Physics-based universal outlier detector for flow statistics

E. Saredi¹*, A. Sciacchitano¹ and F. Scarano¹

¹Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands
* e.saredi@tudelft.nl

Abstract

Outlier detection for PIV velocity fields is still nowadays an active field of research. In the last decades, several image pre-processing and processing algorithms have been developed aiming at increasing the dynamic velocity range of PIV measurements and reducing the measurement uncertainty. Nevertheless, PIV velocity fields are still often characterised by the presence of outliers, which potentially hamper the correct interpretation of the flow physics and negatively affect the evaluation of the flow statistics. The outlier detection strategies presented in literature are mainly based on the statistical analysis of the velocity vector with its immediate neighbour. Most of these algorithms have been demonstrated to be effective for instantaneous flow fields, where the errors associated with the outliers are order of magnitude larger than the expected measurement uncertainty. However, these approaches are not as effective for the flow statistics, where the outliers yield errors of the same order of the measurement uncertainty. To overcome this limitation, this paper proposed an outlier detection approach based on the agreement of the flow statistics to the constitutive equations, more specifically to the turbulent kinetic energy (TKE) transport equation. The focus is posed on the ratio between the local advection terms of TKE and a robust estimation of the TKE production along the local streamline. It is demonstrated that, in presence of outliers, the proposed principle yields a clear separation between the correct and the erroneous vectors. In order to assess the performance of the proposed principle, three different test cases are considered. For all of them, the results are compared with a reference outlier detection methodology, namely the universal outlier detection method proposed by Westerweel and Scarano (2005). The proposed turbulence transport-based approach exhibits higher performance in terms of percentage of outliers correctly identified in the flow statistics.

1 Introduction

In Particle Image Velocimetry (PIV), outliers are spurious vectors that exhibit large unphysical variations in magnitude and direction from neighbouring valid vectors (Westerweel, 1994). Several factors can lead to the appearance of outliers. Among them, unwanted light reflections, lack of seeding and inadequate processing methodology are the main ones (Hart, 2000; Lazar et al., 2004; Sciacchitano, 2019; among others). Several methods for outlier detection have been presented in the literature, with the focus on the analysis of instantaneous flow fields. The most common approach relies on the analysis of the difference between a vector and its neighbours, either in the temporal or in the spatial domain. In this category of outlier detection algorithms, the approach proposed by Westerweel and Scarano (2005) is nowadays widely applied. The latter approach builds upon the median filter presented by Westerweel (1994) and achieves a larger degree of
generality by introducing a normalisation of the median residual with respect to an estimate of the local velocity variation. The authors proved the applicability of their normalised median test in a wide range Reynolds numbers from $10^4$ to $10^6$. Duncan et al. (2010) extended this approach for unstructured data as those obtained by Particle Tracking Velocimetry (PTV) or adaptive interrogation algorithms. These methodologies proved to be effective for isolated outliers, whereas they suffer from under-detection of outliers when a cluster of false vectors is present in the data. This phenomenon is common when low-seeding areas, if cross-correlation based algorithms are employed, or strong reflections on objects are present in the raw images (Masullo and Theunissen, 2016). In order to overcome this limitation, Masullo and Theunissen (2016) presented a new method based on the combination of a first spatial coherence test and a modified Gaussian-weighted distance-based averaging median. The introduction of the former parametrises the extent of the region inside which the vectors are tested. This methodology showed a gain in the achievable spatial resolution and outlier detection probability, at a cost of an increased computational burden.

In addition to the spatial domain, also the time domain has been exploited to identify erroneous vector in PIV data. Proper orthogonal decomposition (POD) has been considered as a filter to detect outlier (Wang et al., 2015; Raiola et al., 2015). In these cases, the low-order POD modes are used in order to reconstruct the velocity field at each time step and the difference between the original velocity and the reconstructed field is used as flag for the erroneous vector. This approach has shown a higher effectiveness for detecting clustered outliers with respect to the normalized median test approach.

The advent of Shake-The-Box (STB, Schanz et al., 2016) has established a new state-of-the-art to perform volumetric measurements. Particles are tracked individually across frames and the temporal information is used then to improve the accuracy of the estimation of the particle position at the various time-steps. The result of this operation is an ensemble of particle tracks inside the measurement volume. Several techniques have been presented to obtain the instantaneous velocity field $\mathbf{u}$ or the time-averaged velocity field $\overline{\mathbf{u}}$ on a Cartesian grid, such as ensemble averaging (Agüera et al., 2016) or functional binning (Godbersen and Schröder, 2020). The aforementioned outlier detection strategies are typically effective for the instantaneous velocity when the latter is expressed on a Cartesian grid. In contrast, when the ensemble-average operation is conducted to determine the time-average velocity field, the magnitude of the outliers often decreases significantly because of their unsteady nature. As a consequence, the outliers’ magnitude can become of the same order of that of the correct vectors, thus decreasing the effectiveness of the aforementioned outlier detection methods and potentially leaving a large percentage of invalid vectors undetected.

Considering the constitutive equations is another possible approach to the outlier detection problem. Song et al. (1999) made use of the continuity equation calculated on a Delaunay tessellation in order to detect spurious vectors. Due to the presence of outliers, unphysical nonzero values of the velocity divergence appear in the incompressible flow field. The outliers can be identified based as regions of high velocity divergence magnitude; however, for this purpose, the definition of a user-defined threshold is necessary, so as to limit the occurrence of false positive (as deductible from the data presented by Azijli and Dwight, 2015).

Inspired by the two works above, in the present work a robust physics-based approach is proposed that employs the turbulent kinetic energy transport equation to detect the outliers in the flow statistics.
2 Physics-based outlier detection principle

This work proposes an outlier detection approach for flow statistics. The measured velocity is expressed via Reynolds decomposition in terms of mean and fluctuating component:

\[ u_i = \bar{u}_i + u'_i \]  

(1)

where the subscript \( i \) indicates the component of the velocity (\( i = 1, 2, 3 \) in a three-dimensional flow).

In the following discussion, an outlier is defined as a velocity vector whose time-average value or its fluctuations depart from the exact value to an extent exceeding the measurement uncertainty. From the fluctuating component of equation 1, \( u'_i \), the turbulent kinetic energy (TKE), expressed as \( k = \frac{1}{2} (u'^2_1 + u'^2_2 + u'^2_3) \), can be obtained. Its behaviour is described by the associated transport equation (Hinze, 1975):

\[ \frac{Dk}{D\tau} = P + T - \varepsilon \]  

(2)

where:

\[ \frac{Dk}{D\tau} = \frac{\partial k}{\partial \tau} + A \]  

(3)

\( A = \bar{u}_i \frac{\partial k}{\partial x_i} \) is the transport by advection,

\[ P = \overline{-u'_i u'_j \frac{\partial \bar{u}_i}{\partial x_j}} \]  

(4)

\( T \) contains all the transport terms of turbulence,

\[ \varepsilon = \nu \frac{\partial u'_i}{\partial x_j} \frac{\partial u'_j}{\partial x_i} \]  

(5)

is the pseudo-dissipation.

Many studies have focused the attention on the description of the relative importance of the different terms of the TKE equation. In Ikhennicheu et al. (2020) a list of the most recent works is reported where the TKE budget has been experimentally evaluated. In this work, turbulent shear flows are considered, such as jets, mixing layers, channel flows and boundary layers. For this kind of flows, the turbulent quantities change slowly together with the mean flow; the production and the dissipation of the turbulent kinetic energy are in balance, i.e. \( P \approx \varepsilon \) (Pope, 2000). Furthermore, equation 2 can be further simplified since in many flows the turbulent transport of \( k \) is negligible (Nieuwstadt et al., 2016). From this, it is possible to conclude that \( P \) acts as upper bound of the variation of TKE along a streamline, since, neglecting the transport term, a non-null dissipation can yield only to a reduction of TKE.

In order to explain the proposed outlier detection principle, it is here considered a time-average velocity field around an airfoil described by \( \bar{u}, u' \) and \( k \), whose streamline are schematically represented in figure 1. The airfoil is set at high angle of attack, which causes the separation of the flow along the suction side of the airfoil. The edges of the separated region are characterized by a high shear and then, as consequence, a high turbulence production. In this case, the presence of four regions of outliers is hypothesized: region A, in front of the airfoil; regions B and C, caused by the shadow created by the light refraction at leading and trailing edges of the (transparent) airfoil, and region D, located in the shear region. Along any time-average streamline of the flow, the evolution of \( k \) must comply with equation 2.
Figure 1: Schematic representation of the outlier detection principle based on turbulent transport.

Based on the above considerations, the normalized admissible TKE transport $\rho_{TT}$ is defined which accounts for the variation of the turbulent kinetic energy along a streamline, relative to the TKE production:

$$\rho_{TT} = \frac{|A|}{|P|} < Thr$$

Values of $\rho_{TT}$ larger than a threshold $Thr$ (theoretically close to unity, in practice larger than 1 to account for the noise in the measurements) are physically inadmissible and therefore correspond to the presence of outliers. In order to increase the robustness of the evaluation of $P$ with respect to the presence of spurious measurement data, the production term in equation 8 is substituted by $P_m$, that represents the median of its neighbourhood. Furthermore, to avoid elevated values of $\rho_{TT}$ in regions where $P$ tends to zero, a minimum normalization level $\gamma_{TT}$ is assumed that represents the uncertainty on the determination of the numerator of equation 8, similarly to what proposed by Westerweel and Scarano (2005). With these two modifications, equation 8 becomes:

$$\rho_{TT} = \frac{|A|}{|P_m|+\gamma_{TT}} < Thr$$

Equation 9 is proposed in this work as a means to identify outliers in statistical data from PIV or LPT. In regions of valid vectors, the normalised ratio $\rho_{TT}$ remains below the threshold, meaning that the variation of the kinetic energy due to the advection along a streamline is smaller than the kinetic energy production (multiplied by the parameter $Thr$). Instead, values of $\rho_{TT}$ larger than $Thr$ indicate a variation of the turbulent kinetic energy greater than its production, which is physically inadmissible.

Assuming an uncertainty level on the measured velocity of the order of 5% of $U_\infty$, the term $\gamma_{TT}$ can be estimated:

$$\gamma_{TT} = \frac{(0.05U_\infty)^3}{\Delta}$$

(10)
being $\Delta$ the vector pitch of the velocity field. In the next section, the approach is assessed using three different datasets comprising both planar and volumetric PIV measurements.

## 3 Experimental assessment

In order to assess the proposed principle, three different experimental datasets have been considered: the 2D velocity field around a two-dimensional wing with NACA 0012 section, the near-wake of a truncated cylinder and the near wake of the Ahmed body. Figure 2 shows three instantaneous raw images of the three considered datasets. The three datasets have been acquired with different systems: in the 2D-wing case, a single-camera two-component PIV system has been used (Adatrao et al, 2021); the images for the near-wake of the cylinder have been acquired with four cameras placed at a large tomographic aperture (Schneiders et al., 2016); finally, the near-wake of the Ahmed body has been investigated using the Robotic Volumetric PIV system (Jux et al., 2018), which is composed by four cameras placed at a low tomographic angle (Saredi et al., 2020).

In order to assess the performance of the turbulence transport-based principle, its results are compared to the ones obtained by the universal outlier detection, hereafter referred as UOD, proposed by Westerweel and Scarano (2005). In their work, the authors proposed to identify outliers based on the value of a normalized residual $r_0^*$ defined as:

$$r_0^* = \frac{|U_0 - U_m|}{r_m + \gamma_{UOD}}$$

where $U_0$ is the velocity vector considered, $U_m$ the median of its neighbours, $r_m$ the median of the neighbour residuals defined as $r_i = |U_i - U_m|$ and $\gamma_{UOD}$ the minimum normalization level, set to 0.1 pixel. In the original paper, velocity measurements were considered outliers when the normalized residual $r_0^*$ exceed the value of 2; the same threshold is used in this analysis.

---

**Figure 2:** Raw images of the three experimental datasets considered in this work. Left: flow around a 2D-wing NACA 0012; center: near-wake of a truncated cylinder; right: near-wake of the Ahmed body.

The sources of outliers differ among the considered test cases. For the 2D-wing case (figure 2, left) the presence of the wing body and the shadow created by the refraction at the leading edge create outlier regions in the velocity field, as it can be seen in figure 3 (top-left), where the instantaneous velocity field is presented. Considering the case of the cylinder, a strong reflection on the object surface is visible in the raw image presented in figure 2 (center). This reflection cannot be eliminated with pre-processing and causes the appearance of outlier instantaneous tracks along the line of sight of the considered camera, as visible in the highlighted region in figure 3 (center-left). For the Ahmed body flow, studied with a coaxial system (Schneiders et al., 2018), an excess of particles coupled with the background noise in the middle of the images causes the appearance of spurious tracks at the center of the measured volume close to the cameras (figure 3, bottom-left). The outliers that appear in the instantaneous fields, represented by the velocity vector on a
Cartesian grid in the 2D-wing case and particle tracks for the other two cases, corrupt the correspondent time-average and ensemble-average velocity fields, presented for all the cases in the right column of figure 3. For the 2D-wing case, the region correspondent to both the shadow and the perspective section of the wing are characterized by erroneous low velocity. This effect is clearly visible in the area correspondent to the shadow. For the cylinder case, the detrimental effect of laser reflections that generates invalid vectors inside the cylinder geometry is evident. Furthermore, a region of flow characterized by quasi-null streamwise velocity can be spotted in the area correspondent to the erroneous tracks, indicated with an ellipse in figure 3 (center-left). This can be explained by the fact that the erroneous particles generated by the surface reflection yield random tracks whose mean velocity is approximately null. Similarly, the presence of spurious tracks in the near-wake of the Ahmed body produces an unphysical region of deceleration and acceleration at $X = 0.2$ m in the free-stream region of the ensemble-averaged flow field (figure 3, bottom-right).

Figure 3: Left column: Instantaneous streamwise velocity field $u$ represented on a Cartesian grid and at particle locations for the planar case and for the 3D cases, respectively. In all the cases, the areas interested by erroneous measurements are highlighted. Right column: time-averaged and ensemble-averaged velocity field $\bar{u}$ for the 2D case and the 3D cases, respectively. For the near-wake of the Ahmed body, the region labelled A indicates the area considered correct for the determination of $\eta_{fp}$, presented in figure 6.
The presence of outliers in the instantaneous velocity fields does not influence only the time-averaged field $\bar{u}$ but also the fluctuating component of the velocity $u'$. In order to evaluate this effect along all the three velocity components, the turbulent kinetic energy is here considered, in the form of $\sqrt{k}/U_\infty$. For all the cases analysed, the outlier regions present a sharp increase of turbulent kinetic energy with respect to the surrounding (figure 4). It can be noted that the extension and the shape of the outlier regions result more visible in the flow fluctuations, visualised by means of the turbulent kinetic energy in figure 4, than in the time-averaged flow fields (figure 3-right). Nevertheless, the turbulent kinetic energy alone can hardly be used as a criterion for outliers detection, because its physically admissible values vary greatly among the experiments.

![Figure 4: Contour of normalized turbulent kinetic energy $\sqrt{k}/U_\infty$ for the three experimental datasets: 2D-wing (left), cylinder (centre) and Ahmed body (right). For the near-wake of the Ahmed body, the region labelled B indicates the area considered erroneous for the determination of $\eta_{id}$, presented in figure 6.](image)

In order to assess the performance of the proposed principle, the two outliers indicators, namely the normalized admissible TKE variation of the current turbulence transport-based approach ($\rho_{TT}$ of equation 9 and the residual $r_0^*$ of Westerweel and Scarano (2015) are presented in figure 5 for all the three considered cases. When considering the universal outlier detection, it should be remembered that the presence of false vectors is flagged by the residual $r_0^*$ exceeding the value of 2. For both the 2D case and the cylinder case, the universal outlier detection fails to detect the outlier region. In the former, only a small region at the leading edge of the shadow is flagged as erroneous. In the latter, the area affected by reflection is not detected as erroneous, with only some vectors at the edges of the flow domain indicated as outliers. The UOD shows better results when applied to the data of the near-wake of the Ahmed body. In this case, a portion of the erroneous region in the free-stream is correctly detected.

The right column of figure 5 presents the spatial distribution of the ratio $\rho_{TT}$ for all the considered cases. For all of them, $\rho_{TT}$ reaches the highest value at the edges of the erroneous region, with $\rho_{TT} > 10$, an order of magnitude difference with respect to the adjacent correct region. Considering the 2D-wing case, the edges of the shadow region and the front and top edges of the area occupied by the prospective view of the wing body are clearly detected as faulty vectors. When the cylinder case is considered, the edges of the erroneous region are clearly detected. It is remarked that the indicator $\rho_{TT}$ exhibits a large separation between the erroneous region ($\rho_{TT}$ typically exceeding 10) and the correct portion of the flow ($\rho_{TT} < 1$).

When the third case is considered, the near-wake of the Ahmed body, the proposed method correctly detects the boundary of the erroneous region highlighted in 4 (right). Also in this case, the separation in terms of $\rho_{TT}$ between the correct and the erroneous region is larger than one order of magnitude. In addition, two regions close to the Ahmed body are highlighted as
erroneous. Unfortunately, their validity cannot be determined since both the velocity field presented in figure 3 (bottom-right) and the TKE presented in figure 4 (right) do not show unphysical behaviour. This suggests the presence of false positives in those regions.

Figure 5: (Left column) Normalized residual $r_0^*$ of the universal outlier detection algorithm proposed by Westerweel and Scarano (2005) for all the considered datasets. Top: 2D-wing; center: near-wake of a truncated cylinder; bottom: near-wake of an Ahmed body. (Right column) Normalized admissible TKE variation $\rho_{TT}$ defined by equation 9 for all the considered datasets. Top: 2D-wing; center: near-wake of a truncated cylinder; bottom: near-wake of an Ahmed body.

In order to quantitatively evaluate the performance of the turbulence transport-based principle in comparison with the reference methodology, results are presented in terms of detection and false positive rates $\eta_d$ and $\eta_{fp}$, respectively. The former is defined as the number of spurious vectors correctly flagged as erroneous divided by the total number of outliers. The false positive rate $\eta_{fp}$ is defined as the ratio between the number of correct vectors flagged as outliers and the total number of correct vectors in the considered portion of the flow. To evaluate the robustness of the two methodologies, the two ratios have been evaluated varying the thresholds between 0 and 4 for the UOD, and between 0 and 10 for the turbulence transport-based approach. To perform this analysis, the dataset regarding the near-wake of the Ahmed body has been considered. The calculation of $\eta_{fp}$ requires the knowledge of the number of correct vectors that are wrongly flagged as outliers. In order to evaluate this, a volume containing only correct vectors has been selected by visual inspection. The planar section of the considered volume is indicated as region A in figure 3.
On the contrary, the calculation of $\eta_d$ requires the knowledge of the erroneous vectors. For this purpose, the area circled and indicate as region B in figure 4 (right) has been considered. Since the proposed principle aims to detect the edges of erroneous regions, only the vectors located at the edge of the faulty area have been considered. Once isolated this group of vectors, the amount of them flagged as erroneous by the UOD and by the turbulence transport-based outlier detection has been calculated. The results obtained in terms of $\eta_d$ and $\eta_{fp}$ by the two methods varying the correspondent threshold is presented in figure 6. It has to be noted that the optimal values of $\eta_{fp}$ and $\eta_d$ are 0 and 1, respectively. It is expected that, when increasing the value of the threshold of the two algorithms, both correct detection probability and the false positive rate decrease. Both the methodologies presents the expected behaviour. However, for the UOD approach the correct-detection and false-positive curves decrease at similar rates, whereas for the turbulence transport-based approach the false-positive curve drops much more rapidly. As a consequence, when the typical threshold of $r_0 = 2$ is used for the UOD method, only 5% of the correct vectors are detected as outliers (5% over-prediction), but less than 50% of the outliers are identified (only 45% correct prediction). Conversely, using the turbulence transport-based approach with a threshold of 5 yields 5% over-prediction and 95% correct prediction. This result supports the conclusion that the median test based on the velocity field is not able to detect outliers when the flow statistics are considered, while the proposed turbulence transport-based approach is able to detect the outliers in the flow statistics.

Figure 6: Comparison of the detection ratio $\eta_d$ and the false positive ratio $\eta_{fp}$ obtained by universal outlier detection and the turbulence transport-based approach proposed in this paper. The data are referred to the near-wake of Ahmed body dataset.

4 Conclusions

A novel outlier detection criterion is presented that invokes the physical principle of turbulence transport and applies to the flow statistics evaluated by PIV or PTV. The approach relies on the fact that the measured flow field should satisfy the turbulent kinetic energy transport equation. However, outliers often produce un imbalance between the terms of the transport equation. The ratio between the measured turbulent kinetic advection rate and the production term along a trajectory ($\rho_{TT}$) is introduced to define a criterion for the detection of the boundaries of regions of erroneous data. In order to guarantee the robustness of the method, the production term is substituted with the median of its neighbours. The approach has been tested on three different datasets, obtained with three different techniques: planar PIV, large-aperture 3D-LPT and small-
aperture 3D-LPT. The proposed turbulence transport-based outlier detection principle compares favourably with the state-of-the-art universal outlier detection when considering diverse experimental cases. In order to compare quantitively the performance of the proposed method with the reference, the dataset regarding the near-wake of the Ahmed body has been considered. Considering the two optimal threshold, $r_0^\ast = 2$ for UOD of Westerweel and Scarano (2005) and $\rho_{TT} = 5$ for the turbulence transport-based approach, the results of the two methods are compared in terms of detection ratio $\eta_d$ and false positive ratio $\eta_{fp}$. Both of the methods yielded a similar over-prediction of the outliers (about 5%); however, while the algorithm from Westerweel and Scarano could only detect less than 50% of the outliers, the proposed turbulence transport-based outlier detector was capable to identify 95% of the outliers. In case of large regions of outliers, once the boundary of the outlier regions has been identified with the proposed methodology, standard contour tracing algorithms can be applied in order to isolate the of the erroneous regions.

References
Adatrao S, Bertone M, Sciachitano A (2021) Multi-$\Delta t$ approach for peak-locking error correction and uncertainty quantification in PIV. Measurement Science and Technology 32:054003
Schneiders JFG, Caridi GCA, Sciachitano A, Scarano F (2016) Large-scale volumetric pressure from tomographic PTV with HFSB tracers. Experiments in fluids 57:164