High-resolution PIV with uncertainty estimation with KNN

Iacopo Tirelli[1,*], Andrea Ianiro[1], Stefano Discetti[1]

1: Aerospace Engineering Research Group, Universidad Carlos III de Madrid, Av. de la Universidad 30, Leganés, 28911, Madrid, Spain

*Corresponding author: iacopo.tirelli@uc3m.es

Keywords: PIV processing, PIV, PTV, KNN, Data-driven measurement enhancement, Uncertainty quantification.

ABSTRACT

We introduce a novel approach to improving the resolution of Particle Image Velocimetry (PIV) measurements. The method merges information from different non-time resolved snapshots exploiting similarity of flow regions in different time instants. The main hypothesis is that the identification of similar flow structures at different time instants is feasible if a sufficiently large ensemble of statistically-independent snapshots is available. Merging individual vectors from different snapshots with similar flow organisation allows an artificial increase of the available information. This paves the way to refining of the interrogation region, i.e. increasing spatial resolution. The similarity can be enforced on a local scale, i.e. morphologically-similar regions are sought only among subdomains corresponding to the same flow region. The identification of locally-similar snapshots is implemented with an unsupervised K-nearest neighbours search in the space of significant flow features. Such features are identified with Proper Orthogonal Decomposition (POD) in subdomains of the original low resolution data. The refined bin size will depend on the number of "sufficiently close" snapshots: the more neighbours are identified, the higher is the "virtual" particle image density available, and consequently the smaller is the bin size.The statistical dispersion of the velocity vectors within the bin is then exploited in the estimation of the uncertainty. The optimal number of neighbours is the one corresponding to the minimum uncertainty. The method is tested and validated against datasets with a progressively increasing level of complexity: two virtual experiments based on direct simulations of the wake of a fluidic pinball and a turbulent channel flow; experimental data collected in a turbulent boundary layer. For further details and more extensive comparison against the state of art the reader is referred to the pre-print of the full article available at `https://arxiv.org/abs/2205.02766`.

1. Introduction

The characterisation of turbulent flows poses an exceptional challenge for measurement techniques. The spatial and temporal scales involved in turbulence dynamics span a range whose extent increases with the Reynolds number. Particle Image Velocimetry (PIV) stands as a powerful

tool for this task as it provides a spatial (and temporal, if sufficiently fast hardware is available) description of the turbulent flow structures. PIV provides useful information that can be exploited for diagnostics, as well as for validation of models and numerical simulations, among others.

The range of observable scales in PIV measurements depends on mean particle spacing that can be achieved on the images with a reasonable particle image density (normally of the order of 0.01-0.1 particles per pixels) and on hardware limitations. Adrian (1997) defines the ratio between the largest and the smallest measurable scales as Dynamic Spatial Range (DSR). The largest one can be increased by employing larger camera sensors, and/or increasing the field of view by reducing the optical magnification. On the other hand, the smallest one depends on the capability of the particles to sample the flow field, i.e. on the particle concentration.

In turbulent flows, the ratio between the large scales and the Kolmogorov spatial scale broadens with $Re^{3/4}$ (Pope, 2000), thus pushing towards developing strategies to increase the DSR of PIV. This has led to the development of high-accuracy high-spatial-resolution PIV processing techniques to reduce the size of the smallest measurable scale, including cross-correlation-based multi-step image deformation methods (Scarano, 2001) with weighting windows (Nogueira et al., 1999; Astarita, 2007), adaptive-resolution techniques (Di Florio et al., 2002; Theunissen et al., 2006; Astarita, 2009; Novara et al., 2012), or methods exploiting time coherence in time-resolved measurements (Hain & Kähler, 2007; Sciacchitano et al., 2012; Cierpka et al., 2013; Lynch & Scarano, 2013; Schanz et al., 2016; Beresh, 2021). Ultimately, the mean particle spacing on the images sets a Nyquist limit for spatial sampling, which is difficult to overcome (or even approach) if time resolution is not available.

It is though common practice in PIV experiments to capture large sequences of samples, most often statistically-independent from each other. This opens up the possibility to improve the spatial resolution and the measurement accuracy by employing statistical information. Remarkable examples include ensemble-correlation, also called single-pixel correlation (Westerweel et al., 2004; Scharnowski et al., 2012; Avallone et al., 2015), and ensemble-particle-averaging (Cowen & Monismith, 1997; Kähler et al., 2012; Agüera et al., 2016), often referred to as Ensemble Particle Tracking Velocimetry (EPTV). EPTV obtains dense clouds of vectors by superposition of instantaneous realisations. Bin averaging of such distributions delivers local probability distribution functions (PDF), from which high-resolution statistical moments can be inferred. Once the number of particles needed for acceptable convergence of the PDF is fixed, increasing the number of snapshots allows to reduce the bin size, thus achieving higher spatial resolution. On the downside, such improvement is obtained at expense of giving up instantaneous information.

The recent advances in data-driven and machine-learning algorithms open the way to the development of resolution-enhancement methods based on the analysis of the statistical distribution of the available samples. Optical flow estimators based on deep learning techniques (Cai et al., 2019; Lagemann et al., 2021) have shown promising results, although their robustness and generalizability are still under investigation. Recently, the majority of successful super-resolution algorithms are based on Generative Adversarial Networks (GANs, Goodfellow et al., 2014), as Z. Deng et al.

(2019), that achieve an increase in spatial resolution of up to 8 times with super-resolution GANs (Ledig et al., 2017). Güemes et al. (2022) recently introduced a new concept of Randomly Seeded super-resolution GANs (RaSeedGAN) that achieves similar resolution enhancement and does not need a paired low-high resolution dataset for training as it exploits directly the sparse particle measurements as a high-resolution target. The main drawback of neural-network-based methods is that they require experienced user for training, and the uncertainty quantification is still difficult.

In this work we propose a novel methodology that merges information from different snapshots, and directly embeds uncertainty quantification in the process. The measurement domain is split in subdomains in order to search, for each snapshot, the closest neighbours of each subdomain and to blend their corresponding information. This process can be considered as a local ensemble particle averaging performed on each region, but only among snapshots which are identified as neighbours in the statistical distribution. The K-nearest neighbour search in a reduced-dimensionality space, obtained by performing a POD (Proper Orthogonal Decomposition) analysis in each subdomain separately. The statistical dispersion among the particles identified in the nearest neighbours, and used for averaging, is exploited to provide an estimation of the measurement uncertainty. The details of the proposed method are discussed in Sec. 2, where all the steps are highlighted and their theoretical background is explained. In order to analyze the algorithm performance, in Sec. 3 the algorithm is validated with two different datasets with increasing complexity, the flow around the fluidic pinball and the flow in a turbulent channel. In Sec. 4 the proposed methodology is also tested on an experimental dataset.

2. Methodology

Figure 1 sketches the flowchart of the proposed algorithm for high-resolution field reconstruction. The method builds upon two hypotheses:

- the identification of similar subdomains is feasible if a sufficiently large ensemble of velocity fields is available;
- the particles randomly sample the flow velocities in statistically-independent snapshots. This provides randomised sampling of flow structures which are assessed to be similar, even if occurring at different time instants in a fixed location.

The similarity between realisations in the same local flow region is assessed in a low-order feature space obtained by performing a POD analysis on each subdomain separately, and taking into account the most energetic local POD modes. In the remainder of the paper we refer to this process as "local POD". The K-nearest neighbour (KNN) algorithm is employed as tool to highlights the "nearest" candidates in the feature space described by POD. For these reasons, the method is referred to as KNN-PTV.

The steps of the method are outlined in the following.

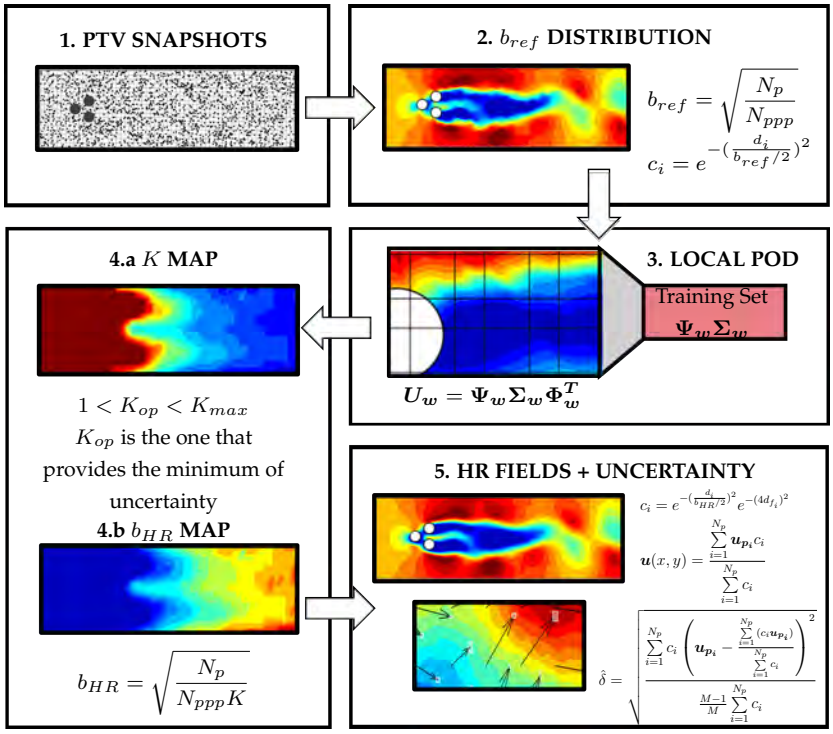

Figure 1. Flowchart of KNN-PTV algorithm.

Step 1: PTV analysis of original images

PTV analysis on individual fields provides scattered vectors which are later used to feed KNN-PTV. Super-resolution PTV (Keane et al., 1995) provides robust particle matching for PTV on dense images.

Step 2: Building reference binned distribution

Velocity fields for feature identification are built either with standard PIV algorithms or from the PTV data from Step 1 with traditional methods (interpolation, spatial averaging, etc.). In this work, we use a weighted average, as described in the following..

In this process, we select the bin size for the weighted average to be sufficiently large in order to minimize empty spots. Assuming that the binning is a simple moving-average operation, in order to obtain N_p particles per spot it should be set a bin size (assumed square) such that:

$$b_{ref} = \sqrt{\frac{N_p}{N_{ppp}}}, \tag{1}$$

where N_{ppp} is the particle density, expressed in particles per pixel, and b_{ref} is the bin size of the reference binned fields. In our implementation, N_p is set equal 10. The velocity corresponding to each bin $\boldsymbol{u}(x, y)$ is computed as the weighted average of the velocities of the particles falling in it, $\boldsymbol{u_{p_i}}$. The weight coefficients c_i are set as a function of the distance d_i of the i^{th} particle from the bin centre:

$$\boldsymbol{u}(x,y) = \frac{\sum_{i=1}^{N_p} \boldsymbol{u_{p_i}} c_i}{\sum_{i=1}^{N_p} c_i}, \quad c_i = e^{-\left(\frac{d_i}{b_{ref}/2}\right)^2}. \tag{2}$$

Step 3: Local Proper Orthogonal Decomposition

The similarity between different snapshots is enforced at a local level. The entire domain is split into subdomains, each containing N_v vectors. Regarding the shape of the subdomains there are no restrictions, so, without leading generality, we employ squared subdomains with size comparable to an interrogation window of a standard cross-correlation analysis, to mimic the PIV interrogation process also in terms of size. In order to provide a sufficiently large number of vectors, the bins in Step 1 are located with high overlap. Throughout this work, we adopted a grid distance of 4 pixels, and subdomains of 40×40 pixels. To reduce the dimension of the search space and to guarantee robustness to noise, the similarity between subdomains in different snapshots is assessed in low-order coordinates. To this purpose, POD is performed on each subdomain over the ensemble of realisations. Only the the first r temporal modes are employed, corresponding to the most energetic flow features. Each subdomain, or window, is treated as an independent realisation and all the time samples for each subdomain are employed to build a snapshot matrix $\boldsymbol{U_w}$, of size $N_t \times (N_v N_c)$, with N_t being the total number of snapshots, and N_c the number of velocity components. A Singular Value Decomposition leads to:

$$\boldsymbol{U_w} = \boldsymbol{\Psi_w \Sigma_w \Phi_w^T}. \tag{3}$$

The matrices $\boldsymbol{\Psi_w}$ and $\boldsymbol{\Phi_w}$ represent the temporal and spatial modes, while $\boldsymbol{\Sigma_w}$ collects the singular values. It must be remarked that the operation in Eq. 3 is performed on the fluctuating velocity fields after subtracting the average of ensemble PTV velocity fields at b_{ref}. This is necessary to avoid that the KNN search is strongly biased by the average flow field, that is coincident with the first mode. For each specific subdomain the features for KNN search are built by $\boldsymbol{\Theta_w} = \boldsymbol{\Psi_w \Sigma_w}$, truncated at rank r, which is set here as the number of modes collecting 90% of energy. This rule of thumb is in qualitative agreement with the elbow criterion for the cases studied in this work. The KNN algorithm searches the closest neighbours in the dataset according to its features $\boldsymbol{\theta_{w_i}} = \boldsymbol{\psi_{w_i} \sigma_{w_i}}$, corresponding to the POD temporal coefficients at the i^{th} time instant. The process is repeated for each subdomain of the snapshot.

Step 4: Optimal-K computation

A dense clouds of vectors for each snapshots is obtained by merging vectors from the K similar instantaneous realisations within each subdomain. This operation is equivalent to increase locally the particle image density by a factor K. The bin size can thus be reduced accounting to this artificial increase of N_{ppp}, leading to:

$$b_{HR} = \sqrt{\frac{N_p}{N_{ppp} K}}. \tag{4}$$

In our implementation N_p is still fixed to 10 (following classical rules of thumb from PIV theory).

The size of the bin depend on the selected value of K. Low K values lead to moderate bin size reduction (i.e. increase of spatial resolution); on the other hand, high K values implies to include more neighbours, thus increasing the risk of merging snapshots with lower similarity in the process. The minimisation of the dispersion of the velocity values within each bin represents the criterion for the selection of K. We discuss it in the following step.

Step 5: Velocity field reconstruction and uncertainty estimation

The velocity for each bin is evaluated as a weighted average of the contributions of all vectors falling in the bin for the K neighbours. The difference from Step 2, is that here vectors belonging to different snapshots are used and weighted taking into account the distance d_i of the vector to the bin centre and the distance between the neighbours in the feature space d_{f_i}.

$$\boldsymbol{u}(x,y) = \frac{\sum_{i=1}^{N_p} \boldsymbol{u_{p_i}} c_i}{\sum_{i=1}^{N_p} c_i}, \quad c_i = e^{-\left(\frac{d_i}{b_{HR}/2}\right)^2} e^{-\left(\alpha \frac{d_{f_i}}{\|\theta_i\|}\right)^2}. \tag{5}$$

The coefficient α is set empirically equal to 4 to give more relevance to the spatial distance with respect to the distance in the feature space. It must be remarked that in Eq. 5 the u_{p_i} represent the original velocity vectors, mean flow included.

Interestingly enough, in analogy to the particle disparity method proposed by Sciacchitano et al. (2013), the statistical dispersion of the vectors in Eq. 5 can be exploited as an indicator of uncertainty. In our case, the disparity vectors are set as the difference between the value assigned by Eq. 5 and the velocity vector used for its evaluation. The instantaneous uncertainty $\hat{\delta}$ is defined as the weighted standard deviation of the disparity vectors:

$$\hat{\delta} = \sqrt{\frac{\sum_{i=1}^{N_p} c_i \left(\boldsymbol{u_{p_i}} - \frac{\sum_{i=1}^{N_p} (c_i \boldsymbol{u_{p_i}})}{\sum_{i=1}^{N_p} c_i} \right)^2}{\frac{M-1}{M} \sum_{i=1}^{N_p} c_i}}, \tag{6}$$

where M is the number of nonzero weights. The expanded uncertainty U is related to $\hat{\delta}$ according to the definition given in Coleman & Steele (2018):

$$U = t\,\hat{\delta}, \tag{7}$$

where the factor t is a coverage factor that comes from the tabulation of T-Student distribution, employed to approximate the $\hat{\delta}$ distribution. The role of expanded uncertainty is to associate a level of confidence to the estimation, or in other words, bounds that within a certain probability containing the true value.

During Step 4 of the process the uncertainty minimisation is performed to select an optimal K for each subdomain, that is equivalent to search for $arg\,min_K\,\hat{\delta}$. In this sense the algorithm becomes "adaptive", i.e. it locally selects an optimal bin size according to uncertainty minimisation. This operation could, in principle, be carried out on each snapshot individually.In practice, this is computationally intensive. Furthermore, as will be discussed in the validation, the sensitivity of the uncertainty to K is rather low. For this reason, the approach adopted here is to establish a fixed K map computed on statistical grounds on limited number of snapshots.

3. Validation

3.1. Fluidic pinball

The first test case is the flow around and in the wake of three cylinders with equal diameter D, allocated to form with their centres an equilateral triangle with a side length equal to $3D/2$. This configuration is referred as fluidic pinball (N. Deng et al., 2020). A total amount of 4737 snapshots are generated from a DNS data at $Re = 130$. The domain is $x/D \in [-5D, 19D]$ and $y/D \in [-4D, 4D]$. Synthetic PTV results are generated in this domain considering a random particle distribution with a particle image density N_{ppp} of 0.02 particles per pixel and a resolution of 25 pixels$/D$. Particle images are generated for standard PIV analysis, with Gaussian-shaped particles having a maximum intensity of 100 counts and a diameter of 2 pixels.

The KNN-PTV is fed with the exact position of the particles. We are conscious that this approach neglects the random noise due to particle positioning in real images, but it is necessary in this first stage of validation to isolate the error due to blending snapshots from other sources related to the image pairing process.

Following Fig. 1, the first step is building the reference low-resolution distribution to perform the local POD analysis. According to Eq. 1, $b_{ref} = 23$ pixels. The local POD is performed on subdomains containing 10×10 vectors (i.e. 40×40 pixels since the high-resolution grid is set with a spacing of 4 pixels) with an overlap between adjacent subdomains of approximately 75%. For each of them, we compute the corresponding snapshot matrices and perform the POD analysis.

With the selected rank for each window, the optimal K selection is carried out: it consists in spanning K between 1 ($b = bref$) and 32 ($b = 4$ pixels, according to Eq. 4) for each subdomain to identify the K minimising the uncertainty from Eq. 6. It can be expected that high number of neighbours K (thus high particle image density) implies the reduction of modulation effects, as it allows the use of smaller bin size. On the other hand, it also implies merging velocity vectors from more

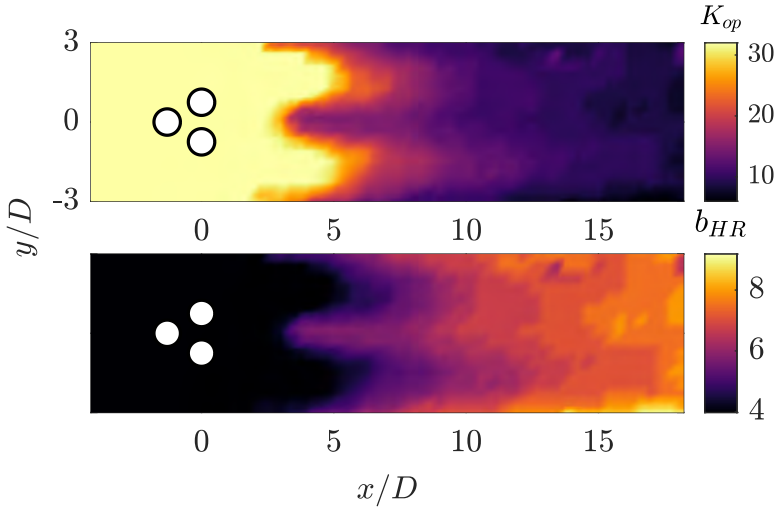

Figure 2. Map of the optimal K (top) and the corresponding map of bin size according to Eq. 4 in pixels (bottom).

"distant" (in the feature space) snapshots, which increases the random error due to non-perfect correspondence in the flow structure. For each subdomain the optimal K is thus estimated. For all bins contained within a subdomain we use the same value of K. The maps shown in Fig. 2 are obtained following the criteria described in the previous section. For this test case, the domain can be ideally split in three different regions. In the upstream region, thanks to the availability of more neighbours very similar to each other, it is possible to reduce the bin size. This is a consequence of the fact that the flow exhibits limited variability. Instead, further from the near field downstream of the cylinder, the flow becomes more chaotic, reducing the similarity between the nearest neighbours. Furthermore, as modulation effects become less important, the random error introduced by small differences between neighbours arises. Both effects reduce the possibility of including neighbours, thus not allowing to push toward smaller bin size.

Figure 3 reports a qualitative comparison between the proposed KNN-PTV, a standard PIV with an interrogation window of 32×32 pixels and the DNS as ground-truth. A clear improvement in spatial resolution with respect to the PIV can be observed.

The adopted metric to quantify the error is the normalised root mean square error δ_{RMS}, evaluated as:

$$\delta_{RMS} = \frac{\sqrt{\frac{\sum_{i=1}^{N_t}(U_i - U_{DNS_i})^2 + (V_i - V_{DNS_i})^2}{N_t}}}{U_\infty}, \tag{8}$$

where U and V are the estimated velocity vectors, U_{DNS} and V_{DNS} are the corresponding vectors from DNS, N_t is the number of snapshots and U_∞ is the freestream velocity, the parameter chosen to obtain a dimensionless estimation, in this case 1 pixel.

Table 1 reports the parameter $\langle \delta_{RMS} \rangle$, i.e. a spatial mean of the above mentioned δ_{RMS}. The KNN-PTV is carried out on datasets with a different number of images (1000, 2000, 3000 and 4737). We can observe that KNN-PTV is able to progressively reduce the error by increasing the number of

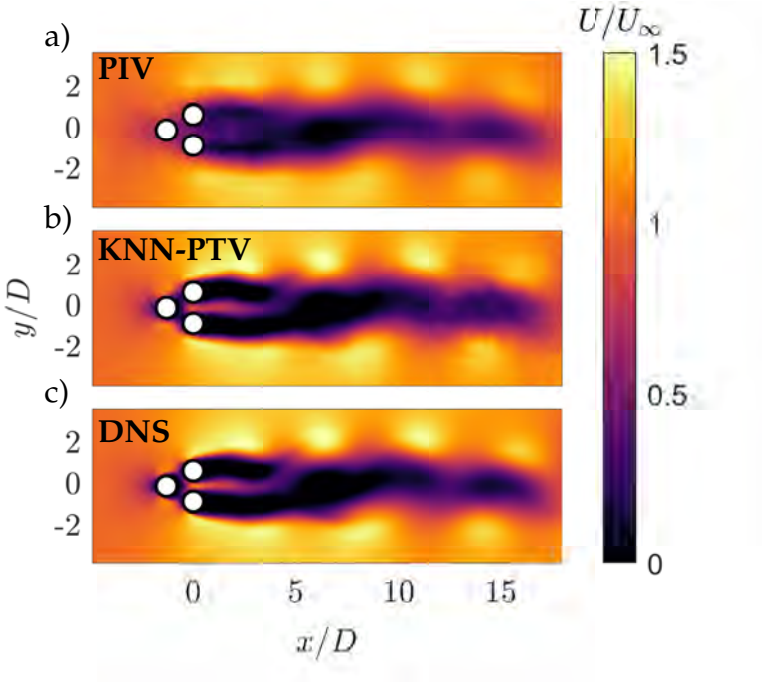

Figure 3. Instantaneous streamwise velocity field contours estimated with: a) standard PIV with interrogation window of 32×32 pixels, b) KNN-PTV. The reference field from the original DNS is included for comparison (c).

samples.

Table 1. Spatial average of the root mean square error $\langle \delta_{RMS} \rangle$, evaluated with Eq. 8.

Method	Nt	$\langle \delta_{RMS} \rangle$	bin size [pixels]
KNN-PTV	4737	0.0299	adaptive
KNN-PTV	3000	0.0308	adaptive
KNN-PTV	2000	0.0316	adaptive
KNN-PTV	1000	0.0337	adaptive
PIV	4737	0.0830	32

The last step for this test case is the assessment of the metric for the uncertainty estimation. To this purpose we employ the weighted standard deviation of the disparity vectors from Eq. 6. The approach follows the disparity method proposed by Sciacchitano et al. (2013), with the significant difference that the vectors are here weighted according to their distance from the bin centre and from the snapshot in the feature space.

An easy way is the evaluation of the "uncertainty effectiveness" (Timmins et al., 2012), i.e. the expanded uncertainty explained in Sec. 2. In this case we evaluate the exact coverage factor t for a 95% confidence from the T-Student tabulation (Coleman & Steele, 2018) taking into account the exact number of particles involved in the computation of velocity for each bin. For this test case the value of uncertainty effectiveness computed on both the velocity components and averaged over

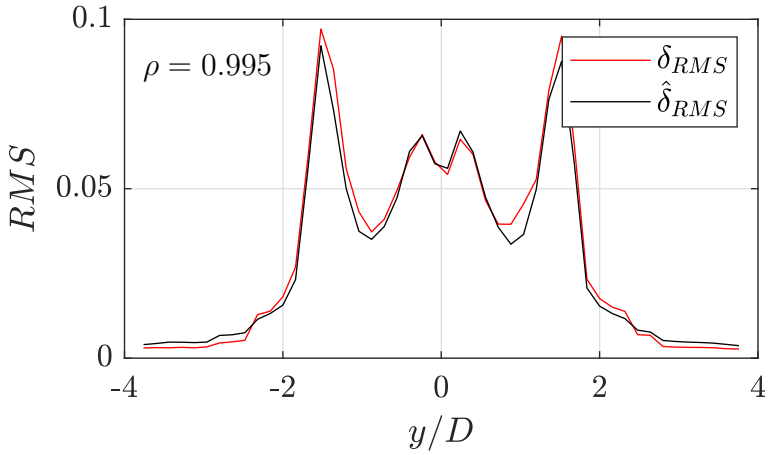

Figure 4. Comparison between δ_{RMS} and $\hat{\delta}_{RMS}$ at x/D = 1. In the top-left corner of the figure the correlation coefficient ρ is also reported.

100 snapshots is 90%, i.e. the uncertainty estimated by KNN-PTV is slightly non-conservative.

Following Sciacchitano et al. (2013), the assessment of the uncertainty estimation can be carried out also from a statistical perspective in time and space. A comparison between the real root mean square error, δ_{RMS}, versus the estimated one, $\hat{\delta}_{RMS}$, on a fixed streamwise position corresponding to $x/D = 1$ is shown in Fig. 4. here it is evident that the estimated uncertainty profile follows very closely the real error, confirmed by a correlation coefficient $\rho = 0.994$, supporting the meaningfulness of the proposed uncertainty quantification method.

3.2. Channel Flow

The second synthetic test case is based on a DNS of a turbulent channel flow, available at the Johns Hopkins Turbulence Database (`http://turbulence.pha.jhu.edu/`). The dimensions of the channel are: 2 half-channel-heights h from wall to wall, $3\pi h$ in the span-wise direction and $8\pi h$ in the stream-wise direction. For further details about the simulation settings the reader is referred to Li et al. (2008). In this simulated experiment, we extract subdomains of $2h \times h$ with a separation of $2h$ and $0.25h$ in the streamwise and wall-normal directions, respectively. The total number of generated snapshots is 11856 with a resolution of 512 pixels$/h$ and a particle image density equal to 0.01 particles per pixel. Following the same approach of the pinball test case, PIV velocity fields with an interrogation window of 32×32 are generated.

Figure 5 shows the profiles of K (black square) and b_{HR} (green squarer), computed as mean of the values on the map along the streamwise direction. In addition, the standard deviations of these values are reported as vertical lines. The trend that the algorithm exhibits in this picture is to select higher K values near the wall, allowing locally to increase the resolution. The reason is that increasing the number of neighbours (thus reducing bin size) reduces the bias errors due to finite spatial resolution on the mean field, while on the other hand increases the random error due to dissimilarities between the identified neighbours. The trade-off between these two sources of

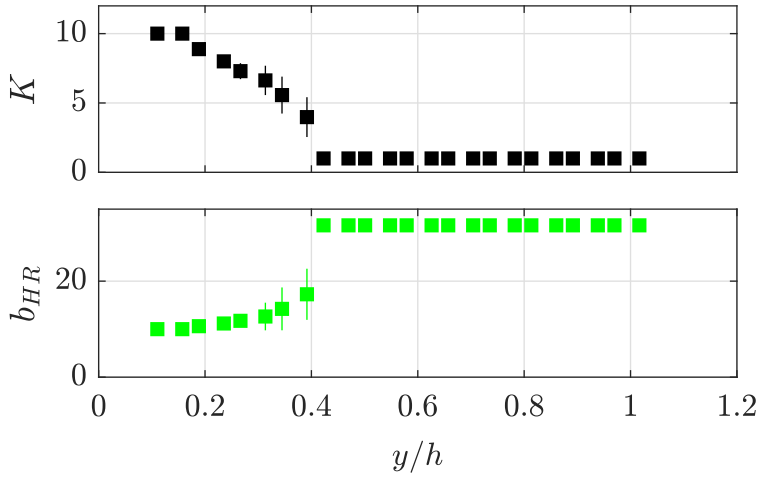

Figure 5. Profile of the optimal K, on the top, and of the corresponding bin size, b_{HR}, on the bottom. The values of K and b_{HR} are represented by squares (black and green respectively) and evaluated as the mean along the streamwise direction. The vertical lines represent the standard deviation of these values.

error occur at larger K in regions where the systematic error on the mean flow is expected to be higher, i.e. in the near-wall region.

A qualitative comparison of KNN-PTV against a traditional cross-correlation-based PIV process is reported in Fig. 6. Their quantitative comparison is reported in terms of $\langle \delta_{RMS} \rangle$, computed as in Eq. 8 (in which U_∞ is equal to 7.5 pixels). The value of $\langle \delta_{RMS} \rangle$ measured for the proposed method is 0.0207, that compared to the one achieved by PIV, 0.0248. In this case the improvement of KNN-PTV with respect to the standard PIV is less evident than in the pinball case.

The uncertainty validation is carried out similarly to Sec. 3.1. The uncertainty effectiveness in this case has been measured to be 95%, in excellent agreement with the theoretical value.

The statistical assessment in the time domain is shown for $(x/h, y/h) = (1, 1)$ and $(x/h, y/h) = (1, 0.1)$ in Fig. 7. In the region far from the wall the estimation of the uncertainty is quite accurate, with a high degree of correlation with the statistical distribution of errors. In the near-wall region the PDFs are wider, as expected due to the higher intensity of the velocity fluctuations and stronger velocity gradients. In this case the agreement is slightly weaker.

4. Experimental validation: turbulent boundary layer

The algorithm is tested on experimental data from Guemes et al. (2019). The experiments are carried out in a Göttingen-type wind tunnel with test section length of $1.5m$ and cross-sectional area of $0.4 \times 0.4\ m^2$. The freestream turbulence intensity is below 1% for velocities up to 20 m/s. A turbulent boundary layer develops on a smooth methacrylate flat plate. All the details about the experiment are reported in Guemes et al. (2019).

Velocity vectors are extracted with a super-resolution PTV approach (Keane et al., 1995). A multi-

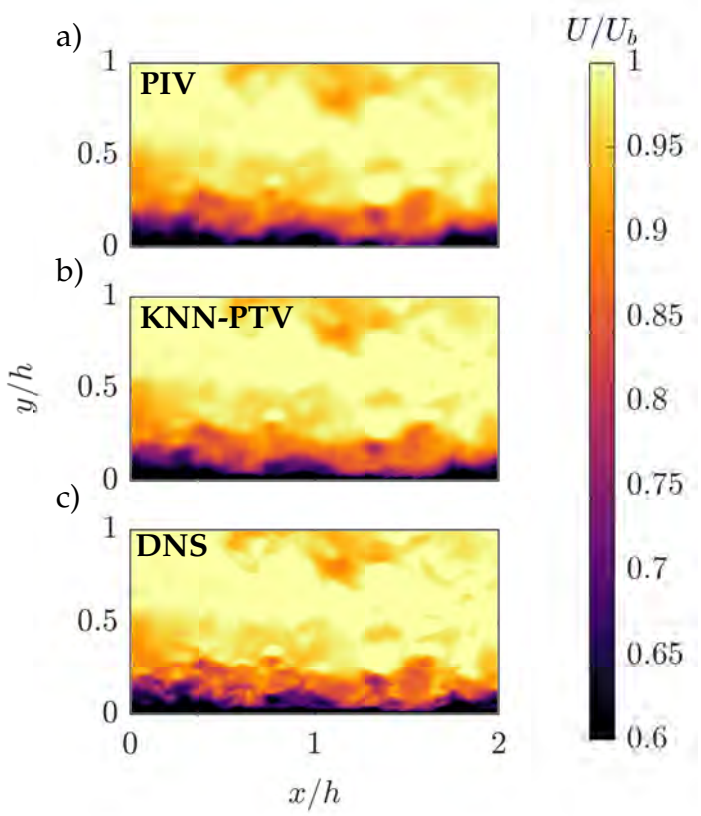

Figure 6. Comparison of the same snapshot from: a) PIV with interrogation window of 32×32, b) KNN-PTV and c) the reference DNS.

step image deformation algorithm (Scarano, 2001) based on high-accuracy interpolation schemes (Astarita & Cardone, 2005; Astarita, 2007) is used to determine the predictor for the biased search.

In order to build a solid high-resolution reference for comparison, we reduce the number of particles available for the KNN-PTV by a factor of 10. The starting point for the building of binned matrices is the result of a PIV processing with an interrogation window of 128×128 and overlap of 25%, while the high-resolution fields for comparison are obtained with a PIV with an interrogation window of 32×32 and an overlap of 25%. The total amount of snapshots employed is 30000.

This time, the large number of snapshots and grid points, and consequently the computational cost for estimating a full map of K, might be intimidating at first glance. The solution can be the enforcing of statistical homogeneity in the streamwise direction, thus requiring only to compute a profile. The computational cost is significantly reduced, paying a small price of an increase of error.

In Fig. 8 a qualitative comparison of the streamwise velocity fields obtained with KNN-PTV and the low-resolution PIV analysis is reported. The results are reported in dimensionless form using the boundary layer thickness δ_{99} and the freestream velocity U_∞, equal to 24.7 mm and 15.5 m/s, respectively. As high-resolution target the results from the PIV analysis with interrogation window of 32×32 pixels are included for reference. Also in this case, KNN-PTV exhibits the capability to

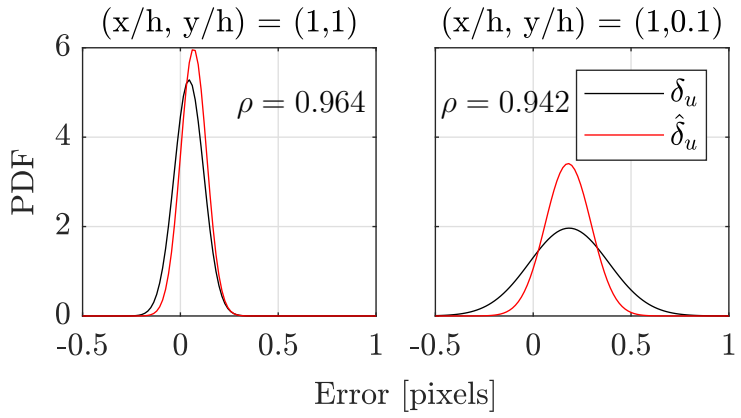

Figure 7. PDF of the instantaneous error δ_u and $\hat{\delta_u}$ for two different points: on the top $(x/h, y/h) = (1, 1)$, on the bottom $(x/h, y/h) = (1, 0.1)$. The correlation coefficient ρ between estimated uncertainty and error distribution is included.

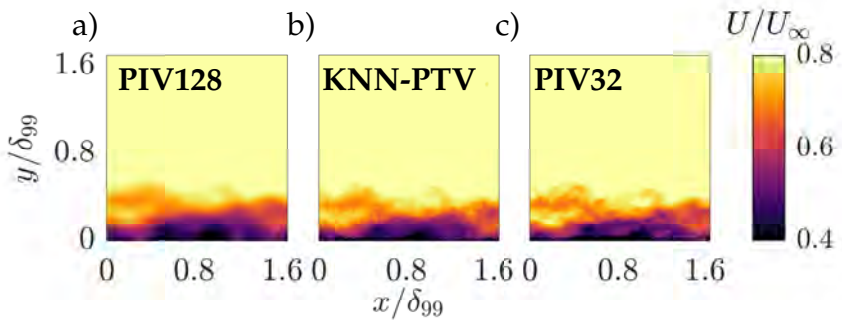

Figure 8. Comparison between: a) PIV with interrogation window of 128×128 pixels, b) KNN-PTV reconstruction, c) PIV with interrogation window of 32×32. The parameter δ_{99} for the dimensionless coordinates is the boundary layer thickness.

recover small scales that are filtered out by the original PIV analysis with 128×128 pixels interrogation window. The non-dimensional error $\langle \delta_{RMS} \rangle$ for KNN-PTV is 0.0170, while for PIV analysis with interrogation window of 128×128 pixels is equal to 0.0186.

The uncertainty validation is carried out with the same procedure of the previous section. The measured uncertainty effectiveness is 92% against the theoretical 95%, which is again a satisfactory agreement. Figure 9 shows a comparison between the profiles along the wall-normal direction of δ_{RMS} and $\hat{\delta}_{RMS}$. The two curves show a very good agreement, with a correlation coefficient equal to $\rho = 0.959$. As expected, higher uncertainty are observed in the near-wall region due to the strongest velocity gradients.

The statistical distribution of real and estimated errors is performed at $x/\delta_{99} = 0.8$ and $y/\delta_{99} = (0.1, 1.6)$, located respectively near and far the wall. The two PDFs in Fig. 10 confirm the trend of the synthetic case. There is a larger disagreement in the near-wall region, with a significant difference in systematic error. This can be ascribed in part by the use of experimental data as "ground-truth". Indeed, the reference data are also affected by finite spatial resolution in the near-

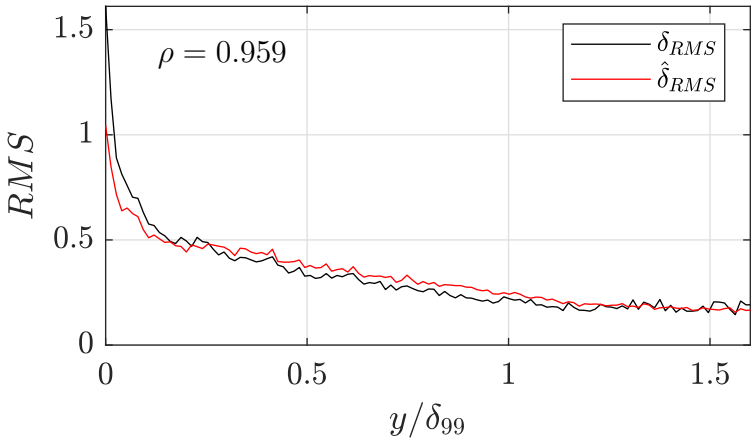

Figure 9. Comparison between δ_{RMS} and $\hat{\delta}_{RMS}$ at $x/\delta_{99} = 0.8$.

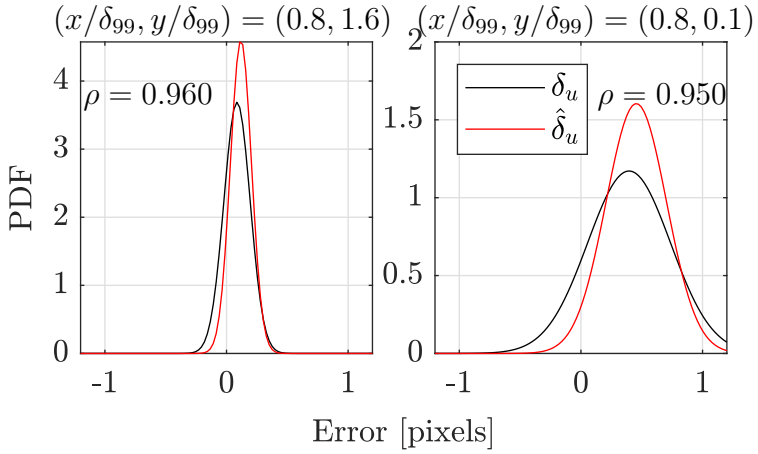

Figure 10. PDF of the instantaneous error δ_u and $\hat{\delta_u}$ for two different points: on the top $x/D = 0.8$ and $y/D = 1.6$, on the bottom $x/D = 1$ and $y/D = 0.1$. For each one is reported also the correlation coefficient ρ.

wall region and by measurement noise.

5. Conclusions

A novel data-driven approach for the enhancement of PIV spatial resolution has been proposed. The method enforces local similarity thanks to a merging information (i.e. measured velocity vectors from PTV) of different non-time-resolved snapshots. The domain is split into subdomains, and for each of them a feature space is built to identify similar realisations. KNN is exploited to identify neighbouring subdomains in the feature space.The uncertainty is computed employing statistical dispersion of the velocity values of the vectors identified in each subdomain. Our method, named as KNN-PTV, provides a complete tool for high-resolution measurements with uncertainty quantification directly embedded in the process.

As proofed by the validation, KNN-PTV features the corresponding novelty and interesting properties:

- Superior resolution if compared to standard PIV and techniques based on interpolating data from PTV. Similarly to EPTV methods must be sought a trade-off between dataset size and desired performance.
- Independently on the flow, the method achieves a robust resolution enhancement. With a local approach, and only using information from few significant neighbours per bin, the method exploits all the advantages of the flexibility of a locally linear embedding.
- Uncertainty estimation is directly embedded in the process, which is a key feature for application and usage of measured data in models for uncertainty propagation.

In conclusion, KNN-PTV presents as a promising end-to-end tool for high-resolution measurements with embedded uncertainty quantification. The proposed framework provides the flexibility to adapt locally the spatial resolution. The method is very easy to implement, as it requires little if no expertise at all in the training phase. Once the vectors are available, the performances are weakly sensitive to the parameter choice, thus easing significantly its applicability. We foresee applications in the field of wall-bounded flows and, more in general, in moderate-to-high Reynolds number flows where time resolution is often not available. Future research efforts might be directed towards extension to volumetric measurements and/or inclusion of physical constraints to improve the accuracy. For further details and comparison against the state of art the reader is referred to the pre-print of the full article available at `https://arxiv.org/abs/2205.02766`.

Acknowledgment

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 949085). The authors warmly acknowledge N. Deng, B. Noack, M. Morzynski and L. Pastur for providing the code for the fluidic pinball simulations and A.Güemes for the experimental dataset.

References

Adrian, R. (1997). Dynamic ranges of velocity and spatial resolution of particle image velocimetry. *Measurement Science and Technology*, *8*(12), 1393.

Agüera, N., Cafiero, G., Astarita, T., & Discetti, S. (2016). Ensemble 3d PTV for high resolution turbulent statistics. *Measurement Science and Technology*, *27*(12), 124011.

Astarita, T. (2007). Analysis of weighting windows for image deformation methods in PIV. *Experiments in Fluids*, *43*(6), 859–872.

Astarita, T. (2009). Adaptive space resolution for PIV. *Experiments in Fluids*, *46*(6), 1115–1123.

Astarita, T., & Cardone, G. (2005). Analysis of interpolation schemes for image deformation methods in PIV. *Experiments in Fluids*, *38*(2), 233–243.

Avallone, F., Discetti, S., Astarita, T., & Cardone, G. (2015). Convergence enhancement of single-pixel PIV with symmetric double correlation. *Experiments in Fluids*, *56*(4), 1–11.

Beresh, S. J. (2021). Time-resolved particle image velocimetry. *Measurement Science and Technology*, *32*(10), 102003.

Cai, S., Zhou, S., Xu, C., & Gao, Q. (2019). Dense motion estimation of particle images via a convolutional neural network. *Experiments in Fluids*, *60*(4), 1–16.

Cierpka, C., Lütke, B., & Kähler, C. J. (2013). Higher order multi-frame particle tracking velocimetry. *Experiments in Fluids*, *54*(5), 1–12.

Coleman, H. W., & Steele, W. G. (2018). *Experimentation, Validation, and Uncertainty Analysis for Engineers*. John Wiley & Sons.

Cowen, E., & Monismith, S. (1997). A hybrid digital particle tracking velocimetry technique. *Experiments in Fluids*, *22*(3), 199–211.

Deng, N., Noack, B. R., Morzyński, M., & Pastur, L. R. (2020). Low-order model for successive bifurcations of the fluidic pinball. *Journal of Fluid Mechanics*, *884*.

Deng, Z., He, C., Liu, Y., & Kim, K. C. (2019). Super-resolution reconstruction of turbulent velocity fields using a generative adversarial network-based artificial intelligence framework. *Physics of Fluids*, *31*(12), 125111.

Di Florio, D., Di Felice, F., & Romano, G. (2002). Windowing, re-shaping and re-orientation interrogation windows in particle image velocimetry for the investigation of shear flows. *Measurement Science and Technology*, *13*(7), 953.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*.

Guemes, A., Ianiro, A., & Discetti, S. (2019). Experimental assessment of large-scale motions in turbulent boundary layers. In *13th international symposium on particle image velocimetry.*

Güemes, A., Sanmiguel Vila, C., & Discetti, S. (2022). Super-resolution GANs of randomly-seeded fields. *arXiv preprint arXiv:2202.11701*.

Hain, R., & Kähler, C. (2007). Fundamentals of multiframe particle image velocimetry (PIV). *Experiments in Fluids*, *42*(4), 575–587.

Kähler, C. J., Scharnowski, S., & Cierpka, C. (2012). On the resolution limit of digital particle image velocimetry. *Experiments in Fluids*, *52*(6), 1629–1639.

Keane, R., Adrian, R., & Zhang, Y. (1995). Super-resolution particle imaging velocimetry. *Measurement Science and Technology*, *6*(6), 754.

Lagemann, C., Lagemann, K., Mukherjee, S., & Schröder, W. (2021). Deep recurrent optical flow learning for particle image velocimetry data. *Nature Machine Intelligence*, *3*(7), 641–651.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... others (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 4681–4690).

Li, Y., Perlman, E., Wan, M., Yang, Y., Meneveau, C., Burns, R., ... Eyink, G. (2008). A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence. *Journal of Turbulence*(9), N31.

Lynch, K., & Scarano, F. (2013). A high-order time-accurate interrogation method for time-resolved PIV. *Measurement Science and Technology*, *24*(3), 035305.

Nogueira, J., Lecuona, A., & Rodriguez, P. (1999). Local field correction PIV: on the increase of accuracy of digital PIV systems. *Experiments in Fluids*, *27*(2), 107–116.

Novara, M., Ianiro, A., & Scarano, F. (2012). Adaptive interrogation for 3D-PIV. *Measurement Science and Technology*, *24*(2), 024012.

Pope, S. B. (2000). *Turbulent flows*. Cambridge University Press.

Scarano, F. (2001). Iterative image deformation methods in PIV. *Measurement Science and Technology*, *13*(1), R1.

Schanz, D., Gesemann, S., & Schröder, A. (2016). Shake-The-Box: Lagrangian particle tracking at high particle image densities. *Experiments in Fluids*, *57*(5), 1–27.

Scharnowski, S., Hain, R., & Kähler, C. J. (2012). Reynolds stress estimation up to single-pixel resolution using PIV-measurements. *Experiments in Fluids*, *52*(4), 985–1002.

Sciacchitano, A., Scarano, F., & Wieneke, B. (2012). Multi-frame pyramid correlation for time-resolved PIV. *Experiments in Fluids*, *53*(4), 1087–1105.

Sciacchitano, A., Wieneke, B., & Scarano, F. (2013). PIV uncertainty quantification by image matching. *Measurement Science and Technology*, *24*(4), 045302.

Theunissen, R., Scarano, F., & Riethmuller, M. L. (2006). An adaptive sampling and windowing interrogation method in PIV. *Measurement Science and Technology*, *18*(1), 275.

Timmins, B. H., Wilson, B. W., Smith, B. L., & Vlachos, P. P. (2012). A method for automatic estimation of instantaneous local uncertainty in particle image velocimetry measurements. *Experiments in Fluids*, *53*(4), 1133–1147.

Westerweel, J., Geelhoed, P., & Lindken, R. (2004). Single-pixel resolution ensemble correlation for micro-PIV applications. *Experiments in Fluids*, *37*(3), 375–384.