

PAPER • OPEN ACCESS

Functional binning: improving convergence of Eulerian statistics from Lagrangian particle tracking

To cite this article: Philipp Godbersen and Andreas Schröder 2020 *Meas. Sci. Technol.* **31** 095304

View the [article online](#) for updates and enhancements.

Functional binning: improving convergence of Eulerian statistics from Lagrangian particle tracking

Philipp Godbersen¹  and Andreas Schröder¹

German Aerospace Center (DLR), Institute of Aerodynamics and Flow Technology, Göttingen, Germany

E-mail: Philipp.Godbersen@dlr.de

Received 28 January 2020, revised 3 April 2020

Accepted for publication 21 April 2020

Published 24 June 2020



CrossMark

Abstract

In the analysis of Lagrangian particle tracking data, ensemble averaging with spatial bins is used to generate Eulerian flow statistics. Due to the scattered nature of the particles over independent snapshots, the possible spatial resolution is directly dependent on the measured particle position accuracy and the amount of available data. This requires a balance between convergence of the underlying statistic and the bin resolution. Current binning approaches use the velocity information of the particle positions at single time steps directly and do not exploit the additional information available from the temporal filtering of the tracking process. We present a novel functional approach to the binning procedure that extracts all available information from the particle tracks and improves convergence speed. For a given experiment this allows for higher resolution of flow statistics than classical approaches or alternatively to reduce the necessary amount of data required for a given resolution. Furthermore, uncertainty measures from the particles position, velocity and acceleration can be propagated directly by weighting coefficients.

Keywords: spatial binning, ensemble averaging, turbulence statistics, Lagrangian particle tracking, Shake-the-Box

(Some figures may appear in colour only in the online journal)

1. Introduction

Ensemble averaging of flow fields measured for a time series of snapshots is a key component in the calculation of flow statistics. Measurement methods based on seeding the flow with small tracer particles allow to relate their movement to the flow for Stokes numbers $St \ll 1$. While particle image velocimetry (PIV) can be used to determine the flow fields for snapshots, it is limited by the use of correlation windows which inherently introduces a low pass filtering effect on resolvable flow structures and velocity fluctuations respective gradients. For building statistics the spatial resolution can be improved upon

by involving an ensemble averaging approach directly in the image correlation process [5] allowing for single pixel resolution statistics of the mean flow [8] and Reynolds stresses [17]. However this approach is limited to those statistics that can be extracted from the ensemble image correlation and no method to evaluate multi point statistics has been presented yet. One common aspect of the correlation based approaches is that their extension to three-dimensional measurements is challenging. The reconstruction of intensity fields in 3D for correlation is very time consuming and limits resolution.

An alternative for use in ensemble averaging can be found using Lagrangian particle tracking approaches. Here the individual particles in the flow are tracked, which can provide point measurements of individual fluid elements. Due to the individual tracking such approaches are more readily extended into three dimensions [11]. Particle tracking velocimetry (PTV) can be used to recover turbulent flow statistics



Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

by spatial binning and subsequent ensemble averaging [7]. Lagrangian approaches are very well suited for ensemble averaging as particle positions are distributed in a stochastic manner between individual snapshots. Instead of the binning resolution being predefined by correlation-windows in an Eulerian grid like for PIV, the maximum resolution for an ensemble average from scattered particles is theoretically limited only by the amount of snapshots available and the particle position measurement error. In practical applications the achievable resolution is limited. The classical PTV measurement techniques require low particle seeding densities to allow for correct identification of particle tracks which poses a challenge when a large amount of particles is required.

Novel approaches to Lagrangian particle tracking improve these conditions drastically. The Shake-the-Box (STB) method [16] allows for the accurate measurement of the position, velocity and acceleration of particles at high seeding densities with low position error by using optical transfer functions [15], iterative particle reconstruction [20] and the available temporal information. The greatly increased amount of particles per snapshot allows users to quickly acquire large amounts of particle tracking data while the increased positioning accuracy increases the upper bound for possible bin resolutions. As Lagrangian methods allow to exploit possible symmetry or homogeneity conditions for some investigated flows to increase the effective particle density the STB method makes sub-pixel sized bins for ensemble averaging viable [18].

For slow flow velocities, high speed cameras allow to build long time-resolved tracks of particles through the entire measurement volume. For faster flows approaches using double-frame cameras are used, capturing short independent bursted image sequences at low repetition rates. Measurements can be performed either using two pulses [4, 6], or with a multi-pulse approach [13, 14]. With multi-pulse STB each snapshot consists of a sequence of four images allowing short tracks to be fitted to the recovered particle sequences. The greater amount of data available in each snapshot leads to better tracking and decreases errors when compared to two-pulse based particle tracking. Within this paper we are using a multi-pulse STB approach with four pulses but the introduced method can also be applied to methods using two pulses. For both multi-pulse and time-resolved variants, a particle distribution with the associated velocities and accelerations is available for each snapshot in time and both variants are processed similarly for bin averaging.

Improvements in bin averaging methods so far have primarily focused on improving the binning procedures for these single data points. For a given position of a particle and a known regular binning grid, a particle can easily be assigned the bin whose statistic it contributes to via an integer truncated division:

$$\text{Bin}_{\text{index}} = \text{div}(x - x_0, \Delta x). \quad (1)$$

The velocity information of that particle is then added to the desired statistic within the bin. Commonly used are single pass, constant memory algorithms [19] for the calculation of flow statistics such as the Reynolds stress tensor. Particles are

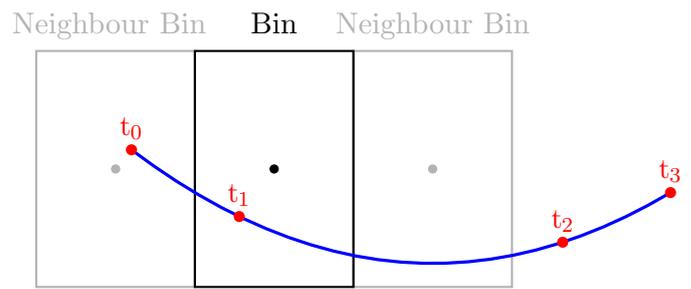


Figure 1. Sketch of the classical binning technique utilizing an evaluation at the particle positions.

either directly assigned to a certain bin simply by the partition given in 1 or further processed by additional weighting based on the particle position. To minimize the impact of the spatial averaging imposed by the size of the bin extent, particles can be weighted by their distance to their respective midpoint. An early example of such an approach is shown by [1] who use a Gaussian weight on particles manually digitized from photographic film. A different approach is taken by [2] who introduce a method based on a polynomial fit of particle positions collected within each bin. All particles of a bin are gathered and their mean velocity is expressed as a polynomial with respect to their position relative to the bin center. The fitted polynomial is then evaluated at the bin center to give the mean velocity value for that bin.

All these approaches have in common that they consider flow information at individual particle positions. In the following we want to motivate a different approach. This paper focuses on the multi-pulse STB method but the described approaches are suitable for the bin averaging of any kind of tracking data.

For high resolution statistics utilizing a large number of snapshots the bin size can be set very low. In these cases the bin size can get small compared to the track length even for the comparatively short multi-pulse STB tracks.

For the current applied methods we want to distinguish two different approaches with respect to the use of the available particles. The classic approach uses all of the fitted particles of the multi pulse track as shown in figure 1 with the track in blue and the fitted particles in red. Such an approach is close to the approach used for binning classic PTV evaluations such as presented by [12]. The multi-pulse data considered here is only evaluated on a short segment and with unequal particle spacing. A different way, which is currently in use for MP-STB, considers only the midpoint of a fitted track to provide a single virtual particle with low error at the cost of less particles that are available for binning. This is shown in figure 2. The first approach will be referred to as ‘Classic’ and the second as ‘Midpoint’ within this paper.

As can be seen in the figures even though the track passes through all of the shown bins, the fitted particles only contribute to some of them. When using the midpoint approach this is especially apparent as its midpoint actually lies in the neighboring bin so it will only contribute to the statistic there.

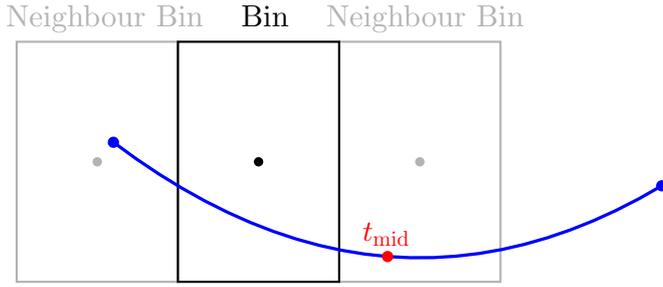


Figure 2. Sketch of the midpoint binning technique utilizing an evaluation at the midpoint of the fitted track.

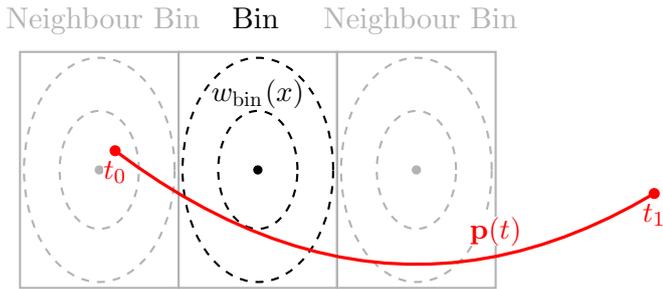


Figure 3. Sketch of a binning technique utilizing the functional approach. The entire track is evaluated and bins are represented by a spatial weight function shown here by the dashed contour lines.

Using just the midpoint ignores the additional data which is available due to the tracking and functional filtering scheme. It is not just individual particles that are available, but a track in form of a parametric function over time, e.g. 2nd/3rd order B-spline or polynomial fits along pulses.

One way this could be improved upon would be to sample additional points along the track, for example at the points closest to each bin center. While the closest point approach does utilize the track function to provide better data points for long tracks, it still does not extract all possible information. Still only a few points along a track are considered, even if more are available now and with better positions. The approach ignores the continuous nature of the available track. This motivates a functional approach where the entire continuous parametric track is considered. Figure 3 shows a graphical representation of such an approach which will be detailed in the following section.

2. Method

In the following the method is introduced as used for a multi-pulse STB setup in order to present a concrete implementation. These steps can be transferred analogously to time-resolved STB or different tracking approaches. Each track of a multi-pulse STB evaluation is represented by, for example, a quadratic polynomial for each dimension, giving the parametric functions:

$$p_i(t) = \alpha_i t^2 + \beta_i t + \gamma_i, \quad (2)$$

$$v_i(t) = \dot{p}_i(t) = 2\alpha_i t + \beta_i, \quad (3)$$

$$a_i(t) = \dot{v}_i(t) = 2\alpha_i, \quad (4)$$

for the position $p_i(t)$, the velocity $v_i(t)$ and acceleration $a_i(t)$ for the component $i \in (x, y, z)$ over time t . Instead of evaluating these functions at the midpoint t_{mid} as done for the midpoint bin averaging approach they are used in their entirety. A given bin is represented by a spatial weighting function. In the following the weight function:

$$w_{\text{bin}}(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)/c$$

$$\text{with } c = (2\pi)^{k/2} \det(\boldsymbol{\Sigma})^{1/2}, \quad (5)$$

the probability density function of a multivariate Gaussian distribution of dimension k centered at the bin center $\boldsymbol{\mu}$ with its shape determined by the co-variance matrix $\boldsymbol{\Sigma}$ is used. The use of a Gaussian allows for an intuitive parametrization based on the co-variance matrix but any other appropriate weighting function can be used. The current implementation uses the same diagonal $\boldsymbol{\Sigma}$ for all bins with the standard deviations σ_x , σ_y and σ_z set proportional to the bin extent for each coordinate direction. This is not a restriction imposed by the method as fully individual $\boldsymbol{\Sigma}$ for each bin are possible. This would for example enable adaptively shaping bins according to mean flow field gradients. It is also not required that the bin centers remain arranged in a regular grid as done in figure 3. Bins are fully independent of each other and can be freely positioned in space using their $\boldsymbol{\mu}$ value. For better comparison with other bin methods and easier visualization bins are placed in a rectangular grid for this paper.

In addition to the spatial weighting function $w_{\text{bin}}(\mathbf{x})$ assigned to each bin, a temporal weighting function $w_{\text{track}}(t)$ is assigned to each track. This allows to introduce additional information about the data quality along the track into the binning process. A detailed analysis of uncertainty quantification in Lagrangian particle tracking approaches lies outside of the scope of this paper, a comprehensive approach through the entire measurement chain is presented by [3]. Instead, we use a simpler approach starting from a known constant positional error of triangulated particles and consider only the track fitting process that then follows. An uncertainty propagation analysis through the tracking process allows to quantify the changing uncertainty along the track over time. The positional error for the triangulated particles is estimated based on realistic synthetic experiments and assumed to be uncorrelated across the separate pulses. While there are some sources of error that are correlated between the particles within a track, such as errors in the camera calibration affecting close by particles similarly, their influence is negligible compared to the random sources.

This known uncertainty is then propagated through the polynomial fit and the derivation to obtain a local uncertainty for the velocity along the track. For the four-pulse scheme

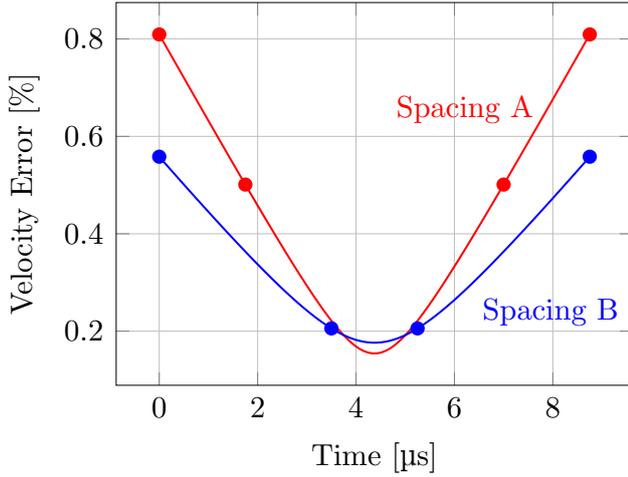


Figure 4. Example of the error analysis result used for determining the track weighting function $w_{\text{track}}(t)$ for a track with a duration of $8.75 \mu\text{s}$ for two different pulse spacings. The circles mark the time instants of the pulses.

evaluated here, this results in a value for the uncertainty that is smallest at the midpoint of the track and increases towards both ends. An example for such an evaluation is shown in figure 4 for two different pulse spacing strategies. Previous strategies (Spacing A in figure) for pulse spacing have among other aspects focused on reducing the error at the midpoint as this is the only location considered in binning making use only of the point with the highest velocity accuracy. With functional binning the integral error over the entire track length is of relevance and one could optimize pulse spacing with this in mind. Current practice for multi exposure based multi-pulse STB has changed towards greater spacing between the initial two pulses due to tracking requirements [13] (Spacing B in figure). This results in an error distribution that is slightly worse at the midpoint but exhibits a flatter trajectory over the rest of the track resulting in a lower integral error. Such a distribution is thus even better suited for the functional binning approach and is indeed close to the optimal pulse spacing when considering only the integral error. This optimal pulse spacing can be found by fixing the endpoints of the time interval and minimizing the integral error function in terms of the placement of the two middle pulses. The result places both pulses exactly on the midpoint of the time interval, a solution which might minimize the integral error but is not feasible for use in an actual tracking approach. While modifying the pulse spacing would be beneficial for the application of functional binning, it is not a necessity and all results shown use a non-optimized pulse spacing (Spacing A).

The calculated error function in figure 4 can then directly be translated into a weighting function $w_{\text{track}}(t)$. For the current implementation one single weight function is used for all tracks. If future improvements in uncertainty quantification within the STB method allow for individual position error estimates for each particle, this could be directly introduced here resulting in individual weighting functions tailored for each track.

The total weight function is then given by a combination of the two weighting functions, with the spatial weighting function 5 evaluated at the parametrized track position $\mathbf{x} = \mathbf{p}(t) = [p_x(t), p_y(t), p_z(t)]$:

$$w(t) = w_{\text{bin}}(\mathbf{p}(t)) w_{\text{track}}(t). \quad (6)$$

The mean velocity value $\bar{\mathbf{V}}$ of a track from time t_0 to t_1 with respect to this continuous weighting function can be expressed using integrals as

$$\bar{\mathbf{V}} = \frac{\int_{t_0}^{t_1} \mathbf{v}(t) w(t) dt}{\int_{t_0}^{t_1} w(t) dt}, \quad (7)$$

condensing the available velocity data from the track into a single value for a given bin.

When building the bin average over many tracks, each of the values needs to be weighted according to the properties of the track data it was calculated from. Tracks which pass closer to the bin center should be weighted higher than ones further away from it. Additionally the individual uncertainty values of the tracks need to be considered. Even if all tracks follow the same weight function this leads to differences as a low certainty outer region of one track should be weighted lower than the higher certainty middle region of a different track even if both are at the same distance to the center of the bin. These concerns are directly captured in the integral of the weight functions $w(t)$ used for the mean values of the individual tracks which are therefore used as weights for the bin averaging process. For an individual track with id n this leads to the velocity value $\bar{\mathbf{V}}_n$ and accumulated weight function \bar{W}_n :

$$\bar{\mathbf{V}}_n = \frac{\int_{t_0}^{t_1} \mathbf{v}_n(t) w_n(t) dt}{\int_{t_0}^{t_1} w_n(t) dt}, \quad \bar{W}_n = \int_{t_0}^{t_1} w_n(t) dt. \quad (8)$$

A weighted velocity mean \mathbf{M}_V for a bin over all N tracks can therefore be calculated by the following equation:

$$\mathbf{M}_V = \frac{\sum_{n=1}^N \bar{\mathbf{V}}_n \bar{W}_n}{\sum_{n=1}^N \bar{W}_n} = \frac{\sum_{n=1}^N \int_{t_0}^{t_1} \mathbf{v}_n(t) w_n(t) dt}{\sum_{n=1}^N \int_{t_0}^{t_1} w_n(t) dt}. \quad (9)$$

The actual implementation uses a different approach than just summing and dividing the terms in order to reduce numerical errors. A constant memory streaming weighted mean algorithm is used instead [9]. Other relevant bin averaging statistics can also be calculated at this point using the results given by 8 such as the co-variance matrix for the Reynolds stresses or other higher statistical moments. The approach is not limited to the analysis of velocities either, other quantities derived from the particle motion such as the acceleration can be considered analogously.

Depending on the choice of weight function, the integrals given in 8 generally do not have an analytical solution in terms of standard mathematical functions. If a constant track weight were used as is typical for time-resolved STB tracks, meaning

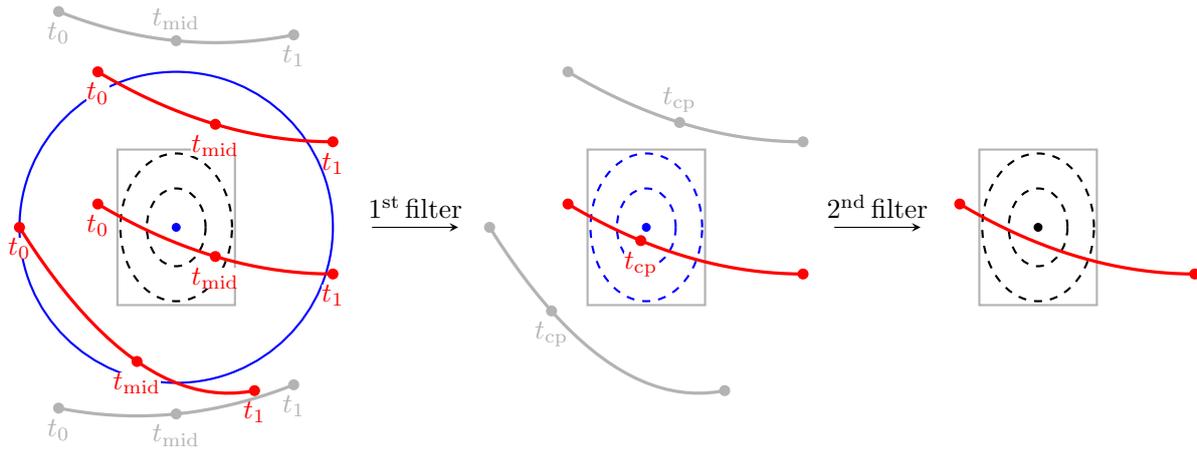


Figure 5. Double stages filtering process to reduce the number of tracks that need to be integrated. Tracks that are eliminated at each stage are shown in gray, filter regions in blue. The first stage removes far away tracks measured by their midpoints as a rough pre-filter. In the second stage the support of the weight function is used together with the closest point along the track to accurately determine relevant tracks.

only the spatial Gaussian weight is present, the upper term in by 7 can be expressed in terms of the error function:

$$\int \frac{1}{2}(2\alpha t + \beta) \exp(-(\alpha t^2 + \beta t + \gamma)^2) dt = \frac{1}{4}\sqrt{\pi} \operatorname{erf}(\alpha t^2 + \beta t + \gamma) + C. \quad (10)$$

But even for this simplification there is no analytical solution for the integral in the lower term containing only the weight function. Therefore a numerical solution for the integrals is required. Since all the functions can be continuously evaluated, a Gaussian quadrature rule is used for the integration.

While (9) provides a complete mathematical description of the procedure to calculate a mean velocity bin average, care must be taken to reduce computational effort as the numerical solution of the integrals is comparatively expensive. In the midpoint binning approach each particle only contributes to a single bin that is easily identified. The computational effort therefore only scales with the number of particles N_p not that of the bins N_b . A naive implementation of the functional method would require each bin to consider every particle resulting in a method that scales with the number of bins as well, being proportional to $N_b N_p$. Even without the increased effort of a numeric integration this would be infeasible. It is clear however, that only tracks within a certain neighborhood of a given bin need to be considered. The known spatial weighting function is therefore used as an indication to filter out tracks which are far away from the bin. Only tracks whose closest point to the bin center lies within the support of the weight function are considered. While the Gaussian window function for a bin chosen here technically does not have a compact support, a threshold can be set proportional to the standard deviation beyond which any contribution would be negligible. An analytical expression for the quadratic distance $d(t)$ of a track described by $\mathbf{p}(t)$ to the bin center at $\mathbf{p}_0 = [x_0, y_0, z_0]$ is given

$$\begin{aligned} d(t) &= \|\mathbf{p}(t) - \mathbf{p}_0\|^2 \\ &= (p_x(t) - x_0)^2 + (p_y(t) - y_0)^2 \\ &\quad + (p_z(t) - z_0)^2. \end{aligned} \quad (11)$$

The minimum distance is found by using the derivative

$$\begin{aligned} d'(t) &= 2((\beta_x + 2\alpha_x t)(\gamma_x + \beta_x t + \alpha_x t^2 - x_0) \\ &\quad + (\beta_y + 2\alpha_y t)(\gamma_y + \beta_y t + \alpha_y t^2 - y_0) \\ &\quad + (\beta_z + 2\alpha_z t)(\gamma_z + \beta_z t + \alpha_z t^2 - z_0)) \\ &= At^3 + Bt^2 + Ct + D = 0, \end{aligned} \quad (12)$$

requiring the solution of a cubic equation which can be done analytically. To reduce the computation time even further, a double staged filtering process is used. Before using the closest point filter described above, a rough pre-filter based on a Kd-tree range search of track midpoints is employed. Since only the midpoints are used this will be an inaccurate approximation of track distance so a generous exclusion range based on some multiple of the expected maximum track length should be used. The only goal is to use the very fast tree based search to exclude tracks which are so far away they could not possibly contribute to the considered bin. The only cost of setting the exclusion range too far is an increase in computational cost while setting it too low could exclude relevant data. Depending on the mean track length it could be advantageous to sample multiple points along the track for this filter stage in order to enable tighter bounds for the exclusion range.

With this combined setup the tracks are therefore directed through multiple stages of increasing computational complexity. The entirety of the tracks is sent to the first filter, requiring a Kd-tree based search. The relevant subset of the tracks is then directed to the next filter stage, requiring the solution of a cubic polynomial for each track which is done analytically. This final subset of the tracks is then directed to the actual binning stage, requiring the numerical calculation of integrals using Gaussian

quadrature rules. An initial implementation used an adaptive Gauß–Kronrod quadrature to ensure a low error for the integration result. As the considered functions are smooth a Gauß–Legendre quadrature using a suitable fixed number of nodes is used instead for faster performance. Additionally, since currently all tracks use the identical weight term $w_{\text{track}}(t)$ in (6), this term can be precomputed by directly including it in the determination of the quadrature rule leaving only the evaluation of the spatial weighting term $w_{\text{bin}}(\mathbf{x})$.

3. Application

The three different methods are evaluated on synthetic data based on a known ground truth velocity field:

$$u = \frac{\sinh(y)}{\cosh(y) + \varepsilon \cos(x - ct)} + c \quad \text{and} \quad (13)$$

$$v = \frac{\varepsilon \sin(x - ct)}{\cosh(y) + \varepsilon \cos(x - ct)}. \quad (14)$$

This field is taken from an example given by [10] and is based on the so called Stuart vortex representing a vortex moving along a shear layer. Within this work, the coefficients used are $c = 3$ and $\varepsilon = 0.7$. Synthetic four-pulse tracks are generated for each snapshot by advecting randomly placed particles using a differential equation solver. Each track is then sampled at the four pulses according to the prescribed timing strategy and artificial noise is added to represent particle position measurement error.

This approach allows to compare the values calculated by bin-averaging the synthetic snapshots to the known mean values of the velocity field in order to define an error value for the binning statistic. All velocity comparisons shown here use the u -component unless otherwise noted. The local ground truth mean value is calculated using the integral over the time domain:

$$\bar{u} = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} u(t) dt, \quad (15)$$

which can then also be used to calculate a ground truth value for the fluctuations:

$$\overline{u'u'} = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} (u(t) - \bar{u})^2 dt. \quad (16)$$

The synthetic data is generated for the spatial domain $\Omega = [-2\pi, 2\pi] \times [-\pi, \pi]$ and over the time domain $T = [0, 2\pi]$ equivalent to one period of the function. Snapshots are generated for random times within this interval.

To provide a good analog to actual applications, the parameters for the synthetic measurement are chosen such that they resemble a real world measurement.

The desired parameters from table 1 are used to appropriately scale the synthetic setup, leading to a scaling factor of approximately 32 px per length unit and 76 tracks per snapshot. A particle positioning error of 0.1 px is used for the generation of the synthetic tracks. Both methods are then applied

Table 1. Parameters used for the synthetic measurement study. These represent realistic values as found in a real experiment.

Parameter	Value
Particle image density	0.03 ppp
Number of snapshots	10,000
Length of tracks	30–80 px
Positioning error	0.1 px
Relative pulse spacing	(0.0, 0.2, 0.8, 1.0)

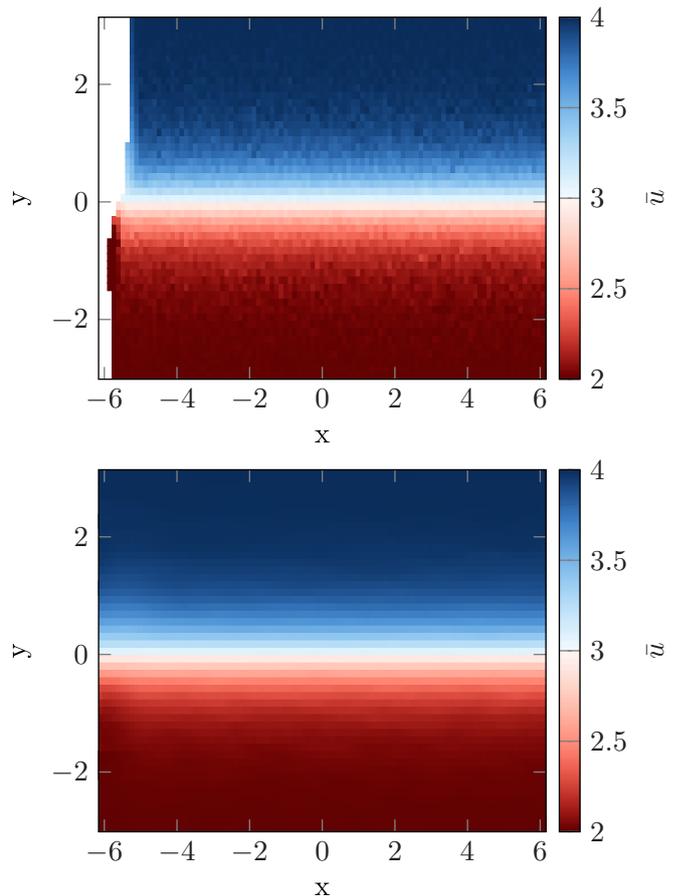


Figure 6. Mean velocity field for a bin size of 4 px calculated using 10,000 images with the the midpoint (top) and the functional binning approach (bottom). Bins with missing data are blanked out.

using high resolution two-dimensional bins with a side length of 2 px. An additional evaluation is conducted using larger bins with a size of 4 px.

Figure 6 shows the recovered mean velocity field using the 4 px bin size for both methods. Apart from the larger amount of noise for the midpoint method a region of missing data near the inflow region is also apparent, the size of which is increased for higher velocities. This is due to the initialization of synthetic particles only within the binning domain. When only the track midpoint is considered, bins that are closer to the inflow boundary that half the maximum track length show either no entries at all or a result biased towards the shorter, and therefore slower, tracks which still register. This issue is not present itself for the functional method as it considers the full length of the track. While this aspect highlights a nice feature of the

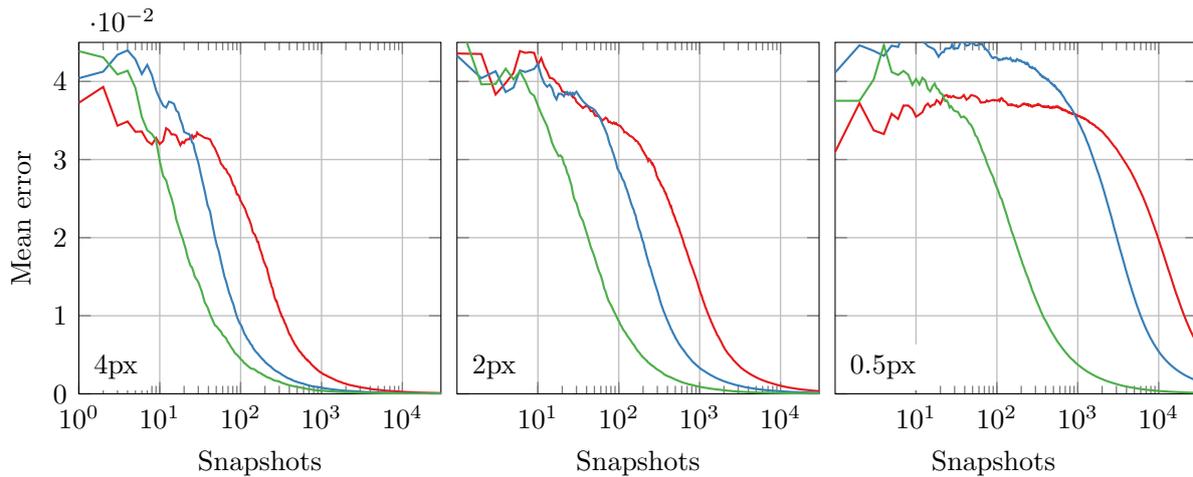


Figure 7. Evolution of the mean squared error of mean velocities over the number of snapshots used for the bin averaging shown for the different methods at three different bin resolutions. — Midpoint; — Classic; — Functional.

functional approach, it is often not as relevant in experimental investigations as it can be avoided by using a measurement domain larger than the binning domain of interest. To provide a fairer comparison between the two methods, all following calculations use a synthetic measurement domain that is large enough to prevent this issue from occurring.

While the visual difference between the two mean velocity fields in figure 6 provides some initial indication on the relative performance of the two methods, a systematic evaluation is facilitated through the available ground truth values.

The mean squared error over the entire field of the mean velocity can be used to quantify the quality of bin-wise reconstruction of the true mean field. Calculating the error over the amount of snapshots used allows to analyze the convergence behavior of the binning methods as shown in figure 7. The figure shows the results for the different methods for both binning resolutions and additionally for a resolution of 0.5 px. The improvement in convergence speed offered by the functional approach is immediately apparent. Even for the 2 px binning resolution convergence is reached well before the full number of snapshots while the midpoint approach barely does so using the full amount. The classic approach converges better than the midpoint variant as expected by the different amount of particles used. For the resolution of 0.5 px which has been included here as an extra evaluation one can see that the functional approach is less affected by the increase in resolution providing a viable approach for sub-pixel resolution ensemble averaging.

In practice one would not necessarily require full convergence of the desired statistics due to the diminishing returns of increasing the number of snapshots. If the error level the midpoint method exhibits using 5000 snapshots is deemed acceptable, less than 500 snapshots are required to reach the same error level using the functional method for this example. The classic method performs better than the midpoint approach and requires a level of about 2000 snapshots. The left figure shows the results for the coarser bins with 4 px length. It can be seen that the convergence penalty of moving from 4 px to the

smaller bin size is smaller in the functional method than in the classic one. This is inherent in the functional approach. While the total number of bins has increased and their size decreased, the ratio of track length to bin size has increased. The number of bins one given track contributes to thereby increases, counteracting some of the impact of using a finer resolution. This effect depends on the dimension of the binning grid as the region of influence of a track is in form of a line. In 1D the lines for both resolutions should lie on top of each other as bins are fully aligned with tracks in one dimension. For the 2D case as considered here, some decrease is expected as tracks are still quasi one dimensional while the bin size decreases in two dimensions. In 3D this effect will be stronger. As the same consideration applies for the classic approach, the benefit of using the functional approach is not diminished even in higher dimensions.

Statistical independence of functional binning is assured by the ratio of track length to bin size and the sparsity of the scattered track data available in 2D or 3D. The time length of four-pulse tracks consists of statistical dependent data typically with a length in the range of Kolmogorov time scales τ_η . The size of bins is a small fraction of the spatial extent of the tracks.

In addition to the analysis of the convergence of the mean, the dataset also allows to study the behavior of the convergence of velocity fluctuations. Figure 8 shows this using the same approach used for the mean velocities. The development of the error shows the advantage of the midpoint approach over the classic one. Although initially converging slower, the midpoint approach surpasses the classic approach after some time leading to higher accuracy for the fluctuation statistics. The new functional approach shows a significantly lower error than the other considered methods. To improve the classic approach, various additional weighting methods can be applied. We consider a spatial weighting by the distance from the bin center as well as a version that additionally includes an error based weight for the particles along a track. The results are shown in figure 9. While improvement in the error of the

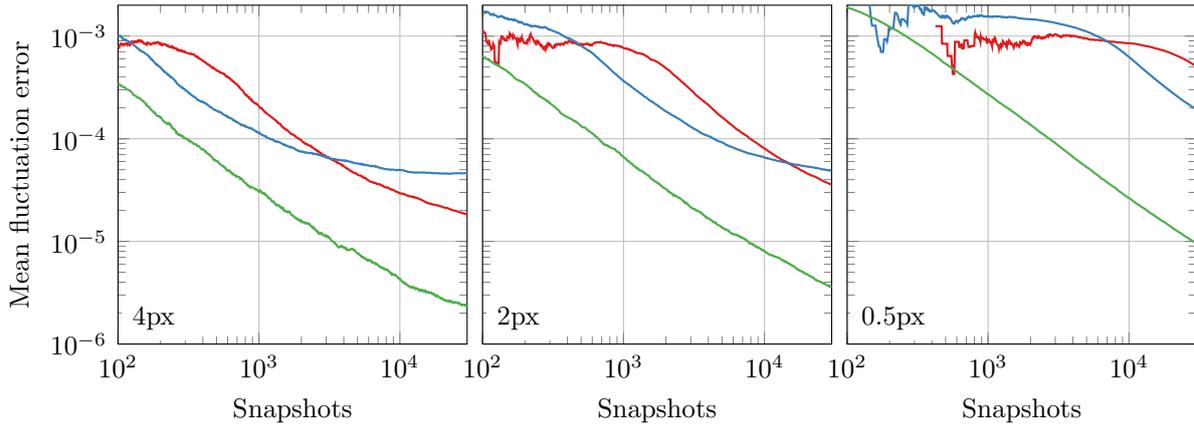


Figure 8. Evolution of the mean squared error of velocity fluctuations over the number of snapshots used for the bin averaging shown for the different methods at three different bin resolutions. — Midpoint; — Classic; — Functional.

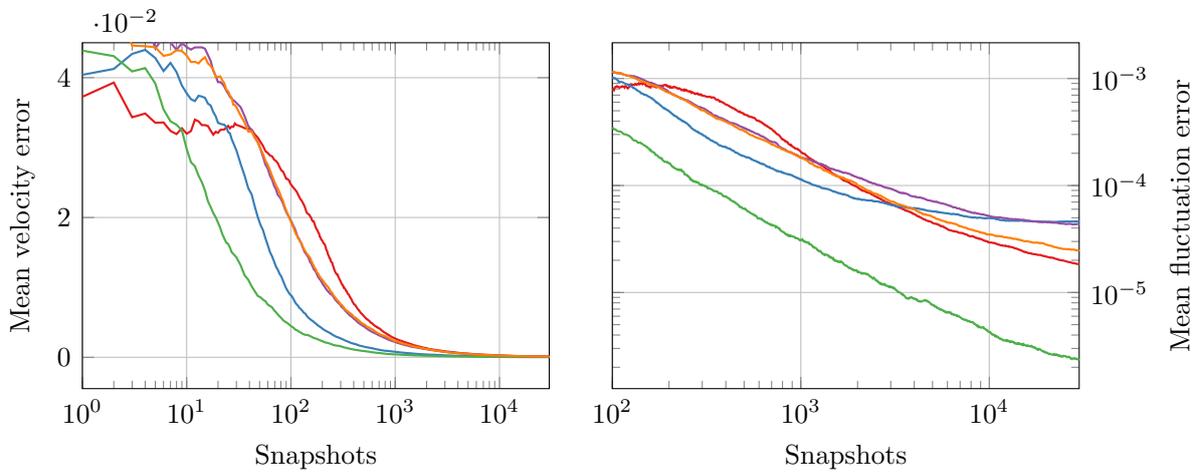


Figure 9. Evolution of the mean squared error over the number of snapshots used for the bin averaging with 4 px resolution shown for the different methods using the velocity mean (left) and velocity fluctuation (right). — Midpoint; — Classic; — Functional; — Classic with spatial weight; — Classic with error and spatial weight.

fluctuations is evident, the weighting approaches also have a negative impact on the convergence of the mean velocity. For all compared methods the functional approach remains superior.

To provide a better overview of the improvement afforded by the functional approach, figures 10 and 11 show the ratio of error of the midpoint approach and the different classic approaches to the error of the functional approach for the mean velocity and velocity fluctuations respectively. For the mean velocity value there is an initial development region where many bins have only few contributing particles. Due to differences in convergence speed there is a possible peak and the ratio then stabilizes towards a nearly constant value. The classic approach with both weighting strategies shows a higher ratio than the classic approach without weighting as expected from the results in figure 9.

For the velocity fluctuation a different picture emerges. The midpoint method approaches the lowest ratio to the functional method for a sufficiently large number of snapshots. A strong rise of the ratio for the classical method for a large amount of images is due to the larger inherent velocity error of this

approach. As seen in figure 9 the error plateaus for a large number of images and additional data does not significantly improve the result anymore. Since the functional approach is still improving at this stage, the ratio between them keeps rising. The weighting strategies can improve upon this somewhat by taking the error into consideration.

The reason for the differing performance of the approaches can be seen in the amount of particles contributing to the statistic for a given bin. Figure 12 shows such an analysis for the bin at the center of the domain. As expected the midpoint based approach registers the least amount of contributing particles, focusing on fewer entries with a higher quality. The classic approach with its use of four particles per track quadruples the number of entries for the bin. Using the functional approach a drastically higher number of tracks are able to contribute to the statistic. It should be noted that this is the raw number of contributing tracks independent of the weight value assigned to them. Therefore even tracks which are assigned a minuscule weight value are therefore included. Including the weighting of the tracks in this figure would seem like an obvious solution to better relate the functional method to the others but doing

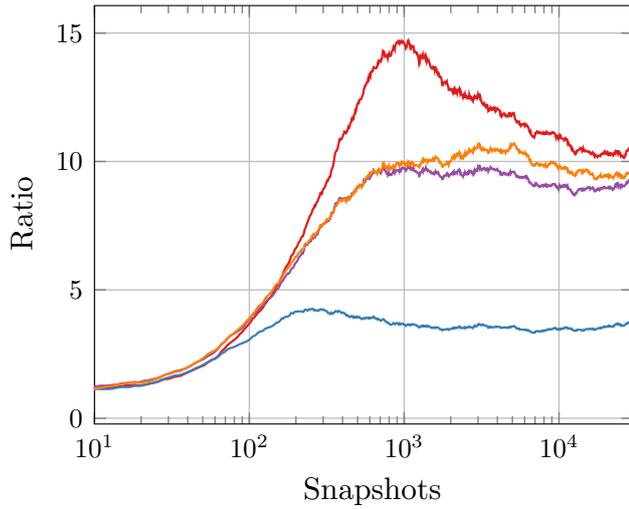


Figure 10. Ratio of mean velocity error of different binning approaches at 2 px resolution to that of Functional binning. — Midpoint; — Classic; — Classic with spatial weighting; — Classic with spatial and error weighting;.

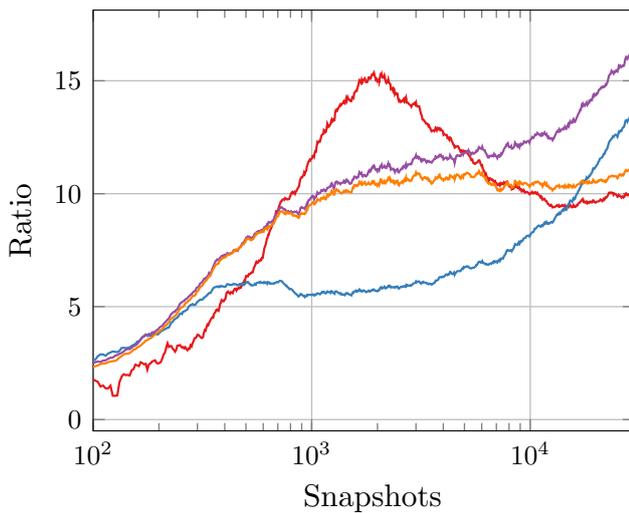


Figure 11. Ratio of mean fluctuation error of different binning approaches at 2 px resolution to that of functional binning. — Midpoint; — Classic; — Classic with spatial weighting; — Classic with spatial and error weighting;.

so would be detrimental for this aim. The individual weights used during the evaluation are consistent with respect to each other but there is no obvious normalization that would allow for comparison to the individual particles used in the discrete approaches.

The effect of the improved convergence behavior can be seen when examining a slice through the mean velocity field calculated without the full number of snapshots available as shown in figure 13 for 2000 snapshots. The functional approach results in a profile that is much closer to the true profile without much of the noisy artifacts still visible in the midpoint result. Figure 14 shows a similar comparison for the velocity fluctuation, again highlighting a better convergence

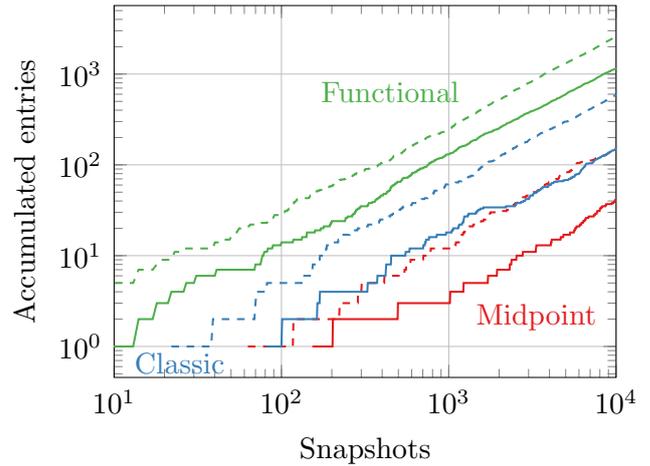


Figure 12. Number of contributions to the central bin over the number of snapshots shown for all methods at two bin resolutions — — — 4 px; — — — 2 px. The axis starts at 10 snapshots for this visualization.

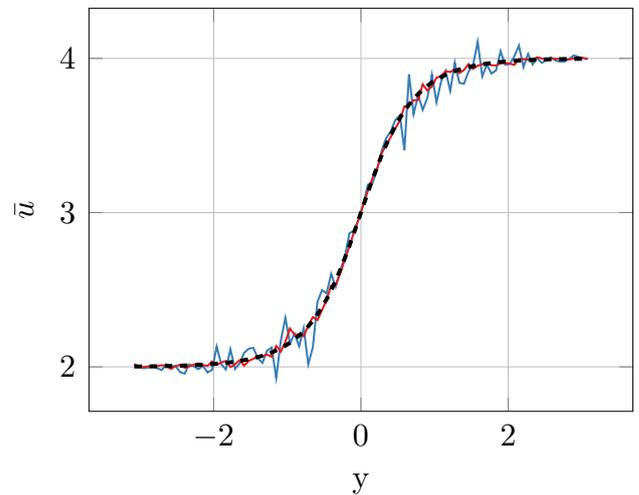


Figure 13. Slice through velocity mean field for bins of size 2 px using only 2000 images. — Midpoint; — Functional; — — — Truth.

towards the true result. When using the full amount of 30 000 snapshots as shown in figure 15, the functional approach follows the true solution very well except for some remaining deviations at the peaks. The classical solution, while visibly improved, still displays noisy artifacts and overestimates the fluctuation at the peaks. Using the classical solution with spatial and error weighting improves the situation as known from figure 9 but deviations are still visible. The midpoint solution shows more accurate results at the peaks but is still somewhat noisy. A better solution is provided by the functional approach with a smooth profile and accurate values at the peaks. Some small deviations are still visible.

One aspect that needs to be considered is the increase in computational effort. However the synthetic data set considered here is not well suited for such an evaluation. While the chosen parameters represent an actual measurement well in terms of the relative sizes, the amount of data involved is

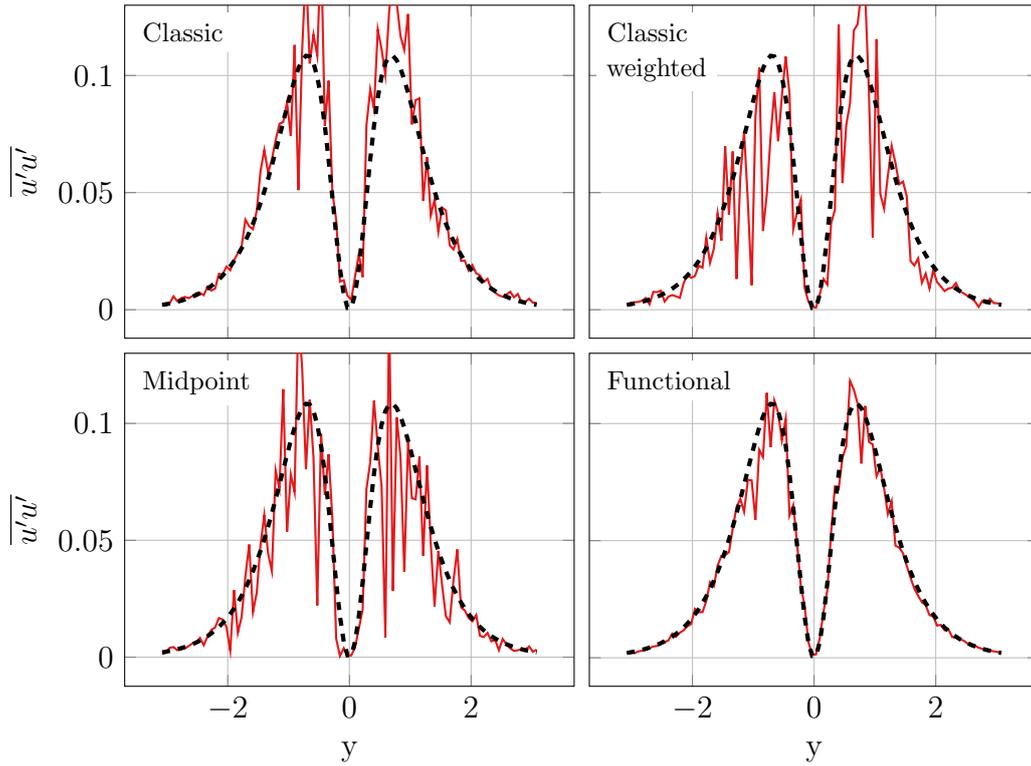


Figure 14. Slice through velocity variance field for bins of size 2 px using only 2000 images for different binning methods. — Binning result; - - - Truth.

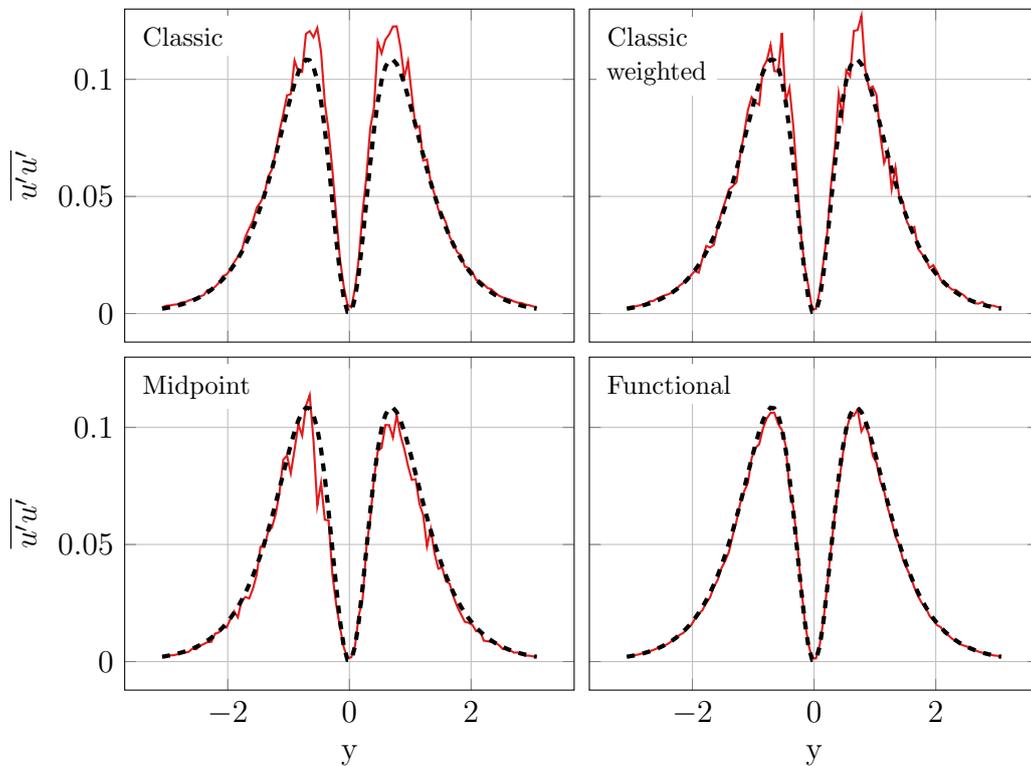


Figure 15. Slice through velocity variance field for bins of size 2 px using 30 000 images for the different binning methods. — Binning result; - - - Truth.

not representative. Only 76 tracks are used per snapshot due to the small image size. For a better example of the time requirement of the functional binning approach, it is applied to a real world 3D multi-pulse STB data set using 30 000 snapshots and approximately 60 000 tracks per snapshot. The results of this processing using 3D bins on a 16-core CPU are shown in table 2 for two bin sizes to highlight the impact of the amount of bins on the processing time. For the classic binning approach the run time is fairly independent of the amount of bins. This is the expected result as the approach is only dependent on the number of tracks used. The small difference seen here can be explained by variances in the time to read in the files from storage. The short run time of 10 to 13 minutes is a reflection of the simple assignment of each track midpoint to a single bin. Only a small amount of the time is dedicated to the actual binning, reading in the data from the file is the dominant factor. In contrast the functional binning requires much more time, here the actual computation effort dominates. For the larger bin size functional binning takes approximately 12 times longer than the classic approach. The time required scales with the number of bins as can be seen in the comparison with the finer bin size. Here the processing takes approximately 46 times longer than the classic approach for a total duration of 10 hours. Although the time required increases with the number of bins used, it does not do so directly. The time taken per individual bin decreases for greater bin numbers due to the efforts to consider only relevant tracks.

Reducing the computational requirements in the application of functional binning is an ongoing development effort. One immediate improvement can be realized by careful consideration of the spatial weight function. Within this paper, a Gaussian weight has been used to highlight a very flexible and intuitively parametrizable weight function. The ability to choose arbitrary weight functions provides opportunities to better adapt bins to the desired flow statistics. In many cases this flexibility is not necessary and simpler weight functions can be used to avoid the comparatively expensive evaluation of the exponential function within the Gaussian. One of such weights considered is the quartic function:

$$f(u) = \begin{cases} \frac{15}{16}(1 - u^2)^2 & \text{for } |u| < 1 \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

used for an anisotropic radial basis function with center μ , scaling s and \odot as the elementwise multiplication operator:

$$w_{\text{bin}}(\mathbf{x}) = f(\|(\mathbf{x} - \mu) \odot \mathbf{s}\|). \quad (18)$$

The scaling should be set based on the bin extent in each coordinate direction. Such a spatial weight then allows some anisotropy through stretching the radial basis in each coordinate direction but can no longer be tilted in arbitrary directions. For many applications this would be sufficient and current grid based binning methods face the same restrictions. The advantage in this weight function lies in the restriction to polynomial operations which can be computed much cheaper than the exponential required for the Gaussian weight with initial

Table 2. Runtime comparison for a real world 3D multi-pulse STB data set with 10 000 snapshots.

	Midpoint		Functional	
	8 px	4 px	8 px	4 px
Size	8 px	4 px	8 px	4 px
Bins	1.6 M	12.8 M	1.6 M	12.8 M
Runtime	10 m	13 m	1 h 56 m	10 h 1 m
Time per bin	356 μs	60 μs	4380 μs	2816 μs

tests showing an additional 20% reduction on the numbers shown in table 2. While there is a large increase in the time required to calculate a bin average from tracking data for the functional approach, one should consider the total processing times that are required. The midpoint binning approach is extremely fast so previously there was no need to approach it as a post-processing step with significant time or computing resource requirements. Functional binning now requires to view the ensemble averaging as a more elaborate process. Instead of the results being available within several minutes, an evaluation might now need to run over night. Considering the large amount of effort typically involved in acquiring the experimental track data, such run times are still acceptable. Especially if the improvement in convergence is taken into account. If we consider the improvement in figure 7 we can use 1000 snapshots for the functional binning of that measurement to reach similar convergence than the midpoint method with 10 000. As seen in figure 10 the error ratio of the various methods to the functional approach varies between 10 and 4. As the STB processing of a single typical multi-pulse snapshot takes approximately four minutes, the reduction of the necessary STB processing time from over 666 hours by a factor of ten easily dwarfs the increase in binning time. An additional factor is that this comparison was conducted using three-dimensional bins on a three-dimensional measurement. If only two-dimensional bins from such a measurement are desired because profiles are averaged along one dimension, the amount of bins decreases drastically thereby reducing the run time difference between the functional and classic approach. If desired, this difference can also be improved upon by including more computational resources. As the individual snapshots are independent of each other, the binning can easily be calculated in batches in parallel on different machines with near perfect scaling and the statistics then merged together.

4. Conclusion and outlook

By use of a functional approach to the spatial binning of particle tracking data better convergence of the desired statistics is achieved. The functional approach extracts all of the available information of the track and allows for easy inclusion of auxiliary data such as uncertainty quantification. The method provides a very generalized framework based on the weight functions that allows to tailor the evaluation to the individual needs.

An concrete implementation for synthetic multi-pulse STB data, short particle tracks, using Gaussian shape functions was shown and the benefits in convergence demonstrated.

Many other flow field statistics are at their core also based on a spatial binning approach and can equally be transferred into a functional binning framework. Besides the improvements in convergence evident in the calculation of the mean, the functional approach also has as an additional advantage when calculating statistics that involve considering pairs of particles as done for the calculation of two-point statistics and of spatial gradients respective velocity differences for dissipation rate. There the distance between particles is used to calculate two-point correlation functions. A focus lies on close distances which are relevant for the calculation of wave number spectra and length scales such as the Taylor micro-scale. Using the classical approach the resolution close to the origin is limited, especially for time-resolved tracking. Data is only available at the actual reconstructed particle positions themselves and therefore particle pairs cannot get arbitrarily close to each other due to imaging requirements. Particle images have a size of several pixels on the sensor and positioning error increases once they start to overlap. This introduces a lower bound for the achievable resolution near the origin. Using a functional approach for the tracks this can be improved upon since the parametric functions can get arbitrarily close to each other even if the relevant particles stay sufficiently far away from each other as shown in figure 16. The distance should never reach exactly zero as this would imply a collision but much closer approaches than suggested by the particle images are possible due to the diffraction limited imaging. While reaching very close distances is possible, the positioning error of the particles should be taken into account to achieve meaningful results which again imposes a lower bound for the minimum distance. As this error is in the order of 0.1 px the bound is much smaller than the one imposed by the approximately 3 px particle diameter, thereby allowing for increased resolution near the origin for two-point statistics.

The distance between two parametric functions can be expressed as another parametric function as shown in figure 17, directly allowing for functional binning in the relative distance space as is necessary for two-point statistics. Providing similar benefits as shown here for the velocity mean and fluctuations, this would allow for better resolution and convergence of two-points statistics. The binning in the relative space is not just filled by single points but along the full length of the function. As both tracks are parametrized via the same time variable it is ensured that no temporal dependence is introduced by differing time instants.

Other possible improvements can be introduced even without direct usage of the functional binning approach, using just the continuous nature of parametric tracks over time. Evaluations that require triplets or tetrahedrons of particles within a certain distance to each other for the calculation of statistics over spatial gradients are one example. Suitable pairings can be searched for along the entire length of the track, thereby increasing the chance of encountering a fitting one.

Apart from improvements to the individual statistics, the method also allows for a more flexible approach to the binning itself. The classic approach generally assumes a regular Cartesian grid for its binning domain, which can pose a challenge when flow around complex geometries needs to be

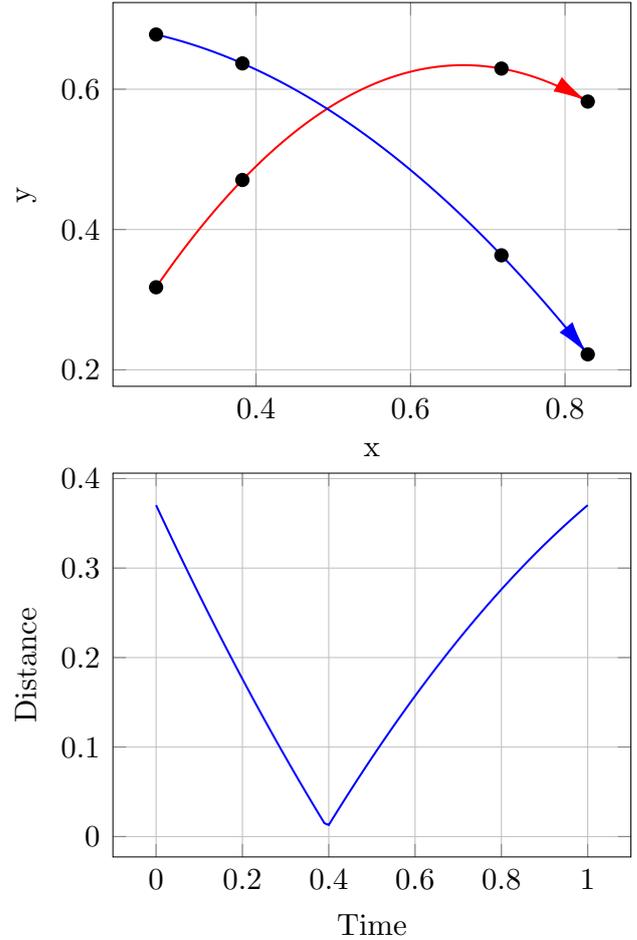


Figure 16. Sketch of two particle tracks in space and the distance between them over time. Pulse timings are shown by the black circles.

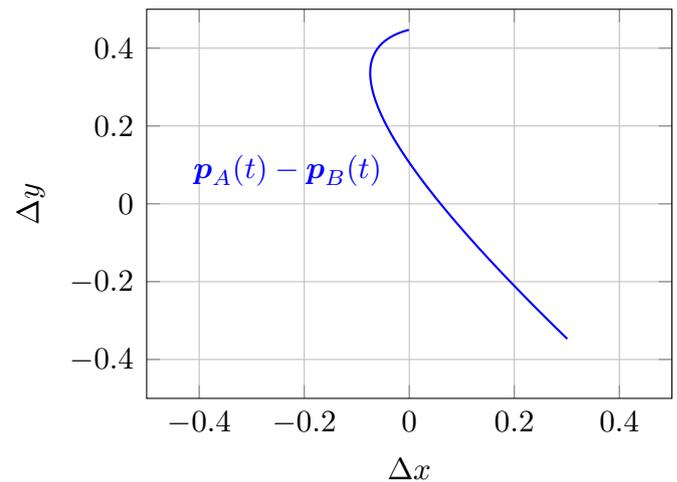


Figure 17. Sketch of the path of the position difference between two tracks in the relative space. Functional binning in this space would allow to calculate two-point correlations.

evaluated. Deviating from such a grid is possible but generally requires abandoning the very simple formula for particle to bin association given in (1) which drastically impacts performance. In functional binning a different approach already

has to be used by design thereby allowing for fully flexible placement and sizing of each bin without additionally reducing performance. The individual shaping of the weight functions for each bin allows for adaptive schemes that could orient them according to local mean flow properties such as velocity gradients. Depending on the nature of the flow field it could be advantageous to perform a rough fast initial binning using the classic approach to gather some preliminary information and then use these statistics to select appropriate local bin shapes for a functional binning evaluation. When working with complex geometries, bins in a region close to the wall could additionally be shaped to better conform to the local geometry.

Acknowledgments

The project leading to this journal article has received funding in the frame of the project HOMER from the European Union's Horizon 2020 research and innovation program under grant agreement No. 769237.

ORCID iD

Philipp Godbersen  <https://orcid.org/0000-0002-0917-4897>

References

- [1] Agüí J C and Jiménez J 1987 On the performance of particle tracking *J. Fluid Mech.* **185** 447–68
- [2] Agüera N, Cafiero G, Astarita T and Discetti S 2016 Ensemble 3d PTV for high resolution turbulent statistics *Meas. Sci. Technol.* **27** 124011
- [3] Bhattacharya S and Vlachos P P 2019 Volumetric particle tracking velocimetry (PTV) uncertainty quantification arXiv:1911.12495
- [4] Cornic P, Leclaire B, Champagnat F, Le Besnerais G, Cheminet A, Illoul C and Losfeld G 2020 Double-frame tomographic PTV at high seeding densities *Exp. Fluids* **61** 1–24
- [5] Delnoij E, Westerweel J, Deen N, Kuipers J and van Swaaij W 1999 Ensemble correlation PIV applied to bubble plumes rising in a bubble column *Chem. Eng. Sci.* **54** 159–5171
- [6] Fuchs T, Hain R and Kähler C J 2016 Double-frame 3d-PTV using a tomographic predictor *Exp. Fluids* **57** 174
- [7] Kasagi N and Nishino K 1991 Probing turbulence with three-dimensional particle-tracking velocimetry *Exp. Thermal Fluid Sci.* **4** 601–12
- [8] Kähler C J, Scholz U and Ortmanns J 2006 Wall-shear-stress and near-wall turbulence measurements up to single pixel resolution by means of long-distance micro-PIV *Exp. Fluids* **41** 327–41
- [9] Kraemer G, Gutwin M, Weidner J and Gans F 2019 WeightedOnlineStats.jl: v0.4.0 (Github)
- [10] Luchtenburg D, Noack B and Schlegel M 2009 An introduction to the POD Galerkin method for fluid flows with analytical examples and MATLAB source codes *Berlin Institute of Technology Technical Report* **11**
- [11] Maas H G, Gruen A and Papantoniou D 1993 Particle tracking velocimetry in three-dimensional flows *Exp. Fluids* **15** 133–46
- [12] Malik N, Dracos T and Papantoniou D 1993 Particle tracking velocimetry in three-dimensional flows *Exp. Fluids* **15** 279–94
- [13] Novara M, Schanz D, Geisler R, Gesemann S, Voss C and Schröder A 2019 Multi-exposed recordings for 3D Lagrangian particle tracking with multi-pulse Shake-the-Box *Exp. Fluids* **60** 44
- [14] Novara M, Schanz D, Reuther N, Kähler C J and Schröder A 2016 Lagrangian 3D particle tracking in high-speed flows: Shake-the-Box for multi-pulse systems *Exp. Fluids* **57** 128
- [15] Schanz D, Gesemann S, Schröder A, Wieneke B and Novara M 2012 Non-uniform optical transfer functions in particle imaging: calibration and application to tomographic reconstruction *Meas. Sci. Technol.* **24** 024009
- [16] Schanz D, Gesemann S and Schröder A 2016 Shake-the-Box: Lagrangian particle tracking at high particle image densities *Exp. Fluids* **57** 70
- [17] Scharnowski S, Hain R and Kähler C J 2011 Reynolds stress estimation up to single-pixel resolution using PIV-measurements *Exp. Fluids* **52** 985–1002
- [18] Schröder A, Schanz D, Michaelis D, Cierpka C, Scharnowski S and Kähler C J 2015 Advances of PIV and 4D-PTV 'Shake-the-Box' for turbulent flow analysis - the flow over periodic hills *Flow, Turbul. Combust.* **95** 193–209
- [19] Welford B P 1962 Note on a method for calculating corrected sums of squares and products *Technometrics* **4** 419–20
- [20] Wieneke B 2012 Iterative reconstruction of volumetric particle distribution *Meas. Sci. Technol.* **24** 024008