sensors

MDPI

Article

Multivariate Frequency and Amplitude Estimation for Unevenly Sampled Data Using and Extending the Lomb–Scargle Method

Martin Seilmayer [1,*,†], Thomas Wondrak [2,†] and Ferran Garcia [3,†]

[1] Staatliche Studienakademie Bautzen, Duale Hochschule Sachsen, Löbauer Strasse 1, 02625 Bautzen, Germany
[2] Department of Magnetohydrodynamics, Helmholtz-Zentrum Dresden-Rossendorf, Bautzner Landstraße 400, D-01328 Dresden, Germany; t.wondrak@hzdr.de
[3] Department of Fluid Mechanics, Universitat Politècnica de Catalunya-BarcelonaTech, Av. Víctor Balaguer 1, Vilanova i la Geltrú, 08800 Barcelona, Spain; fernando.garcia-gonzalez@upc.edu
* Correspondence: martin.seilmayer@dhsn.de; Tel.: +49-3591-353-214
† These authors contributed equally to this work.

Abstract

The Lomb–Scargle method (LSM) constitutes a robust method for frequency and amplitude estimation in cases where data exhibit irregular or sparse sampling. Conventional spectral analysis techniques, such as the discrete Fourier transform (FT) and wavelet transform, rely on orthogonal mode decomposition and are inherently constrained when applied to non-equidistant or fragmented datasets, leading to significant estimation biases. The classical LSM, originally formulated for univariate time series, provides a statistical estimator that does not assume a Fourier series representation. In this work, we extend the LSM to multivariate datasets by redefining the shifting parameter τ to preserve the orthogonality of trigonometric basis functions in $\mathbb{R}^n$. This generalization enables simultaneous estimation of the frequency, phase, and amplitude vectors while maintaining the statistical advantages of the LSM, including consistency and noise robustness. We demonstrate its application to solar activity data, where sunspots serve as intrinsic markers of the solar dynamo process. These observations constitute a randomly sampled two-dimensional binary dataset, whose characteristic frequencies are identified and compared with the results of solar research. Additionally, the proposed method is applied to an ultrasound velocity profile measurement setup, yielding a three-dimensional velocity dataset with correlated missing values and significant temporal jitter. We derive confidence intervals for parameter estimation and conduct a comparative analysis with FT-based approaches.

Keywords: multivariate data analysis; spectroscopic; uneven sampling

check for updates

Academic Editor: Anastasios Doulamis

Received: 9 September 2025
Revised: 13 October 2025
Accepted: 17 October 2025
Published: 23 October 2025

Citation: Seilmayer, M.; Wondrak, T.; Garcia, F. Multivariate Frequency and Amplitude Estimation for Unevenly Sampled Data Using and Extending the Lomb–Scargle Method. *Sensors* **2025**, *25*, 6535. https://doi.org/10.3390/s25216535

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In numerous signal processing applications, the characterization of a physical process via its power spectrum or its amplitude and phase spectra is of fundamental importance. Typically, the recorded signal is sampled at uniform intervals over a finite observation period and stored as discrete sampled data. Spectral analysis serves multiple purposes, including the identification of characteristic frequencies for process recognition, the selection or attenuation of specific spectral components, and the precise estimation of the amplitude and phase at specified frequencies. Conventional approaches predominantly rely on the discrete Fourier transform (DFT) or its computationally efficient companion, the fast Fourier transform (FFT), both of which decompose the data into Fourier series coefficients from which the power spectrum can be derived [1,2]. These transformations are

inherently invertible, computationally efficient, and straightforward to implement. However, their applicability is constrained by the requirement of equidistant and complete [1] sampling—limitations that pose challenges for data acquisition, in general. In practical applications, missing or corrupted data are often handled by substituting zero values, a common but problematic approach. While not the intention, zero values provide some sort of information and do not mean "not acquired". Subsequently, zero values introduce a systematic error that leads to increased inaccuracies in frequency and amplitude estimation, ultimately affecting the reliability of parameter estimation. While the DFT naturally extends to higher dimensions—facilitating applications such as image reconstruction in magnetic resonance imaging [3], image filtering [4], and higher-order spectral filtering in spatiotemporal domains [5]—its performance deteriorates in the presence of non-uniform sampling. In ultrafast nuclear magnetic resonance (NMR) spectroscopy, non-uniform sampling enables the substantial acceleration of chemical analysis [6,7], albeit at the cost of increased missing data, reduced sensitivity, and signal attenuation. These limitations are inherent to DFT-based and similar techniques, as they exhibit reduced accuracy when applied to incomplete datasets with data gaps [8]. Although alternative methods, such as the "non-uniform Fourier transform" [9–11], attempt to address these issues, they remain reliant on classical orthogonal mode decomposition (e.g., FFT, DFT, wavelet transform) and thus inherit systematic errors, as discussed in Section 2.2. The recent publication by Wei and Yang [12] discusses the associated theory, complexity, and applications. They also compare different realizations of the non-uniform FFT.

The application of techniques based on Fourier series to signals with non-equidistant sampling is difficult, since it usually requires zero padding (replacing missing values with zeros) on a grid and, in the general case, resampling or interpolation of the data to a regular grid [13,14]. The effect of a gap-filling method was investigated extensively by Munteanu et al. [8]. They concluded that without a corrective measure, errors such as the amplitude minimization—which depends on the total gap density—cannot be avoided when FFT or DFT is applied. On the other hand, interpolation could be utilized to shift non-uniformly sampled data to a regular grid, such that the new interpolated data points comprise the signal information and a projection of the accompanying noise. In the general case, such distortions or noise are not band-limited, leading to biased estimates that consist of the local and aliased errors (e.g., noise, outliers, missing values) of the surrounding data.

Astrophysical data, for example, are affected by gaps, missing values, and uneven sampling. When ground-based radio telescopes are exploring areas in space, there may exist time intervals in which the antenna is not pointing towards the object of interest due to the rotation of the Earth. The result is an incomplete dataset. Non-uniform (random) sampling emerges; for example, the irregular appearance of objects, such as sunspots, is measured as a binary quality depending on the time and location. Section 5.3 provides a working example, which analyzes the frequencies and periods in latitude and time from the two-dimensional dataset of sunspot observations. Another technical example for random sampling (with highly variable sampling frequency) is asynchronous data acquisition in large sensor networks, which are used in Smart Homes, Industry 4.0, and automated driving; see Geneva et al. [15], Cadena et al. [16], Sudars [17]. Here, the time series provides missing values and data gaps originating from time periods in which either strong noise prevents the measurement or the source of the signal to be measured is not within the range of the sensor.

The Lomb–Scargle method (LSM) was developed especially for ground-based non-uniformly sampled one-dimensional data, from which the amplitude spectrum is calculated [18,19]. The main advantage of this method is that it can directly estimate the spectrum without the iterative optimization of trigonometric models, as discussed in Section 2.1. A

fast version for one-dimensional signals was presented by Townsend [20] and Leroy [21]. As discussed in Mathias et al. [22], the computational complexity of estimating such a periodogram is significant. Consequently, even when using accelerated algorithms, refer to [20], the calculation remains a computationally intensive problem with a complexity of $\mathcal{O}(M \cdot N)$, if N equals the number of samples and M the number of processed frequencies. Additionally, Zechmeister and Kürster [23] introduced a generalized Lomb–Scargle (GLS) method, which incorporates individual measurement errors for each sample and an extra constant term per frequency. Initially, all developments presented here will be based on the traditional Lomb–Scargle method. However, other related methods like GLS can be extended to multivariate analyses as well with the subsequently presented method. An implementation can be found in the corresponding R-Package (Version 2.0 from 2021) [24].

An extension of the LSM to two- or three-dimensional time series has not yet been presented. Especially for large multivariate datasets, this direct approach would be comparably faster than the present iterative procedures proposed in (Babu and Stoica [25], Chapter 9). The advantages of multivariate LSM are demonstrated in Section 5 by means of a 3D ultrasound flow profile measurement and the 2D analysis of sunspot time series data.

In the case of the analysis of flow profile measurements, which were obtained in an experiment investigating the magnetorotational instability in liquid metals [26], this method would be highly desirable. Due to the complex experimental setup and the weak signal-to-noise ratio, the flow profile measurements contain several time intervals in which distortions are dominant; see Seilmayer et al. [27]. These time intervals have to be rejected, leading to a time series with invalid (missing) data points, from which the multivariate version of the LSM is able to determine the parameters of the characteristic traveling wave.

In order to demonstrate the method and to motivate the basic idea of the multivariate version of the LSM, we compare it—in terms of the necessary conditions and error (noise) behavior—with the traditional orthogonal mode decomposition (OMD) with trigonometric basis functions, which is the essence of the classical Fourier transform. The LSM better fits the conditions of an arbitrary finite length of the sampling series in comparison with the DFT because the LSM reduces the error in the model parameter estimation. In contrast to the traditional approach, which was derived from a statistical point of view (see the appendix of Scargle [19]), the presented method is deduced from a technical point of view and focuses on its application. This leads to a slight change in the scaling of the model parameters, which is discussed in Section 4.2. However, the introduced procedure includes all the benefits, such as arbitrary sampling, fragmented data, and good noise rejection.

The starting point of this paper is the analysis of a continuous 1D signal $s : \mathbb{R} \to \mathbb{R}$, which is composed of an arbitrary and finite set of individual frequency components $\omega_i, 0 \leq i \leq M$. Without loss of generality, the band limitation is assumed, and there exists an upper maximum frequency $\omega_{\max}$ with $\omega_i < \omega_{\max}, 0 \leq i \leq M$, which mimics an intrinsic low pass filter, characteristic of the measurement device. Since the measurement time is finite, the value of s is only known in the time interval $[0, T]$ with $T \in \mathbb{R}$ and $T > 0$. The m-dimensional extension of this signal is $S : \mathbb{R}^m \to \mathbb{R}$. If such a continuous signal is sampled, the pair $(\hat{s}_i, t_i) \in \mathbb{R}^2, 0 \leq i \leq N-1$ represents the measured 1D value and the corresponding instant in time, whereas the pair $(\hat{S}_i, \vec{t}_i) \in \mathbb{R}^{m+1}, 0 \leq i \leq N-1$ represents a measured value and the m-dimensional location ($\vec{t}_i$ space/time) for the i-th sample (N is the number of samples). All the methods presented herein were implemented in a package written in R and are published on CRAN [24]. The reader will find more elaborate examples in the package and in the Supplementary Materials.

2. Mathematical Model and Comparison of OMD and LSM

In order to delineate the differences between the trigonometric OMD and LSM, we start with the basic model of a periodic signal as a sum of the signals of different frequencies $\omega_k, 0 \leq k \leq M$, with the corresponding amplitude $A_k \in \mathbb{R}$ and phase shift $\varphi_k \in \mathbb{R}$. The corresponding trigonometric model function

$$y(t) = \sum_{k=0}^{M} A_k \cos(\omega_k t + \varphi_k) = \sum_{k=0}^{M} \Big(a_k \cos(\omega_k t) + b_k \sin(\omega_k t) \Big) \tag{1}$$

describes an infinite, stationary, and steady process $y : \mathbb{R} \to \mathbb{R}$, with the coefficients $a_k, b_k \in \mathbb{R}$ for the defined frequency ω_k and the identities $A_k = \sqrt{a_k^2 + b_k^2}$, as well as $\varphi_k = \tan^{-1}(b_k/a_k)$. Furthermore, the trigonometric model above consists of an arbitrary number $M \in \mathbb{N}_0$ of frequency components. If the given signal $s(t)$ is described by the defined model from Equation (1), the model misfit ϵ is given by

$$\epsilon(t) = s(t) - y(t). \tag{2}$$

Thus, the signal can now be described by inserting Equation (1) into Equation (2) and defining the misfit ϵ_k for each discrete frequency k with $\epsilon(t) = \sum_k \epsilon_k(t)$ as follows:

$$s(t) = \sum_{k=0}^{M} \Big(a_k \cos(\omega_k t) + b_k \sin(\omega_k t) + \epsilon_k(t) \Big). \tag{3}$$

The defined misfit originates from the measurement uncertainties or parametric errors from a_k, b_k. In the general case, $\epsilon(t)$ can be any function or distribution. The challenge is to precisely determine the model parameters a_k and b_k for a given signal $s(t)$ achieving minimal $\epsilon(t)$. This can be accomplished by one of the three methods described in the following sections: (i) least-square fit; (ii) orthogonal mode decomposition; (iii) Lomb–Scargle method (LSM).

2.1. Least-Square Fit

The optimal fit is reached by least-square fitting, resulting in a minimum ϵ, which was shown by Mathias et al. [22], Barning [28]. Since such procedures are iterative, the convergence of the algorithm might need a large number of function evaluations of Equation (3). Therefore, a direct version such as the LSM is preferable. The LSM becomes equivalent to a least-square fit of a sinusoidal model [18,28].

2.2. Trigonometric OMD

Generally, two functions $f, g : \mathbb{R} \to \mathbb{R}$ are said to be orthogonal on the interval $[a, b] \subset \mathbb{R}$, if the following condition holds; refer to Weisstein [29]:

$$\int_a^b f(x) g(x) \, \mathrm{d}x = 0. \tag{4}$$

By selecting $f(x) = \sin(x)$ and $g(x) = \cos(x)$, the integration leads, by exploiting the identity $\cos(x)\sin(x) = \frac{1}{2}\sin(2x)$, to $\frac{1}{4}\Big(\cos(2a) - \cos(2b)\Big) = 0$, which is only zero if $\cos(2a) = \cos(2b)$. This is true for any $a \in \mathbb{R}$ if the length of the interval $[a, b]$ is a multiple of the period 2π, such that $2b = 2a + 2\pi k$ with $k \in \mathbb{N}^+$. It is interesting to note that this interval can be shortened to one half of the period if $\cos(2a)$ is zero. By exploiting this feature of trigonometric functions, the individual model coefficients are calculated by

multiplying the sine or the cosine to the measured data $s(t)$ and integrating over all times as shown by (Cohen [1], Chapter 15):

$$a_k = \frac{2}{T}\int_{-\infty}^{\infty} s(t)\cos(\omega_k t)\,\mathrm{d}t \tag{5}$$

$$b_k = \frac{2}{T}\int_{-\infty}^{\infty} s(t)\sin(\omega_k t)\,\mathrm{d}t. \tag{6}$$

For a measured signal, the integration can only be performed over the interval $[0, T]$. It is obvious that an error is introduced if $T \neq 2\pi n$. In order to investigate the properties of the finite integration, we focus on the cosine term (Equation (5)), since the analysis of the sine term (Equation (6)) is similar. By setting the integral boundaries to the finite time interval and expressing the signal using Equation (3), the integral in Equation (5) can be written as

$$\int_0^T s(t)\cos(\omega_k t)dt = \int_0^T \Big(a_k \cos^2(\omega_k t) + b_k \sin(\omega_k t)\cos(\omega_k t) + \epsilon_k(t)\cos(\omega_k t)\Big)\,\mathrm{d}t. \tag{7}$$

Using the identities $\cos^2(x) = \frac{1}{2}(1+\cos(2x))$ and $\sin(x)\cos(x) = \frac{1}{2}\sin(2x)$, Equation (7) becomes

$$\frac{2}{T}\int_0^T s(t)\cos(\omega_k t)\,\mathrm{d}t = a_k\Bigg(1 + \underbrace{\frac{1}{T}\int_0^T \Big(\cos(2\omega_k t) + \frac{b_k}{a_k}\sin(2\omega_k t)\Big)\mathrm{d}t}_{\text{truncation error } \epsilon_k^{\mathrm{T}}}\Bigg) + \underbrace{\frac{2}{T}\int_0^T \epsilon_k(t)\cos(\omega_k t)\,\mathrm{d}t}_{\text{random error } \epsilon_k^{\mathrm{FS}}}. \tag{8}$$

The equation above indicates that the coefficient a_k is effected by two errors ϵ_k^{T} (truncation) and ϵ_k^{FS} (random), which might be nonzero for an arbitrary T. Thus, if both errors are neglected, the integral on the left hand side turns into an approximation of a_k. Taking the consideration about the orthogonality described in Equation (4) into account, ϵ_k^{T} only becomes zero if the integration time T is an integer multiple of π/ω_k. This means for the Fourier series that a given integration time T (observation window) defines the lowest allowed frequency ω_0. Furthermore, only multiples of this fundamental frequency $\omega_k = k\omega_0$ with $k \in \mathbb{N}$ are allowed in the Fourier series because $\epsilon_k^{\mathrm{T}} = 0$ in this case.

The technical realization of OMD is called "quadrature demodulation", which is the discrete version of Equation (8) by replacing the integral over $s(t)$ into a sum over the discrete sampled values $\hat{s}_n, 0 \leq n \leq N-1$ and setting the truncation error to zero. For a time-discrete signal with N samples and equidistant sampling with a constant sampling period T_s, and $t_n = nT_s$, the parameter a_k (corresponding to the frequency ω_k) is approximated by

$$a_k \approx \frac{2}{N}\sum_{n=0}^{N-1} \hat{s}_n \cos(\omega_k t_n). \tag{9}$$

According to the previous considerations, the truncation error ϵ_k^{T} is larger than zero if the measurement range $T = NT_s$ is not exactly a multiple of π/ω_k, as shown in Figure 1.

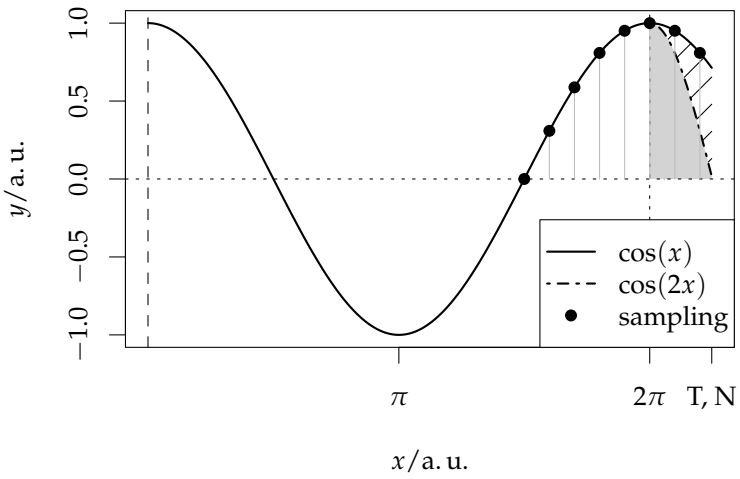


Figure 1. Example of the truncation error of a signal with the time period of 2π, which is generated by the last two samples. The value is indicated by the gray shaded area. The axes are scaled in arbitrary units (a.u.).

In this example, a signal with a time period of $t = 2\pi$ is sampled with $N = 23$, $T_s = 2\pi/20$. The total integration time is $T = 2\pi + 1/5\pi$, slightly longer than the period of the signal, which is indicated by the two additional sampling points after $t = 2\pi$. The truncation error equals the gray shaded area.

Generally speaking, the sampling error in the discrete version can be estimated by

$$\frac{2}{N}\sum_{n=0}^{N-1}\hat{s}_n\cos(\omega_k t_n) \approx a_k\Big(1+\underbrace{\frac{\Delta\varphi}{T}}_{\epsilon_k^{\mathrm{T}}}\Big)\pm\underbrace{\Phi_{1-\alpha}\frac{2\sigma_0}{\sqrt{N}}}_{\epsilon_k^{\mathrm{FS}}}, \tag{10}$$

where $\Delta\varphi = \min_i(T - \pi i/\omega_k)$, $i \in \mathbb{N}$. The random error is modeled by the α-quantile of the sampling error distribution $\Phi_{1-\alpha}$ related to the underlying process with its specific standard deviation σ_0 (e.g., JCGM [30]). This estimates the confidence interval of a_k and b_k, respectively (see the Supplementary Material for details).

The above considerations lead to the following four statements, which are derived in detail in Section 2 of the Supplementary Materials: (i) the maximum absolute value of the truncation error is bounded $\left|\epsilon_k^{\mathrm{T}}\right| \leq 0.2$ if $T \gtrsim \pi/\omega_k$, which is consistent with the experimental findings from Thompson and Tree [31]; (ii) ϵ_k^{T} is independent from the total number of samples N for a constant time T, which makes the OMD a non-consistent estimator for model parameters a_k and b_k; (iii) by increasing the sample rate, the random error decreases by $\mathcal{O}(N^{-0.5})$ but scales with twice the standard deviation of the noise distribution; and (iv) the DFT can be derived from Equation (10) via a restriction to equidistant sampling with constant T_s.

Concluding, the truncation error—which is an intrinsic feature of the OMD—produces a systematic deviation from the true value depending on the difference between the sampling time interval and the corresponding period of the frequency of interest. In contrast, the random error diminishes with an increasing sampling rate $T_\mathrm{s}^{-1} = N/T$. A detailed discussion on this topic can be found in Thompson and Tree [31], Jerri [32], Shannon [33]. In the case of multivariate and randomly sampled data, the recent work of Al-Ani and Tarczynski [34] suggests two calculation schemes for estimating the Fourier transform in the general case. The continuous time Fourier transform estimation (similar to Equation (9)) takes samples with arbitrary spacing as the general approach, in contrast to the second proposed discrete time Fourier transform ($t_n = nT_\mathrm{s}$ and $\omega_k = k\omega_0$) estimation scheme. Here, the data are projected onto a regular grid, which may provide the locations of missing

data. Both schemes consider the sampling pattern t_n and its power spectrum distribution function $p(t_n)$. However, even for such sophisticated methods, the main conceptual drawbacks (see points (i) and (ii)) remain, which motivates the subsequent Lomb–Scargle method as a non-OMD method and its extension to multivariate data.

2.3. Lomb–Scargle Method

As described in the previous section, the disadvantage of the OMD is that the cosine and sine are not orthogonal on arbitrary intervals. By introducing an additional parameter $\tau \in \mathbb{R}$ into Equation (4), we show that

$$\int_a^b \cos(x-\tau)\sin(x-\tau)\,\mathrm{d}x = 0 \tag{11}$$

holds for arbitrary intervals $[a, b]$. By utilizing the trigonometric identities $\cos^2(\phi) - \sin^2(\phi) = \cos(2\phi)$ and $\cos(\phi)\sin(\phi) = \frac{1}{2}\sin(2\phi)$, in order to remove the differences in the arguments, the integral can be transformed to

$$\int_a^b \left(\frac{1}{2}\sin(2x)\cos(2\tau) - \frac{1}{2}\sin(2\tau)\cos(2x)\right)\mathrm{d}x = 0.$$

If this integral is set to zero, the following condition holds

$$\cos(2\tau)\int_a^b \sin(2x)\,\mathrm{d}x = \sin(2\tau)\int_a^b \cos(2x)\,\mathrm{d}x,$$

which can be transformed to

$$\frac{\int_a^b \sin(2x)\,\mathrm{d}x}{\int_a^b \cos(2x)\,\mathrm{d}x} = \tan(2\tau). \tag{12}$$

Therefore, from Equation (12), the parameter τ verifying Equation (11) may be found. In the case of equidistant sampling, the value of τ can be directly calculated from the integration boundaries by

$$\tau = \frac{b-a}{2}.$$

Based on this general consideration and in order to remove the truncation error, for each frequency ω_k, the time shifting parameter $\tau_k \in \mathbb{R}, 0 \le k \le M$ was introduced by Lomb and Scargle [18,19] into the model given in Equation (1); so,

$$s(t) = \sum_{k=0}^{M}\left(a_k\cos\big(\omega_k(t-\tau_k)\big) + b_k\sin\big(\omega_k(t-\tau_k)\big) + \epsilon_k(t)\right). \tag{13}$$

The parameter τ_k can be calculated by

$$\tan(2\omega_k\tau_k) = \frac{\int_a^b \sin(2\omega_k t)\,\mathrm{d}t}{\int_a^b \cos(2\omega_k t)\,\mathrm{d}t}, \tag{14}$$

similar to Equation (12). For the time-discrete version, the integrals transform into a sum, resulting in

$$\tan(2\omega_k\tau_k) = \frac{\sum_{n=0}^{N-1}\sin(2\omega_k t_n)}{\sum_{n=0}^{N-1}\cos(2\omega_k t_n)}. \tag{15}$$

The parameters a_k can be determined, beginning with Equation (7), but by factorizing them by $\cos^2(x)$, instead of expanding $\cos^2(x)$ to $\frac{1}{2}(1+\cos(2x))$ as performed for Equation (8).

In the following, the method is delineated for the discrete set $\hat{s}_i$, since it is applied to sampled data. These data are multiplied by $\cos(\omega_k(t_n - \tau_k))$, resulting in the following equation for a single frequency ω_k:

$$\sum_{n=0}^{N-1} \hat{s}_n \cos(\phi_k) = a_k \sum_{n=0}^{N-1} \cos^2(\phi_k) + b_k \underbrace{\sum_{n=0}^{N-1} \sin(\phi_k)\cos(\phi_k)}_{\text{truncation error } \epsilon_k^{\mathrm{T}}} + \underbrace{\sum_{n=0}^{N-1} \epsilon_k(t_n)\cos(\phi_k)}_{\text{random error } \epsilon_k^{\mathrm{s}}}, \tag{16}$$

where $\phi_k = \omega_k(t_n - \tau_k)$. The sum of the terms $\cos(\phi_k)\sin(\phi_k)$ vanishes because of the proper selection of τ_k, according to Equation (15). The sum of the terms $\epsilon_k(t_n)\cos(\phi_k)$ describes the modulated noise distribution function. The value of the parameters a_k can be directly obtained by dividing by $\sum_{n=0}^{N-1}\cos^2(\phi_k)$. This gives

$$\frac{\sum_{n=0}^{N-1} \hat{s}_n \cos(\phi_k)}{\sum_{n=0}^{N-1}\cos^2(\phi_k)} = a_k + \underbrace{\frac{\sum_{n=0}^{N-1} \epsilon_k(t_n)\cos(\phi_k)}{\sum_{n=0}^{N-1}\cos^2(\phi_k)}}_{\epsilon_k^{\mathrm{LS}}}. \tag{17}$$

The residual error term on the right-hand side, ϵ_k^{LS}, consists of a random distributed part divided by a sum over the square of the cosine. In contrast to the OMD, the estimation error of the parameter a_k only depends on noise and is independent from the realization of sampling.

The next step is to determine ϵ_k^{LS} in terms of a confidence interval with respect to α-quantile of the sampling error distribution $\Phi_{1-\alpha}$, as performed for the OMD in Equation (10). Given a normal distributed error function with zero expectation value $\mu = 0$ and a standard deviation σ independent from time ($\epsilon_k \leftrightarrow \mathcal{N}(0, \sigma)$), the error can be factorized (Parzen [35], Theorem 4A, p. 90), leading to

$$\epsilon_k^{\mathrm{LS}} = \Phi_{1-\alpha}\frac{\sigma_0}{\sqrt{N}}\frac{\sum_{n=0}^{N-1}\cos(\phi_k)}{\sum_{n=0}^{N-1}\cos^2(\phi_k)} \leq \Phi_{1-\alpha}\frac{\sigma_0}{\sqrt{N}}e_{\max},$$

estimating the most probable limits of ϵ_k^{LS}. In the above equation, $e_{\max}$ does not depend on the frequency or the shifting parameter and is defined as

$$e_{\max} = \max_{\beta\in[0,2\pi]}\left[\frac{\int_0^\beta \cos(\phi)\,\mathrm{d}\phi}{\int_0^\beta \cos^2(\phi)\,\mathrm{d}\phi}\right] = \max_{\beta\in[0,2\pi]}\left[\frac{4\sin(\beta)}{(2\beta + \sin(2\beta))}\right] = \frac{4}{\pi}.$$

The final parameter estimation gives

$$a_k = \frac{\sum_{n=0}^{N-1} \hat{s}_n \cos(\omega_k(t_n - \tau_k))}{\sum_{n=0}^{N-1}\cos^2(\omega_k(t_n - \tau_k))} \pm \underbrace{\frac{4}{\pi}\Phi_{1-\alpha}\frac{\sigma_0}{\sqrt{N}}}_{\Delta a_k}, \tag{18}$$

where the confidence interval $\Delta a_k = \frac{4}{\pi}\Phi_{1-\alpha}\frac{\sigma_0}{\sqrt{N}}$ converges to zero when increasing the number of samples N, for a fixed time interval. This property qualifies the LSM as a consistent estimator of the amplitude and phase [22]. In addition, the confidence interval Δa_k for the LSM is smaller than the value given in Equation (10) for the OMD (this also occurs for the confidence interval of the parameters b_k, $\Delta b_k = \Delta a_k$). We notice that the definition of a_k given in Equation (18) is slightly different from the original definition given by Lomb [18], but both definitions converge for large N (see the Supplementary Materials).

The confidence interval for the amplitude, A_k, can be deduced by propagating the error Δa_k:

$$\Delta A_k = \left|\frac{\partial A_k}{\partial a_k}\right|\Delta a_k + \left|\frac{\partial A_k}{\partial b_k}\right|\Delta b_k = \frac{4}{\pi}\Phi_{1-\alpha}\sqrt{\frac{2}{N}}\sigma_0.$$

In a similar way, the confidence interval for the phase φ_k can be defined by

$$\varphi_k = \tan^{-1}\left(\frac{b_k}{a_k}\right) \pm \Delta\varphi = \tan^{-1}\left(\frac{b_k}{a_k}\right) \pm \frac{4}{\pi}\Phi_{1-\alpha}\sqrt{\frac{2}{N}}\frac{\sigma_0}{A_k}.$$

It is interesting to note that the confidence interval for the phase φ_k decreases for increasing amplitude.

3. Spectrum from the Lomb–Scargle Method

In order to determine the spectrum with the LSM, we assume the recorded signal is described by many frequencies. The most significant frequency is then represented by a peak in the frequency spectrum of a certain width and height. The width is determined by the frequency resolution Δf, which equals $1/T$. From this point of view, the precision of a frequency estimation changes only with the observation length T and seems to be independent from the number of samples N and the signal quality.

The quality Σ is measured by a signal-to-noise ratio-like expression

$$\Sigma = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}\frac{(s_n - y_n)^2}{\sigma_n^2}} \approx \frac{\sqrt{\sum_l A_l^2}}{\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}(\epsilon(t_n))^2}}, \tag{19}$$

with A_l counting the *significant* amplitudes. In Equation (19), y_n is the fitted model, σ_n is the uncertainty per sample, with $\sigma = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}\sigma_n^2}$, and $\epsilon(t_n)$ is the noise per sample. Following VanderPlas [36], we apply Bayesian statistics and assume that every peak is Gaussian-shaped, i.e., $\mathrm{e}^{P(f_{\max}\pm\Delta f)} \propto \mathrm{e}^{-\Delta f^2/\left(2\sigma_f^2\right)}$. It follows that a significant peak $A^2_{\max} = A^2(f_{\max})$ appears at $f_{\max}$, in such a way that $A^2_{\max}/2 = A^2(f_{\max} \pm \Delta f)$ is valid. Here, A^2 is related to the power spectral density $P(f) \propto a_k^2 + b_k^2$. The frequency uncertainty (or standard deviation) is then given by

$$\sigma_f \approx \Delta f\sqrt{\frac{2}{N\Sigma^2}},$$

so that a significant peak is located in the interval $f_{\max} \pm \sigma_f$. This approach suggests that increasing the number of samples in a fixed interval T enhances the precision of $f_{\max}$ by reducing σ_f. However, if the original signal contains two frequencies with a distance in the range of $1/T$, it cannot be excluded, even for the LSM, that these peaks merge together into a single peak with small σ_f, according to Kovács [37].

As a statistical measure, the probability $\mathcal{P}(P_k > P_0)$ states that there is no peak P_k larger than a reference value P_0 of the best fit. From here, the statistical significance of a single frequency ω_k can be deduced as the so-called false alarm probability (FAP) with

$$\mathrm{FAP} = \begin{cases} 1 - (1 - \mathcal{P}(P_k > P_0))^M, & \text{if } \mathcal{P}(P_k > P_0) \approx 1 \\ M\mathcal{P}(P_k > P_0), & \text{if } \mathcal{P}(P_k > P_0) \ll 1 \end{cases}, \tag{20}$$

where M denotes the number of independent (fundamental) frequencies present in the signal. Horne and Baliunas [38] carried out an extensive study about the number of independent frequencies (and the maximum detectable frequency). They found an empirical approximation

$$M = -6.362 + 1.193N + 0.00098N^2,$$

which is a compromise between the conservative $N/2$ and the artificially large minimal distance value. A detailed discussion on FAP and the independent frequencies can be found in the studies by Baluev [39–41]. Furthermore, the Supplemetary Materials provide a detailed discussion on calculating power spectral density (PSD), $P_k = \frac{N}{4\sigma_0^2}(a_k^2 + b_k^2)$, and its standardized companion $p_k \in [0,1]$ according to Zechmeister and Kürster [23] and Hocke [42].

Summarizing the properties of the LSM: (i) there is no truncation error $\epsilon_\mathrm{T} = 0$; (ii) the LSM provides better noise rejection compared to the OMD, $\epsilon^{\mathrm{LS}} < \epsilon^{\mathrm{FS}}$; (iii) the specific sampling pattern influences the spectrum calculated using LSM or another OMD-based method. The reversal of this effect is beyond the scope of this work.

4. The Multivariate Lomb–Scargle Method

A signal S depending on m-independent variables represents a function $S(\vec{t}) : \mathbb{R}^m \to \mathbb{R}$ with the input vector described by $\vec{t} = [t_1, t_2, \ldots, t_m]$. The model function for a multivariate LSM is gained by replacing the scalar arguments of the cosine and sine of the univariate model function of Equation (13) with vectors, resulting in

$$Y(\vec{t}) = \sum_{k=0}^{M} \left(a_k \cos\left(\vec{\omega}_k \cdot \left(\vec{t} - \vec{\tau}_k\right)\right) + b_k \sin\left(\vec{\omega}_k \cdot \left(\vec{t} - \vec{\tau}_k\right)\right)\right).$$

In this case, the shifting parameter $\vec{\tau}_k \in \mathbb{R}^m, 0 \leq k \leq M$ is a vector and, in principle, hard to calculate. However, if the argument of the cosine is expanded, it is obvious that the scalar product $\vec{\omega}_k \cdot \vec{\tau}_k \in \mathbb{R}$ does not depend on the time; thus, the cosine argument can be written as $\vec{\omega}_k \cdot \vec{t} - \tau_k^*$ with $\tau_k^* = \vec{\omega}_k \cdot \vec{\tau}_k$. The determination of τ_k^* is similar to that of τ shown in Equation (14); however, there are some differences in the equations, since the phase, instead of the time coordinate, is shifted now. This determination is described in the following.

4.1. Derivation of the Shifting Parameter

This section shows that shifting the phase, instead of the time, does not affect the Lomb–Scargle algorithm. The derivation of the shifting parameter for the multivariate case is delineated for the time-discrete signal $\hat{S} = \left\{(\hat{S}_i, \vec{t}_i) \in R^{m+1}, 0 \leq i \leq N-1\right\}$. Starting with the orthogonality condition

$$\sum_{n=0}^{N-1} \sin\left(\vec{\omega}_k \cdot \vec{t}_n - \tau_k^*\right) \cos\left(\vec{\omega}_k \cdot \vec{t}_n - \tau_k^*\right) = 0$$

and applying the trigonometric identities to remove the differences in the arguments, we have

$$\sum_{n=0}^{N-1} \left[\left(\underbrace{\cos\left(\vec{\omega}_k \cdot \vec{t}_n\right)}_{\mathrm{c}_n} \underbrace{\cos\left(\tau_k^*\right)}_{\mathrm{c}_\varnothing} + \underbrace{\sin\left(\vec{\omega}_k \cdot \vec{t}_n\right)}_{\mathrm{s}_n} \underbrace{\sin\left(\tau_k^*\right)}_{\mathrm{s}_\varnothing} \right) \left(\underbrace{\sin\left(\vec{\omega}_k \cdot \vec{t}_n\right)}_{\mathrm{s}_n} \underbrace{\cos\left(\tau_k^*\right)}_{\mathrm{c}_\varnothing} - \underbrace{\cos\left(\vec{\omega}_k \cdot \vec{t}_n\right)}_{\mathrm{c}_n} \underbrace{\sin\left(\tau_k^*\right)}_{\mathrm{s}_\varnothing} \right) \right] = 0.$$

By using the defined abbreviations and rearranging the summation (the terms $\mathrm{c}_\varnothing$ and $\mathrm{s}_\varnothing$ do not depend on n), the following equation is easily deduced:

$$\frac{\mathrm{c}_\varnothing \, \mathrm{s}_\varnothing}{{\mathrm{c}_\varnothing}^2 - {\mathrm{s}_\varnothing}^2} = \frac{\sum_{n=0}^{N-1} \mathrm{c}_n \, \mathrm{s}_n}{\sum_{n=0}^{N-1} (\mathrm{c}_n^2 - \mathrm{s}_n^2)}.$$

The fraction on both sides can be further simplified by applying $\cos^2(x) - \sin^2(x) = \cos(2x)$, $\cos(x)\sin(x) = \frac{1}{2}\sin(2x)$, and the definition of the tangent. This results in

$$\tan(2\tau_k^*) = \frac{\sum_{n=0}^{N-1} \sin(2\vec{\omega}_k \cdot \vec{t}_n)}{\sum_{n=0}^{N-1} \cos(2\vec{\omega}_k \cdot \vec{t}_n)}. \tag{21}$$

Notice that in comparison with the one-dimensional LSM, the frequency ω_k is missing on the left side in Equation (21).

4.2. Parameter Estimation

Similar to the procedure for the LSM for a discrete signal in one dimension (see Equation (16)), a discrete multivariate signal $(\hat{S}_i, \vec{t}_i)$ with N samples is multiplied by $\cos(\vec{\omega}_k \cdot \vec{t}_i - \tau_k^*)$, resulting in

$$\sum_{n=0}^{N-1} \hat{S}(\vec{t}_n)\cos(\vec{\omega}_k \cdot \vec{t}_n - \tau_k^*) = a_k \sum_{n=0}^{N-1} \cos^2(\vec{\omega}_k \cdot \vec{t}_n - \tau_k^*) + b_k \sum_{n=0}^{N-1} \underbrace{\sin(\vec{\omega}_k \cdot \vec{t}_n - \tau_k^*)\cos(\vec{\omega}_k \cdot \vec{t}_n - \tau_k^*)}_{=0, \text{ if orthogonal}} + \sum_{n=0}^{N-1} \underbrace{\epsilon_k(\vec{t}_n)\cos(\vec{\omega}_k \cdot \vec{t}_n - \tau_k^*)}_{\text{random error}}.$$

The determination of the parameter a_k and its confidence interval Δa_k are similar to the one-dimensional case shown in Equation (18):

$$\frac{\sum_{n=0}^{N-1} y(\vec{t}_n)\cos(\vec{\omega}_k \cdot \vec{t}_n - \tau_k^*)}{\sum_{n=0}^{N-1} \cos^2(\vec{\omega}_k \cdot \vec{t}_n - \tau_k^*)} = a_k, \quad \Delta a_k = \frac{4}{\pi}\Phi_{1-\alpha}\frac{\sigma_0}{\sqrt{N}}.$$

The parameter b_k and its confidence interval Δb_k are calculated analogously:

$$\frac{\sum_{n=0}^{N-1} y(\vec{t}_n)\sin(\vec{\omega}_k \cdot \vec{t}_n - \tau_k^*)}{\sum_{n=0}^{N-1} \sin^2(\vec{\omega}_k \cdot \vec{t}_n - \tau_k^*)} = b_k, \quad \Delta b_k = \Delta a_k.$$

We recall that the α-quantile of the sampling error distribution $\Phi_{1-\alpha}$ (with standard deviation σ) has been used to model the random error. The power spectral density, as well as the false alarm probability, are calculated in the same way as the one-dimensional LSM method (see the Supplementary Materials).

An implementation of the multivariate Lomb–Scargle method is available, in the `spectral` (R) package published on CRAN [24], to give the user access to a multidimensional analysis with an easy-to-use interface.

5. Application

The selected applications are ordered by the increasing complexity of the sampling and the data themselves. First, a regular grid with missing values for synthetic test data is considered. Second, Ultrasound Doppler Velocimetry (UDV) measurement data, which contain jitter and missing values, are analyzed. Finally, a general data scenario given by an astrophysical 2D set of sunspots, which appear freely in time and space, is investigated.

5.1. Synthetic Test Data

We start with a simple test case, where the input signal is a simple two-dimensional plain wave $z = \cos\left(2\pi(xf_x + yf_y) + \frac{\pi}{4}\right)$, with $f_x = 3.25$ and $f_y = 6.32$ representing the dimensionless frequencies. These fixed frequencies are selected in such a way that $f_x, f_y \neq k \cdot \omega_0$, with $\omega_0 = 2\pi/T_{x,y}$ and $T_{x,y} = 1$. This ensures the situation where the signal

frequency does not map to any test frequency $k \cdot \omega_0$ if DFT is performed, which results in the worst-case scenario for all DFT-based methods. The aim is to enlarge the truncation error ϵ_T in case of the DFT for frequencies that do not match.

The 2D sampling of $x, y \in [0, 1]$ is performed with $\delta x, \delta y = 0.025$. With respect to the chosen frequencies f_x and f_y, it becomes evident that both frequencies do not fit to the data range by an integer fraction. Data gaps are introduced by removing randomly distributed and uncorrelated grid points covering 60% of the total dataset. The latter is an exaggerated example of how data with a high proportion of missing values can be analyzed using multivariate LSM. Therefore, Figure 2a shows the non-uniform distributed input data. Gray areas indicate missing values (non-available numbers "nan"). As the LSM requires an appropriate input vector of frequencies, we choose for both variables $f_x, f_y \in [-10, 10]$ with a resolution of $\delta f_x, \delta f_y = 0.025$. This ensures that the frequency space is sampled sufficiently densely.

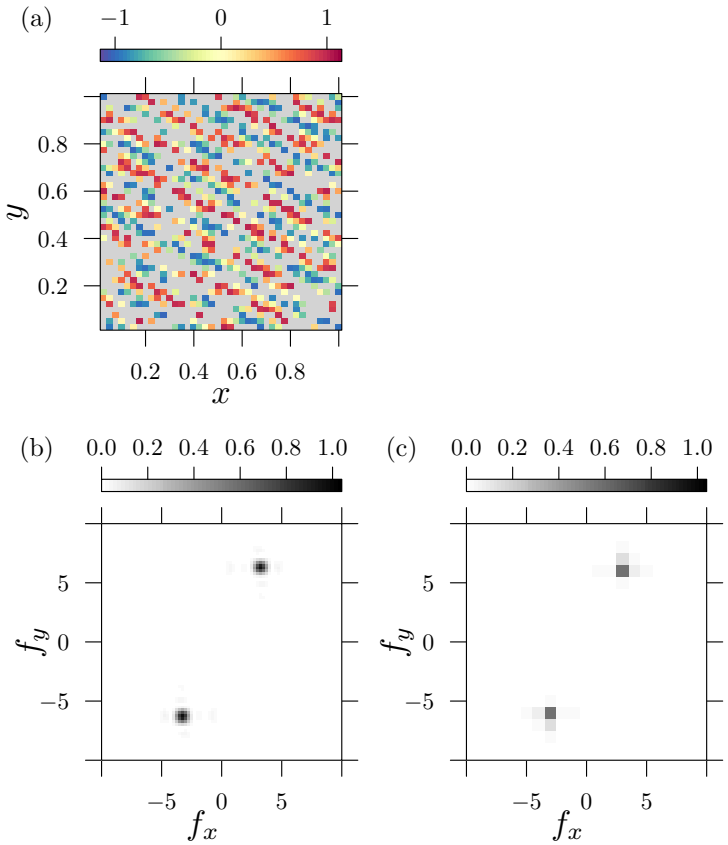


Figure 2. Comparison of the PSD calculated with the LSM and with the DFT approach for two-dimensional input data with missing values: (**a**) Sampled data $z = \cos(2\pi(xf_x + yf_y) + \pi/4)$, for $x, y \in [-1, 1]$ with sampling increment $\delta x, \delta y = 0.025$, $f_x = 3.25$, and $f_y = 6.32$. Gray areas represent missing values. (**b**) PSD calculated with the LSM and (**c**) PSD calculated with the DFT. Notice that the truncation error and the sparse resolution leads to a rough approximation of the frequency and to a reduced amplitude (by 50%) in the PSD calculated using the DFT.

The corresponding power spectral density (PSD) from the LSM (see the Supplementary Materials for the definition), shown in Figure 2b, displays the maxima at the first and third quadrants, which represent a wave traveling upwards. The figure illustrates that even with a lot of missing data, it is possible to properly detect the periodic signal. In the present case, the value of $\mathrm{psd}(f_x, f_y) \approx 1$ indicates a perfect fit to the corresponding sinusoidal model. For comparison, Figure 2c depicts the standardized PSD

$$\mathrm{psd}_{\mathrm{FT}}(\vec{\omega}) = \frac{N}{N-1} \cdot \frac{1}{2\sigma_0^2} \left(\frac{2|\mathcal{F}(z(x,y))(\vec{\omega})|}{1 - \frac{N_{\mathrm{zero}}}{N}} \right)^2$$

calculated from a DFT denoted by the operator $\mathcal{F}\big(z(x,y)\big)(\vec{\omega}) = \int z(x,y) \mathrm{e}^{\mathrm{i}(\omega_x x + \omega_y y)} \mathrm{d}x\, \mathrm{d}y$. Here, missing values (nan) are handled by zero padding, which means that the introduced gaps are filled with zeroes. Since the DFT represents a conservative transformation (mapping) into the spectral domain, zero padding leads to a significant reduction in the average signal energy. A proper rescaling to the number of zeros N_{zero} (missing values) is required to ensure approximated amplitude estimations. Zero padding always changes the character of the input signal, in such a way that the corresponding DFT treats the zeros as if they were part of the original signal. The resulting PSD then becomes different from the original. Since the signal frequencies f_x, f_y do not fit into an integer-spaced scheme of the data range, the DFT suffers drawbacks, as briefly discussed in Section 2.2. The effects of leakage and the misfit of frequencies can be seen in Figure 2c, in terms of a rather coarse resolution and a lower PSD value.

5.2. Three-Dimensional UDV Flow Measurement

The second example is taken from previous experimental flow measurements [26] devoted to investigating the azimuthal magneto-rotational instability (AMRI). The experiment consists of a cylindrical annulus containing a liquid metal between the inner and the outer wall. The inner wall rotates at a frequency $\omega_{\mathrm{in}} = 2\pi \cdot 0.05\,\mathrm{Hz}$, whereas the outer wall rotates at a frequency $\omega_{\mathrm{out}} = 2\pi \cdot 0.013\,\mathrm{Hz}$. The flow is driven by the shear, since $\omega_{\mathrm{in}} - \omega_{\mathrm{out}} > 0$. Two ultrasound sensors mounted on the outer cylinder, on opposite sides of each other and at the same radius, measure the axial velocity component v_z along the line of sight parallel to the rotation axis of the cylinder. Additionally, the liquid metal flow is exposed to a magnetic field $B_\varphi \propto r^{-1}$ (r is the radial coordinate), originating from a current I_{axis} on the axis of the cylinder. A typical time series of the axial velocity (along the measuring line) is displayed in Figure 3. The existence of periodic patterns (large blue inclined stripes) in this figure indicates a traveling wave propagating in the fluid. In cylindrical geometry, a traveling wave is described in terms of the drift frequency $\omega = 2\pi f$, the vertical wave numbers k, and the azimuthal wave numbers m. A close look at Figure 3 shows that this wave is not axially symmetric, and the leading azimuthal wave number is $m = 1$. This is indeed proven in the following through the LSM analysis.

The measured time series for v_z depends on the time t_n, depth d_n, and the angular positions of the sensors $\varphi_n^{\mathrm{S}_1} = \omega_{\mathrm{out}} t_n$ and $\varphi_n^{\mathrm{S}_2} = \varphi_n^{\mathrm{S}_1} + \pi$, where the subscript n indicates the sampling. The azimuthal angles depend on the time, since the sensors are attached to the outer wall. This induces the mapping $v_z = f(t_n, d_n, \varphi_n)$, where $\varphi_n = \{\varphi_n^{\mathrm{S}_1}, \varphi_n^{\mathrm{S}_2}\}$ is meant to be one of the two sensors at each sampling instance. Finally, the measured velocity component is a mapping $v_z : \mathbb{R}^3 \mapsto \mathbb{R}$, which fits perfectly to the proposed multivariate LSM.

Figure 4 illustrates the result of the LSM decomposition into several azimuthal m-modes with the corresponding amplitudes. The $m = 0$ mode contains a stationary structure, at a frequency $f \approx 0$, which originates from sensor mis-alignments and (thermal) side effects in the flow. The minor non-stationary components, here the two point-symmetric peaks, originate from crosstalk (or alias projections, see labels on Figure 4) of the $m = 1$ mode. This is mainly for two reasons: First, the sensors do not behave identically (e.g., misalignment or different sensitivity) and respond slightly differently to the same flow signal. This means that one of the sensors projects a little more energy into the data than the other, which leads to a "leakage" effect and a weak signal in the $m = 0$ mode. Second, the rotation of the sensors acts as an additional sampling frequency f_{out}. In consequence, the spectrum of this regular sampling function folds with the spectrum of the observed process,

leading to alias images (copies) of the original process spectrum into other frequency ranges (i.e., ms). In the present multivariate analysis, these alias structures exhibit point symmetry in the $m = 0$ panel, indicating corresponding traveling waves. It is evident that these aliases originate as a systematic error from the LSM method and, due to their symmetry, effectively cancel each other out. One could also argue that the multivariate LSM enables a robust identification of alias amplitudes that do not belong to the observed process.

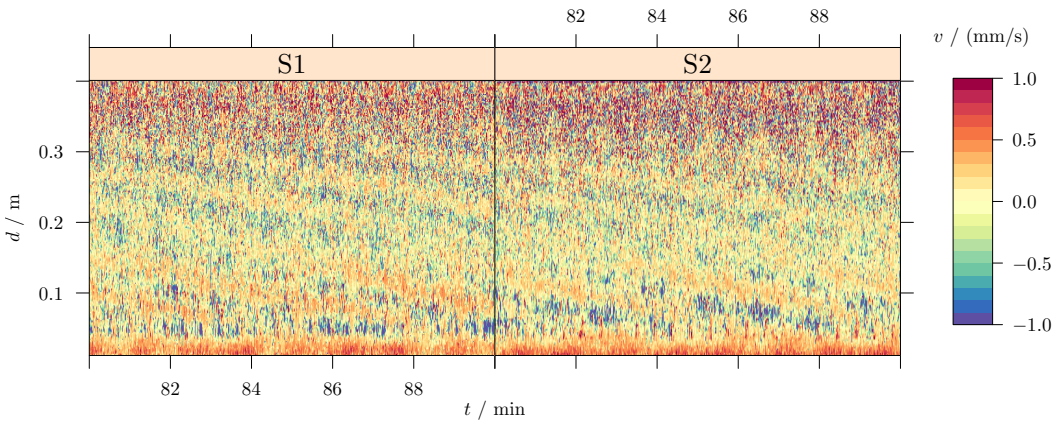


Figure 3. Sensor data in the co-rotating reference frame. Data are taken as a time series from two rotating sensors. The measurement time of the sensor data is folded with the individual phase location; so, $\varphi_n^{S_1} = \omega_{\text{out}} t_n$ (sensor1, S1) and $\varphi_n^{S_2} = \varphi_n^{S_1} + \pi$ (sensor2, S2) describe the two ordinates.

The AMRI wave itself is located in the $m = 1$ panel, with a characteristic frequency in time (f) and in the vertical direction (k). Here, two components can be identified: the dominant wave at $f \approx 9\,\text{mHz}$ and $k \approx 20\,\text{m}^{-1}$ and a minor counterpart at $f \approx 4\,\text{mHz}$ and $k \approx -20\,\text{m}^{-1}$. The signals observed in the $m = 3$ and $m = 5$ panels might be due to aliases because of the rotation of the inner cylinder at a frequency $f_\text{i} = 0.05\,\text{Hz}$.

The advantage of "high" dimensional spectral decomposition is the improved noise rejection. With respect to the raw data, given in Figure 3, it is obvious that high-frequency noise is present in the data. Depending on the exact distribution of this noise, its energies spread over a certain range of frequencies. If we could select a representative depth and angle so that $d_n, \varphi_n = \text{const.}$, the velocity v_z would only depend on time. In the subsequent one-dimensional analysis, the noise would accumulate along the single frequency ordinate, probably hiding the signal of interest. Taking the higher-order analysis distributes the noise energies over multiple domain variables. For the present example, this means that the distortions are projected into higher frequencies f, higher ms, and larger ks. Since the signal of interest remains in the same spectral corridor, its signal amplitude becomes clearer because of the "reduced" local noise.

5.3. Analyzing 2D Sunspot Data

Since the beginning of the 17th century, systematic visual observations of sunspots have been available. This enables the investigation of dynamic processes taking place in the sun. The appearance of sunspots on the sun's surface depends on the level of solar activity; therefore, it renders some fundamental features of the underlying solar dynamo, such as the 11 yr solar cycle. Since the beginning of the 1820s, observational data have been available, in terms of a two-dimensional time series, as displayed in Figure 5. The data exhibit a periodic wing-like pattern, essentially symmetric with respect to the sun's equator, forming the sunspot butterfly diagram. This diagram summarizes the individual sunspot groups appearing at a certain time and latitude on the Sun, for the past 190 years. Additionally, Figure 5 depicts the assigned field polarity order indicated by the sunspot color (gray or black). Given a year Y and a mean latitude L (degrees), the field polarity $P(Y, L) \in \{-1, 1\}$

gives the arrangement of the leading north/south polarity of a sunspot group. This polarity is changing approximately every 11 years, which is called Schwabe's cycle. The full period of 22 years is called the Hale cycle. The segmentation, shown in Figure 5, was obtained following the suggestions from Leussu et al. [43], with some simplifications leading only to minor mis-assignments for individual sunspot groups.

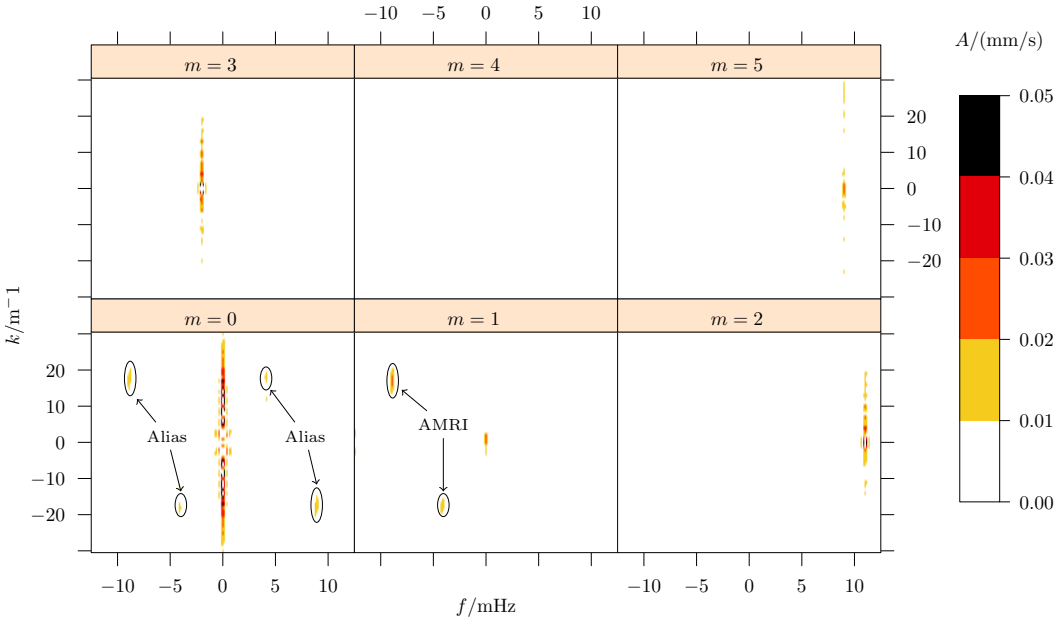


Figure 4. Spectrum of sensor data. The amplitude spectrum is calculated from the data in Figure 3 with the LSM. The weak alias peaks in $m = 0$ panel originate from the sensor mismatch and aliasing effect due to the outer rotation. The latter acts as a sampling frequency and, therefore, projects the AMRI wave to $m = 0$.

The sunspot data are taken "as-is" from Leussu et al. [44], which originate from several sources with different levels of quality, i.e., the Royal Greenwich Observatory–USAF/NOAA(SOON) (Available at http://solarscience.msfc.nasa.gov/greenwch.shtml, (accessed on 10 October 2025)) [45,46], Schwabe [47], and Spoerer [48] datasets. The reader might also refer to the historic publications of Spoerer [49], Spoerer and Maunder [50]. A detailed discussion of the data collection is given by Leussu et al. [43,44] and the referenced literature.

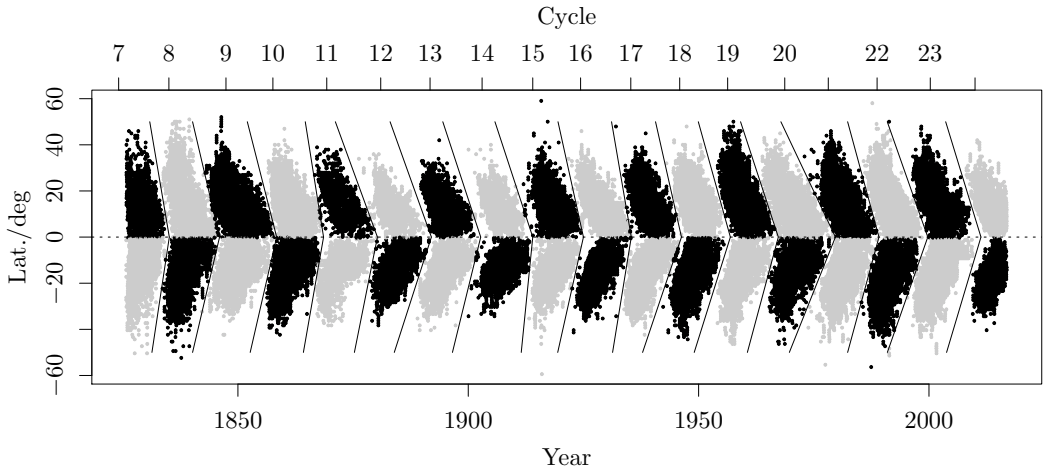


Figure 5. Butterfly diagram of the sunspot data. The separation of the individual wings took place according to the procedure presented in Leussu et al. [43]. The colored patches indicate the changing polarity of each cycle. The tilted segmentation lines depict the optimized borders between two cycles.

In this section, the LSM spectral decomposition from this unevenly sampled binary dataset is obtained in order to identify typical periods. In contrast to the two examples analyzed in Sections 5.1 and 5.2, the sunspot dataset consists of real, arbitrary sampling in both time and space, providing the most general scenario input data for the LSM. The starting point for the subsequent analysis is the simplified segmentation of the butterfly diagram, reassigning the field polarity. The segmentation takes place by optimizing the distance d (in the (L, Y)-plane) between each sunspot and the segmentation line

$$L(Y) = \begin{cases} m_1(Y - T_0) & m_1 < 0, L > 0 \\ m_2(Y - T_0) & m_2 > 0, L \leq 0 \end{cases},$$

which is a piecewise linear function with intersection point $(T_0, 0)$. The individual slopes, with $|m_1|, |m_2| < 12\,\mathrm{deg/yrs}$, are assigned on the northern and southern hemispheres, respectively. The final segmentation, shown in Figure 5 consists of 18 individual cycles subdivided by a ">"-shaped separation area.

The LSM spectral decomposition is based on the definition of a two-dimensional wave model

$$P(Y, L) \sim A\cos(\omega_{\mathrm{Y}} Y + \omega_{\mathrm{L}} L + \tau^*) + B\sin(\omega_{\mathrm{Y}} Y + \omega_{\mathrm{L}} L + \tau^*) + \epsilon, \tag{22}$$

where P is given by the assigned patch polarity. To show the advantage and robustness of the LSM, the raw data are employed without any further preprocessing. The selected frequencies $\omega_{\mathrm{Y},n}, \omega_{\mathrm{L},n} \propto n^{-1}$ are given on a rectangular grid with inverse distance so that the periods are distributed uniformly alongside the consecutive counter n. Furthermore, the frequency resolution close to $|\omega_{\mathrm{L}}| = 0$ is selected as finer to better resolve this region.

Figure 6 gives the resulting LSM spectral decomposition in a reciprocal log-scale plot to analyze the different periods. Each of the selected peaks (numbered black dots) provides an FAP value of $p < 10^{-10}$, meaning that these periods are significantly different from the noise level. Each of these points originates from a local frequency refinement to achieve the local maximum amplitude. The binary order information $\{-1, 1\}$ induced by the sunspots introduces higher harmonics into the spectrum, which also appears with a significantly low FAP value.

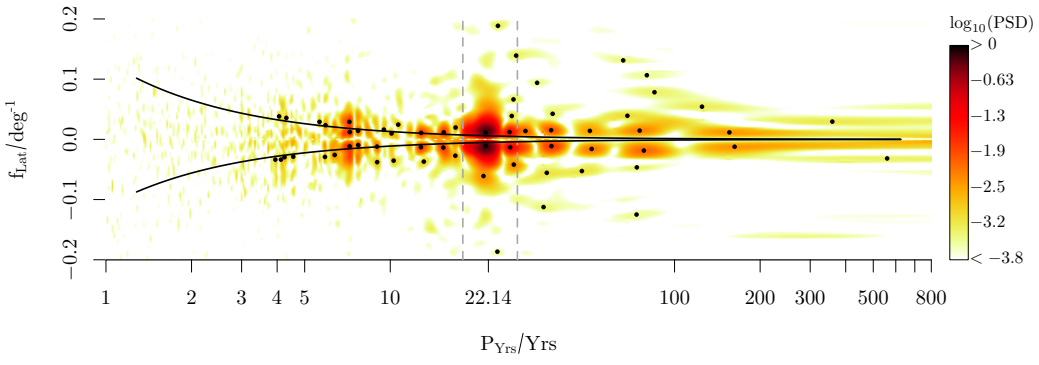


Figure 6. Two-dimensional LSM spectrum of the sunspot data. The curved line follows the path line $v = (f_{\mathrm{L}} P_{\mathrm{Y}})^{-1} = \text{const.}$, indicating a wave structure with a certain time dependence. The range of the solar cycle period is given by the vertical dashed gray lines.

Figure 6 provides good agreement between the peaks found and the commonly known periods. Table 1 summarizes the major outcome in comparison with the literature. For example, the 22 year Hale cycle varies from 18 to 28 years (Usoskin [51]), which is identified by the two main peaks. Since the solar cycle is modulated—we refer the reader to Hathaway [52]—it is

natural that a broad spectrum with many harmonics is present. These subsequent patterns are related to a set of local maxima, mainly collapsing on a horizontal line at $|f_{\mathrm{Lat}}| \approx 1.4 \cdot 10^{-2}\,\mathrm{deg}^{-1}$, indicating that the complex structure of the individual wing shapes of the butterfly diagram with different widths, heights, and orientations corresponds to a main period $P_{\mathrm{Hale}} = 21.634\,\mathrm{yrs}$. Next, we can identify typical periods, which are related to the Gleissberg process (see Table 1), or the short periods in the range of $P \approx 7.2\,\mathrm{yrs}$, which are consistent with the data provided by Prestes et al. [53] and Kane [54] and partly with Deng et al. [55].

The latitudinal dimension of the spectrum spreads the peaks vertically, which is an advantage over a purely 1D analysis, where all signals would be projected on the $f_{\mathrm{Lat}} = 0$ line. In addition, complex patterns such as the lines of constant phase velocities (dashed curve of Figure 6) may be used to analyze modulations in the dynamics of the sunspot motion.

Table 1. Common peaks. Period values are given in years. Values above 100 yrs are affected by the low period resolution caused by the limited time span of sunspot data. The results from Prestes et al. [53] refer to the Schwabe cycle and, therefore, are doubled.

Process	Common Period	From Spectrum	
Eddy	515	559	(1)
	350	359	(1)
Hale (Schwabe)	22.14 (11.07)	21.63 ± 2.5	(2)
	18 . . . 28		(2)
Gleissberg	88 (80 . . . 150)	78, 85, 125, 156	(1), (2)
–	126	125	(3)
–	2×3.6	7.2	(4)
–	2×3.9	7.7	(4), (5)

(1) McCracken et al. [56]; (2) Usoskin [51]; (3) Ogurtsov et al. [57]; (4) Prestes et al. [53] (Figure 6); (5) Kolotkov et al. [58].

6. Conclusions

In the present work, a multidimensional extension of the Lomb–Scargle method was developed. The key aspect is the redefinition of the phase argument to $\vec{\phi}_{\mathrm{new}} = \vec{\omega} \cdot \vec{t} - \tau^*$. We suggest using a modified shifting parameter τ^* in contrast to the traditional approach, which shifts the ordinate $\phi_{\mathrm{orig}} = \omega \cdot (t - \tau)$ instead of the phase. This enables multivariate modeling with a single scalar value τ^* for all independent variables, as there is always a shifting parameter, τ^*, for which $\int_a^b \sin(\phi(t)) \cos(\phi(t))\ \mathrm{d}t$ vanishes on any interval $[a, b]$.

With respect to the evaluation of the measurement results, the noise rejection $\epsilon_{\mathrm{LS}} \propto 4N^{-0.5}/\pi$ and confidence intervals (Δa_k and Δb_k) of the model parameters a_k and b_k are delineated. To emphasize the advantages of the LSM, it is compared with the traditional Fourier mode decomposition. The systematic error ϵ_{T} does not vanish for FT-based methods with the increasing number of samples on a fixed interval T, which finally leads to the common leakage effect. Here, the signal amplitudes are distributed on an area of neighboring spectral components; i.e., individual bins or pixels. We conclude that the standard orthogonal mode-related procedures do not represent a consistent estimator for model parameters a_k and b_k, whereas the introduction of τ^* in the LSM leads to a consistent estimator with $\epsilon_{\mathrm{T}} = 0$, even for higher dimensions. Finally, it was shown that the LSM converges to the true model parameters with an increasing number of samples and provides better noise rejection ($\epsilon_{\mathrm{LS}} < \epsilon_{\mathrm{FS}}$) as well.

The examples from Section 5 underline the strengths of the developed procedure in a consecutive way. First, the application on ideal two-dimensional test data shows the ability to analyze fragmented time series. Here, the sampling remains regular, meaning $t = nT_0$ with $n \in \mathbb{N}$, but with $n_i - n_{i+1} \neq 1$ as an incomplete set of locations to describe

the missing values. A second quasi-similar situation is given in the experimental dataset of UDV measurements, for which a certain level of jitter is expected. The dimensionality in this example was extended to $\mathbb{R}^3$ to decompose the wave parameters in frequency f, symmetry m, and spatial frequency k.

As the third analysis, sunspot datasets represent the most general case of uneven sampling. Taking only the time series of positions into account, the LSM is able to calculate the spectrum on an individual frequency grid, pronouncing the low frequencies close to zero. The assignment of $\{-1, 1\}$ to each data point is a minor modification, such that the data can be assumed to be binary raw data. The spectral decomposition with the LSM shows the characteristic footprint of waves and its dynamics present in the butterfly diagram. The time–frequency spectrum itself provides the commonly known frequencies, i.e., Hale Cycle or Gleissberg process. The second frequency f_{Lat} provides information about the minor peaks, which spread into the latitudinal domain. Moreover, from these values, the sunspot drift motion may be calculated using a characteristic v-line with its side bands.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/s25216535/s1, (i) All R-scripts calculating the presented Figures, (ii) raw data set and processed data set from the MRI experiment at HZDR, (iii) sun spot data and processed data set including tabulated significant frequencies from the sun spot data set. References [59–61] is cited in the Supplementary Materials.

Author Contributions: Conceptualization, M.S., F.G. and T.W.; methodology, M.S.; software, M.S.; validation, M.S., F.G. and T.W.; formal analysis, M.S., F.G. and T.W.; data curation, M.S.; writing—original draft preparation, M.S.; writing—review and editing, M.S., F.G.; visualization, M.S.; supervision, T.W.; project administration, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data required for this study are either cited or included in the Supplementary Materials section of this publication.

Acknowledgments: The authors would like to thank Rainer Arlt from the Astrophysical Institute, Potsdam, for providing the sunspot data and Andre Gieseke from the HZDR for many fruitful discussions. F. Garcia kindly acknowledges the Serra Húnter and SGR (2021-SGR-01045) programs of the Generalitat de Catalunya (Spain).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cohen, L. *Time-Frequency Analysis*; Prentice Hall Signal Processing Series; Prentice Hall PTR: Englewood Cliffs, NJ, USA, 1995.
2. Oppenheim. *Discrete-Time Signal Processing*, 2nd ed.; Prentice Hall: Englewood Cliffs, NJ, USA, 1999.
3. Haacke, E. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*; John Wiley and Sons: Hoboken, NJ, USA, 1999.
4. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 3rd ed.; Prentice Hall: Englewood Cliffs, NJ, USA, 2008.
5. James, J. *A Student's Guide to Fourier Transforms: With Applications in Physics And Engeneering*, 3rd ed.; Cambridge University Press: Cambridge, UK, 2011.
6. Frydman, L.; Lupulescu, A.; Scherf, T. Principles and Features of Single-Scan Two-Dimensional NMR Spectroscopy. *J. Am. Chem. Soc.* **2003**, *125*, 9204–9217. [CrossRef]
7. Giraudeau, P.; Frydman, L. Ultrafast 2D NMR: An Emerging Tool in Analytical Spectroscopy. *Annu. Rev. Anal. Chem.* **2014**, *7*, 129–161. [CrossRef]
8. Munteanu, C.; Negrea, C.; Echim, M.; Mursula, K. Effect of data gaps: Comparison of different spectral analysis methods. *Ann. Geophys.* **2016**, *34*, 437–449. [CrossRef]

9. Fessler, J.A. *Iterative Tomographic Image Reconstruction Using Nonuniform Fast Fourier Transforms*; The University of Michigan: Ann Arbor, MI, USA, 2002; p. 11.
10. Fessler, J.; Sutton, B. Nonuniform fast fourier transforms using min-max interpolation. *IEEE Trans. Signal Process.* **2003**, *51*, 560–574. [CrossRef]
11. Liu, Q.; Nguyen, N. An accurate algorithm for nonuniform fast Fourier transforms (NUFFT's). *IEEE Microw. Guid. Wave Lett.* **1998**, *8*, 18–20. [CrossRef]
12. Wei, D.; Yang, J. Non-Uniform Sparse Fourier Transform and Its Applications. *IEEE Trans. Signal Process.* **2022**, *70*, 4468–4482. [CrossRef]
13. Press, W.H.; Rybicki, G.B. Fast Algorithms for Spectral Analysis of Unevenly Sampled Data. *Astrophys. J.* **1989**, *338*, 277–280. [CrossRef]
14. Greengard, L.; Lee, J.Y. Accelerating the Nonuniform Fast Fourier Transform. *Siam Rev.* **2004**, *46*, 443–454. [CrossRef]
15. Geneva, P.; Eckenhoff, K.; Huang, G. Asynchronous Multi-Sensor Fusion for 3D Mapping And Localization. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 1–6. [CrossRef]
16. Cadena, C.; Carlone, L.; Carrillo, H.; Latif, Y.; Scaramuzza, D.; Neira, J.; Reid, I.; Leonard, J.J. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Trans. Robot.* **2016**, *32*, 1309–1332. [CrossRef]
17. Sudars, K. Data acquisition based on nonuniform sampling: Achievable advantages and involved problems. *Autom. Control Comput. Sci.* **2010**, *44*, 199–207. [CrossRef]
18. Lomb, N.R. Least-Squares Frequency Analysis of Unequally Spaced Data. *Astrophys. Space Sci.* **1976**, *39*, 447–462. [CrossRef]
19. Scargle, J.D. Studies in Astronomical Time Series Analysis. II - Statistical Aspects of Spectral Analysis of Unevenly Spaced Data. *Astrophys. J.* **1982**, *263*, 835–853. [CrossRef]
20. Townsend, R.H.D. Fast Calculation of the Lomb-Scargle Periodogram Using Graphics Processing Units. *Astrophys. J. Suppl. Ser.* **2010**, *191*, 247. [CrossRef]
21. Leroy, B. Fast calculation of the Lomb-Scargle periodogram using nonequispaced fast Fourier transforms. *Astron. Astrophys.* **2012**, *545*, A50. [CrossRef]
22. Mathias, A.; Grond, F.; Guardans, R.; Seese, D.; Canela, M.; Diebner, H.H.; Baiocchi, G. Algorithms for Spectral Analysis of Irregularly Sampled Time Series. *J. Stat. Softw.* **2004**, *11*, 1–30. [CrossRef]
23. Zechmeister, M.; Kürster, M. The Generalised Lomb-Scargle Periodogram. A New Formalism for the Floating-Mean And Keplerian Periodograms. *Astron. Astrophys.* **2009**, *496*, 577–584. [CrossRef]
24. Seilmayer, M. *Common Methods of Spectral Data Analysis*; Helmholtz-Zentrum Dresden-Rossendorf: Dresden, Germany, 2021. Available online: http://CRAN.R-project.org/package=spectral (accessed on 10 October 2025).
25. Babu, P.; Stoica, P. Spectral analysis of nonuniformly sampled data–A review. *Digit. Signal Process.* **2010**, *20*, 359–378. [CrossRef]
26. Seilmayer, M.; Galindo, V.; Gerbeth, G.; Gundrum, T.; Stefani, F.; Gellert, M.; Rüdiger, G.; Schultz, M.; Hollerbach, R. Experimental Evidence for Nonaxisymmetric Magnetorotational Instability in a Rotating Liquid Metal Exposed to an Azimuthal Magnetic Field. *Phys. Rev. Lett.* **2014**, *113*, 024505. [CrossRef]
27. Seilmayer, M.; Gundrum, T.; Stefani, F. Noise Reduction of Ultrasonic Doppler Velocimetry in Liquid Metal Experiments with High Magnetic Fields. *Flow Meas. Instrum.* **2016**, *48*, 74–80. [CrossRef]
28. Barning, F.J.M. The Numerical Analysis of the Light-Curve of 12 Lacertae. *Bull. Astron. Institutes Neth.* **1963**, *17*, 22.
29. Weisstein, E.W. Orthogonal Functions, From MathWorld–A Wolfram Resource. 2025. Available online: https://mathworld.wolfram.com/OrthogonalFunctions.html (accessed on 16 October 2025).
30. *JCGM 100:2008 E*; Evaluation of Measurement Data–Guide to the Expression of Uncertainty in Measurement. Joint Committee for Guides in Metrology: Paris, France, 2008.
31. Thompson, J.K.; Tree, D.R. Leakage error in Fast Fourier analysis. *J. Sound Vib.* **1980**, *71*, 531–544. [CrossRef]
32. Jerri, A.J. The Shannon Sampling Theorem—Its Various Extensions And Applications: A Tutorial Review. *Proc. IEEE* **1977**, *65*, 1565–1596. [CrossRef]
33. Shannon, C.E. Communication in the Presence of Noise. *Proc. IRE* **1949**, *37*, 10–21. [CrossRef]
34. Al-Ani, M.; Tarczynski, A. Evaluation of Fourier transform estimation schemes of multidimensional signals using random sampling. *Signal Process.* **2012**, *92*, 2484–2496. [CrossRef]
35. Parzen, E. *Stochastic Processes*; Holden Day Series in Probability and Statistics; Holden-Day: San Francisco, CA, USA, 1962.
36. VanderPlas, J.T. Understanding the Lomb-Scargle Periodogram. *Astrophys. J.* **2017**, *236*, 16. [CrossRef]
37. Kovács, G. Frequency Shift in Fourier Analysis. *Astrophys. Space Sci.* **1981**, *78*, 175–188. [CrossRef]
38. Horne, J.H.; Baliunas, S.L. A Prescription for Period Analysis of Unevenly Sampled Time Series. *Astrophys. J.* **1986**, *302*, 757. [CrossRef]
39. Baluev, R.V. Assessing the Statistical Significance of Periodogram Peaks. *Mon. Not. R. Astron. Soc.* **2008**, *385*, 1279–1285. [CrossRef]

40. Baluev, R.V. Detecting Multiple Periodicities in Observational Data with the Multifrequency Periodogram—II. Frequency Decomposer, a Parallelized Time-Series Analysis Algorithm. *Astron. Comput.* **2013**, *3–4*, 50–57. [CrossRef]
41. Baluev, R.V. Detecting Multiple Periodicities in Observational Data with the Multifrequency Periodogram – I. Analytic Assessment of the Statistical Significance. *Mon. Not. R. Astron. Soc.* **2013**, *436*, 807–818. [CrossRef]
42. Hocke, K. Phase Estimation with the Lomb-Scargle Periodogram Method. *Ann. Geophys.* **1998**, *16*, 356–358.
43. Leussu, R.; Usoskin, I.G.; Arlt, R.; Mursula, K. Properties of sunspot cycles and hemispheric wings since the 19th century. *Astron. Astrophys.* **2016**, *592*, A160. [CrossRef]
44. Leussu, R.; Usoskin, I.G.; Pavai, V.S.; Diercke, A.; Arlt, R.; Denker, C.; Mursula, K. Wings of the butterfly: Sunspot groups for 1826–2015. *Astron. Astrophys.* **2017**, *599*, A131. [CrossRef]
45. Clette, F.; Svalgaard, L.; Vaquero, J.M.; Cliver, E.W. Revisiting the Sunspot Number. A 400-Year Perspective on the Solar Cycle. *Space Sci. Rev.* **2014**, *186*, 35–103. [CrossRef]
46. Willis, D.M.; Wild, M.N.; Warburton, J.S. Re-examination of the Daily Number of Sunspot Groups for the Royal Observatory, Greenwich (1874–1885). *Sol. Phys.* **2016**, *291*, 2519–2552. [CrossRef]
47. Arlt, R.; Leussu, R.; Giese, N.; Mursula, K.; Usoskin, I.G. Sunspot positions and sizes for 1825-1867 from the observations by Samuel Heinrich Schwabe. *Mon. Not. R. Astron. Soc.* **2013**, *433*, 3165–3172. [CrossRef]
48. Diercke, A.; Arlt, R.; Denker, C. Digitization of sunspot drawings by Spörer made in 1861-1894: Digitization of sunspot drawings by Spörer made in 1861-1894. *Astron. Nachrichten* **2015**, *336*, 53–62. [CrossRef]
49. Spoerer, G. Mémoires et observations. Sur les différences que présentent l'hémisphère nord et l'hémisphère sud du soleil. *Bull. Astron. Ser. I* **1889**, *6*, 60–63.
50. Spoerer, F.W.G.; Maunder, E.W. Prof. Spoerer's researches on Sun-spots. *Mon. Not. R. Astron. Soc.* **1890**, *50*, 251. Available online: https://ui.adsabs.harvard.edu/abs/1890MNRAS..50..251S (accessed on 16 October 2025).
51. Usoskin, I.G. A history of solar activity over millennia. *Living Rev. Sol. Phys.* **2017**, *14*, 3. [CrossRef]
52. Hathaway, D.H. The Solar Cycle. *Living Rev. Sol. Phys.* **2015**, *12*, 4. [CrossRef]
53. Prestes, A.; Rigozo, N.R.; Echer, E.; Vieira, L.E.A. Spectral analysis of sunspot number and geomagnetic indices (1868–2001). *J. Atmos. Sol.Terr. Phys.* **2006**, *68*, 182–190. [CrossRef]
54. Kane, R.P. Quasi-biennial and quasi-triennial oscillations in geomagnetic activity indices. *Ann. Geophys.* **1997**, *15*, 1581–1594. [CrossRef]
55. Deng, H.; Mei, Y.; Wang, F. Periodic variation and phase analysis of grouped solar flare with sunspot activity. *Res. Astron. Astrophys.* **2020**, *20*, 022. [CrossRef]
56. McCracken, K.; Beer, J.; Steinhilber, F.; Abreu, J. The Heliosphere in Time. *Space Sci. Rev.* **2013**, *176*, 59–71. [CrossRef]
57. Ogurtsov, M.G.; Nagovitsyn, Y.A.; Kocharov, G.E.; Jungner, H. Long-Period Cycles of the Sun's Activity Recorded in Direct Solar Data and proxies. In *Solar Physics*; Springer: Berlin/Heidelberg, Germany, 2002; p. 24.
58. Kolotkov, D.Y.; Broomhall, A.M.; Nakariakov, V.M. Hilbert–Huang transform analysis of periodicities in the last two solar activity cycles. *Mon. Not. R. Astron. Soc.* **2015**, *451*, 4360–4367. [CrossRef]
59. Cumming, A.; Marcy, G.W.; Butler, R.P. The Lick Planet Search: Detectability and Mass Thresholds. *Astrophys. J.* **1999**, *526*, 890. [CrossRef]
60. Mortier, A.; Faria, J.P.; Correia, C.M.; Santerne, A.; Santos, N.C. BGLS: A Bayesian formalism for the generalised Lomb-Scargle periodogram. *Astron. Astrophys.* **2015**, *573*, A101. [CrossRef]
61. Eyer, L.; Bartholdi, P. Variable Stars: Which Nyquist Frequency? *Astron. Astrophys. Suppl. Ser.* **1999**, *135*, 1–3. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.