Digital Signal Processing 76 (2018) 22–33

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

A fair review of non-parametric bias-free autocorrelation and spectral methods for randomly sampled data in laser Doppler velocimetry

Nils Damaschke [a], Volker Kühn [b], Holger Nobach [c,*]

[a] University of Rostock, Faculty of Computer Science and Electrical Engineering, Institute of General Electrical Engineering, Albert-Einstein-Straße 2, 18059 Rostock, Germany
[b] University of Rostock, Faculty of Computer Science and Electrical Engineering, Institute of Communications Engineering, Richard-Wagner-Straße 31, 18119 Rostock-Warnemünde, Germany
[c] Max Planck Institute for Dynamics and Self-Organization, Am Faßberg 17, 37077 Göttingen, Germany

ARTICLE INFO

Article history:
Available online 1 February 2018

Keywords:
Correlation
Spectrum
Signal processing
Random sampling
Laser Doppler

ABSTRACT

This paper presents a comparison of currently available methods for non-parametric and bias-free estimation of the autocorrelation function and power spectral density from randomly sampled data. The primary motivation is the processing of velocity data obtained using laser Doppler techniques in turbulent flows. However, the methods are applicable to various other cases of random sampling, including those with small deviations from the ideal Poisson process. Whilst these methods have been compared in the literature before, a fair comparison of their relative performance requires that they be tested under identical conditions. This includes identical use of special processing options and identical processing parameters. This has not been achieved in the literature to date. An example application on publicly available laser Doppler data shows agreement between the results obtained with the different methods. Under this fair comparison, the 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

Discrete sampling of continuous signals usually reduces the information available. However, there is no alternative if one wants a numerical representation of the original signal suitable for statistical analysis through computer programs. Uniform sampling is usually the preferred means of achieving this goal. It has the advantage of enabling the application of well-established methods for statistical analysis with the temporal resolution defined by the chosen sampling frequency. This kind of sampling has been comprehensively investigated and is reflected in the Shannon–Nyquist theorem [1]. For non-uniformly sampled signals, high-resolution information is preserved since some samples are taken in pairs with short separations. These pairs of samples are taken sporadically. The average sampling rate can be smaller. The overall information of the signal is reduced, similar to uniform sampling, but there is no specific cut-off frequency or resolution limit. On the other hand, for a given average sampling rate, short sampling distances must be balanced by longer sampling distances. For the same average data rate, nonuniform sampling yields a less effective representation of the information in comparison to uniform sampling, requiring longer data sets to obtain a given statistical confidence. Furthermore, the processing routines must account for the irregular sampling times, making them computationally expensive.

Laser Doppler velocimetry (LDV) [2–4] is a good example, of a measurement process with nonuniform sampling. 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. 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

* Corresponding author.
E-mail addresses: nils.damaschke@uni-rostock.de (N. Damaschke), volker.kuehn@uni-rostock.de (V. Kühn), holger.nobach@nambis.de (H. Nobach).

https://doi.org/10.1016/j.dsp.2018.01.018

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 [5,6]. One way to tackle this issue is to apply weighting factors to the individual samples. The weighting factors should take into account the varying probability density of measurement events. Common weighting methods are discussed in Section 2.1. During the passage through the measurement volume each 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, making appropriate treatment of systematic errors necessary, which occur in statistical functions derived from these data.

The statistical analysis of LDV data requires procedures, which are suitable under these conditions. Various successful processing methods have been developed in the past, including demonstration of specific test cases where one particular method is superior to others. However, comparisons between the processing methods are confounded by different boundary conditions, processing parameters and optional processing steps. So far, differences between the processing principles themselves have not been discovered uniquely. To do so, variations in respective processing steps need homogenization. Important are for instance the application of a common weighting scheme, the treatment of self-products, normalization attempts and a common method to transform the correlation estimate into a corresponding spectrum including an effective reduction of the spectral resolution with most efficient use of information available. The result is that some of the established autocorrelation and spectral estimators yield equivalent estimates, where the values of the functions differ between the methods for a particular data set, but the estimates have similar statistical properties in terms of systematic errors and estimator variance. This indicates that these methods are comparably efficient at using the information of the signal. Further, this gives evidence that the various processing methods are interchangeable and that the preference of one particular method will not affect the statistical characteristics of the results.

The available methods to obtain estimates of the autocorrelation function and the power spectral density from irregularly sampled data are reviewed in the following sections. However, to achieve the inter-comparability, processing options are shared between the various processing methods. Where required, modifications from the original literature will be introduced in detail. For all procedures used here Python source code is online available at [7]. Note that the given estimators of the autocorrelation function and the power spectral density are useful for nonuniform observations of stationary processes only.

The present article consequently follows the goal of non-parametric and bias-free estimates. Complete bias correction 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 $\Delta\tau$), 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 is regularization. A broad overview of present approaches is given in [8]. Since this inevitably introduces a bias to both the correlation function and the corresponding spectrum, regularization is unusual in LDV data processing. It is not investigated here, where bias-free estimation has priority. However, it 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 introduced here.

2. Processing methods

The velocity $u(t)$ as a function of time is assumed to be 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. Appropriate weighting schemes are discussed in the following section.

2.1. Mean value

The mean value of the velocity $u(t)$ is defined as

$$\mu = \langle u(t) \rangle \tag{1}$$

with the expectation $\langle \cdot \rangle$. For a given data set, the mean value instead is estimated as the ensemble mean of the samples u_i, assuming both ergodicity and sufficiency. Considering appropriate weights w_i, the mean value of the measured data set can be obtained as

$$\bar{u} = \frac{\sum_{i=0}^{N-1} w_i u_i}{\sum_{i=0}^{N-1} w_i}. \tag{2}$$

If the weighting scheme is appropriate, this estimator can be bias-free. Only a random error remains due to finite sampling.

Various weighting schemes have been introduced and tested with laser Doppler data. At first glance, velocity weighting ($w_i = 1/(|u_i|\,\gamma_i)$) [5,9] seems to be a suitable weighting scheme. Originally, $|u_i|$ is understood as the magnitude of the three-dimensional velocity vector and γ_i is the projection area of the ellipsoidal measurement volume into the direction of the velocity vector. This weighting scheme specifically tackles the source of the statistical bias, namely the higher mean data rate at higher convection velocities. The effort required to measure all three components of the velocity is enormous. Therefore, for practical measurement systems, where only one or two velocity components are available, the missing components have been modelled. In other cases, the missing components have not been considered at all. However, these reductions are not suited to provide reliable bias correction in variable turbulent flow cases [6]. Additionally, this weighting method has been found to lead to significant systematic errors in the estimated statistical values if the data are superimposed with noise. Last, it fails if the spatial concentration of tracer particles varies with the velocity. This can happen if the seeding is generated at a fixed Eulerian point with a constant particle generation rate.

Inter-arrival time weighting ($w_i = t_i - t_{i-1}$) [10] is robust even against altering correlations between the instantaneous data rate and the velocity [11]. However, the efficiency of this weighting method depends on the reduced data rate, which is the number of samples per integral timescale [12] (also called the data density). Equivalent results occur for sample-and-hold processing [13].