Paper D2

A Macro for Producing Graphs used in Assessing a Spectrum

David W. Meek
USDA-ARS-MWA National Soil Tilth Laboratory
Ames, IA 50011

ABSTRACT

This paper is intended for those familiar with the discrete Fourier transform and PROC SPECTRA from SAS/ETS®[1]. The procedure is called through a macro and the output variables are used to develop three diagnostic graphs and a montage. The first graph displays the $\log_{10}$ of the spectra, a Hanning weight interpolation of the spectra, and the 95% confidence bands of the spectra as a function of the $\log_{10}$ of the frequency. The second graph is an overlay of two related plots. The first plot displays the spectral density as a function of the wave number. The values are represented by black points with the ordinates plotted using the black left-hand vertical axis scale. The second plot displays the periodgram fraction as a function of the wave number. The values are represented by gray-scale needles with the ordinates plotted using the gray right-hand vertical axis scale. The third graph displays the normalized cumulative periodgram including the 95% confidence limits as a function of the wave number. A table with spectral tests is also presented. The sun spot record from 1700 -1991 is used to generate an example set of graphs and a table.

INTRODUCTION

In analyzing and modeling the historical or observational records of phenomena there often is interesting periodic behavior to consider. Spectral analysis via the discrete Fourier transform is a likely method to be used for this purpose. Unfortunately PROC SPECTRA from SAS/ETS® does not itself produce any printer, screen, or ODS graphics. It can, however, produce an output file with the Fourier transform coefficients, periodgram values, and spectral density values for further analysis and graphics (SAS, 1993). This paper presents a SAS macro that produces further output and graphics to assist in the analysis of a spectrum.

THE DATA SET

The number of observed sunspots is of historical and scientific interest. The data are publically available through many sources. In this case, the serially complete annual record from 1700 to 1991 was obtained from Iowa State University, Ames, IA. The series has a well-known 11 year cycle. The series, shown in Figure 1, has a slight trend based on several non-parametric correlation measures such as Kendall's τ_b; hence, it is not reasonably stationary. To detrend the series for spectral analysis, it is subjected to a high-pass filter, here the first difference operator (Chapter 2 in Chatfield, 2004). The resulting stationary series is used hereafter. The SAS variable name used here for this series is *dif1_nss*.

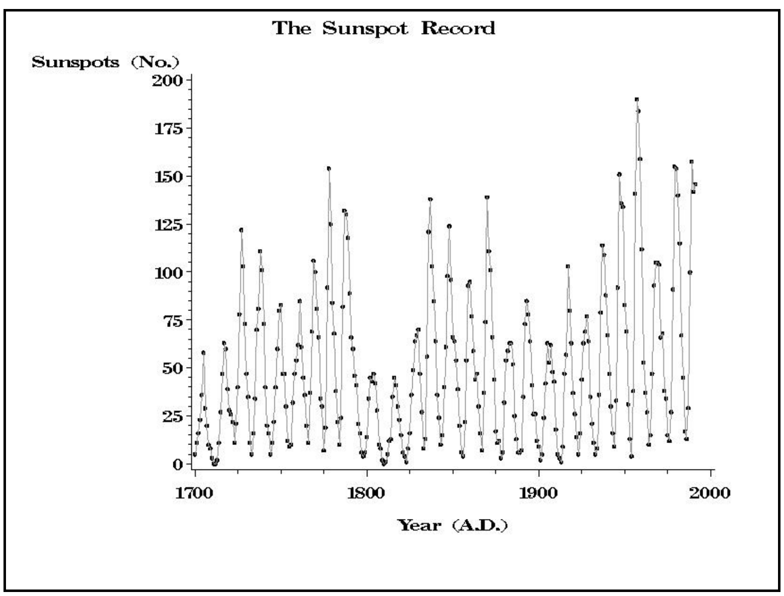

Figure 1. The original serially complete sunspot time series.

[1]The mention of a trade name is for informational purposes only and is not an endorsement by the USDA-ARS.

THE SAS MACRO

All the output shown in this paper is produced from a SAS macro that requires a SAS data set and variable of interest to be defined in a data step. It is mainly intended for exploration and initial analysis on a personal computer. The time variable used is always the observation index which is serially complete and equally spaced. The variable of interest must also be serially complete, i.e., no missing values. To use the Fourier coefficients for a particular application model, the time index may be scaled to the unit of interest. Assuming the SAS data set is named *a0*, the variable of interest is named *dif1_nss*, the chosen output file name is *a0_spout*, and the macro code is compiled, the routine is invoked as follows:

```
%mspectrumv(a0, dif1_nss, a0_spout)
```

The output file produced by proc spectra includes the p, s, coef, and whitetest options. The file is printed in the output window. It is not saved but the user can choose to do so. If desired each coefficient can be tested individually assuming it is from a χ^2 distribution (Chapter 7 in Fuller, 1976). As an alternative, a second output file is also made and again printed in the output window. It is named *mtests*. Here a list indexed by wave number (k), period, and frequency is produced that includes the Fourier coefficients in radial form with the amplitude and phase angle and two overall term test statistics. First, a determination estimate for each k term, R^2_k, is estimated by using Parseval's Theorem (Chapter 7 in Chatfield, 2004). This test statistic is really a normalized periodgram value (scaled from 0 to 1). Next a formal F-test and associated probability value are produced for each term following material presented in Chapter 7 of Fuller (1976).

ENHANCED OUTPUT

THE LOG-LOG SPECTRAL DENSITY

Standard output from the procedure includes frequency and period from the time index and the periodgram and spectral density from the Fourier series coefficients. Common gplot and printer plots for this output are discussed on pp 762 - 765 in Chapter 15 of the *SAS® /ETS User's Guide* (SAS, 1993). Many engineering and other applications like to consider the spectral density and frequency graph on a $\log_{10}$-$\log_{10}$ scale plot. Figure 2 shows such a plot for the spectra of the detrended sunspot series. Here frequency is defined as 1/period rather than use the SAS 2π/period default definition. Here the value 0.1 corresponds to 10 years in the original time series. Also included are the 95% confidence bands for the spectra following principles in Chapter 7 of Fuller (1976) and a simple interpolation using the Hanning weight (¼,½,¼) following principles in Chapter 7 of Chatfield (2004). The plot is labeled "(A)" because it is the first panel in the montage that the macro produces at the completion of its call.

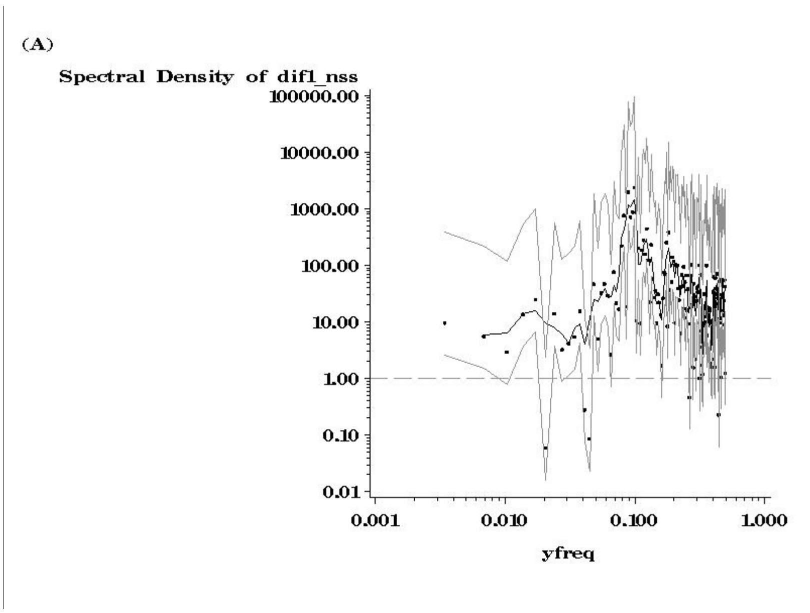

Figure 2. The $\log_{10}$(spectral density) versus the $\log_{10}$(frequency) plot for the detrended sunspot spectra. The solid black line is a Hanning weight interpolation and the solid gray lines are the 95% confidence limits of the spectral density. Notice the group of large spectral density values for frequencies between 0.08 to 0.1 (about 10 to 12 years) are about an order of magnitude larger than all the rest. The highest peak is just before 0.1 (10 years).

PLOTS OF SPECTRAL DENSITY AND PERIODGRAM FRACTION VERSUS WAVE NUMBER

In many disciplines, especially the engineering and physical sciences, it is common to see spectral plots as a function of wave number. The graph of spectral density as a function of wave number (denoted k) for the detrended sunspot series is shown in Figure 3. The black dots are scaled to the left-hand black vertical axis. The Hanning weight interpolation is included. Finally the R_k^2 values are plotted as gray needles with the vertical coordinates scaled to the right hand gray colored vertical axis. This plot is labeled "(B)" because it is the second panel in the montage.

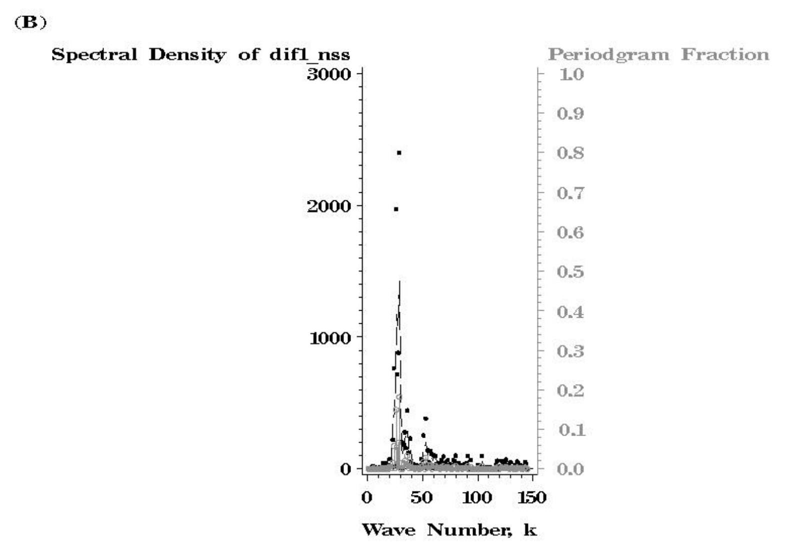

Figure 3. Spectral density and periodgram fraction versus wave number on an overlay plot. The black line is the Hanning weight interpolation. Notice the relatively higher values for $24 \le k \le 29$ (again about 10 to 12 years) with k = 29 being the single largest, $R_{29}^2 \approx 0.18$.

THE NORMALIZED CUMULATIVE PERIODGRAM

Another useful plot of the normalized periodgram values is one that accumulates the normalized values over increasing wave number and plots the cumulative periodgram values between 95% confidence bands (Chapter 7 in Fuller, 1976). When the cumulative curve is entirely inside the confidence bands there are no significant cycles in the time series. Figure 4 shows this plot for the spectra of the detrended sunspot data; in it the normalized cumulative periodgram value axis is abbreviated "Nrm Cum Prdgrm". This plot is labeled "(C)" because it is the third panel in the montage. The resulting montage with all 3 panels is shown in Figure 5.

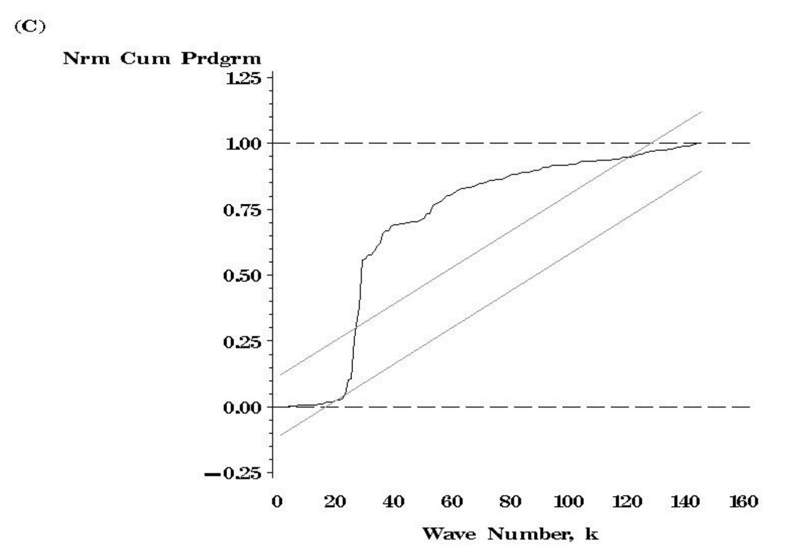

Figure 4. The normalized cumulative periodgram for the detrended sunspot series. Notice the curve increases rapidly at k=24 (≈ 12 year period), exceeding the upper band at k=27 but then increases relatively slowly after k=29 (≈ 10 year period).

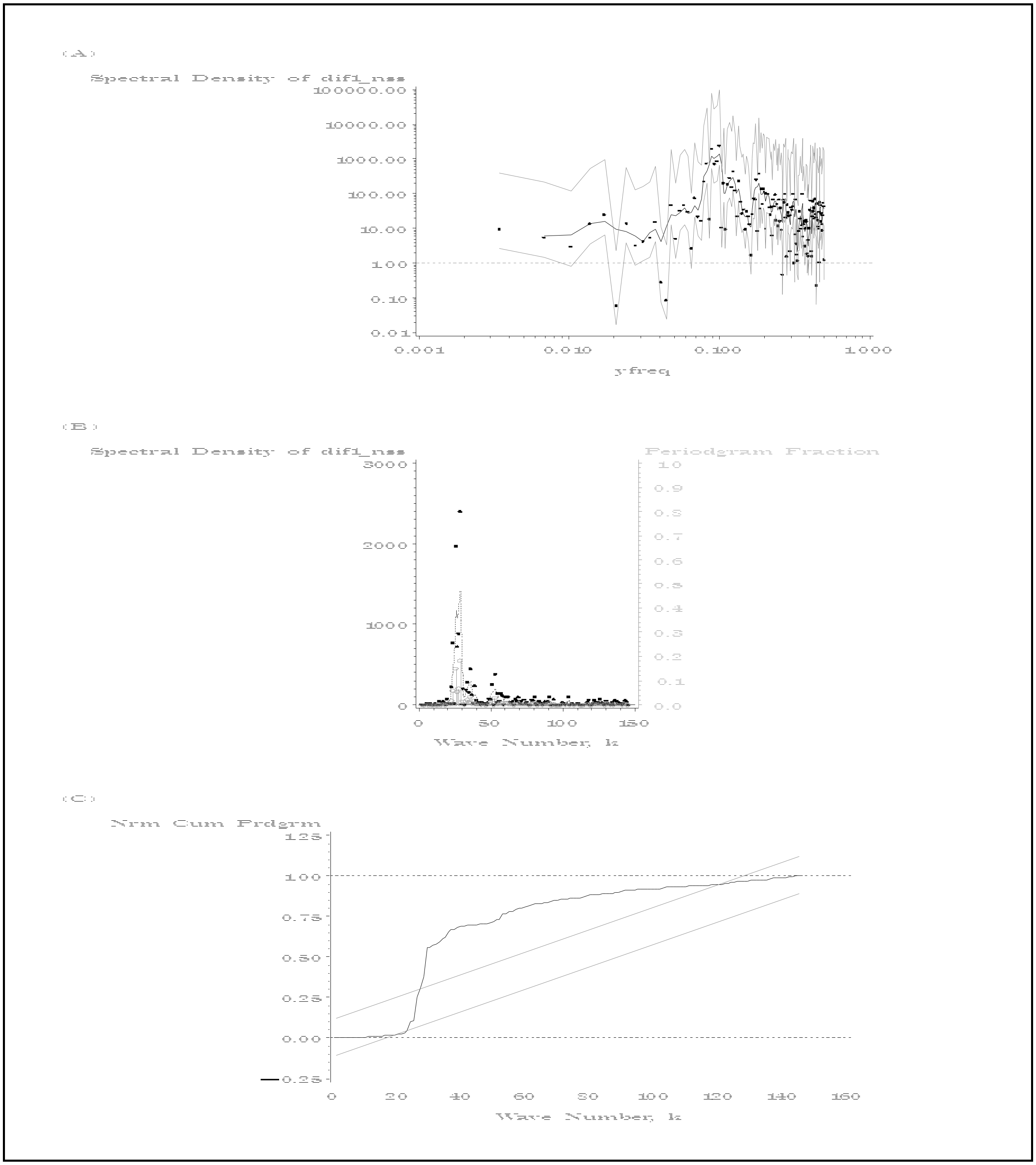

Figure 5. The montage combining all of the spectral graphs presented in Figures 2, 3, and 4.

TABLED OUTPUT

Table 1 presents the first 35 lines of printout from the second output file printed in the output window. For this table the comparatively significant terms in $24 \leq k \leq 29$ are shown in bold type. The user can also save this file if desired. The SAS dataset name for it is *mtests*.

Table 1. Selected printout from the mtests SAS file.

Wave Number, k	Frequency / 0 to 2 Pi	Period / Sequence Units	Term Coef. / Radius	Term Coef. / Phase Angle (rad)	Term R^2 (%)	Term F Statistic	Term F Alfa Probability
0	0.0000	.	49.26918	0	.	.	.
1	0.0215	292	8.410652	-0.4193	2.163	3.206	0.0420
2	0.0430	146	7.178975	-1.24512	1.576	2.322	0.0999
3	0.0646	97.33333	16.90984	1.069528	8.744	13.894	0.0000
4	0.0861	73	3.546235	0.232429	0.385	0.560	0.5719
5	0.1076	58.4	8.513724	-0.20014	2.217	3.287	0.0388
6	0.1291	48.66667	8.218935	-1.54267	2.066	3.059	0.0485
7	0.1506	41.71429	6.142283	-0.37445	1.154	1.692	0.1859
8	0.1721	36.5	6.419026	-1.13159	1.260	1.850	0.1590
9	0.1937	32.44444	4.858007	-1.01917	0.722	1.054	0.3498
10	0.2152	29.2	7.085325	-1.37788	1.535	2.261	0.1061
11	0.2367	26.54545	1.022216	-1.18167	0.032	0.046	0.9547
12	0.2582	24.33333	4.207675	1.557259	0.541	0.789	0.4551
13	0.2797	22.46154	2.882572	1.394015	0.254	0.369	0.6915
14	0.3012	20.85714	3.649502	-1.56616	0.407	0.593	0.5533
15	0.3228	19.46667	1.089361	1.179715	0.036	0.053	0.9487
16	0.3443	18.25	2.39459	0.761726	0.175	0.255	0.7753
17	0.3658	17.17647	4.277002	-0.00485	0.559	0.816	0.4433
18	0.3873	16.22222	2.034877	-0.57526	0.127	0.184	0.8322
19	0.4088	15.36842	3.606761	-1.50269	0.398	0.579	0.5610
20	0.4304	14.6	3.830586	0.815583	0.449	0.654	0.5209
21	0.4519	13.90476	3.088175	0.366766	0.292	0.424	0.6548
22	0.4734	13.27273	4.022687	-0.98475	0.495	0.721	0.4871
23	0.4949	12.69565	7.560854	1.023428	1.748	2.580	0.0775
24	**0.5164**	**12.16667**	**12.65483**	**1.129057**	**4.897**	**7.467**	**0.0007**
25	0.5379	11.68	5.307378	-1.45294	0.861	1.260	0.2852
26	**0.5595**	**11.23077**	**18.51618**	**1.310905**	**10.485**	**16.983**	**0.0000**
27	**0.5810**	**10.81481**	**17.71359**	**-1.04285**	**9.595**	**15.390**	**0.0000**
28	**0.6025**	**10.42857**	**19.38818**	**1.327275**	**11.495**	**18.833**	**0.0000**
29	**0.6240**	**10.06897**	**23.95895**	**1.537327**	**17.554**	**30.873**	**0.0000**
30	0.6455	9.733333	1.81407	0.127279	0.101	0.146	0.8642
31	0.6671	9.419355	6.447052	-0.63303	1.271	1.867	0.1565
32	0.6886	9.125	2.840551	-1.27305	0.247	0.359	0.6989
33	0.7101	8.848485	6.285309	-1.08722	1.208	1.773	0.1716
34	0.7316	8.588235	4.422008	0.15722	0.598	0.872	0.4191
				*			
				*			
				*			

DISCUSSION AND CONCLUSIONS

This macro was developed to help researchers at the author's agency explore time series data for possible cyclical behavior. The sound assessment of periodic behavior in a time series can be difficult because it is subject to many influences that may require further consideration. In the case of the sunspot series, the output from this macro displays something occurring between a 10 and a 12 year period but nothing definite was determined in the initial exploration. The 11 year cycle is often shown by using more general versions of the Hanning filter. The selection of an appropriate filter for a given time series requires experience, knowledge of both statistics and the specific data, and careful consideration. When sound judgement and methods are employed, however, useful information and insight may be revealed. Further information and the macro code are available from the author.

REFERENCES

Chatfield, C., 2004. The analysis of time series, an introduction, 6th ed. Chapman&Hall/CRC Press, Boca Raton, FL. 333 pp.

Fuller, W. 1976. An introduction to statistical time series. J. Wiley and Sons. New York, NY. 470 pp.

SAS®/ETS User's Guide, 1993, Version 6, 2nd ed., No. 56010. SAS Inst. Inc., Cary, NC

ACKNOWLEDGMENTS

The sunspot record was obtained from Professor (Emeritus) R. Carlson, The Department of Agronomy, Iowa State University, Ames, IA 50011. This work was supported by the USDA-ARS National Soil Tilth Laboratory, Ames, IA, Dr. J.L. Hatfield, Director. Thanks to Drs. Rob Malone and Jeremy Singer, USDA-ARS MWA, Ames, IA; Deb Palmquist, USDA-ARS MWA, Peoria, IL; and Delayne Stokke, Wells Fargo Bank, Des Moines, IA for comments.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Name: David Meek
Enterprise: USDA-ARS-MWA National Soil Tilth Laboratory
2150 Pammel Dr.
Ames, IA 50011
Work phone: (515) 294-2246
Fax: (515) 294-8125
Email: Dave.Meek@ars.usda.gov
Web: None