

# A fast all-in-one method for automated post-processing of PIV data

Damien Garcia

Received: 6 November 2009 / Revised: 17 September 2010 / Accepted: 23 September 2010 / Published online: 10 October 2010  
© Springer-Verlag 2010

**Abstract** Post-processing of PIV (particle image velocimetry) data typically contains three following stages: validation of the raw data, replacement of spurious and missing vectors, and some smoothing. A robust post-processing technique that carries out these steps simultaneously is proposed. The new all-in-one method (DCT–PLS), based on a penalized least squares approach (PLS), combines the use of the discrete cosine transform (DCT) and the generalized cross-validation, thus allowing fast unsupervised smoothing of PIV data. The DCT–PLS was compared with conventional methods, including the normalized median test, for post-processing of simulated and experimental raw PIV velocity fields. The DCT–PLS was shown to be more efficient than the usual methods, especially in the presence of clustered outliers. It was also demonstrated that the DCT–PLS can easily deal with a large amount of missing data. Because the proposed algorithm works in any dimension, the DCT–PLS is also suitable for post-processing of volumetric three-component PIV data.

## 1 Introduction

Non-intrusive velocity measurement has become a sine qua non for quantitative analysis in several domains involving

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s00348-010-0985-y) contains supplementary material, which is available to authorized users.

---

D. Garcia (✉)  
LBUM, CRCHUM, Research Centre,  
University of Montreal Hospital, 2099 Alexandre de Sève,  
room Y-1619, Montreal, QC H2L 2W5, Canada  
e-mail: Damien.Garcia@crchum.qc.ca;  
Damien.Garcia@BiomeCardio.com  
URL: www.biomecardio.com

complex fluid dynamics. Particle image velocimetry (PIV) and phase-contrast magnetic resonance imaging (PC-MRI) are regarded as the most reliable non-invasive techniques to capture time-resolved velocity information of large flow fields. Although PC-MRI is widely devoted to the study of cardiovascular hemodynamics (Vennemann et al. 2007), PIV is the most popular velocity measurement technique and is extensively used in manifold areas as aerodynamics, hydraulics, biology, combustion and microfabricated systems (Raffel et al. 2007). This technique can provide two- or three-dimensional (2-D or 3-D) vectorial representations of the flow with high accuracy on an evenly gridded region of interest. PIV data, however, are almost always subject to background Gaussian noise, outliers and missing data, which not only alter the visualization of the velocity distribution and streamlines but also greatly affect flow quantities based on differentiation such as strain, vorticity and vortex identification. Post-processing algorithms are therefore necessarily required before any analysis of the information offered by the PIV technique.

Post-processing of PIV data usually consists in three consecutive steps: (1) validation of the data, i.e. detection of the outliers, (2) replacement of the incorrect and missing values, and (3) data smoothing. Outlier identification is the most critical procedure and has been the subject matter of several recent papers (Hart 2000; Liang et al. 2003; Liu et al. 2008; Pun et al. 2007; Shinneebe et al. 2004; Westerweel and Scarano 2005). The most common validation techniques proposed in standard commercial software are the global histogram filter, the dynamic mean value operator and the normalized median test (Raffel et al. 2007). Alternative processes based on cellular neural network or bootstrapping have also been shown to provide satisfactory detections but they are relatively time-consuming (Liang et al. 2003; Pun et al. 2007) and rarely used in practice. Most of the current

outlier detectors require the use of thresholding parameters whose selection remains somewhat arbitrary. An inappropriate threshold may lead to undetected spurious vectors or, on the contrary, may eliminate data that are actually not invalid. Westerweel demonstrated that the normalized median test can get round this drawback by using a single threshold value that has been shown to be suitable for a large range of PIV configurations (Westerweel and Scarano 2005). Once most of the outliers have been identified, they are generally replaced by using a median, bilinear or spline interpolation. Because experimental noise also alters the PIV data, the whole velocity field is finally smoothed using a small  $2 \times 2$  or  $3 \times 3$  average kernel before any differential operations (Raffel et al. 2007). As an alternative approach to the conventional methods, kriging interpolation has recently been successfully applied to PIV data and it has been shown that Gaussian ordinary kriging smoothes and fills gaps efficiently while being slightly sensitive to outliers (Gunes and Rist 2007). The kriging procedure requires the use of a so-called variogram parameter that somewhat regulates the amount of smoothness. The appropriate choice of such variogram, however, must be based on prior data set acquired in similar conditions.

To reduce the above-mentioned limitations of the current algorithms, a fast all-in-one robust process allowing simultaneous validation, replacement and smoothing of PIV data is described. According to (Westerweel 1994), a typical 2-D PIV velocity field can be expressed as:

$$V_{kl} = \gamma_{kl}(\mathbf{V}_{kl}^0 + \varepsilon_{kl}) + (1 - \gamma_{kl}) \mathbf{U}_{kl}, \quad (1)$$

where the subscript  $(k, l)$  refers to the position in the PIV flow field,  $\gamma$  is a binary discrete random variable that takes a value of 0 or 1 and  $\mathbf{V}^0$  is the actual velocity field to be sought. During PIV measurements, this velocity field is marred by a Gaussian experimental noise ( $\varepsilon$ ), with zero mean and unknown variance, and randomly distributed spurious vectors ( $\mathbf{U}$ ). The technique described herein can provide a reliable estimate of  $\mathbf{V}^0$  (noted  $\hat{\mathbf{V}}$ ) using a single automated process. The proposed method based on a penalized least squares approach, combined with the discrete cosine transform (DCT), allows automatic smoothing of multidimensional data that may include outlying and missing data (Garcia 2010b). In this manuscript, we demonstrate the effectiveness of the DCT-based penalized least squares method, acronymed DCT-PLS, for post-processing of experimental PIV data.

## 2 The DCT-based penalized least squares method

The DCT-based penalized least squares (DCT-PLS) method has been thoroughly described for evenly spaced data in one and higher dimensions (Garcia 2010b). Because

typical PIV data are equally gridded, the DCT-PLS can be efficiently applied to such data, as briefly introduced in the present section. For simplicity, only planar two-component velocity fields will be considered in this study although the method can be immediately applied to volumetric three-component velocity fields (see Fig. 13 for one example).

### 2.1 Automatic smoothing with the DCT-PLS

Let us write the 2-D raw  $m \times n$  velocity field  $\mathbf{V} = (V_x, V_y)$ , where  $V_x$  and  $V_y$  are the usual Cartesian velocity components. Assuming first that the PIV data are corrupted by experimental noise only (i.e. no outliers, no missing data), the DCT-PLS yields a smoothed velocity field  $\hat{\mathbf{V}}$  from the following equation (Garcia 2010b):

$$\hat{\mathbf{V}} = \text{IDCT2}(\Gamma \circ \text{DCT2}(\mathbf{V})), \quad (2)$$

where DCT2 and IDCT2 refer to the type II 2-D discrete cosine transform and inverse discrete cosine transform, respectively, and  $\circ$  stands for the Schur (elementwise) product. The filtering matrix  $\Gamma$  is defined by

$$\Gamma_{kl} = \left( 1 + s \left( 4 - 2 \cos \frac{(k-1)\pi}{m} - 2 \cos \frac{(l-1)\pi}{n} \right)^2 \right)^{-1}. \quad (3)$$

Equation 3 includes a real positive scalar  $s$  that controls the degree of smoothing: as the  $s$  parameter increases, the smoothing of  $\hat{\mathbf{V}}$  also increases. An unsuitable selection for  $s$  could lead to under- or over-smoothed velocities. To avoid any subjectivity in the choice of the amount of smoothness, this parameter can be determined by minimizing the so-called generalized cross-validation (GCV) score (Craven and Wahba 1978). In the particular case of the DCT-PLS, the GCV score can be simplified as demonstrated in (Garcia 2010b). Minimization of the GCV score helps to optimize the trade-off between bias and variance: the bias measures how well the smoothed velocity field  $\hat{\mathbf{V}}$  approximates the true velocity field  $\mathbf{V}^0$ , and the variance measures how well the original experimental data field  $\mathbf{V}$  can be estimated by  $\hat{\mathbf{V}}$ . Excellent results are expected if the additive noise ( $\varepsilon$ ) in Eq. 1 is Gaussian white (Wahba 1990). It has also been reported that the GCV remains fairly suitable even with non-homogenous variances and non-Gaussian errors (Wahba 1990). The GCV criterion makes the DCT-PLS fully automated and is very well adapted to PIV data as shown further in the next sections.

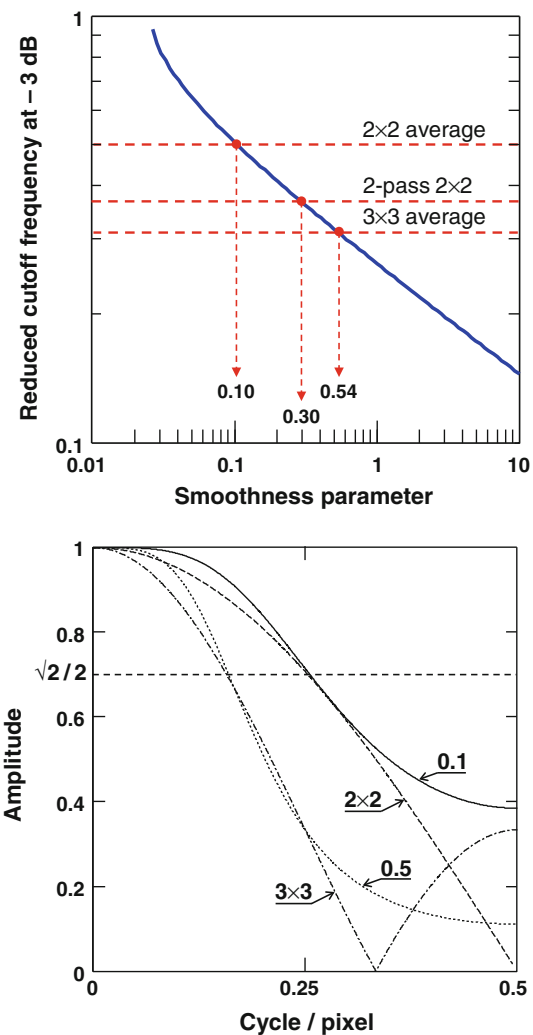
### 2.2 Spatial frequency analysis of the DCT-PLS: effect of the smoothness parameter

As stated in the latter paragraph, smoothness of the velocity field increases when  $s$  increases. When  $s$  runs from

0 to  $\infty$ , the output runs from the unchanged (original) data to the arithmetic mean of the velocities. In the spectral domain, increasing the smoothness parameter reduces the low-pass filter bandwidth. An exaggeratedly high  $s$  can lead to the loss of high spatial frequency components as may occur with turbulence or high shear flow. One thus has to ensure that the optimal  $s$ , i.e. the  $s$  value that minimizes the GCV score, prevents over-smoothing. Craven and Wahba have shown that the  $s$  value that minimizes the GCV score is also a fair minimizer of the mean square error (MSE) between the estimator and the actual value (Craven and Wahba 1978). In the context of PIV data, the optimal  $s$  thus provides the best guess  $\hat{V}$  of  $V^0$  (see Eq. 1) in the MSE sense. This assertion is essentially true with a Gaussian white noise. To better appreciate how  $s$  influences the smoothing process, Fig. 1 (top panel) compares the  $s$ -dependent cut-off frequencies at  $-3$  dB of the DCT-PLS with those of small kernel average smoothers. As shown in this figure, the DCT-PLS has a cut-off frequency identical to that of  $2 \times 2$ , 2-pass  $2 \times 2$  and  $3 \times 3$  kernel average smoothers for  $s = 0.1$ , 0.3 and 0.5, respectively. Using a similar approach as in (Foucaut and Stanislas 2002), the spatial frequency response (SFR) of the DCT-PLS was also determined for  $s = 0.1$  and 0.5 and compared with the SFR of  $2 \times 2$  and  $3 \times 3$  moving average filters (Fig. 1, bottom panel). Ideally,  $s$  should be smaller than 0.5 in order the DCT-PLS to be compatible with the conventional smoothing methods. It is worth mentioning, however, that the optimal  $s$  highly depends upon the variance of the additive noise. Optimal values larger than 0.5 are thus expected with low-SNR (signal-to-noise ratio) PIV data. In that case, the smoothness parameter can be set manually at a low level, say smaller than 0.5; one must be aware, however, that this can lead to under-smoothing if the PIV data have a low SNR.

### 2.3 Replacement of the outlying data with the DCT-PLS

Besides experimental noise, PIV data typically contain spurious vectors mostly due to the loss of correlation signal in case of insufficient particle seeding, out-of-plane motions or strong gradients. If neglected, the outlying vectors may make the smoothed PIV field abnormally distorted. To overcome this drawback, the DCT-PLS uses a robust procedure that is very little affected by the outliers. This method consists in constructing weights with a specified weighting function by using the current residuals and updating them, from iteration to iteration, until the residuals remain unchanged. It is worth noting that the DCT-PLS does not contain any “outlier detector” based on a specific predicate such as used in the normalized median



**Fig. 1** *Top panel:* Reduced cut-off frequency of the DCT-PLS (at  $-3$  dB) as a function of the smoothness parameter ( $s$ ). The cut-off frequencies of common average filters are also represented (dashed horizontal lines). *Bottom panel:* Spatial frequency responses (SFR) of the DCT-PLS for two  $s$  values (0.1 and 0.5). The SFR of the  $2 \times 2$  and  $3 \times 3$  average filters are depicted for comparison. The dashed horizontal line corresponds to the  $-3$  dB limit

test. The algorithm rather assigns a low weight to outlying values in order to reduce their influence. The robust version of DCT-PLS is given by the following implicit formula (see (Garcia 2010b) for more details):

$$\hat{V} = \text{IDCT2}(\Gamma \circ \text{DCT2}(W^{bs} \circ (V - \hat{V}) + \hat{V})), \quad (4)$$

This equation can be solved using a fixed point iteration with  $\hat{V} = V$  as initial conditions. The matrix  $W^{bs}$  contains the bisquare weights which depend upon the so-called studentized residuals (Heiberger and Becker 1992). In the case of the DCT-PLS, the studentized residuals can be extremely simplified to make the algorithm faster (Garcia 2010b).

## 2.4 Dealing with missing values and masks

Missing values can also occur due to masking materials, laser reverberations or defects in the CCD camera sensor. In addition, the region of interest (ROI) and the area enclosed by the PIV raw images do not always coincide. Missing vectors and areas to be masked thus need to be extracted before any post-processing. The DCT-PLS can easily cope with missing and masked data by introducing a second weight matrix,  $W^m$ , defined by  $W_{kl}^m = 0$  if the velocity in pixel  $(k, l)$  is masked or missing, 1 otherwise. Equation 4 thus becomes:

$$\hat{V} = \text{IDCT2}(\Gamma \circ \text{DCT2}(W^m \circ W^{bs} \circ (V - \hat{V}) + \hat{V})). \quad (5)$$

Equation 5 is the generalized equation of the robust DCT-PLS applied to PIV velocity data.

## 2.5 Matlab code

The use of Eq. 5 provides a post-processed velocity field by removing noise, missing data and outliers conjointly and automatically. Further details regarding the smoothing algorithm can be found in (Garcia 2010b). Because the procedure mostly involves DCTs, the numerical scheme can be made very fast. A Matlab code written with Matlab R2008b is given in the supplemental material (see “Appendix”). Updated versions can be downloaded from the author’s personal website (Garcia 2010a). The efficacy of the DCT-PLS is demonstrated in the following sections by post-processing simulated and experimental PIV fields. Several results obtained with the DCT-PLS are compared with those issued from the conventional post-processing methods.

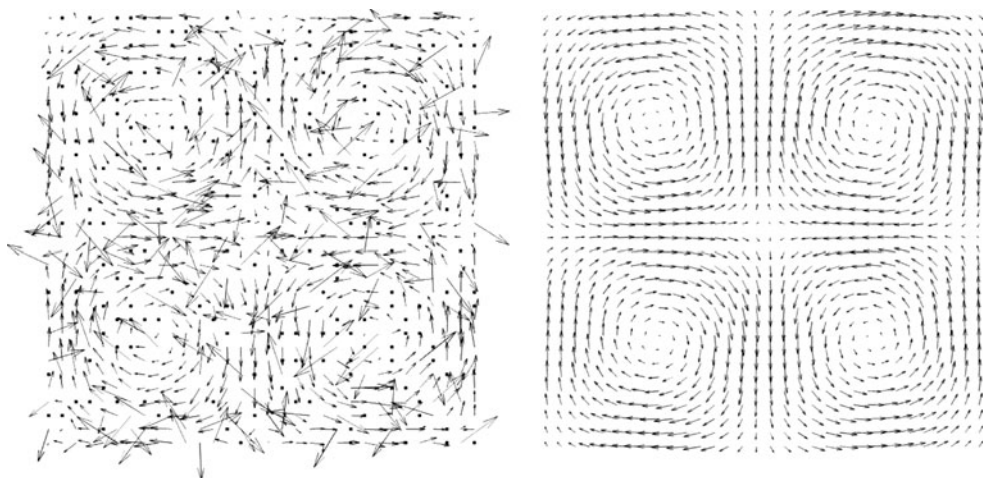
## 3 Processing of outlying and noisy velocities: conventional methods vs. DCT-PLS

Following Shinnee et al. (2004), and in order to quantify the accuracy of the proposed method, the DCT-PLS was tested on simulated vortical cellular flows with additional noise and outliers. The original  $m \times n$  velocity field ( $V^0$ ), containing four vortical cells, was defined by the following Cartesian components:

$$V_{kl}^0 = V_{\max} \left\{ \begin{aligned} &\sin\left(2\pi\frac{k-1}{m-1}\right) \cos\left(2\pi\frac{l-1}{n-1}\right); \\ &-\cos\left(2\pi\frac{k-1}{m-1}\right) \sin\left(2\pi\frac{l-1}{n-1}\right) \end{aligned} \right\}, \quad (6)$$

where  $V_{\max}$  stands for the maximal velocity. Figure 2 illustrates such a vortical flow in the particular case of  $(m, n) = (32, 32)$  which has been made exaggeratedly corrupt by the addition of random noise, spurious vectors and missing data (Fig. 2, left). Despite high deterioration of the velocity data, the DCT-PLS is clearly able to recover the imprint of the actual flow field (Fig. 2, right). To better represent real PIV measurements, numerical tests were performed on  $32 \times 32$  fields with random noise and spurious vectors as defined below. A value of 10 pixels/pulse-interval was chosen for the maximal velocity  $V_{\max}$ . Gaussian noise with mean zero and constant variance was first included on both x- and y-components of the original velocity field. Two variances were tested:  $(0.01 \times V_{\max})^2$  and  $(0.1 \times V_{\max})^2$ . Spurious (outlying) velocity vectors were then added using either scattered or clustered patterns (Shinnee et al. 2004). Spurious data were defined as vectors whose x- and y-components were drawn independently from a uniform distribution on the interval  $[-2V_{\max}, 2V_{\max}]$ . By way of example, two noisy synthetic

**Fig. 2** Vortical cellular flow ( $32 \times 32$ ) corrupted by Gaussian noise, outliers and missing data before (left) and after (right) smoothing by the DCT-PLS. The normalized root mean squared error (NRMSE, see Eq. 7) was 14% here



**Fig. 3** Vortical cellular flow ( $32 \times 32$ ) corrupted by a Gaussian noise of variance  $(0.01 \times V_{\max})^2$  and 15% of outlying vectors. Two configurations were tested in this study: scattered (*left*) and clustered (*right*) spurious vectors



PIV fields with 15% of spurious vectors in both scattered and clustered configurations are illustrated in Fig. 3. In the scattered configuration (Fig. 3, left), the spurious velocities were randomly and uniformly distributed throughout the whole field. In practice, however, spurious vectors often appear at the edges of the PIV field, near model surfaces, on low seeded areas, and on regions poorly or highly illuminated. Under these conditions, spurious vectors are found in clusters. In the simulated clustered organizations (Fig. 3, right), the groups were positioned randomly and their sizes were chosen from a Poisson distribution with mean six. Four scenarios were thus analyzed: 2 Gaussian noises  $\times$  2 spurious arrangements. For each scenario, the relative amount of spurious vectors was ranged from 0 to 25% with an increment of 2.5% which represented a total of  $4 \times 11 = 44$  cases. Finally, 100 PIV fields were created based on a Monte-Carlo method for each of these cases. The simulated PIV data were analyzed using both a conventional post-processing and the DCT-PLS described in this paper. The conventional method consisted in (1) validating the data by means of the normalized median test using 2.0 for the threshold and 0.1 for the noise level, as recommended (Raffel et al. 2007; Westerweel and Scarano 2005), (2) replacing the invalid vectors by the median of the  $3 \times 3$  neighborhood and (3) smoothing the entire field using a  $3 \times 3$  averaging kernel. These successive conventional processes were performed with the free package for particle image velocimetry, JPIV, written by Vennemann (2008, 2009). The simulated PIV fields were also post-processed by means of the new DCT-PLS technique, which was made fully automated using the above-mentioned GCV score. The DCT-PLS was entirely programmed in Matlab language (The MathWorks, Natick, MA, USA; see also “Appendix”). The performances of the two methods were compared by using the normalized root

mean squared error (NRMSE) between the original ( $V^0$ ) and the post-processed ( $\hat{V}$ ) velocity fields given by:

$$\text{NRMSE} = \frac{\|\hat{V} - V^0\|_F}{\|V^0\|_F} \times 100, \quad (7)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm. The median, lower and upper quartiles of the NRMSE were plotted as a function of the percentage of spurious vectors for each of the four scenarios (Fig. 4).

Figures 4a and c show that the conventional method resulted in  $\sim 8$  and  $\sim 12\%$  errors with the scattered patterns and noise variances of  $(0.01 \times V_{\max})^2$  and  $(0.1 \times V_{\max})^2$ , respectively. In the scattered configuration, the errors related to the conventional method (CONVL) were independent of the number of spurious vectors. This means that the normalized median test (Westerweel and Scarano 2005) was highly efficient in detecting the scattered spurious vectors, even in large number. The remaining errors were thus mostly due to the noise which was not completely removed by the smoothing kernels. Because the normalized median test identifies the outliers by comparing the vectors with their neighbors, it lost some effectiveness with the clustered arrangements beyond 10% of outliers (Fig. 4b and d). The normalized errors with the DCT-PLS method increased when increasing the percentage of spurious vectors. The DCT-PLS provided better results than the conventional method with the clustered outlying vectors (Fig. 4b and d). With scattered outliers and low-variance noise, the conventional method yielded lower errors beyond 15% of spurious data (Fig. 4a). High-quality PIV fields, however, typically contain no more than 10% of spurious velocities, in which case the DCT-PLS provides improved post-processed velocity fields in comparison with the conventional method ( $<7\%$  vs.  $<13\%$  see

**Fig. 4** Normalized root mean squared errors (NRMSE, see Eq. 7) between the original and the post-processed vortical flow fields, as a function of the percentage of outlying vectors, in the following situations: **a** and **b**: Gaussian noise with standard deviation (STD)  $1\% V_{\max}$ , with scattered (**a**) and clustered (**b**) spurious vectors; **c** and **d**: Gaussian noise with standard deviation  $10\% V_{\max}$ , with scattered (**c**) and clustered (**d**) spurious vectors. CONVL stands for the conventional method, while DCT-PLS represents the DCT-based penalized least squares method

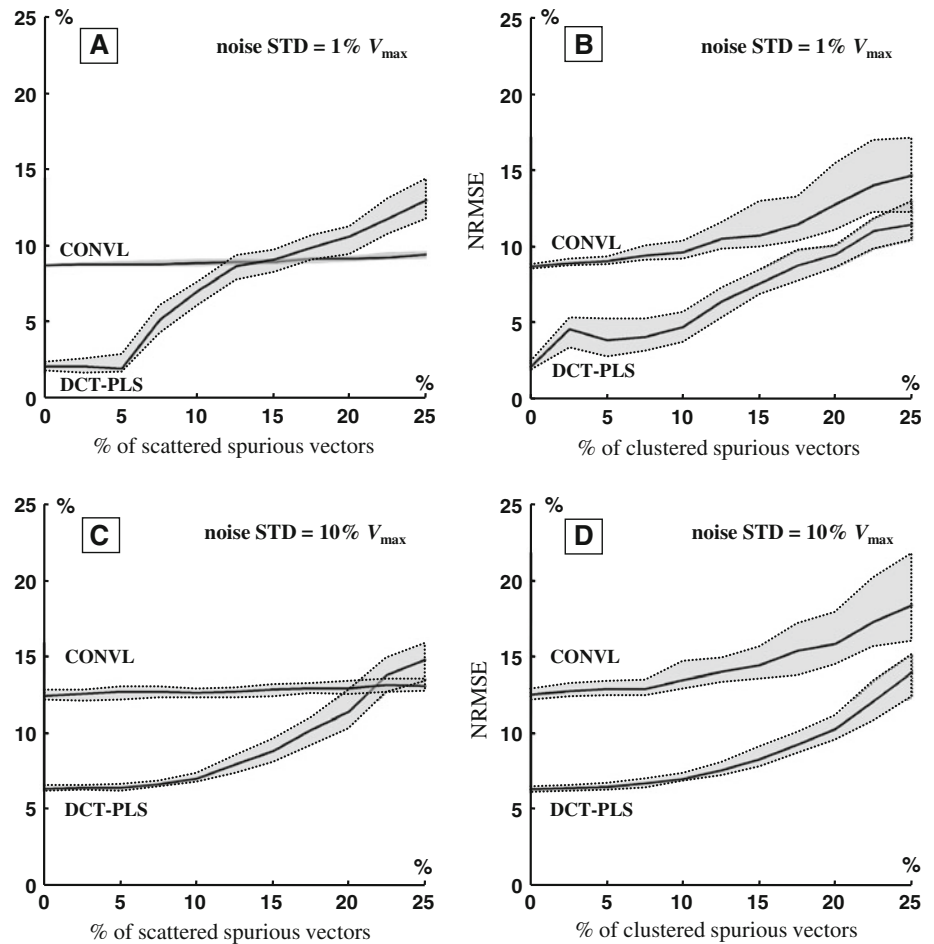
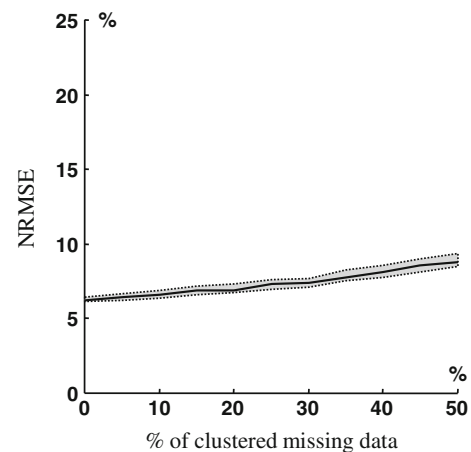


Fig. 4a–d). It is worth noting that the DCT-PLS method does not consider only the neighborhood of the data to be replaced but actually takes the entire field into account. This makes the DCT-PLS little dependent upon the outliers' distribution.

#### 4 The DCT-PLS with missing values and masks

##### 4.1 Dealing with occurrence of missing data

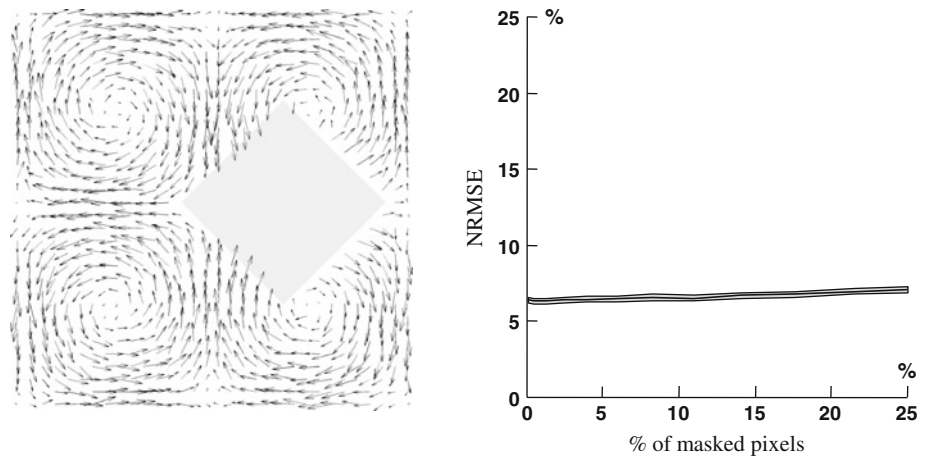
Dropouts or missing data can sometimes occur in PIV measurements due to experimental hazards or minor material defects. How the DCT-PLS can deal with missing values was analyzed on the  $32 \times 32$  vortical flow (Eq. 6) with additive Gaussian noise of variance  $(0.1 \times V_{\max})^2$ . An amount of 0–50% (using an increment of 5%) of clustered missing data was included within the original PIV field using random Poisson distributions with mean six, as described in the previous section. One hundred Monte-Carlo simulations per configuration were performed. The median, lower and upper quartiles of the corresponding NRMSE (see Eq. 7) are given in Fig. 5. The NRMSE



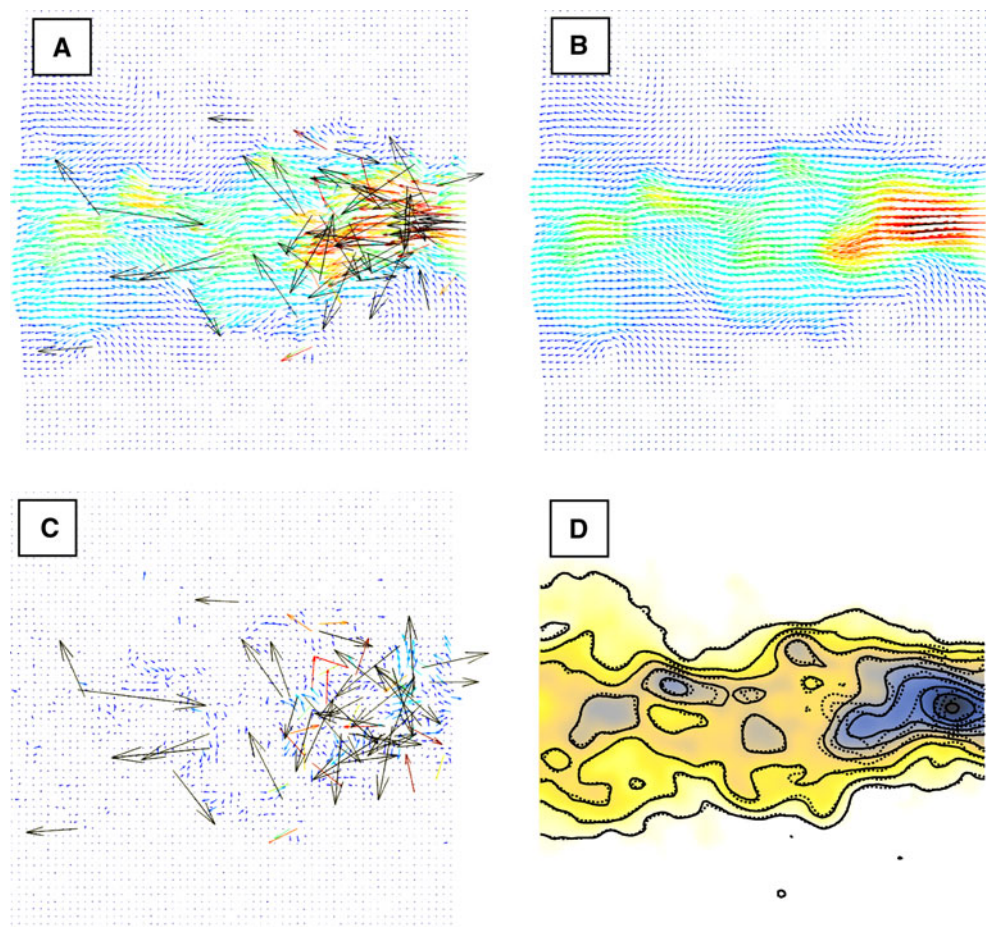
**Fig. 5** Effect of clustered missing data on the post-process of the vortical flow (noise STD =  $10\% V_{\max}$ ) with the DCT-PLS: normalized root mean squared errors (between the post-processed and original velocity fields) as a function of percentage of missing vectors

remained relatively low ( $<8\%$ ) even with 50% of missing vectors and was mostly influenced by the additive noise. These results confirm that the DCT-PLS is highly robust to missing data.

**Fig. 6** Effect of masks on the post-process of the vortical flow (noise  $STD = 10\% V_{max}$ ) with the DCT-PLS. *Left panel*: one example with a diamond-shaped mask of radius 8 pixels. *Right panel*: normalized root mean squared errors (between the post-processed and original velocity fields) as a function of percentage of masked pixels



**Fig. 7** Experimental PIV data measured in a turbulent jet (from Westerweel et al., see text for details). **a** Raw PIV velocities. **b** PIV velocities post-processed with the DCT-PLS. **c** Residuals (velocities in **a**–velocities in **b**). **d** Contour plots of the velocity magnitudes (*solid lines*: using the DCT-PLS, *dashed lines*: using the conventional method). The color background represents the velocity magnitudes obtained with the DCT-PLS (Fig. 7b)

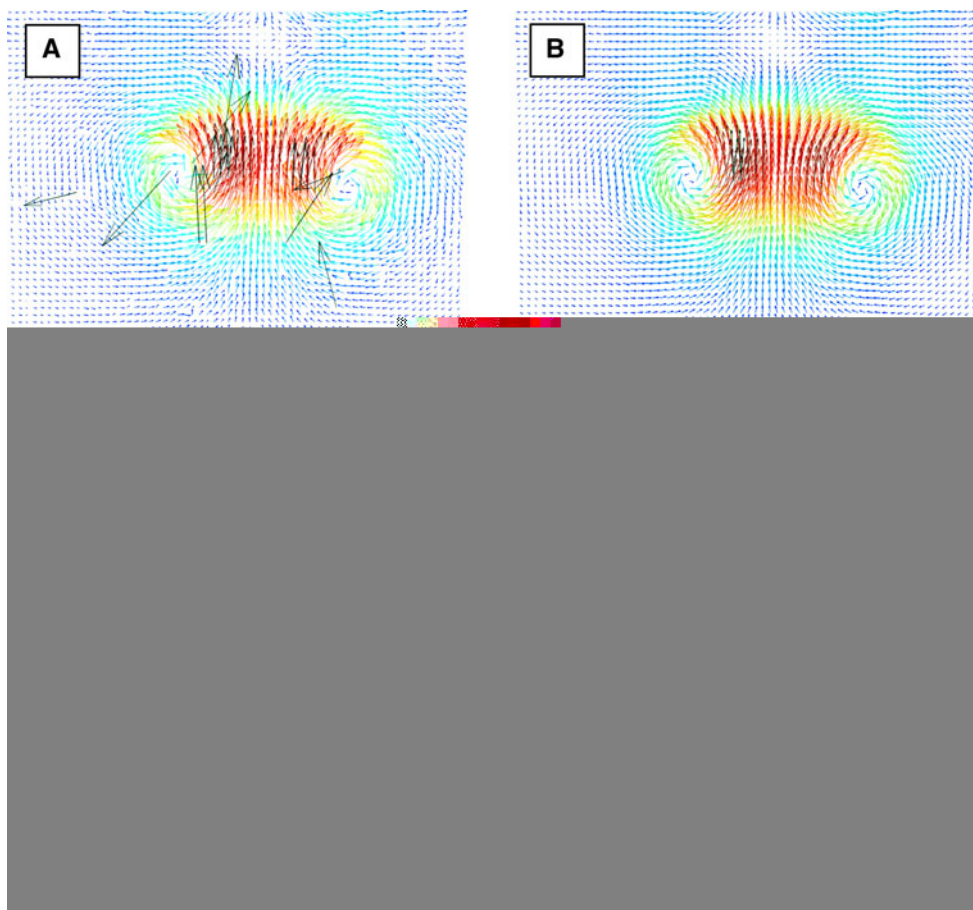


#### 4.2 Application of the DCT-PLS with masks

The fast DCT and inverse DCT transforms only work with data gridded evenly on rectangles (or N-orthotopes in N dimensions). In the case of a masked PIV field, the region of interest becomes non-rectangular (or non-orthotopic).

However, as described by Eq. 5, masking can be taken into account by associating nil weights to masked pixels so that they do not influence the DCT-PLS process. To illustrate that the DCT-PLS can readily deal with masks, several series of Monte-Carlo simulations were completed as described in the previous paragraph. Instead of including

**Fig. 8** Experimental PIV data of a vortex pair (from Willert et al. see text for details). **a** Raw PIV velocities. **b** PIV velocities post-processed with the DCT-PLS. **c** Residuals (velocities in **a**–velocities in **b**). **d** Contour plots of the velocity magnitudes (*solid lines*: using the DCT-PLS, *dashed lines*: using the conventional method). The color background represents the velocity magnitudes obtained with the DCT-PLS (Fig. 8b)



missing values, however, randomly positioned diamond-shaped masks of varying radii (from 0 to 11 pixels) were added. Figure 6a illustrates such a PIV field with a diamond-shaped mask of radius 8. The NRMSE obtained with masked PIV fields (Fig. 6b) shows that the masks have very little effect on the smoothing process.

## 5 The DCT-PLS method with experimental PIV data

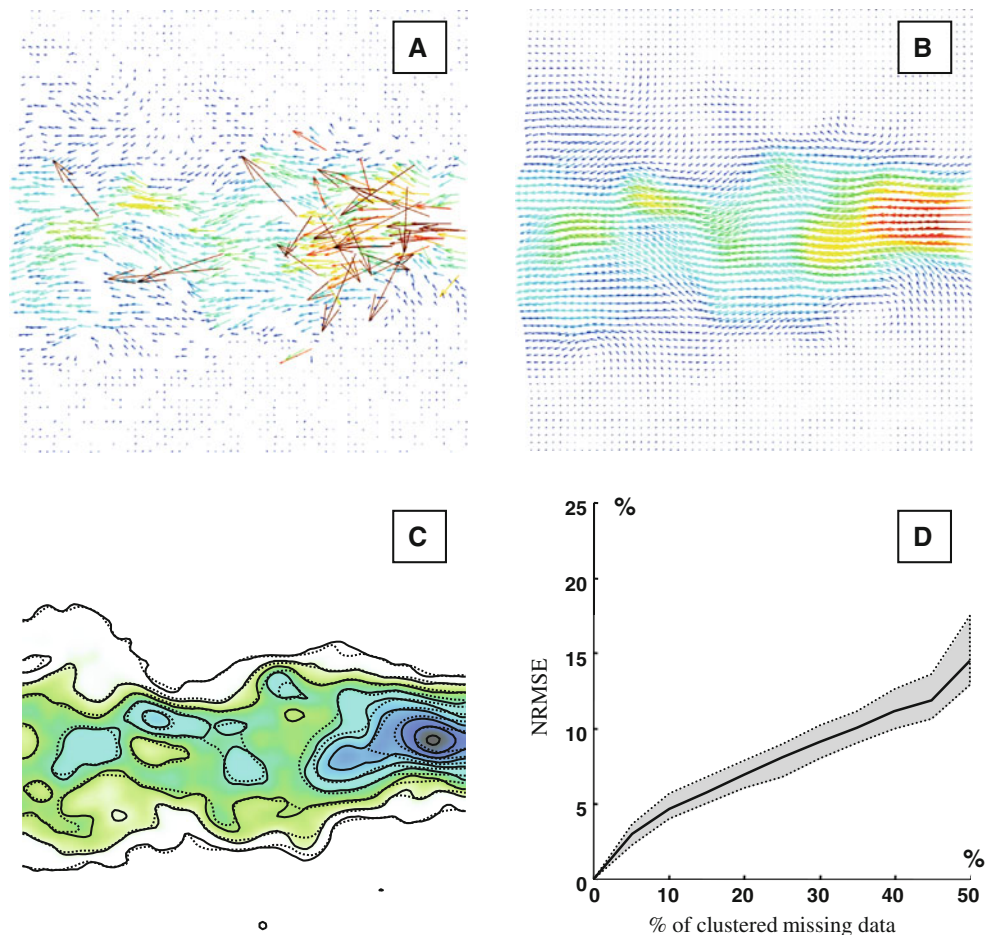
To verify that the new method proposed in this paper is also suitable to real PIV data, the DCT-PLS was tested on two experimental PIV image series freely accessible online. The first image pair, provided by Westerweel (2005), represents a free turbulent jet and is available from PIV challenge (Okamoto 2003). The second image series was acquired in a vortex pair (Willert and Gharib 1997) and is downloadable from the PIVTEC website (PIVTEC 2009). The image pairs were evaluated by PIVview2C Demo (version 2.4, PIVTEC GmbH, Göttingen, Germany) with a  $32 \times 32$  (turbulent jet, Fig. 7a) or  $16 \times 16$  (vortex pair, Fig. 8a) interrogation window using a standard single-pass cross-correlation and a

50% window overlap. The raw pixel shift data were saved in ASCII files and post-processed with the DCT-PLS. As illustrated by Figs. 7a and 8a, the raw PIV fields contained spurious vectors that were all eliminated by the DCT-PLS (Figs. 7b and 8b). A closer look at the residuals (Figs. 7c and 8c) reveals that the DCT-PLS method also removed spurious velocity components of small magnitude. It is worth noticing (Figs. 7d and 8d) that significant differences appeared between the smoothed fields returned by the DCT-PLS and the conventional method (normalized median filter +  $3 \times 3$  average kernel). This difference is particularly notable in the jet inflow (Fig. 7d), which is likely due to the presence of numerous outliers (see Fig. 7a). The overall normalized differences between the two methods (defined as:  $2\|\mathbf{V}_{\text{DCT-PLS}} - \mathbf{V}_{\text{CONVL}}\|_F / (\|\mathbf{V}_{\text{DCT-PLS}}\|_F + \|\mathbf{V}_{\text{CONVL}}\|_F)$ ) were 7.6 and 3.6% for the jet and vortex patterns, respectively.

### 5.1 Missing data

As in Sect. 4, an amount of 5–50% (with an increment of 5%) of clustered missing data was included within the

**Fig. 9** Efficiency of the DCT–PLS with missing data. **a** 50% of vectors have been randomly removed from the raw PIV velocities illustrated in Fig. 7a. **b** The DCT–PLS reconstructed the velocity field with reasonable accuracy (compared with Fig. 7b). **c** Contour plots of the velocity magnitudes (*solid lines*: using the DCT–PLS on the raw PIV data of Fig. 7a, *dashed lines*: using the DCT–PLS on the modified raw PIV data of Fig. 9a). The color background represents the velocity magnitude obtained with the original raw data (Fig. 7b). **d** Effect of the percentage of missing vectors on the normalized root mean squared error (NRMSE, see Eq. 8) between the post-processed PIV fields with missing vectors (turbulent jet, see Fig. 9b) and the post-processed PIV field without missing data (see Fig. 7b)



original PIV data of the turbulent-free jet using random Poisson distributions with mean six. Again, 100 Monte-Carlo simulations per arrangement were carried out. The post-processed PIV fields ( $\hat{V}$ ) with missing vectors were compared to the post-processed PIV field without missing data ( $\hat{V}_{0\%}$ ) by means of the following NRMSE:

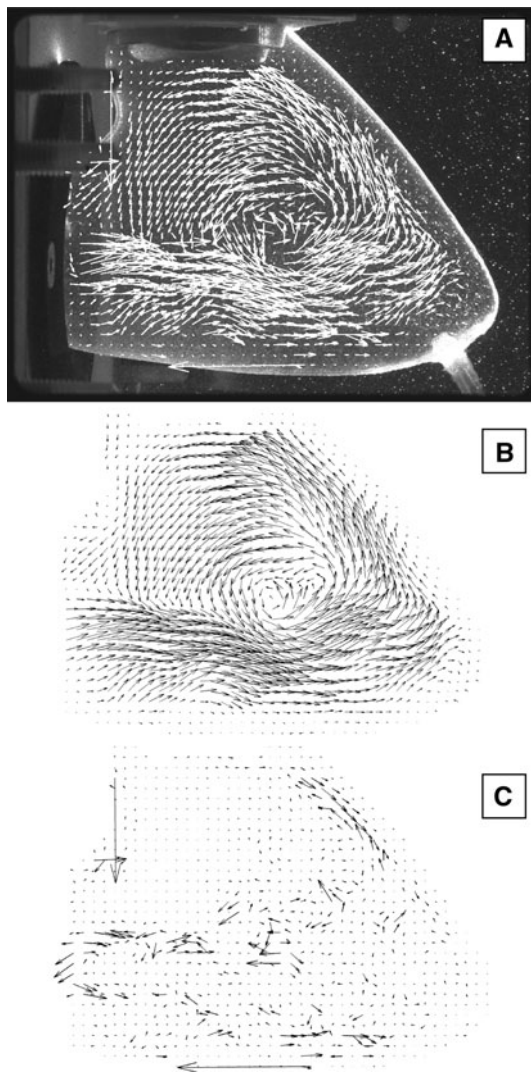
$$\text{NRMSE} = \frac{\|\hat{V} - \hat{V}_{0\%}\|_F}{\|\hat{V}_{0\%}\|_F} \times 100. \quad (8)$$

The median, lower and upper quartiles of the NRMSE were plotted as a function of the percentage of missing data. An example with 50% of missing values is given by Fig. 9a (see also Fig. 7a for comparison). Although this case represents an unrealistic PIV pattern, it clearly illustrates that the DCT–PLS can efficiently cope with missing vectors. In spite of high degradation, the velocity field was indeed reconstructed with reasonable accuracy (Fig. 9b–c). As expected, the normalized error increased when increasing the number of missing vectors but remained less than 10%

with an amount of missing velocities as large as 30% (Fig. 9d).

## 5.2 Masked region of interest

The DCT–PLS was finally used to post-process PIV raw velocities measured in an in vitro left ventricle, supplied by the cardiovascular biomechanics team of the IRPHE (Institut de recherche sur les phénomènes hors équilibre, Marseilles, France) (Garcia et al. 2010; Tanné et al. 2010). These PIV data decipher the intraventricular velocity field during early diastole (ventricular relaxation). As depicted in Fig. 10a, we were interested in the flow delimited by the left ventricular wall. The blood-mimicking fluid enters the left ventricle through the mitral valve and forms a large vortex that remains until the onset of left ventricular contraction. Figure 10b and c represent the smoothed velocity field resulting from the DCT–PLS and the residuals, respectively. These results demonstrate that the DCT–PLS worked rightly within the non-rectangular cardiac cavity.



**Fig. 10** Application of the DCT-PLS approach with PIV measurements performed in a left ventricular model (IRPHE, Marseilles, France). From top to bottom: **a** raw PIV data, **b** PIV velocity field post-processed with the DCT-PLS, **c** residuals

## 6 Limitation of the DCT-PLS and application to 3-D data

### 6.1 The DCT-PLS with high velocity gradients

The examples provided in the present manuscript show that the DCT-PLS can post-process PIV fields very efficiently in numerous instances. One noteworthy limitation of the algorithm, however, should be addressed: better results will be obtained if the velocity field to be sought ( $\mathbf{V}^0$ , see Eq. 1) is sufficiently smooth (i.e. if it contains continuous derivatives in both directions up to some order over the whole domain of interest). This may not be the case in the presence of supersonic flows where velocity discontinuities may occur. In such situation, the DCT-PLS can tend to

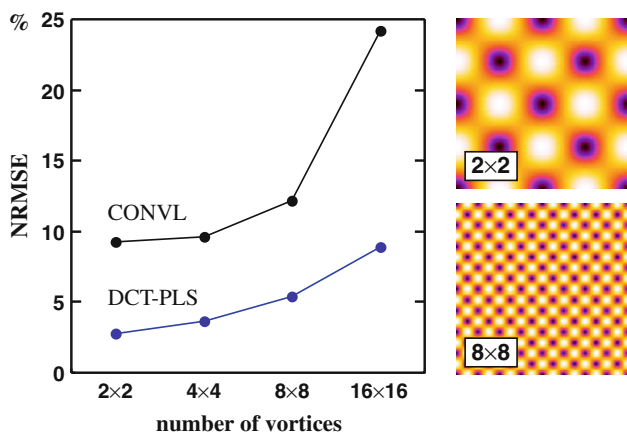
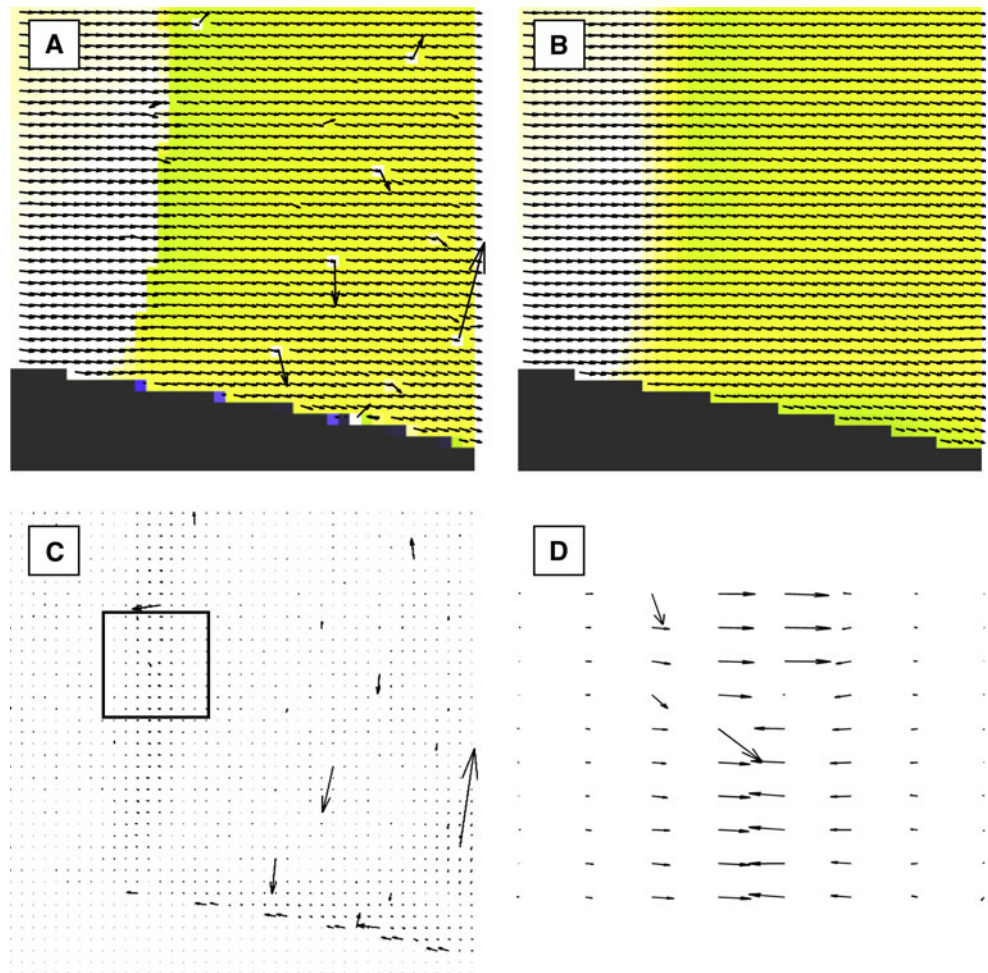
over-smooth the velocity step gradients that appear along the shock waves. As depicted in Fig. 1 and discussed above, the DCT-PLS behaves similarly to a  $2 \times 2$  or  $3 \times 3$  moving average (in terms of cut-off frequency) when  $s$  is in the range  $[0.1-0.5]$ . As a consequence, the original PIV field must preferably not be too noisy so that the optimal smoothness parameter remains lower than 0.5. The optimal  $s$  (issued from the GCV score minimization) calculated with the three experimental data (Figs. 7, 8 and 10) were all less than 0.5 (0.18, 0.38 and 0.21). Although this remains to be confirmed with more velocity fields, the results reported in this study tend to show that the optimal  $s$  values are in the range  $[0.1-0.5]$  for experimental subsonic PIV data, which ensures that over-smoothing, should be avoided in most cases. Alternatively, the DCT-PLS can be used with a very low smoothness parameter (without GCV score minimization) to reduce the effect of smoothing. Nonetheless, even with a low  $s$  value, the DCT-PLS may over-smooth high step gradients. As an example, Fig. 11a depicts a transonic flow containing velocity discontinuities (courtesy of Dr. Kompenhans and Dr. Schroeder from DLR's institute of aerodynamics and flow technology, Göttingen, Germany) and illustrates the effect of the DCT-PLS on such PIV data (Fig. 11b). Over-smoothing clearly appears along the shock wave (Fig. 11c-d), which may limit the use of the DCT-PLS in the presence of supersonic flows.

Besides the presence of high step gradients, the DCT-PLS can also be limited if the region of interest is not adequately sampled (i.e. is characterized by a low spatial resolution). As shown in Fig. 12, the NRMSE indeed increases as the number of vortices in the vortical cellular flow (see Fig. 2) augments for a given resolution (field size =  $128 \times 128$ , noise variance =  $(0.1 \times V_{\max})^2$ , 10% of scattered spurious vectors). This result is not surprising since this would occur with any non-parametric smoother, like the  $3 \times 3$  moving average filter. It must be noticed, however, that the DCT-PLS still provides better outputs than the conventional method (see Fig. 12).

### 6.2 Application to volumetric PIV data

Volumetric three-component PIV is now available (Elsinga et al. 2006; Pereira et al. 2006) and can provide a complete three-dimensional velocity field in a rectangular parallelepiped. Although this paper is focused on two-dimensional velocity fields exclusively, the algorithm is immediately applicable to three and higher dimensions (Garcia 2010b). To test its suitability to post-process three-dimensional (3-D) velocity field, the DCT-PLS was finally used on a 3-D mock flow altered by noise, missing data and outliers. This synthetic flow, of size  $35 \times 41 \times 15$ , was created using the “wind” data available in Matlab (Fig. 13). Several

**Fig. 11** Over-smoothing due to the DCT-PLS with a supersonic flow (courtesy of Dr. Kompenhans and Dr. Schroeder, DLR’s institute of aerodynamics and flow technology). **a** Raw PIV data. **b** PIV velocity field post-processed with the DCT-PLS. **c** Difference between the raw and post-processed fields. **d** Close-up that reveals the presence of over-smoothing along the shock wave



**Fig. 12** *Left panel:* effect of the number of vortices in the vortical cellular flow on the normalized root mean squared error (NRMSE): comparison between the DCT-PLS and the conventional method (CONVL). Velocity field size =  $128 \times 128$ , noise standard deviation =  $0.1 \times V_{max}$ , percentage of spurious vectors = 10%. *Right panels:* two velocity amplitude fields (before being corrupted) are represented for an illustrative purpose ( $2 \times 2$  and  $8 \times 8$  vortices)

Monte-Carlo simulations were also performed to quantify the effect of scattered spurious vectors (methodology similar to Sect. 3). As shown on Figs. 13 and 14, the new post-processing method can efficiently smooth 3-D velocity fields as well.

### 7 Conclusion

Conventional post-processing of PIV velocity fields typically requires three consecutive steps, namely validation, replacement of the incorrect data and smoothing. The DCT-based penalized least squares (DCT-PLS) method described in this paper can handle experimental noise, outliers and missing data, simultaneously. As demonstrated by our results, the DCT-PLS is very well adapted to raw PIV data due to its efficiency, accuracy and rapidness. In addition, contrary to other methods, the DCT-PLS is an automatic all-in-one process and does not require any subjective adjustment of thresholding parameters or kernel

sizes. More importantly, the DCT-PLS is independent of the outliers' distribution, while the conventional median test is less efficient when the spurious vectors are clustered, as often occurs in practice. Because it makes use of the fast

noise STD = 10%

noise STD = 1%  $V_{\max}$

discrete cosine transform, the DCT-PLS is as computationally fast as the conventional methods and much faster than alternative methods recently published.

## Appendix

The supplemental material contains one Matlab program (*pppiv*, Post-Processing of PIV data) that includes all the properties described in this paper. It carries out automatic and robust post-processing of 2-D PIV data. Enter “help *pppiv*” in the Matlab command window to obtain a detailed description, the syntax for *pppiv* and one example. Updated versions of *pppiv* are also downloadable from the author's personal website (Garcia 2010a).

## References

- Craven P, Wahba G (1978) Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer Math* 31:377–403
- Elsinga G, Scarano F, Wieneke B, van Oudheusden B (2006) Tomographic particle image velocimetry. *Exp Fluids* 41:933–947

- Foucaut JM, Stanislas M (2002) Some considerations on the accuracy and frequency response of some derivative filters applied to particle image velocimetry vector fields. *Meas Sci Technol* 13:1058–1071
- Garcia D (2010a) BioméCardio website. <http://www.biomecardio.com>
- Garcia D (2010b) Robust smoothing of gridded data in one and higher dimensions with missing values. *Comput Stat Data Anal* 54:1167–1178
- Garcia D, del Álamo JC, Tanné D, Yotti R, Cortina C, Bertrand E, Antoranz JC, Rieu R, Garcia-Fernandez MA, Fernandez-Aviles F, Bermejo J (2010) Two-dimensional intraventricular flow mapping by digital processing conventional color-Doppler echocardiography images. *IEEE Trans. Med. Imaging* 29:1701–1713
- Gunes H, Rist U (2007) Spatial resolution enhancement/smoothing of stereo—particle-image-velocimetry data using proper-orthogonal-decomposition—based and Kriging interpolation methods. *Phys Fluids* 19:064101–064119
- Hart DP (2000) PIV error correction. *Exp Fluids* 29:13–22
- Heiberger RM, Becker RA (1992) Design of an S function for robust regression using iteratively reweighted least squares. *J Comput Graph Stat* 1:181–196
- Liang DF, Jiang CB, Li YL (2003) Cellular neural network to detect spurious vectors in PIV data. *Exp Fluids* 34:52–62
- Liu Z, Jia L, Zheng Y, Zhang Q (2008) Flow-adaptive data validation scheme in PIV. *Chem Eng Sci* 63:1–11
- Okamoto K (2003) PIV challenge. <http://www.piv.jp/challenge/>
- Pereira F, Stüer H, Graff EC, Gharib M (2006) Two-frame 3D particle tracking. *Meas Sci Technol* 17:1680–1692
- PIVTEC (2009) PIVTEC GmbH. <http://www.pivtec.com>
- Pun CS, Susanto A, Dabiri D (2007) Mode-ratio bootstrapping method for PIV outlier correction. *Meas Sci Technol* 18:3511–3522
- Raffel M, Willert C, Wereley S, Kompenhans J (2007) Post-Processing of PIV Data in Particle image velocimetry. A practical guide pp 177–208. Springer, Berlin
- Shinneeb AM, Bugg JD, Balachandar R (2004) Variable threshold outlier identification in PIV data. *Meas Sci Technol* 15:1722–1732
- Tanné D, Bertrand E, Kadem L, Pibarot P, Rieu R (2010) Assessment of left heart and pulmonary circulation flow dynamics by a new pulsed mock circulatory system. *Exp Fluids* 48:837–850
- Vennemann P (2008) Particle image velocimetry for microscale blood flow measurement (Thesis)
- Vennemann P (2009) JPIV. <http://www.jpiv.vennemann-online.de/>
- Vennemann P, Lindken R, Westerweel J (2007) In vivo whole-field blood velocity measurement techniques. *Exp Fluids* 42:495–511
- Wahba G (1990) Estimating the smoothing parameter in *Spline models for observational data* (Society for Industrial Mathematics, Philadelphia), 45–65
- Westerweel J (1994) Efficient detection of spurious vectors in particle image velocimetry data. *Exp Fluids* 16:236–247
- Westerweel J, Scarano F (2005) Universal outlier detection for PIV data. *Exp Fluids* 39:1096–1100
- Willert CE, Gharib M (1997) The interaction of spatially modulated vortex pairs with free surfaces. *J Fluid Mech* 345:227–250