

# On feature extraction for noninvasive kernel estimation of left ventricular chamber function indices from echocardiographic images

Ricardo Santiago-Mozos<sup>a,\*</sup>, José Luis Rojo-Álvarez<sup>a,d</sup>, J. Carlos Antoranz<sup>c</sup>, Mar Desco<sup>c</sup>, Daniel Rodríguez-Pérez<sup>c</sup>, Raquel Yotti<sup>b</sup>, Javier Bermejo<sup>b</sup>

<sup>a</sup> Department of Signal Theory and Communications, Universidad Rey Juan Carlos, Fuenlabrada, Madrid, Spain

<sup>b</sup> Department of Cardiology, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón, Facultad de Medicina, Universidad Complutense de Madrid, Spain

<sup>c</sup> Department of Mathematical Physics and Fluids, Universidad Nacional de Educación a Distancia, Spain

<sup>d</sup> Prometeo, Department of Electric and Electronic Engineering, Universidad de las Fuerzas Armadas ESPE, Ecuador



## ARTICLE INFO

### Article history:

Available online 15 January 2015

### Keywords:

Color Doppler M-mode imaging  
Kernel methods  
Left ventricular function  
Elastance  
LV relaxation  
Cosine transform

## 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 ( $\tau$ ). 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 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 nonlinear versions of the above methods provided significant improvement in the estimation quality. Our results showed that very few input features suffice for providing a good (medium) quality estimator for  $E_{max}$  (for  $\tau$ ), 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  $\tau$  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.

© 2014 Elsevier Inc. All rights reserved.

## 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 ( $\tau$ ) 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  $\tau$  requires invasive catheterization of the LV using high-fidelity micromanometers. For these reasons, neither  $E_{max}$  nor  $\tau$  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  $\tau$ . 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  $\tau$  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 [27]. Similarly,  $\tau$  can be approximated by the peak reverse end-ejection pressure difference with reasonable accuracy [26]. Importantly, using a learning from sam-

\* Corresponding author.

E-mail addresses: ricardo.santiago.mozos@urjc.es (R. Santiago-Mozos), joseluis.rojo@urjc.es (J.L. Rojo-Álvarez), jcantoranz@dfmf.uned.es (J. Carlos Antoranz), mmDESCO@dfmf.uned.es (M. Desco), daniel@dfmf.uned.es (D. Rodríguez-Pérez), raquel.yotti@salud.madrid.org (R. Yotti), javier.bermejo@salud.madrid.org (J. Bermejo).

ples approach we have obtained similar approximations without the need of complex fluid-dynamic modeling [13]. 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  $\tau$  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  $\tau$  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 [20]. Given a training set  $\{(\mathbf{x}_i, y_i), i = 1, \dots, n\}$  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

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

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:

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

where  $\mathcal{L}$  is in our case the robust  $\epsilon$ -Huber loss function [13], which increases the flexibility modeling outliers, given by

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

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

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

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

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

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

$$\xi_i, \xi_i^* \geq 0 \quad (7)$$

for  $i = 1, \dots, n$ , and where  $\xi_i, \xi_i^*$  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 [20], we obtain the following solution:

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

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  $\alpha_i, \alpha_i^*$  have a non-zero value and their associated sample is named *support vector* (SV), because it influences function  $f$ . Three parameters ( $C, \epsilon, \delta$ ) need to be tuned for linear SVR (as well as the kernel width  $\sigma$  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 [10] 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)  $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_g]$  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:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T \quad (9)$$

where  $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_r]$  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 [10] 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 [7] and Hadi and Ling [5] 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:

$$\mathbf{Y} = \mathbf{T}\mathbf{B} + \mathbf{F} \quad (10)$$

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

$$\hat{\mathbf{Y}} = \mathbf{T}\hat{\mathbf{B}} = \mathbf{X}\mathbf{P}\hat{\mathbf{B}} \quad (12)$$

$$\hat{\mathbf{Y}}_t = \mathbf{X}_t\mathbf{P}\hat{\mathbf{B}} \quad (13)$$

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 [15,17], the nonlinear kernel PCR (KPCR) performs kernel PCA [19] 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 [14]. One parameter ( $g$ ) needs to be tuned for PCR (and the kernel width for Gaussian KPCR).

## 2.3. PLS estimation

PLS for regression (see [16,24] 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:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \quad (14)$$

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

where  $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_g]$  and  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_g]$  contain the extracted latent vectors as columns, and  $g$  is the number of latent components extracted;  $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_g]$  and  $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_g]$  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 [23] finds  $\mathbf{w}_i, \mathbf{c}_i$  vectors such that

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

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 [6] or using other approaches as SIMPLS [8] that directly finds  $\mathbf{w}_i$  vectors. The loadings can be computed as

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

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

$$\mathbf{X} = \mathbf{X} - \mathbf{t}_i\mathbf{p}_i^T \quad (17)$$

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

and the algorithm is repeated until the required number of components is reached (see [24] 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:

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

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{B} = \mathbf{T}\mathbf{C}^T \quad (20)$$

$$\hat{\mathbf{Y}}_t = \mathbf{X}_t\mathbf{B} \quad (21)$$

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 [18] to the predictor matrix  $\mathbf{X}$  and then perform a linear PLS in the feature space [17]. The NIPALS-based kernel version results:

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

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

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  $\tau$ . All measures were averaged by performing the experiment three times on each pig. For  $E_{max}$  estimation, 9 pigs were considered that jointly sum up to 274 beats. For  $\tau$  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) [27]. 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 [1]. 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 down-sampled to a grid of  $32 \times 32$  pixels. The details of the experimental setup have been reported in [27].

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  $\tau$ ).

### 3.2. Experiment 1: linear SVR on raw images input space

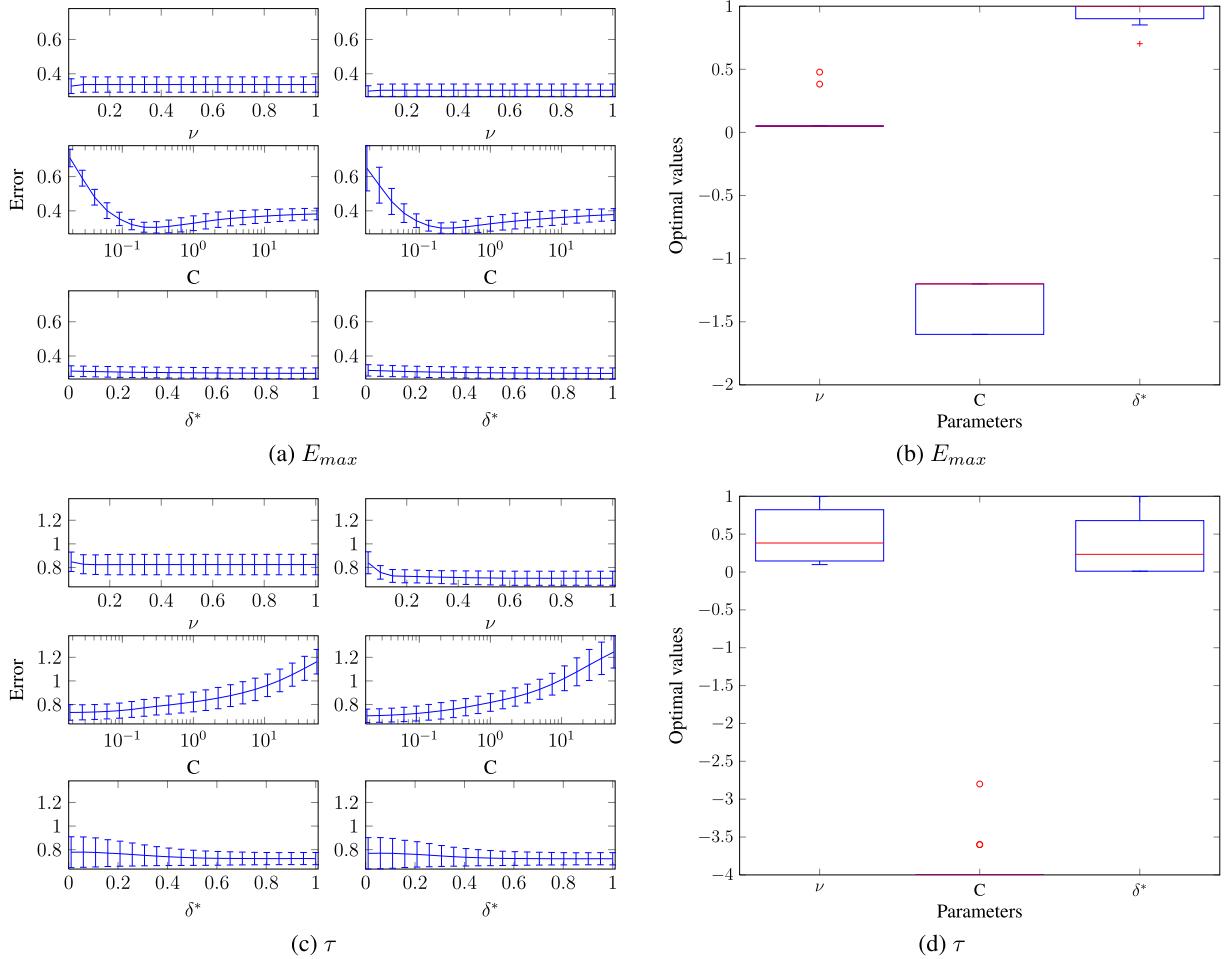
In this experiment, we analyzed the raw  $32 \times 32$  CDMM images, in a leave-one-individual-out scheme, with linear SVR. All the images were standardized by dividing them by the constant that made its maximum absolute value one.

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  $\delta C$  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  $\delta C$  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 cannot be modeled.

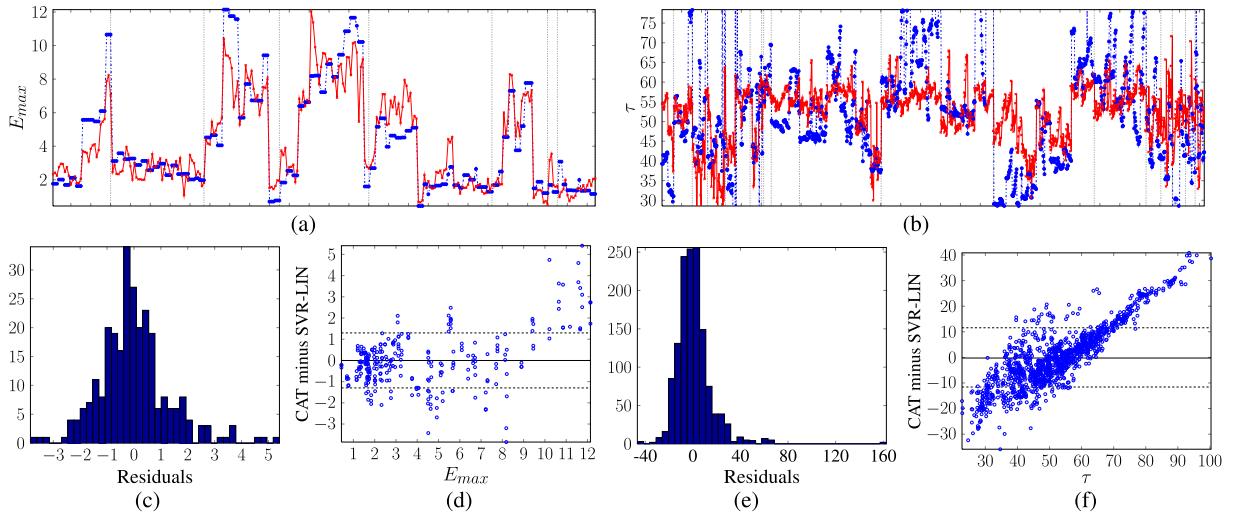
We started with the prediction of  $E_{max}$  by using the raw CDMM images. The average result (for all individuals) of the linear search of SVR parameters is shown in Fig. 1, which also shows the boxplot of the best values for each individual. Parameter  $\nu$  seemed to have little impact on the accuracy, and only  $C$  and  $\delta^*$  seemed to have some effect. However, when looking at the optimal parameters,  $\nu$  was close to zero for the prediction of 7 out of nine animals,  $C$  had a little variation, and  $\delta^*$  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 Fig. 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 Fig. 3(a), and their mean and standard deviation in Fig. 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.



**Fig. 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  $E_{max}(\tau)$  results.

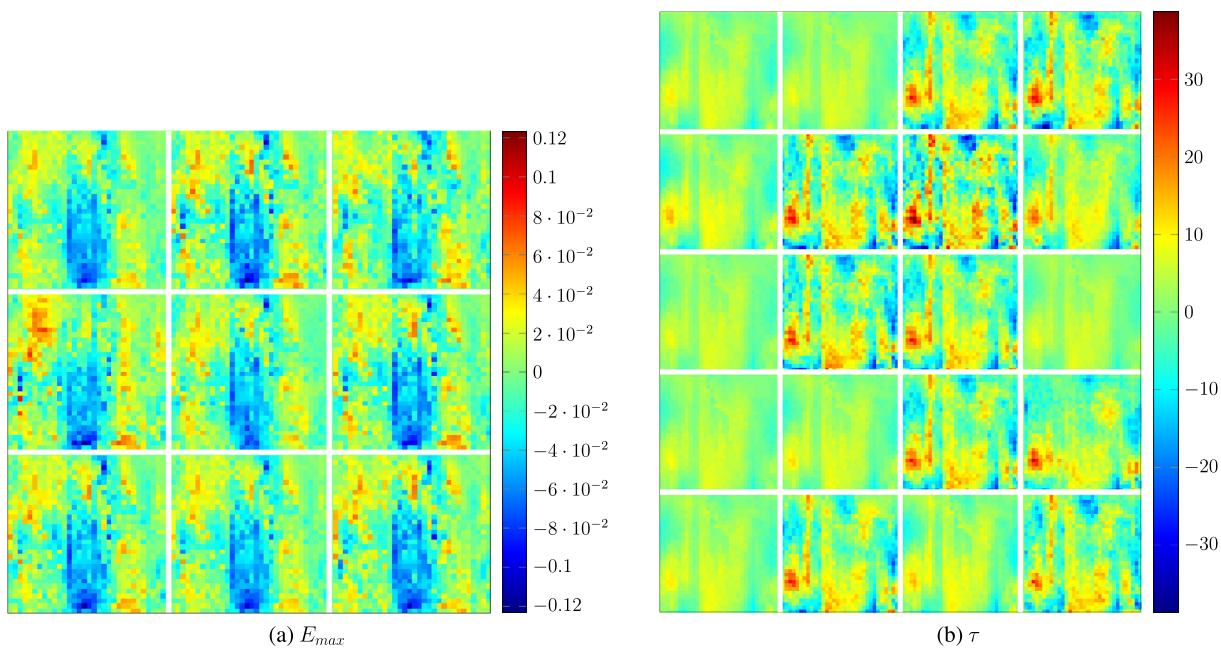


**Fig. 2.** SVR-LIN on raw images. (a) and (b) are the predictions for  $E_{max}$  and  $\tau$  respectively; catheter measurement is plotted in blue (dot-line) and prediction in red (continuous), vertical dotted lines divide the individuals. (c) and (d), and (e) and (f) are the residuals and Bland-Altman plots respectively for  $E_{max}$  and  $\tau$ . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

For  $\tau$  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 Fig. 1. In this case, the boxplot shows that both  $\nu$  and  $\delta^*$  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  $\nu$  from 0.2 to 1 had almost no effect.

The regression weights for  $\tau$ , represented as images, are shown in Fig. 3(b) and their mean and standard deviation in Fig. 4(b).



**Fig. 3.** SVR-LIN on raw images. (a) Regression coefficients to predict  $E_{max}$  for individuals 1–9. (b) Regression coefficients to predict  $\tau$  for individuals 1–20.

**Table 1**

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

Method	$E_{max}$ mmHg/ml	$\tau$ ms
	$\bar{E} \pm \text{desv}$	$\bar{E} \pm \text{desv}$
SVR-LIN	<b><math>0.919 \pm 0.32</math></b>	<b><math>9.79 \pm 4.7548</math></b>
PCR	$0.9465 \pm 0.2980$	$11.0005 \pm 5.5693$
SPCR	$0.9679 \pm 0.3151$	$11.8448 \pm 5.0939$
PLS	$0.92685 \pm 0.31652$	$10.6805 \pm 5.3261$

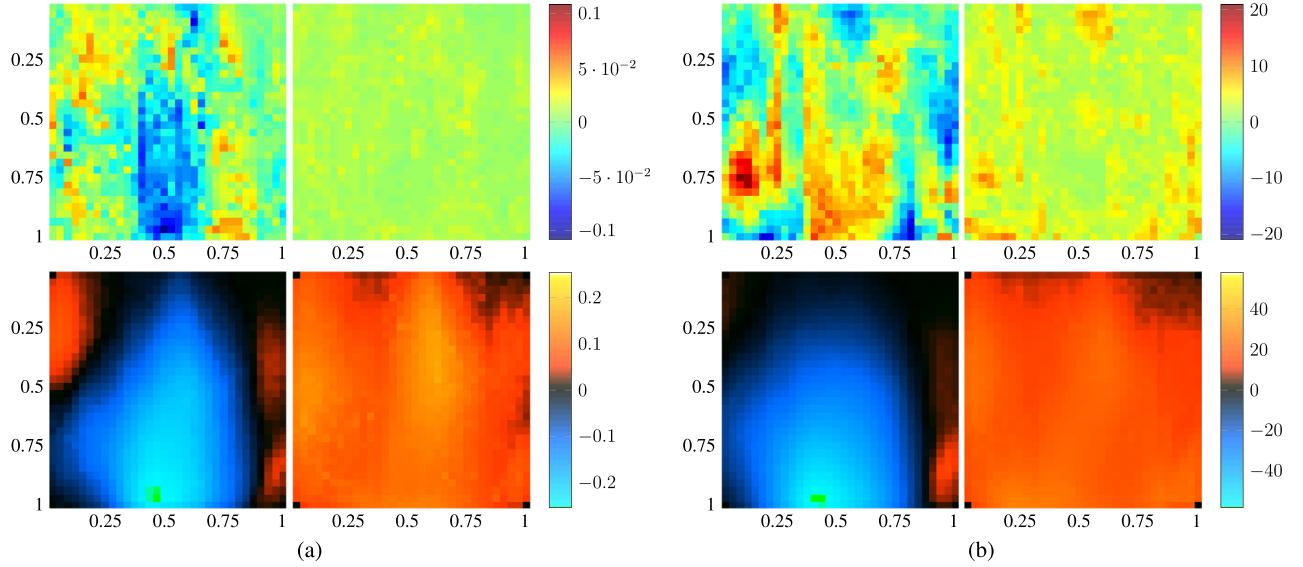
There were several patterns in the regression weights that depended at least on the selected training parameters: patterns like the ones for predicting animals 1, 2, 9, 12, 13, 14, 17, 19 resulted from training with  $\nu < 0.15$  (most samples do not penalize) and  $\delta^* > 0.5$  values (the quadratic 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 Fig. 4. The prediction and the residuals are shown in Fig. 2. We observe that the estimation provided by the SVR-LIN for  $\tau$  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  $\tau$  values. This systematic error dependent of the value of  $\tau$  (that could be modeled by a line and partially removed) demonstrates the inability of the model to follow high  $\tau$  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  $\tau$  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  $\tau$ , and the maximum absolute error was 162.6 ms which was an outlier as we can see in Fig. 2, the next maximum error was just above 60 ms.

### 3.3. Experiment 2: linear SVR on DCT input space

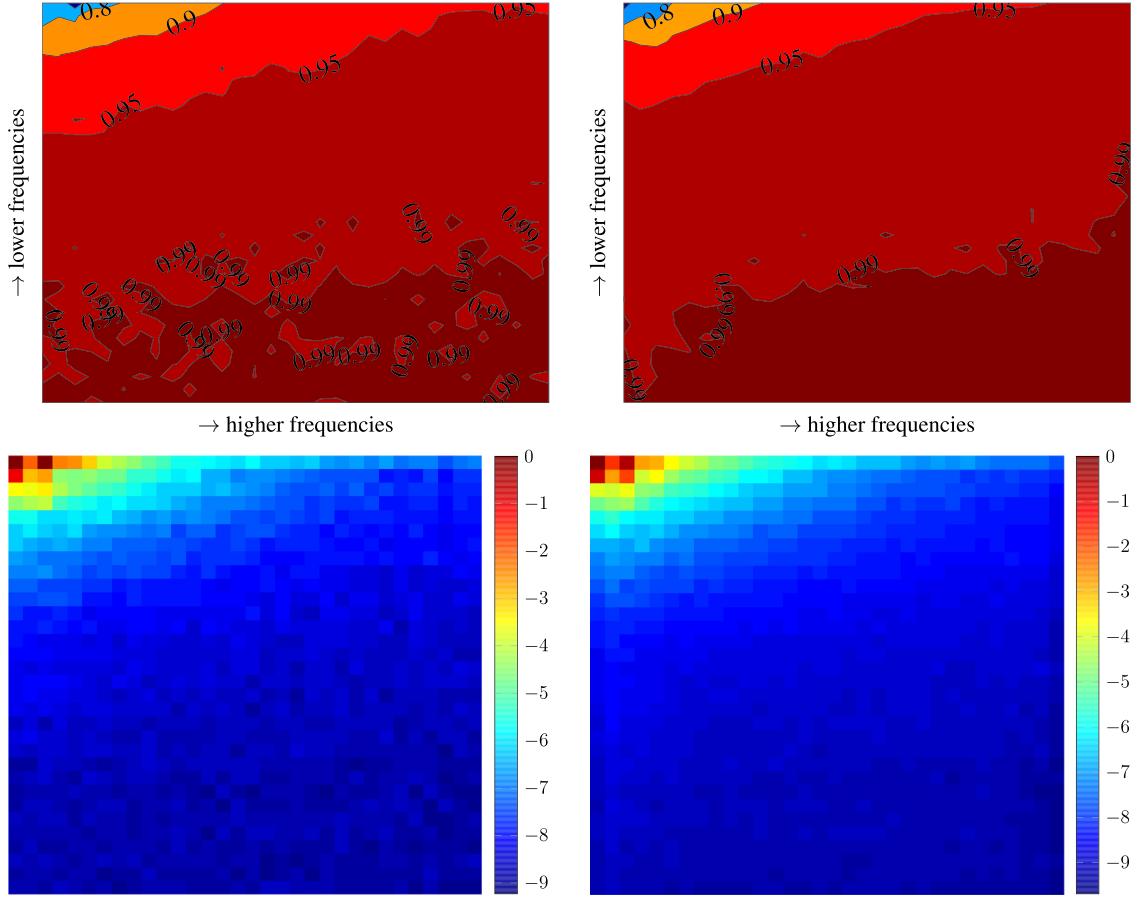
We also tested the estimation of  $E_{max}$  and  $\tau$  using the SVR with the Discrete Cosine Transform (DCT) coefficients instead of the raw images. DCT coefficients represent the frequency content of CDMM images, which resulted smooth and could be efficiently represented with few coefficients, e.g., {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  $\tau$  (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.

Fig. 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 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 Fig. 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



**Fig. 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  $\tau$  (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.

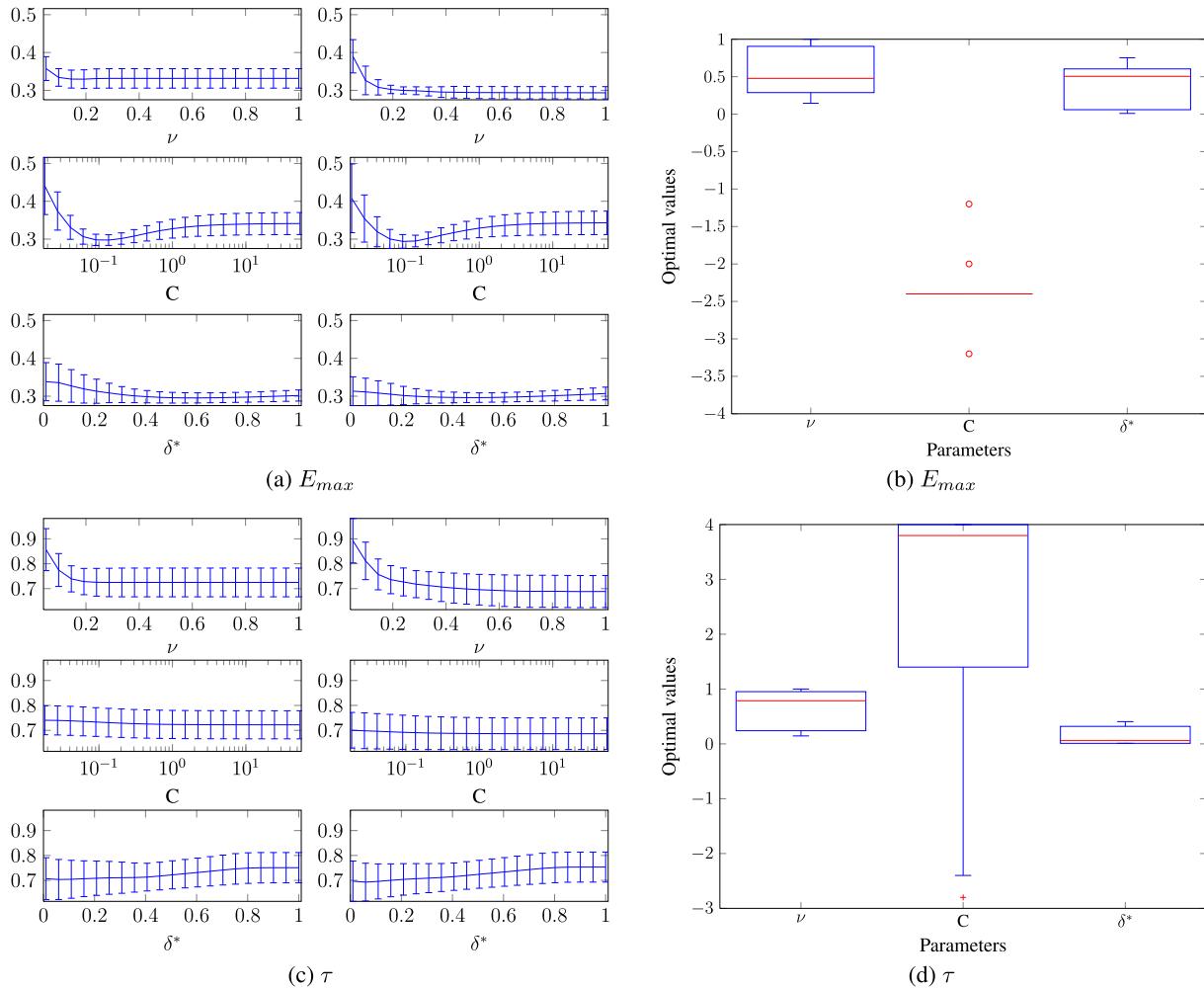


**Fig. 5.** First row present DCT coefficients cumulative energy.  $E_{\max}$  (left). {70, 75, 80, 85, 90, 95}% of the energy was respectively retained with {4, 6, 9, 15, 37, 176} out of 1024 coefficients. For  $\tau$  (right) the same energy percentages were retained with {4, 5, 7, 11, 23, 110} out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

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



**Fig. 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% ( $\tau$  retaining 85%) of energy.

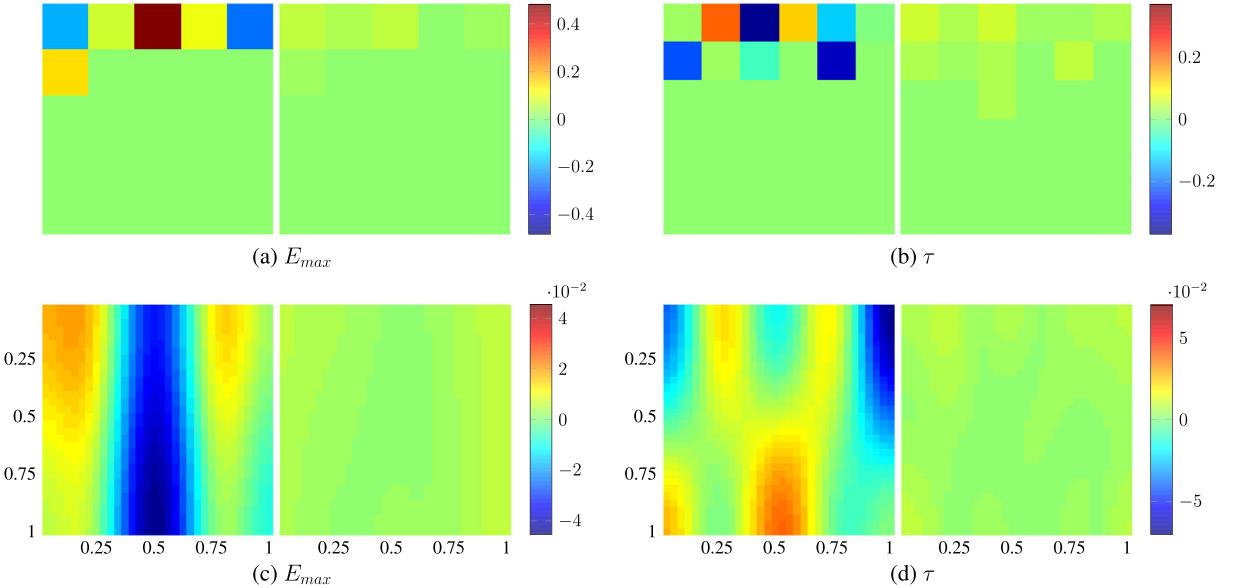
zero-mean unit-variance standardization). The optimum parameters of the search of SVR parameters (see Fig. 6) were different than the ones obtained using the raw images, as the input space has now only 6 dimensions. We found, however, some similarities, described below. In the parameter search for  $E_{max}$ ,  $\nu$  showed an approximately flat error from 0.2 to 1 (it was almost flat from 0 to 1 in Fig. 1) and the most chosen value was approximately 0.5 (it was close to 0 in Fig. 1);  $C$  showed a minimum around  $10^{-1}$  (same shape that in Fig. 1) and its most chosen optimum value was close to  $e^{-2.4} \approx 0.09$  (lower than in Fig. 1);  $\delta^*$  seemed to have more effect than in Fig. 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 Fig. 1). Therefore, the models for DCT with 75% of the energy had in general more support vectors, i.e., were more complex (higher  $\nu$ ), had a shorter quadratic zone in the loss function (lower  $\delta^*$ ) 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 Fig. 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  $\alpha$  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 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  $\tau$  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  $\nu$  and  $\delta^*$  (see Fig. 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  $\nu$  (more support vectors), higher  $C$  (more penalty to errors) and lower  $\delta^*$  (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 Fig. 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 Fig. 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  $\tau$  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



**Fig. 7.** SVR-LIN+DCT: we show in (a) and (b) a zoom on the non-zero DCT coefficients. Both for  $E_{max}$  and  $\tau$  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  $\tau$  in image space. The predictors for  $E_{max}$  and  $\tau$  retain respectively 75% and 85% of the energy.

**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.

Method	Energy (%)	$E_{max}$ mmHg/ml		$\tau$ ms	
		#Coef	$\bar{E} \pm$ desv	#Coef	$\bar{E} \pm$ desv
RAW	100	1024	<b>0.919 ± 0.323</b>	1024	<b>9.7929 ± 4.7548</b>
DCT1	70	4	1.13 ± 0.501	4	10.989 ± 5.2658
DCT1	75	6	<b>0.949 ± 0.323</b>	5	11.128 ± 5.6244
DCT1	80	9	0.994 ± 0.369	7	11.519 ± 6.0716
DCT1	85	15	1.05 ± 0.352	11	<b>9.9107 ± 4.0324</b>
DCT1	90	37	0.958 ± 0.318	23	10.121 ± 4.4954
DCT1	95	176	1.10 ± 0.515	110	10.513 ± 4.6229
DCT1	99	760	1.37 ± 0.802	698	12.626 ± 6.0274
DCT1	99.5	883	1.36 ± 0.838	849	12.899 ± 5.9656
DCT1	99.9	993	1.31 ± 0.868	985	13.11 ± 5.713
DCT2	70	4	1.1027 ± 0.5197	4	11.0601 ± 5.3992
DCT2	75	6	0.9393 ± 0.3372	5	11.0699 ± 5.5055
DCT2	80	9	0.9666 ± 0.3697	7	11.1006 ± 5.7420
DCT2	85	15	0.9736 ± 0.3462	11	10.4328 ± 4.5058
DCT2	90	37	0.9252 ± 0.3169	23	10.0855 ± 4.8154
DCT2	95	176	0.9251 ± 0.3237	110	<b>10.0439 ± 4.8735</b>
DCT2	99	760	0.9232 ± 0.3237	698	10.0427 ± 4.9079
DCT2	99.5	883	<b>0.9222 ± 0.3233</b>	849	10.0464 ± 4.9098
DCT2	99.9	993	0.9241 ± 0.3242	985	10.0700 ± 4.8990

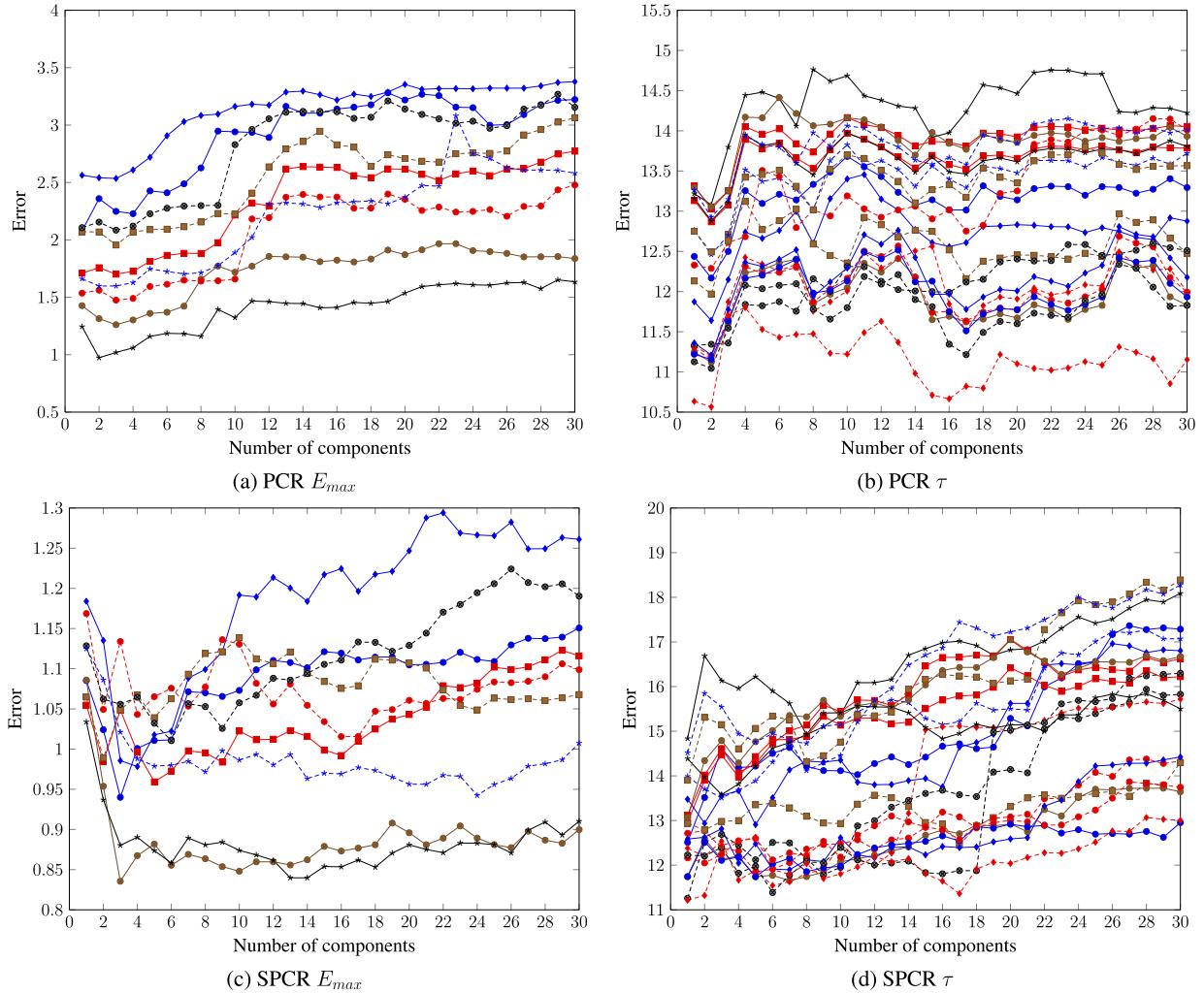
and the residuals were similar to the ones resulted when using linear SVR applied to raw images (see Fig. 2).

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  $\tau$ . 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  $\tau$ ) of the energy using only 6 (11 for  $\tau$ ) 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 method-

ology 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 Fig. 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 Fig. 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 Fig. 9. We



**Fig. 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  $E_{max}$ .

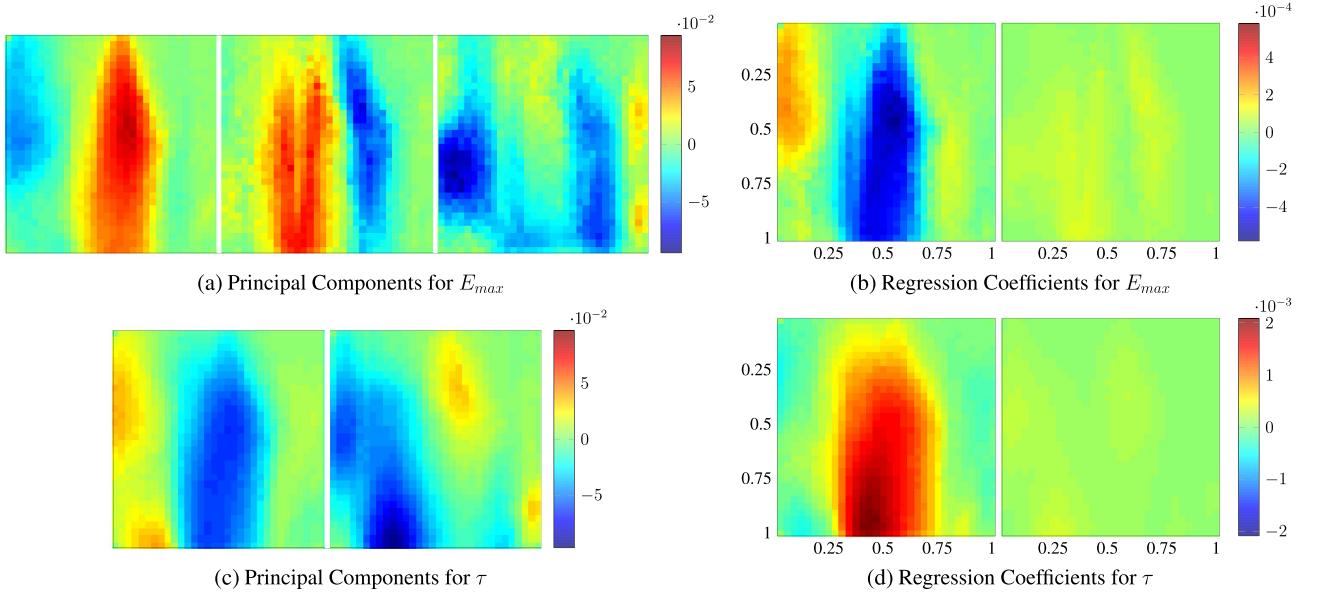
observe that the predictor was dominated by the first principal component. In addition, the predictor had some similarities with the Fig. 4 of SVR on raw predictions, both cases had a clear negative zone on the mid-ejection time, and a clear positive zone in the upper left of the figure. In addition, 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 Fig. 7, and their performance was almost the same.

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

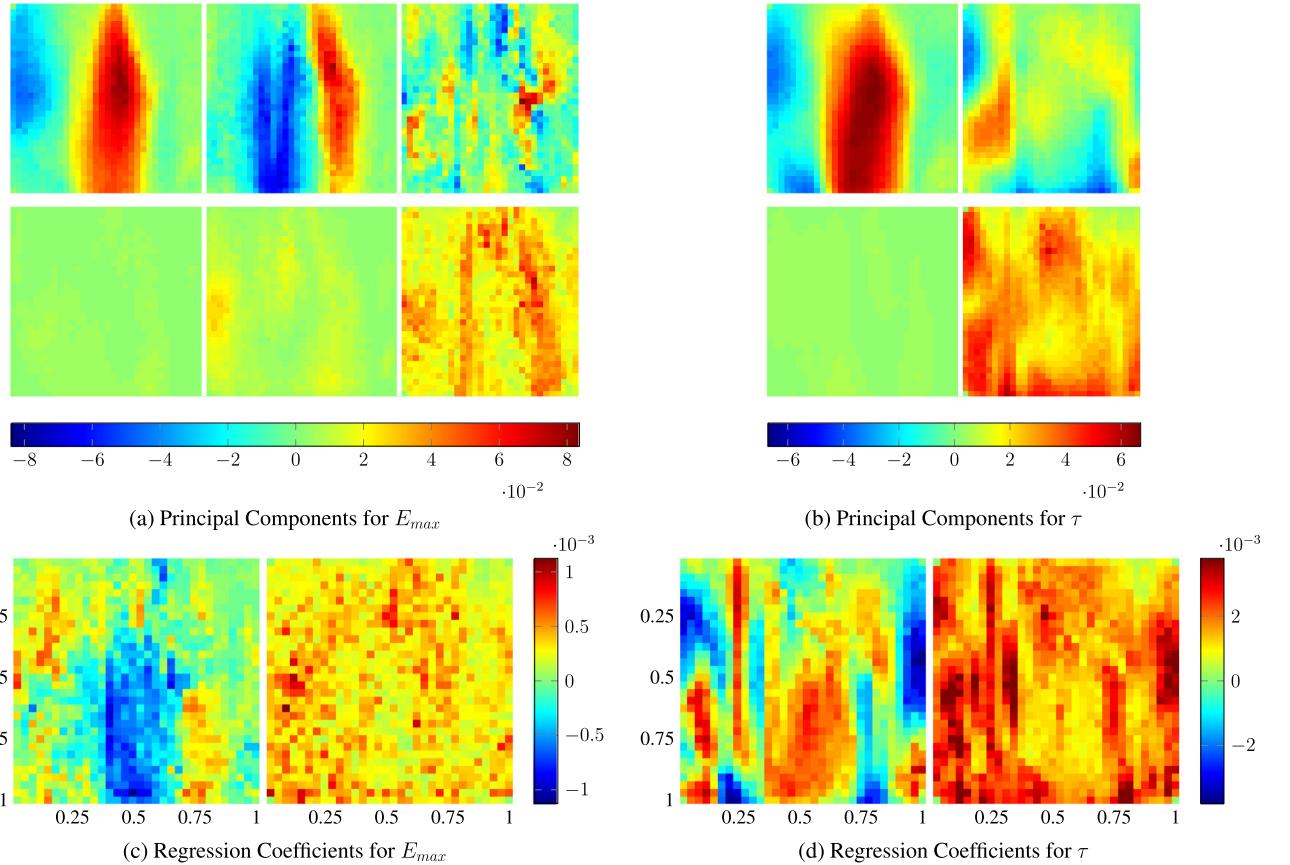
Another possibility to select the principal components uses the correlation with the variable to estimate (SPCR). Following this approach, we show the search for best number of components in Fig. 8(c). Here we observe that the selected number of components ranged from 2 to 24, being higher than 10 in 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 components extracted from the data to train each individual predictor are shown in Fig. 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 Fig. 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 Fig. 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 SPCR when compared to PCR was due to overestimations.

Later, we considered the prediction of  $\tau$  with PCR. We show the search for the optimum number of components in Fig. 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 Fig. 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 Fig. 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



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



**Fig. 10.** SPCR on raw CDMM images: (a) and (b) are the principal components for  $E_{max}$  and  $\tau$ , 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  $\tau$ .

previous cases, the residuals had heavier tails than DCT+SVR-LIN and RAW+SVR-LIN.

We also considered the results of predicting  $\tau$  with SPCR. The search for optimum coefficients is shown in Fig. 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 compo-

nents extracted from the data to train each individual predictor are shown in Fig. 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 is shown in Fig. 10, which was similar to the one provided by RAW+SVR-LIN (see Fig. 4). The prediction, residuals, and BA plot (not shown) were similar to the previous ones, but the predic-

**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.

Animal	PCR		SPCR		Difference	
	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	3.14e-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

tion 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 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  $\tau$ . 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 [Figs. 9 and 10](#), where we see that first principal component for  $\tau$  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 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 IV indices from raw CDMM images with PLS using the same methodology of previous sections. We first considered the prediction of  $E_{max}$ . [Fig. 11](#) shows the search for the optimal number of dimensions, which turned

**Table 4**

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

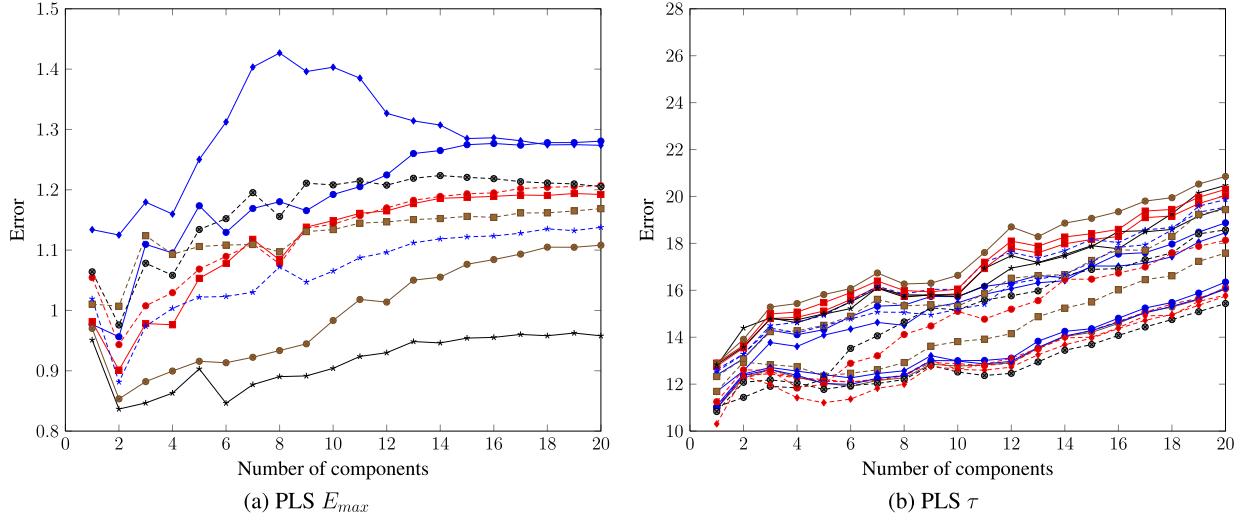
Animal	PCR		SPCR		Difference	
	Comps	MAE	Comps	MAE	Subs	%
1	2	4.99	1	4.50	0.490	9.82
2	2	7.74	1	7.42	0.322	4.15
3	2	3.78	7	5.64	-1.86	-49.3
4	2	4.93	1	10.3	-5.38	-109
5	2	14.2	5	12.4	1.78	12.5
6	2	8.96	5	10.2	-1.26	-14.0
7	2	14.5	17	16.2	-1.70	-11.8
8	1	25.0	1	23.3	1.61	6.46
9	2	14.4	3	12.4	2.02	14.0
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

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 [Fig. 12](#). The first loading was almost the same for all predictors, the second loading was similar for all predictors with little variations. The mean of the predictors for the 9 individuals is shown in [Fig. 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 [Fig. 13](#). Prediction is similar to the one provided by SVR-LIN (see [Fig. 2](#)) but slightly higher error, as we see one sample above 6 mmHg/ml in the BA plot.

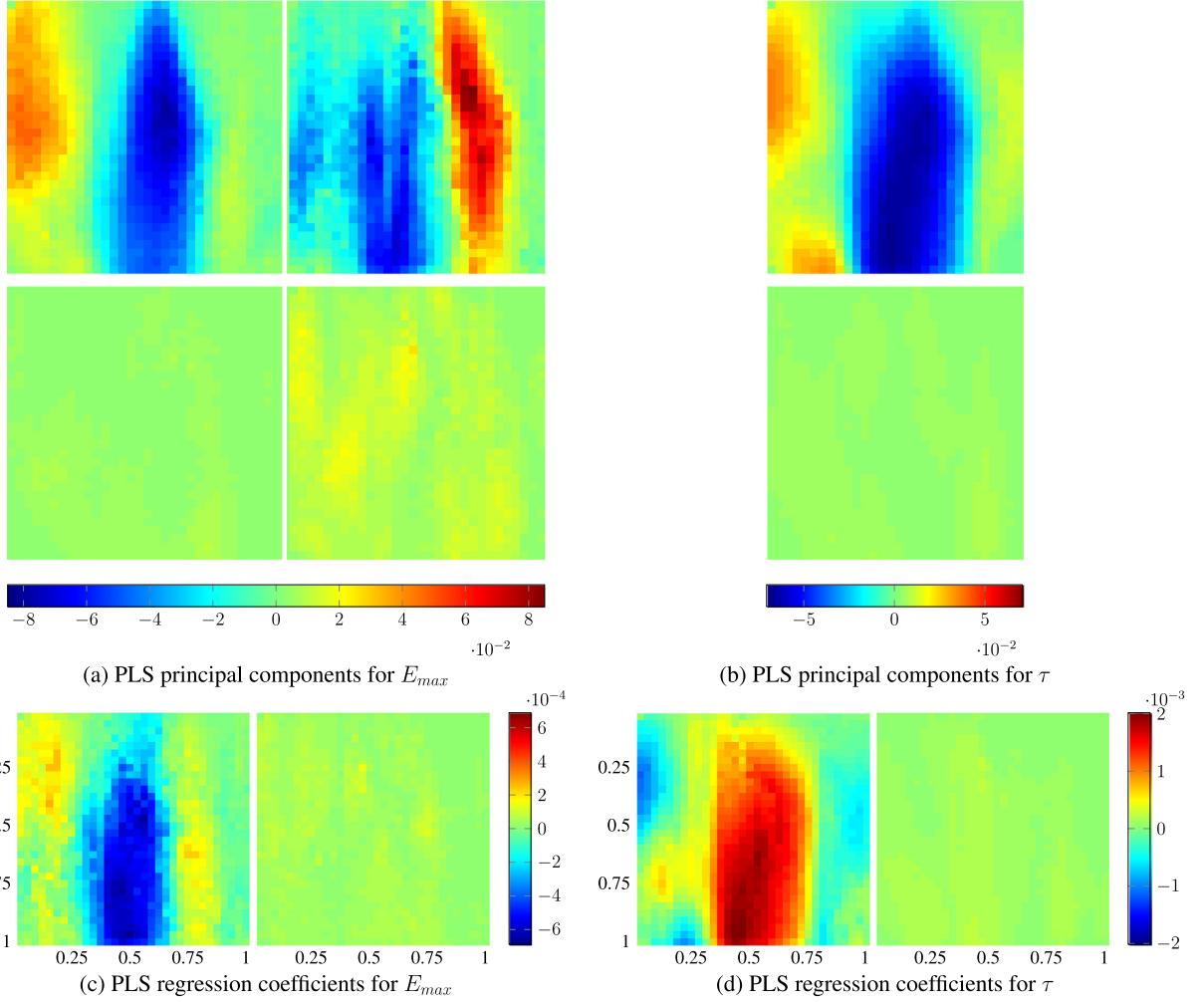
The  $\tau$  case was similar, [Fig. 11](#) shows the search for the optimal number of dimensions, which turned to be one for all cases. The mean of the loadings is shown in [Fig. 12](#), where we can see a great agreement for all individuals. The same happened with the predictors. The mean of all predictors was similar to the one obtained by PCR (see [Fig. 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 variance was untouched to not distort the images. The results of the prediction of  $E_{max}$  and  $\tau$  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



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

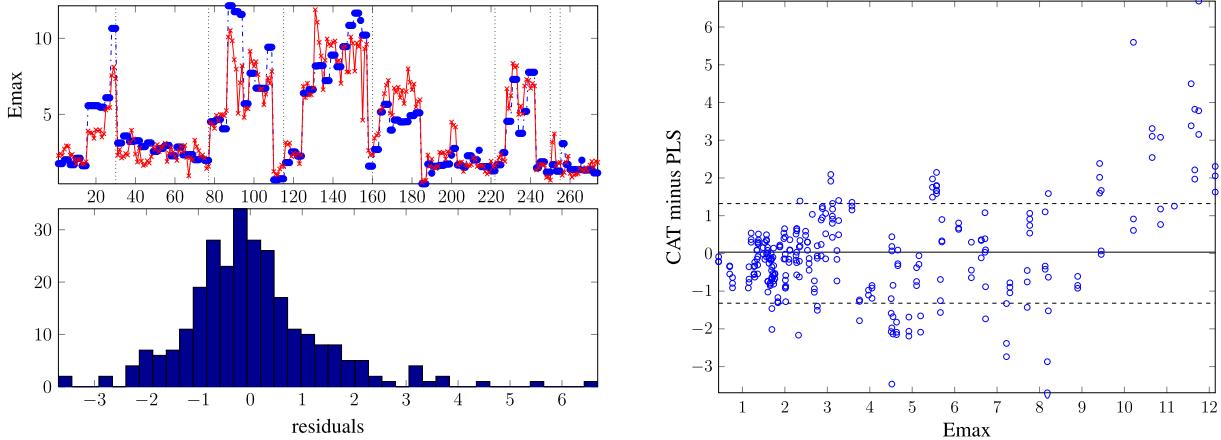


**Fig. 12.** PLS on raw CDMM images: (a) and (b) are the principal components for  $E_{max}$  and  $\tau$ , first row is the mean for all the individuals and second row standard deviation. (c) and (d) are the mean and standard deviation of the individual's predictors for  $E_{max}$  and  $\tau$ .

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.  $\tau$  prediction, which is harder, showed that DCT pre-filtering improved the 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 con-



**Fig. 13.** PLS: prediction and BA plot,  $E_{max}$ .

**Table 5**

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

Energy	PCR		PLS	
	$E_{max}$	$\tau$	$E_{max}$	$\tau$
	$\bar{E} \pm \text{desv}$	$\bar{E} \pm \text{desv}$	$\bar{E} \pm \text{desv}$	$\bar{E} \pm \text{desv}$
70	1.1804 ± 0.5023	11.0864 ± 5.4807	1.1522 ± 0.4576	10.7368 ± 5.2904
75	1.0317 ± 0.3788	<b>10.7564 ± 5.4532</b>	0.9621 ± 0.3146	10.7363 ± 5.2306
80	1.0226 ± 0.3606	10.9202 ± 5.1626	0.9500 ± 0.3125	10.7546 ± 5.2493
85	0.9700 ± 0.3219	11.1555 ± 5.5242	0.9548 ± 0.3344	<b>10.5215 ± 4.7425</b>
90	<b>0.9459 ± 0.2965</b>	11.0081 ± 5.5590	0.9310 ± 0.3147	10.7020 ± 5.3114
95	0.9634 ± 0.3142	10.9995 ± 5.5662	0.9329 ± 0.3163	10.6818 ± 5.3233
99	0.9467 ± 0.2980	11.0002 ± 5.5690	0.9274 ± 0.3167	10.6805 ± 5.3258
99.5	0.9466 ± 0.2980	11.0003 ± 5.5692	0.9271 ± 0.3166	10.6804 ± 5.3259
99.9	0.9465 ± 0.2980	11.0004 ± 5.5693	0.9269 ± 0.3166	10.6804 ± 5.3260
100	0.9465 ± 0.2980	11.0005 ± 5.5693	<b>0.9268 ± 0.3165</b>	10.6805 ± 5.3261

sider the performance of  $E_{max}$  estimators using raw CDMM images ([Table 1](#) vs [Table 6](#)), we observe that linear methods outperformed 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  $\tau$  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  $\tau$  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. [Fig. 14](#) compares RAW+SVM-LIN vs DCT2+SVM-RBF. We observe that nonlinear method did much better in the case of  $\tau$ ; 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 better than nonlinear SVM for high dimensional input spaces like text categorization [25] or fMRI data [4,9,12]. This is supported by two theoretical facts: Cover's Theorem [3] and Vapnik–Chervonenkis VC-dimension [22]. Cover's The-

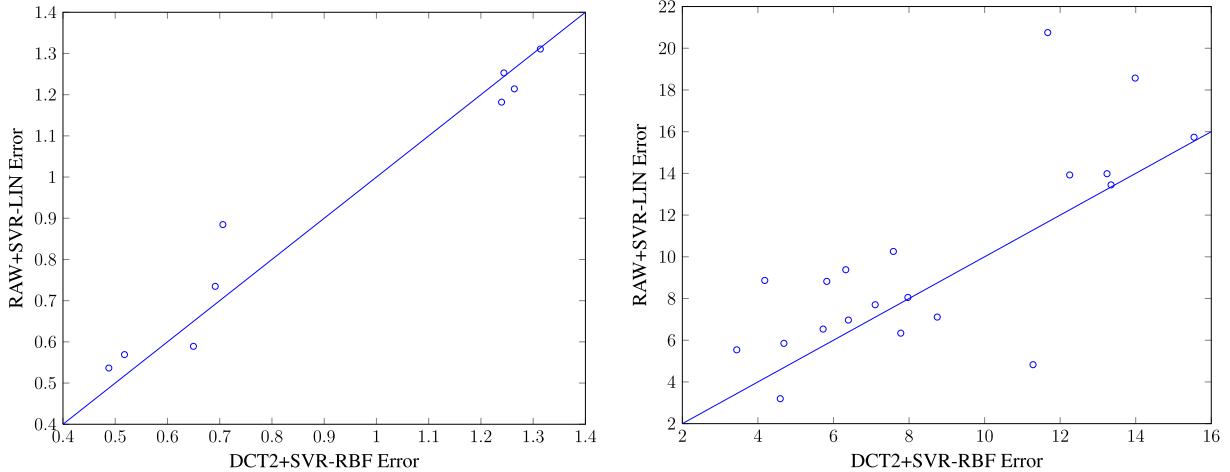
**Table 6**

Experiment 6: Nonlinear methods comparison.

Method	Energy (%)	$E_{max}$ mmHg/ml	$\tau$ ms
		$\bar{E} \pm \text{desv}$	$\bar{E} \pm \text{desv}$
RAW+SVM-RBF		0.94189 ± 0.39114	8.8799 ± 3.7268
RAW+SVM-RBF <sup>a</sup>		0.96978 ± 0.38364	8.5903 ± 3.8616
RAW+KPCR		1.0201 ± 0.32389	9.815 ± 3.8849
RAW+KSPCR		0.96864 ± 0.42345	9.1868 ± 3.2948
RAW+KPLS		0.94409 ± 0.3352	9.4474 ± 3.4883
DCT1+SVM-RBF	70	1.1867 ± 0.70409	10.052 ± 3.9406
DCT1+SVM-RBF	75	0.93137 ± 0.37818	9.8281 ± 4.1519
DCT1+SVM-RBF	80	0.98092 ± 0.34885	9.415 ± 3.4601
DCT1+SVM-RBF	85	0.98462 ± 0.37802	9.0974 ± 3.4904
DCT1+SVM-RBF	90	1.0793 ± 0.43061	9.3801 ± 3.5928
DCT1+SVM-RBF	95	1.2822 ± 0.61532	9.4047 ± 3.9132
DCT1+SVM-RBF	99	1.4617 ± 0.74496	9.9832 ± 3.4502
DCT1+SVM-RBF	99.5	1.425 ± 0.74683	10.149 ± 3.4265
DCT1+SVM-RBF	99.9	1.4263 ± 0.71899	10.061 ± 3.702
DCT2+SVM-RBF	70	1.0891 ± 0.66012	9.8645 ± 3.644
DCT2+SVM-RBF	75	0.92784 ± 0.36901	9.7182 ± 3.8304
DCT2+SVM-RBF	80	0.94927 ± 0.38176	9.2847 ± 3.3639
DCT2+SVM-RBF	85	0.94783 ± 0.36022	<b>8.5838 ± 3.693</b>
DCT2+SVM-RBF	90	<b>0.90155 ± 0.35303</b>	8.7809 ± 3.4463
DCT2+SVM-RBF	95	0.92159 ± 0.36189	8.9537 ± 3.4956
DCT2+SVM-RBF	99	0.94251 ± 0.40221	8.8741 ± 3.5639
DCT2+SVM-RBF	99.5	0.94772 ± 0.40618	8.7835 ± 3.5932
DCT2+SVM-RBF	99.9	0.94492 ± 0.40414	8.777 ± 3.615

<sup>a</sup> 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.

orem 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



**Fig. 14.** Comparison of best linear method against best nonlinear method mean absolute error for each individual. (Left)  $E_{max}$  result, (right)  $\tau$  (result). Better nonlinear performance is for points above diagonal.

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  $\infty$ . 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  $\tau$ . 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 [2]. 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 [22], 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 heteroscedasticity** 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  $\tau$ , the heteroscedasticity is considerable, rendering the method unable to accurately estimate high values of  $\tau$ . 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  $\tau$  estimation [13].

**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  $\tau$  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  $\tau$ , where their estimation capabilities are really close to SVR-LIN (3% more error for  $E_{max}$  and 1% for  $\tau$ ). DCT+SVR-LIN estimators use much less coefficients (6 for  $E_{max}$ , 11 for  $\tau$ ), therefore, it could be worthy to interpret their estimators instead of SVR-LIN ones, especially for  $\tau$ , 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  $\tau$ , 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  $\tau$  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  $\tau$  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 cannot 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. [11,21]). 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  $\tau$  estimation provided by RAW+SVR-LIN have heavy-tails (see Fig. 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  $\tau$ . 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.

## 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 RD12/0042 (Red de Investigación Cardiovascular), 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.

## Appendix A. Supplementary material

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.dsp.2014.12.012>.

## References

- [1] J. Bermejo, J.C. Antoranz, R. Yotti, M. Moreno, M.A. García-Fernández, Spatio-temporal mapping of intracardiac pressure gradients. A solution to Euler's equation from digital postprocessing of color Doppler M-mode echocardiograms, Ultrasound Med. Biol. 27 (2001) 621–630, [http://dx.doi.org/10.1016/S0301-5629\(01\)00349-0](http://dx.doi.org/10.1016/S0301-5629(01)00349-0).
- [2] V. Cherkassky, X. Shao, F.M. Mulier, V.N. Vapnik, Model complexity control for regression using VC generalization bounds, IEEE Trans. Neural Netw. 10 (1999) 1075–1089, <http://dx.doi.org/10.1109/72.788648>.
- [3] T.M. Cover, Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, IEEE Trans. Electron. Comput. EC-14 (1965) 326–334, <http://dx.doi.org/10.1109/PGEC.1965.264137>.
- [4] D.D. Cox, R.L. Savoy, Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex, NeuroImage 19 (2003) 261–270, [http://dx.doi.org/10.1016/S1053-8119\(03\)00049-1](http://dx.doi.org/10.1016/S1053-8119(03)00049-1).
- [5] A.S. Hadi, R.F. Ling, Some cautionary notes on the use of principal components regression, Am. Stat. 52 (1998) 15–19, <http://dx.doi.org/10.1080/00031305.1998.10480530>.
- [6] A. Höskuldsson, PLS regression methods, J. Chemom. 2 (1988) 211–228, <http://dx.doi.org/10.1002/cem.1180020306>.
- [7] I.T. Jolliffe, A note on the use of principal components in regression, J. R. Stat. Soc., Ser. C, Appl. Stat. 31 (1982) 300–303, <http://dx.doi.org/10.2307/2348005>.
- [8] S. de Jong, SIMPLS: An alternative approach to partial least squares regression, Chemom. Intell. Lab. Syst. 18 (1993) 251–263, [http://dx.doi.org/10.1016/0169-7439\(93\)85002-X](http://dx.doi.org/10.1016/0169-7439(93)85002-X).
- [9] S. LaConte, S. Strother, V. Cherkassky, J. Anderson, X. Hu, Support vector machines for temporal classification of block design fMRI data, NeuroImage 26 (2005) 317–329, <http://dx.doi.org/10.1016/j.neuroimage.2005.01.048>.
- [10] W.F. Massy, Principal components regression in exploratory statistical research, J. Am. Stat. Assoc. 60 (1965) 234–256, <http://dx.doi.org/10.2307/2283149>.
- [11] J. Mourão Miranda, A.L.W. Bokde, C. Born, H. Hampel, M. Stetter, Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data, NeuroImage 28 (2005) 980–995, <http://dx.doi.org/10.1016/j.neuroimage.2005.06.070>.
- [12] M. Misaki, Y. Kim, P.A. Bandettini, N. Kriegeskorte, Comparison of multivariate classifiers and response normalizations for pattern-information fMRI, NeuroImage 53 (2010) 103–118, <http://dx.doi.org/10.1016/j.neuroimage.2010.05.051>.
- [13] J.L. Rojo-Alvarez, J. Bermejo, V.M. Juarez-Caballero, R. Yotti, C. Cortina, M.A. Garcia-Fernandez, J.C. Antoranz, Support vector analysis of color-Doppler images: a new approach for estimating indices of left ventricular function, IEEE Trans. Med. Imaging 25 (2006) 1037–1043, <http://dx.doi.org/10.1109/TMI.2006.875437>.
- [14] R. Rosipal, M. Girolami, L.J. Trejo, On kernel principal component regression with covariance inflation criterion for model selection, Technical report, Department of Computing and Information Systems, University of Paisley, 2001.
- [15] R. Rosipal, M. Girolami, L.J. Trejo, A. Cichocki, Kernel PCA for feature extraction and de-noising in nonlinear regression, Neural Comput. Appl. 10 (2001) 231–243, <http://dx.doi.org/10.1007/s521-001-8051-z>.
- [16] R. Rosipal, N. Krämer, Overview and recent advances in partial least squares, in: C. Saunders, M. Grobelnik, S. Gunn, J. Shawe-Taylor (Eds.), Subspace, Latent Structure and Feature Selection, in: Lecture Notes in Computer Science, vol. 3940, Springer, Berlin, Heidelberg, 2006, pp. 34–51, <http://dx.doi.org/10.1007/11752790>.
- [17] R. Rosipal, L.J. Trejo, Kernel partial least squares regression in reproducing kernel Hilbert space, J. Mach. Learn. Res. 2 (2001) 97–123.
- [18] B. Schölkopf, A. Smola, Learning with Kernels, MIT Press, Cambridge, MA, USA, 2001.
- [19] B. Schölkopf, A.J. Smola, K.R. Müller, Kernel principal component analysis, in: B. Schölkopf, C.J.C. Burges, A.J. Smola (Eds.), Advances in Kernel Methods—Support Vector Learning, MIT Press, Cambridge, MA, 1999, pp. 327–352.
- [20] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222, <http://dx.doi.org/10.1023/B:STCO.0000035301.49549.88>.

- [21] S. Song, Z. Zhan, Z. Long, J. Zhang, L. Yao, Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data, *PLoS ONE* 6 (2011) e17191, <http://dx.doi.org/10.1371/journal.pone.0017191>.
- [22] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
- [23] H. Wold, Nonlinear estimation by iterative least squares procedures, in: F. David (Ed.), *Research Papers in Statistics*, Wiley, New York, 1966, pp. 411–444.
- [24] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.* 58 (2001) 109–130, [http://dx.doi.org/10.1016/S0169-7439\(01\)00155-1](http://dx.doi.org/10.1016/S0169-7439(01)00155-1).
- [25] Y. Yang, An evaluation of statistical approaches to text categorization, *Inf. Retr.* 1 (1999) 69–90, <http://dx.doi.org/10.1023/A:1009982220290>.
- [26] R. Yotti, J. Bermejo, Y. Benito, J.C. Antoranz, M.M. Desco, D. Rodríguez-Pérez, C. Cortina, T. Mombiela, A. Barrio, J. Elízaga, F. Fernández-Avilés, Noninvasive estimation of the rate of relaxation by the analysis of intraventricular pressure gradients, *Circ. Cardiovasc. Imaging* 4 (2011) 94–104, <http://dx.doi.org/10.1161/CIRCIMAGING.110.960369>.
- [27] R. Yotti, J. Bermejo, M.M. Desco, J.C. Antoranz, J.L. Rojo-Álvarez, C. Cortina, C. Allué, H. Rodríguez-Abella, M. Moreno, M.A. García-Fernández, Doppler-derived ejection intraventricular pressure gradients provide a reliable assessment of left ventricular systolic chamber function, *Circulation* 112 (2005) 1771–1779, <http://dx.doi.org/10.1161/CIRCULATIONAHA.104.485128>.

**Ricardo Santiago-Mozos** received the Ph.D. degree in signal processing and communications from the Carlos III University of Madrid, Madrid, Spain, in 2009. He has just finished a Postdoctoral Fellowship of the Spanish Program for Recruitment and Incorporation of Human Resources at the University Rey Juan Carlos of Madrid, Madrid, Spain. He has been a Spanish Foundation for Science and Technology Postdoctoral Fellow at the Machine Learning and Data Mining Group in National University of Ireland, Galway, Ireland, where he had also completed a Marie Curie Fellowship. His research interests include machine learning and medical applications and he has coauthored more than 20 papers appearing in refereed journals and conference proceedings.

**José Luis Rojo-Álvarez** (born in Bilbao, Spain, 1972) received the Telecommunication Engineering Degree in 1996 from University of Vigo, Spain, and the Ph.D. in Telecommunication in 2000 from the Polytechnical University of Madrid, Spain. Since 2006, he has been an Associate Professor in the Department of Signal Theory and Communications, University Rey Juan Carlos, Madrid, Spain. He has published more than 80 papers in JCR journals and more than 140 (inter)national conference communications. He has participated in more than 50 projects (with public and private fundings), and directed more than 10 of them, including several actions in the National Plan for Research and Fundamental Science. He is a Senior Researcher at the Prometeo program in Ecuador (Army University).

His main research interests include statistical learning theory, digital signal processing, and complex system modelling, with applications to digital communications and to cardiac signal and image processing.

**J. Carlos Antoranz** graduated in Physics in 1977 at Universidad Autónoma Madrid and obtained his Ph.D. in 1982 at Spanish Open University. He works as Physics Full Professor since 1996 at UNED. He has been visiting Professor at University of Oslo, Norway (Mathematics), Kysuhu University, Japan (Physics), IBM Research Center in Kingston, USA, Non-linear Center at Los Alamos National Laboratory, USA, Institute de Mécanique des Fluides, Marseille, France. He has published more than 200 papers.

**Mar Desco** received her Ph.D. in Medicine and Surgery from Universidad Complutense de Madrid. Her interests include anatomical pathology and circulatory mechanic assistance. She has published more than 10 journal articles.

**Daniel Rodríguez-Pérez** received his M.Sc. (1998) and Ph.D. (2005) in Physics from UNED. He is a Senior Lecturer at UNED Medical Physics Master. He works, among other areas, in mathematical and physical models of biomedical signals and processes. He has coauthored more than 20 journal papers.

**Raquel Yotti** (Madrid 1973) is a clinical cardiologist with expertise in cardiac imaging that coordinates the Complex Structural Heart Disease Program at the Hospital General Universitario Gregorio Marañón. She obtained the Ph.D. degree in Medicine in 2006 from the Universidad Complutense Madrid. Her research has dedicated to the development and clinical application of new diagnostic tools based on cardiac imaging, to improve the characterization of physiology and hemodynamics in patients with different cardiovascular diseases. She is author of more than 30 articles in this field.

**Javier Bermejo** (London, 1966) is the Director of Noninvasive Cardiology and Cardiovascular Imaging of the Hospital General Universitario Gregorio Marañón, Madrid, Spain, and Associate Professor of Medicine in the Universidad Complutense de Madrid in 2013. His research has focused on the invasive and noninvasive assessment of cardiac hemodynamics. He leads a multidisciplinary research team integrated by clinical cardiologists, surgeons, engineers and physicists. He has received specific training in a number of fields of biomedical engineering, including advanced biostatistics, computer algorithm programming, image processing, and basic biomechanics. Dr. Bermejo is author of more than 80 articles in the field and external reviewer for major scientific societies and more than 20 scientific journals. Current areas of research are focused on imaging-derived biomechanics in the field of heart failure and valvular heart disease.