International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

Efficacy of spectral windows in detecting periodicity of the monthly average sunspot data

Tarek A Elghazali [1,*], Mohammad M Sulayman [2] and Ibrahim M Elbakosh [3]

[1] *Department of Statistics, Faculty of Science, University of Benghazi, Benghazi, Libya.*
[2] *Department, Faculty of Engineering Science, University of Bright Star, Bright, Libya.*
[3] *Department of Health Informatics, Faculty of Public Health, University of Benghazi, Benghazi, Libya.*

International Journal of Science and Research Archive, 2024, 13(02), 2344–2352

Publication history: Received on 28 October 2024; revised on 07 December 2024; accepted on 09 December 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.2.2399

Abstract

This study aimed to determine the effectiveness of different spectral windows, such as Daniel, Tukey, Hamming, Parzen, and Bartlett, in identifying the periodicity of monthly average sunspot data (Wolf-Sunspot, 1961–1980). The goal was to comprehend the recurring patterns of sunspot activity, which is crucial in anticipating solar activity and its potential impact on Earth. By employing these windows, the accuracy of this method in uncovering the underlying periodicity is evaluated.

Keywords: Spectral analysis; Monthly average sunspot data; Periodicity; Spectral windows

1. Introduction

Sunspot activity displays periodic patterns essential for predicting solar activity and its potential impact on Earth. Sunspot data has drawn the attention of many time-series analysts due to its notable feature of having cycles of activity that differ in duration, occurring with an average periodicity of approximately 11 years. This distinctive attribute prompted the scientist Yule to apply the autoregressive method to analyze this data, as well as the scientists Box and Jenkins (1970)(1). Spectral analysis, a frequency-domain analysis of time series, is utilized to understand and uncover these patterns. However, the raw periodogram obtained from particular observations is an asymptotically unbiased estimator for the spectral function but not a consistent estimator, resulting in significant fluctuations in the periodogram graph (2). The lack of consistency is expected in Fourier series representation since N parameters need to be estimated from N observations, regardless of the series size (3). Spectral (Tuckey, 1950; Daniell, 1946; Bartlett, 1946; Hamming, 1989; and Parzen, 1961) windows can be used to estimate the spectral density function of the series, which involves smoothing the periodogram (4–8). This study examines the effectiveness of these windows in evaluating the spectral density function, which identifies the periodicity of monthly average sunspot data (Wolf-Sunspot, 1961-1980). It assesses the accuracy of this approach in capturing the underlying periodic patterns. The performance of the windows is assessed by calculating their standard deviations and comparing them based on the smallest standard deviation value. The window with the smallest standard deviation value is considered to have the best performance.

It is worth indicating that there are several key differences between this study and previous ones:

- We have used a significantly larger dataset, comprising 20 years' average monthly data, which amounts to approximately 240 values. In contrast, previous researchers relied on datasets that did not exceed 100 values.

* Corresponding author: Tarek A Elghazali

- Our analysis takes a different approach (non-parametric) by employing the frequency domain instead of the time-domain method (parametric) used by previous researchers. This allows us to avoid depending on probabilistic models to describe the data (1).
- Our study benefits from the use of advanced statistical packages such as STATISTICA and SPSS software for conducting calculations. These tools were not available to previous researchers studying the same data, enabling us to achieve more precise calculations.

Finally, it is important to note that we standardized the span (the width or range of frequencies covered within a window) of all windows to observe the impact of window shape solely.

1.1. Some Definitions

- **Stochastic Process:** A stochastic process is a family of random variables $\{X_t, t \in T\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$.
- **Stationarity:** The time series $\{X_t, t \in \mathbb{Z}\}$, with index set $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, is said to be stationary if i) $E|X_t|^2 < \infty$ for all $t \in \mathbb{Z}$, ii) $E(X_t) = m$ for all $t \in \mathbb{Z}$, and iii) $\gamma_X(r,s) = \gamma_X(r+t, s+t)$ for all $r, s, t \in \mathbb{Z}$.
- **Strict Stationarity:** The time series $\{X_t, t \in \mathbb{Z}\}$ is said to be strictly stationary if the joint distributions of $(X_{t_1}, \dots, X_{t_k})'$ and $(X_{t_1+h}, \dots, X_{t_k+h})'$ are the same for all positive integers k and for all $t_1, \dots, t_k, h \in \mathbb{Z}$.
- **Stationary up to Order m:** The time series $\{X_t, t \in \mathbb{Z}\}$ is said to be stationary up to order m if all the joint moments up to order m of $(X_{t_1}, \dots, X_{t_k})'$ exist for all positive integers k and for all $t_1, \dots, t_k, h \in \mathbb{Z}$ and equal the corresponding joint moments up to order m of $(X_{t_1+h}, \dots, X_{t_k+h})'$ (2,9).
- **Gaussian Time Series:** The process $\{X_t\}$ is a Gaussian time series if and only if the distribution functions of $\{X_t\}$ are all multivariate normal (10).
- **White Noise:** The process $\{X_t\}$, $t = 0, \pm 1, \pm 2, \dots$, is called a white noise process if it consists of a sequence of uncorrelated random variables, i.e. if for all $s \neq t$, $cov\{X_s, X_t\} = 0$ (11).

2. Spectral Analysis

Spectral analysis is a statistical method used to analyze and interpret trends, cycles, and patterns in a data series over time. This technique is based on Fourier analysis, which decomposes a time series into a set of sinusoidal components (sin and cosine), each with a specific frequency, amplitude, and phase (12). Unlike other time series analysis methods (time-domain), spectral analysis does not specify a parametric model for the data and then estimate this model; instead, it estimates the spectral function (the spectrum) without any prior assumptions, although the estimates have to be adjusted according to the series' properties. The spectral analysis is particularly useful in fields such as economics, meteorology, geophysics, and engineering, where it is essential to understand the periodic behaviour of time series data (2). In addition, it provides a powerful tool for identifying hidden periodicities, detecting them, and understanding the underlying dynamics of a time series (13). However, the use of spectral analysis is subject to certain conditions:

- The time series data must be stationary, meaning that its statistical properties do not change over time (12). This includes the constant mean, variance, and autocorrelation structure. Non-stationary data can often be transformed into stationary data through differencing or detrending.
- The time series should be infinite or at least very long. In practice, this condition is often relaxed, but it may lead to inaccuracies in the spectral estimates (2).
- The time series should be regularly sampled. Irregularly sampled data can lead to complications in the spectral analysis, although methods have been developed to handle such data (14).

2.1. Some Concepts in the Frequency Domain

- **Fourier Series:** Fourier analysis in its crudest form is primarily about finding the best fit of the form of a function by a sum of sine and cosine functions, which is known as Fourier series expansion. Let a function $f(x)$ be defined on the interval $(-\pi, \pi)$, and let it satisfy a certain number of Dirichlet conditions. These conditions guarantee that $f(x)$ is reasonably 'well-behaved,' which means that over the interval$(-\pi, \pi)$, $f(x)$ is absolutely integrable, has a finite number of discontinuities, and possesses a finite number of maxima and minima. Consequently, $f(x)$ can be approximated by the Fourier series $(a_0/2) \sum_{r=1}^{k} [a_r \cos rx + b_r \sin rx]$ where the constant a_0 , a_r and b_r are called a Fourier series coefficients, $a_0 = (1/\pi) \int_{-\pi}^{\pi} f(x)dx$; $a_r = (1/\pi) \int_{-\pi}^{\pi} f(x) \cos rx \, dx, r = 0, 1, 2, \dots$; $b_r = (1/\pi) \int_{-\pi}^{\pi} f(x) \sin rx \, dx, r = 0, 1, 2, \dots$

- It can be demonstrated that this Fourier series converges to $f(x)$ as $k \to \infty$, except at points of discontinuity, where it converges to the midpoint of the step change. To apply Fourier analysis to discrete time series, the Fourier series representation of $f(x)$ needs to be examined when $f(x)$ is defined solely on the integers $1, 2, \ldots, N$. Instead of presenting the formula, we illustrate that the necessary Fourier series arises intuitively by analyzing a straightforward sinusoidal model (2,3).
- **Fourier Transform:** In mathematics, integrals of the form $F(\alpha) = \int_a^b f(x)\, k(\alpha, x)\, dx$ often occur. The function $F(\alpha)$ is said to be the integral transform of $f(x)$ by the kernel $k(\alpha, x)$. Kernels associated with Fourier transforms is given by $F(\beta) = (1/\sqrt{2\pi}) \int_a^b f(\mathrm{x}) \exp{(i\beta x)}\, dx$.
- **The Spectral Density Function:** Let $X(t)$ be a zero continuous parameter stationary process with power spectral density function $h(\omega)$ and corresponding autocovariance function $C(\tau)$. In this case, $h(\omega)$ is the Fourier transform of (τ) $h(\omega) = (1/2\pi) \int_{-\infty}^{\infty} \exp(-i\omega\tau)\, C(\tau)\, d\tau$. If $\int_{-\infty}^{\infty} |C(\tau)|\, d\tau < \infty$ holds and $R(\tau)$ is continuous at $\tau = 0$ (continuity everywhere), $C(\tau)$ can be obtained by using the inverse Fourier transform of $h(\omega)$, i.e. $C(\tau) = \int_{-\infty}^{\infty} \exp(i\omega\tau)\, h(\omega)\, d\omega$

In the case of real valued process, the complex term in the above integral vanishes and we obtain

$h(\omega) = (1/2\pi) \int_{-\infty}^{\infty} \cos\,(\omega\tau)\, R(\tau)\, d\tau$. It is evident that $h(\omega)$ is an even function of ω, and $C(\tau)$ can be derived from the relation $C(\tau) = (1/2\pi) \int_{-\infty}^{\infty} \cos\,(\omega\tau)\, h(\omega)\, d\omega$. The function $h(\omega)$ represents the average contribution to the total power (variance) from components in $X(t)$ within the frequency range ω and $\omega + d\omega$ (15).

Properties of $h(\omega)$: The function $f(\omega)$ has the following properties:

$$\int_{-\pi}^{+\pi} h(\omega) d\omega = C(0) = \sigma_x^2$$

$$h(\omega) \geq 0.$$

$$h(-\omega) = h(\omega)$$

for all ω, i.e. $h(\omega)$ is even function (10).

- **The Periodogram:** The periodogram $I_N(\omega)$ for $X_1, X_2, \ldots, X_n$ for all ω in the range $-\pi \leq \omega \leq \pi$ is defined as follows:
- $I_N(\omega) = [A_1(\omega)]^2 + [A_2(\omega)]^2$, where $A_1(\omega) = \left(\sqrt{2/N}\right) \sum_{i=1}^{N} X_i \cos \omega t$, and $A_2(\omega) = \left(\sqrt{2/N}\right) \sum_{i=1}^{N} X_i \sin \omega t$.
- $I_N(\omega)$ could be written in the form $(2/N)\ |\sum_{i=1}^{N} X_i \exp{(-i\omega t)}\ |$. In addition, $I_N(\omega)$ is defined for all ω in $(-\pi, \pi)$ practice; it is computed only at discreate sets of frequencies.

The periodogram is therefore computed at the frequencies $0, 2\pi/N\,, 4\pi/N\,, 6\pi/N$, i.e.

$$I_p = I_N(\omega_p), \omega_p = 2\pi p/N\,, p = 0, 1, \ldots, [N/2] \text{ (2).}$$

- **The Modified Periodogram:** The function $I_N^*(\omega) = (1/4\pi)\, I_N(\omega)$ where $I_N(\omega)$ is ordinary periodogram, and the function $I_N^*(\omega)$ is known as the modified periodogram due to its close link to the periodogram. Most time series analysts call it just the periodogram. However, the alternative form of $I_N^*(\omega)$ is $(1/2\pi) \sum_{s=-(N-1)}^{N-1} \hat{C}\,(s) \cos(s\omega)$.

In other words $I_N^*(\omega)$ is just the sample version of $h(\omega)$ which justifies calling $I_N^*(\omega)$ the sample spectral density function. It can be shown that the periodogram is an asymptotically unbiased estimate of $h(\omega)$ for large N as follows:

$$E\ [I_N^*(\omega)] = (1/2\pi) \sum_{s=-(N-1)}^{N-1} E\ [\hat{C}\,(s)\,] \cos(s\omega) = (1/2\pi) \sum_{s=-(N-1)}^{N-1} \left[1 - \frac{|s|}{N}\right]\, C\,(s) \cos(s\omega).$$

Using the relation $E\left[\hat{C}(r)\right] = (1/N)\sum_{I=1}^{N-|r|} R\,(r) = [1-(|r|/N)]\,C(r)$. As N becomes large the factor $[1-(|r|/N)] \longrightarrow 1$ for all s. Hence $E\,[I_N^*(\omega)] = (1/2\pi)\sum_{s=-\infty}^{\infty} C(s)\cos\,(s\omega) = h(\omega)$ (16,17).

- **Properties of the Periodogram:** The raw periodogram is generally a poor estimate of the spectral density function for the following reasons:
    - As $N \longrightarrow \infty$, the variance of the periodogram does not go to zero, thus showing that $I_N^*(\omega)$ is not a consistent estimator for $h(\omega)$.
    - For any fixed neighboring frequencies, $\omega_1, \omega_2, cov\,[I_N^*(\omega_1), I_N^*(\omega_2)]$ decrease as N increases, which causes the periodogram to have an erratic and widely fluctuating pattern. Priestley (1981) gave a reasonable expansion for the inconsistency of $I_N^*(\omega)$ by pointing out that $I_N^*(\omega)$ includes all the sample autocovariance extending from $s = 0$ to $N-1$, and hence no matter how large N becomes, $I_N^*(\omega)$ always involves the tail of the sample autocovariance function, which is a poor estimate of the corresponding theoretical autocovariance since in the tail region $C(s)$ is based only on a small number of pairs of observations.

Finally, it is important to note that the periodogram, in its raw form, suffers from a problem known as leakage. This means that the frequencies can "leak" into adjacent frequencies, affecting the accuracy of the analysis.

There are three approaches to address the issue of leakage:

- *Padding the Series:* by padding the data, you can apply a finer frequency resolution to the analysis.
- *Tapering the Series:* by tapering the data before analysis, you can reduce the effects of leakage.
- *Smoothing the Periodogram:* smoothing the periodogram can help identify the overall frequency region that contributes most significantly to the periodic nature of the data.

Shumway & Stoffer (2018) discussed the problem of the leakage and related topics. Hence, To minimize leakage from the periodogram, time series analysts often use an estimator of $h(\omega)$, based on the periodogram, expressed as $\hat{h}_a = (1/2\pi)\sum_{s=-M}^{M} \hat{C}\,(s)\cos(s\omega)$. Here, M is an integer less than $(N-1)$, and its specific value should be determined from the data provided. Moreover, In the subject literature, an estimate $\hat{h}_a$ is known as a truncated periodogram, and M is called the truncation point (2).

2.2. Spectral Windows

Spectral windows are essential tools in spectral analysis, designed to reduce the leakage effect and enhance the accuracy of frequency estimation. In time series analysis, spectral windows play a crucial role in mitigating or minimizing the interference that can result in spectral leakage. Some of the most commonly used windows are the Tukey window, Daniel window, Hamming window, Bartlett window, and Parzen window, each distinguished by its specific features and applications.

Spectral windows play crucial roles in a time series analysis, and their purpose is to mitigate or minimize the interference that would otherwise manifest as spectral leakage. Some of the most used windows are the Tukey window, Daniel window, Hamming window, Hanning window, and Parzen window, each with its own unique features due to its use.

- **Tukey Window:** This is unique in its ability to trade off the frequency accuracy estimation for minimizing spectral leakage. This consists of a rectangular window with a cosine taper, which controls side lobes in effect (19).
- **Daniel Window:** It is used for particular time series predicting applications to enhance the frequency resolution and damp or attenuate noise (5).
- **Hamming Window:** This window is applied in many cases, owing to its capability to reduce side lobes, hence improving the accuracy of the frequency estimates. This is highly used in digital signal processing (7).
- **Bartlett Window:** This triangular window provides a simple yet effective means of reducing spectral leakage. It is often used in applications with desirable computational simplicity (20).
- 5) Parzen Window: Smoothing of data - The Parzen window technique is mostly used in non-parametric density estimation, but it can equally further the cause of increasing the accuracy of frequency estimation. It does so through kernel density estimations that incorporate observational weights, whereby closer observations tend to influence the estimator averages (21).

For more comprehensive information, you may refer to the works of Jenkins & Watts (1968); Priestley (1981); Warner (1998); Chatfield & Xing (2019); and Bloomfield (2004) (1–3,12,23) **Results**

This section presents the main contribution of the paper, which is to evaluate the effectiveness of spectral windows in detecting periodicity in monthly average sunspot data. We used various windows, including Daniel, Tukey, Hamming, Parzen, and Bartlett.

Initially, we checked the stationarity of the series. Upon plotting the series (Fig. 1), it was observed that non-stationarity existed.

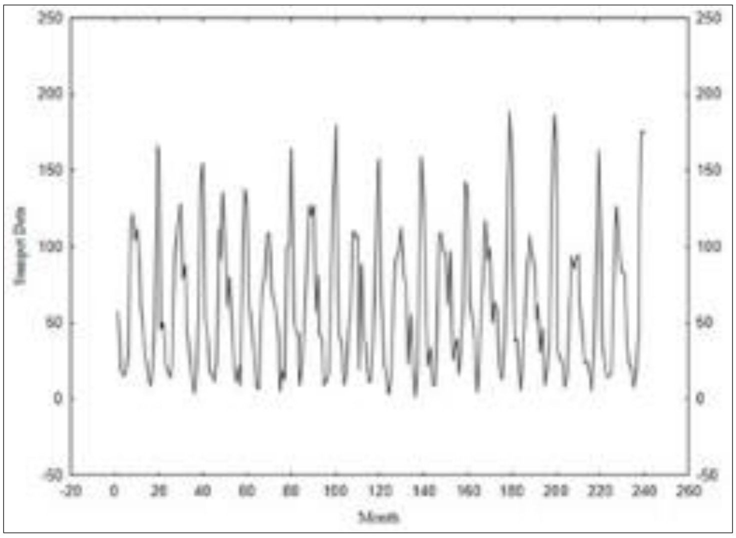

Figure 1 Time Plot of the Monthly Average Sunspot Data

Therefore, we processed the data using the first successive differences until the series became stationary. After ensuring the stationarity of the series, we generated a periodic chart of the variables (Fig. 2). It was observed that some values in the series varied, but this did not necessarily mean that these were the only differences in the series. There might be some other differences that we were unable to observe clearly.

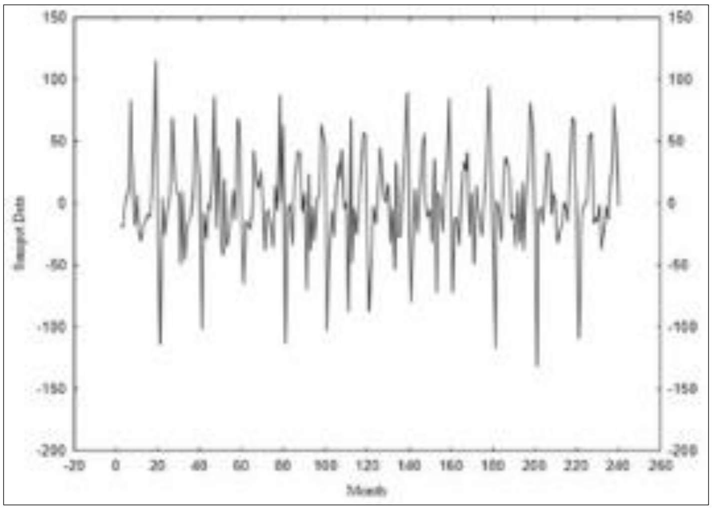

Figure 2 Time Plot of the Transformed Monthly Average Sunspot Data

Therefore, we applied natural logarithmic transformation to the periodic chart, which provided us with more explanations of the differences in the series (Fig. 3).

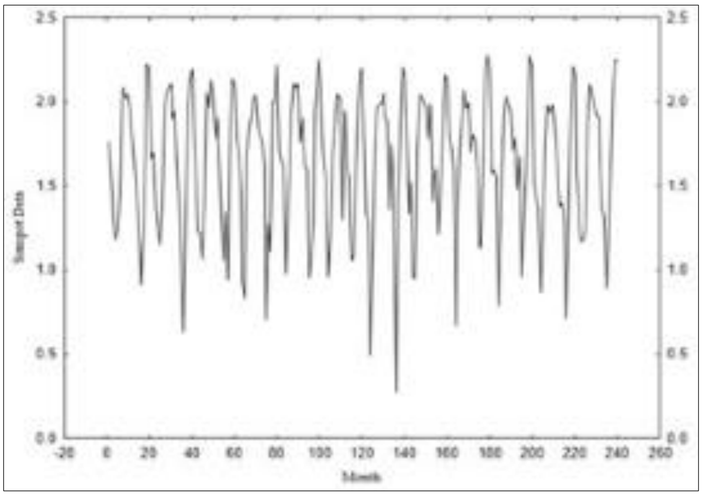

Figure 3 Time Plot of Log10 Monthly Average Sunspot Data

The analysis was conducted using the STATISTICA statistical package. The following tasks were performed:

- Calculation of the mean and standard deviation of the spectral density estimate (smoothed periodogram) and the weights used in smoothing. The smoothing was done using well-known windows in spectral estimation, specifically Daniell, Tukey, Hamming, Parzen, and Bartlett.
- Plotting the unsmoothed (raw) periodogram on a log scale in Fig. 4 and applying smoothing to the periodogram to eliminate random fluctuations and obtain spectral density estimates (Fig. 5).

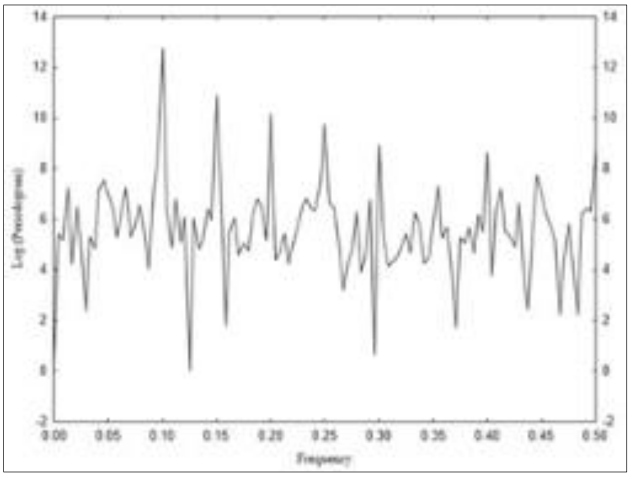

Figure 2 Periodogram of Log10 Monthly Average Sunspot Data

The analysis was conducted using the STATISTICA statistical package. The following tasks were performed:

- Calculation of the mean and standard deviation of the spectral density estimate (smoothed periodogram) and the weights used in smoothing. The smoothing was done using well-known windows in spectral estimation, specifically Daniell, Tukey, Hamming, Parzen, and Bartlett.
- Plotting the unsmoothed (raw) periodogram on a log scale in Fig. 4 and applying smoothing to the periodogram to eliminate random fluctuations and obtain spectral density estimates (Fig. 5).

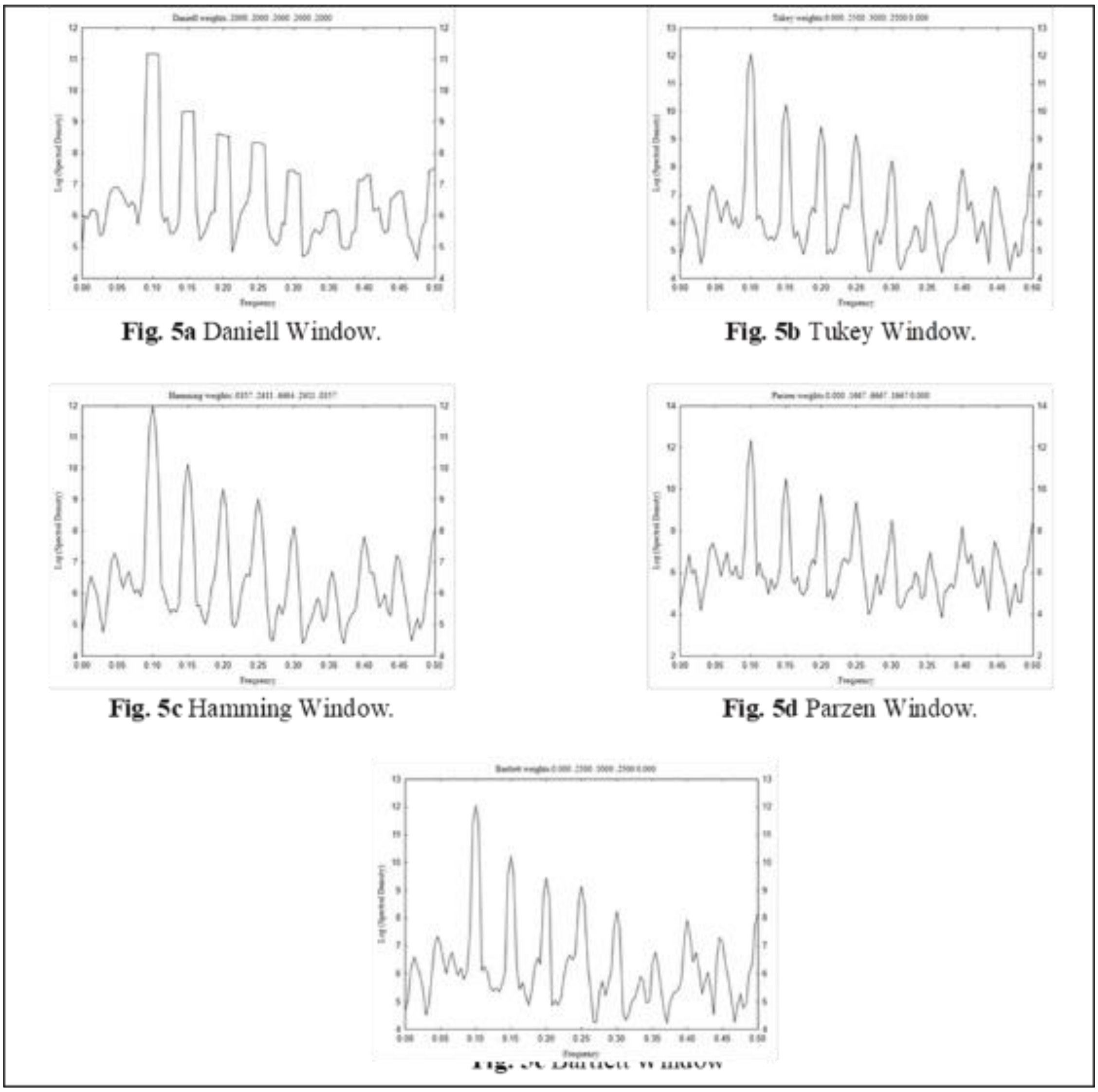

Figure 5 Spectral Density Estimates of Sunspot Average Monthly Data

Table 1 presents the results of spectral density estimates of the monthly average sunspot data for comparison purposes. The summary statistics such as mean value, standard deviation, and weights are organized by spectral windows - Daniell, Tukey, Hamming, Parzen, and Bartlett.

Table 1 Summary Statistics of Spectral Density Estimates of the Monthly Average Sunspot Data

Spectral Windows	**Mean**	**S.D.**	**Weights**
Daniell	0.322289	1.006384	0.2000, 0.2000, 0.2000. 0.2000, 0.2000
Tukey	0.322354	1.399481	0.0000, 0.2500, 0.5000, 0.2500, 0.0000
Hamming	0.322342	1.285780	0.0357, 0.2411, 0.4464, 0.2411, 0.0357
Parzen	0.322440	1.618444	0.0000, 0.1667, 0.6667, 0.1667, 0.0000
Bartlett	0.322354	1.399481	0.0000, 0.2500, 0.5000, 0.2500, 0.0000

3. Discussion

After carefully analyzing our results (Table 1), we have found that the performance of various windows is nearly identical for the monthly average sunspot series. However, it appears that the Daniell's window is slightly better at estimating the spectral density function, as it has a smaller standard deviation. This indicates that the Daniell's window is perhaps more precise and has less leakage compared to other windows. This can be seen in Fig. 5a.

It is worth noting that the spectral analysis function drawings, which used spectral windows, clearly indicated the existence of a peak at iteration $F = 0.1$. This peak corresponds to a 10-month cycle, which means that our study discovered recurring patterns in the monthly sunspot data over 10 months.

4. Conclusion

We conducted study to assess the effectiveness of spectral windows in detecting the periodicity of monthly average sunspot data. Our study indicates that spectral windows are capable of capturing the dominant periodic components of sunspot data. However, the choice of spectral window size plays a crucial role as different sizes capture different periodicities. Further studies are required to optimize the selection of spectral window sizes and explore the multi-scale nature of sunspot periodicity.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Box, G.E.P. & Jenkins GM. Time Series Analysis: Forecasting and Control. Handbook Of Medical Statistics. San Francisco: Holden-Day (revised edn., 1976).; 1970.

[2] Priestley MB. Spectral Analysis and Time Series. 1981;1 and 2(London: Academic Press.).

[3] Chatfield C, Xing H. The Analysis of Time Series: An Introduction with R, Seventh Edition. Seventh Ed. CRC press; 2019.

[4] Tuckey JW. The sampling Theory of Power Spectrum Estimates. In: Symposium on Applications of Autocorrelation Analysis to Physical Problems US Office of Naval Research. 1950. p. 47–67.

[5] Daniell PJ. Discussion on Symposium on autocorrelation in time series. J R Stat Soc Ser C (Applied Stat. 1946;8(1).

[6] Bartlett MS. On the theoretical specification and sampling properties of autocorrelated time-series. Suppl to J R Stat Soc. 1946;8(1):27–41.

[7] Hamming RW. Digital Filters. Third Edit. Mineola, New York: Dover Publications, inc; 1989.

[8] Parzen E. An approach to time series analysis. Ann Math Stat. 1961;32(4):951–89.

[9] Brockwell, Peter J & Davis RA. Time Series: Theory and Methods. Springer; 2009.

[10] Elghazali T. A. On the Use of the Bootstrap Methods in Uncovering the Sampling Distribution of Threshold Value Estimates. (Doctoral Dissertation. A Thesis Submitted to the Faculty of Science). University of the Witwatersrand, Johannesburg, Soth Africa; 2019.

[11] Elghazali T. A. Bootstrapping Some Linear Time Series Models. (Master Thesis. A Thesis Submitted to the Faculty of Science). University of Benghazi, Benghazi, Libya; 2008.

[12] Bloomfield P. Fourier Analysis of Time Teries: An Introduction. Second Edi. North Carolina: John Wiley & Sons, INC; 2004.

[13] Brillinger R. D. Time Series: Data Analysis and Theory. classics edn. Philadelphia: SIAM.; 2001.

[14] Scargle JD. Studies in Astronomical Time Series Analysis.II. Statistical Aspects of Spectral Analysis of Unevenly Spaced Data. Astrophys J. 1982;263:835–53.

[15] Rao, T Subba & Gabr M. An Introduction to Bispectral Analysis and Bilinear Time Series Models. Springer Science & Business Media; 2012.

[16] Hannan E. j. Time Series Analysis. London: Methuen; 1960.

[17] Koopmans LH. The spectral analysis of time series. Elsevier; 1995.

[18] Shumway RH, Stoffer DS. Time Series Analysis and Applications. EZ Edition. Time Series Analysis and Applications. 2018.

[19] Tukey JW. A survey of sampling from contaminated distributions. Contrib to Probab Stat. 1960;448–85.

[20] Bartlett MS. The use of transformations. J R Stat Soc Ser B. 1946;8(2):28–38.

[21] Parzen E. On the estimation of a probability density function and the mode. Ann Math Stat. 1962;33(3):1065–76.

[22] Jenkins GM. Spectral analysis and its applications. Holden-Day, Inc, San Fr Card Nr 67-13840. 1968;

[23] Warner RM. Spectral analysis of time-series data. Guilford Press; 1998.