Pakistan Academy of Sciences

Research Article

Parametric and Non-Parametric Spectral Signal Processing Techniques for Estimation of Periodicity in Sunspot Numbers

Waseem Iqbal*, Muhammad Shoaib, Jameel Ahmad, Muhammad Asim Butt, Abdullah Khalid, and Muhammad Adnan

School of Engineering, Department of Electrical Engineering,
University of Management & Technology, Lahore, Pakistan.

Abstract: Sunspots occur due to magnetic disturbances on the surface of the sun. The sunspot activity effects weather on earth and also affect the earth temperature. In this research paper, various spectral estimation techniques for estimation of universal cyclic behavior of sunspot numbers are discussed. Spectral analysis has been based on two different approaches, namely parametric and non-parametric estimation. The performance of various parametric and non-parametric spectral estimation methods has been compared and frequency of occurrence of sunspots is calculated. MATLAB computer simulations have been extensively used for various estimator settings to arrive at correct results. The results show that the parametric spectral estimation techniques show better and consistent performance as compared to non-parametric spectral estimation techniques.

Keywords: Sunspot, Periodogram, Solar Activity, Blackman-Tuckey, Yule-Walker Method, Parametric and Non-Parametric Spectral Estimation.

1. INRODUCTION

In astronomy two things are of great importance, namely the measurement of the periodicity in the occurrence of the event and the effect of the event on planet earth. Sunspots are astronomical events which appears in the form of dark spots due to magnetic disturbances and solar activity on the surface of sun. These sunspots as shown in Fig. 1 are clearly visible from earth. Original concept was given in 1848 by the Swiss astronomer Johann Rudolph Wolf, who introduced a daily measurement of sunspot number. From the studies, it was revealed that there is a periodic cyclic behavior in the occurrence of these sunspots. Spectrum or periodontics are the main tools to estimate the periodicity of the data. Sunspot also occurs in clusters. Heinrich Schwabe also reported his results in his article [1] about days of observation in a year and number of clusters on sun. Wolf and Wolfer calculated daily sunspot number

by multiplying the number of groups by ten and then adding the product to total count of individual spots. Results, however, vary greatly, since the measurement strongly depends on observer interpretation and experience and on the stability of the Earth's atmosphere above the observing site. Today, sunspot numbers are more relative as many factors influence the accuracy such as; weather condition, place of observation, instruments used for observing solar activity. Sunspot numbers are smoothed and weighted average of measurements is considered from a set of network observations around the planet. Smoothed data is compared with predictions and finally results contain several different types of tables. The relative sunspot number was defined by Wolf in 1856 as $R = K (10g + s)$, where g is the number of sunspot groups and s is the total number of distinct spots. The scale factor K (usually less than unity) depends on the observer and is intended to affect the conversion to
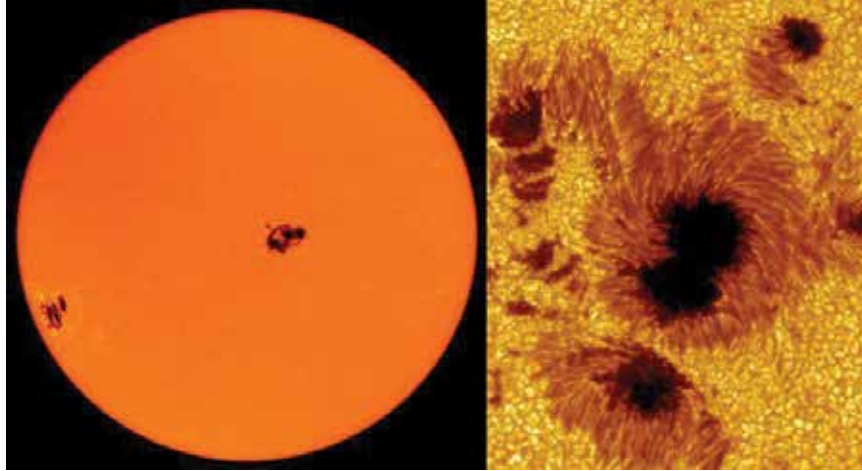
Fig. 1. Sunspots

the scale originated by Wolf.

Interest has grown over time to investigate daily sunspot numbers and study long term solar activity due to coronal mass ejections (CME), magnetic storms and their effect on global climate, weather changes and seismic activity. Investigating sunspot activity on solar surface and determining sunspot cyclic quasi-periodic behavior [2, 4] has many interesting effects on radio communication (RF) and satellite communication. There is a strong evidence that temperature variation on the surface of earth is also due to sunspot activity besides many other factors. Efficient spectrum estimation techniques [5] are available which can be used for estimating sunspot time series. GU Yule (1927) characterized the sunspot numbers as a "disturbed harmonic function" and proposed a parametric model with reference to Wolf sunspot numbers. Neural Network techniques have been used for forecasting sun spot numbers [6, 7] and efficient neural forecaster schemes have been proposed. The effect of sunspot numbers on high frequency radio waves has been investigated recently in [8]. Satellite communication is also affected by solar activities occurring on the surface of sun as it heavily depends on ionospheric stability conditions at various times of the year and seasonal changes that affect high frequencies in GHz range. This research work is an effort to utilize different parametric and non-parametric techniques to estimate the estimate the frequency of occurrence of sunspot numbers.

2. SOURCE OF DATASET FOR SUNSPOT NUMBERS

The daily sunspot number data were obtained from the NOAA's National Geophysical Data Center. NOAA's National Geophysical Data Center (NGDC) [9] provides scientific products, and
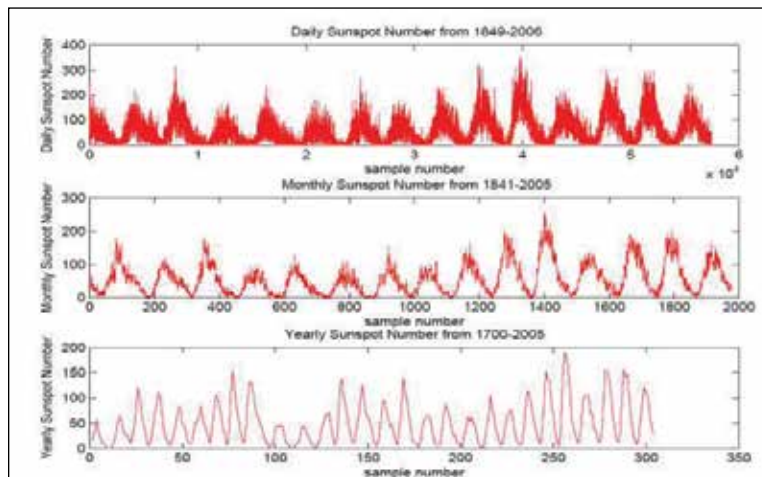
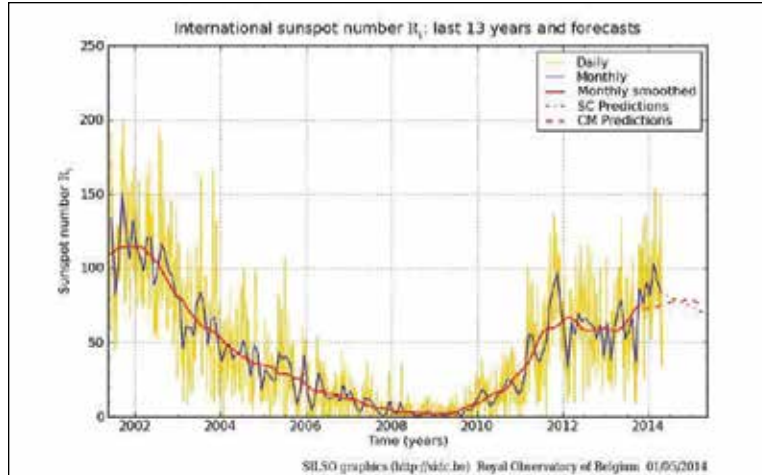

Fig. 2. Data Record for Sunspot Number

Fig. 3. (a) Data Record for monthly and yearly Sunspot Number (1700-2014)
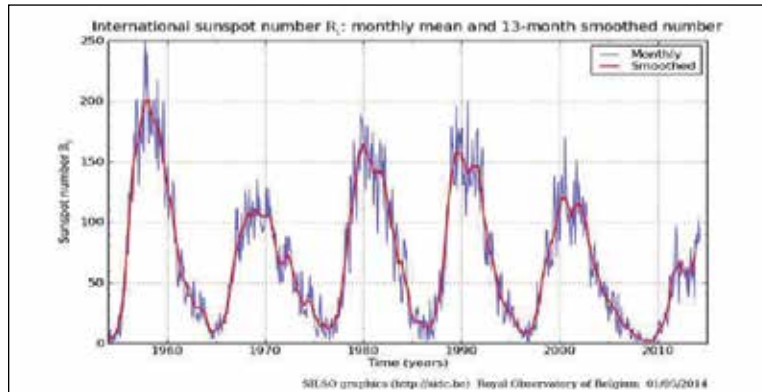

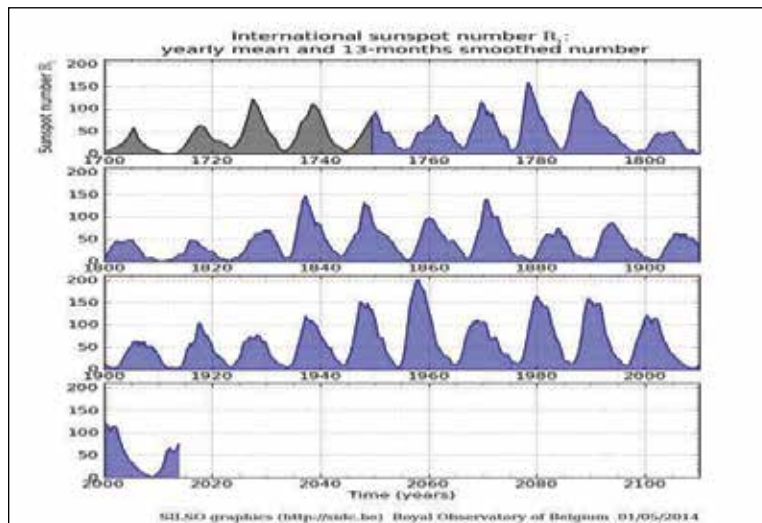**Fig. 3. (b)** Data Record for monthly and yearly Sunspot Number (1700-2014)

Fig. 3. (c) Data Record for monthly and yearly Sunspot Number (1700-2014)

services for geophysical data from the Sun to the Earth. Three sets of these International Sunspot Numbers are available:

  ✓   Daily sunspot numbers (1849-2006).

  ✓   Monthly sunspot numbers (1841-2005).
  ✓   Yearly mean sunspot numbers (1700- 2014).

    The yearly mean sunspot numbers were used to generate the spectral analysis results. More recent

data is also available from 2002 up to year 2014 from Royal Observatory of Belgium (http: //www. sidc.be/silso/.). The data is shown graphically in Fig. 2 and Fig. 3. Daily, monthly and yearly sunspot number for different range of years are plotted in Fig. 2 and Fig. 3.

3. MATERIALS AND METHODS

There are two approaches for analyzing the sunspot data, namely time domain and frequency domain approaches, both are described in the following sections.

3.1. Time Domain Approach

The autocorrelation sequence (ACS) of a stationary random signal [10] depends on the lag (or shift) between two time instants of a signal. The formal definition of ACS or auto-covariance of is given below.

$$r(k) = E\{x(t)x^*(t-k)\} \qquad (1)$$

Where E{.} is an expectation operator.

We expect that ACS estimates produced from N samples of a stationary random signal data should look substantially like ACS estimates over the same range of lag values if estimated from any other segment of N data samples, including the adjacent N data samples. Thus, if non-stationary data is suspected, a simple determination of the maximum N for quasi-stationary statistical behavior is to compare two ACS-estimates-vs-lag plots produced from two adjacent N-point data segments. If the plots are substantially similar, then N point data segments are continuously increased. The comparison of ACS estimates from an adjacent pair of N-point data segments are done until an N value is reached in which significant differences in the ACS estimates can be observed. This will be the threshold value between quasi-stationary and non-stationary statistical behavior.

3.2. Frequency Domain Approach

If the ACS is stationary, then the Fourier transform (FT) of the ACS (or the power spectral density (PSD)) is also stationary. In this case, the shape of the plotted PSD, like the ACS plot, will not change over time. If time varying frequency content is observed, then the random signal has non-stationary statistical

behavior. Again, one could experimentally vary the analysis interval duration N until differences in successive PSD estimate plots are noted, indicating the threshold between quasi-stationary and non-stationary statistical behavior. Using the sunspot, an experimental determination through MATLAB scripts was made. k=250 samples were the threshold between quasi-stationary and stationary statistical behavior. The time-vs-frequency gram was created by using a sample spectrum (magnitude of FFT) for each analysis intervals of 250 samples, with 125 sample overlap is shown on the next pages.

3.3. Problem Statement

The power spectral density (PSD) of a zero-mean stationary stochastic process is defined as [10]:

$$P_x\left(e^{j\omega}\right) = \sum_{-\infty}^{+\infty} r_x(k)e^{-jk\omega} \qquad (2)$$

Our problem is to find an estimate of the power spectrum, of a discrete processsfrom a finite record of observations where n=0 to N-1 of a single realization of sunspot numbers.is auto-correlation function of. Different parametric, non-parametric and frequency estimation approaches are used to obtain a good spectral estimate for sunspot numbers. The Periodogram is used as the benchmark to compare the results of the sunspot cycle period estimate by various techniques.

3.4. Analysis and Results Using Non-Parametric Methods

The simplest way to estimate the power spectrum of a signal is to use non-parametric methods which are also known as window-based methods. Window-based methods are the most fundamental type of spectral estimation approaches without requiring signal modeling. One of the most attractive properties of the window-based methods is their simplicity for implementation. However, these methods suffer from low frequency resolution especially when the length of observation data is short. Non- Parametric spectral estimation techniques are:

- ✓ The Periodogram
- ✓ Biased and Unbiased autocorrelation estimate
- ✓ Welch's Method: Averaging Modified

Periodogram
- ✓ Blackman-Tukey   approach: Periodogram smoothing

3.4.1. The Periodogram

The Periodogram is based on the power spectrum for finite data x(k), k = 0, 1… N-1. In Fourier domain, the Periodogram P can be expressed as

$$P_{per}\left(e^{j\omega}\right) = \frac{1}{N}\left|\sum_{k=0}^{N-1} x(k)e^{-jk\omega}\right|^{2} \tag{3}$$

The sunspot data is noisy and has a DC offset and has non-zero mean. So the first step was to remove the mean and filter the data. The Periodogram of our monthly sunspot data is shown in Fig. 4 and results are described in Table 1. A

large 0 Hz DC component due to positive numeric property of sunspot number is apparent in Fig. 4. In order to study the spectral estimation further it was necessary to remove the large DC component and to remove the noise from the data. The sunspot

Table 1. Result of the Periodogram Spectral Estimate

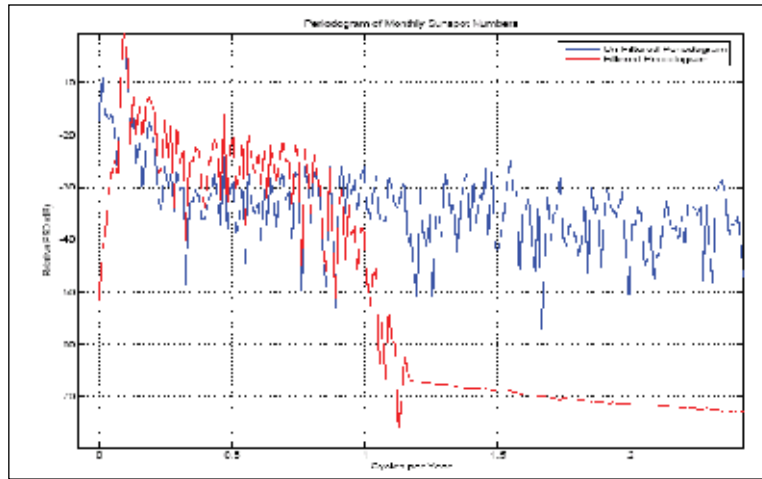|	
Sampling Frequency, Fs	12 samples/year
Low pass Filter Order	66
High pass filter order	148
Cycle revealed from Periodogram	10.667 year/cycle

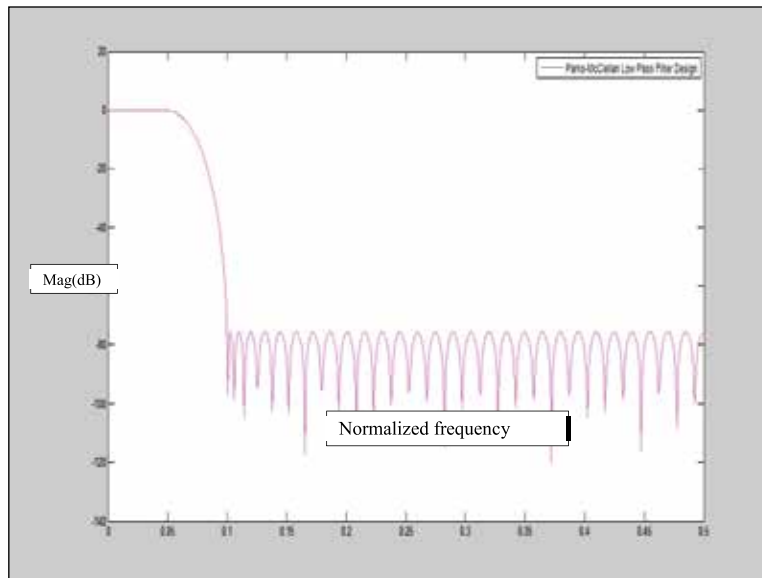**Fig. 4.** Periodogram: before filtering and after filtering

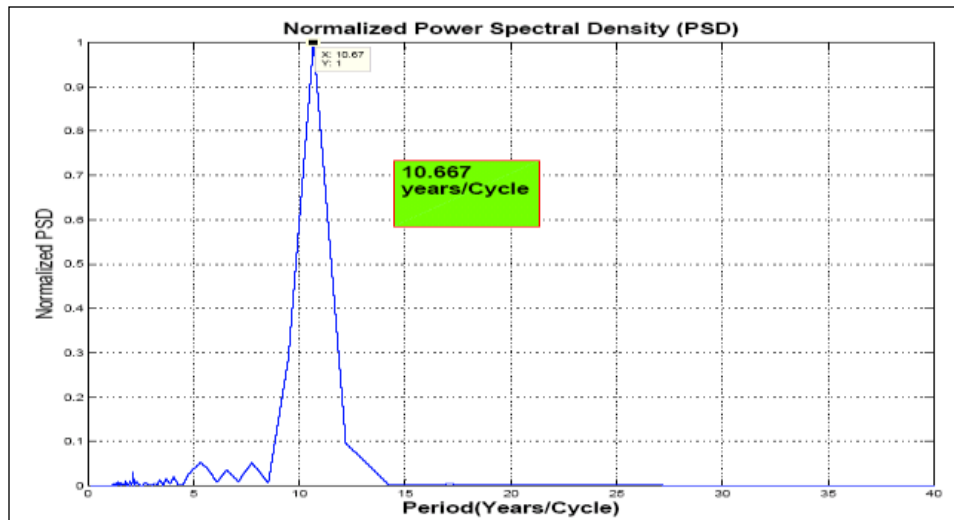**Fig. 5.** Parks-McClellan Low pass filter (Order 66)

Fig. 6. Cycle revealed from Periodogram

numbers were filtered using low pass and high pass filters. Linear phase filters [11, 12] using Parks-McClellan filter structures as shown in Fig. 5 were used to realize low pass and high pass filters. MATLAB functions firpmord and firpm were used for simulation. The classical 11-year sunspot period frequency of 0.09 cycles per year is also clearly apparent in Fig. 6.

3.4.2. Biased and Unbiased Autocorrelation Spectral Estimate

Both biased and unbiased autocorrelation function were obtained from sunspot data. The autocorrelation function shown in Fig. 7 was generated for stationary analysis of the sunspot number time series and for observations regarding this series' periodicity. The autocorrelation function for the sunspot number data clearly demonstrates periodicity with small cycles inside of larger cycles. Looking closer at the period of the smaller cycles, 11-year period can be clearly seen in the autocorrelation and monthly mean sunspot number plots. The autocorrelation estimate does not decay significantly for large correlation lag and so memory of the time series does not disappear quickly. Fig. 7 also shows that biased estimate is more trustworthy
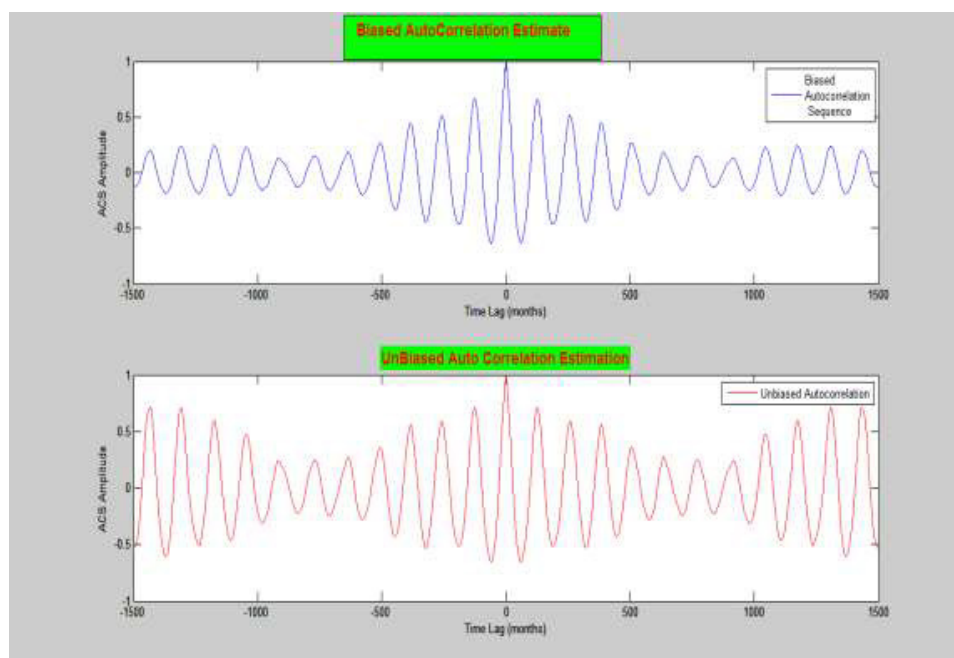

Fig. 7. Biased and un-Biased Autocorrelation Sequence

than unbiased estimate.

3.4.3. The Welch-Bartlett Method: Averaging Modified Periodogram

There are two methods available if we modify the original Periodogram. Any window can be used other than rectangular window. The modified Periodogram can be [10] expressed as

$$P_{Mod}\left(e^{j\omega}\right) = \frac{I}{NU}\left|\sum_{n=-\infty}^{\infty} x(n)w(n)e^{-jn\omega}\right|^2 \qquad (4)$$

Where $U = \frac{1}{N}\sum_{n=0}^{N-1}|w(n)|^2$, N is the length of the window

Bartlett's method uses Periodogram averaging. However, the difficulty of this approach is that uncorrelated realizations of a process are generally not available. is partitioned into K non-overlapping

sequences x$_i$ (n) of length L, where $N = KL$.

$$P_B((e^{j\omega}) = \frac{1}{N}\sum_{i=0}^{K-1}\left|\sum_{n=0}^{L-1} x(+iL)e^{-jn\omega}\right|^2 \qquad (5)$$

Welch method is a refined version of Bartlett method. In Welch method data window is overlapped and a windowing function is also applied on data segment prior to computation of Periodogram. Smooth Periodogram as shown in Fig. 8 is achieved but the period estimate was approximately the same as the Autocorrelation estimate. The percentage of overlap changed the period estimate slightly but the best results were with overlaps of 40% to 50%. The period estimate was sensitive to segment length and segment length values of about 10% (=200/1977) of the data length resulted in the best period estimate.

Table 2. Parameter Setting and Result of the Welch-Bartlett Spectral Estimate

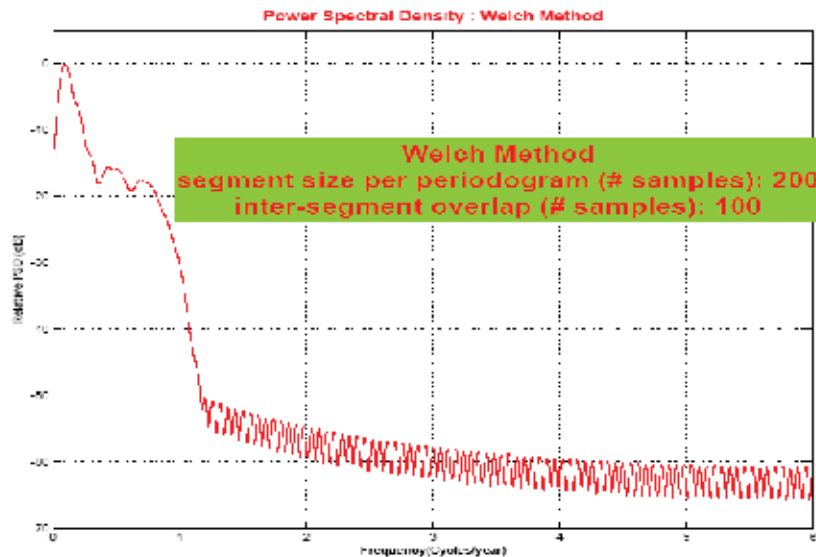| **Spectral estimator parameters:** |
|---|
| Window choices: 0--none, 1--Hamming, 2--Nuttall |
| Enter number of window choice: 1 |
| Enter segment size per Periodogram (# data samples): 200 |
| Enter inter-segment overlap (# data samples): 100 |
| Performance with lag (L=200): 50% overlap |
| K= (2N/L)-1=18.77~ 19 sequences of length L |
| Resolution= (1.28*2*pi)/L=0.02 |
| Variability=9/16*(L/N) =0.056 |
| Revealed period=0.09961 corresponds to 10.039 years/cycle |

Fig. 8. Welch-Bartlett Power spectral estimate

As the segment length increased, the Welch method returned results similar to the Periodogram. Bartlett, Hamming and Nutall windows were used to arrive at the period estimate. Final Estimator settings for MATLAB simulations are shown belowand results are summarized in Table 2.
Fs =12,
N=1977 (data samples),
Window=Bartlett
$F_{nyquist}$=6, T =0.0833
num_psd =2048.

3.4.4. The Blackman-Tukey (BT) Method
Classical Blackman-Tukey auto/cross spectral estimate based on the Fourier transform of the auto/cross correlogram (ACS or CCS estimate) has been used. This method is also called Periodogram smoothing and uses a window function to in order to decrease the contribution of unreliable estimate to the Periodogram.

The PSD via Blackman-Tukey can be calculated using expression in eq. (6).

$$P_{BT}\left(e^{j\omega}\right) = \sum_{k=-(M-1)}^{M-1} w(k)r_x(k)e^{-jk\omega} \qquad (6)$$

The power spectrum has smaller variance. However, there is some resolution loss since a small number of autocorrelation estimates are used to form the estimate of the power spectrum. The Blackman-Tukey method of smoothing the Periodogram is shown in Fig. 9. This has resulted in a lower sunspot cycle period estimate than the Periodogram but worse than the modified Periodogram. Various ACS

lags which control window size were tried but the period estimate became worse when the window length exceeded 20% of the data length.

4. ANALYSIS SETTINGS

We have used the following setting for different estimators:

- ✓ Autocorrelation Temporal Sequence Estimate. Both biased and unbiased ACS terms out to lag 500-800 months where it is small but there is strong evidence of cyclic behavior and periodicity.
- ✓ Blackman-Tukey spectral estimate using the 150 ACS estimates with 50% lag. Hamming window is applied to suppress side-lobe artifacts. Parks-McClellan Low-pass filter Order is 66 and high-pass filter order is 120.
- ✓ Autoregressive spectral estimate using modified covariance method gives best resolution and closes at order 48 terms; low order reduces spectral resolution details too much
- ✓ Minimum variance spectral estimate using 60th order
- ✓ Pisarenko spectral estimate uses 20 sinusoids and calculated harmonic frequencies and signal power contribution at these frequencies and is considered the best estimate in Eigen Analysis techniques.

The time-frequency (TF) grams are also shown for non-stationary analysis. The various estimator settings are shown in Fig. 10. The TF gram clearly shows the dominant period at around
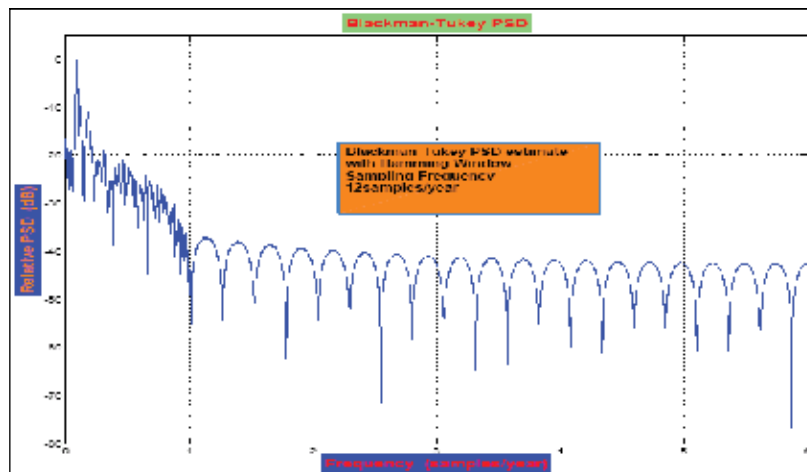

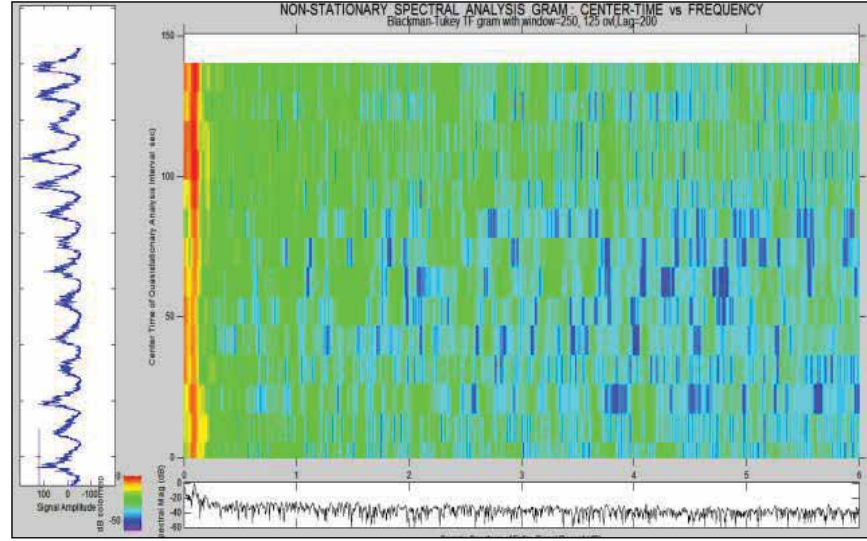**Fig. 9.** Blackman-Tukey PSD Estimate

Fig. 10. Blackman-Tukey 1D Time-Frequency (TF) grams

0.09375 cycles per year with adjacent harmonics of lower magnitude.Final estimator setting for non-stationary analysis:

Interval=200, overlap=50%, Lag=150
Final Estimator setting for stationary analysis
Interval duration (# samples): 1977
Overlap (# samples) between analysis intervals: 0
Spectral estimator parameters:
Window choices: 0--none, 1--Hamming,
2--Nuttall
Enter number of window choice: 1
Enter maximum lag (# samples) for estimated ACS: 300
Dominant Revealed period =0.09375cycles/year (10.667 years/cycle)
2nd revealed cycle=0.1875 cycles/year (5.33 years/cycle) of lower magnitude.

5. RESULTS AND ANALYSIS (USING PAREMETRIC METHODS)

The non-parametric methods for spectrum estimation rely on Discrete Fourier Transform (DFT) and consequently exhibit a frequency resolution where N is the length of the data segment. Increasing the data segment length is usually not possible if data has been collected over a short finite interval or the signal may only be considered stationary over limited intervals. Parametric methods rely on an underlying model that adequately describes the generation of the sampled data. The spectrum calculated from the model produces a high-resolution, smoothed spectrum due to the structure imposed by the model. However, the success of such methods very much depends on how accurate the model is, how accurate the model parameters can be estimated, and how sensitive the model estimation is to perturbations (e.g. noise) present in the data or deficiencies in the model (e.g. incorrect order of model). There are several parametric techniques that may be used to estimate the all-pole parameters. Parametric methods are classified as Autoregressive All-Pole model called AR(p) where p is the order, All-zero Moving average MA(q) process and Auto Regressive-Moving-Average (ARMA) process usually denoted by ARMA (p, q) process. We will mainly concentrate on Auto Regressive process AR (p).

5.1. Autoregressive All-Pole Model

An Autoregressive spectrum estimation requires that an all-pole model be found for the process. The objective is to find the coefficients, a_p, in the AR process equation

$$P_{AR}\left(e^{j\omega}\right) = \frac{|b(0)|^2}{\left|1 + \sum_{k=1}^{p} a_p e^{-jk\omega}\right|^2} \tag{7}$$

There are 5 different method used for AR parameter estimation:

a. The Yule-Walker method (Autocorrelation method)
b. The Lattice (Geometric) method
c. The Lattice Burg method

d. The Covariance method
e. The Modified Covariance Method

5.2. Autoregressive Spectral Estimation by Yule-Walker Method [10]

In this method of all-pole modeling, the AR coefficients, $a_p(k)$ are found by solving a set of

normal equation. $R_p a_p = \varepsilon_p$

Where R_p is an autocorrelation matrix and has Toeplitz structure, so we can use the Levinson-Durbin recursion to solve for the coefficients $a_p(k)$. In this method autocorrelation is first estimated using some windowing function, so the data is extrapolated and produces less resolution than the other methods that do not employ any window. Biased autocorrelation estimate is also preferred over non-biased estimate. If the AR order is too high or over-modeled, then a phenomenon called line splitting may result producing various artifacts resulting in multiple peaks. So it is necessary to estimate an order closing to arrive at a good estimate. The results of this order closing are shown for Yule-Walker method and the power spectrum is shown in Fig. 11 through Fig. 14. The reader is referred to ref. [10, 16] for further details on various techniques. We have omitted certain information due to space limitation and are presenting only the results. There is a strong periodic behavior:
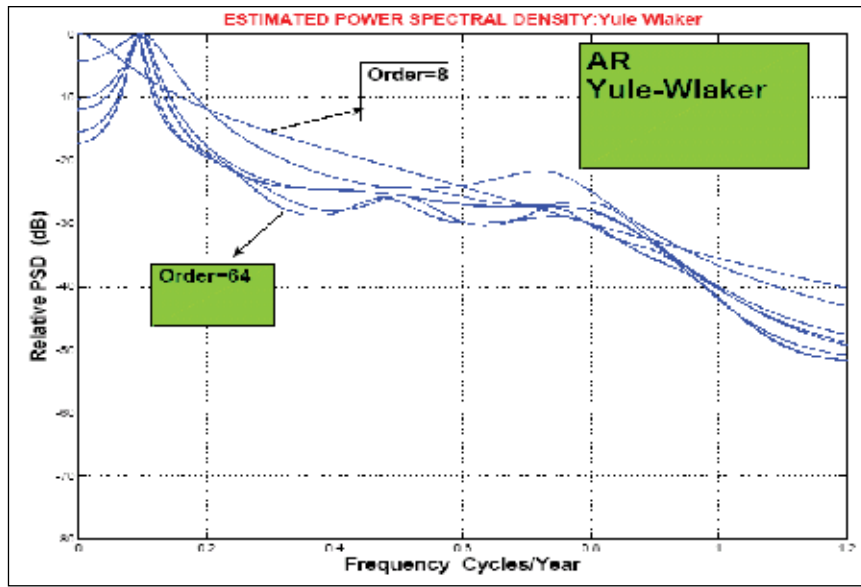

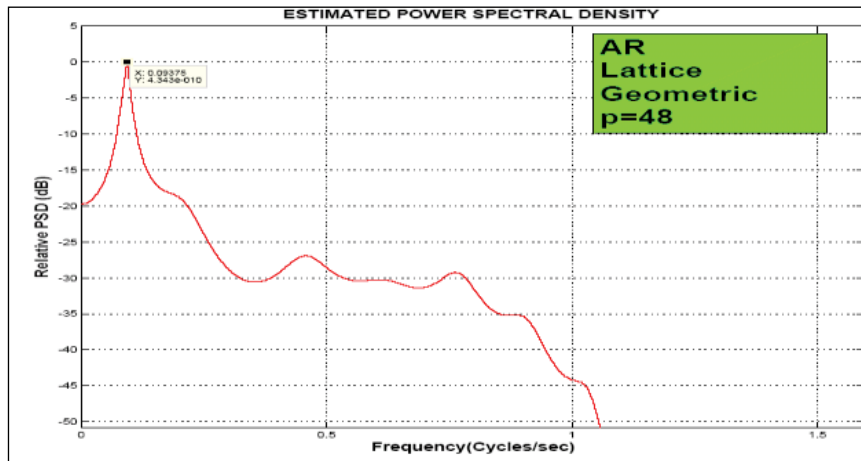**Fig. 11.** Yule-Walker Method for AR spectral estimate

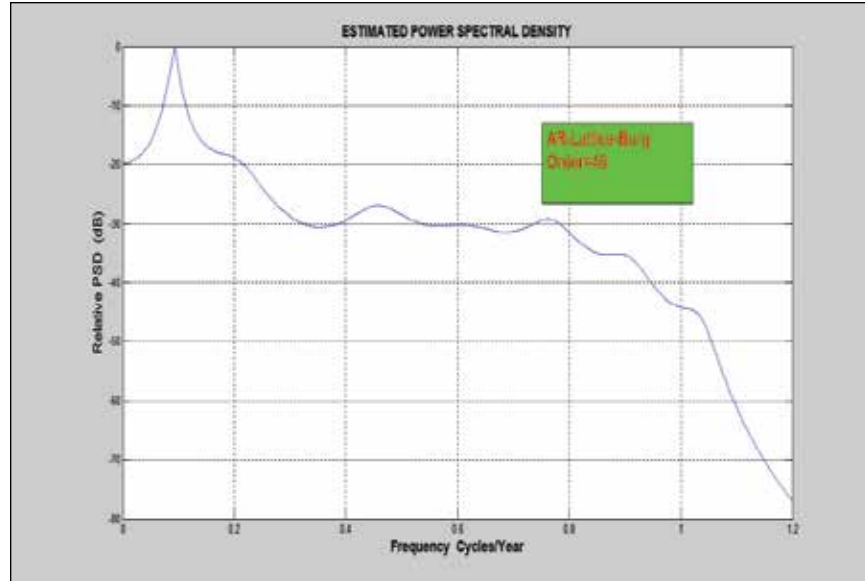**Fig. 12.** AR spectral estimate: The Lattice (Geometric) method

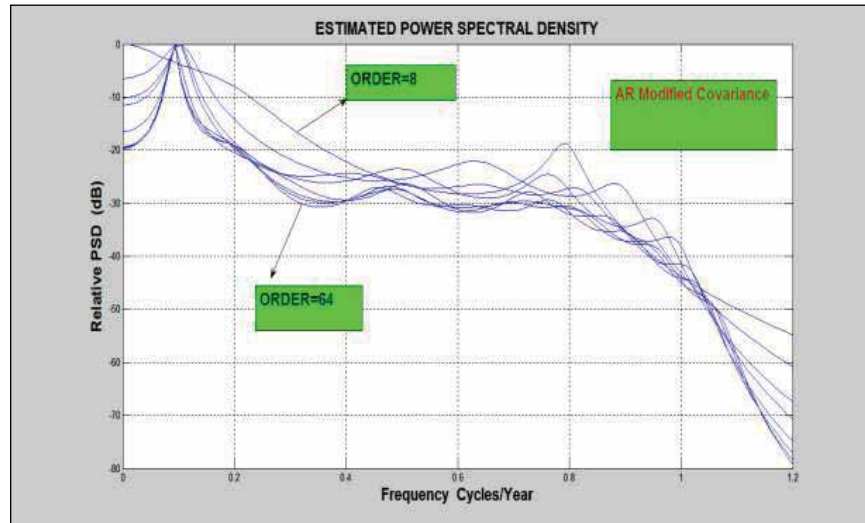**Fig. 13.** Lattice Burg Algorithm PSD estimate

Fig. 14. AR Modified Covariance PSD estimate

Table 3. Performance Comparison of Non-Parametric and Parametric Methods

Non-Parametric Methods	First Dominant period (cycles/year)	Years/cycle
Periodogram	0.09433	10.667
Blackman -Tukey	0.09375	10.66
Welch Method	0.09961	10.3
Min-Variance	0.09523	10.5
Parametric Methods	**First Dominant period(cycles/year)**	**Years/cycle**
Yule-walker	0.09375	10.666
Lattice-Geometric	0.09375	10.666
Lattice-Burg	0.09375	10.666
Modified Covariance	0.09375	10.666

- ✓ Revealed cycle =126/11=10.5 year/cycle
- ✓ Harmonic 1=259/12= 21.583 year/cycle
- ✓ Harmonic 2=385/12=32.083year/cycle

6. CONCLUSION

The sunspots occur due to magnetic disturbances on the surface of the sun and affect the temperature and weather on our earth. It is necessary to find the periodicity by using various parametric and non-parametric techniques. Results of various spectral estimates are shown in Table 3. From this table, it is clear that results of parametric estimation techniques are better than non-parametric estimation techniques of power spectrum estimation. Frequency of occurrence of sunspot numbers from non-parametric methods vary from 10.5 to 11.1 years/cycle whereas from parametric methods this number is approximately 10.66 year/cycle. From this research, it is concluded that parametric methods are a preferred choice for estimating the periodicity and PSD of sunspot numbers. Accurate prediction of sunspot numbers and their cyclic behavior can help us in predicting the accuracy of weathers conditions, design of satellite communication systems, predicting magnetic storms and their effect on global climate and seismic activity.

7. REFERENCES

1. Schwabe, S. H. Solar observations during 1843. *In: Astronomische Nachrichten*, 20(495): 234–235 (1843).
2. Schuster, A. On the Periodicities of Sun spots. *Philosophical Transactions of the Royal Society*. 206: 69–100 (1906).
3. Yule, G.U. On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers. *Philosophical Transactions of the Royal Society*. 226: 267–298 (1927).
4. Le, G. M. & J. L. Wang. Wavelet Analysis of Several Important Periodic Properties in the Relative Sunspot Numbers. *Chinese Journal of Astronomy and Astrophysics*. 3: 391–394 (2003).
5. Hayes, M. H. Statistical Digital Signal Processing and Modeling. John Wiley and Sons, Inc, (1996).
6. Samin, R.E., M.S. Saealal., A. Khamis., S. Isa., R.M. Kasmani. Forecasting sunspot numbers with Feedforward Neural Networks (FNN) using 'Sunspot Neural Forecaster' system. *2011 International Conference on Electrical, Control and Computer Engineering (INECCE)*. 1-5 (2011).
7. Guan, X., L. Sun., F. Yu & X. Li. Sunspot number time series prediction using neural networks with quantum gate nodes. *2014 11th World Congress on Intelligent Control and Automation (WCICA)*. 4647-4650 (2014).
8. Malik, R.A., M. Abdullah., S. Abdullah., M.J. Homam. The influence of sunspot number on high frequency radio propagation. *2014 IEEE Asia-Pacific Conference on Applied Electromagnetics (APACE)*. 107-110 (2014).
9. http://www.ngdc.noaa.gov/stp/SOLAR   [accessed on July 1, 2017]
10. Stoica, P. & R.L. Moses. Spectral Analysis of Signals. Prentice Hall, 1st edition, (2005).
11. Shon, S. & K. Mehrotra. Performance comparison of autoregressive estimation methods. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '84)*. 9: 581–584 (1984).
12. Mitra, S.K. Digital Signal Processing: A Computer-Based Approach. McGraw-Hill Science/Engg. / Math, 2nd edition (2001).
13. Manolakis, D.G., V.K. Ingle & S.M. Kogan. Statistical and Adaptive Signal Processing. Artech House (2005).
14. Bounar, K. H., E.W. Cliver., & V. Boriakoff. A prediction of the peak sunspot number for solar cycle 23. *Solar Physics*. 176: 211 (1997).
15. Cliver, E.W., V. Boriakoff., & K.H. Bounar. The 22-year cycle of geomagnetic and solar wind activity. *Journal of Geophysical Research,* DOI: 10.1029/96JA02037 (1996).
16. Du, Z. L. & H. N. Wang. The relationships of solar flares with both sunspot and geomagnetic activity. *Research in Astronomy and Astrophysics*. 12(4): (2012).
17. Liu, G. Using entropy to improve the resolution in non-parametric spectral estimation. *Electronic Theses, Treatises and Dissertations. Paper.* 7472: (2013).