A direct spectral estimation method for laser Doppler data using quantization of arrival times

Nils Damaschke¹, Volker Kühn², Holger Nobach³*
1: Institute of General Electrical Engineering, University of Rostock, Faculty of Computer Science and Electrical Engineering, Germany
2: Institute of Communications Engineering, University of Rostock, Faculty of Computer Science and Electrical Engineering, Germany
3: Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany
* Correspondent author: holger.nobach@anambis.de

Keywords: LDV data processing, correlation function, power spectral density, random sampling

ABSTRACT

This paper presents a method for estimating the autocorrelation function and the power spectral density from laser Doppler data with discretized arrival times. The method can be realized by direct estimation of the correlation function or by direct spectral estimation with further transformations to allow appropriate normalization including corrections for some deviations from the ideal Poisson sampling process like processor dead times. The method also makes use of processing steps, some of them initially developed for other estimation methods, like sample weighting, treatment of self-products, the above mentioned normalization or an effective reduction of the spectral resolution with most efficient use of information available. An example application on publicly available laser Doppler data shows agreement between the results obtained with competing methods. Furthermore, under this fair comparison, some methods converge in terms of their systematic and random errors, indicating that they are comparably efficient at using the available information content of the randomly sampled signal. The results also identify that the available methods are interchangeable and indicate a possible replacement for the current best-practice procedure in the laser Doppler community.

1. Introduction

In burst-mode Laser Doppler velocimetry (LDV) [4,21,62] the fluid flow under investigation is sampled by randomly distributed tracer particles carried along by the flow. Particles crossing the measurement volume of the system lead to individual estimates of the arrival time and the velocity, yielding a nonuniformly sampled data set. Additionally, the system evaluates the transit time (also called the residence time), which is the time the particle needs to cross the measurement volume.

An ideal Poisson process would result in an exponential distribution of inter-arrival times of consecutive samples. Unfortunately, LDV introduces deviations from this ideal random sampling. Particles with very short separations may lead to interfering signals. To avoid
subsequent errors, such signals are identified and rejected by the measurement system. Therefore, the data set has a certain minimum distance between consecutive particle arrivals. The corresponding minimum interval plus other temporal delays of the measurement system is known as the processor dead time. These delays set the effective limit of the temporal resolution achievable by the measurement system. Typically, this limit is much higher than the mean data rate in practical applications. However, it influences the distribution of sampling intervals and it affects the statistical characteristics of the sampled velocity signal.

The sampling rate increases with increasing velocity since more particles pass through the measurement volume at higher volume flux. This introduces a bias to all statistical quantities derived from the obtained data ensemble [22, 35]. During the passage through the measurement volume the particle generates a Doppler signal. The extraction of a velocity value from this Doppler signal comes with a certain estimation uncertainty. This random error produces an additional white noise superimposed upon the velocity samples.

The statistical analysis of LDV data requires procedures, which are suitable under these conditions. The following procedures to derive estimates of the autocorrelation function and that of the power spectral density have found acceptance with irregularly sampled LDV data. For a review see [20].

- Direct spectral estimation derives the spectrum (spectrogram) directly from the given samples. The randomly sampled signal is interpreted as a series of Dirac pulses with gaps between them. Assuming the gaps are filled with a constant value of zero, then the Fourier transform is defined for this continuous time-limited signal. Representatives of direct estimation from LDV data can be found in [14, 25, 26, 27, 28, 33, 46, 57] with variations in deriving appropriate normalization factors for the spectra, treatment of self-products, in applying window functions and weighting schemes. Even though direct spectral estimation is a pioneering method for the processing of randomly sampled data, it has initiated a new debate recently. One consequence is the further development and adaptation of advanced methods, originally developed for other processing types [15, 16, 17, 40, 41, 55, 56, 57]. Practical implementations of direct spectral estimation usually perform processing steps on the corresponding correlation function. Therefore, heavy use is made of the Wiener-Khinchin theorem [32], formulating the correspondence between the correlation function and the spectrum via the Fourier transform.

- The slotting technique primarily derives an estimate of the correlation function (correlogram). The slotting technique is performing ensemble averages for a specific lag time by binning pairs of samples falling into a certain interval $\Delta \tau$ of inter-arrival times.
Examples of the slotting technique can be found in [14, 25, 28, 33, 34, 37, 39, 44, 46, 50, 51, 52, 54] with variations in applying windowing, weighting schemes, use of self-products, slot boundaries and special processing options like local normalization, fuzzy slotting and variable windowing. Since the technique needs to determine the time between the arrivals of every two of the data samples, this method primarily is computational heavy. This drawback is significantly reduced by defining the total support of the correlation estimate to be significantly smaller than the duration of the data set \( T \), e.g. \( K \) different lag times equidistantly spaced with intervals of \( \Delta \tau \) with \( K \Delta \tau \ll T \). This way only pairs of samples must be counted, which are within the maximum time delay specified. The support of the correlation function of course must include the entire range, where correlations between the samples occur. The correlation function finally can be transformed into a spectrum by applying the discrete Fourier transform incorporating again the Wiener-Khinchin theorem [32].

- Interpolation of the randomly sampled data set followed by a uniform resampling yields a new data set, which can be processed using classical signal processing methods for uniformly sampled data. Various interpolation schemes have been investigated [3, 5, 8, 11, 13, 18, 19, 31, 43, 48, 49, 53, 58]. Unfortunately, at low mean data rates, the interpolation introduces a significant dynamic error to the autocorrelation function and to the power spectrum [3, 11]. The spectral distribution of the interpolated signal then becomes a fingerprint of the interpolation method used [12]. This systematic error becomes acceptable only at very high data rates, about five times higher than the highest significant frequency in the spectrum of the underlying process. To correct this dynamic error at lower mean data rates, a deconvolution attempt known as refinement can be applied [36, 43, 45, 47]. However, the derivation of the refinement strictly relies on the idealized randomness of the sampling. Even small deviations like processor dead times will cause remaining systematic errors after deconvolution. Suitable normalization attempts for appropriate correction are not available for the interpolation method. Individual sample weighting is also not available for this estimation method. However, interpolation implicitly yields more resamples of the interpolated values if the instantaneous data rate is small. Effectively, this is a kind of intrinsic inter-arrival time weighting. Unfortunately, this one is efficient only at high enough data rates, at least ten samples per integral timescale. For the estimation and removal of the errors due to superimposed noise model-free [42] as well as model-based procedures [45] exist.
For interpolation neither implementations of appropriate weighting schemes nor normalization procedures have been realized so far. Since these two processing details are essential for bias-corrections, interpolation lacks behind the other two methods. On the other hand, direct spectral estimation and the slotting technique can reach equivalent estimation quality in terms of systematic and random errors. This requires, that they are used under equivalent conditions and with identical processing options like sample weighting, removal of self-products, normalization and limitation of the support of the correlation function [40, 41]. Reaching equivalent estimation quality, there remains the question of the computational effort of the different estimation methods. While the slotting technique needs to sum all pairs of samples within a certain time interval, direct spectral estimation needs to perform a continuous time Fourier transform. Therefore, both methods need extensive computations.

Recently, quantization of arrival times has been found very useful to accelerate the calculation of appropriate normalization factors for direct spectral estimation [15]. However, quantization of arrival times has been investigated previously with the slotting technique [46]. The results have been found to be different but of equivalent estimation quality. This has triggered to extend the application of arrival-time quantization also to the entire direct spectral estimation. Quantization of arrival times leads to a quasi-uniform sampling of the data set. This way one may take advantage of fast computer routines like the fast Fourier transform, which accelerates the computation of the autocorrelation function and the spectrum significantly. The final procedure yields yet another estimation of the autocorrelation function and the power spectral density. However, comparison to results from the slotting technique and those from direct spectral estimation show equivalence in terms of systematic and random errors.

2. Arrival-time quantization

Assuming the velocity \( u(t) \) as a function of time is sampled irregularly at instances \( t_i \), yielding a data set of \( N \) samples \( u_i = u(t_i) \) with \( i = 0 \ldots N - 1 \). For each measured value \( u_i \) a weight \( w_i \) is introduced to suppress the bias associated with the correlation between the instantaneous convection velocity and the conditional expectation of the sampling rate. The procedures given here can make direct use of transit-time weighting [14, 30], which will be followed through the rest of the paper.

The random particle-arrival times \( t_i \) are getting discretized with a constant temporal increment of \( \Delta \tau \). The result is a quasi-uniformly sampled data set. The time instances are sampled sparsely with many missing instances in the gaps between the original samples. By masking the missing
instances with corresponding weights set to zero, new uniformly sampled data sets can be defined: $u'_i$ for the weighted velocity samples and $w'_i$ for the series of weights

$$
u'_i = \sum_{j=0}^{N-1} w_j u_j$$

$$|i\Delta \tau - t_j| < \Delta \tau / 2$$

$$\tag{1}$$

$$w'_i = \sum_{j=0}^{N-1} w_j$$

$$|i\Delta \tau - t_j| < \Delta \tau / 2$$

$$\tag{2}$$

with $i = 0 \ldots N' - 1$, where $N' = 2T / (\Delta \tau)$ and $T$ is the duration of the original data set, where implicitly the data sets have been extended by zero padding. This uniform resampling is comparable to the interpolation technique, but without the interpolated values between the original samples. The prime symbol indicates that these are new data sets with uniform resampling. Note that $u'_i$ already includes the appropriate weights. Furthermore, multiple original samples falling into the same time interval are linearly superimposed (summation of their values). For these uniformly sampled series $u'_i$ and $w'_i$ standard procedures can be used for the primary processing. With discretized arrival times, the discrete Fourier transform then can be obtained using a fast Fourier transform.

The autocorrelation function can be computed directly from the time-discretized series as

$$R(\tau_k) = \frac{\sum_{i=0}^{N'-1} u'_i u'_{i+k} - \sum_{i=0}^{N-1} w'_i u'^2_i}{\sum_{i=0}^{N'-1} w'_i w'_{i+k} - \sum_{i=0}^{N-1} w'^2_i}$$

$$\tag{3}$$

with $\tau_k = k \Delta \tau$. Note the removal of self-products from both the numerator and the denominator. This removal avoids problems due to different probability densities of the occurrences of self- and cross-products of the original data samples. At the same time, systematic errors due to white noise components in the data values are suppressed [33].

Alternatively, the correlation function can also be obtained via direct spectral estimation incorporating the Wiener-Khinchin theorem [32]. In this case, the primary spectra are obtained through

$$S_u(f_j) = T \frac{|U'(f_j)| - \sum_{i=0}^{N-1} w'^2_i u'^2_i}{|W'(0)| - \sum_{i=0}^{N-1} w'^2_i}$$

$$\tag{4}$$
\[ S_w(f_j) = T \left| \frac{W'(f_j)}{\sum_{i=0}^{N-1} w_i^2} \right| - \sum_{i=0}^{N-1} w_i^2 \]  
\[ \frac{1}{\sum_{i=0}^{N-1} w_i^2} \]  

(5)

with

\[ U'(f_j) = \text{DFT}\{u'_i\} = \sum_{i=0}^{N'-1} u'_i \exp(-2\pi if_j t_i) \]  
\[ W'(f_j) = \text{DFT}\{w'_i\} = \sum_{i=0}^{N'-1} w'_i \exp(-2\pi if_j t_i) \]  

(6)

(7)

and with the imaginary unit \( i \) and \( f_j = j/(N' \Delta \tau) \) and \( j = -\lceil N'/2 \rceil \ldots \lceil (N' - 1)/2 \rceil \) incorporating the discrete Fourier transform (DFT). Note again the removal of self-products to correct the systematic error of the spectrum due to the random sampling. At the same time, both systematic errors due to white noise components in the data values and the problem of different probability densities of self- and cross-products are suppressed. Because this method is only a variant of direct spectral estimation with discretized arrival times, weighting schemes can be applied as before, especially the preferable transit-time weighting. The autocorrelation functions of the two primary spectra are then obtained via the inverse discrete Fourier transform (IDFT).

\[ R_u(\tau_k) = \frac{1}{\Delta \tau} \cdot \text{IDFT}\{S_u(f_j)\} = \frac{1}{N' \Delta \tau} \cdot \sum_{j=-\lceil N'/2 \rceil}^{\lceil (N' - 1)/2 \rceil} S_u(f_j) \exp(2\pi if_j \tau_k) \]  

(8)

\[ R_w(\tau_k) = \frac{1}{\Delta \tau} \cdot \text{IDFT}\{S_w(f_j)\} = \frac{1}{N' \Delta \tau} \cdot \sum_{j=-\lceil N'/2 \rceil}^{\lceil (N' - 1)/2 \rceil} S_w(f_j) \exp(2\pi if_j \tau_k) \]  

(9)

with \( \tau_k = k \Delta \tau \) for \( k = -\lceil N'/2 \rceil \ldots \lceil (N' - 1)/2 \rceil \). The normalized autocorrelation function, directly comparable to Eq. (3) as with the slotting technique is

\[ R(\tau_k) = \frac{R_u(\tau_k)}{R_w(\tau_k)} \]  

(10)
Up to this point, the autocorrelation function still has $N'$ values. The spectrum corresponding to this long autocorrelation function has a high spectral resolution for the price of a high estimator variance. To avoid disadvantages of block subdivision and averaging of the data set or modulation of the data set by windowing functions, and to make use of maximum information available, the support of the correlation function is truncated in a post-processing step. This finally leads to a reduced resolution and a reduced estimator variance of the spectrum. Assuming that the autocorrelation function, computed for $K$ lag times with $K \ll N'$ is negligible outside the interval $k = -[K/2]...[(K - 1)/2]$. Then the corresponding power spectral density can be obtained from the shorter autocorrelation function via the DFT as

$$S(f_j) = \Delta \tau \cdot \text{DFT}\{R(\tau_k)\} = \Delta \tau \cdot \sum_{k=-[K/2]}^{|(K-1)/2|} R(\tau_k) \exp(-2\pi if_j \tau_k)$$  \hspace{1cm} (11)

for $f_j = j/(K\Delta \tau)$ and with $j = -[K/2]...[(K - 1)/2]$. These basic procedures have been extended by further optional methods, namely forward-backward inter-arrival time weighting, fuzzy techniques, local normalization and correction for short data sets or blocks respectively, which are available as processing options with the program source code at [1]. Forward-backward inter-arrival time weighting is efficient only if the mean data rate is sufficiently high, at least about ten samples per integral timescale. If transit times are available, transit-time weighting should be given preference. With a recent adaptation of fuzzy slotting [44, 52] to direct spectral estimation it has been determined to effectively low-pass filter the data [41]. Since this is not useful for a non-parametric estimation, the application can no longer be recommended. Local normalization [51, 54] can reduce the estimator variance of the correlation estimate at lag times with large correlation and also the estimator variance of the spectrum. There have been no cases reported, where local normalization has reduced accuracy.

If the mean value is obtained from the randomly sampled signal and subtracted out from the data set before correlation and spectral estimation, the reduced degree of freedom will lead to a systematic error in both, the correlation estimate and the spectrum. In this case, the estimator is only asymptotically bias-free, meaning that this error vanishes for long enough data sets even without an appropriate correction. Since the method does not require block subdivision, the remaining error should be small enough with long enough data sets or blocks respectively. Due to the temporal limitation of the correlation function as a post-processing step, the estimator can handle large data blocks or the entire data set. If block subdivision is still used, the block size should be chosen sufficiently large to keep this bias small. However, an exact correction for the
covariance estimate is introduced in [59]. The principle requires adaptation to the case of weighted averages of irregularly sampled data. A practical and useful zeroth-order approximation (constant value) of the correction for LDV data can be found in [40]. For shorter data sets, the constant-value correction is a helpful means to reduce this particular bias. However, the correction can be used by default also with longer data sets without any risk of an overcorrection or making the bias worse.

All these extensions have been implemented as options to the programs without giving details in the present article. If required, detailed descriptions and ready-for-use programs can be found at [1].

3. A fair comparison

Thus far, the above processing methods have preferably been used with different processing options, e.g. with different weighting schemes, different treatment of self-products, different normalization attempts, different transforms between the correlation function and the spectrum, window functions, block subdivision and averaging etc.. Of course, this leads to different results and depending on a clever combination of such processing parameters for one particular method and a bad combination for another one, the comparison between different estimation methods can be arbitrarily influenced as desired. A fair comparison of the methods requires the adaptation of such processing options to all methods as well as the application of identical processing parameters. Therefore, the processing principles above for estimating the autocorrelation function and the power spectral density have been homogenized in terms of boundary conditions, optional processing techniques and all relevant processing parameters. Only this can lead to a fair comparison of the various. For all procedures used here Python source code is online available at [1].

3.1 Data Weighting

To suppress the statistical bias originated from the correlation between the instantaneous data rate and the convection velocity, effective data weighting is used with all the methods where available. For example, the advantage of transit-time weighting [14, 30] has been known for decades as well as the application of this weighting scheme to direct spectral estimation [28]. Other estimation methods have been introduced and preferentially applied with different weighting schemes. This has led to the wrong impression that the methods can only be used
with particular weighting schemes. However, the weighting schemes can be applied to the various estimation methods interchangeably. Detailed investigations and comparisons between various weighting schemes [22, 23, 24, 61] have shown that transit-time weighting should always be the first choice. Accordingly, transit-time weighting is used with all estimation methods for the following comparison. This holds for the slotting technique and for direct spectral estimation including arrival-time quantization, where the direct application of weighting schemes is possible [14, 15, 27, 28, 37, 39, 40, 41, 57]. For the interpolation method, individual data weighting is not available. Instead interpolation inherently performs a quasi-inter-arrival time weighting, which of course will lead to a different behavior of this processing method compared to the other three methods.

3.2 Mean removal and correction for short data sets

In flow measurements, the correlation function and the spectrum are typically understood as those of the fluctuations about the mean value. Since the true mean value is typically unknown, the estimated mean value obtained from the same data set is subtracted out instead. Applying appropriate weighting, preferably transit-time weighting, the mean estimator is widely accepted to yield bias-free estimates. However, the estimation has an uncertainty and subtracting the estimated mean value from the original signal reduces the power of the original signal by this uncertainty. A systematic error of the correlation estimate results. An exact correction for the covariance estimate is introduced in [59]. However, the principle requires adaptation to the case of weighted averages of irregularly sampled data. A practical and useful zeroth-order approximation (constant value) of the correction for LDV data can be found in [40] including the correction of the variance estimate. The methods used here all are asymptotically bias-free, meaning that this error vanishes for long enough data sets even without an appropriate correction. Since the methods do not require a block subdivision, the remaining error can be kept small with long enough data sets. If block subdivision is still used, the block size should be chosen sufficiently large. For shorter data sets, the constant-value correction is a helpful means to reduce this bias. It can also be applied by default to longer data sets. There is no risk of an overcorrection or making the bias worse.
3.3 Removal of self-products

Both, the slotting technique as well as direct spectral estimation, originally sum over all pairs of samples including both, self-products of single samples and cross-products of pairs of different samples. Unfortunately, the lack of a fixed spacing between samples, inherent to the measurement method, leads to different probabilities of self- and cross-products. The removal of all self-products from summations during the signal processing avoids averages over events of different probabilities. At the same time, systematic errors of the autocorrelation function and the spectrum due to uncorrelated noise superimposed upon the data are suppressed [28, 37, 50]. Furthermore, direct spectral estimators applied to the randomly sampled signal show a systematic error caused by the fact that the spectral characteristics of the randomly sampled signal deviate from those of the underlying process. In [26] an analysis of this error is given including the correction by removing the self-products.

Accordingly, self-products are removed from all summations in the investigated methods where possible, which is the case for direct spectral estimation, the slotting technique and the arrival-time quantization. For interpolation, appropriate corrections for errors due to noise requires a separate post-processing step. There are both model-based [45] and model-free procedures [42] for the estimation and removal of the errors due to noise. Following the intention of a model-free estimation, model-free estimation and removal of noise contribution has been used here for consistency.

3.4 Normalization

With the slotting technique, the estimated correlation values for each lag time are the result of an averaging over a finite number of pairs. This procedure includes a normalization of the sum of cross-products by the number of cross-products individually for each lag time. This holds also for the application of weighting schemes. This normalization process is capable to tackle the influence of deviations of the sampling process from ideal random sampling, e.g. due to processor dead times.

For direct spectral estimation, the spectrum of the randomly sampled signal equals the convolution of the spectrum of the underlying, continuous signal and that of the random sampling function. Therefore, a spectrum corrected for such deviations from the purely random sampling, can be obtained through a deconvolution of the spectrum of the randomly sampled signal and that of the sampling function [15]. This procedure is advantageously applied to the
respective spectra after removing all self-products, to avoid remaining errors due to different probabilities of self- and cross-products or due to noise. The procedure is suitable for moderate deviations from ideal random sampling only. However, it is sufficient to correct the influence of processor dead times. The deconvolution can be efficiently performed through a normalization of the autocorrelation function of the randomly sampled data set with the autocorrelation function of the sampling function incorporating the Wiener-Khinchin theorem. Of course, weighting schemes must be considered in all of these processing steps.

For the interpolation method, suitable normalization attempts are not available, which will also contribute to a different behavior of this processing method compared to the other three methods.

### 3.5 Local normalization

An alternative normalization, known as local normalization [51, 54], derives the correlation coefficient as a function of the lag time. Here, the autocorrelation estimate of each individual slot is normalized by a “local” estimate of the variance, based on the same selection of data samples used for the autocorrelation estimate. Local normalization has recently been adapted to direct spectral estimation [41]. Therefore, it is also available for arrival-time quantization and can be used as an option with the investigated processing methods, except for the interpolation, where local normalization is not available yet.

This technique can reduce the estimator variance of the autocorrelation estimate at lag times with large correlation and finally the estimator variance of the spectrum. Note, that in the case of noise superimposed to the data some of the involved summations get biased, while others remain bias-free. The result is a bias in the normalized correlation coefficient. This bias cancels out in a re-normalization process incorporating an estimate of the variance from the entire data set [37, 38, 39, 41]. There have been no further cases reported, where local normalization has reduced accuracy.

### 3.6 Truncation of the support of the correlation function and no too short block subdivision

The direct estimation of the power spectral density from a single data set has a very high estimator variance. A common means to reduce the variance is a subdivision of the data set into shorter blocks. The average of the power spectral densities of data blocks then has a significantly smaller variance. This method is known as Bartlett’s method [6, 7]. A disadvantage of Bartlett’s
method is that correlations between the samples at the end of one block and the beginning of the next are not counted. Furthermore, the wrap-around error may be increased if the assumption is made that the signal respectively the block is periodic. For too short blocks this may lead to significant deviations. In contrast, for longer blocks the reduction of the estimator variance becomes less effective. With Welch’s method [60], where the statistical functions from overlapping blocks get averaged, correlations between block boundaries are counted. However, this also partially generates redundancy. This fact is taken into account by applying windowing functions to the data blocks prior to their statistical analysis. Unfortunately, the additional modulation of the signal introduces an additional bias to the estimates of the statistical functions. A direct count and subtraction of cross-products between the beginning and the end of each block and addition of cross-products between the end of one and the beginning of the next block can more accurately correct the effects of block subdivision. However, this is computationally costly.

A powerful alternative without the necessity of block subdivision is the reduction of the spectral resolution in a post-processing step. First, the spectrum is obtained for the entire data set or with sufficiently long blocks to avoid systematic errors. The estimator variance is accordingly high. Such a high-resolution spectrum can be transformed through the inverse discrete Fourier transform into a corresponding autocorrelation function. The high spectral resolution translates into a large support of the autocorrelation function. Reducing the spectral resolution corresponds to a shorter support of the correlation function. Assuming a random process with arbitrarily long but finite memory, the autocorrelation function is zero at longer lag times. In this case the autocorrelation function can be shortened to the extent of the longest lasting correlation without losing information. Finally, the spectrum can be obtained from the shorter autocorrelation function via the discrete Fourier transform. Due to the shorter support of the autocorrelation function, the spectral resolution is reduced accordingly, leading also to a significantly lower estimator variance of the spectrum without introducing new errors from too short a block subdivision. The advantage of this method compared to usual block subdivision is that the correlations of all samples up to a certain maximum lag time are considered from the entire data set or from sufficiently long blocks. Further suppression of the wrap-around error e.g. via the application of a window function is not needed. This way, one also avoids the spectrum becoming smeared by the modulation of the signal by the window function. This method is identical to Blackman-Tukey [9, 10] used with a rectangular window applied to the correlation function estimated. Furthermore, this method has been previously investigated by Bartlett and mentioned in [6], yielding results comparable to block averaging. Since longer data records need
to be Fourier transformed, truncation of primarily long correlation estimates is computationally more expensive. However, it is superior in efficiently using the information available. Truncating the support of the autocorrelation function is applied to all methods to reduce the estimator variance of the spectrum, including the interpolation method, where it has originally been used for LDV data processing [38]. The slotting technique inherently supports this processing step by defining a number $K$ of lag times with the interval $\Delta \tau$ to be derived, independent from and typically significantly smaller than the length of the data set. To finally achieve interchangeability, the support $K\Delta \tau$ of the estimated correlation function must be chosen identical for all methods.

3.7 No fuzzy slotting

For the slotting technique, a distribution of products over the two neighboring slots depending on the sub-bin lag time has been developed called fuzzy slotting [44, 52]. Initially it was meant to gain sub-bin resolution for the autocorrelation estimate and it has been part of the best-practice recommendations so far. However, with its recent adaptation to direct spectral estimation it has been determined to effectively low-pass filter the data [41]. Those investigations give indication that also the original fuzzy slotting technique has such characteristic. Since this is not useful for a non-parametric estimation, the application can no longer be recommended. Nonetheless, it is still available as a processing option with the programs at [1].

3.8 No window functions

The truncation of the support of the correlation function is applied to all methods introduced above. Further suppression of the wrap-around error e.g. via the application of a window function is not needed. This way, one also avoids the spectrum becoming smeared by the modulation of the signal by the window function.

3.9 No variable windowing

For a further reduction of the estimator variance of the spectrum, a frequency-dependent variable windowing of the autocorrelation function has been used successfully when transformed into the spectrum [25, 50]. The corresponding spectrum looks smooth and the estimator variance is reduced, especially at higher frequencies. This technique is fully
independent from random sampling and can be applied to any transformation of an autocorrelation function into a spectrum. However, the choice of window function and the frequency dependent scaling parameter substantially influence the result. Furthermore, the leakage effect arises due to modulation by the window function, leading to systematic errors in the spectrum. Finally, the obtained spectrum violates the correspondence to the autocorrelation function via the Wiener-Khinchin theorem. Since a smeared spectrum and corresponding systematic errors are not useful for the goal of non-parametric and bias-free estimates, the variable windowing technique is not used in this comparison, where bias-free estimation has priority. Instead, the transformation of the correlation functions (with limited support) via the discrete Fourier transform into a power spectral density is used for all methods investigated.

3.10 Regularization

The present paper consequently follows the goal of non-parametric and bias-free estimates of the correlation function and the spectrum. However, the advanced bias corrections may potentially lead to correlation matrices, which violate the non-negative definiteness. As a consequence, negative values may occur in the corresponding power spectral density. Since the introduced procedures yield bias-free and consistent estimates of both, the correlation function as well as the spectrum (except for averaging over the fundamental time intervals), averages over multiple estimates of the functions or estimates from longer data records will converge towards the correct functions of the underlying process. The ultimate solution to achieve non-negative definiteness is regularization. Since this inevitably introduces a bias to both the correlation function and the corresponding spectrum, regularization is not investigated here, where bias-free estimation has priority. However, regularization can be added on demand as an intermediate processing step prior to the final transformation into the spectrum or as a post-processing step past the procedures used here.

4 Application

LDV data from a turbulent flow behind a rectangular grid in a closed loop wind tunnel is presented as an example. Hot-wire data from a constant-temperature anemometer (CTA), taken at the same position of the flow as the LDV data, is available as a reference. Descriptions of the experiment can be found in [29] and the data is available at [2]. The available LDV data consist of ten data sets with 100 000 samples each, including arrival times, velocities, and transit times.
The characteristic parameters of the experimental data and the fundamental signal processing parameters are given in Table 1. The odd numbers of the fundamental time step $\Delta \tau$ and the number of time steps $K$ comes from the temporal resolution of 187.5 kHz of the LDV arrival time acquisition. To avoid varying probability of pairs of samples at the respective lag times, the fundamental time interval is an integer multiple of the primary arrival-time step.

The second experiment made use of the same flow and measurement technique, however, with a cylinder in the flow as an obstacle providing a van-Kármán vortex street in the wake. Because of the long lasting correlation, the support of the correlation function used for the transform into the spectrum needs to be much longer than in the first experiment. The support has been used ten times as long, still the correlation is not entirely decayed at the maximum lag time. With longer support of the correlation function investigated, the spectral resolution further increases, finally increasing also the random error. Besides this, also the graphical presentation becomes challenging. In such cases of long lasting correlations, a compromise must be found between systematic and random errors, possibly including also variations of the measurement time and application of window functions.

With 2 000 samples and a mean duration of about 2 400 integral timescales (data set I) and 2 500 integral timescales (data set II) each, the data blocks were long enough to make corrections for short data sets not needed. However, correction for short data sets has been enabled by default. There is no risk of an overcorrection or making the bias worse, if applied also to longer data sets. Each data block has been processed with the aforementioned methods utilizing the programs available at [1] with transit-time weighting. To compare barely the processing methods, none of the optional processing steps have been used, namely no local normalization, no fuzzy techniques and also no windowing.

The averages over the considered data blocks yield empirical estimates of the mean autocorrelation function and the mean spectrum and their respective estimator variances, shown in Fig. 1. Roughly, the results show agreement between the investigated processing methods. No significant deviation from the references can be quickly identified. A more detailed inspection finds certain systematic errors for the interpolation method. A major part of this error originates from the fact that the interpolation method cannot make use of the transit times for data weighting. With this data rate, the implicit inter-arrival time weighting is not sufficient, resulting in a remaining statistical bias. At the same time, the estimator variance of this method is different in comparison to the other methods.

A deeper inspection discovers that the spectral density at high frequencies from the LDV data is a little below those of the CTA system. Possibly, the particles are not completely passive at these
frequencies and they cannot follow the flow exactly with no slip. This speculation complies with the observation in Fig. 2, where the first harmonic peak of the base frequency of the vortex street is suppressed by the LDV data while the peak at ground frequency is reproduced exactly in amplitude.

Except for this damping, which probably originates in the inertia of the tracer particles, the three methods, direct spectral estimation, the slotting technique and arrival-time quantization, show no obvious systematic errors and their estimator variance is comparable. This evidences that all three methods have comparable efficiency in terms of bias correction and efficient usage of the information available. The equivalence of the three methods, and to a lesser extent the interpolation method, in terms of systematic errors and the achieved estimator variance shows that the methods are interchangeable. The backlog demand on processing options of the interpolation method is illustrated in Tab. 3. Due to interchangeability, the preference of one or the other method can be done considering other criteria such as the programming effort, the memory usage or the processing time. In the latter case, quantization of arrival times is the optimum choice. Due to the quasi-uniform sampling, it can gain from fast computer routines like the fast Fourier transform.

5 Conclusion

An alternative estimator for the autocorrelation function and the power spectral density has been introduced based on quantization of arrival times. The method combines advantageous processing options like individual sample weighting, preferably transit-time weighting, removal of self-products, normalization, optionally local normalization, truncation of the support of the correlation function and optionally correction for short data sets. It has been shown that this estimator yields results, which are comparable to direct spectral estimation and to the slotting technique in terms of systematic and random errors. Of course, this requires the estimator to be applied under equivalent conditions in terms of optional processing steps and processing parameters to obtain a fair comparison. The equivalence of the results in terms of their statistical properties indicates the methods are equally efficient at using the information available. Therefore, the selection of one specific procedure to process a certain data set will not affect the statistical characteristics of the results. The preference for one particular method can be based on more practical criteria like programming effort, memory usage or processing time. In this respect, quantization of arrival times is a good candidate. While the slotting technique needs to sum all pairs of samples within a certain time interval and direct spectral estimation needs to
perform a continuous-time Fourier transform, arrival-time quantization leads to a quasi-uniform sampling of the data set. This way one may take advantage of fast computer routines like the fast Fourier transform, which accelerates the computation of the autocorrelation function and the spectrum significantly without losing accuracy or efficiency.

Acknowledgements

The work of Palle Gjelstrup, Finn E. Jørgensen (Dantec Dynamics), and Knud Erik Meyer (Danmarks Tekniske Universitet) performing the experiment and providing the data is gratefully acknowledged.

References


Fig. 1 a) Empirical mean autocorrelation function, b) empirical estimator variance of the autocorrelation function, c) empirical mean one-sided power spectral density and d) empirical estimator variance of the spectrum from experimental LDV data behind a grid. The results from the hot-wire anemometer (constant-temperature anemometer, CTA) data are given for comparison.
Fig. 2  a) Empirical mean autocorrelation function, b) empirical estimator variance of the autocorrelation function, c) empirical mean one-sided power spectral density and d) empirical estimator variance of the spectrum from experimental LDV data behind an obstacle. The results from the hot-wire anemometer (constant-temperature anemometer, CTA) data are given for comparison.
Table 1 Parameters of the experimental data behind a grid and fundamental signal processing parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of data sets</td>
<td>10</td>
</tr>
<tr>
<td>Number of samples per data set</td>
<td>100 000</td>
</tr>
<tr>
<td>Mean duration of each data set</td>
<td>22.4 s</td>
</tr>
<tr>
<td>Temporal resolution (arrival times)</td>
<td>187.5 kHz</td>
</tr>
<tr>
<td>Mean velocity</td>
<td>31 m/s</td>
</tr>
<tr>
<td>Velocity variance</td>
<td>1.5 m²/s²</td>
</tr>
<tr>
<td>Turbulence intensity</td>
<td>4.4 %</td>
</tr>
<tr>
<td>Mean data rate</td>
<td>4450 /s</td>
</tr>
<tr>
<td>Integral timescale</td>
<td>0.2 ms</td>
</tr>
<tr>
<td>Δτ</td>
<td>0.044 ms</td>
</tr>
<tr>
<td>K</td>
<td>182</td>
</tr>
</tbody>
</table>

Table 2 Parameters of the experimental data behind an obstacle and fundamental signal processing parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of data sets</td>
<td>10</td>
</tr>
<tr>
<td>Number of samples per data set</td>
<td>100 000</td>
</tr>
<tr>
<td>Mean duration of each data set</td>
<td>47.6 s</td>
</tr>
<tr>
<td>Temporal resolution (arrival times)</td>
<td>187.5 kHz</td>
</tr>
<tr>
<td>Mean velocity</td>
<td>26 m/s</td>
</tr>
<tr>
<td>Velocity variance</td>
<td>55 m²/s²</td>
</tr>
<tr>
<td>Turbulence intensity</td>
<td>29 %</td>
</tr>
<tr>
<td>Mean data rate</td>
<td>2100 /s</td>
</tr>
<tr>
<td>Integral timescale</td>
<td>0.38 ms</td>
</tr>
<tr>
<td>Δτ</td>
<td>0.044 ms</td>
</tr>
<tr>
<td>K</td>
<td>1818</td>
</tr>
</tbody>
</table>
Table 3 Summary of available and required processing options

<table>
<thead>
<tr>
<th>Processing option</th>
<th>Direct spectral estimation</th>
<th>Slotting technique</th>
<th>Interpolation method</th>
<th>Arrival-time quantization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transit-time weighting (essential)</td>
<td>✓</td>
<td>✓</td>
<td>not available</td>
<td>✓</td>
</tr>
<tr>
<td>Forward-backward inter-arrival time weighting (fallback, sufficient only at high data density)</td>
<td>✓</td>
<td>✓</td>
<td>inherent</td>
<td>✓</td>
</tr>
<tr>
<td>Constant-value approximation for bias correction for short data records</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Removal of self-products (essential)</td>
<td>✓</td>
<td>✓</td>
<td>not available</td>
<td>✓</td>
</tr>
<tr>
<td>Normalization (essential)</td>
<td>✓</td>
<td>✓</td>
<td>not available</td>
<td>✓</td>
</tr>
<tr>
<td>Local normalization</td>
<td>✓</td>
<td>✓</td>
<td>not available</td>
<td>✓</td>
</tr>
<tr>
<td>Truncation of the support of the correlation function</td>
<td>✓</td>
<td>inherent</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Fuzzy slotting/fuzzy time quantization (not recommended anymore)</td>
<td>✓</td>
<td>✓</td>
<td>not available</td>
<td>✓</td>
</tr>
</tbody>
</table>