TIME SERIES ANALYSIS OF LONG TERM FULL DISK OBSERVATIONS OF THE Mn I 539.4 nm SOLAR LINE

S. Danilović¹, I. Vince¹,², N. Vitas² and P. Jovanović¹

¹Astronomical Observatory, Volgina 7, 11160 Belgrade 74, Serbia and Montenegro
²Department of Astronomy, Faculty of Mathematics, University of Belgrade
Studentski trg 16, 11000 Belgrade, Serbia and Montenegro

(Received: March 14, 2005; Accepted: April 18, 2005)

SUMMARY: The equivalent width and central depth data of Mn 539.4 nm solar spectral line, observed in the period from 1979 to 1992 at Kitt Peak Observatory, was analyzed in pursuit for periodic changes. As the observations are highly unevenly sampled, tests if the periods really exist in the observed data were needed. Two different methods for spectral analysis were applied to synthesized data sampled in the same way as observations. Comparison of these results with results obtained from the observed data showed that the parameters display at least three periodic changes with the periods of: 11-years, quasi-biannual and 27-days.

Key words. Sun: activity – Methods: statistical

1. INTRODUCTION

Solar activity covers a range of phenomena at time scales ranging from seconds and minutes (solar flares and solar mass ejections), through months (evolution of active regions), to the 11 years activity cycle. Temporal changes at different levels of solar atmosphere are reflected well in changes of corresponding activity indices. Establishing the existence of real periods in solar observational data and relating them to real physical phenomena has been studied for a long time, but still many periods remain unexplained.

One of the unexplained phenomena is the variability of the parameters of the Mn I 539.4 nm spectral line with solar activity which is recognized on the basis of full-disk observations at Kitt Peak Observatory (Livingston and Wallace 1987). Changes of 1.12% and 1.05% are found in equivalent width and central depth of the line respectively (Vince and Erkapic 1998). Based on observations of the line performed at Astronomical Observatory in Belgrade relative variations of 1.4% and 2.3% of the same parameters with the solar activity were obtained (Danilovic and Vince 2004). In this study, we search for periodic changes of these parameters at different time ranges. This will enable us to understand what kinds of processes are involved and what is their contribution in change of the line parameters.

For this purpose, taking into account time sampling of observations, we have chosen two methods for spectral analysis that are frequently used and known to give the best results in recovering harmonic content of the gapped time series (Carbonell et al. 1992). One of them is Lomb-Scargle modified periodogram (Scargle 1982, Horne and Baliunas 1986), the tool of power spectrum analysis. It has been widely used in analysis of various solar activity indices (for example Özguc et al. 2002). The other is the CLEAN algorithm (Roberts 1987) which proved to be very good at revealing real frequencies in a messy spectrum of synthetic data with or without noise (Roberts 1987, Baisch et al. 1999), or applied...
on real data (Carbonell et al. 1992., Crane 2001). To check if found periodic signals are really present in observations, we applied these methods on both synthesized (artificial) and real data and compared the results.

2. TIME SAMPLING OF THE Mn DATA

Observations of the Mn I 539.4 nm solar spectral line in the full disk irradiance spectrum are performed in period 1979-1992 at Kitt Peak Observatory with the the double-pass spectrometer (13.5 m focal length) at the MacMath Telescope. Due to specific illumination of the spectrograph, the resolving power is reduced to about 60,000 (Livingston et al. 1997). The scintillation noise is the dominant random error that provides a signal to noise ratio of about 1000 (Livingston 1992). The equivalent width and central depth of the spectral line are determined (Livingston 1997). Data sets of these parameters (hereafter Mn data) cover period of 4993 days, with frequency of a few days per month. The minimal sampling rate is 1 day, but the evenly sampled series are maximum 4 days long, with gaps between them going from 2 to 81 days. In this way the period of 4993 days is covered with only 461 observational days. We analyzed how this data distribution affects the power spectrum using the sampling function. The sampling function is defined as a sum of Dirac delta functions with the same time distribution as in observations (Roberts 1987). If data set consists of N values sampled at times $t_r$, $r=1,N$, Fourier transform of sampling function will be:

$$W(\nu) = \frac{1}{N} \sum_{r=1}^{N} e^{-2\pi i \nu t_r}. \quad (1)$$

This is the spectral window function. Convolution of this function with Fourier transform of real function of time gives the spectrum of observed time series. How many spurious peaks will arise and how much smearing will be introduced in power spectrum is controlled by window function. In the case of evenly sampled data window function is proportional to sinc function with main lob determined by the time interval covered by the data. The width of the main lob defines the frequency resolution of spectrum and the existence of sidelobs as a result has a spectral leakage from one frequency to other in its vicinity. The absolute value of a spectral window function of a data set is shown in Fig. 1a. The data set is evenly sampled and covers the same time interval of 4993 days as the Mn data. Because of daily sampling the maximal frequency which can be recovered is 0.5 day$^{-1}$, one half of sampling rate and the width of the main lob is $2 \times (1/4993)$. If, for example, 10 successive points of every 30 are removed from evenly sampled data, window function will be as it is presented on Fig. 1b. In this case, besides the main peak there are peaks at frequencies $\pm k/30$ day$^{-1}$ where k is an integer. Intensities of the sidelobs are modulated by the abs(sinc $(2\pi \times 20)$) function. This window function will produce false peaks in power spectrum that will be displaced from the real peak by $\pm k/30$ day$^{-1}$. Further, if we randomly chose and remove 3000 points from previous data, various peaks will appear in window function (Fig. 1c). Their intensities depend on number of data removed. This will result in very noisy spectrum.

![Fig. 1. Window functions: evenly sampled data (a), data with 10 of 30 points missing (b), data with 10 of 30 point missing plus around 3000 more randomly chosen (c). Mn observations (d).](image-url)
Fig. 1d displays the window function of the Mn observations. By comparing it with the window function in Fig. 1c it can be concluded that besides randomness in missing data there is at least a monthly period in the Mn data sampling because of four prominent peaks near the main lob at approximately $\pm 1/30$ and $\pm 2/30$ day$^{-1}$. The presence of some less intensive peaks in their vicinity suggests that there are more periodicities in observational time sampling. These peaks will produce satellite peaks in spectrum and number of little peaks will generate very noisy spectrum with intensive spectral leakage.

3. TEST OF SPECTRAL ANALYSIS TECHNIQUES

In order to test if recovered periods in the Mn data spectrum are real or not, synthetic data are generated with periods and amplitudes similar to ones find in equivalent width of the Mn I 539.4 nm spectral line (MNW data):

$$y(t) = A_0 + \sum_{i=1}^{3} A_i \cos(2\pi(t - \langle t \rangle)\nu_i), A_0 = 79. \quad (2)$$

Chosen frequencies are 0.00025, 0.00115 and 0.03666 day$^{-1}$ with amplitudes 0.4, 0.086 and 0.048 respectively. All cosinusoids are shifted so that synthetic data look like data set of equivalent widths of the Mn 539.4 nm spectral line. The $\langle t \rangle$ is the mean sample time, calculated as an average of sample times. We will call this data set Synth data. Synth data are computed only for days covered by the Mn data. Thus, the window function is the same as for real data and its influence on known periods can be analyzed. It was especially interesting to see if periodicity due to the rotation of the Sun can be identified. As an illustration what classical methods would recover from Mn data we linearly interpolated Synth data. In that way evenly sampled data are produced with sampling interval of 5 days. To obtain power spectrum, Maximal entropy method (MEM) is used on interpolated synthetic data. Result is shown in Fig. 2a. Largest peak is shifted from 0.00025 day$^{-1}$ to 0.000279 day$^{-1}$. Period at 0.00115 day$^{-1}$ is in its place but there are no peaks in vicinity of frequency of the third signal. Instead, a small one is at 0.0038 day$^{-1}$.

Results of Lomb-Scargle modified periodogram (Scargle 1982, Horne and Baliunas 1986, Press et al. 1992) are presented in Fig. 2b. This method is a generalization of the Discrete Fourier transform that can be applied on unevenly sampled data. For time series $x_i = x(t_i), i = 1..N$ the normalized periodogram at the angular frequency $\omega$ is calculated:

![Fig. 2. Spectral analysis of synthetic data set without the noise: linear interpolation plus MEM (a), Lomb-Scargle periodogram (b), CLEAN method after 50 (c) and after 2000 iterations (d).]
\[
P(\omega) = \frac{1}{2\pi^2} \left\{ \frac{\sum_{i=1}^{N} (x_i - \bar{x}) \cos \omega(t_i - \tau)}{\sum_{i=1}^{N} \cos^2 \omega(t_i - \tau)} \right\}^2 + \frac{\sum_{i=1}^{N} (x_i - \bar{x}) \sin \omega(t_i - \tau)}{\sum_{i=1}^{N} \sin^2 \omega(t_i - \tau)} \right\}^2.
\]

where \( \bar{x} \) and \( \sigma \) are mean and variance of the data, respectively, and \( \tau \) is defined as:

\[
\tan (2\omega\tau) = \frac{\sum_{i=1}^{N} \sin 2\omega t_i}{\sum_{i=1}^{N} \cos 2\omega t_i}.
\]

Calculation of the periodogram is performed for frequencies \( \omega_k = 2\pi k/T, k = 1...N \) where \( T \) is the total time interval covered by the data. Any oversampling would add no information but would only act as interpolation and smooth the spectrum (Scargle 1982). That is why normalized periodogram will be equal to \( NA^2/4\sigma^2 \) where \( A \) is amplitude of signal. In the case of evenly sampled data, modified periodogram is reduced to classical one (Scargle 1982). Modification of the periodogram is performed in such a way that the statistical distribution of periodogram is preserved in the case where data are unevenly sampled. This means that if there is no periodic signal but only a Gaussian noise, distribution of peaks with heights for one frequency will be exponential. Level of significance for whole spectrum, i.e. the probability that highest peak in a periodogram sampled at \( M \) independent frequencies is of height \( z \) or higher is given by false alarm probability:

\[
F = 1 - \left[1 - e^{-z} \right]^M.
\]

In a case of evenly sampled data, the number of independent frequencies is given by the formula of Horne and Baliunas (1986). When data are unevenly sampled, this number is reduced and hence false alarm probability is smaller. In order to estimate number of independent frequencies for \( Mn \) data set, simple Monte Carlo simulation is done: 1000 data sets of pure Gaussian noise are generated, the largest value in periodogram is identified for every data set and distribution of largest peaks with height and, hence, cumulative function is determined. The same statistics can be obtained by another method proposed by Delache et al. (1985). Instead of generating noise, the same can be done by randomizing the real data 1000 times. In this way periodic changes are destroyed and the noise is preserved. If data are enough shuffled, both method would give the same result, as it is so in our case. The resulting cumulative functions are fitted with Eq. 5 and approximate value of 475 is obtained as the number of independent frequencies. Level of significance of 99% is about 10.8, which means that there is the probability of 99% that peak higher than 10.8 is in fact a signal and not a noise. In Fig. 2b there are two peaks above the level of significance which match the first two simulated ones. The amplitude of the first period is estimated well and amplitude of the second one is reduced to 13% of its real value. Several other peaks can be seen in the vicinity of 0.033 days\(^{-1}(=1/30)\). These are artefact due to window function.

Much better result is obtained using CLEAN method (Roberts 1987), the one-dimensional version of cleaning algorithm developed for aperture synthesis in radioastronomy. This method performs iterative subtraction of periodic signals from frequency domain. Fourier transform of the input data (dirty spectrum) is, as mentioned, convolution of Fourier transform of true function and spectral window function. This means that dirty spectrum of a simple cosinusoid will be:

\[
D(\upsilon) = aW(\upsilon - \hat{\upsilon}) + a^*W(\upsilon + \hat{\upsilon}),
\]

where \( W \) is the spectral window function, \( \hat{\upsilon} (= \omega/2\pi) \) is the frequency and \( a \) is the complex amplitude of the periodic signal (* represents complex conjugation). The first step in CLEAN algorithm is identification of the highest peak in dirty spectrum of the observed data. The second step is removing contribution of signal with this period in form of Eq. 6 from dirty spectrum. The amplitude of the signal is calculated according to:

\[
a(\hat{\upsilon}) = \frac{D(\hat{\upsilon}) - D^*(\hat{\upsilon})W(2\hat{\upsilon})}{1 - |W(2\hat{\upsilon})|^2}.
\]

In this way the residual spectrum is obtained and identified period is added to the so called clean spectrum. This step is repeated on the residual instead of the dirty spectrum until the stopping condition is reached. Because of an error that can occur when frequency is determined in the presence of several periodic signals, only a fraction \( g \) (gain factor) of this period is removed in one iteration step, so the error is reduced in proceeding iteration. Value of 0.1 is chosen for gain factor in this research which increases the number of necessary iterations. This means that for gain factor 0.1, ten iterations would be needed to clean one frequency. If we assume that every sampled frequency contains one independent periodicity, then number of iterations needed for extracting all of them would be (number of sampled frequencies)/(gain factor). This is however never the case and the number of iterations should be less than this. On the other hand, to iterate over this number is pointless. In this research Baisch’s implementation (Baisch 1999) is used which predetermines the maximum number of iteration steps as a stopping criteria.

The results of CLEAN method applied to Synth data after 50 and 2000 iterations are shown...
in Figs. 2c and 2d. These are final spectra obtained by convolution of the clean spectrum with so called clean beam (the Gaussian fitted to central peak of spectral window function). The frequency resolution of the clean spectrum is inversely proportional to the total time interval covered by the data. The last residual spectrum is added so that noise level can be preserved. Spectrum is calculated at frequencies $k/(4T)$, where $T$ is the total time interval covered by the data and $k$ is integer that takes values $-n, -1, 0, 1...n$. Integer $n$ is determined by the Nyquist frequency which is inversely proportional to the minimal separation between observed points. Thus, for the Mn data, maximum frequency for calculation of the spectral window function is 0.5 day$^{-1}$ and for spectrum it is 0.25 day$^{-1}$. Intervals of frequencies shown in Figs. 2c and 2d are chosen so that peaks at certain frequencies can be clearly recognized. It is obvious that an increase in the number of iterations produces noiseless spectrum with well reproduced frequencies and amplitudes. Because of the fact that CLEAN algorithm deals with complex spectrum, information about the phase of periods is not lost and it is possible to reconstruct missing data set by the inverse Fourier transform of the clean spectrum. If comparison of the reconstructed and the synthesized data sets is performed, it can be concluded that difference decreases when the number of iterations is increased (Fig. 3). It appears, as Baisch stated (1999), that result is insensitive to over-iteration, but only for the noiseless data, as it will be explained in what follows.

In order to analyze how noise would influence the result, the Gaussian noise is added to Synth data. Standard deviation of the noise is 30% of the standard deviation of Synth data. This percentage is so chosen that synthesized data approximately mimic variability of MNW data. It is important to note that standard deviation of noise is 0.09 which is slightly more than amplitude of the second and third signal in Synth data. Days with no Mn observation are removed from this noisy data set and then spectral analysis is performed, but only with CLEAN method. Other methods were not used because results they produced were not satisfactory even in the case of noiseless data.

![Fig. 3. Change of standard deviation of the difference between original synthetic data and data reconstructed with CLEAN after various number of iterations.](image)

![Fig. 4. Spectral analysis of synthetic data set with noise: CLEAN method after 100 (a) and after 16000 iterations (b).](image)

![Fig. 5. Change of standard deviation of the difference between original synthetic data and data reconstructed with CLEAN after various number of iterations.](image)
Figs. 4a and 4b shown clean spectra after 100 and 16000 iterations. It is evident that noise in spectrum is reduced with iterations, but is not completely removed. There are small peaks at frequencies 0.00331 and 0.00746 which could be sidelobs of the larger component. They disappear if frequency of the second periodic signal is changed. There are also peaks of similar heights at higher frequencies (i.e. smaller periods) which are visible in Fig. 6b. Most prominent are below periods of 16 days. Amplitudes at these high frequencies are increased with number of iterations. As a consequence the standard deviation of the difference of Synth and data recovered from noisy Synth data is increased after 16000 iterations (Fig. 5). In the time domain this produces increment of sharp spikes with increment of the number of iteration. Fig. 6a shows how well data recovered after 16000 iterations follow variability of noisy Synth data in time domain. We have chosen this number of iterations as optimal. Harmonics content is then well reproduced, and noise of recovered data is minimal. As one can see in Fig. 6b, recovered amplitudes of signals are reduced, in the case of the third one even to 50%, but these peaks are still most prominent in the spectrum.

4. SPECTRAL ANALYSIS OF REAL DATA

In Fig. 7 results of CLEAN method applied to real Mn data are shown. Interval between the presented frequencies is chosen so that comparison with previous results is possible. It is evident that spectra of the real data are more complex because real processes cannot be simulated by a perfect sine curve, as it is done for synthetic data, but by the multiple components which will mimic their quasi stationary and quasi periodic behavior. Figs. 7a-c and 7d-f show three phases in time analysis for both parameters of MnI 539.4 nm line: equivalent width (MNW) and central depth (MNC), respectively. In Figs. 7a and 7d dirty spectra, spectra before any cleaning, are presented. Figs. 7b and 7e show spectra after two iterations with gain factor 1 and Figs. 7c and 7f after 16000 iterations with gain factor 0.1. Several peaks in vicinity of 0.033 $^{-1}$, similar to ones visible in Fig. 2b are present in dirty spectrum of both time series. They are consequence of time sampling and, as one can see in figures below, they are removed after 2 iterations with gain factor 1, along with two periods at lower frequencies. In spectra obtained after 16000 iteration presence of small peaks, especially at higher frequencies can be seen. These peaks are at the same frequencies as in spectrum of noisy synthetic data which we identified as a consequence of the noise.

Whole frequency interval can be seen in Figs. 8b and 8d for MNW and MNC data, respectively. Resemblance of Figs. 6a and 6b with Figs. 8a and 8b is now obvious. Multiple peaks are present in vicinity of period of 27 days in MNW and MNC spectra because of its quasi-periodic nature due to birth, evolution and death of active regions. In order to support the statement about the existence of 27-day period in Mn data more firmly, we split time series into shorter ones, approximately 500 days long. The length of shorter time series is chosen so that any influence of periods at lower frequencies can be avoided and at the same time the best spectral resolution can be provided. Linear trends were removed from every time series and then CLEAN algorithm was used for spectral analysis. Fig. 9 shows results of this procedure. Different line styles correspond to different time intervals which are listed chronologically. Only
last two intervals overlap so that whole time range of the Mn observations can be covered. As one can see in these figures, spectra for three (1. 1979 - 5.
1980, 12. 1988 - 4. 1990 and 5. 1990 - 9. 1991) of ten time intervals indicate the presence of 27-day periodic change. These intervals approximately match the maximum activity periods of solar cycles 21 and 22. It is possible that some modification of periods around 27 days due to time sampling exists but there is no doubt about the existence of that period.

On the other hand, existence of periods smaller than 20 days is not certain especially when accumulation of noise in case of synthetic data is taken in account. The same can be stated for periods around 170 and 350 days which could be just sidelobs of more prominent one at around 870 days. This period can be identified as quasi-biannual oscillation (QBO) and as such is recognized in various solar indices as in the geophysical data (for example Djurovic and Paquet 1993). The highest peak is period already identified in Mn data due to 11 yr solar activity cycle. Extraction of other potentially real periods must be the subject of a more detailed research. Comparison of recontracted and observed data for both parameters is presented in Figs. 8a and 8c.

Fig. 7. Spectral analysis of MNW ((a) - dirty spectrum, (b) - after two iterations with gain factor 1, (c) - after 16000 iterations) and MNC data ((d) - dirty spectrum, (e) - after two iterations with gain factor 1, (f) - after 16000 iterations).
Fig. 8. Reconstructed data of (a) MNW and (c) MNC (circles – observed data, solid line – reconstructed) and CLEAN spectra after 16000 iterations of MNW (b) and MNC (d).

Fig. 9. Spectral analysis of approximately 500 days long parts of MNW (a) and MNC (b) data sets (different line styles correspond to different time intervals which are listed chronologically in the legend).
5. CONCLUSION

Spectral analysis of highly unevenly sampled observations of the Mn 539.4 nm spectral line parameters is performed in this study. In view of the irregular time sampling of these data, the classical methods for spectral analysis cannot be used. For this reason, the two most frequently used methods for spectral analysis of unevenly sampled data are chosen: the Lomb-Scargle normalized periodogram and the CLEAN method. In order to test the ability of these methods to yield good results two data sets are synthesized as a sum of tree cosine curves with periods and amplitudes similar to the ones found in Mn data. Noise is added to one of them and then data are removed from both data sets, so that the time sampling and, hence, spectral window function are the same as in the case of the Mn observations. Better results are obtained with CLEAN method, which is used to generate the spectrum of Mn data. Reconstruction of missing data is performed in this way and more detailed analysis of spectra becomes possible with classical methods for spectral analysis as for example the wavelet transform.

Finally, the conclusion is that in Kitt Peak data for MnI 539.4 nm we have clearly isolated 11-year, QBO and rotational period in periods of maximal solar activity. Such result is expected because it is shown that parameters of this line are correlated with both CaII index (Livingston and Wallace, 1987) and MgII index of solar activity (Danilovic et al, in preparation). On the other hand, the question how solar activity reflects on the profile of the Mn I 539.4 nm line still remains, is the recovered change the direct image of solar activity or just mimics the change in the core of the MgII k line due to optical pumping, as it is proposed by Doyle et al (2001).

Acknowledgements – Ministry of Science, Technology and Development of the Republic of Serbia (Contract No. 1951) supported our work. We obtained the observational data of Mn 539.4 nm solar line thanks to W. Livingston, for which we are very grateful. Also, we would like to thank the referee for useful suggestions which improved the paper.

REFERENCES

ВРЕМЕНСКА АНАЛИЗА ДУГОРОЧНИХ ПОСМАТРАЊА НЕУТРАЛНОГ МАНГАНА НА 539.4 nm

S. Danilović¹, I. Vince¹,², N. Vitas² and P. Jovanović¹

¹Astronomical Observatory, Volgina 7, 11160 Belgrade 74, Serbia and Montenegro
²Department of Astronomy, Faculty of Mathematics, University of Belgrade
Studentski trg 16, 11000 Belgrade, Serbia and Montenegro

UDK 523.9-3555.3
Оригинални научни рад

Подаци о еквивалентној ширини и централној дубини спектралне линије неутрализованог манган на 539.4 nm, посматрање у периоду од 1979 до 1992 на астрономском обсерваторији Кит Пик, анализирани су у потрази за периодичним променама. Због чињенице да су посматрања нееквидистативно узоркована, било је потребно извршити тест да би се утврдило да ли периоdıчне промене заиста постоје у подацима или не. Две различите методе су примењене на синтетизоване податке који су узорковани на исти начин као и посматрања. Поређење резултата добијених на основу синтетизованих и на основу реалних података показује најмање три периодичне промене ових параметара: 11-годишњу, квази-дводељебану и 27-дневну промену.