



# Asymptotically-unbiased nonparametric estimation of the power spectral density from uniformly-spaced data with missing samples

Cédric Chavanne

Institut des sciences de la mer, Université du Québec à Rimouski, Rimouski, Québec, Canada

**Correspondence:** Cédric Chavanne ([cedric.chavanne@ensta.org](mailto:cedric.chavanne@ensta.org))

Received: 26 June 2025 – Revised: 29 January 2026 – Accepted: 31 January 2026 – Published: 20 February 2026

**Abstract.** The nonparametric estimation of the power spectral density of uniformly-spaced data with missing samples is revisited. Classical estimators, such as the standard periodogram and the Lomb-Scargle periodogram, are biased when samples are missing. The classical method to obtain an asymptotically-unbiased estimator is to take the finite Fourier transform of the standard unbiased estimator of the autocorrelation function. However, the latter estimator is not necessarily positive semidefinite, so its finite Fourier transform can yield negative power spectral density values at some frequencies. To avoid this problem, Gao et al. (2021) have proposed taking the absolute value of the finite Fourier transform of the standard unbiased estimator of the autocorrelation function to estimate the power spectral density of data with missing samples. We show that the estimator of power spectral density proposed by Gao et al. (2021) is even more biased than classical estimators and should not be used for quantitative analysis of spectral characteristics such as spectral slope in log-log space. We illustrate this using both synthetic data from fractional Brownian processes and actual data from a laboratory experiment of decaying turbulence in an active grid-generated air flow, to which we apply synthetic Bernoulli and batch-Bernoulli sampling functions to simulate missing samples. In fact, negative values of power spectral density estimates for particular realizations of a random process with missing samples should be retained, so that when sufficiently averaged the estimate will be nonnegative, and will not contain the bias induced from taking absolute values as Gao et al. (2021) propose. It is also proposed here to use the circular unbiased estimator of the autocorrelation function, the finite Fourier transform of which yields a power spectral density estimator identical to the standard periodogram estimator in the absence of missing samples. Its advantages are reduced variance and reduced computing memory usage compared to the finite Fourier transform of the standard unbiased estimator of the autocorrelation function. Both power spectral density estimators, when sufficiently averaged, are able to recover the  $-5/3$  spectral slope of the decaying turbulence data even when 50 % of the data are missing. A Matlab implementation of the proposed estimator is provided.

## 1 Introduction

Spectral analysis, i.e., estimating how the total power of a signal is distributed over frequency, finds applications in many diverse fields, such as meteorology, oceanography, astronomy, seismology, radar and sonar systems, medicine and economics, to name a few (Stoica and Moses, 2005). In particular, in meteorology and oceanography, a spectral analysis of turbulent flow quantities such as velocity and temperature is often performed to diagnose turbulence regimes based

on the slope of the power spectral density (PSD) in log-log space (e.g. Obukhov, 1941; Corrsin, 1951; Grant et al., 1962; Charney, 1971; Nastrom and Gage, 1985). The slope is usually determined by least-squares fitting a straight line to the PSD in log-log space, which requires an unbiased PSD estimator to properly diagnose turbulence regimes. The first non-parametric PSD estimator was the periodogram (Schuster, 1898). Wiener (1930) and Khintchine (1934) showed that PSD can also be estimated by taking the Fourier transform of the signal autocorrelation function, called the correl-

ogram estimator. Unfortunately, since processes are always observed over a finite period of time, both the finite periodogram and correlogram estimators are biased due to spectral leakage (Stoica and Moses, 2005). However, both estimators are asymptotically unbiased, which means that as the observed period of time becomes longer, the bias becomes smaller. The main problem with these estimators is that they are inconsistent, since their variance does not decrease as the observed period of time becomes longer. Several methods have been proposed in the literature to decrease the variance of the estimated PSD (Bartlett, 1948; Blackman and Tukey, 1958; Welch, 1967), but they have the drawback of increasing its bias.

The estimation of PSD becomes even more difficult when samples are missing within the observed period of time. The simplest way to deal with this problem is to compute the periodogram or the standard biased correlogram using only the available data, which is equivalent to replacing the missing samples with zeros. This increases both the bias and the variance of the PSD estimators. More effective methods have been developed with different approaches dedicated to processes with periodic (deterministic) signals and to random processes with continuous spectra. For the former, the most commonly-used method is the Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982). For the latter, which is more relevant for turbulence measurements, two different classes of methods have been proposed: parametric methods, which postulate a model for the data and therefore a functional form for its spectrum, and nonparametric methods, which make no assumption on the functional form of the spectrum. Nonparametric methods consist of either reconstructing missing samples through interpolation (e.g. Plantier et al., 2012), or refining the correlogram estimator using only the available data, without replacing missing samples with zeros. The latter method appears to have been first proposed by Jones (1962) for the special case of periodically missing samples, and subsequently generalized to more general missing data patterns (Parzen, 1963; Scheinok, 1965; Bloomfield, 1970; Jones, 1971; Efromovich, 2014; Damaschke et al., 2024), and to irregularly-sampled data (Gaster and Roberts, 1975; Clinger and Van Ness, 1976; Babu and Stoica, 2010).

A seemingly major problem with the latter approach is that, when some samples are missing, the estimated autocorrelation function is no longer necessarily positive semidefinite, potentially leading to negative PSD values at some frequencies. Although this problem has been acknowledged for a long time (e.g., Jones, 1971), a recently published work attempted to bypass the oddity of negative spectral density estimates by taking the absolute value of the Fourier transform of the standard unbiased estimator of the autocorrelation function (Gao et al., 2021, see line 5 of their Matlab functions `acf2psd1D` and `acf2psd2D` in their Figs. D2 and D3, respectively). Unfortunately, doing so does not provide an unbiased PSD estimator, as will be illustrated below. The code published by Gao et al. (2021) has been used in a few other

published works (Li et al., 2021; Ma et al., 2024; Robache et al., 2025) in which PSD are estimated from data with missing samples. A primary purpose of this paper is to refute the practice of replacing negative PSD estimates with their absolute values, since after averaging, the PSD will be biased to a degree which depends on the number and distribution of missing samples. Power spectral slopes are used in these papers to diagnose turbulence regimes. Potentially biased PSD estimates may therefore distort spectral slopes and lead to erroneous diagnostics.

Although most of the literature on the subject has focused on reducing the variance of PSD estimators (see Efromovich, 2018, for a recent account), following Damaschke et al. (2024), we will focus here on the bias of PSD estimators, assuming that their variance can be sufficiently reduced by averaging a large number of independent estimates. The main purpose of this paper is to stress that to obtain an asymptotically-unbiased estimation of the PSD of a signal with missing samples, it is actually necessary that the estimator yield sometimes negative values. Since the estimator is asymptotically-unbiased, by averaging a sufficiently large number of independent PSD estimates, the averaged values should converge toward the true (positive) PSD of the signal at all frequencies, provided that the sampling function is independent of the signal and that the period of observation is sufficiently long.

The paper is organized as follows. Section 2 gives some theoretical background about Wiener–Khinchin’s theorem for both infinite (Sect. 2.1) and finite data (Sect. 2.2), and proposes a new PSD estimator for data with missing samples (Sect. 2.3). Section 3 illustrates the effects of missing samples on PSD estimation using synthetic data from fractional Brownian motion processes (Sect. 3.1) and actual data from a laboratory experiment of decaying turbulence in an active grid-generated air flow (Sect. 3.2). Section 4 proposes some discussion and conclusions. The derivation of Wiener–Khinchin’s theorem for finite discrete data is given in Appendix A, and the algorithm to compute the proposed PSD estimator is detailed and generalized to compute the cross-spectral density of two random processes in Appendix B.

## 2 Theoretical background

### 2.1 Wiener–Khinchin theorem for infinite continuous data

The Wiener–Khinchin theorem (Wiener, 1930; Khintchine, 1934) states that the auto-spectral density function  $S_{xx}(f)$  of an infinite continuous stationary random process  $x(t)$  (real or complex) is the infinite Fourier transform of the autocorrelation function  $R_{xx}(\tau)$  of the process, a result often used to define the auto-spectral density function (e.g. Stoica and Moses, 2005; Bendat and Piersol, 2011) by

$$S_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-i2\pi f\tau} d\tau, \tag{1}$$

where the auto-correlation function  $R_{xx}(\tau)$  is defined by

$$R_{xx}(\tau) = E[x(t + \tau)x^*(t)], \tag{2}$$

where  $E$  is the expected value operator and  $*$  denotes complex conjugation.  $S_{xx}(f)$  is finite, provided that

$$\int_{-\infty}^{\infty} |R_{xx}(\tau)| d\tau < \infty, \tag{3}$$

which requires that the mean value of  $x(t)$  be zero.

Another definition of the auto-spectral density function is based on the finite Fourier transform of  $x(t)$ , which is defined by

$$X(f, T) = \int_{-T/2}^{T/2} x(t) e^{-i2\pi ft} dt, \tag{4}$$

where  $T$  is a finite time interval. Then, the auto-spectral density function is defined by

$$S_{xx}(f) = \lim_{T \rightarrow +\infty} E[S_{xx}(f, T)], \tag{5}$$

where

$$S_{xx}(f, T) = \frac{1}{T} X(f, T) X^*(f, T). \tag{6}$$

The Wiener–Khinchin theorem states that Eqs. (1) and (5) are mathematically equivalent.

### 2.2 Wiener–Khinchin theorem for finite discrete data

The Wiener–Khinchin theorem is formally valid only for infinite stationary random processes. In practice, processes are observed over finite time intervals, which leads to a subtle modification of the theorem. The derivation of the Wiener–Khinchin theorem for finite discrete data is given in Appendix A, and the main result is presented here.

Consider a time series of  $N$  samples  $\{x_n\}, n = 0, \dots, N-1$ , sampled at equally spaced time intervals  $\Delta t$  from an infinite, continuous, and stationary random process  $x(t)$ , such that  $x_n = x(n\Delta t)$ , spanning the time period  $T = N\Delta t$ . Its discrete finite Fourier transform is defined by

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi \frac{k}{N} n \Delta t} \quad k = 0, \dots, N-1. \tag{7}$$

The periodogram estimator of the auto-spectral density of  $x$  is defined by (Stoica and Moses, 2005)

$$S_{xx,k}^p = \frac{1}{T} X_k X_k^*. \tag{8}$$

The Wiener–Khinchin theorem for finite discrete data states that the periodogram estimator of the auto-spectral density of  $x$  is the finite Fourier transform of the circular estimator  $R_{xx}^c$  of the autocorrelation function of  $x$ ,

$$S_{xx,k}^p = \sum_{n=0}^{N-1} R_{xx,n}^c e^{-i2\pi \frac{k}{N} n \Delta t} \quad k = 0, \dots, N-1, \tag{9}$$

where  $R_{xx,n}^c$  is defined by (Bendat and Piersol, 2011)

$$R_{xx,n}^c = \begin{cases} \frac{1}{N} \sum_{m=0}^{N-1} x_m x_m^* & \text{for } n = 0, \\ \frac{1}{N} \left( \sum_{m=0}^{N-1-n} x_{m+n} x_m^* + \sum_{m=N-n}^{N-1} x_{m-n} x_m^* \right) & \text{for } n = 1, \dots, N-1. \end{cases} \tag{10}$$

An important property of  $R_{xx}^c$  is that it is positive semidefinite, which ensures that its Fourier transform yields only positive PSD values.  $R_{xx}^c$  is also an unbiased estimator.

Applying the Wiener–Khinchin theorem for infinite data (Eq. 1) directly to finite data leads to the standard correlogram estimator:

$$S_{xx,k}^b = \sum_{n=-(N-1)}^{N-1} R_{xx,n}^b e^{-i2\pi \frac{k}{2N-1} n \Delta t} \quad k = -(N-1), \dots, N-1, \tag{11}$$

where  $R_{xx,n}^b$  is the standard biased estimator of the auto-correlation function of  $x$ , defined by (Stoica and Moses, 2005)

$$R_{xx,n}^b = \begin{cases} \frac{1}{N} \sum_{m=0}^{N-1-n} x_{m+n} x_m^* & \text{if } n \geq 0, \\ \frac{1}{N} \sum_{m=-n}^{N-1} x_{m+n} x_m^* & \text{if } n < 0. \end{cases} \tag{12}$$

Like  $R_{xx}^c$ ,  $R_{xx}^b$  is positive semidefinite, but it is a biased estimator. Because of this, another correlogram estimator is sometimes used:

$$S_{xx,k}^u = \sum_{n=-(N-1)}^{N-1} R_{xx,n}^u e^{-i2\pi \frac{k}{2N-1} n \Delta t} \quad k = -(N-1), \dots, N-1, \tag{13}$$

where  $R_{xx,n}^u$  is the standard unbiased estimator of the auto-correlation function of  $x$ , defined by (Stoica and Moses, 2005)

$$R_{xx,n}^u = \begin{cases} \frac{1}{N-n} \sum_{m=0}^{N-1-n} x_{m+n} x_m^* & \text{if } n \geq 0, \\ \frac{1}{N+n} \sum_{m=-n}^{N-1} x_{m+n} x_m^* & \text{if } n < 0. \end{cases} \tag{14}$$

Although  $R_{xx}^u$  is an unbiased estimator, it is not necessarily positive semidefinite, potentially producing negative values for  $S_{xx}^u$ .

2.3 Nonparametric estimation of the power spectral density from data with missing samples

Consider now an incomplete record of  $N - N'$  samples spanning the time period  $T = N\Delta t$ , with  $N' < N$  missing samples. Skipping missing samples to estimate the PSD using any of the above estimators is equivalent to replacing the missing samples with zeros. Records with missing samples can therefore be represented as the product of the uninterrupted observed process  $x$  with a sampling series  $\{w_n\}$ ,  $n = 0, \dots, N - 1$ , defined by

$$w_n = \begin{cases} 1 & \text{if } x_n \text{ is recorded,} \\ 0 & \text{otherwise,} \end{cases} \quad (15)$$

yielding the recorded series  $y_n = w_n x_n$ .

Since the Fourier transform of  $y$  is the convolution of the Fourier transform of  $x$  with the Fourier transform of  $w$ , the latter will influence the PSD estimation using the periodogram estimator (Eq. 8). Wiener–Khinchin’s theorem shows that the resulting PSD will be biased since, with missing values replaced by zeros, the circular autocorrelation estimator (Eq. 10) becomes biased, as there are less than  $N$  terms available in the summations. But it also provides a way to obtain an unbiased PSD estimator for data with missing samples, as first proposed by Jones (1962) (who used  $R_{xx}^u$  rather than  $R_{xx}^c$  as proposed here), as follows:

$$S_{xx,k}^c = \sum_{n=0}^{N-1} R_{xx,n}^c e^{-i2\pi \frac{k}{N} n \Delta t} \quad k = 0, \dots, N - 1, \quad (16)$$

where  $R_{xx,n}^c$  is replaced by

$$R_{xx,n}^c = \begin{cases} \frac{1}{N_n} \sum_{m=0}^{N-1-n} y_m y_m^* & \text{for } n = 0, \\ \frac{1}{N_n} \left( \sum_{m=0}^{N-1-n} y_{m+n} y_m^* + \sum_{m=N-n}^{N-1} y_{m+n-N} y_m^* \right) & \text{for } n = 1, \dots, N - 1, \end{cases} \quad (17)$$

where  $N_n$  is the number of available data pairs for lag  $n$ , given by

$$N_n = \begin{cases} \sum_{m=0}^{N-1-n} w_m^2 & \text{for } n = 0, \\ \left( \sum_{m=0}^{N-1-n} w_{m+n} w_m + \sum_{m=N-n}^{N-1} w_{m+n-N} w_m \right) & \text{for } n = 1, \dots, N - 1. \end{cases} \quad (18)$$

Following Efromovich (2014), when no data is available for a specific lag  $n$  ( $N_n = 0$ ), the autocorrelation for this lag cannot be estimated and is set to zero. When at least one pair of data is available for all lags, Eq. (17) is an unbiased estimator of the autocorrelation of  $x$ , provided that the sampling function  $w$  is independent of  $x$ .

We will also consider the standard unbiased estimator  $S_{xx}^u$  (Eq. 13) by replacing Eq. (14) with

$$R_{xx,n}^u = \begin{cases} \frac{1}{M_n} \sum_{m=0}^{N-1-n} y_{m+n} y_m^* & \text{if } n \geq 0, \\ \frac{1}{M_n} \sum_{m=-n}^{N-1} y_{m+n} y_m^* & \text{if } n < 0, \end{cases} \quad (19)$$

where  $M_n$  is the number of available data pairs for the lag  $n$ , given by

$$M_n = \begin{cases} \sum_{m=0}^{N-1-n} w_{m+n} w_m & \text{if } n \geq 0, \\ \sum_{m=-n}^{N-1} w_{m+n} w_m & \text{if } n < 0. \end{cases} \quad (20)$$

An advantage of using  $R_{xx}^c$  (Eq. 17) instead of  $R_{xx}^u$  (Eq. 19) is that  $N_n \geq M_n$  for all lags  $n$ .

The estimator proposed by Gao et al. (2021) is

$$S_{xx,k}^G = \left| \sum_{n=-(N-1)}^{N-1} R_{xx,n}^u e^{-i2\pi \frac{k}{2N-1} n \Delta t} \right| \quad k = -(N - 1), \dots, N - 1, \quad (21)$$

where  $R_{xx,n}^u$  is given by Eq. (19).

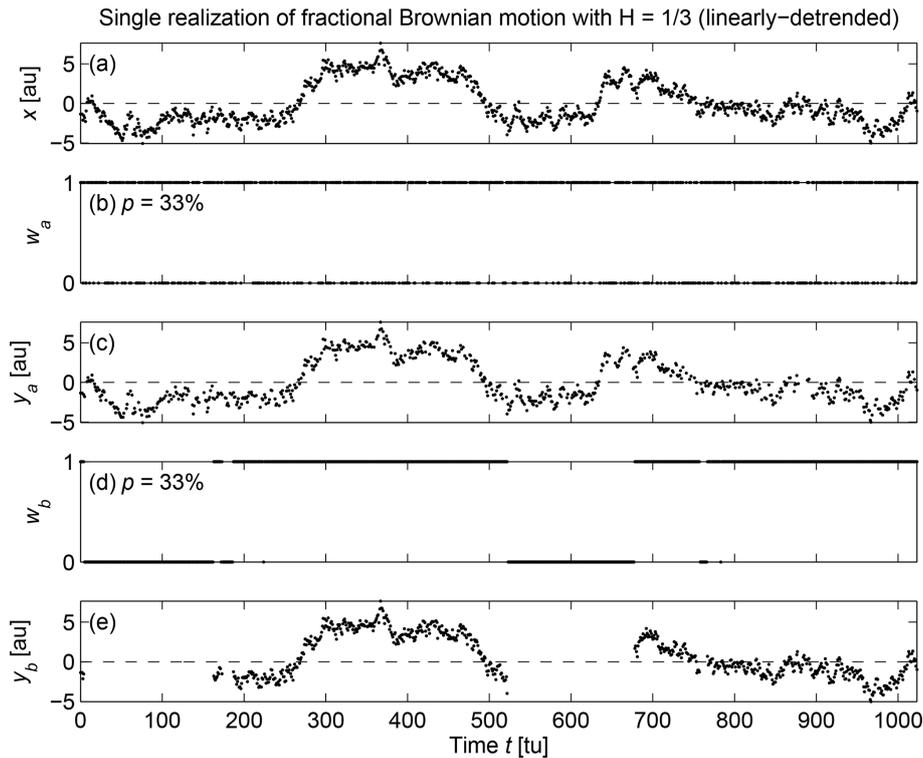
Finally, we will also consider the standard biased estimator  $S_{xx}^b$  (Eq. 11) by replacing Eq. (12) with

$$R_{xx,n}^b = \begin{cases} \frac{1}{M_0} \sum_{m=0}^{N-1-n} y_{m+n} y_m^* & \text{if } n \geq 0, \\ \frac{1}{M_0} \sum_{m=-n}^{N-1} y_{m+n} y_m^* & \text{if } n < 0. \end{cases} \quad (22)$$

This is equivalent to estimating the autocorrelation function by simply replacing missing samples with zeros and correcting for the energy loss due to the sampling window  $w$ . Dividing by a constant value  $M_0$  instead of  $M_n$  ensures that  $R_{xx}^b$  is positive semidefinite, but it is a biased estimator.

A problem when using Eq. (16) with Eq. (17), or Eq. (13) with Eq. (19), when samples are missing, is that negative values can be obtained for the PSD at some frequencies, due to the fact that the unbiased estimators of the autocorrelation (Eq. 17) and Eq. (19) are not necessarily positive semidefinite. This motivated the use of the absolute value for the estimator (Eq. 21) proposed by Gao et al. (2021). In the next section, we illustrate that this seemingly undesirable property of Eqs. (17) and (19) is actually necessary to obtain unbiased PSD estimates when averaging a very large number of independent realizations.

For comparison, we will also consider the widely-used Lomb-Scargle periodogram estimator  $S_{xx}^L$  (Lomb, 1976; Scargle, 1982) obtained using Matlab’s function `plomb`. An efficient algorithm for computing Eqs. (16)–(18) using the fast Fourier transform, along with a generalization to estimate the cross-correlation and cross-spectral density of two stationary random processes, is provided in Appendix B and at <https://doi.org/10.5281/zenodo.18405064> (Chavanne, 2026), along with a similar algorithm for computing Eq. (13) with Eqs. (19)–(20) by zero-padding the time series with  $N$  zeros to separate the two contributions from  $R_{xx}^u$  to  $R_{xx}^c$  (see Eq. A15 in Appendix A and Fig. 11.6 in Bendat and Piersol, 2011). Therefore, an additional advantage of using  $R_{xx}^c$  (Eq. 17) instead of  $R_{xx}^u$  (Eq. 19) is that it requires half the memory usage to compute.



**Figure 1.** Example (one realization) for  $p = 33\%$  missing samples of (a) original fractional Brownian motion (with  $H = 1/3$ ) series  $x$ , (b) sampling function with Bernoulli missing samples  $w_a$ , (c) observed series  $y_a$ , (d) sampling function with batch-Bernoulli missing samples  $w_b$ , and (e) observed series  $y_b$ . For better visualization, missing samples are not shown in (c) and (e), rather than being set to zero.

### 3 Applications

#### 3.1 Fractional Brownian motion synthetic data

To illustrate the effects of missing samples on PSD estimation, we first generate synthetic random data using Matlab's `wfbm` function, which generates fractional Brownian motion signals for a given Hurst parameter  $H$  ( $0 < H < 1$ ) and length  $N$ , following the algorithm proposed by Abry and Sellan (1996). Although the fractional Brownian motion is not a stationary random process, its spectral density exists for  $0 < H \leq 1/2$  and is proportional to  $f_k^{-(2H+1)}$ , which corresponds to the limit of the periodogram estimate when  $N \rightarrow \infty$  (Kuleshov and Grudin, 2013). We test two values of the Hurst parameter,  $H = 1/3$ , which gives a spectral slope  $-(2H + 1) = -5/3$ , and  $H = 1/2$ , which gives a spectral slope  $-(2H + 1) = -2$ . The first slope value is chosen to illustrate the case of isotropic homogeneous turbulence in the inertial range, if the frequency  $f$  represents the wavenumber (Kolmogorov, 1941; Obukhov, 1941, 1949; Corrsin, 1951). The second slope value is chosen to illustrate the case of discontinuities associated with, e.g. atmospheric or oceanic fronts (Boyd, 1992). For each value of  $H$ , we generate  $M$  series of  $N$  samples, and the expected value operator  $E$  is approximated by the arithmetic average over the  $M$  realizations. Here, we chose  $N = 2^{10}$  (yielding almost

three decades in frequency) and  $M = 10^4$ . Each series were linearly-detrended using the Matlab function `detrend`. An example of a linearly-detrended series for  $H = 1/3$  is shown in Fig. 1a.

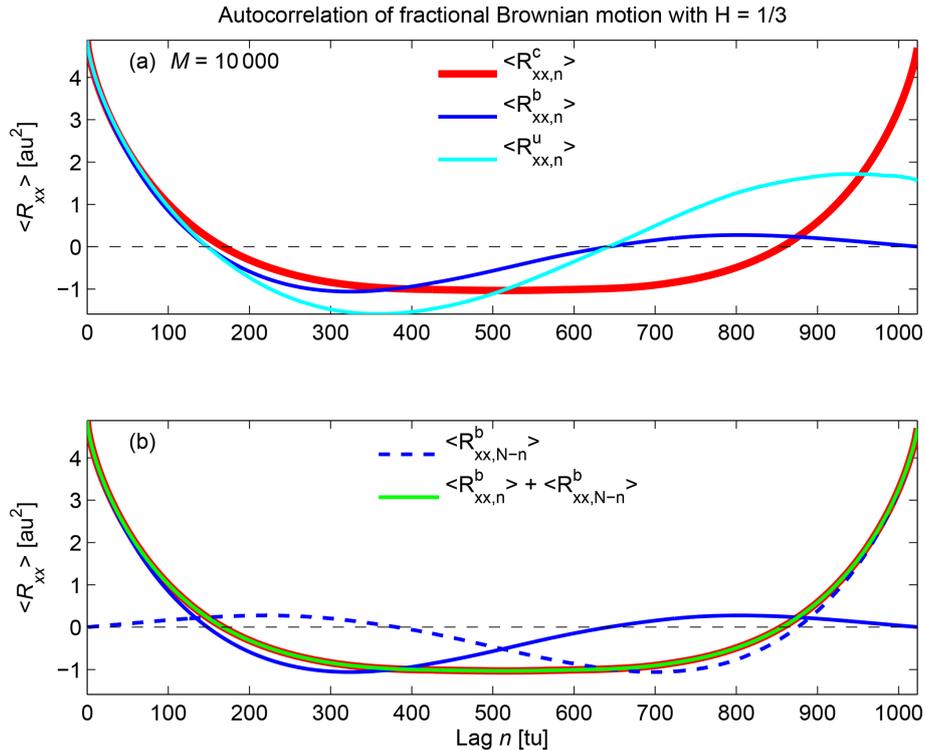
##### 3.1.1 Effects of autocorrelation function estimator

The autocorrelation functions of  $x$  for  $H = 1/3$  estimated by the different estimators are shown in Fig. 2a. Equation (A14) is illustrated in Fig. 2b for real-valued series.

The PSD of  $x$  for  $H = 1/3$  and  $H = 1/2$  estimated by the different estimators are shown in Fig. 3. With no missing sample, all estimators recover the theoretical spectral slopes over most of the frequency domain for both  $H = 1/3$  and  $H = 1/2$ , but  $S_{xx}^u$  is noisier than the other estimators, especially at high frequencies, even when averaged over  $M = 10^4$  realizations.

##### 3.1.2 Effects of missing samples

We now turn to illustrate the effects of missing samples on PSD estimation. For each of the  $M$  series, we generate two sets of sampling functions,  $w_a$  with randomly missing samples occurring independently from each other (Fig. 1b) and  $w_b$  with batches of consecutively missed samples of randomly varying lengths (Fig. 1d). The number of missing sam-



**Figure 2.** (a) Estimates (averaged over  $M = 10^4$  realizations) of circular unbiased autocorrelation  $R_{xx,n}^c$  (red), standard biased autocorrelation  $R_{xx,n}^b$  (blue), and standard unbiased autocorrelation  $R_{xx,n}^u$  (cyan). Equation (A14) is illustrated in (b), where the flipped standard biased autocorrelation  $R_{xx,N-n}^b$  (dashed blue) and the sum  $R_{xx,n}^b + R_{xx,N-n}^b$  (green) are also shown.

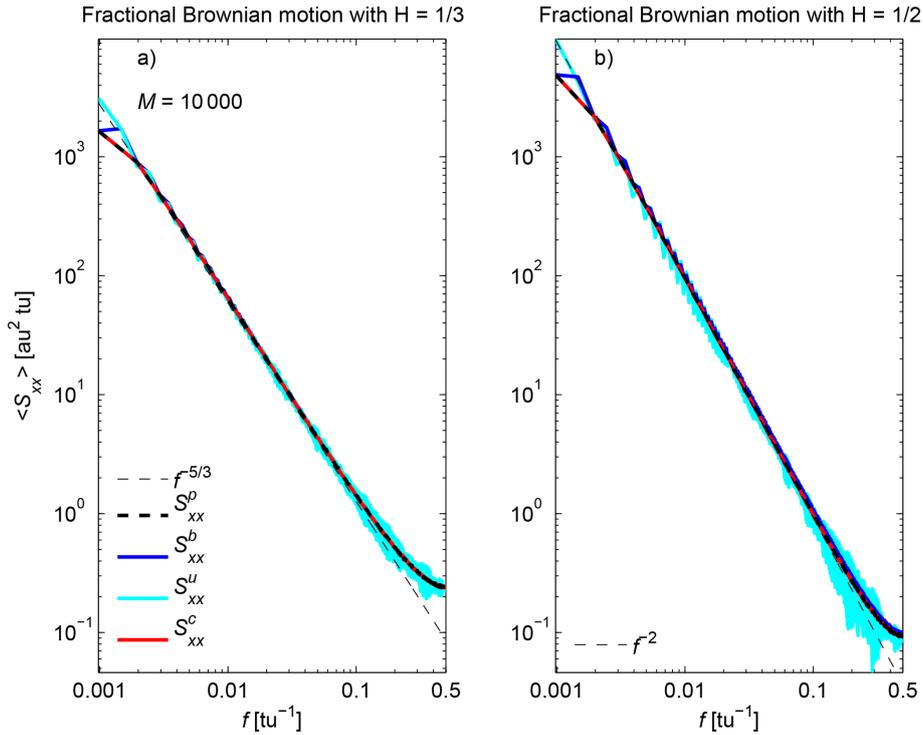
ples  $N'$  is set to  $p$  percent of  $N$  for both sampling functions. In order for  $N'$  to maintain the same value for each realization of  $w_a$  and  $w_b$ , they are generated as follows. For  $w_a$ , initially set to  $w_{a,n} = 1, n = 1, \dots, N$ , we generate  $N'$  values  $j_n, n = 1, \dots, N'$ , comprised between 2 and  $N - 1$  (to ensure that there are available samples at the beginning and end of the time series), sampled uniformly at random, without replacement, using the Matlab function `rand-sample`, and set  $w_{a,j_n} = 0, n = 1, \dots, N'$ . For  $w_b$ , we sequentially generate a series of lengths of batches of missing samples by generating a first length  $L_1^m$ , randomly taken between 1 and  $N'$  using `rand-sample`, then a second length  $L_2^m$ , randomly taken between 1 and  $N' - L_1^m$ , and so on until  $\sum_{n=1}^{N_L} L_n^m = N'$ , where  $N_L$  is the number of batches of missing samples. The series  $L_n^m, n = 1, \dots, N_L$ , is then randomly reordered using the Matlab function `randperm`. We also generate a series of lengths of batches of available samples,  $L_n^a, n = 1, \dots, N_L + 1$  (to ensure that there are available samples at the beginning and end of the time series) with a similar procedure, such that  $\sum_{n=1}^{N_L+1} L_n^a = N - N'$ . Finally, we intersperse available and missing samples with their respective batch lengths as follows:  $w_{b,n} = 1, n = 1, \dots, L_1^a, w_{b,n} = 0, n = L_1^a + 1, \dots, L_1^a + L_1^m$ , etc.  $w_a$  is called a Bernoulli missing

mechanism and is an example of missing completely at random (MCAR), while  $w_b$  is called a batch-Bernoulli missing mechanism and is an example of missing at random (MAR) (Rubin, 1976; Efromovich, 2018; Little and Rubin, 2019).

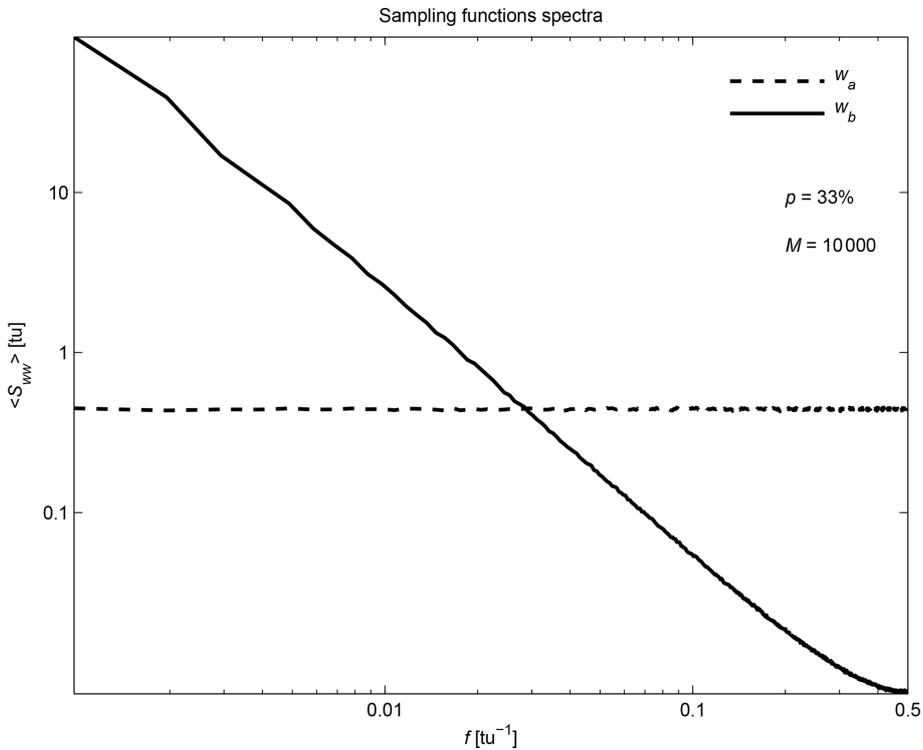
In this section, we set  $p = 33\%$ . The resulting observed series (with missing samples not shown instead of replaced by zeros) are shown in Fig. 1c and e. PSD of the sampling functions are computed using Eq. (8) and averaged over the  $M$  realizations, showing that  $w_a$  has a white spectrum while  $w_b$  has a red spectrum (Fig. 4).

PSD of the observed series are computed using the different estimators presented in Sect. 2.3 and averaged over the  $M$  realizations. Spectra for  $H = 1/3$  are shown in Fig. 5. For both missing data schemes, both  $S_{xx}^u$  and  $S_{xx}^c$  recover the  $-5/3$  spectral slope over most of the frequency range, but these estimates get noisy at high frequencies, even when averaged over  $M = 10^4$  realizations, especially for  $S_{xx}^u$ .  $S_{xx}^b$  and  $S_{xx}^L$  perform similarly, with a slightly biased spectral slope for  $w_b$  and PSD leveling at high frequencies for  $w_a$ . In contrast,  $S_{xx}^G$  is strongly biased for both missing data schemes.

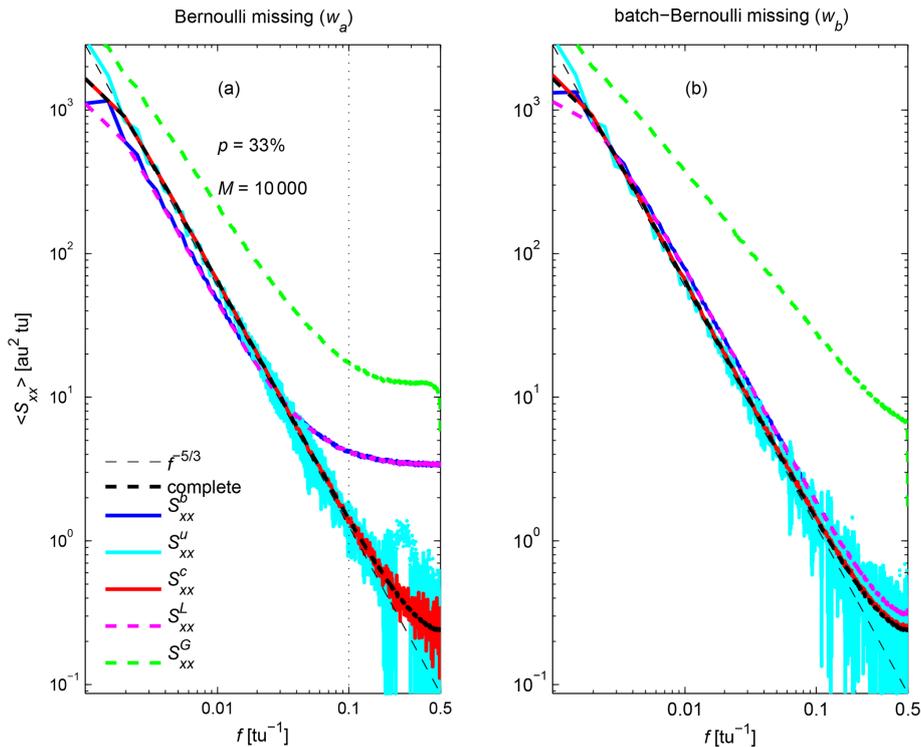
The unbiasedness of  $S_{xx}^u$  and  $S_{xx}^c$  comes with an increased variance of the estimator (Jones, 1962) and potential negative values at some frequencies (Jones, 1971), as illustrated in Fig. 6 for  $f = 0.1$ . Although PSD are always positive for  $S_{xx}^b, S_{xx}^L$ , and  $S_{xx}^G$ , the average values of these estimates



**Figure 3.** PSD estimates (averaged over  $M = 10^4$  realizations) of fractional Brownian motion for (a)  $H = 1/3$  and (b)  $H = 1/2$ , estimated by  $S_{xx}^p$  (thick black),  $S_{xx}^b$  (blue),  $S_{xx}^u$  (cyan), and  $S_{xx}^c$  (red). The theoretical spectral slopes are shown for reference (dashed black).



**Figure 4.** PSD (averaged over  $M = 10^4$  realizations) of the sampling functions  $w_a$  (dashed line) and  $w_b$  (solid line) for  $p = 33\%$  missing samples.



**Figure 5.** PSD (averaged over  $M = 10^4$  realizations) of the complete (without missing samples) fractional Brownian motion with  $H = 1/3$  (thick black), and of the observed series with missing samples using  $S_{xx}^b$  (blue),  $S_{xx}^u$  (cyan),  $S_{xx}^c$  (red),  $S_{xx}^L$  (magenta), and  $S_{xx}^G$  (green) for the sampling functions (a)  $w_a$  and (b)  $w_b$  with  $p = 33\%$  missing samples. The theoretical spectral slope is shown for reference (dashed black). The vertical dotted line in (a) indicates the frequency ( $f = 0.1$ ) at which the PSD histograms are shown in Fig. 6.

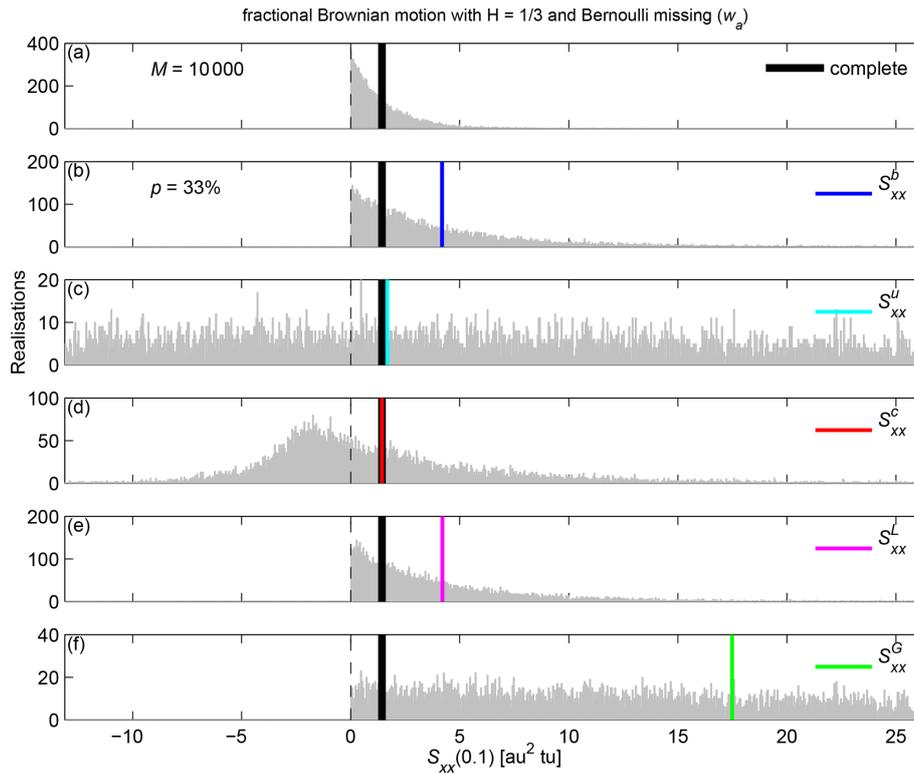
are biased relative to the value obtained without missing samples. In contrast, negative values can be obtained with  $S_{xx}^u$  and  $S_{xx}^c$ , but the average values of these estimates are close to the value obtained without missing samples.  $S_{xx}^u$  has the largest variance among all estimators tested.

Although negative PSD values are not physically meaningful and may appear to be an undesirable feature of the unbiased estimators, they are necessary for the unbiasedness of the estimators. Indeed, if one uses the absolute value of Eq. (13) to convert negative values into positive values (Fig. 6f), as proposed by Gao et al. (2021), the resulting average PSD (Fig. 5, green) are even more biased than those obtained with classical estimators such as  $S_{xx}^b$  and  $S_{xx}^L$ .

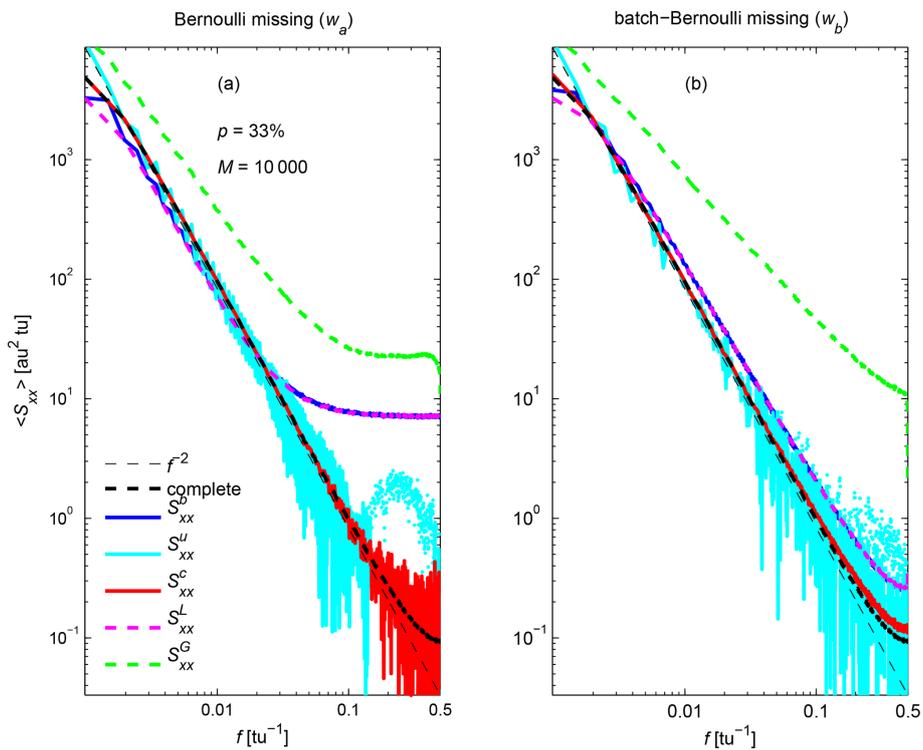
To investigate the sensitivity of the different estimators to the spectral slope, Fig. 7 shows the different estimated PSD for fractional Brownian motion with  $H = 1/2$ . The results are very similar to the case of  $H = 1/3$  (Fig. 5), except that both unbiased estimators  $S_{xx}^u$  and  $S_{xx}^c$  are even noisier at high frequencies. Of course, noise could be reduced by increasing  $M$ , but in practice the number of independent realizations available is usually limited. Further averaging can be performed by averaging PSD estimates within larger frequency bins (i.e., band-averaging). This is illustrated using laboratory turbulence data in the next section.

### 3.2 Case study: decaying turbulence in an active grid-generated air flow

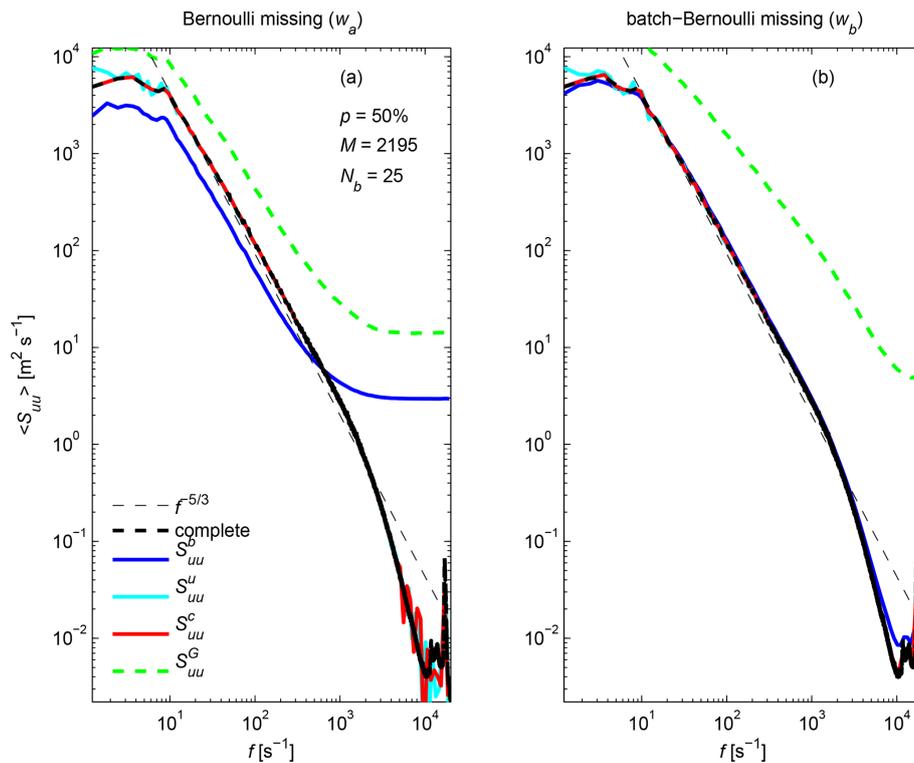
We now turn to an example involving actual data: air flow velocities measured downstream of an active turbulence-generating grid in the return-type Corrsin wind tunnel experiment reported by Kang et al. (2003). The Reynolds number is  $Re_\lambda \approx 720$ . Velocities were measured using an X-wire probe at a sampling rate of 40 kHz mounted at a downstream distance of  $20L$ , where  $L = 0.152$  m is the mesh size of the grid. Following Kang et al. (2003), the 15 min duration velocity record was divided into  $M = 2195$  segments (with 50% overlap) of  $N = 2^{15}$  samples each. Velocity records for each segment were linearly-detrended using the Matlab function `detrend`. While Kang et al. (2003) multiplied each segment by a Bartlett window prior to estimating PSD, here, following Damaschke et al. (2024), we do not apply any window so as to obtain asymptotically-unbiased estimates. The original record is complete (no missing sample), and the PSD of the streamwise velocity component is estimated using the periodogram estimator  $S_{xx}^p$ , averaged over the  $M = 2195$  realizations (Fig. 8, thick black). The spectrum is very similar to that shown in Fig. 3 of Kang et al. (2003), with a spectral slope close to  $-5/3$  over two decades ( $10 \text{ s}^{-1} < f < 10^3 \text{ s}^{-1}$ ).



**Figure 6.** Histograms of  $M = 10^4$  PSD estimates at a particular frequency ( $f = 0.1$ ) of (a) the complete (without missing samples) fractional Brownian motion with  $H = 1/3$ , and of the observed series with  $p = 33\%$  Bernoulli missing samples using (b)  $S_{xx}^b$ , (c)  $S_{xx}^u$ , (d)  $S_{xx}^c$ , (e)  $S_{xx}^L$ , and (f)  $S_{xx}^G$ . Vertical solid lines show the average values.



**Figure 7.** Same as Fig. 5 for fractional Brownian motion with  $H = 1/2$ .



**Figure 8.** Same as Fig. 5 for decaying turbulence in an active grid-generated air flow from Kang et al. (2003), with  $p = 50\%$  missing samples. Spectra for data with missing samples have been band-averaged over  $N_b = 25$  frequency bins per decade.

To investigate the effects of missing samples on PSD estimation, we generate Bernoulli and batch-Bernoulli sampling functions  $w_a$  and  $w_b$ , respectively, with the same algorithm used in the previous section. Here, we test the more stringent case of  $p = 50\%$  missing samples, a large value beyond which time series are usually considered unusable for conventional spectral analysis. Due to the larger fraction of missing samples and the smaller number of realizations than in the previous section, the spectral estimates obtained with  $S_{xx}^u$  and  $S_{xx}^c$  are very noisy (not shown). Therefore, we further averaged all spectral estimates by band-averaging over  $N_b = 25$  frequency bins per decade (Fig. 8). With band-averaging,  $S_{xx}^u$  and  $S_{xx}^c$  have similar performances and recover the  $-5/3$  spectral slope over the two decades. In contrast,  $S_{xx}^G$  is strongly biased for both missing data schemes, while  $S_{xx}^b$  is biased only for the Bernoulli missing data scheme.  $S_{xx}^L$  is not shown because it would have to be multiplied by a factor of  $2 \times 10^4$  to be comparable to the PSD without missing sample, and would then be very similar to  $S_{xx}^b$ .

#### 4 Discussion and conclusions

Wiener–Khinchin’s theorem for finite discrete data states that the periodogram estimator (Eq. 8) is the finite discrete Fourier transform of the circular estimator of the autocor-

relation function of the signal  $R_{xx}^c$  (Eq. 9), while the finite discrete Fourier transform of the standard biased estimator of the autocorrelation function  $R_{xx}^b$  (Eq. 11), usually considered in the literature (e.g. Stoica and Moses, 2005), yields spectral estimates with twice the frequency resolution of the periodogram estimator. Since  $R_{xx}^b$  is a biased estimator of the autocorrelation function, the standard unbiased estimator  $R_{xx}^u$  is sometimes used to estimate the PSD (Eq. 13). This is the estimator commonly used to estimate the PSD of a signal with missing samples, as first proposed by Jones (1962). Here, we instead propose using the unbiased estimator  $R_{xx}^c$ , since the number of data pairs available for each lag ( $N_n$ ) is always greater or equal to the number ( $M_n$ ) available for  $R_{xx}^u$ , which decreases the variance of the PSD estimator. Furthermore, computing  $R_{xx}^c$  using the fast Fourier transform algorithm requires half the memory usage necessary for computing  $R_{xx}^u$ .

The effects of missing samples on the PSD estimation are illustrated in Sect. 3.1 using fractional Brownian motion synthetic data with Hurst parameters  $H = 1/3$  and  $H = 1/2$ , with constant power spectral slopes in log-log space of  $-5/3$  and  $-2$ , respectively, along with synthetic Bernoulli (missing completely at random) and batch-Bernoulli (missing at random) sampling functions with  $p = 33\%$  missing samples. The results show that the standard estimator  $S_{xx}^b$ , as well as the commonly-used Lomb–Scargle estimator  $S_{xx}^L$ , are biased

and do not recover the original spectral slopes throughout the frequency range. In contrast,  $S_{xx}^u$  and  $S_{xx}^c$  are unbiased and recover the original spectral slopes throughout the frequency range, but they have a much larger variance than the biased estimators, especially for  $S_{xx}^u$ , with possible negative values for individual realizations. Although this may be seen as an undesirable property of these estimators, we show that this is actually a necessary feature to obtain unbiased estimates. Indeed, trying to eliminate the negative values by taking their absolute value (as, e.g., Gao et al., 2021, did) leads to strongly biased estimates and distorted spectral slopes.

Similar results are obtained using actual data from a laboratory experiment of decaying turbulence in an active grid-generated air flow (Kang et al., 2003), along with synthetic Bernoulli and batch-Bernoulli sampling functions with  $p = 50\%$  missing samples. To reduce the high level of noise of  $S_{xx}^u$  and  $S_{xx}^c$ , all spectral estimates were further band-averaged over 25 frequency bins per decade. With band-averaging,  $S_{xx}^u$  and  $S_{xx}^c$  have similar performances and recover the  $-5/3$  spectral slope over the same frequency range as for the original data, while the other estimators remain biased.

To obtain an unbiased PSD estimator, Damaschke et al. (2024) propose a combination of three methods, the first using  $S_{xx}^u$ , the second restricting the domain of the autocorrelation function (maximum lag much smaller than  $N - 1$ ) to reduce the variance of the spectral estimates, and the third correcting the autocorrelation estimate after removal of the estimated mean value of the signal. Their first method is almost the same as that proposed here, except that we use  $S_{xx}^c$  instead of  $S_{xx}^u$ . Their second method is similar to that proposed by Efromovich (2014) and provides unbiased PSD estimates for random processes with finite memory, for which the autocorrelation function becomes zero at lags beyond the memory length, provided the maximum lag considered is equal to or longer than the memory length. Here, our synthetic data do not have finite memory, as the autocorrelation function  $S_{xx}^u$  does not tend to zero for large lags (Fig. 2a). Since actual processes do not necessarily have finite memory, as in the turbulence data considered in Sect. 3.2, we do not restrict the domain of the autocorrelation function, accepting the large variance of the resulting PSD estimates. Finally, since actual processes also do not necessarily have zero mean value, the correction proposed by Damaschke et al. (2024) should be used to obtain strictly unbiased PSD estimates. However, even without applying their correction, the PSD estimator proposed here is asymptotically unbiased as  $N \rightarrow \infty$ , so the correction proposed by Damaschke et al. (2024) becomes unnecessary for long time series, such as those considered here ( $N \geq 2^{10}$ ).

As with the periodogram estimator  $S_{xx}^p$ , one problem with  $S_{xx}^c$  is that it assumes that the observed process is periodic with period  $T = N\Delta t$ . However, this is rarely the case for actual processes, including those considered here, and the lack of periodicity leads to spectral leakage of signal fre-

quencies different from the resolvable Fourier frequencies, which contaminates the estimated PSD. For the sake of obtaining asymptotically-unbiased estimates, this problem has been ignored here. Spectral leakage can be reduced by windowing the time series prior to estimating PSD (e.g., Stoica and Moses, 2005) with any of the proposed estimators.

### Appendix A: The Wiener–Khinchin theorem for finite discrete data

Consider a time series of  $N$  samples  $\{x_n\}$ ,  $n = 0, \dots, N - 1$ , sampled at equally spaced time intervals  $\Delta t$  from an infinite, continuous, and stationary random process  $x(t)$ , such that  $x_n = x(n\Delta t)$ , spanning the time period  $T = N\Delta t$ . Its discrete finite Fourier transform is defined by

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi \frac{k}{N} n \Delta t} \quad k = 0, \dots, N - 1, \tag{A1}$$

at discrete Fourier frequencies

$$f_k = \frac{k}{T}. \tag{A2}$$

The periodogram estimator of the auto-spectral density of  $x$  is defined by (Stoica and Moses, 2005)

$$S_{xx,k} = \frac{1}{T} X_k X_k^*. \tag{A3}$$

Using Eq. (A1), Eq. (A3) becomes

$$S_{xx,k} = \frac{\Delta t}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} x_n x_m^* e^{-i2\pi \frac{k}{N} (n-m)}. \tag{A4}$$

Let  $n' = n - m$  replace the variable  $n$  in Eq. (A4). The summation limits for  $n'$  are  $n' = -m$  when  $n = 0$ , and  $n' = N - 1 - m$  when  $n = N - 1$ , modifying the summation domain as illustrated in Fig. A1. Then, Eq. (A4) becomes

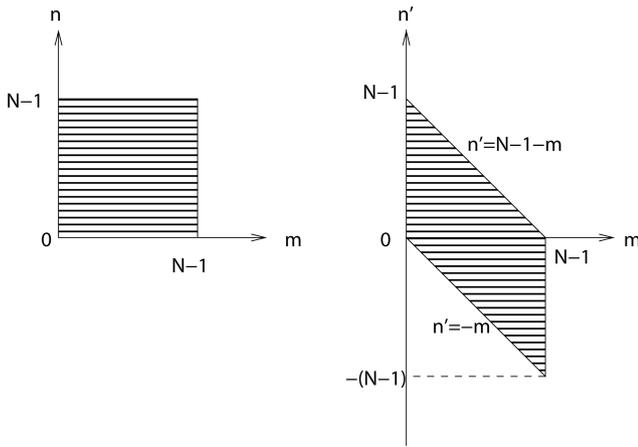
$$S_{xx,k} = \frac{\Delta t}{N} \left( \sum_{n'=-m}^{-1} \sum_{m=0}^{N-1} x_{m+n'} x_m^* e^{-i2\pi \frac{k}{N} n'} + \sum_{n'=0}^{N-1} \sum_{m=0}^{N-1-n'} x_{m+n'} x_m^* e^{-i2\pi \frac{k}{N} n'} \right). \tag{A5}$$

Defining the standard biased estimator of the autocorrelation function of  $x$  as follows (Stoica and Moses, 2005)

$$R_{xx,n'}^b = \begin{cases} \frac{1}{N} \sum_{m=0}^{N-1-n'} x_{m+n'} x_m^* & \text{if } n' \geq 0, \\ \frac{1}{N} \sum_{m=-n'}^{N-1} x_{m+n'} x_m^* & \text{if } n' < 0, \end{cases} \tag{A6}$$

Eq. (A5) becomes

$$S_{xx,k} = \sum_{n'=-m}^{N-1} R_{xx,n'}^b e^{-i2\pi \frac{k}{N} n'} \Delta t \quad k = 0, \dots, N - 1. \tag{A7}$$



**Figure A1.** Change of variables for the derivation of the Wiener-Khinchin theorem (adapted from Bendat and Piersol, 2011).

$R_{xx,n'}^b$  is a biased estimator, since its expected value is

$$E[R_{xx,n'}^b] = \frac{N - |n'|}{N} R_{xx}(n' \Delta t). \tag{A8}$$

An unbiased estimator, called the standard unbiased estimator of the autocorrelation function of  $x$  (Stoica and Moses, 2005), can therefore be defined by

$$R_{xx,n'}^u = \frac{N}{N - |n'|} R_{xx,n'}^b. \tag{A9}$$

The right-hand side of Eq. (A7) looks like the finite Fourier transform of  $R_{xx}^b$ , but the summation is carried over  $2N - 1$  rather than  $N$  samples as in Eq. (A1). Consequently, the finite Fourier transform of  $R_{xx}^b$  yields  $2N - 1$  Fourier frequencies  $f_k = \frac{k}{2T - \Delta t}$ ,  $k = -(N - 1), \dots, N - 1$ .

To estimate the auto-spectral density of  $x$  at the same Fourier frequencies as in Eq. (A2), let  $n = n' + N$  replace  $n'$  for  $n' < 0$  and  $n = n'$  replace  $n'$  for  $n' \geq 0$  in Eq. (A5), which becomes

$$S_{xx,k} = \frac{\Delta t}{N} \left( \sum_{n=1}^{N-1} \sum_{m=N-n}^{N-1} x_{m+n-N} x_m^* e^{-i2\pi \frac{k}{N} n} + \sum_{n=0}^{N-1} \sum_{m=0}^{N-1-n} x_{m+n} x_m^* e^{-i2\pi \frac{k}{N} n} \right). \tag{A10}$$

Defining the circular unbiased estimator of the autocorrelation function of  $x$  as follows (Bendat and Piersol, 2011)

$$R_{xx,n}^c = \begin{cases} \frac{1}{N} \sum_{m=0}^{N-1} x_m x_m^* & \text{for } n = 0, \\ \frac{1}{N} \left( \sum_{m=0}^{N-1-n} x_{m+n} x_m^* + \sum_{m=N-n}^{N-1} x_{m+n-N} x_m^* \right) & \text{for } n = 1, \dots, N - 1, \end{cases} \tag{A11}$$

Eq. (A10) becomes

$$S_{xx,k} = \sum_{n=0}^{N-1} R_{xx,n}^c e^{-i2\pi \frac{k}{N} n} \Delta t \quad k = 0, \dots, N - 1. \tag{A12}$$

$R_{xx,n}^c$  is an unbiased estimator, since the summations in Eq. (A11) are always carried over  $N$  data pairs, and therefore

$$E[R_{xx,n}^c] = R_{xx}(n \Delta t). \tag{A13}$$

The Wiener-Khinchin theorem (Eq. A12) states that the periodogram estimator of the auto-spectral density of  $x$  is the finite Fourier transform of the circular unbiased estimator of the autocorrelation function of  $x$ . It is called circular because, when using the finite discrete Fourier transform, the time series  $x(t)$  is considered periodic of period  $T = N \Delta t$ , then  $x_{m+n-N} = x_{m+n}$  so that the second part of the right-hand side of Eq. (A11) for  $n \geq 1$  represents the remainder of the autocorrelation terms for indices that go beyond  $N - 1$ , for which the beginning of the time series is used again (see Fig. 11.4 in Bendat and Piersol, 2011).

$R_{xx,n}^c$  can also be expressed as

$$R_{xx,n}^c = R_{xx,n}^b + R_{xx,N-n}^{b*}, \tag{A14}$$

or equivalently

$$R_{xx,n}^c = \frac{N - n}{N} R_{xx,n}^u + \frac{n}{N} R_{xx,N-n}^{u*}. \tag{A15}$$

In the limit of infinitely long time series,

$$\lim_{N \rightarrow +\infty} R_{xx,n}^c = \lim_{N \rightarrow +\infty} R_{xx,n}^u. \tag{A16}$$

**Appendix B: Asymptotically-unbiased nonparametric estimation of the cross-spectral density from uniformly-spaced data with missing samples**

The algorithm (implemented with Matlab) to obtain an asymptotically-unbiased estimate of the cross-spectral density of two stationary random processes from uniformly-spaced data with missing samples is detailed below.

Consider two series of  $N$  samples  $\{x_n\}$  and  $\{y_n\}$ ,  $n = 0, \dots, N - 1$ , sampled at equally spaced time (or distance) intervals  $\Delta t$  from two stationary random processes  $x(t)$  and  $y(t)$ , such that  $x_n = x(n \Delta t)$  and  $y_n = y(n \Delta t)$ , spanning the time period (or spatial length)  $T = N \Delta t$ . Suppose that the series have  $N_x < N$  and  $N_y < N$  missing values, respectively.

If the time series are not taken from random processes with zero mean values, an estimate of their mean values should be removed from the time series:

$$\begin{aligned} x &= x - \text{nanmean}(x); \\ y &= y - \text{nanmean}(y). \end{aligned}$$

Next, missing values should be replaced by zeros and sampling functions defined by

$$w_{x,n} = \begin{cases} 1 & \text{if } x_n \text{ is recorded,} \\ 0 & \text{otherwise,} \end{cases} \tag{B1}$$

and

$$w_{y,n} = \begin{cases} 1 & \text{if } y_n \text{ is recorded,} \\ 0 & \text{otherwise,} \end{cases} \quad (\text{B2})$$

which is obtained by:

```
% replace missing samples with zeros
I = find(isnan(x));
x(I) = 0;
J = find(isnan(y));
y(J) = 0;
% sampling functions
wx = ones(size(x));
wx(I) = 0;
wy = ones(size(y));
wy(J) = 0.
```

Generalizing Eq. (8), the discrete finite cross-spectral density of  $x$  and  $y$  is defined by

$$S_{xy,k}^b = \frac{1}{T} X_k Y_k^* \quad k = 0, \dots, N-1, \quad (\text{B3})$$

where  $X$  and  $Y$  are the discrete finite Fourier transforms of  $x$  and  $y$ , respectively. With missing samples replaced with zeros, Eq. (B3) is a biased estimator of the cross-spectral density of  $x$  and  $y$ .

Generalizing the Wiener–Khinchin theorem (Eq. 9), the circular cross-correlation function of  $x$  and  $y$  is defined as the inverse Fourier transform of  $S_{xy}^b$ :

$$R_{xy,n}^{c,b} = \sum_{k=0}^{N-1} S_{xy,k}^b e^{i2\pi \frac{k}{N} n \Delta t} \quad n = 0, \dots, N-1. \quad (\text{B4})$$

With missing samples replaced with zeros, Eq. (B4) is a biased estimator of the circular cross-correlation function of  $x$  and  $y$ , and is obtained by:

$$Rb = \text{ifft}(\text{fft}(x) .* \text{conj}(\text{fft}(y)) / \text{length}(x)).$$

Generalizing Eq. (18) using the Wiener–Khinchin theorem, the number of data pairs available to estimate the circular cross-correlation function of  $x$  and  $y$  is given by:

$$N_{xy,n} = \sum_{k=0}^{N-1} W_{x,k} W_{y,k}^* e^{i2\pi \frac{k}{N} n \Delta t} \quad n = 0, \dots, N-1, \quad (\text{B5})$$

where  $W_x$  and  $W_y$  are the discrete finite Fourier transforms of  $w_x$  and  $w_y$ , respectively.  $N_{xy}$  is obtained by:

$$Npairs = \text{round}(\text{real}(\text{ifft}(\text{fft}(wx) .* \text{conj}(\text{fft}(wy)))));$$

where the “round” and “real” operators are used to obtain real integer values for  $N_{xy,n}$  even with machine precision errors.

Then, the unbiased estimator of the circular cross-correlation function of  $x$  and  $y$  is defined by

$$R_{xy,n}^{c,u} = \frac{N}{N_{xy,n}} R_{xy,n}^{c,b} \quad n = 0, \dots, N-1. \quad (\text{B6})$$

If  $N_{xy,n} = 0$ ,  $R_{xy,n}^{c,u}$  is set to zero.  $R_{xy}^{c,u}$  is obtained by:

```
Ru = Rb * length(x) ./ Npairs;
% if no data pair is available for a lag, set Ru to zero
Ru(find(Npairs == 0)) = 0.
```

Finally, the unbiased estimator of the cross-spectral density of  $x$  and  $y$  is defined by

$$S_{xy,k}^u = \sum_{n=0}^{N-1} R_{xy,n}^{c,u} e^{-i2\pi \frac{k}{N} n \Delta t} \quad k = 0, \dots, N-1, \quad (\text{B7})$$

and is obtained by:

$$Su = \text{fft}(Ru).$$

**Code and data availability.** Matlab software to generate synthetic data, analyse data and reproduce figures is available at <https://doi.org/10.5281/zenodo.18405064> (Chavanne, 2026). Data used and generated for this paper are provided as supplementary material. The decaying turbulence data from Kang et al. (2003) was obtained from <https://engineering.jhu.edu/meneveau/hot-wire-data.html> (last access: 26 September 2025).

**Supplement.** The supplement related to this article is available online at <https://doi.org/10.5194/ascmo-12-59-2026-supplement>.

**Competing interests.** The author has declared that there are no competing interests.

**Disclaimer.** Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

**Acknowledgements.** The author thanks Daniel Bourgault and Christopher Bouillon for reading and commenting on an initial draft of the manuscript, as well as two anonymous reviewers and Associate Editor Dan Cooley for their constructive comments and suggestions.

**Financial support.** This research has been supported by the Natural Sciences and Engineering Research Council of Canada (grant no. RGPIN-2024-04826).

**Review statement.** This paper was edited by Dan Cooley and reviewed by two anonymous referees.

## References

- Abry, P. and Sellan, F.: The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation, *Appl. Comput. Harmon. Anal.*, 3, 377–383, 1996.
- Babu, P. and Stoica, P.: Spectral analysis of nonuniformly sampled data – a review, *Digit. Sig. Process.*, 20, 359–378, 2010.
- Bartlett, M. S.: Smoothing periodograms from time-series with continuous spectra, *Nature*, 161, 686–687, 1948.
- Bendat, J. S. and Piersol, A. G.: *Random data: analysis and measurement procedures*, John Wiley & Sons, Hoboken, NJ, 2011.
- Blackman, R. B. and Tukey, J. W.: The measurement of power spectra from the point of view of communications engineering – Part I, *Bell Syst. Tech. J.*, 37, 185–282, 1958.
- Bloomfield, P.: Spectral Analysis with Randomly Missing Observations, *J. Roy. Stat. Soc.*, 32, 369–380, 1970.
- Boyd, J. P.: The energy spectrum of fronts: time evolution of shocks in Burgers' equation, *J. Atmos. Sci.*, 49, 128–139, 1992.
- Charney, J. G.: Geostrophic Turbulence, *J. Atmos. Sci.*, 28, 1087–1095, 1971.
- Chavanne, C.: Asymptotically-unbiased nonparametric estimation of the power spectral density from uniformly-spaced data with missing samples, Zenodo [data set], <https://doi.org/10.5281/zenodo.18405064>, 2026.
- Clinger, W. and Van Ness, J. W.: On unequally spaced time points in time series, *Ann. Stat.*, 736–745, 1976.
- Corrsin, S.: On the spectrum of isotropic temperature fluctuations in an isotropic turbulence, *J. Appl. Phys.*, 22, 469–473, 1951.
- Damaschke, N., Kühn, V., and Nobach, H.: Bias-free estimation of the covariance function and the power spectral density from data with missing samples including extended data gaps, *EURASIP J. Adv. Sig. Process.*, 2024, 17, <https://doi.org/10.1186/s13634-024-01108-4>, 2024.
- Efromovich, S.: Efficient non-parametric estimation of the spectral density in the presence of missing observations, *J. Time Ser. Anal.*, 35, 407–427, 2014.
- Efromovich, S.: Missing and modified data in nonparametric estimation: with R examples, Chapman and Hall/CRC, <https://doi.org/10.1201/9781315166384>, 2018.
- Gao, Y., Schmitt, F. G., Hu, J., and Huang, Y.: Scaling Analysis of the China France Oceanography Satellite Along-Track Wind and Wave Data, *J. Geophys. Res.-Oceans*, 126, e2020JC017119, <https://doi.org/10.1029/2020JC017119>, 2021.
- Gaster, M. and Roberts, J.: Spectral analysis of randomly sampled signals, *IMA J. Appl. Math.*, 15, 195–216, 1975.
- Grant, H., Stewart, R., and Moilliet, A.: Turbulence spectra from a tidal channel, *J. Fluid Mech.*, 12, 241–268, 1962.
- Jones, R. H.: Spectral analysis with regularly missed observations, *Ann. Math. Stat.*, 33, 455–461, 1962.
- Jones, R. H.: Spectrum estimation with missing observations, *Ann. Inst. Stat. Math.*, 23, 387–398, 1971.
- Kang, H. S., Chester, S., and Meneveau, C.: Decaying turbulence in an active-grid-generated flow and comparisons with large-eddy simulation, *J. Fluid Mech.*, 480, 129–160, 2003.
- Khinchine, A.: Korrelationstheorie der stationären stochastischen Prozesse, *Math. Ann.*, 109, 604–615, 1934.
- Kolmogorov, A.: The local structure of turbulence in incompressible viscous fluid for very large Reynolds number, *Dokl. Akad. Nauk SSSR*, 30, 299–303, 1941.
- Kuleshov, E. and Grudin, B.: Spectral density of a fractional Brownian process, *Optoelectronics, Instrum. Data Process.*, 49, 228–233, 2013.
- Li, X., Huang, Y., Wang, G., and Zheng, X.: High-frequency observation during sand and dust storms at the Qingtu Lake Observatory, *Earth Syst. Sci. Data*, 13, 5819–5830, <https://doi.org/10.5194/essd-13-5819-2021>, 2021.
- Little, R. J. and Rubin, D. B.: *Statistical analysis with missing data*, John Wiley & Sons, ISBN 0470526793, 2019.
- Lomb, N. R.: Least-squares frequency analysis of unequally spaced data, *Astrophys. Space Sci.*, 39, 447–462, 1976.
- Ma, Y., Cheng, W., Huang, S., Schmitt, F. G., Lin, X., and Huang, Y.: Hidden turbulence in van Gogh's *The Starry Night*, *Phys. Fluids*, 36, 095140, <https://doi.org/10.1063/5.0213627>, 2024.
- Nastrom, G. and Gage, K.: A Climatology of Atmospheric Wavenumber Spectra of Wind and Temperature Observed by Commercial Aircraft, *J. Atmos. Sci.*, 42, 950–960, 1985.
- Obukhov, A.: Spectral energy distribution in a turbulent flow, *Izv. Akad. Nauk. SSSR, Ser. Geogr. i. Geofiz.*, 5, 453–466, 1941.
- Obukhov, A. M.: Structure of the temperature field in a turbulent flow, *Izv. Akad. Nauk SSSR, Ser. Geogr. Geofiz.*, 13, 58–69, 1949.
- Parzen, E.: On spectral analysis with missing observations and amplitude modulation, *Sankhyā: The Indian Journal of Statistics, Series A*, 25, 383–392, 1963.
- Plantier, G., Moreau, S., Simon, L., Valière, J.-C., Le Duff, A., and Bailliet, H.: Nonparametric spectral analysis of wideband spectrum with missing data via sample-and-hold interpolation and deconvolution, *Digit. Sig. Process.*, 22, 994–1004, 2012.
- Robache, K., Schmitt, F. G., and Huang, Y.: Scaling and intermittent properties of oceanic and atmospheric  $p\text{CO}_2$  time series and their difference in a turbulence framework, *Nonlin. Processes Geophys.*, 32, 35–49, <https://doi.org/10.5194/npg-32-35-2025>, 2025.
- Rubin, D. B.: Inference and missing data, *Biometrika*, 63, 581–592, 1976.
- Scargle, J. D.: Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data, *Astrophys. J.*, 263, 835–853, 1982.
- Scheinok, P. A.: Spectral analysis with randomly missed observations: the binomial case, *Ann. Math. Stat.*, 36, 971–977, 1965.
- Schuster, A.: On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena, *Terr. Magnet.*, 3, 13–41, 1898.
- Stoica, P. and Moses, R.: *Spectral analysis of signals*, Prentice Hall, Upper Saddle River, NJ, ISBN 0-13-113956-8, 2005.
- Welch, P. D.: The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, *IEEE T. Audio Electroacoust.*, 15, 70–73, 1967.
- Wiener, N.: Generalized harmonic analysis, *Acta Math.*, 55, 117–258, 1930.