JKAU: Sci., **Vol. 21** No. 2, pp: 371-379 (2009 A.D. / 1430 A.H.); DOI: 10.4197 / Sci. 21-2.14

Data Analysis of the Annual Mean of the Sunspot Numbers for the Years 1945-2004

M.A. Sharaf and **Amr A. Sharaf** [1]

Department of Astronomy, Faculty of Science,
[1]Department of Computer Science, Faculty of Computing Science and Information Technology, King Abdulaziz University,
Jeddah, Saudi Arabia

Abstract. In the present paper, an accurate and full data analysis was developed for the most recent annual mean of the sunspot numbers for the years 1945-2004. The precision criteria of the analysis are very satisfactory and were supported graphically and computationally by many tests.

1. Introduction

A growing mass of evidence suggests that solar activity affects our weather and long-term variations of the sun's energy output affect our climate. The literatures on this subject covers a period of more than 120 years, and many distinguished scientists have contributed, (see, *e.g*. the extraordinary number of articles on the site: httt://adsabs.harvard.edu). Moreover almost every large solar-terrestrial symposium now includes at least one session on sun weather/climate investigations. On the other hand, the basic measure of the solar activity is the number of the sunspot visible on the solar disk at any given time, the more spots, the more active is the sun[1].

Now, if the sunspots are a key factor, that is, a good usable indicator of solar activity for sun-weather relationships, an obvious condition must be met before sunspot numbers can be used to predict changes in weather and climate. The sunspots themselves must be predictable. In fact it is very important to have a full understanding of sunspot predictability for sun weather purposes.

Due to the serious rule of the sunspot in our climate, we devolved in the present paper, an accurate and full data analysis for the most recent annual mean of the sunspot numbers for the years 1945-2004[2]. The precision criteria of the analysis are very satisfactory and were supported graphically and computationally by many tests.

2. Basic Formulations

2.1 *Some Basic Statistics*

For data analysis of the present paper we used some basic statistics, of these are:

1. Descriptive statistics; 2. Location statistics; 3. Dispersion statistics; 4. Shape statistics.

2.2 *Autocorrelation*

Autocorrelation is important in time series analysis. Let ρ_k be the autocorrelation at lag k. An estimate of ρ_k is[3]

$$\rho_k = \sum_{t=1}^{N-k}(x_t - \overline{m})(x_{t+k} - \overline{m}) / \sum_{t=1}^{N}(x_t - \overline{m})^2 \ , \tag{1}$$

where $\overline{m}$ is the mean of the data $x_1, x_2, \cdots, x_N$

2.3 *Linear Least Squares Fit*

Let y be represented by the general linear expression of the form:

$$\sum_{i=1}^{n} c_i \varphi_i(x) \tag{2}$$

where φ's are linear independent functions of x. Let $\mathbf{c}$ be the vector of the exact values of the c's coefficients, and $\hat{\mathbf{c}}$ the least squares estimators of $\mathbf{c}$ obtained from the solution of the normal equations of the form $\mathbf{G}\hat{\mathbf{c}} = \mathbf{b}$. The coefficients matrix $\mathbf{G}(n \times n)$ is symmetric positive definite, that is, all its eigen values $\lambda_i ; i = 1,2,\cdots,n$ are positive. Let $E(z)$ denotes the expectation of z and σ^2 the variance of the fit, defined as:

$$\sigma^2 = \frac{q_n}{(N-n)} \tag{3}$$

where

$$q_n = (\mathbf{y} - \mathbf{\Phi}^T \hat{\mathbf{c}})^T (\mathbf{y} - \mathbf{\Phi}^T \hat{\mathbf{c}}), \tag{4}$$

N is the number of observations, **y** is the vector with elements y_k and $\mathbf{\Phi}(n \times N)$ has elements $\Phi_{ik} = \Phi_i(x_k)$. The transpose of a vector or a matrix is indicated by the superscript "T".

According to the least squares criterion, it could be shown that[4]:

1. The estimators $\hat{\mathbf{c}}$ obtained by the least squares method gives the minimum of q_n.

2. The estimators $\hat{\mathbf{c}}$ of the coefficients $\mathbf{c}$, obtained by the least squares method, are unbiased; *i.e.* $E(\hat{\mathbf{c}}) = \mathbf{c}$.

3. The variance-covariance matrix $\mathrm{Var}(\hat{\mathbf{c}})$ of the unbiased estimators $\hat{\mathbf{c}}$ is given by:

$$\mathrm{Var}(\hat{\mathbf{c}}) = \sigma^2 \mathbf{G}^{-1}, \tag{5}$$

where $\mathbf{G}^{-1}$ is the inverse of the matrix $\mathbf{G}$.

Finally, it should be noted that, if the precision is measured by probable error e, then :

$$e = 0.6745\,\sigma \tag{6}$$

2.4 *Confidence Interval*

A confidence interval gives a bound within which a parameter is expected to lie with a certain probability (in the present paper we used a probability of 0.95). Interval estimation of a parameter is often useful in observing the accuracy of an estimator as well as in making statistical inferences about the parameter in question.

As for examples, we compute the confidence interval for *a single observed response* at each of the values of the independent variables. In this way we get a region that is likely to contain all possible observations, also, we compute the confidence interval for *the mean response* at each of the values of the independent variables. In this way we get a region that is likely to contain the fitted curve.

3. Data Analysis

3.1 *Source*

The data are the, monthly mean northern and southern sunspot numbers for the years 1945-2004[2], and were taken from Hipparcos main catalogue:

http://cdsweb.ustrasbg.fr/cats/Cats.htx., its table number at CDS (Strasbourg astronomical data centre) is: (J/A+A/447/735 Hemispheric Sunspot numbers 1945-2004).

3.2 *Numerical Results*

The formulations of Section 2 were applied to the above data and the results are displayed in Appendix A.

In concluding the present paper, an accurate and full data analysis was developed for the most recent annual mean of the sunspot numbers for the years 1945-2004[2]. The precision criteria of the analysis are very satisfactory and were supported graphically and computationally by many tests.

References

[1] **Herman, J.R.** and **Goldberg, A.R.,** *Sun, Weather and Climate*, Dover Publications, Inc., New York (1985).

[2] **Temmer, M., Rybak, J., Bendik, P., Veronig, A., Vogler, F., Otruba, W., Poetzi W.** and **Hanslmeier, A.,** *Astron.Astrophys*, 447,735 (2006).

[3] **Box, G.E.P.** and **Jenkins, G.M.,** *Time Series Analysis, Forecasting and Control,* Holden-Day, San Francisco (1976).

[4] **Kopal, Z.** and **Sharaf, M.A.,** *Apss.*, 70, 77 (1980).

Appendix A

Analysis of the Annual Mean of the Sunspot Number for the Years 1945-2004.

I- Statistics of the Annual Mean of Sunspot Number for the Years:1945-2004

I-1-Basic Descriptive Statistics

- The median (central value) =67.8458
- The variance =2716.75

I-2-Location Statistics

- The geometric mean =55.8696
- The harmonic mean =34.1104
- The root mean square =92.9255

I-3-Dispersion Statistics

- The variance of sample mean =45.2792
- The standard error of sample mean =6.72898
- The coefficient of variation =0.674944
- The mean deviation =45.3949
- The median deviation =39.6583
- Sample range =186.95

I-4-Shape Statistics

- Skewness =0.355971
- The Pearson skewness 2 =0.539827
- The kurtosis = 1.94387
- The kurtosis excess =-1.05613

II-Histogram of the Annual Mean of Sunspot Number for the Years:1945-2004

IĨ Histogram of the Annual Mean of Sunspot Number for the Years 1945̃ 2004.

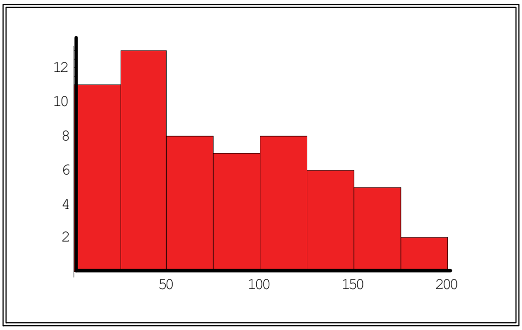
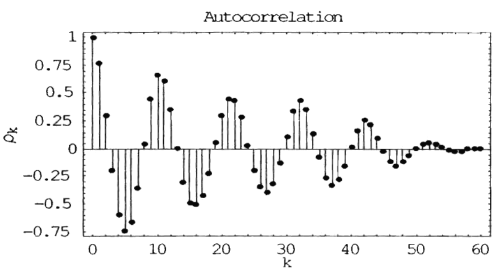


Fig. 1. Histogram of the annual mean of sunpots for the years 1945-2004.

IIĨ Autocorrelation of the Annual Mean of Sunspot Number for the Years 1945̃ 2004.

Let ρ_k be the autocorrelation at the lag k, the numerical values and the graphical representation of (k, ρ_k) are shown in what follows.

k	ρ_k	k	ρ_k	k	ρ_k
0	1.	20	0.299021	41	0.166811
1	0.763791	21	0.450376	42	0.254088
2	0.302193	22	0.437026	43	0.212681
3	-0.202832	23	0.283836	44	0.103153
4	-0.594885	24	0.02969	45	-0.0190735
5	-0.749984	25	-0.202411	46	-0.11989
6	-0.661014	26	-0.34807	47	-0.152225
7	-0.351045	27	-0.398801	48	-0.122647
8	0.047653	28	-0.316488	49	-0.056684
9	0.445595	29	-0.134125	50	-0.000872152
10	0.658903	30	0.115222	51	0.0471603
11	0.601296	31	0.338727	52	0.061452
12	0.349641	32	0.436938	53	0.0481578
13	0.00863991	33	0.348853	54	0.0191134
14	-0.29977	34	0.132993	55	-0.0140444
15	-0.488993	35	-0.0799108	56	-0.0262683
16	-0.509135	36	-0.261099	57	-0.0259294
17	-0.418518	37	-0.330991	58	0.000069799
18	-0.218409	38	-0.281937	59	0.0101748
19	0.056291	39	-0.157789	60	0
		40	0.0142989		



IV. Least Squares Fit of the Annual Mean of Sunspot Number for the Years 1945-2004.

IV. The Fitted Equation

$$y = c_1 + c_2\sin(2mx) + c_3\cos(6mx) + c_4\sin(6mx) + c_5 \sin(8mx); \quad m = 0.098$$

IV. 2 *The c's Coefficients and Their Probable Errors*

c_1=74.0888±1.54733

c_2=14.0579±2.12677

c_3=-36.014±2.17468

c_4=54.7758±2.19384

c_5=-16.2089±2.18633

IV. 3 Confidence Intervals of the c's Coefficients

Coefficient of	Value	Confidence interval
1	74.0888	{69.4914, 78.6862}
sin(2mx)	14.0579	{7.73897, 20.3769}
cos(6mx)	–36.014	{–42.4753, –29.5527}
sin(6mx)	54.7758	{48.2575, 61.294}
sin(8mx)	–16.2089	{–22.7048, –9.71295}

IV. 4 The Data and the Fit

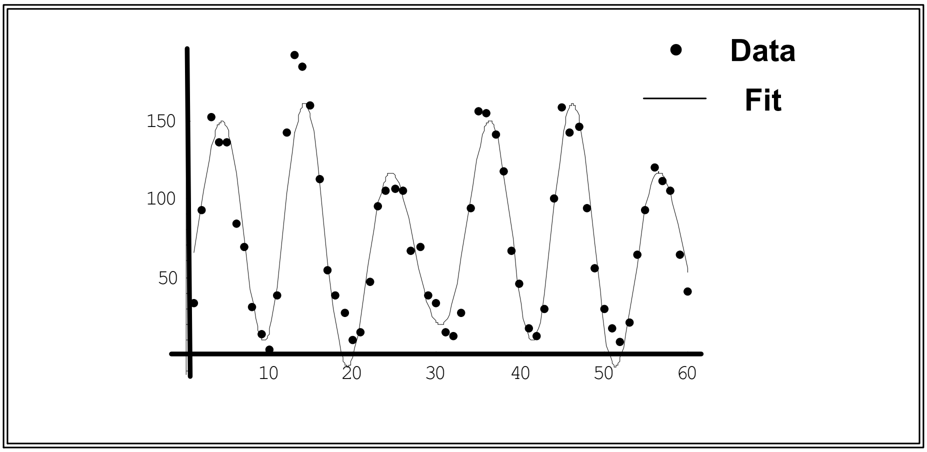


Fig. 2. Comparsion between the raw and the fitted data.

V. Confidence Region of the Data

The region that is likely to contain all possible observations is illustrated in the following figure.

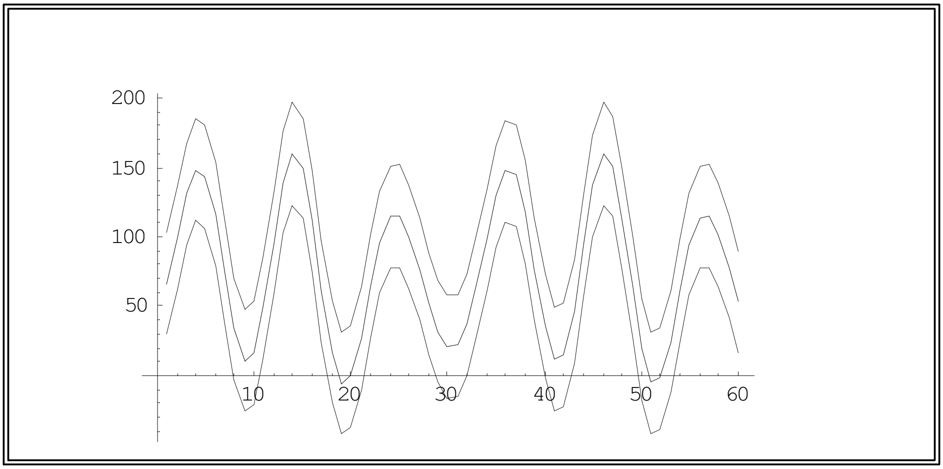


Fig. 3. Confidence region of the data.

VI. Confidence Region of the Curve

The region that is likely to contain the fitted curve is illustrated in the following figure.

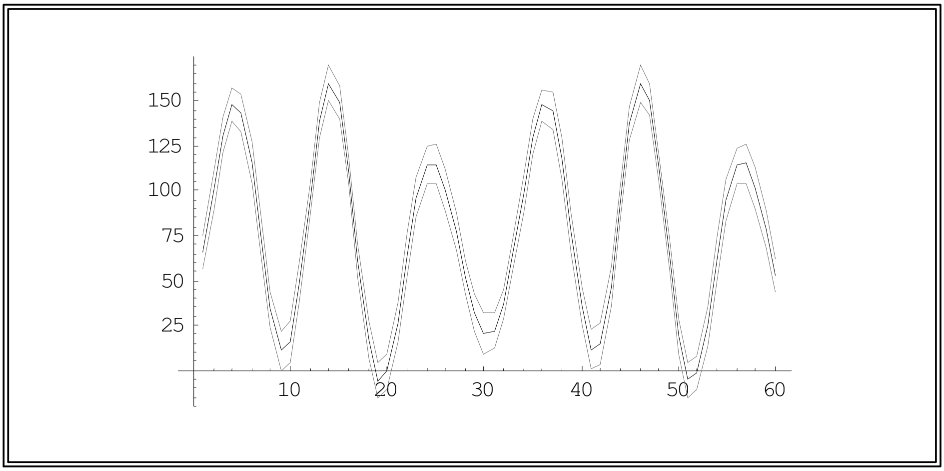


Fig. 4. Confidence region of the curve.

تحليل بيانات المتوسط السنوي للبقع الشمسية للسنوات ١٩٤٥-٢٠٠٤م

محمد عادل شرف، و عمرو عبد العزيز شرف

قسم العلوم الفلكية، كلية العلوم،

وقسم علوم الحاسبات، كلية الحاسبات وتقنية المعلومات،

جامعة الملك عبد العزيز، جدة – المملكة العربية السعودية

المستخلص. تم في هذا البحث تشييد تحليل دقيق وشامل لبيانات المتوسط السنوي للبقع الشمسية للسنوات ١٩٤٥-٢٠٠٤م، فقد كانت معايير الدقة للتحيل مرضية للغاية كما تبين بيانيًا وتحليليًا بالعديد من الاختبارات.