbf{w}_i$, $\mathbf{p}_i$ and $\mathbf{c}_i$ as columns.

To extend PLS to nonlinear problems, it is common to apply the kernel-trick (Schölkopf and Smola, 2001) to the predictor matrix $\mathbf{X}$ and then perform a linear PLS in the feature space (Rosipal and Trejo, 2001). The NIPALS-based kernel version results:

1.- $\mathbf{t}_i = \mathbf{K}\mathbf{u}_i/||\mathbf{K}\mathbf{u}||$

2.- $\mathbf{c}_i = \mathbf{Y}^T\mathbf{t}_i$

3.- $\mathbf{u}_i = \mathbf{Y}\mathbf{c}_i/||\mathbf{Y}\mathbf{c}_i||$

where $\mathbf{K}$ is the kernel matrix of the training data. Steps 2.- and 3.- can be merged and a kernel might also be applied to the dependent variables.

3. Experiments

The experimental design for data processing is as follows. In *Experiment 1*, linear SVR was built from raw input space (CDMM images without transformations), both for estimating E_{max} and τ. Attention was paid to the free parameters search. Error analysis and model diagnostic is given. Weights are plotted and analyzed from a physiological perspective. In *Experiment 2*, we used DCT coefficients of raw images as input space. Linear SVR was used to analyze the resulting images and the dependence on the retained energy and two normalization schemes were examined. In *Experiment 3*, we analyzed raw images using PCR. This method was one of the simplest alternatives to SVR as it reduces the dimensionality with PCA and then estimates using OLS. We compared the supervised and unsupervised alternatives to select the principal components. In *Experiment 4* raw images were analyzed using PLS. We can see PLS as a dimensionality reduction method guided by the variable to be estimated, followed by a prediction using OLS, and accordingly, we wanted to compare the gain of doing supervised dimensionality reduction as opposed to PCR. In *Experiment 5* we tested whether

PCR and PLS methods, which include an implicit dimensionality reduction, could be benefited by using, as input space, the DCT coefficients of the CDMM images. Finally, in *Experiment 6* we compared linear methods with their nonlinear (Gaussian) counterparts.

To compare the performance of different algorithms, we follow a leave-one-individual-out procedure, which has as many iterations as individuals and is explained as follows: 1) each individual is selected as test in one iteration and is only used to evaluate the estimator capabilities. 2) The test individual is used neither to train the estimator nor to search for its optimal parameters. 3) The rest of individuals (training individuals) are used to search for the optimal parameters of the model by using cross-validation, where each individual is assigned only to a fold. 4) All the training individuals are used to create a model with the best parameters of the previous step. 5) The test individual is used to evaluate that model.

This evaluation procedure, which follows best machine learning practices, provides appropriate error estimates as test individuals are not known for the training procedure and controls overfitting by not mixing individuals in the cross-validation folds, which induces conservative selection of parameters.

We denote the mean of the absolute error (MAE) of the estimation of index Y for individual m as:

$$\bar{E}_m = 1/N_m \sum_{i=1}^{N_m} |\hat{Y}_m[i] - Y_m[i]| \ , \tag{22}$$

where N_m is the number of cases recorded for individual m, $Y_m[i]$ and $\hat{Y}_m[i]$ are respectively the measured and estimated indices. We use as error figure $\bar{E}$, defined as the mean of the error defined by (22) for all individuals.

3.1. Materials

This study used data from an experimental protocol with 20 minipigs, whose left ventricular outflow tract (LVOT) flow velocity profile is similar to normal adult humans. Both CDMM images and invasive tracings (using catheters) were simultaneously acquired for each animal in a high fidelity setup. Animals were anesthetized and their heart conditions were manipulated (drugs and vena cava occlusion) to get different values for E_{max} and τ. All measures where averaged by performing the experiment three times on each pig. For E_{max} estimation, 9 pigs where considered that jointly sum up to 274 beats. For τ estimation, 20 pigs were considered that jointly sum up to a total of 1362 beats.

We detail further the images. CDMM images (DICOM format) were obtained from an epicardial approach using a phase-array broadband 2.0–4.0 MHz transducer on a Sequoia C-256 system (Siemens AG) (Yotti et al., 2005). These CDMM tracings display the 1-dimensional spatiotemporal map of ejection flow velocity along the long axis of the LV. Velocity values were obtained from the color values of each pixel using a previously validated decoding and de-aliasing algorithm (Bermejo et al., 2001). Spatial and temporal calibration was obtained from the DICOM metadata, while velocity calibration was obtained by reading the images scale limits. The beginning and end of ejection, as well as the positions of the LV apex and outflow tract, were visually identified in each image. CDMM recordings were then aligned and cropped slightly beyond these limits, interpolated using a bivariate tensor product spline, and then downsampled to a grid of 32×32 pixels. The details of the experimental setup have been reported in (Yotti et al., 2005).

We considered this data in an animal per animal basis, i.e., we always considered together the beats of an animal whether it is used for training or for testing purposes. Our data for each animal consist in the set of its measurements for each beat, namely, CDMM image and invasively measured values (E_{max} and τ).

3.2. Experiment 1: Linear SVR on Raw Images Input Space

In this experiment, we analyzed the raw 32×32 CDMM images, in a leave-one-individual-out scheme, with linear SVR. All the images were standardized by dividing them by the constant that made its maximum absolute value one.

A relevant matter is parameter search. We followed a sequential grid search approach where: (1) a parameter was optimized in a grid while the remaining parameters were kept static; (2) the parameter was set to the value providing minimum error; (3) the next parameter was considered. The search finished when no parameter changed after a round. The SVR parameters were sampled in the following ranges: $\nu \in [0.05, 1]$, $C \in [e^{-4}, e^{4}]$ and $\delta \in [10^{-2}, 1]/C \times \max(|E_{max}|)$. Given that it can be shown that δC is the length of the quadratic segment of the loss function (3), $\delta C \in [10^{-2}, 1] \times \max(|E_{max}|)$. We defined $\delta^* = \delta C / \max(|E_{max}|)$, which varies between 0.01 and 1, and relates to the length of the quadratic zone of (3) with the maximum of the target variable. As E_{max} is standardized to be zero-mean and unit-variance previously to create the model, this selection of δC provides a quadratic penalty in the loss function that can be selected to be as small $0.01\max(|E_{max}|)$ or as high as $\max(|E_{max}|)$. Larger lengths for the quadratic zone are not considered as we do not want quadratic loss for large residuals that are even clearly outliers or samples that can not be modeled.

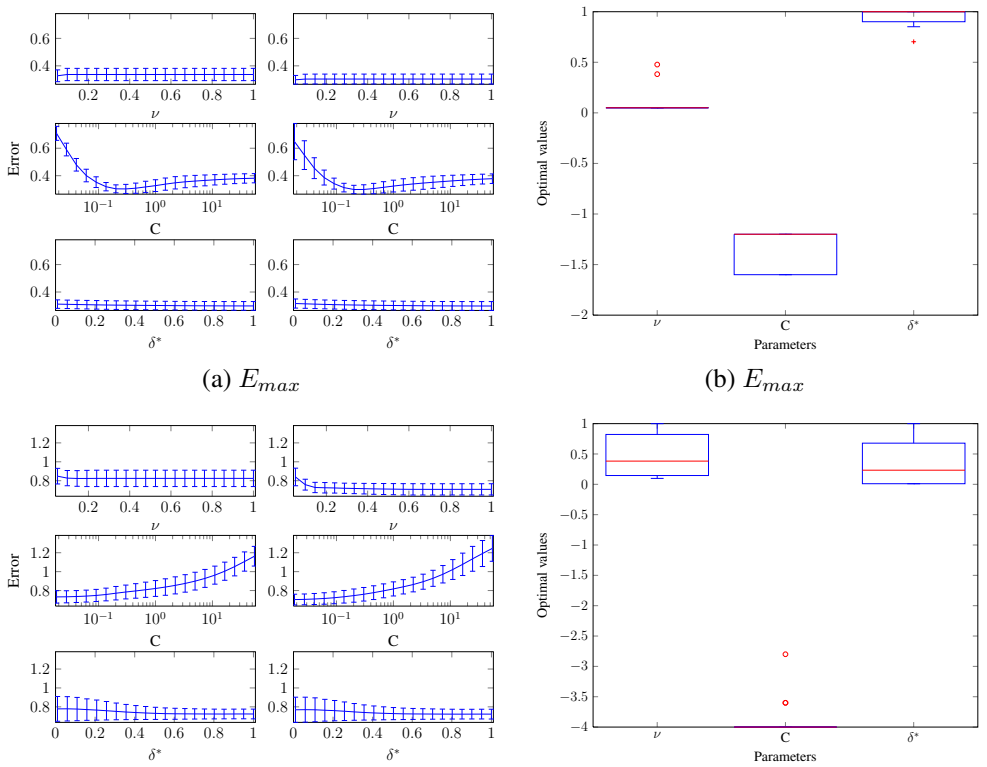
We started with the prediction of E_{max} by using the raw CDMM images. The average result (for all individuals) of the linear search of SVR parameters is shown in Figure 1, which also shows the boxplot of the best values for each individual. Parameter ν seemed to have little impact on the accuracy, and only C and δ^* seemed to have some effect. However, when looking at the optimal parameters, ν was close to zero for the prediction of 7 out of nine animals, C had a little variation, and δ^* was close to 1 for almost all cases, which informs that almost all residuals were penalized in the quadratic zone of the loss function.

The prediction and the residuals are shown in Figure 2, where we observe that the SVR was accurately following E_{max}. The residuals were not Gaussian, as we see heavy tails in the histogram. The Bland-Altman (BA) plot compares the true value, i.e. the gold standard, to the difference between the true and estimated values, with the aim of detecting systematic errors, biases, heteroscedasticity or outliers. In the BA plot for E_{max} we observe that SVR trended to underestimate values above 9 mmHg/ml, which shows a mixture of heteroscedasticity and bias. The averaged absolute error and its standard deviation is shown in Table 1. SVR-LIN was the best linear predictor for raw images. E_{max} had a mean of 4.2 mmHg/ml and a standard deviation of 3.1 mmHg/ml. The average estimation error was about a 20% (30%) of the mean (standard deviation) of E_{max}, and the maximum absolute error was 5.4 mmHg/ml.

With linear kernel, predictor f (1) becomes a linear combination of the training images, and its absolute value for each image pixel represents the importance of that pixel for the prediction of the haemodynamical variable. The regression weights, represented as images, for each test case are shown in Figure 3(a), and their mean and standard deviation in Figure 4(a). A strong agreement on these weights is found in most machines, where we see that the center and lower center of the image is contrasted with its surroundings, hence, the predictor contrasted the region near the apex and near the LVOT at the beginning and near the end of the normalized ejection time with the region near the LVOT in the middle of the normalized ejection time.

For τ prediction, we used the same ranges for the parameters as for E_{max} prediction. We show the average result of the linear search of SVR parameters and the boxplot of the best values in Figure 1. In this case, the boxplot shows than both ν and δ^* varied dramatically from predicting each case, and on the other hand, C was almost the same for all cases. The three parameters seemed to have impact in the accuracy, but variations of ν from 0.2 to 1 had almost no effect.

The regression weights for τ, represented as images, are shown in Figure 3(b) and their mean and standard deviation in Figure 4(b). There were several patterns



(a) E_{max} (b) E_{max}

(c) τ (d) τ

Figure 1: SVR-LIN on raw images. (a) and (c) show the averages of parameter searches. (b) and (d) show the boxplot with optimal parameters for each animal: $\nu \in [0.05, 1]$, $C \in [\exp(-4), \exp(4)]$ (represented in natural log scale) and normalized $\delta^* \in [1e{-}2, 1]$. First (second) row represents E_{max} (τ) results.

in the regression weights that depended at least on the selected training parameters: patterns like the ones for predicting animals 1, 2, 9, 12, 13, 14, 17, 19 resulted from training with $\nu < 0.15$ (most samples do not penalize) and $\delta^* > 0.5$ values (the quadratic zone of the loss function is large); patterns like 3, 6, 7, 10, 11, 15, 18, 20 resulted for training with $\nu > 0.5$ (most samples penalize) and $\delta^* < 0.06$ values (the quadratic zone of the loss function is very small so almost all the residuals were penalized with a linear cost). When looking at the coefficients themselves, we see a great agreement just above the lower-left corner where the maximum of the coefficients are for all images. There is also quite agreement on the zones with the larger negative coefficients. Both observations are observed in Figure 4. The prediction and the residuals are shown in Figure 2. We observe

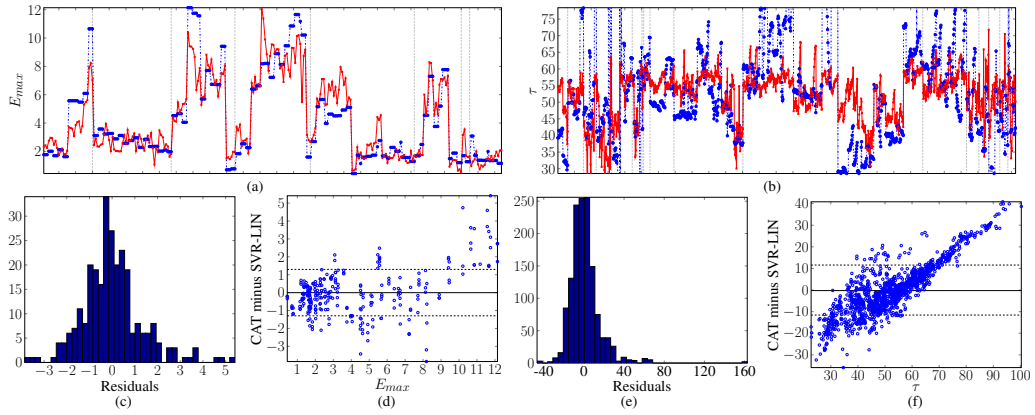
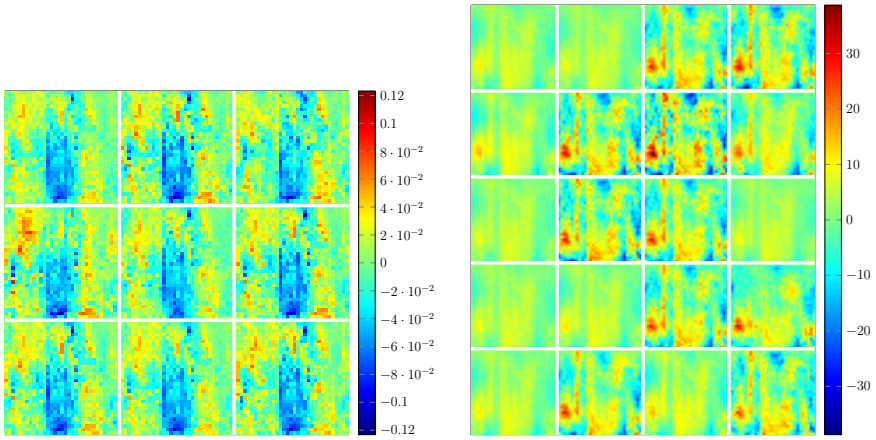


Figure 2: SVR-LIN on raw images. (a) and (b) are the predictions for E_{max} and τ respectively; catheter measurement is plotted in blue (dot-line) and prediction in red (continuous), vertical dotted lines divide the individuals. (c) and (d), and (e) and (f) are the residuals and Bland-Altman plots respectively for E_{max} and τ.

	E_{max} mmHg/ml		τ ms	
Method	$\bar{E}$	$\pm$ desv	$\bar{E}$	$\pm$ desv
SVR-LIN	**0.919**	$\pm$ **0.32**	**9.79**	$\pm$ **4.8**
PCR	0.947	$\pm$ 0.30	11.0	$\pm$ 5.6
SPCR	0.968	$\pm$ 0.32	11.8	$\pm$ 5.1
PLS	0.927	$\pm$ 0.32	10.7	$\pm$ 5.3

Table 1: Comparison of Experiment 1, 3 and 4. Estimation error and standard deviation. RAW CDMM images.

that the estimation provided by the SVR-LIN for τ was less accurate than for E_{max}. In addition, the BA plot shows an almost-linear behavior in the difference of prediction and gold standard for high τ values. This systematic error dependent of the value of τ (that could be modeled by a line and partially removed) demonstrates the inability of the model to follow high τ values and the need to include other covariates. The average absolute error and its standard deviation is shown in Table 1. SVR-LIN was again the best linear predictor for raw images. Measured τ had a mean of 53.2 ms and a standard deviation of 16.6 ms. The average estimation error was about a 18% (59%) of the mean (standard deviation) of τ, and the maximum absolute error was 162.6 ms which was an outlier as we can see in Figure 2, the next maximum error was just above 60 ms.



(a) E_{max} (b) τ

Figure 3: SVR-LIN on raw images. (a) Regression coefficients to predict E_{max} for individuals 1–9. (b) Regression coefficients to predict τ for individuals 1–20.

3.3. Experiment 2: Linear SVR on DCT Input Space

We also tested the estimation of E_{max} and τ using the SVR with the Discrete Cosine Transform (DCT) coefficients instead of the raw images. DCT coefficients represent the frequency content of CDMM images, which resulted smooth and could be efficiently represented with few coefficients, e.g., $\{70, 75, 80, 85, 90, 95\}\%$ of the energy was respectively retained with only $\{4, 6, 9, 15, 37, 176\}$ out of 1024 coefficients in the case of E_{max} (9 individuals), and with $\{4, 5, 7, 11, 23, 110\}$ out of 1024 coefficients in the case of τ (20 individuals). To obtain the coefficients needed to preserve an energy amount, we first calculated the DCT transform of each image. Then, we obtained the energy in the DCT domain and accumulated the energy content of all images in frequency domain. Finally, we sorted the frequencies in descending energy-content order and kept the minimum number of coefficients that provided the required amount of energy.

Figure 5 shows a contour plot with the cumulative energy content of the DCT coefficients in decreasing order. We observe that energy-content was located in the low frequencies in most images. In addition, it shows the absolute value of DCT coefficients in log scale. We observe that there were 4 orders of magnitude (e^{-9}) among maximum and minimum coefficient.

DCT coefficients were standardized (zero-mean, unit-variance) as a preprocessing step to the SVR. Alternatively, it was also considered to subtract the mean but not changing the variance. The first approach is inspired in typical preprocessing of independent variables, which is common practice in machine learning and

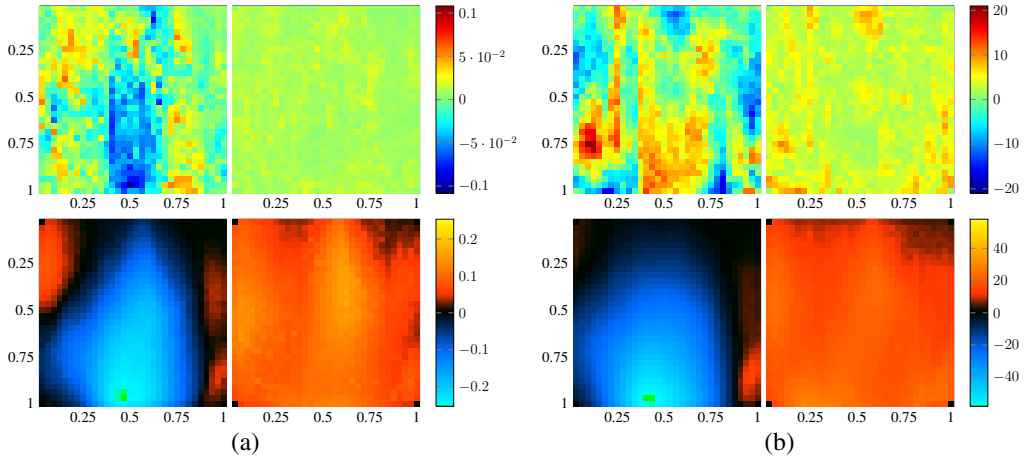


Figure 4: SVR-LIN on raw images. For E_{max} (left panel) the first row shows mean (left) and standard deviation (right) of regression coefficients predicting each test case; and the second row shows mean (left) and standard deviation (right) of all CDMM images. For τ (right panel) the structure is repeated. The images are represented in normalized axis: the y-axis represents the distance to the apex and the x-axis represents the normalized time from diastole.

statistics. The second approach implicitly makes the assumption that information is related to variance, thus avoiding emphasizing low variance components, which might be regarded as noise. Both approaches are compared below. A closer analysis of the DCT coefficients shown in Figure 5 in the E_{max} case reveals, on the one hand, that only 22 (152) DCT coefficients (frequencies) absolute means were above the maximum absolute mean divided by 100 (1000); on the other hand, only 40 (282) frequencies variance were above the maximum variance divided by 100 (1000). Considering both facts together, to use a zero-mean unit-variance standardization has the drawback of amplifying the importance of many frequencies with little contribution on the data, which can lead to poor modeling, especially in the case of small datasets. Table 2 shows this behavior with large errors when more than 176 components were considered. By subtracting only the mean, low energy components are not emphasized, and to model with all components does not penalize the results. The problem of the later approach is that higher-variance components have more influence in the model, but variance does not necessarily mean information for the estimation at hand.

For E_{max} we focused on the DCT coefficients that capture 75% of the energy (this figure provided the best prediction for the zero-mean unit-variance standardization). The optimum parameters of the search of SVR parameters (see Figure 6)

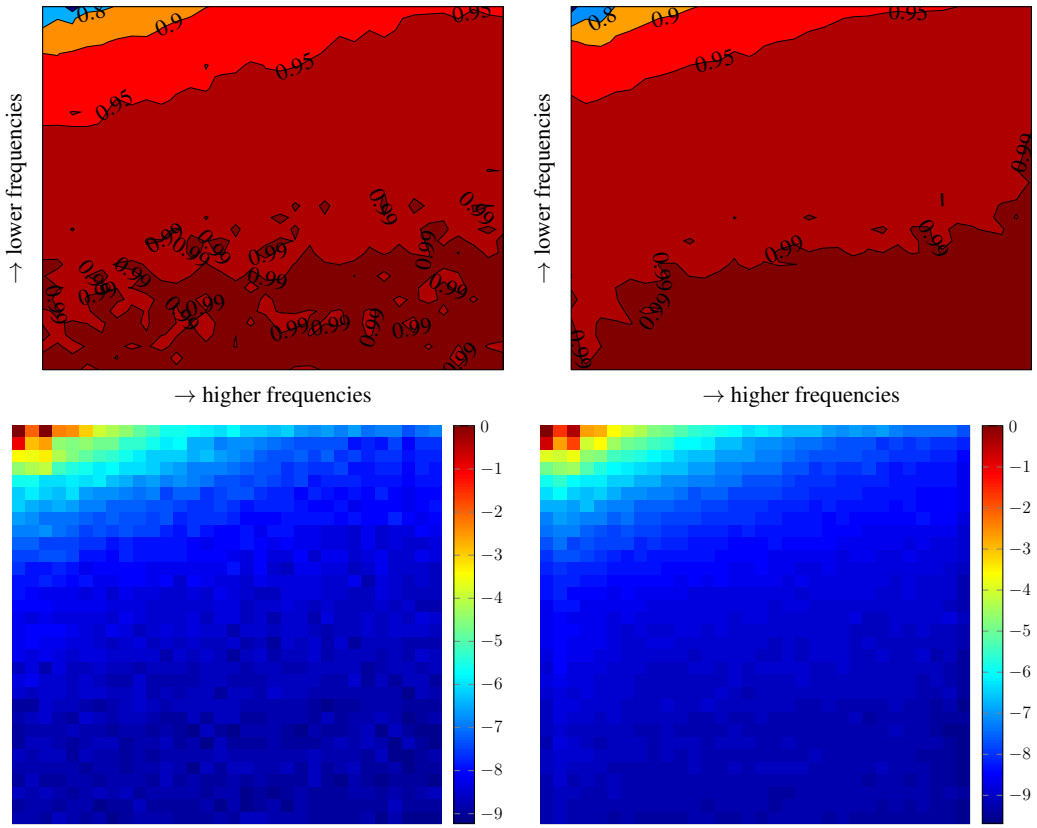
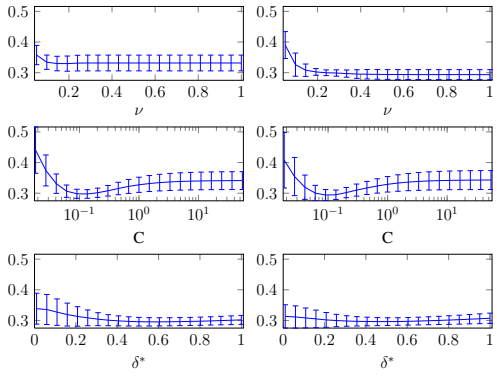
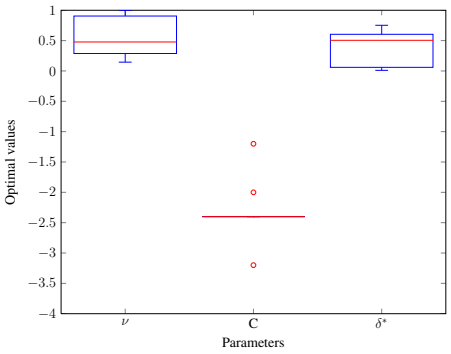


Figure 5: First row present DCT coefficients cumulative energy. Emax (left). $\{70, 75, 80, 85, 90, 95\}\%$ of the energy was respectively retained with $\{4, 6, 9, 15, 37, 176\}$ out of 1024 coefficients. For τ (right) the same energy percentages were retained with $\{4, 5, 7, 11, 23, 110\}$ out of 1024 coefficients. Second row represent the absolute value of DCT coefficients in natural log scale for E_{max} (left) and τ (right).

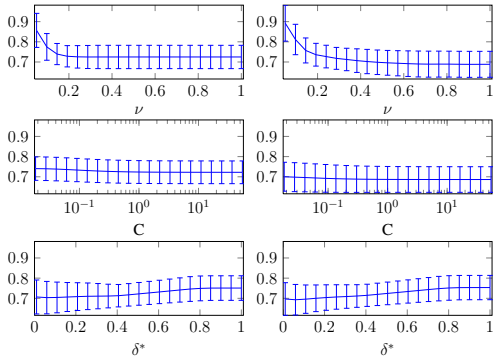
were different than the ones obtained using the raw images, as the input space has now only 6 dimensions. We found, however, some similarities, described below. In the parameter search for E_{max}, ν showed an approximately flat error from 0.2 to 1 (it was almost flat from 0 to 1 in Figure 1) and the most chosen value was approximately 0.5 (it was close to 0 in Figure 1); C showed a minimum around 10^{-1} (same shape that in Figure 1) and its most chosen optimum value was close to $e^{-2.4} \approx 0.09$ (lower than in Figure 1); δ^* seemed to have more effect than in Figure 1 and presented a minimum around 0.5, which was also the value most selected as optimum (compare with 1, which was most selected in Figure 1). Therefore, the models for DCT with 75% of the energy had in general more sup-



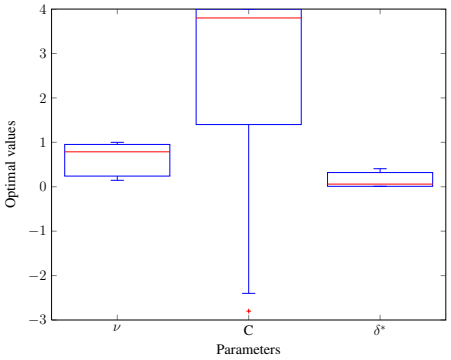
(a) E_{max}



(b) E_{max}



(c) τ



(d) τ

Figure 6: SVR-LIN on DCT coefficients. (a) and (c) show the averages of parameter searches. (b) and (d) show the boxplot with optimal parameters for each animal: $\nu \in [0.05, 1]$, $C \in [\exp(-4), \exp(4)]$ (represented in natural log scale) and normalized $\delta^* \in [1e{-}2, 1]$. First (second) row represents the results for E_{max} retaining 75% (τ retaining 85%) of energy.

port vectors, i.e., were more complex (higher ν), had a shorter quadratic zone in the loss function (lower δ^*) and less penalty to errors (lower C) than the models used for raw images.

The mean for all individuals of the regression weights to be applied to the selected 6 DCT frequencies are shown as images on Figure 7. In addition, to have a better understanding of the predictor, we show the mean for all individuals of the regression coefficients to be applied to the original images. These coefficients were obtained from combining (with the SVR α weights) the images calculated as the IDCT (inverse DCT) of the standardized DCT coefficients of the samples that became support vectors. The prediction is (roughly) made by adding the

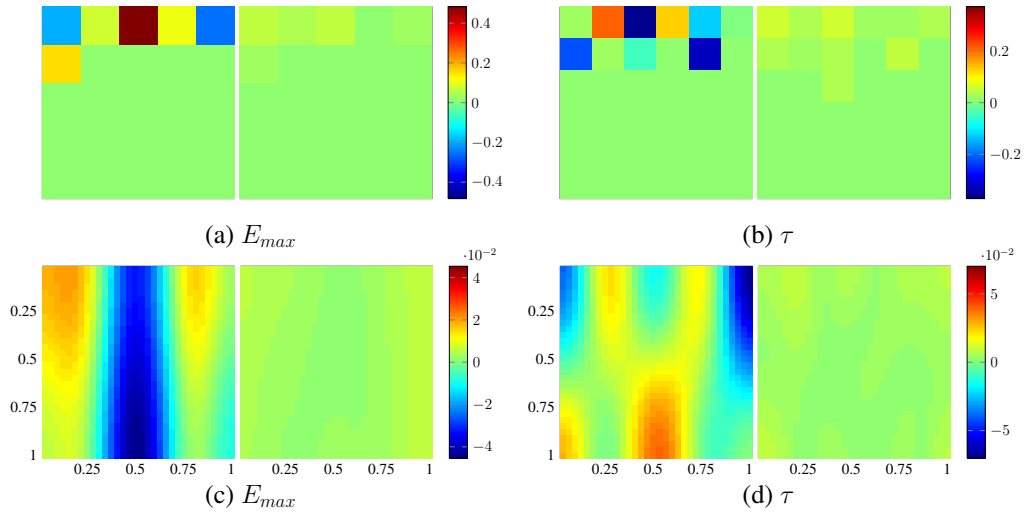


Figure 7: SVR-LIN + DCT: we show in (a) and (b) a zoom on the non-zero DCT coefficients. Both for E_{max} and τ we show the mean (left) and standard deviation of (right) of regression coefficients for prediction of each individual. (c) and (d) show the same predictors mean and standard deviation for both E_{max} and τ in image space. The predictors for E_{max} and τ retain respectively 75% and 85% of the energy.

beginning and the end of the ejection and subtracting mid-ejection. The prediction and the residuals (not shown) had a great agreement with the obtained from the raw images, yielding a small increase in the maximum error.

For τ we focused on the DCT coefficients that captured 85% of the energy (again, the one that provided the best prediction for the zero-mean unit-variance standardization). Parameters ν and δ^* (see Figure 6) seemed to be the more influential ones in the search process, which explain the large variation of the optimal values selected for C. Models using DCT coefficients with 85% of energy had higher ν (more support vectors), higher C (more penalty to errors) and lower δ^* (smaller quadratic zone) than the models used for raw images.

The mean for all individuals of the regression weights to be applied to the selected 11 DCT coefficients are shown in Figure 7, where we also show the regression coefficients to be applied to the original images. The pattern, a linear combination of low frequency DCT image basis components, is a low pass approximation to the one shown in Figure 4. The prediction error of the DCT predictor is slightly higher than the one of the raw predictor, but DCT predictor is easier to interpret. We observe that the predictor for τ roughly contrasted the flow at the beginning and end of normalized ejection time near the apex with the flow

at the beginning of ejection time near the LVOT, the mid-ejection time towards the LVOT, and the 1/4 and 3/4 of ejection time towards the apex. The prediction and the residuals were similar to the ones resulted when using linear SVR applied to raw images (see Figure 2).

Table 2 compares SVR-LIN on raw images vs SVR-LIN on DCT coefficients (with both standardization schemes discussed above). SVR-LIN on the raw images performed the best both for predicting E_{max} and τ. However, DCT1 (zero-mean, unit-variance normalization) seemed to be very efficient, as its error was close to the one attained using raw images for 75% (85% for τ) of the energy using only 6 (11 for τ) coefficients instead of 1024 used by raw images. On the other hand, by using DCT2 (zero-mean normalization), we also see that low energy components can increase the accuracy, as in the case of E_{max}.

3.4. Experiment 3: PCR on Raw Images Input Space

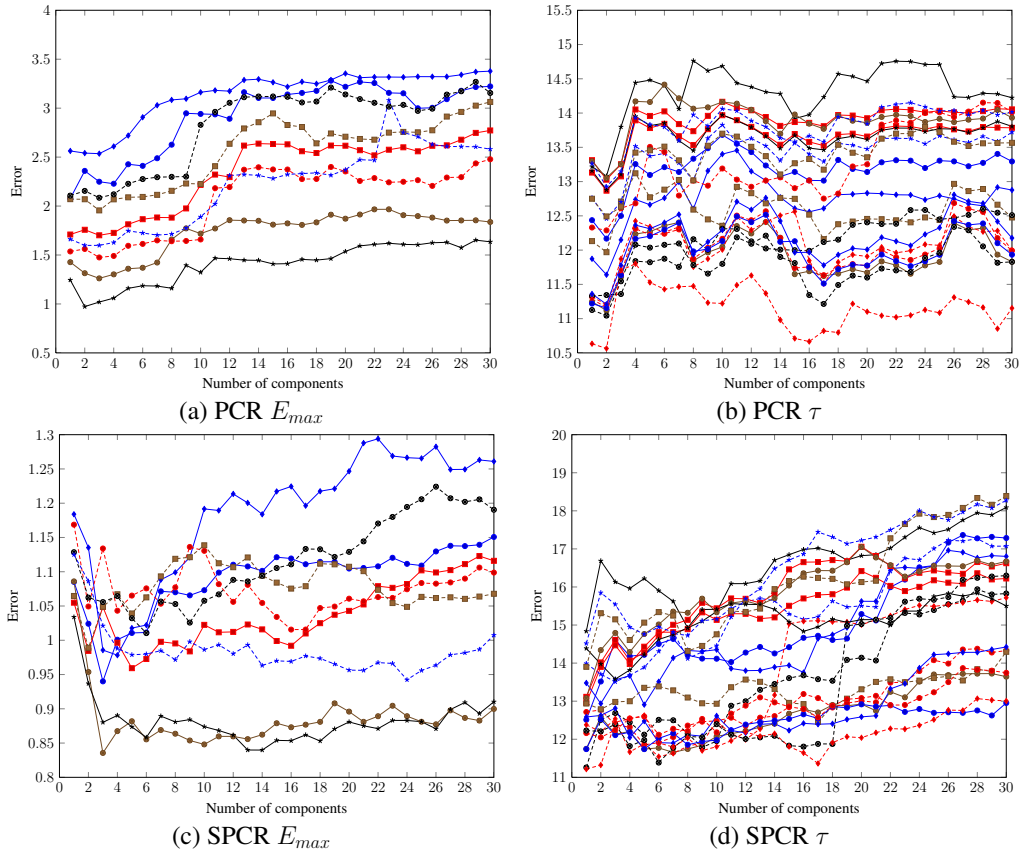
In this experiment, we estimated the reference indices from raw CDMM images with PCR and SPCR using the same methodology of previous sections. No preprocessing was applied to the raw CDMM image in order to make results comparable to the ones obtained by SVR (dividing the images by a constant makes no difference for PCR). We started by considering the prediction of E_{max} with PCR. The search of the optimum number of components is shown in Figure 8. In all cases, the optimum number of components, from the minimum error criterion, was equal to or lower than 3, which was the mode. This was a considerable reduction of dimensionality, as we end adjusting a model with 3 variables (instead of 1024). The principal components are shown in Figure 9. The first one contrasts the normalized mid-ejection time with the beginning of the ejection time region from the middle towards the apex; the second one contrasts two instants around the mid-ejection time; and the third one contrasts the region that surrounds (left, down, right) the middle of the image with the rest of it, especially with the end of the ejection time. The mean of the regression coefficients of the predictors for each individual are shown in Figure 9. We observe that the predictor was dominated by the first principal component. In addition, the predictor had some similarities with the Figure 4 of SVR on raw predictions, both cases had a clear negative zone on the mid-ejection time, and a clear positive zone in the upper left of the figure. In addition, both predictors contrasted the middle of the figure with its surroundings. On the other hand, PCR predictor was smoother than the SVR predictor, which makes it easier to interpret. This predictor was also similar to the obtained from DCT shown in Figure 7, and their performance was almost the same.

		E_{max} mmHg/ml			τ ms		
Method	Energy (%)	#Coef	$\bar{E}$	$\pm$ desv	#Coef	$\bar{E}$	$\pm$ desv
RAW	100	1024	**0.919**	**0.32**	1024	**9.79**	**4.8**
DCT1	70	4	1.13	0.50	4	11.0	5.3
DCT1	75	6	**0.949**	**0.32**	5	11.1	5.6
DCT1	80	9	0.994	0.37	7	11.5	6.1
DCT1	85	15	1.05	0.35	11	**9.91**	**4.0**
DCT1	90	37	0.958	0.32	23	10.1	4.5
DCT1	95	176	1.10	0.52	110	10.5	4.6
DCT1	99	760	1.37	0.80	698	12.6	6.0
DCT1	99.5	883	1.36	0.84	849	12.9	6.0
DCT1	99.9	993	1.31	0.87	985	13.1	5.7
DCT2	70	4	1.10	0.52	4	11.1	5.4
DCT2	75	6	0.939	0.34	5	11.1	5.5
DCT2	80	9	0.967	0.37	7	11.1	5.7
DCT2	85	15	0.974	0.35	11	10.4	4.5
DCT2	90	37	0.925	0.32	23	10.1	4.8
DCT2	95	176	0.925	0.32	110	**10.0**	**4.9**
DCT2	99	760	0.923	0.32	698	10.0	4.9
DCT2	99.5	883	**0.922**	**0.32**	849	10.0	4.9
DCT2	99.9	993	0.924	0.32	985	10.1	4.9

Table 2: Experiment 2. SVR-LIN. Raw is CDMM images without transformations. DCT1 stands for zero-mean unit-variance DCT coefficients normalization and DCT2 for zero-mean untouched-variance DCT coefficients normalization.

Finally the prediction, residuals and BA plot (not shown) were similar to the ones obtained with SVR (Figure 2), but PCR exhibited a heavier right tail in the residuals histogram.

Another possibility to select the principal components uses the correlation with the variable to estimate (SPCR). Following this approach, we show the search for best number of components in Figure 8(c). Here we observe that the selected number of components ranged from 2 to 24, being higher than 10 in 3 out of nine cases. Therefore, we found the optimal number of components highly dependent with the individual that was left-out, which shows the sensitivity of SPCR and its need for more individuals to become more stable. The mean of principal compo-

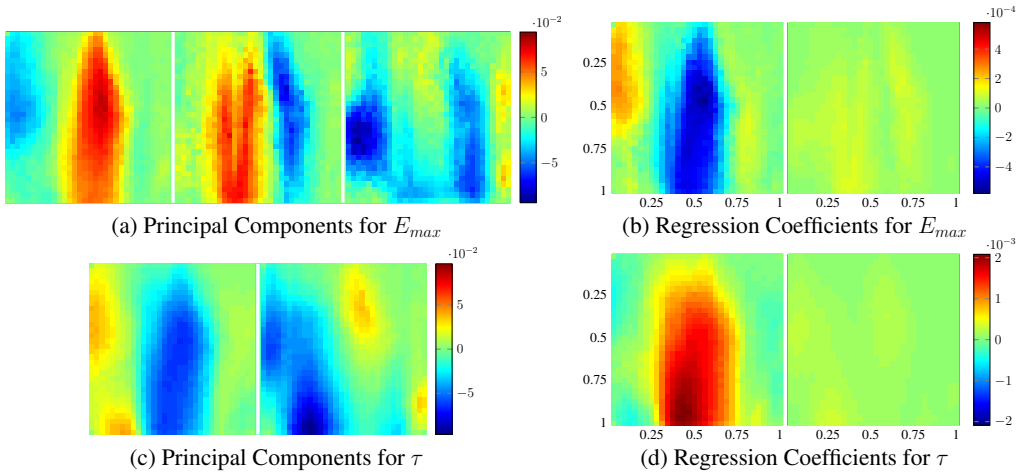


(a) PCR E_{max} (b) PCR τ

(c) SPCR E_{max} (d) SPCR τ

Figure 8: PCR on raw CDMM images: Each line is the search result for optimal number of components during the training of the model for each test individual. (a) and (b) show the results for PCR. (c) and (d) show the results for E_{max}.

nents extracted from the data to train each individual predictor are shown in Figure 10. We see that, in most cases, the first two principal components were similar to the ones obtained by PCR. The rest of components were highly variable from one predictor to another, except the third component, which was similar in most cases (note its variance increase with respect to the other two components). The mean of regression coefficients are shown Figure 10, where we see that predictors changed significantly from one individual to another (note the high variance). The mean of the regression coefficients had some similarity to the mean of SVR-LIN predictors shown in Figure 4. Finally, the prediction, residuals and BA plot (not shown) were similar to the ones given by PCR, the left tail of the SPCR residuals histogram was heavier and reached -4 mmHg/ml; the slight increase in error of



(a) Principal Components for E_{max}

(b) Regression Coefficients for E_{max}

(c) Principal Components for τ

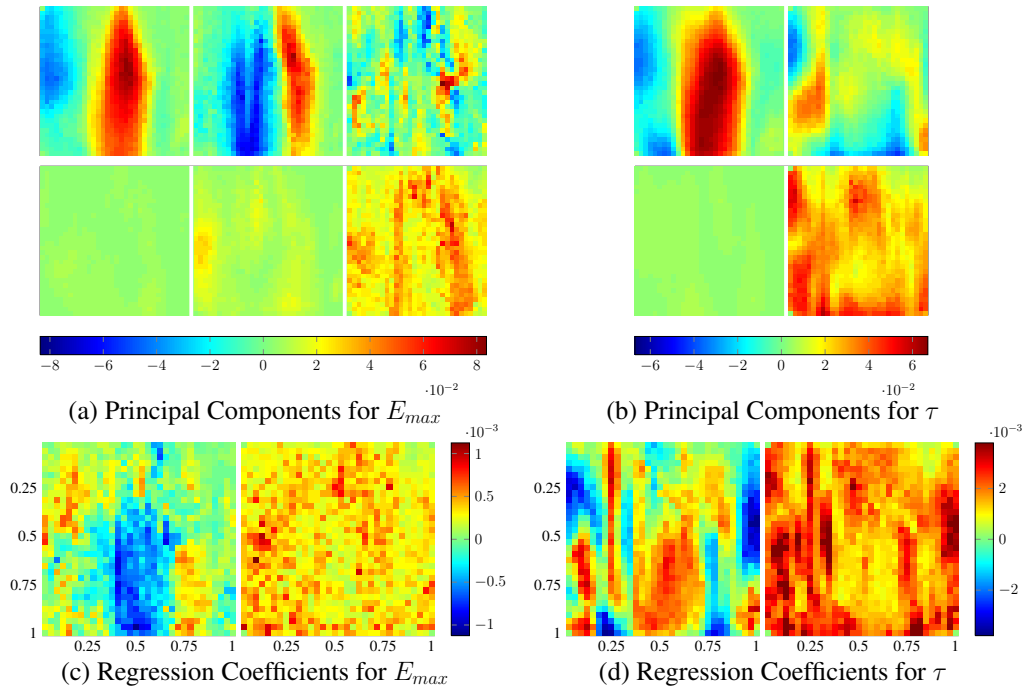
(d) Regression Coefficients for τ

Figure 9: PCR on raw CDMM images: (a) and (c) are the principal components for E_{max} and τ. (c) and (d) are the mean and standard deviation of the individual predictors for E_{max} and τ.

SPCR when compared to PCR was due to over-estimations.

Later, we considered the prediction of τ with PCR. We show the search for the optimum number of components in Figure 8; we found the optimum number of components equals to 2 for all but one of the individuals' predictors, which equals to one. The first 2 principal components are shown in Figure 9. The first one was similar to the first component that resulted in the E_{max} analysis; the second component was different and roughly contrasted the regions above and below the main diagonal, that is, the first part of the ejection time towards the LVOT is contrasted with the second part of the ejection time towards the apex. The mean of the regression components for all individuals are shown in Figure 9; there was a strong agreement for all predictors and they were roughly a fatter version of the first component. This predictor was different to both the resulting from DCT+SVR-LIN and RAW+SVR-LIN. Finally, the prediction, residuals and BA plot (not shown) were similar to previous cases, the residuals had heavier tails than DCT+SVR-LIN and RAW+SVR-LIN.

We also considered the results of predicting τ with SPCR. The search for optimum coefficients is shown in Figure 8. There was again a great variance in the optimum number of components that ranges from 1 to 17, being above 5 for 6 out 20 predictors and 1 the mode. The mean of the principal components extracted from the data to train each individual predictor are shown in Figure 10, where we observe that only the first component was uniformly the common factor (note the high variance of the second component). The mean of these predictors



(a) Principal Components for E_{max}

(b) Principal Components for τ

(c) Regression Coefficients for E_{max}

(d) Regression Coefficients for τ

Figure 10: SPCR on raw CDMM images: (a) and (b) are the principal components for E_{max} and τ, first row is the mean for all the individuals and second row standard deviation. (c) are the mean (left) and standard deviation (right) of the individual's predictors for E_{max}, which (d) represents for τ.

is shown in Figure 10, which was similar to the one provided by RAW+SVR-LIN (see Figure 4). The prediction, residuals, and BA plot (not shown) were similar to the previous ones, but the prediction was worse than previous approaches, having heavier residual tails.

A summary table for error comparison among methods is shown in Table 1, where we observe that both PCR and SPCR accuracies are worse than SVR-LIN. On the other hand, PCR provides easier to interpret predictions. SPCR performs worse than PCR, which we explore below. Table 3 shows a comparison of the mean absolute error on each individual provided by PCR and SPCR for E_{max}. SPCR gave better accuracy on 5 out of 9 individuals, in one of these cases SPCR selected a 2-components model and PCR a 3-components, which shows SPCR ability to choose better components for the estimation at hand. It is interesting to note that in the cases where SPCR performed noticeably worse (animals 6, 8 and 9), at least it doubled the number of components used by PCR. The biggest

	PCR		SPCR		difference	
Animal	Comps	MAE	Comps	MAE	subs	%
1	1	0.623	3	0.584	0.0389	6.24
2	3	0.907	5	0.886	0.0216	2.38
3	3	1.05	3	0.969	0.0841	7.99
4	2	0.527	14	0.524	3.14×10^{-3}	0.596
5	3	1.10	4	1.13	−0.0328	−2.99
6	3	0.602	16	0.714	−0.112	−18.6
7	3	1.36	2	1.24	0.114	8.41
8	3	1.20	6	1.41	−0.217	−18.1
9	2	1.15	24	1.25	−0.0940	−8.15

Table 3: Experiment 3: PCR vs SPCR mean absolute error (MAE) mmHg/ml for each individual E_{max} prediction. We compare the number of components (Comps) and the error difference.

relative difference was in animal 6, 3 components in PCR and 16 in SPCR, which resulted in 19% error increase.

Table 4 shows the same comparison for τ. Overall SPCR only provided better estimations for 8 out of 20 animals. Taking into account that PCR chose 2 components for all animals but one, for which chose 1; one interesting aspect was that SPCR chose 1 component for 7 animals. We can compare PCR, and SPCR means of principal components in Figures 9 and 10, where we see that first principal component for τ is almost the same for both methods and it was uniformly chosen by all predictors. Therefore, SPCR was not selecting more components because the second component with highest correlation with the output (that correspond to the component from 7-th to 18-th highest variance depending on the case) overfitted the output and was kept in the growing sets of components to evaluate. On the other hand, when PCR and SPCR agree on the number of components (animals 8 and 16), SPCR performs better. In the case of animal 8 both principal components are almost the same, the one in PCR has obtained with all data, as the component selection does not involve the variable to be estimated; but only the training data was used to obtain SPCR components. In the case of animal 16 the improvement could be ascribed to the same reason, or to the convenience in this case of the correlation with the target instead of the variance. When the number of components selected by SPCR was higher than the selected by PCR, the performance of SPCR was higher only in 3 out of 11 animals (5, 9 and 14). In these three cases the

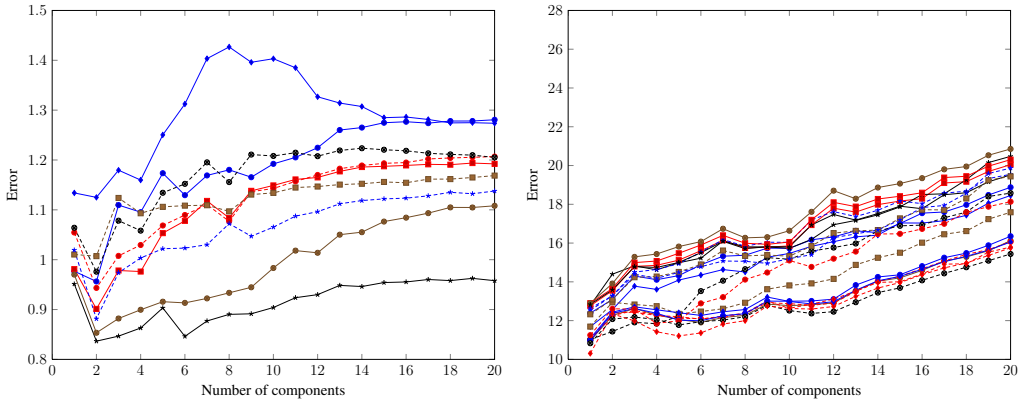
	PCR		SPCR		difference	
Animal	Comps	MAE	Comps	MAE	subs	%
1	2	4.99	1	4.50	0.490	9.82
2	2	7.74	1	7.42	0.322	4.15
3	2	3.78	7	5.64	−1.86	−49.3
4	2	4.93	1	10.3	−5.38	−109
5	2	14.2	5	12.4	1.78	12.5
6	2	8.96	5	10.2	−1.26	−14.0
7	2	14.5	17	16.2	−1.70	−11.8
8	1	25.0	1	23.3	1.61	6.46
9	2	14.4	3	12.4	2.02	14.0
10	2	6.08	6	6.84	−0.761	−12.5
11	2	6.99	5	8.44	−1.46	−20.9
12	2	15.2	1	14.2	0.979	6.45
13	2	10.5	1	12.7	−2.20	−20.9
14	2	7.89	3	7.77	0.128	1.62
15	2	11.7	7	14.8	−3.04	−25.9
16	2	13.3	2	12.0	1.28	9.68
17	2	7.18	1	8.15	−0.965	−13.4
18	2	8.32	6	10.0	−1.70	−20.4
19	2	12.3	8	17.2	−4.97	−40.5
20	2	22.1	1	22.4	−0.202	−0.912

Table 4: Experiment 3: PCR vs SPCR mean absolute error (MAE) ms for each individual τ prediction. We compare the number of components (Comps) and the error difference.

number of selected components (5, 3 and 3) was relatively low compared to the case of animals 3, 19 or 15 (49%, 40% and 26% increase of error, respectively).

3.5. Experiment 4: PLS on Raw Images Input Space

In this experiment, we estimated the LV indices from raw CDMM images with PLS using the same methodology of previous sections. We first considered the prediction of E_{max}. Figure 11 shows the search for the optimal number of dimensions, which turned to be 2 for all cases being independent of the individual left out. This shows how efficiently PLS is able to find the subspace useful for prediction. The mean of the loadings obtained for each predictor is shown in Figure 12. The first loading was almost the same for all predictors, the second loading



(a) PLS E_{max} (b) PLS τ

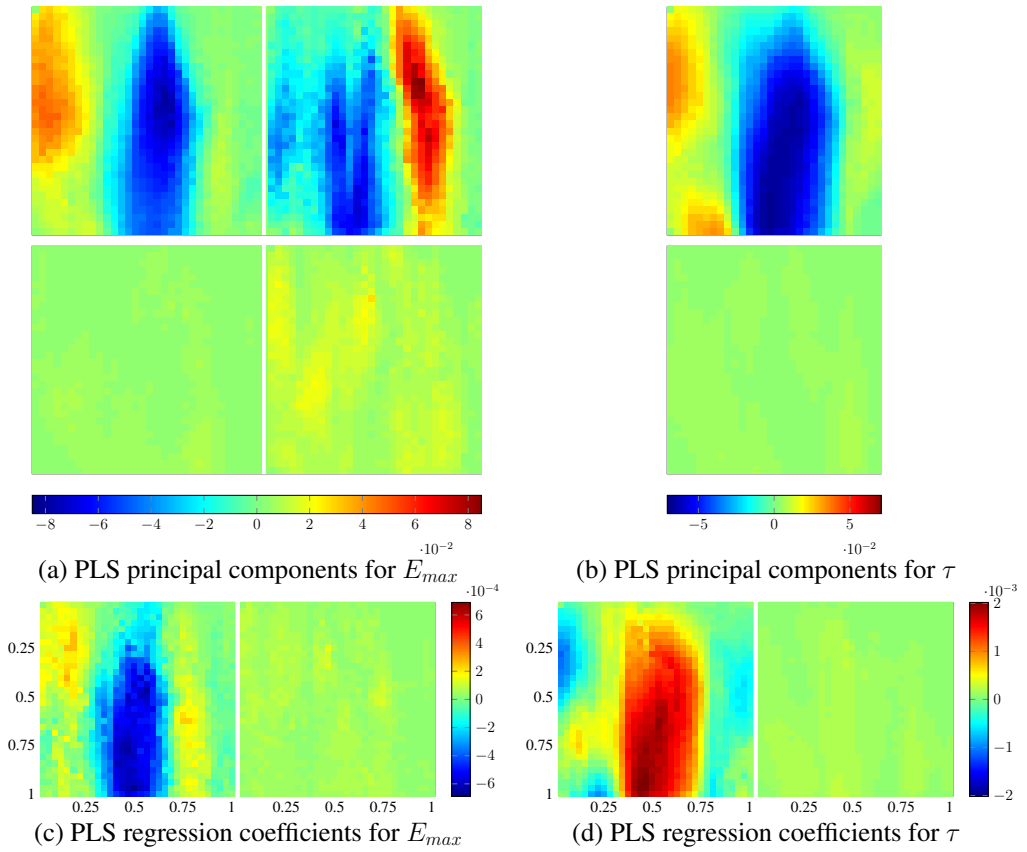
Figure 11: PLS on raw images: Number of components search. Each line is the search result for a different test individual.

was similar for all predictors with little variations. The mean of the predictors for the 9 individuals is shown in Figure 12, which shows a strong similarity among them and their main dependence on the first loading. The predictor was roughly contrasting the normalized mid-ejection time (emphasis near the LVOT) with its surroundings (emphasis at the beginning of the normalized ejection time near the apex and after the normalized mid-ejection time near the LVOT). The prediction, residuals and BA plot is shown in Figure 13. Prediction is similar to the one provided by SVR-LIN (see Figure 2) but slightly higher error, as we see one sample above 6 mmHg/ml in the BA plot.

The τ case was similar, Figure 11 shows the search for the optimal number of dimensions, which turned to be one for all cases. The mean of the loadings is shown in Figure 12, where we can see a great agreement for all individuals. The same happened with the predictors. The mean of all predictors was similar to the one obtained by PCR (see Figure 9). The prediction, residuals and BA (not shown) were similar to previous results. The error was compared with the other methods applied to raw CDMM images in Table 1, attaining the best accuracy after SVR-LIN.

3.6. Experiment 5: PCR and PLS on DCT Images Input Space

In this experiment we estimated the reference indices from the DCT coefficients of the CDMM images using the same methodology as in previous sections. To decide which coefficients to preserve, we used the same energy targets that in Experiment 2. In addition, we also preserved all the coefficients. The coefficients



(a) PLS principal components for E_{max}

(b) PLS principal components for τ

(c) PLS regression coefficients for E_{max}

(d) PLS regression coefficients for τ

Figure 12: PLS on raw CDMM images: (a) and (b) are the principal components for E_{max} and τ, first row is the mean for all the individuals and second row standard deviation. (c) and (d) are the mean and standard deviation of the individual's predictors for E_{max} and τ.

variance was untouched to not distort the images. The results of the prediction of E_{max} and τ are summarized in Table 5. When all the energy is preserved, the prediction results of using the DCT coefficients were the same as using the raw images. This can be interpreted as a basis change that do not affect the results as both PCR and PLS construct a new basis from the input. However, if a fixed amount of energy is preserved, which could be interpreted as some type of filtering in frequency, PCR and PLS could be benefited. For instance, we observe in Table 5 for E_{max} prediction that PCR did slightly better using 90% of the energy given by DCT than using all the energy, which shows that PCR could be benefited by filtering. On the other hand, PLS attained its best result when keeping all the energy. τ prediction, which is harder, showed that DCT pre-filtering improved the

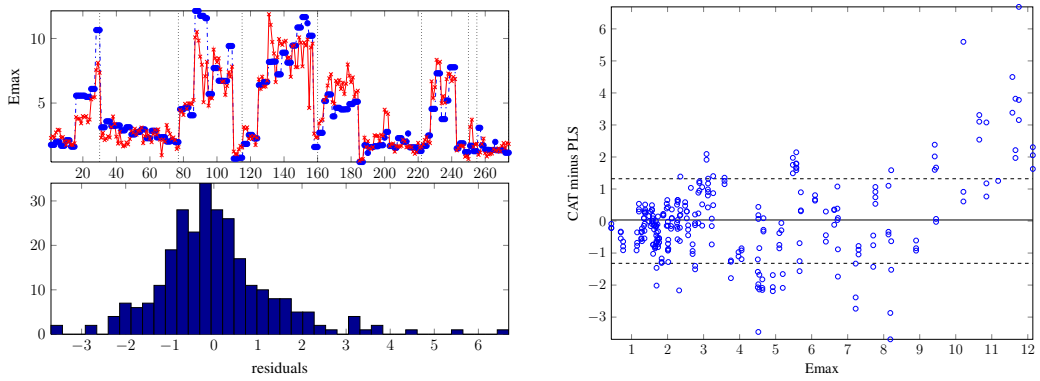


Figure 13: PLS: prediction and BA plot. E_{max}.

	PCR				PLS			
	E_{max}		τ		E_{max}		τ	
Energy	$\bar{E}$	$\pm$ desv	$\bar{E}$	$\pm$ desv	$\bar{E}$	$\pm$ desv	$\bar{E}$	$\pm$ desv
70	1.18	$\pm$ 0.50	11.1	$\pm$ 5.5	1.15	$\pm$ 0.46	10.7	$\pm$ 5.3
75	1.03	$\pm$ 0.38	**10.8**	$\pm$ **5.5**	0.962	$\pm$ 0.31	10.7	$\pm$ 5.2
80	1.02	$\pm$ 0.36	10.9	$\pm$ 5.2	0.950	$\pm$ 0.31	10.8	$\pm$ 5.2
85	0.970	$\pm$ 0.32	11.2	$\pm$ 5.5	0.955	$\pm$ 0.33	**10.5**	$\pm$ **4.7**
90	**0.946**	$\pm$ **0.30**	11.0	$\pm$ 5.6	0.931	$\pm$ 0.31	10.7	$\pm$ 5.3
95	0.963	$\pm$ 0.31	11.0	$\pm$ 5.6	0.933	$\pm$ 0.32	10.7	$\pm$ 5.3
99	0.947	$\pm$ 0.30	11.0	$\pm$ 5.6	0.927	$\pm$ 0.32	10.7	$\pm$ 5.3
99.5	0.947	$\pm$ 0.30	11.0	$\pm$ 5.6	0.927	$\pm$ 0.32	10.7	$\pm$ 5.3
99.9	0.947	$\pm$ 0.30	11.0	$\pm$ 5.6	0.927	$\pm$ 0.32	10.7	$\pm$ 5.3
100	0.947	$\pm$ 0.30	11.0	$\pm$ 5.6	**0.927**	$\pm$ **0.32**	10.7	$\pm$ 5.3

Table 5: Experiment 5. PLS and PCR using DCT coefficients as input space. Energy is shown in %, units for E_{max} and τ are mmHg/ml and ms, respectively.

results both for PCR (75% of the energy) and PLS (85%) of the energy.

3.7. Experiment 6: Linear vs Nonlinear Estimations

In this experiment, we tested the nonlinear counterparts of the algorithms of the previous experiments applying the same methodology. The results are summarized in Table 6. If we consider the performance of E_{max} estimators using raw CDMM images (Table 1 vs Table 6), we observe that linear methods outperformed

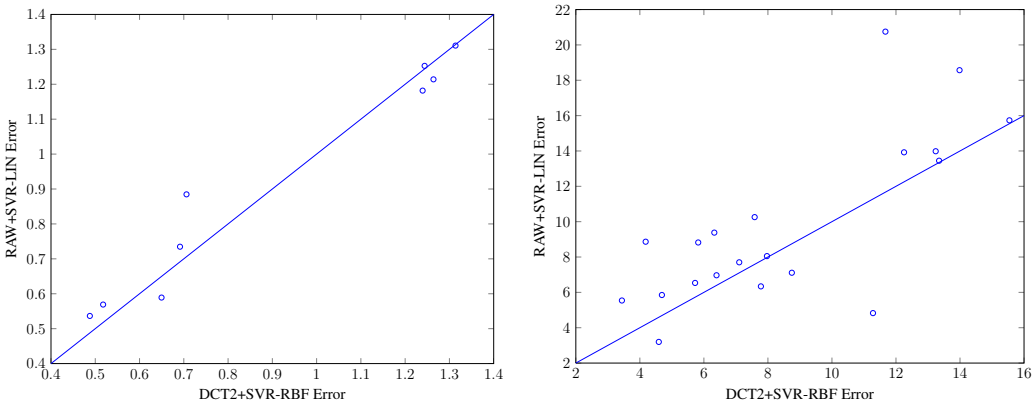


Figure 14: Comparison of best linear method against best nonlinear method mean absolute error for each individual. (left) E_{max} result, (right) τ (result). Better nonlinear performance is for points above diagonal.

their nonlinear counterparts when estimating E_{max}, both method by method and globally. On the other hand, if we consider instead estimators that use DCT coefficients (Table 2 vs Table 6), we found that DCT+SVM-RBF keeping 90% of the energy outperformed all linear and nonlinear methods.

However, we observe that nonlinear methods outperformed their linear counterparts, both method by method and globally, in the estimation of τ using raw CDMM images (Table 1 vs Table 6). The same happened in the DCT space (Table 2 vs Table 6). Globally, the best result for τ was found by DCT+SVM-RBF keeping 85% of the energy, and very close to it by SVM-RBF on the raw CDMM images after zero-mean, unit-variance normalization.

Finally, we compared best linear and nonlinear methods performance for each individual. Figure 14 compares RAW+SVM-LIN vs DCT2+SVM-RBF. We observe that nonlinear method did much better in the case of τ; 16 out of 20 individuals were better estimated by the nonlinear method. For E_{max}, nonlinear method was also better (5 out of 9 cases), but the difference among both methods was only clear for 1 out of 9 individuals, where linear was noticeably worse than nonlinear kernel.

4. Discussion and Conclusion

High-dimensional input spaces and overfitting. Among the kernel methods community, it is generally accepted that linear Support Vector Machine (SVM) is

Method	Energy (%)	E_{max} mmHg/ml $\bar{E}$	± desv	τ ms $\bar{E}$	± desv
RAW+SVM-RBF		0.942	± 0.39	8.88	± 3.7
RAW+SVM-RBF†		0.970	± 0.38	8.59	± 3.9
RAW+KPCR		1.02	± 0.32	9.82	± 3.9
RAW+KSPCR		0.969	± 0.42	9.19	± 3.3
RAW+KPLS		0.944	± 0.34	9.45	± 3.5
DCT1+SVM-RBF	70	1.19	± 0.70	10.1	± 3.9
DCT1+SVM-RBF	75	0.931	± 0.38	9.83	± 4.2
DCT1+SVM-RBF	80	0.981	± 0.35	9.42	± 3.5
DCT1+SVM-RBF	85	0.985	± 0.38	9.10	± 3.5
DCT1+SVM-RBF	90	1.08	± 0.43	9.38	± 3.6
DCT1+SVM-RBF	95	1.28	± 0.62	9.40	± 3.9
DCT1+SVM-RBF	99	1.46	± 0.74	9.98	± 3.5
DCT1+SVM-RBF	99.5	1.43	± 0.75	10.1	± 3.4
DCT1+SVM-RBF	99.9	1.43	± 0.72	10.1	± 3.7
DCT2+SVM-RBF	70	1.09	± 0.66	9.86	± 3.6
DCT2+SVM-RBF	75	0.928	± 0.37	9.72	± 3.8
DCT2+SVM-RBF	80	0.949	± 0.38	9.28	± 3.4
DCT2+SVM-RBF	85	0.948	± 0.36	**8.58**	± **3.7**
DCT2+SVM-RBF	90	**0.902**	± **0.35**	8.78	± 3.4
DCT2+SVM-RBF	95	0.922	± 0.36	8.95	± 3.5
DCT2+SVM-RBF	99	0.943	± 0.40	8.87	± 3.6
DCT2+SVM-RBF	99.5	0.948	± 0.41	8.78	± 3.6
DCT2+SVM-RBF	99.9	0.945	± 0.40	8.78	± 3.6

Table 6: Experiment 6: Nonlinear methods comparison. † indicates that the image pixels were made zero-mean unit-variance. Otherwise, they were only divided by the maximum. DCT1 stands for zero-mean unit-variance DCT coefficients normalization and DCT2 for zero-mean untouched-variance DCT coefficients normalization.

better than nonlinear SVM for high dimensional input spaces like text categorization (Yang, 1999) or fMRI data (Cox and Savoy, 2003; LaConte et al., 2005; Misaki et al., 2010). This is supported by two theoretical facts: Cover's Theorem (Cover, 1965) and Vapnik-Chervonenkis VC-dimension (Vapnik, 1995). Cover's Theorem discusses the probability of separating a set of n samples into two sets in a general setting. He showed that the natural capacity of a transformation having

d degrees of freedom is $2d$. In other words, the expected number of points that can be separated into two sets with high probability by a transformation of d degrees of freedom is $2d$. The VC-dimension of a family of surfaces is the number of points that the family can classify into two classes no matter the points' labels. VC dimension of SVM with linear kernel (hyperplane) in d dimensions is $d + 1$ and VC-dimension of a SVM with Gaussian kernel is ∞. The capacity of SVM with Gaussian kernel is very high which might lead to overfitting.

In our case, we have CDMM images that were resampled to have $d = 1024$ pixels. So we can assume there is high probability that a hyperplane classifies up to $n = 2048$ points in this space. Actually, as VC dimension expresses, we can separate any configuration of 1025 points with hyperplanes. We have 274 and 1362 samples respectively for E_{max} and τ. Therefore, this dimension is enough for a classification task with a linear kernel and the number of samples might be too small to project these samples in a higher-dimension space and then classify them using a hyperplane, which is what SVM with Gaussian kernel does. Therefore, it is easy to fall in overfitting when using nonlinear SVM and the input space is already very high, as the learner is too flexible.

We can qualitatively borrow these ideas and expect a similar behavior in our estimation (regression) problem. VC-dimension has also been proposed for regression linear methods (Cherkassky et al., 1999). The question is how to address an estimation problem where the input space is higher than or comparable to the number of samples. One possibility is to follow the ideas of Structural Risk Minimization (Vapnik, 1995), i.e., create nested subsets of increasing complexity where to find the regressor function. This approach is naturally followed by PCR and PLS regression methods, where there is a step to decide the number of latent components in the data, and then there is a step of search for the best regressor using this number of components. Instead of bounds we used cross-validation to select the adequate number of components. For SVR, we have to find the hyperparameters, but it is not direct to find the parameter sets that form the aforementioned nested subsets of increasing complexity. Therefore, we used the usual cross-validation to select SVR hyperparameters.

Conclusions on hetherocedasticity. The estimation results, though good for E_{max}, are still heteroscedastic (see BA plots for SVR, PCR and PLS), which shows that the model could be improved with extra variables. In the case of τ, the heteroscedasticity is considerable, rendering the method unable to accurately estimate high values of τ. We have to admit that the estimation quality needs to be improved, but still can be of utility in clinical practice. Other variables like heart

rate, peak LV pressure, LV ejection fraction and their squares that could also be estimated noninvasively have shown to improve τ estimation (Rojo-Alvarez et al., 2006).

Interpretability of weights in linear machines with raw input features. One of the main strengths of linear methods, in addition to simplicity, is the possibility of interpret as images what the estimator is doing. These images can be analyzed by an expert to assess whether the method is doing something "reasonable" with the data. Also, these estimator images could give some insights of LV function suggesting new research hypotheses. For E_{max} the estimators given by SVR-LIN and SPCR are quite rough in comparison with the ones given by PCR and PLS. Therefore as the accuracy is almost the same (less than 1% increased error), PLS estimator could be easier to interpret and preferable as it obtained with only two latent factors. For τ the estimator given by PCR and PLS are smoother than the one given by SVR-LIN, however this one estimates much better (error decrease of 9%). So, in this case it could be worthy to interpret SVR-LIN estimator instead. Anyway, as the estimation is not so good, the results should be interpreted with caution.

Conclusions on DCT input spaces. DCT representation seems a good choice for CDMM images, as it gives an insight on their frequency content and allows for effective compression. We represented the estimator as images including standardization. The estimator obtained using DCT+SVR-LIN is somewhat similar to a very smooth version of SVR-LIN estimator, both in case of E_{max} or τ, where their estimation capabilities are really close to SVR-LIN (3% more error for E_{max} and 1% for τ). DCT+SVR-LIN estimators use much less coefficients (6 for E_{max}, 11 for τ), therefore, it could be worthy to interpret their estimators instead of SVR-LIN ones, especially for τ, where PCR and PLS are not so good. In addition, we showed that DCT pre-filtering can improve the results of PCR and PLS, especially in the case of hard prediction problems, such τ, and therefore it can be applied previously to other dimensionality reduction techniques as PCR and PLS.

DCT coefficients standardization should be taken with care. We remove the mean of all coefficients as we do not expect it would be useful for estimation. It could be useful, as an additional criterion to the variance, to select which coefficients to keep. If we decide to standardize the variance of the coefficients with the aim of giving the same importance to all the frequencies, we must be careful selecting the frequencies to include for not amplifying the noise.

Intrinsic image features for input spaces. SPCR could lead to improvements over PCR but we found this improvements more likely in well characterized problems (as E_{max} estimation). Problems where the estimation is hard (as τ estimation) were more suitable for PCR. We conjecture that one of the problems of SPCR is that very low variance components can, by chance, have high correlations with the target variable but no real meaning in the data. A previous filter based on the variance on the number of components that removes components below a threshold could be useful for SPCR. Another important aspect to consider is the way we select the sets of components. We restricted ourselves to nested component-sets, where each set with g components is formed by the g components with highest variance (PCR) or correlation (SPCR). Wrapper feature selection methods could give better component sets at the price of higher search times. The overfitting example given by SPCR on raw data for τ estimation reminds us how careful we should be when the estimation problem at hand uses data in high dimension and we only have a number of points comparable to the dimension. Being conservative in this type of cases (use the variance to select the components as PCR) it is safer and provides better results because we can not select noisy components.

Nonlinear transformations of input spaces. We have also experimentally shown that nonlinear methods could outperform linear methods when a previous feature extraction step is considered to estimation. This behavior is also found in other high-dimensional datasets (e.g. Mourão Miranda et al. (2005); Song et al. (2011)). We got better estimations by using DCT+SVM-RBF than using linear methods. We conjecture, taking into account our previous discussion on complexity, that convenient feature extraction in nonlinear estimation problems enables taking advantage of the flexibility of nonlinear estimation methods by limiting their overfitting capabilities and reducing the search space. Also, not mixing individuals in the cross-validation folds helps to prevent overfitting by inducing conservative parameter selections.

In addition, we observed that transformations of the variable to be estimated could lead to better estimations. The residuals of E_{max} and τ estimation provided by RAW+SVR-LIN have heavy-tails (see Figure 2). Therefore, we tried to estimate $\log(E_{max})$ and $\log(\tau)$ instead and we found a 6% increase of accuracy in the case of E_{max}, and a 4% decrease in accuracy for τ. This is a simple example of how a simple transformation could increase accuracy.

We can conclude that CDMM image-based noninvasive estimates of reference indices of LV function can be improved when using nonlinear machines. This increase in performance involves the selection of alternative lower dimension input

spaces where the power of nonlinear machines can be exploited without falling in overfitting. Overfitting becomes an issue in estimation if flexible architectures are used in high dimensional input spaces. We have shown how DCT input space could be a good alternative to raw images in echocardiography that could benefit from the use of nonlinear machines.

On the other hand in exploratory analysis and when the understanding of the estimator is a must, we found PLS to have a good balance between accuracy and simplicity enabling knowledge gain and expert feedback about the estimator.

5. Acknowledgments

This study was partially supported by research projects TEC2010-19263 (EXCALIVUR) from Ministerio de Ciencia e Innovación, TEC2013-48439-C4-1-R (PRINCIPIAS) from Ministerio de Economía y Competitividad, PRIN13_IYA12 from Universidad Rey Juan Carlos and grants PIS09/02603, PS09/02602, and RD06/0010 (RECAVA), from the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+I), Instituto de Salud Carlos III–Ministerio de Economía y Competitividad, Spain, and Prometeo Project of the Secretariat for Higher Education, Science, Technology and Innovation of Ecuador. Author R. Santiago-Mozos is supported by the Juan de la Cierva Program (Ref: JCI-2011-11150) of the Ministerio de Ciencia e Innovación, Spain.

References

Bermejo, J., Antoranz, J.C., Yotti, R., Moreno, M., García-Fernández, M.A., 2001. Spatio-temporal mapping of intracardiac pressure gradients. A solution to Euler's equation from digital postprocessing of color Doppler M-mode echocardiograms. Ultrasound in Medicine & Biology 27, 621–630. doi:10.1016/S0301-5629(01)00349-0.

Cherkassky, V., Shao, X., Mulier, F.M., Vapnik, V.N., 1999. Model complexity control for regression using VC generalization bounds. IEEE Transactions on Neural Networks 10, 1075–89. doi:10.1109/72.788648.

Cover, T.M., 1965. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Transactions on Electronic Computers EC-14, 326–334. doi:10.1109/PGEC.1965.264137.

Cox, D.D., Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19, 261–270. doi:10.1016/S1053-8119(03)00049-1.

Hadi, A.S., Ling, R.F., 1998. Some cautionary notes on the use of principal components regression. The American Statistician 52, 15–19. doi:10.1080/00031305.1998.10480530.

Höskuldsson, A., 1988. PLS regression methods. Journal of Chemometrics 2, 211–228. doi:10.1002/cem.1180020306.

Jolliffe, I.T., 1982. A note on the use of principal components in regression. Journal of the Royal Statistical Society. Series C (Applied Statistics) 31, 300–303. doi:10.2307/2348005.

de Jong, S., 1993. SIMPLS: An alternative approach to partial least squares regression. Chemometrics and Intelligent Laboratory Systems 18, 251–263. doi:10.1016/0169-7439(93)85002-X.

LaConte, S., Strother, S., Cherkassky, V., Anderson, J., Hu, X., 2005. Support vector machines for temporal classification of block design fMRI data. NeuroImage 26, 317–29. doi:10.1016/j.neuroimage.2005.01.048.

Massy, W.F., 1965. Principal components regression in exploratory statistical research. Journal of the American Statistical Association 60, 234–256. doi:10.2307/2283149.

Mourão Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., Stetter, M., 2005. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. NeuroImage 28, 980–95. doi:10.1016/j.neuroimage.2005.06.070.

Misaki, M., Kim, Y., Bandettini, P.A., Kriegeskorte, N., 2010. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. NeuroImage 53, 103–18. doi:10.1016/j.neuroimage.2010.05.051.

Rojo-Alvarez, J.L., Bermejo, J., Juarez-Caballero, V.M., Yotti, R., Cortina, C., Garcia-Fernandez, M.A., Antoranz, J.C., 2006. Support vector analysis of

color-Doppler images: a new approach for estimating indices of left ventricular function. IEEE Transactions on Medical Imaging 25, 1037–1043. doi:10.1109/TMI.2006.875437.

Rosipal, R., Girolami, M., Trejo, L.J., 2001a. On kernel principal component regression with covariance inflation criterion for model selection. Technical Report. Department of Computing and Information Systems, University of Paisley.

Rosipal, R., Girolami, M., Trejo, L.J., Cichocki, A., 2001b. Kernel PCA for feature extraction and de-noising in nonlinear regression. Neural Computing & Applications 10, 231–243. doi:10.1007/s521-001-8051-z.

Rosipal, R., Krämer, N., 2006. Overview and recent advances in partial least squares, in: Saunders, C., Grobelnik, M., Gunn, S., Shawe-Taylor, J. (Eds.), Subspace, Latent Structure and Feature Selection. Springer Berlin Heidelberg, Berlin, Heidelberg. volume 3940 of *Lecture Notes in Computer Science*, pp. 34–51. doi:10.1007/11752790.

Rosipal, R., Trejo, L.J., 2001. Kernel partial least squares regression in reproducing kernel Hilbert space. The Journal of Machine Learning Research 2, 97–123.

Schölkopf, B., Smola, A., 2001. Learning with kernels. MIT Press, Cambridge, MA, USA.

Schölkopf, B., Smola, A.J., Müller, K.R., 1999. Kernel principal component analysis, in: Schölkopf, B., Burges, C.J.C., Smola, A.J. (Eds.), Advances in Kernel Methods—Support Vector Learning. MIT Press, Cambridge, MA, pp. 327–352.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and Computing 14, 199–222. doi:10.1023/B:STCO.0000035301.49549.88.

Song, S., Zhan, Z., Long, Z., Zhang, J., Yao, L., 2011. Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data. PloS one 6, e17191. doi:10.1371/journal.pone.0017191.

Vapnik, V.N., 1995. The nature of statistical learning theory. Springer-Verlag, New York.

Wold, H., 1966. Nonlinear estimation by iterative least squares procedures, in: David, F. (Ed.), Researh Papers in Statistics. Wiley, New York, pp. 411–444.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58, 109–130. doi:10.1016/S0169-7439(01)00155-1.

Yang, Y., 1999. An evaluation of statistical approaches to text categorization. Information Retrieval 1, 69–90. doi:10.1023/A:1009982220290.

Yotti, R., Bermejo, J., Benito, Y., Antoranz, J.C., Desco, M.M., Rodríguez-Pérez, D., Cortina, C., Mombiela, T., Barrio, A., Elízaga, J., Fernández-Avilés, F., 2011. Noninvasive estimation of the rate of relaxation by the analysis of intraventricular pressure gradients. Circulation. Cardiovascular imaging 4, 94–104. doi:10.1161/CIRCIMAGING.110.960369.

Yotti, R., Bermejo, J., Desco, M.M., Antoranz, J.C., Rojo-Álvarez, J.L., Cortina, C., Allué, C., Rodríguez-Abella, H., Moreno, M., García-Fernández, M.A., 2005. Doppler-derived ejection intraventricular pressure gradients provide a reliable assessment of left ventricular systolic chamber function. Circulation 112, 1771–9. doi:10.1161/CIRCULATIONAHA.104.485128.