

RESEARCH

Open Access



Bias correction for direct spectral estimation from irregularly sampled data including sampling schemes with correlation

Nils Damaschke¹, Volker Kühn² and Holger Nobach^{3*} 

Abstract

The prediction and correction of systematic errors in direct spectral estimation from irregularly sampled data taken from a stochastic process is investigated. Different sampling schemes are investigated, which lead to such an irregular sampling of the observed process. Both kinds of sampling schemes are considered, stochastic sampling with non-equidistant sampling intervals from a continuous distribution and, on the other hand, nominally equidistant sampling with missing individual samples yielding a discrete distribution of sampling intervals. For both distributions of sampling intervals, continuous and discrete, different sampling rules are investigated. On the one hand, purely random and independent sampling times are considered. This is given only in those cases, where the occurrence of one sample at a certain time has no influence on other samples in the sequence. This excludes any preferred delay intervals or external selection processes, which introduce correlations between the sampling instances. On the other hand, sampling schemes with interdependency and thus correlation between the individual sampling instances are investigated. This is given whenever the occurrence of one sample in any way influences further sampling instances, e.g., any recovery times after one instance, any preferences of sampling intervals including, e.g., sampling jitter or any external source with correlation influencing the validity of samples. A bias-free estimation of the spectral content of the observed random process from such irregularly sampled data is the goal of this investigation.

Keywords: Bias-free estimation, Spectrum, Random sampling, Missing samples, Data gaps

1 Introduction

Digital signal processing normally implies a time-limited, non-interrupted sequence of equidistant samples taken from a signal-generating process under investigation. Various reasons may lead to a different sampling: (1) the measurement process may depend on a non-regular, typically stochastic, sampling process, e.g., laser Doppler systems [1–3] in burst mode measure arrival times and velocities of randomly arriving particles carried by a time-variant flow field. (2) The equidistant measurement may be disturbed, leading to either individual missing samples

or longer data gaps. Irregular sampling directly influences the spectral content of the observation. Spectral estimators may correctly estimate, e.g., the power spectral density of the data sequence (the observed signal) from the measurement. However, the spectrum of the observed signal will deviate from the spectrum of the process under investigation. Particularly, the observed signal is the product of the process under investigation multiplied with the sampling function, which is understood as the train of sampling instances of the observed signal, where all values of the sampling function are of unit amplitude (see illustration in Fig. 1). Hence, the spectrum of the observed signal is the convolution of the spectrum of the process with the spectrum of the sampling function. The spectral properties of the sampling function thus have a direct influence on the spectrum of the observed signal. To

*Correspondence: holger.nobach@nambis.de

³Max Planck Institute for Dynamics and Self-Organization, Am Faßberg 17, 37077 Göttingen, Germany

Full list of author information is available at the end of the article

obtain meaningful information about the underlying process, appropriate estimators must consider the irregular sampling process and its spectral composition.

Spectral estimation from randomly sampled signals in continuous time has been investigated in the past, mainly in the context of controlled sampling with induced variation of the sampling times, known as digital alias-free signal processing or sampling jitter [4–7], in the context of astrophysical observations [8–11] or in the context of laser Doppler data processing [12–24] including the specific role of processor dead times [25–27]. More general in their application are investigations in [28–32]. All these estimators are potentially able to handle also equidistant data with missing samples, and some of the given references include this case. However, adaptations are necessary, since random sampling on a continuous domain has a different spectral composition than equidistant sampling with missing samples. Independent and randomly distributed missing samples with otherwise equidistant sampling has also been investigated [31, 33–38]. Correlated data gaps have been investigated only for very specific cases [31, 39–42], without options for generalization or without satisfactory bias correction.

In contrast to theoretical investigations or computer simulations, where purely random sampling can be assumed or realized, strictly independent sampling is not possible in technical systems. For stochastic sampling in continuous time, e.g., laser Doppler systems cannot deliver successive samples with lag times below a certain minimum value. Particles with too low distance will lead to overlapping scatter signals from the measurement volume. The laser Doppler system will remove these overlapping signals to avoid faulty measurements due to possible interference between the signals with an unpredictable phase shift. This way, the size of the measurement volume defines a minimum distance between successive particles and finally a certain minimum lag time, below which the probability of appropriate pairs of measurement events drops rapidly. The result is a distinct measurement error of spectral estimates based on algorithms, which rely on the assumption of purely random and independent sampling instances as shown in [25–27].

Missing data from equidistantly sampled processes may occur as individual outliers, which often are independent from each other. However, if external processes disturb the process under observation, these impinging processes may have a certain correlation and generate missing data samples or gaps, which are not independent. Depending on the individual reasons for missing data samples or data gaps, the correlation and the spectral content of the sampling scheme will differ significantly between application cases. Also here, solutions of bias correction for purely random and independent occurrence of missing samples like in [43] will fail with correlated data gaps.

Universal solutions of bias correction for correlated data gaps with unspecific characteristics are not available so far, neither for continuous time nor for equidistant time. The present article investigates systematic errors of spectral estimators processing directly the sequence of data for different sampling cases with different distributions and with different spectral characteristics of sampling intervals. In continuous time, random sampling is investigated, where in one case purely random and independent sampling instances are taken from a linear stochastic process. For the same process, a minimum time interval between successive samples leads to correlated sampling intervals in another simulation. For equidistant time, also a linear stochastic process is observed. Again, a random occurrence of missing data samples is investigated with independent individual outliers first. In the last sampling scheme, also longer sequences of data points are removed from the original data set, introducing correlation into the pattern of missing samples.

A common procedure to correct systematic errors of direct spectral estimation is introduced and proven to be able to deliver bias-free estimates of the spectra, independent of the spectral characteristics of the sampling scheme or that of the missing data. The required parameters can be obtained by theoretical investigation for sampling processes with a priori known sampling mechanisms. For unknown relations, an alternative procedure is given to obtain the correction parameters numerically, straight from the measured data sequence. The procedures are available as Python source codes as supplementary material to this article together with all data sets under [44].

Note, that the correction procedure is not able to correct aliasing errors. If any aliasing occurs due to significant spectral content of the observed process above the temporal resolution of the sampling scheme, then systematic errors of the estimation will remain. Aliasing has its origin in an insufficient extraction of information from the process under observation, which principally cannot be repaired a posteriori, since the required information is not available from the observed data set any more. Note further, that the bias correction may lead to corresponding correlation matrices, which potentially may violate the non-negative definiteness. As a consequence, negative values occur in the corresponding estimated power spectral densities. Since the introduced procedures yield bias-free estimates, averages over multiple estimates of spectra will converge towards the true spectrum of the underlying process. As a common means to reduce the estimation variance, averaging over spectral estimates from data segments (block averaging) is often applied anyway. The ultimate solution of course, would be regularization. Since this inevitably introduces a bias to the spectrum, regularization is not considered in the present article, where bias-free estimation has priority. If block averaging

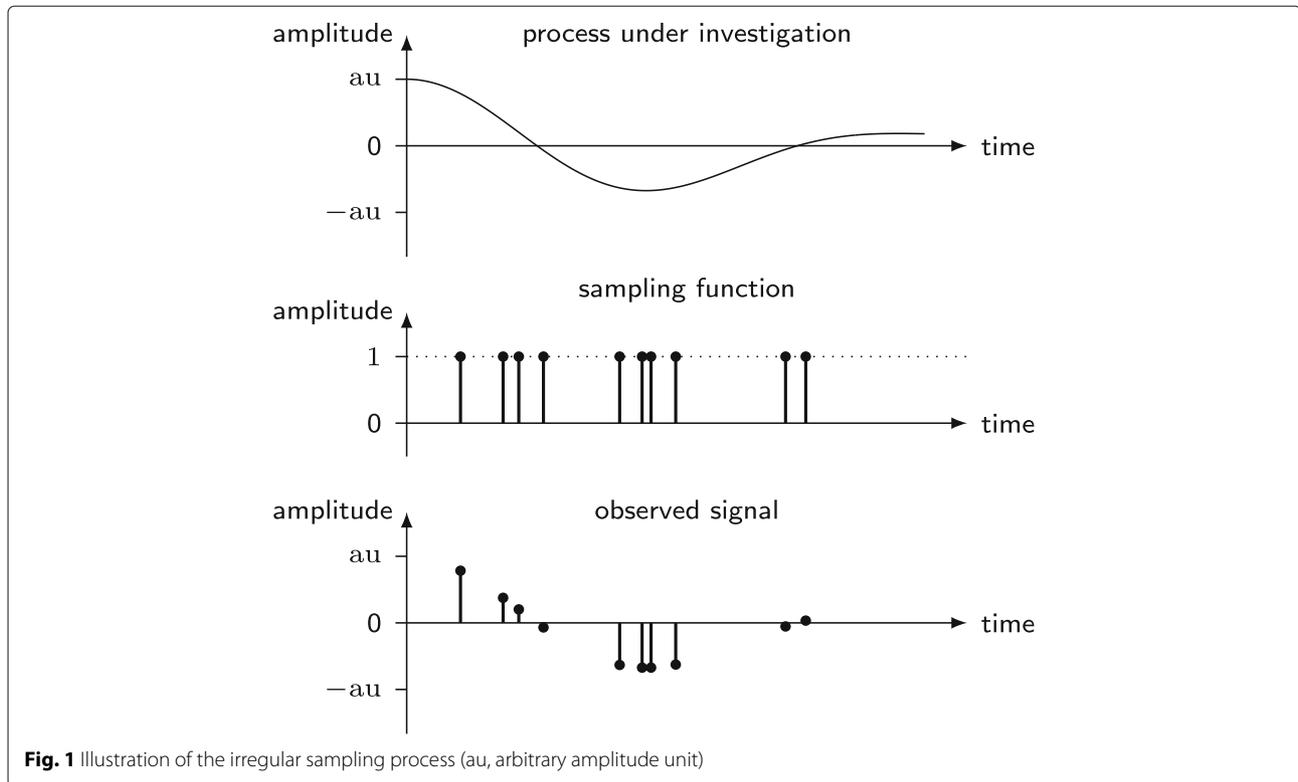


Fig. 1 Illustration of the irregular sampling process (au, arbitrary amplitude unit)

is applied, bias-free estimates are essential to obtain a consistent mean spectrum with vanishing systematic and random errors for an increasing sample size.

2 Methods

2.1 Primary estimates

Let $S_p(f)$ be the power spectral density of an observed, time dependent process $x(t)$. The process is sampled at N ascending time instances t_i , with $i = 1 \dots N$. The entire duration of the signal is T , where $0 \leq t_1 < t_N \leq T$. The sampling is either at random time instances t_i with real values from a continuous domain or at quantized time instances with the fundamental time step Δt , but with missing samples. Note that T is not determined finally by a data set with irregular sampling instances. Only $T \geq t_N$ is fix. The exact duration of the signal can be defined later, according to post-processing requirements like the spectral resolution desired for the discrete numerical spectral estimation.

Let $S_s(f)$ be a bias-free estimate of the power spectral density of the observed signal. For the random sampling in continuous time, the observed signal is understood as the series of measurements $x_i = x(t_i)$. For nominally regular sampling with missing samples, the observed signal is understood as the series of valid values only. Neglecting any possible errors due to a periodic continuation of the signal, the particular estimate of $S_s(f)$ can be obtained, e.g., as the Fourier spectrum

$$S_s(f) = \frac{T}{N^2} \left| \sum_{i=1}^N x_i e^{-2\pi i f t_i} \right|^2 \tag{1}$$

with the imaginary unit i or with other methods like Lomb-Scargle's method [8, 9] or generalized Lomb-Scargle's method [45] for data having a non-zero mean value. Note, that $S_s(f)$ deviates from $S_p(f)$ because irregular sampling influences the spectral content of the observation as outlined below.

Dissolving the square in Eq. (1) leads to

$$S_s(f) = \frac{T}{N^2} \left\{ \left[\sum_{i=1}^N x_i \cos(2\pi f t_i) \right]^2 + \left[- \sum_{i=1}^N x_i \sin(2\pi f t_i) \right]^2 \right\} \tag{2}$$

$$= \frac{T}{N^2} \left\{ \sum_{i=1}^N \sum_{j=1}^N x_i x_j [\cos(2\pi f t_i) \cos(2\pi f t_j) + \sin(2\pi f t_i) \sin(2\pi f t_j)] \right\} \tag{3}$$

$$= \frac{T}{N^2} \left\{ \sum_{i=1}^N \sum_{j=1}^N x_i x_j \cos [2\pi f (t_i - t_j)] \right\}. \tag{4}$$

The last line with the double sum can be rewritten as

$$= \frac{T}{N^2} \left\{ \sum_{i=1}^N x_i^2 + \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N x_i x_j \cos [2\pi f(t_i - t_j)] \right\} \quad (5)$$

with separate summations over self-products x_i^2 on one side and cross-products $x_i x_j$ with $i \neq j$ on the other side.

For uninterrupted equidistant sampling, the probability $P(t_i)$ of getting a valid sample at time t_i is unity at any sampling instance $t_i = i\Delta t$ with the sampling interval Δt and zero between the samples. Since the self-products x_i^2 in Eq. (5) can be built directly from the respective samples x_i , the respective probability of such self-products occurring at time t_i is identical to the probability for the occurrence of the sample itself, namely $P(t_i)$. Therefore, the probability of self-products also is unity at any sampling instance $t_i = i\Delta t$ and zero between the samples.

Cross-products $x_i x_j$ with $i \neq j$ instead require two samples at two different times t_i and t_j . The respective joint probability of getting such cross-products $P(t_i, t_j)$ then is the product of the probability $P(t_i)$ of having the first sample at t_i and the conditional probability $P(t_j|t_i)$ of having another sample at t_j under the condition of having had the first sample at t_i .

$$P(t_i, t_j) = P(t_i)P(t_j|t_i) \quad (6)$$

However, for uninterrupted equidistant sampling, the probability of having another sample with a delay of $t_j - t_i$ that equals an integer multiple of the sampling interval Δt is also unity. Hence, with $t_i = i\Delta t$ and $t_j = j\Delta t$, finally cross-products between different samples and self-products of single samples have the same probability, provided the sampling is equidistant and uninterrupted.

With any kind of irregular sampling, the probability $P(t_i)$ of having a valid sample at a sampling time $t_i = i\Delta t$ is either less than one for nominally equidistant sampling with missing samples or it changes its physical dimension to a probability density for irregular sampling in a continuous time domain. However, the probability of self-products at a certain time t_i is still identical to that probability of the sample itself $P(t_i)$ and also Eq. (6) still holds for cross-products. But with irregular sampling unfortunately, the respective conditional probabilities or conditional probability densities $P(t_j|t_i)$ of getting another sample at time t_j after the sample at time t_i also differ from unity and may additionally vary for different delays $t_j - t_i$. Under these conditions, Eq. (5) averages over self- and cross-products of varying delays, which all have different probabilities or probability densities and contribute differently to the estimate of the spectrum if applied to

irregularly sampled data. This finally leads to the observed bias.

Let further be $R_s(\tau)$ the correlation function of the observed signal. This one corresponds to $S_s(f)$ via the Fourier transform incorporating Wiener-Khinchin's theorem [46]. For a numerical realization via the discrete Fourier transform (DFT) resp. the inverse discrete Fourier transform (IDFT), both R_s and S_s must be sampled regularly. For that, a fundamental time increment $\Delta\tau$ and a frequency increment Δf are defined. While $\Delta\tau$ can be chosen arbitrarily for a randomly sampled signal in continuous time, $\Delta\tau$ equals the primary time step Δt of the signal for equidistant sampling. For the purpose of better clarity, in the following, the time step will be denoted as $\Delta\tau$ commonly, dropping the discrimination between time series with steps of Δt and correlation functions with steps of $\Delta\tau$. The frequency increment in all cases is $\Delta f = \frac{1}{T}$, defined by the assumed duration of the signal T , where T can be chosen either slightly larger than t_N or significantly larger corresponding to an optional zero padding of the signal. In the present study, $T = M\Delta\tau$ is chosen with an integer M , where T is close to t_N , e.g., the difference is within one time step $\Delta\tau$. That leads to the correspondences

$$S_s(f_j) = \Delta\tau \cdot \text{DFT} \{R_s(\tau_k)\} = \Delta\tau \cdot \sum_{k=-\lfloor \frac{M}{2} \rfloor}^{\lfloor \frac{M-1}{2} \rfloor} R_s(\tau_k) e^{-2\pi i f_j \tau_k} \quad (7)$$

$$R_s(\tau_k) = \frac{1}{\Delta\tau} \cdot \text{IDFT} \{S_s(f_j)\} = \frac{1}{T} \cdot \sum_{j=-\lfloor \frac{M}{2} \rfloor}^{\lfloor \frac{M-1}{2} \rfloor} S_s(f_j) e^{2\pi i f_j \tau_k} \quad (8)$$

with $f_j = j\Delta f$, $j = -\lfloor \frac{M}{2} \rfloor \dots \lfloor \frac{M-1}{2} \rfloor$ and $\tau_k = k\Delta\tau$, $k = -\lfloor \frac{M}{2} \rfloor \dots \lfloor \frac{M-1}{2} \rfloor$, where $\lfloor x \rfloor$ is the largest integer smaller than or equal x .

2.2 Bias correction

The sampling schemes investigated here yield either a continuous distribution of sampling intervals (exponential distribution for purely random sampling times or with deviations from the exponential distribution for correlated random sampling) or a discrete distribution (unity at the sampling interval of $\Delta\tau$ and zero otherwise for uninterrupted equidistant sampling or with further integer multiples of $\Delta\tau$ occurring with samples missing). All investigated sampling schemes have in common, that the sampling function, which defines the sampling times t_i , has a certain randomness, namely the sampling instances

occur at random times. For continuously distributed sampling times, the sampling instances themselves are randomly distributed. Regarding equidistant data sets with missing or invalid samples, the randomness applies to the availability or validity of the individual samples. In the following, continuous and discrete distributions of the respective sampling intervals are distinguished to ensure a unique discrimination between these two cases of either sampling in a continuous time domain or nominally equidistant sampling with missing instances. This is apart from possible correlations between the sampling instances resp. between the sampling intervals, which additionally cause deviations from a purely random sampling in each of these two cases.

Let the mean sampling rate be α and let the mean number of samples per time unit $\Delta\tau$ be $\alpha' = \alpha\Delta\tau$. For a discrete distribution of sampling intervals, α' is also the probability of a sample x_i being valid, which complies with $0 \leq \alpha' \leq 1$. Self-products x_i^2 of samples then also occur with the probability of α' as explained above. Cross-products $x_i x_j$ can be made only from two different samples occurring at different time units $t_i = i\Delta\tau$ and $t_j = j\Delta\tau$ with $i \neq j$. Assuming independence between sampling times, the probability of having sampling time t_j after sampling time t_i becomes $P(t_j|t_i) = P(t_j)$, namely α' . The probability of getting cross-products $P(t_j, t_i)$ in Eq. (6) then finally becomes α'^2 . For a continuous distribution of independent sampling intervals with a mean sampling rate of α , the number of samples per time unit $\Delta\tau$ is Poisson distributed with the mean value of $\alpha' = \alpha\Delta\tau$, where α' can also be larger than one. In this case, the mean number of self-products x_i^2 per time unit again is α' , and the mean number of cross-products $x_i x_j$ within two *different* time units again is α'^2 following Eq. (6), provided all sampling instances within the two time units are independent.

Both of these values, the probability of self-products α' and the probability of cross-products from two different time units α'^2 , are identical for the two cases, that of discrete distribution of sampling intervals and that of continuous distribution of sampling intervals for purely random occurrence of samples without correlations between the sampling instances. In contrast to the case of discrete distribution of sampling intervals, for a continuous distribution of sampling intervals, the occurrence of more than one sample within one time unit is possible. Therefore, there also exists a probability of cross-products occurring from different samples within the *same* time unit, which receives further attention at a later point.

Let β_k be the mean number of self- and cross-products per time unit counted in $R_s(\tau_k)$. If only cross-products from different time units were counted and no correlation between the sampling intervals is assumed, from each pair of time units, one would obtain α'^2 pairs of samples on average. Within the measurement time $T = M\Delta\tau$ one

then has $M - |k|$ such pairs of time units of lag time $\tau_k = k\Delta\tau$. For long enough data sets, small enough absolute lag times $|k| \ll M$ or periodic continuation of the signal, one can assume a mean number of M such pairs of time units, each yielding α'^2 pairs of samples on average. Finally, the expected number $M\alpha'^2$ of such cross-products counted in $R_s(\tau_k)$ divided by the length of the data set, which is M time units long, yields the expected mean number of cross-products per time unit $\beta_k = \alpha'^2$.

Let further β'_k be β_k normalized by α'^2 . Then for independent sampling and if only cross-products from different time units are counted β'_k becomes unity. Any other influence, self-products, correlation between sampling instances, or cross-products within one time unit causes β'_k deviating from unity. In any case, β'_k is the ratio of the mean number of pairs of samples expected in $R_s(\tau_k)$ including all effects causing β'_k deviating from unity and the expected number of cross-products only from different time units and from independent sampling instances. A prediction of all coefficients β'_k for a given sampling scheme then can be used to balance the different probabilities of self- and cross-products for each lag time τ_k of the primary estimate of the correlation function $R_s(\tau_k)$ by normalization with β'_k yielding an improved estimate $\hat{R}_s(\tau_k)$ applying

$$\hat{R}_s(\tau_k) = \frac{1}{\beta'_k} R_s(\tau_k). \quad (9)$$

The general procedure for bias correction then requires to transform the primary estimate of the spectrum into the appropriate correlation function. The correlation function then can be corrected for its bias based on appropriate correction factors β'_k , followed by another Fourier transform of the improved correlation function back into a spectrum.

The determination of the values β'_k depends on the particular sampling scheme. An analytical derivation requires a priori knowledge about the specific rules of the particular sampling process. Purely random sampling without any correlations between the sampling instances is a dedicated case, where only β'_0 deviates from unity. This allows to perform the intended bias correction directly on the spectrum, as shown in the following subsection. With correlations between the sampling instances, the correction of the spectrum is more complicated and all values β'_k are required. Analytical derivations for the following test cases with correlations of the sampling instances are given later, namely below the introduction of the particular simulation procedure. These derivations are limited to the shown test cases and they are no longer valid with other sampling schemes. As a universal alternative, the correction coefficients can also be estimated numerically directly from the data taken, as shown in the next but one subsection.

2.3 Purely random sampling with independent sampling instances

For continuous distribution of sampling intervals and with independent sampling instances, all cross-products are independent, which leads to $\beta'_k = 1$ for all $k \neq 0$. For $k = 0$, a number of n samples in a time unit $\Delta\tau$ delivers n^2 cross- and self-products, where the number n of samples in that time unit is Poisson distributed with the probability

$$P(n) = \frac{\alpha'^n}{n!} e^{-\alpha'} \tag{10}$$

and with the mean value of α' . The mean number of cross- and self-products for $k = 0$ then becomes

$$\beta_0 = \sum_{n=0}^{\infty} n^2 P(n) = \alpha' + \alpha'^2, \tag{11}$$

which, after normalization with α'^2 , finally leads to

$$\beta'_k = \begin{cases} 1 + \frac{1}{\alpha'} & \text{for } k = 0 \\ 1 & \text{otherwise.} \end{cases} \tag{12}$$

For discrete distribution of sampling intervals and with independent randomly occurring missed samples, again all cross-products are independent, which leads to $\beta'_k = 1$ for all $k \neq 0$ also here. In contrast to the previous case, where multiple sampling instances were possible within one time unit, in nominally equidistant sampling, only one or zero samples can occur per time interval. Therefore, $R_s(0)$ can include only self-products, which occur with the mean rate of α' . After normalization with α'^2 , this finally leads to

$$\beta'_k = \begin{cases} \frac{1}{\alpha'} & \text{for } k = 0 \\ 1 & \text{otherwise.} \end{cases} \tag{13}$$

In both cases, continuous and discrete distribution of sampling intervals without correlations, only β'_0 deviates from unity. The corresponding correlation function can be corrected at lag time zero by

$$\hat{R}_s(0) = \frac{1}{\beta'_0} R_s(0) \tag{14}$$

Due to the correspondence between the power spectral density and the correlation function given in Eq. (7), a correction of $R_s(0)$ as in Eq. (14) leads to an offset correction of the power spectral density

$$\hat{S}_s(f_j) = S_s(f_j) - \Delta\tau \left(1 - \frac{1}{\beta'_0}\right) R_s(0) \tag{15}$$

for all frequencies f_j .

Fortunately, the procedure directly corrects the spectral estimates, which are the focus of this investigation. On the other hand, the procedure involves the corresponding correlation function. However, only the value $R_s(0)$ is needed. Since the spectral estimates have been obtained directly from the data, a procedure without values of the correlation function would be preferable. In that case, the

transformation of the spectra into corresponding correlation functions could be dropped. For this purpose, the mean signal power

$$P_s = \frac{1}{N} \sum_{i=1}^N x_i^2 \tag{16}$$

is used instead, where $\hat{R}_s(0) = \frac{R_s(0)}{\beta'_0} \approx P_s$. A deviation results from the fact that in $R_s(0)$ or $\hat{R}_s(0)$ in addition to self-products also cross-products of samples within single time units of $\Delta\tau$ may occur, while only self-products are counted in P_s . However, this deviation occurs only for continuous distributions of sampling intervals, and it becomes significant only for large values of α' above one. Fortunately, for continuous distributions of sampling intervals, the interval $\Delta\tau$ can be chosen arbitrarily small, ensuring α' to be sufficiently smaller than one. In the limit of infinitesimal small $\Delta\tau$, $R_s(0)$ and $\beta'_0 P_s$ become the same and the correction of the spectrum becomes

$$\hat{S}(f) = S(f) - \Delta\tau (\beta'_0 - 1) P_s. \tag{17}$$

Using the derivation of β'_0 from above, for random sampling with continuous distribution of sampling intervals this reduces to

$$\hat{S}_s(f) = S_s(f) - \frac{P_s}{\alpha} \tag{18}$$

and for randomly and independent missing samples with discrete distribution of sampling intervals to

$$\hat{S}_s(f) = S_s(f) - P_s \left(\frac{1}{\alpha} - \Delta\tau\right). \tag{19}$$

2.4 Empirical estimates of the correction coefficients

For unknown spectral composition of the sampling function or that of the data gaps, the correction procedure with theoretical values of α' and β'_k is not practicable because the values β_k resp. β'_k are not known a priori. In that case, the number of self- and cross-products and finally β'_k can be estimated for each lag time τ_k individually by directly estimating the correlation function of the sampling function. For that, Eq. (1) for the primary estimate of the spectrum and Eq. (8) for the appropriate correlation function can be reused with the sampling times t_i from the observed signal, this time with all values x_i replaced by a constant value of one. Since Lomb-Scargle's method

has problems with signals of constant value, instead the Fourier spectrum is generally used for the empirical estimation of β'_k , yielding

$$\beta'_k = \text{IDFT} \left\{ \frac{M}{N^2} \left| \sum_{i=1}^N e^{-2\pi i f_j t_i} \right|^2 \right\} = \frac{1}{N^2} \cdot \sum_{j=-\lfloor \frac{M}{2} \rfloor}^{\lfloor \frac{M-1}{2} \rfloor} \left| \sum_{i=1}^N e^{-2\pi i f_j t_i} \right|^2 \cdot e^{2\pi i f_j \tau_k} \quad (20)$$

again with $f_j = j\Delta f, j = -\lfloor \frac{M}{2} \rfloor \dots \lfloor \frac{M-1}{2} \rfloor$ and $\tau_k = k\Delta\tau, k = -\lfloor \frac{M}{2} \rfloor \dots \lfloor \frac{M-1}{2} \rfloor$.

2.5 Mean value

The procedures given here can be applied with no changes to data with or without a mean value. However, in contrast to equidistant sampling, a non-zero mean value in combination with irregular sampling increases the estimation variance of the derived spectra. Therefore, a possible mean value in real data should be estimated and removed from the data before spectral analysis. Unfortunately, the estimation and removal of the mean value is another bias source for the estimated correlation function and finally for the spectrum, as has been analyzed in [47, 48] including appropriate corrections. To avoid interference with additional bias sources, the following test simulations are done with mean-free processes only. Accordingly, only Lomb-Scargle's method is used for direct comparison with

the Fourier spectrum, while generalized Lomb-Scargle's method has not been considered further.

2.6 Simulation

To demonstrate the ability of the estimation routines to derive bias-free estimates of the power spectral density, a moving-average stochastic process of order 200 is generated from white noise, sampled in four different ways and analyzed in Monte-Carlo simulation runs. The coefficients of the moving-average process are chosen such that the generated signals have an artificial spectrum with an exponentially increasing slope and with a distinct dip in the observed frequency range. Each run of the Monte-Carlo simulation generates a signal of such spectral characteristics with a total length of 200 tu (time units). To avoid the influence of further bias sources like incorrect assumptions of a periodic continuation of the signals, each signal fits together on both ends, justifying the assumption made above. The signals have no mean value and a standard deviation of 2 au (amplitude units). Then, four sampling schemes are applied to the time series: For the two cases with sampling instances from continuous time, interpolated values between the discrete samples of the simulated series are obtained by Whittaker-Shannon interpolation formula [49, 50]. (a) Purely random samples are taken from the interpolated time series independent from each other with a mean rate of $\alpha = 0.5 \text{ tu}^{-1}$. (b) Random samples are taken from the interpolated time series including a minimum distance of $d = 0.5 \text{ tu}$ between successive samples roughly mimicking processor dead times in laser Doppler applications. The mean sampling rate remains $\alpha = 0.5 \text{ tu}^{-1}$. For the two test cases with nomi-

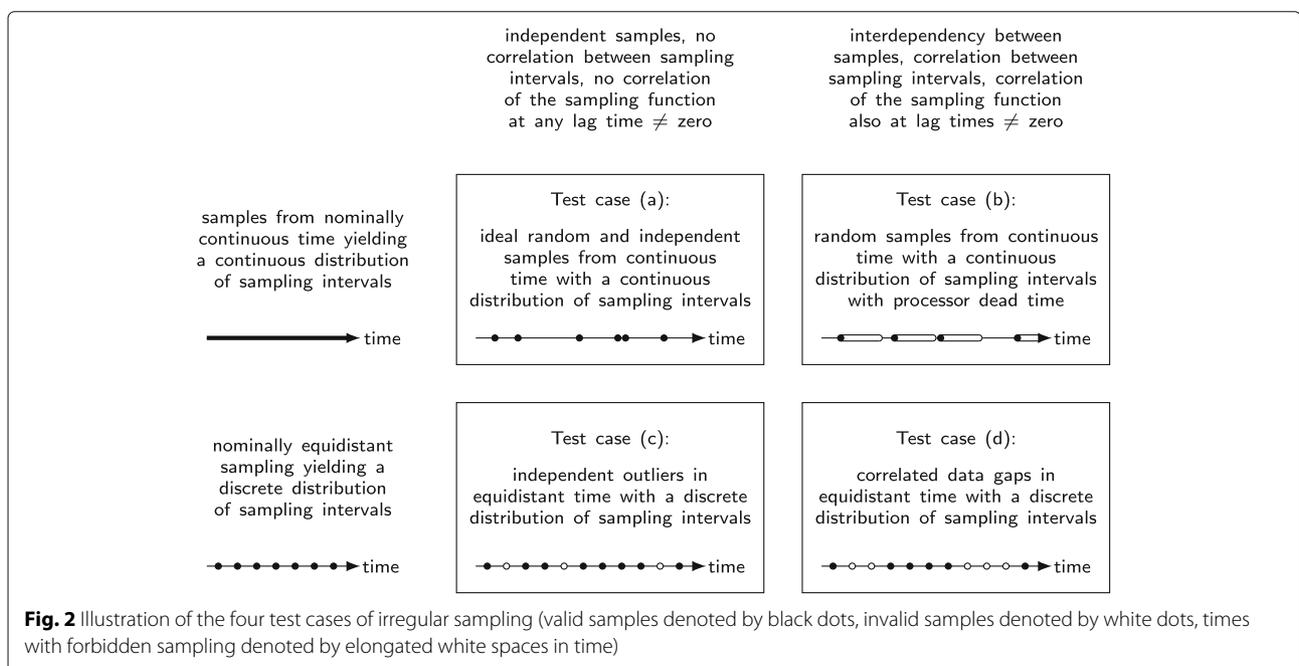


Fig. 2 Illustration of the four test cases of irregular sampling (valid samples denoted by black dots, invalid samples denoted by white dots, times with forbidden sampling denoted by elongated white spaces in time)

nally equidistant sampling, each sample of the simulated time series gets a corresponding weight of either one or zero from a random process to mimic outliers and data gaps. (c) Individual samples are taken out independently, with a probability of 50%. (d) Series of valid and invalid samples are specified, where the state of validity changes with a probability of 20% at each time step. The last procedure also yields 50% invalid samples on average, where the length of valid data or that of invalid data has an exponential distribution with a mean of five samples and the sequence of weights gets correlated. However, in all four cases a mean sampling rate of $\alpha = 0.5 \text{ tu}^{-1}$ is set. Figure 2 illustrates the sampling schemes for the four test cases and the appropriate classification according to the definition in the abstract. The arrangement of the four test cases will remain the same for all following figures in the Section 3.

2.7 Appropriate correction coefficients for the test cases

For the two cases (a) and (c) with purely random sampling

and independent sampling instances, the offset of the spectrum can be determined from the mean data rate and the spectrum can be corrected directly by removing the predicted offset as given in Eqs. (18) resp. (19). More effort is needed to derive the correction coefficients β_k and β'_k for the two cases (b) and (d) with correlation between the sampling intervals. Since all properties of the signal generation including the rules of the sampling process are known, particular bias correction coefficients can be derived analytically. The following derivations are valid for the particular sampling schemes used as examples in the present simulation only. Other sampling schemes need their own derivations.

For case (b) with random sampling in continuous time with a continuous distribution of sampling intervals and with a minimum time d between successive samples, let $P_0(n)$ be the probability to have n samples within one time unit $\Delta\tau$. For n samples in a time unit, the number of pairs of samples (self- and cross-products) is n^2 . The mean number β_0 of pairs between all samples within a time unit

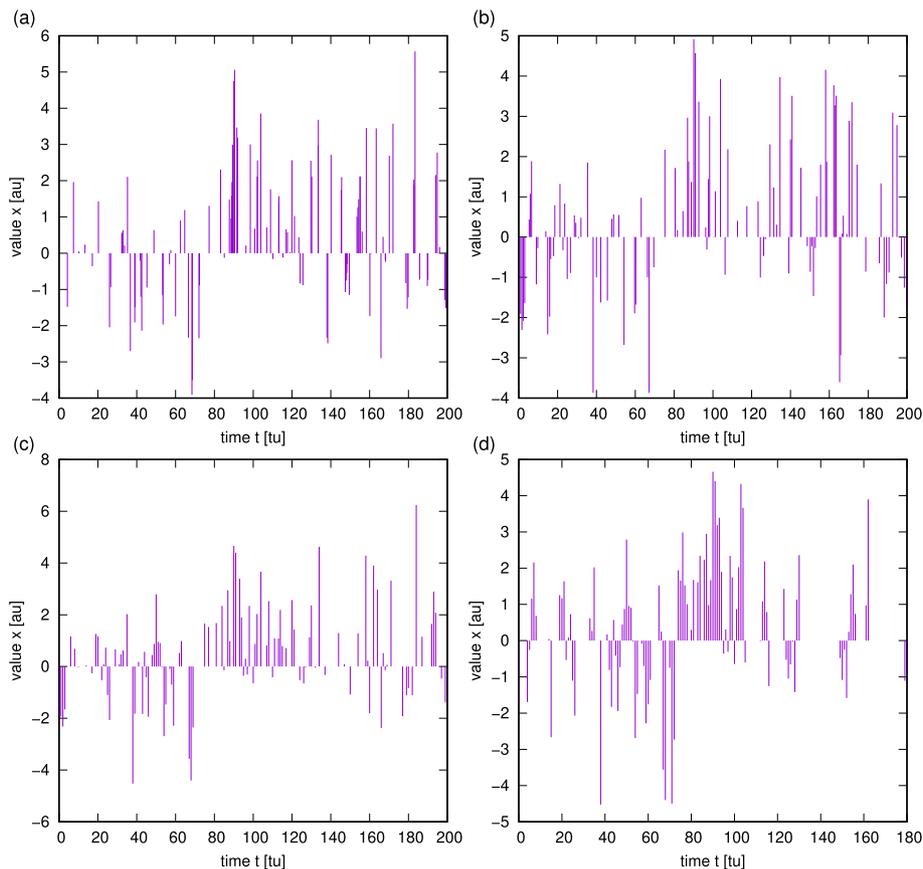


Fig. 3 Single realizations of the signals. **a** Ideal random and independent samples from continuous time with a continuous distribution of sampling intervals, **b** random samples from continuous time with a continuous distribution of sampling intervals with processor dead time, **c** independent outliers in equidistant time with a discrete distribution of sampling intervals, and **d** correlated data gaps in equidistant time with a discrete distribution of sampling intervals (au, amplitude unit; tu, time unit)

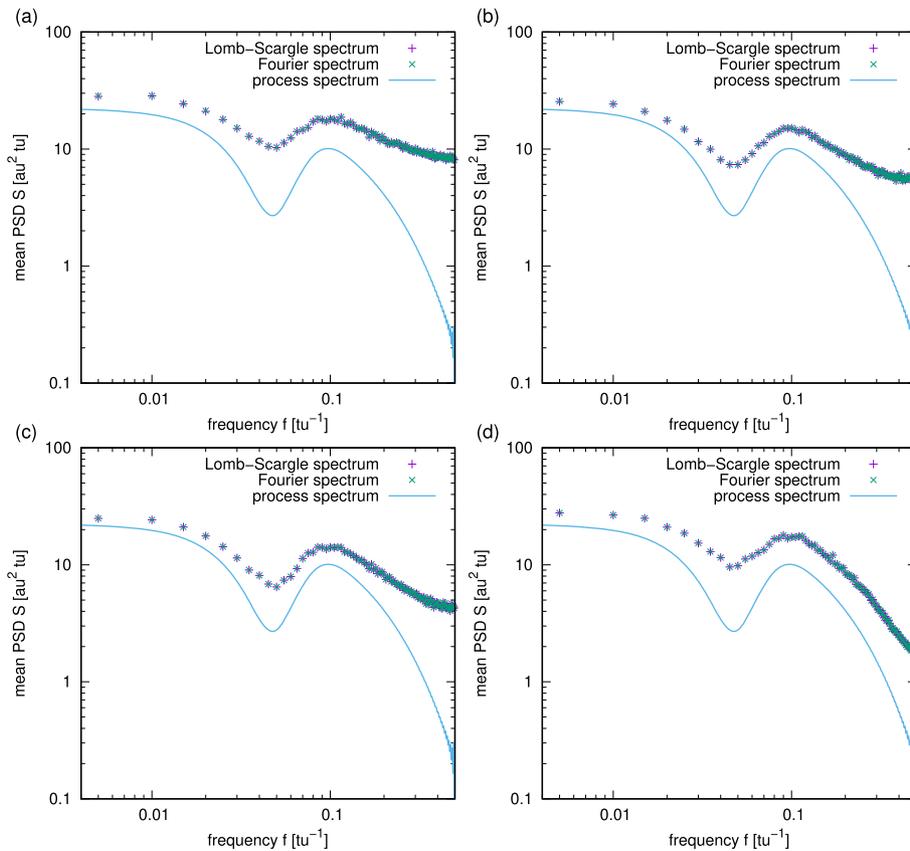


Fig. 4 Mean of the primary estimates (no correction) of the power spectral density from 1000 signals. **a** Ideal random and independent samples from continuous time with a continuous distribution of sampling intervals, **b** random samples from continuous time with a continuous distribution of sampling intervals with processor dead time, **c** independent outliers in equidistant time with a discrete distribution of sampling intervals, and **d** correlated data gaps in equidistant time with a discrete distribution of sampling intervals (au, amplitude unit; tu, time unit)

is then derived by summation over all possible numbers n of samples as

$$\beta_0 = \sum_{n=0}^{\lfloor \frac{1}{d'} \rfloor + 1} n^2 P_0(n) \tag{21}$$

with the normalized minimum time between samples $d' = \frac{d}{\Delta\tau}$. To derive the probabilities $P_0(n)$, also samples before the investigated time unit must be considered, because their delay times influence the probabilities of following samples within the actually investigated time unit. The probability to have no preceding sample effecting the actually investigated time unit is $P_e = 1 - \alpha' d'$ (e - empty). The probability to have the actually investigated time unit fully covered by a delay time of a preceding event is $P_f = \alpha' \max\{0, d' - 1\}$ (f - full). In this case, no events can occur within the actually investigated time unit. This case occurs only, if $d' > 1$. The probability to have the actually investigated time unit partially covered by the delay time from

a preceding sample event is $P_p = \alpha' \min\{1, d'\}$ (p - partial). The probability to have n samples in a time unit $\Delta\tau$ consists of the sum of these three cases, yielding

$$P_0(n) = P_e(n) + P_f(n) + P_p(n), \tag{22}$$

where $P_e(n)$, $P_f(n)$, and $P_p(n)$ are the probabilities to obtain n samples within the investigated time unit with either no preceding sample to be considered, time unit fully covered by the delay time of a preceding sample or partially covered. The probabilities of these three cases finally are

$$P_e(n) = (1 - \alpha' d') [P'(n, 1 - d' \max\{0, n - 1\}) - P'(n + 1, 1 - nd')] \tag{23}$$

$$P_f(n) = \begin{cases} \alpha' \max\{0, d' - 1\} & \text{for } n = 0 \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

$$P_p(n) = \int_0^{\min\{1,d'\}} \alpha' [P'(n, x - d' \max\{0, n-1\}) - P'(n+1, x - nd')] dx \quad (25)$$

where $P'(n, x)$ is the probability to have at least n samples within the (normalized) fraction x of a time unit, which is

$$P'(n, x) = 1 - \sum_{j=0}^{n-1} \frac{(\alpha''x)^j}{j!} e^{-\alpha''x} \quad (26)$$

with the normalized rate of further samples α'' in the remaining time between dead time intervals after assumed samples

$$\alpha'' = \frac{\alpha'}{1 - \alpha'd'} \quad (27)$$

For the mean number β_k of pairs of samples between two different time units ($k \neq 0$), only cross-products between any sample within one time unit and any sample in the other time unit can contribute. Therefore, one sample is assumed at time t_a in one time unit and another sample is assumed at time t_b in a different time unit, which is k time units away from the initial one. Because the autocorrelation is symmetric, only $k > 0$ is discussed. For $k < 0$, $|k|$ can be used instead of k to derive all required parameters. The sample at t_a can occur at any time within its time unit with a mean rate of α . Normalization with the time unit $\Delta\tau$ yields $t'_a = \frac{t_a}{\Delta\tau}$. Then, the sample at t'_a occurs with the mean rate of α' . The occurrence of the sample at $t'_b = \frac{t_b}{\Delta\tau}$ depends on the number n of further samples between t'_a and t'_b . However, reducing the time to the remaining fraction between all dead times of other samples, the mean rate of the sample at t_b is α'' as in the case for $k = 0$ above. The number of further samples n between t'_a and t'_b and their dead times nd' plus the dead time of the initial sample at t'_a , finally leads to the dependence between the sampling instances. Without limitation of generality, t'_a shall occur in time unit number zero. The mean number of pairs of samples between the two time units then becomes

$$\beta_k = \int_{-0.5}^{0.5} \int_{|k|-0.5}^{|k|+0.5} dt'_a dt'_b \alpha' \alpha'' \sum_{n=0}^{\left\lfloor \frac{t'_b - t'_a}{d'} - 1 \right\rfloor} P(n, t'_b - t'_a - (n+1)d') \quad (28)$$

where $P(n, x)$ is the probability to have n further samples within the (normalized) time interval x , which is

$$P(n, x) = \frac{(\alpha''x)^n}{n!} e^{-\alpha''x} \quad (29)$$

From the mean number of pairs of samples β_k the correction coefficients β'_k (for all k , including $k = 0$) can be obtained as

$$\beta'_k = \frac{\beta_k}{\alpha'^2} \quad (30)$$

For case (d) with correlated data gaps in discrete time and with a discrete distribution of sampling intervals, the probability is derived that a valid sample occurs at a given time unit and another one occurs $|k|$ time units away. The probability of having a valid sample at the first instance is $\alpha' = 0.5$ in this simulation. For a given valid sample at the first instance, the probability of having another valid sample at the second instance depends on the number of possible changes n from valid data to invalid data and vice versa. Between the two instances up to $|k|$ changes can occur with a mean number of changes per time unit of $c' = 0.2$ in this simulation. A change occurs at a given time instance with a probability of c' , no change occurs with a probability of $1 - c'$. An odd number of changes yields an invalid sample at the second instance, while an even number of changes yields a valid sample. The changes can occur at any time instance between $1\Delta\tau$ and $|k|\Delta\tau$, where the result depends on their number but not on their order. The mean number β_k of pairs of samples then becomes

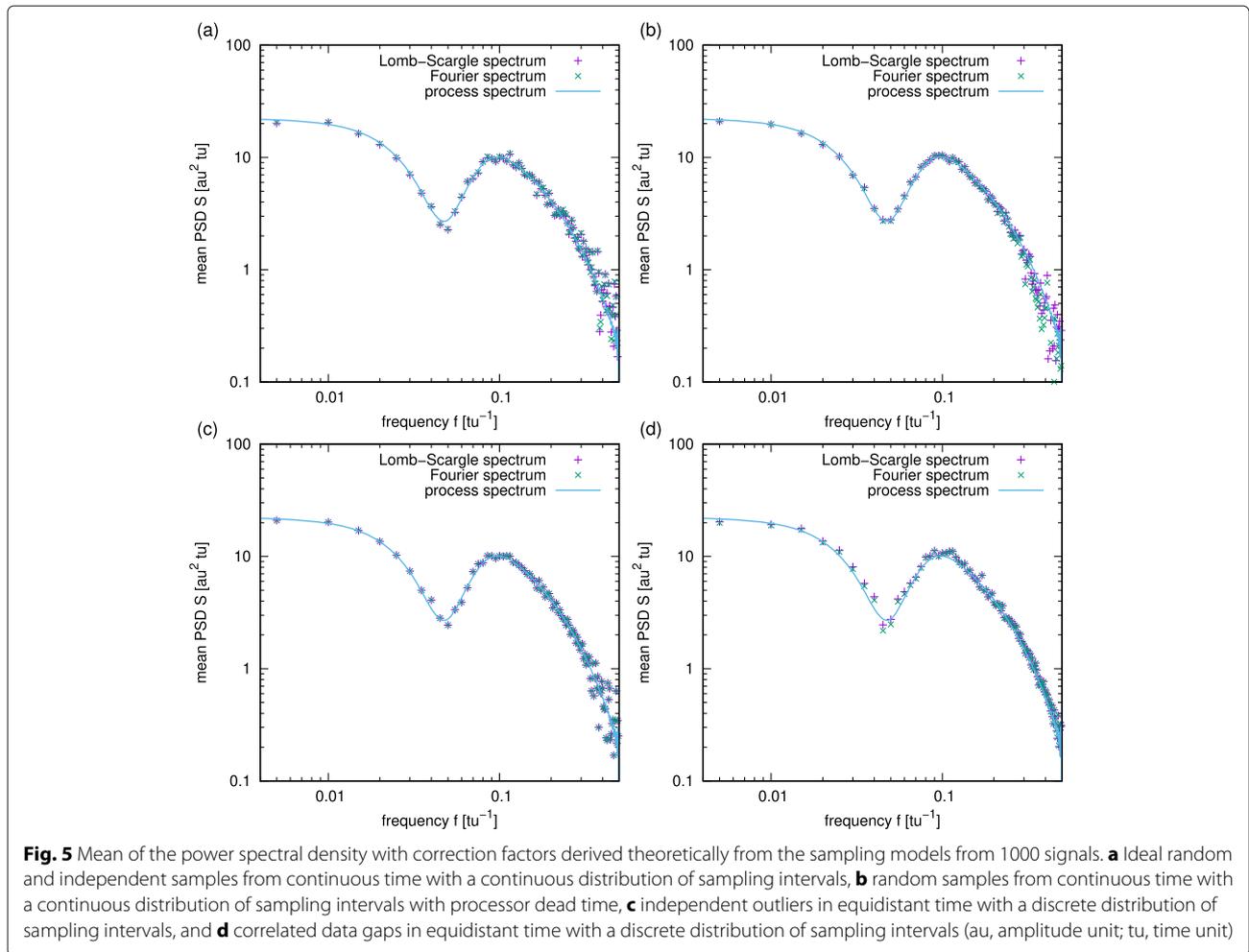
$$\beta_k = \alpha' \sum_{n'=0}^{\lfloor \frac{|k|}{2} \rfloor} \binom{|k|}{n'} (1 - c')^{|k|-n} c'^n = \frac{\alpha'}{2} [1 + (1 - 2c')^{|k|}] \quad (31)$$

with $n = 2n'$, which finally leads to the correction coefficients

$$\beta'_k = \frac{\beta_k}{\alpha'^2} = 1 + (1 - 2c')^{|k|} \quad (32)$$

3 Results and discussion

In Fig. 3, individual realizations of the signals are shown taken from the same original simulation of the discrete time series from the moving-average process. Random sampling at instances from continuous time therefore involves continuous interpolation of the signal before sampling. The plots show the valid samples as impulses to better illustrate the different sampling schemes. In Fig. 3a and b, the sampling instances are chosen from continuous time with a continuous distribution of sampling intervals, where in Fig. 3a the sampling is purely random, while in Fig. 3b, a minimum distance between consecutive samples is complied. In Fig. 3c and d, the sampling is nominally equidistant (with missing samples) yielding a discrete distribution of sampling intervals. While in Fig. 3c, only individual samples (outliers) are taken out independently,



longer sequences of missing samples (data gaps) can be identified in Fig. 3d. However, in all four cases, a mean sampling rate of $\alpha = 0.5 \text{ tu}^{-1}$ is obtained on average.

From the signals simulated, direct spectral estimates are derived, based on the Fourier spectrum and based on Lomb-Scargle's method. In Fig. 4, the mean spectra averaged over 1000 realizations of the simulation and the spectral estimates are shown for the four sampling schemes. A significant bias can be identified in all four sampling cases between the estimate of the power spectral density and the spectrum of the simulated process. Note that this estimate of the power spectral density, however, is a bias-free estimate of the observed signal. The deviation between the estimated spectrum and that of the simulated process is a direct result of the irregular sampling. The sampling scheme changes the spectral content of the observed signal in comparison to the spectrum of the process under observation. Therefore, the spectral composition of the primary estimates directly depends on the characteristics of the sampling scheme. Accordingly, the

four averages of biased primary spectra in Fig. 4a–d also have different spectral characteristics, while no significant differences can be observed between the estimates based on the Fourier spectrum and that based on Lomb-Scargle's method.

From the primary estimates of the power spectral density, the procedures from above are used for bias correction. For random sampling (no correlation) with a continuous distribution of sampling intervals (case a) and for independent individual outliers from equidistant data (case c), the offset of the spectrum is corrected directly as in Eq. (18) resp. Eq. (19) using $\alpha = 0.5 \text{ tu}^{-1}$. For correlated random sampling with a continuous distribution of sampling intervals and for correlated data gaps with a discrete distribution of sampling intervals the spectrum is transformed into the corresponding correlation function first. The latter one is corrected as in Eq. (9) using the appropriate coefficients β'_k , which have been derived according to the models of the different sampling schemes in Eqs. (28)–(30) and (32). Finally, the improved estimates

of the correlation function are transformed back into the corresponding spectra for comparison. The results in Fig. 5 show that for all four sampling schemes, the bias correction succeeds, yielding bias-free estimates of the power spectral density for the two test cases (c) and (d) with discrete distribution for sampling intervals. In cases (a) and (b) with continuous distribution of sampling intervals, the obtained spectra are almost bias-free only. For the highest frequencies resolved, a small aliasing error remains, leading to slightly increased values. Due to the random sampling in continuous time, this alias has no sharp boundary frequency [4] and it is occurring smeared over a certain range of frequencies. However, this error is a result of the insufficient information extraction by the sampling process from the observed process and the bias correction introduced here is not able to add this missing information.

In Fig. 6, the correction coefficients β'_k are derived directly from the data sets as in Eq. (20) and the corresponding correlation functions are corrected as in Eq. (9) and transformed back into the corresponding spectra for

all four sampling schemes. Also, in this case, except for the small remaining aliasing error in cases (a) and (b), the correction succeeds for all investigated sampling schemes.

4 Conclusion

Random sampling of time series causes a systematic error of the spectral estimation compared to the observed process. Lomb-Scargle’s method for the spectral estimation of irregularly sampled time series, often used as a reference, does not show any advantages in this respect compared to a direct spectral estimation using the Fourier transform. The systematic errors caused by the irregular sampling can be analyzed, predicted, and finally corrected using the methods presented in this paper. The presented methods are not limited to certain sampling models. Beyond the present simulation, this is the possibility to obtain bias-free direct spectral estimates from irregularly sampled data, independent of the spectral composition of the sampling scheme. The only requirements for the presented correction method are that the irregular sampling is independent of the values of the observed process and that

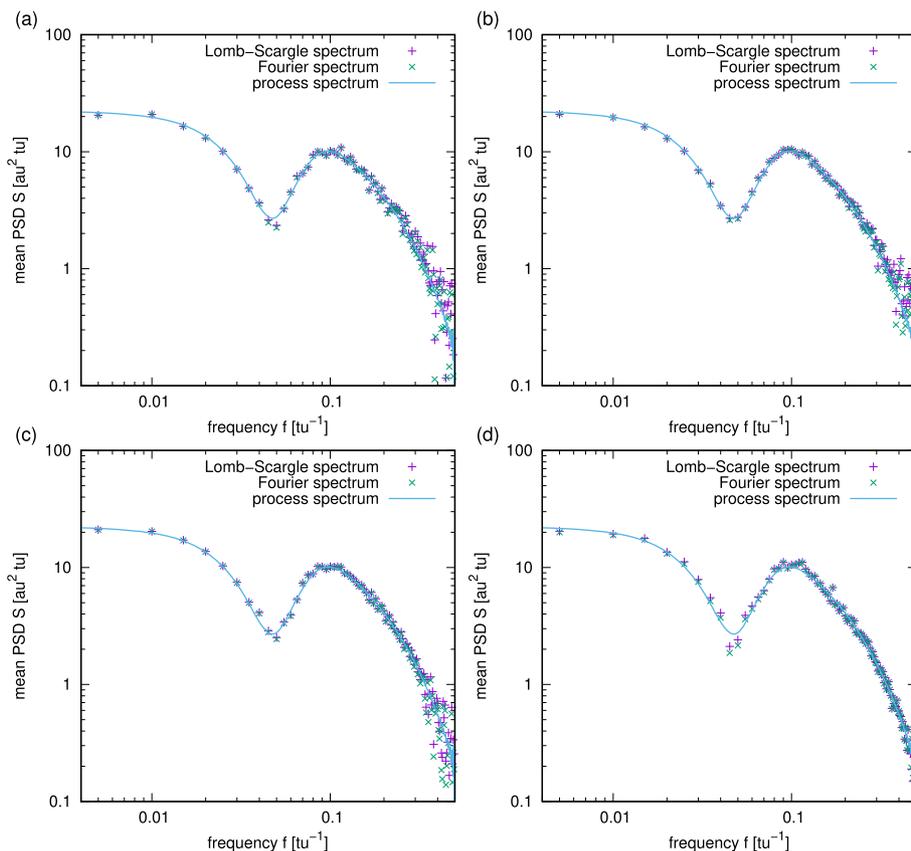


Fig. 6 Mean of the power spectral density with empirical correction factors from 1000 signals. **a** Ideal random and independent samples from continuous time with a continuous distribution of sampling intervals, **b** random samples from continuous time with a continuous distribution of sampling intervals with processor dead time, **c** independent outliers in equidistant time with a discrete distribution of sampling intervals, and **d** correlated data gaps in equidistant time with a discrete distribution of sampling intervals (au, amplitude unit; tu, time unit)

enough pairs of samples can be built for any lag time. If this is given, even static patterns of the irregular sampling function can be corrected, like periodic drop outs. Also, data sequences with significant parts missing can still be processed to appropriate spectra or corresponding correlation functions. Therefore, this is the first universal solution of bias-free spectral and correlation estimation for a broad range of irregular sampling processes.

This work is part of a broader attempt to bias-free estimation of correlation functions and spectra from irregularly sampled data. The authors' experience from data processing in laser Doppler applications inspired to enhance the methods developed there towards more universality and towards an extended range of applications. Next steps are measures to reduce the estimation variance, since the conservation of information about the spectral content seems to be less efficient with irregular sampling than with equidistant sampling. Other methods of spectral analysis like quantization of arrival times or slot correlation, known from laser Doppler applications also are potential candidates for broader use and are object of further investigations.

Abbreviations

DFT: Discrete Fourier transform; IDFT: Inverse discrete Fourier transform; tu: Time unit; au: Amplitude unit

Acknowledgements

Not applicable.

Authors' contributions

The research and the outcome of this specific publication are result of a long cooperation between the authors about fundamentals and applications of signal processing of unevenly sampled data. For the present manuscript, ND contributed to the definition of requirements and applications in laser Doppler velocimetry and related laser-based measurement systems, VK contributed to the generalization of sampling cases, broadening of relevance due to new applications and writing, and HN contributed with methods, programming, simulations, and writing. All authors read and approved the final manuscript.

Funding

This research had no specific funding. Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials

The procedures are available as Python source codes as supplementary material to this article together with all data sets under [44].

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute of General Electrical Engineering, University of Rostock, Faculty of Computer Science and Electrical Engineering, Albert-Einstein-Straße 2, 18059 Rostock, Germany. ²Institute of Communications Engineering, University of Rostock, Faculty of Computer Science and Electrical Engineering, Richard-Wagner-Straße 31, 18119 Rostock-Warnemünde, Germany. ³Max Planck Institute for Dynamics and Self-Organization, Am Faßberg 17, 37077 Göttingen, Germany.

Received: 22 May 2020 Accepted: 11 December 2020

Published online: 18 January 2021

References

1. Y. Yeh, H. Z. Cummins, Localized fluid flow measurements with an He-Ne laser spectrometer. *Appl. Phys. Lett.* **4**, 176–178 (1964). <https://doi.org/10.1063/1.1753925>
2. F. Durst, A. Melling, J. H. Whitelaw, *Principles and practice of laser Doppler anemometry*. (Academic Press, London, 1976)
3. H.-E. Albrecht, N. Damaschke, M. Borys, C. Tropea, *Laser Doppler and phase Doppler measurement techniques*. (Springer, Berlin, 2003). <https://doi.org/10.1007/978-3-662-05165-8>
4. H. S. Shapiro, R. A. Silverman, Alias-free sampling of random noise. *SIAM J. Appl. Math.* **8**, 245–248 (1960)
5. I. Bilinskis, A. Mikelsons, *Randomized signal processing*. (Prentice Hall, University of Michigan, 1992)
6. M. Hajar, M. El Badaoui, A. Raad, F. Bonnardot, Discrete random sampling: theory and practice in machine monitoring. *Mech. Syst. Signal Process.* **123**, 386–402 (2019)
7. H. Semaili, N. Boumaaz, A. Maali, A. Soulmani, A. Ghammaz, J.-F. Diouris, in *2nd International Conference on Wireless Intelligent and Distributed Environment for Communication*, ed. by I. Woungang, S. K. Dhurandher. Exploring the application of random sampling in spectrum sensing (Springer Nature Switzerland AG, Cham, 2019), pp. 143–152
8. N. R. Lomb, Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* **39**, 447–462 (1976). <https://doi.org/10.1007/BF00648343>
9. J. D. Scargle, Studies in astronomical time series analysis. II, statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **263**, 835–853 (1982). <https://doi.org/10.1086/160554>
10. S. Ferraz-Mello, Estimation of periods from unequally spaced observations. *Astrophys. J.* **86**, 619–624 (1986). <https://doi.org/10.1086/112924>
11. G. Foster, Time series analysis by projection. I, statistical properties of Fourier analysis. *Astrophys. J.* **111**, 541–554 (1996). <https://doi.org/10.1086/117805>
12. P. Buchhave, W. K. George Jr, J. L. Lumley, The measurement of turbulence with the laser Doppler anemometer. *Ann. Rev. Fluid Mech.* **11**, 443–503 (1979). <https://doi.org/10.1146/annurev.fl.11.010179.002303>
13. W. K. George, P. D. Beuther, J. L. Lumley, in *Proceedings of the Dynamic Flow Conference, Skovlunde, Denmark, (PO Box 121, DK-2740 Skovlunde)*. Processing of random signals, (1978), pp. 757–800. https://doi.org/10.1007/978-94-009-9565-9_43
14. M. Gaster, J. B. Roberts, Spectral analysis of randomly sampled signals. *J. Inst. Maths. Applics.* **15**, 195–216 (1975). <https://doi.org/10.1093/imamat/15.2.195>
15. W.T. Mayo, in *Proceedings of the Dynamic Flow Conference 1978 on Dynamic Measurements in Unsteady Flows*, ed. by B.W. Hansen. Spectrum Measurements with Laser Velocimeters (Springer, Dordrecht, 1978). https://doi.org/10.1007/978-94-009-9565-9_4
16. M. Gaster, J. B. Roberts, The spectral analysis of randomly sampled records by a direct transform. *Proc. R. Soc. Lond. A.* **354**, 27–58 (1977). <https://doi.org/10.1098/rspa.1977.0055>
17. W. K. George, Quantitative measurement with the burst-mode laser Doppler anemometer. *Exp. Therm. Fluid Sci.* **1**, 29–40 (1988). [https://doi.org/10.1016/0894-1777\(88\)90045-3](https://doi.org/10.1016/0894-1777(88)90045-3)
18. A. K. P. Rajpal, in *Proc. ASME/JSMF Fluids Eng. and Laser Anemometry Conf.* Power spectrum estimates of LDA measurements using Scargle periodogram analysis (American Society of Mechanical Engineers (ASME), Hilton Head Island, 1995), pp. 411–415
19. C. M. Velte, W. K. George, P. Buchhave, Estimation of burst-mode LDA power spectra. *Exp. Fluids.* **55**, 1674 (2014). <https://doi.org/10.1007/s00348-014-1674-z>
20. H. Nobach, Corrections to the direct spectral estimation for laser Doppler data. *Exp. Fluids.* **56**, 109 (2015). <https://doi.org/10.1007/s00348-015-1980-0>
21. P. Buchhave, C. M. Velte, Reduction of noise and bias in randomly sampled power spectra. *Exp. Fluids.* **56**, 79 (2015). <https://doi.org/10.1007/s00348-015-1922-x>
22. H. Nobach, Fuzzy time quantization and local normalization for the direct spectral estimation from laser Doppler velocimetry data. *Exp. Fluids.* **56**, 182 (2015). <https://doi.org/10.1007/s00348-015-2050-3>
23. C. M. Velte, P. Buchhave, W. K. George, in *Proc. 17th Int. Symp. on Appl. of Laser Techn. to Fluid Mechanics*. Power spectrum estimation of randomly sampled signals (Instituto Superior Tecnico, Lisbon, 2014). paper 3.2.3
24. N. Damaschke, V. Kühn, H. Nobach, A fair review of non-parametric bias-free autocorrelation and spectral methods for randomly sampled

- data in laser Doppler velocimetry. *Digit. Signal Proc.* **76**, 22–33 (2018). <https://doi.org/10.1016/j.dsp.2018.01.018>
25. P. Buchhave, C. M. Velte, W. K. George, The effect of dead time on randomly sampled power spectral estimates. *Exp. Fluids*. **55**, 1680 (2014). <https://doi.org/10.1007/s00348-014-1680-1>
 26. P. Buchhave, C. M. Velte, W. K. George, in *Proc. 17th Int. Symp. on Appl. of Laser Techn. to Fluid Mechanics*. The effect of finite measurement volume on power spectra from a burst type LDA (Instituto Superior Tecnico, Lisbon, 2014). paper 3.2.1
 27. C. M. Velte, P. Buchhave, W. K. George, Dead time effects in laser Doppler anemometry measurements. *Exp. Fluids*. **55**, 1836 (2014). <https://doi.org/10.1007/s00348-014-1836-z>
 28. E. Masry, in *Statistical Analysis of Irregularly Observed Time Series in Statistics, Lecture Notes in Statistics*, ed. by E. Parzen. Spectral and probability density estimation from irregularly observed data (Springer, New York, 1983), pp. 224–250. https://doi.org/10.1007/978-1-4684-9403-7_11
 29. A. Mathias, F. Grond, R. Guardans, D. Seese, M. Canela, H. H. Diebner, Algorithms for spectral analysis of irregularly sampled time series. *J. Stat. Softw.* **11**, 1–27 (2004). <https://doi.org/10.18637/jss.v011.i02>
 30. A. Rivoira, G. A. Fleury, A consistent nonparametric spectral estimator for randomly sampled signals. *IEEE Trans. Signal Process.* **52**, 2383–2395 (2004). <https://doi.org/10.1109/TSP.2004.832002>
 31. P. Stoica, N. Sandgren, Spectral analysis of irregularly-sampled data: paralleling the regularly-sampled data approach. *Digit. Signal Proc.* **16**, 712–734 (2006). <https://doi.org/10.1016/j.dsp.2006.08.012>
 32. P. Babu, P. Stoica, Spectral analysis of nonuniformly sampled data — a review. *Digit. Signal Proc.* **20**, 359–378 (2010). <https://doi.org/10.1016/j.dsp.2009.06.019>
 33. R. H. Jones, Spectral estimates and their distributions, part II. *Scand. Actuar. J.* **1962**, 135–153 (1962). <https://doi.org/10.1080/03461238.1962.10405942>
 34. P. A. Scheinok, Spectral analysis with randomly missed observations: the binomial case. *Ann. Math. Stat.* **36**, 972–977 (1965). <https://doi.org/10.1214/aoms/1177700069>
 35. P. Bloomfield, Spectral analysis with randomly missing observations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **32**, 369–380 (1970)
 36. R. H. Jones, Spectrum estimation with missing observations. *Ann. Inst. Stat. Math.* **23**, 387–398 (1971). <https://doi.org/10.1007/BF02479238>
 37. R. Vio, T. Strohmer, W. Wamsteker, On the reconstruction of irregularly sampled time series. *Publ. Astron. Soc. Pac.* **112**, 74–90 (2000). <https://doi.org/10.1086/316495>
 38. M. A. Ghazal, A. Elhassanein, Periodogram analysis with missing observations. *J. Appl. Math. Comput.* **22**, 209–222 (2006). <https://doi.org/10.1007/BF02896472>
 39. R. H. Jones, Spectral analysis with regularly missed observations. *Ann. Math. Stat.* **33**, 455–461 (1962). <https://doi.org/10.1214/aoms/1177704572>
 40. E. Parzen, On spectral analysis with missing observations and amplitude modulation. *Sankhya Indian J. Stat. Ser. A.* **245**, 383–392 (1963)
 41. P. Stoica, E. G. Larsson, J. Li, Adaptive filter-bank approach to restoration and spectral analysis of gapped data. *Astron. J.* **120**, 2163–2173 (2000). <https://doi.org/10.1086/301572>
 42. C. Munteanu, C. Negrea, M. Echim, K. Mursula, Effect of data gaps: comparison of different spectral analysis methods. *Ann. Geophys.* **34**, 437–449 (2016). <https://doi.org/10.5194/angeo-34-437-2016>
 43. G. Plantier, S. Moreau, L. Simon, J.-C. Valière, A. L. Duff, J. Baillet, Nonparametric spectral analysis of wideband spectrum with missing data via sample-and-hold interpolation and deconvolution. *Digit. Signal Process.* **22**, 994–1004 (2012). <https://doi.org/10.1016/j.dsp.2012.05.012>
 44. Supplementary Data Repository. <http://www.nambis.de/publications/jaspdg.html>. Accessed 29 Oct 2020
 45. M. Zechmeister, M. Kürster, The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. *Astron. Astrophys.* **496**, 577–584 (2009). <https://doi.org/10.1051/0004-6361/200811296>
 46. A. Khintchine, Korrelationstheorie der stationären stochastischen Prozesse. *Math. Ann.* **109**, 604–615 (1934). <https://doi.org/10.1007/BF01449156>
 47. T. J. Vogelsang, J. Yang, Exactly/nearly unbiased estimation of autocovariances of a univariate time series with unknown mean. *J. Time Ser. Anal.* **37**, 723–740 (2016). <https://doi.org/10.1111/jtsa.12184>
 48. H. Nobach, Practical realization of Bessel's correction for the bias-free estimation of the auto-covariance and the cross-covariance functions. <http://www.nambis.de/publications/BC4corr17.html>. Accessed 29 Oct 2020
 49. E. T. Whittaker, On the functions which are represented by the expansions of the interpolation-theory. *Proc. R. Soc. Edinburgh*. **35**, 181–194 (1915). <https://doi.org/10.1017/S0370164600017806>
 50. J. M. Whittaker, On the cardinal function of interpolation theory. *Proc. Edinburgh Math. Soc. Ser. 1.* **2**, 41–46 (1927). <https://doi.org/10.1017/S0013091500007318>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)