# Modelling for a Series of Annual Extreme Values of the Daily Sunspot Numbers

**Minoru Tanaka**

*Department of Network and Information, School of Network and Information, Senshu University, Kawasaki 214-8580, Japan*

**Abstract**. This paper deals with a statistical time series modelling for *annual* extreme values of a *daily* averages of the International relative sunspot number. We focus on two time series of annual maximum and minimum sunspot numbers, and individually we consider the prediction of the data based on an AR model and also the estimation of a probability distribution function of the residuals derived from the model.

**Keywords:** annual maxima, annual minima, AR model, daily sunspot number, GEV distribution.

## 1. Introduction

The sun has a great influences on the earth like the cold weather affects the rice crop, and so it must be very important for us to analyze solar activity. A well-known data for the solar yearly activity will be the International sunspot number, which was first reported by Wolf in *1852* (Thoms and Weiss [10]). A series of numbers of yearly averaged sunspot (*from 1700 to 2011*) has about *11*-year period (see Figure A1 in Appendix). The series has been analyzed by many authors and researchers (see, for example, Brockwell and Davis [1], Pourahmadi [9]).

In the previous paper [7], we have considered the prediction of a solar cycles from the yearly sunspot numbers based on an *AR(9)* model. Furthermore, we focused on a *daily* averages of sunspot numbers and also on a time series of *annual maximum* sunspot numbers. We discussed the estimation of a probability distribution function and the prediction of the annual maximum data. Unhappily we found that the used data was not a daily averages of sunspot numbers but the *area* daily averages for the sunspots. In this paper we shall retry modelling and analyzing a true daily sunspot numbers.

We use a daily averages of the International Sunspot Number (published in Solar Influences Data Analysis Center (*SIDC*) in Belgium). It should be noted that daily values for years prior to *1849* are partly missing. The available data is for the period *1 January 1849* through *31 March 2011* plotted in Figure 1. A histogram is given in Figure 2. It is seen that the series has many zeros and the ratio of zero number is about *15*%.

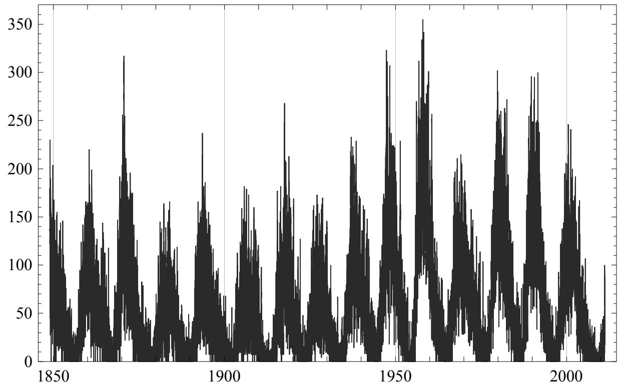


Figure 1. The series of daily sunspot numbers from *1 Jan. 1849* through *2 Mar. 2011*.

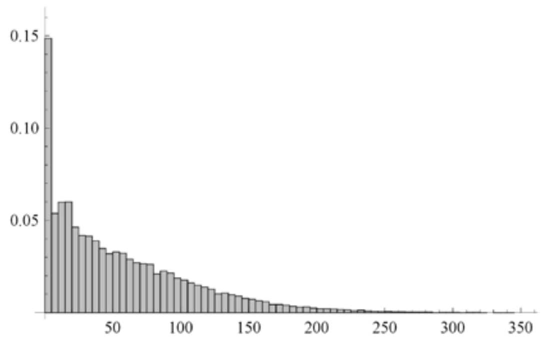


Figure 2. Histogram of the daily sunspot numbers.

Similar to the yearly averaged sunspot numbers, the level of the series seems to oscillate with an approximate period of *11*. But the series of daily sunspot number is fluctuating widely and sharply, and it has many zeros (see Figure 2). So it looks more difficult to get an appropriate model directly for the prediction of the series than that of the yearly averaged sunspot numbers. We then focus on a series of annual extreme values of the daily sunspots, and we consider the problem that the series has a similar model like that of the yearly sunspot numbers.

In Section 2 we treat of a series of annual maximum sunspot numbers obtained from the daily averaged sunspot numbers (for the period *1 January 1849* through *1 December 2010*), and we consider the estimation of a probability distribution function and the prediction of the annual maximum data based on an AR model. Also in Section 3, we focus on a series of annual minimum sunspot numbers obtained from the same daily averaged sunspot numbers, and consider the estimation of a distribution function and the prediction of the data.

This paper is supported by the computer software *Mathematica V8.0* and its application *Time Series Pack for Mathematica* ([5]).

## 2. Annual maximum data of the daily sunspot numbers

We deal with annual *maximum* sunspot numbers, which is the block maximum (extreme value) data obtained from the daily sunspot numbers. The series is plotted in Figure 3 from *1849* to *2010*. It also seems to oscillate with the period of *11* similar to that of the yearly averaged sunspot data.

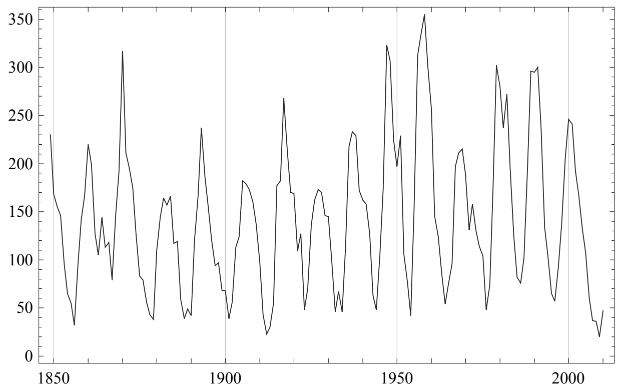


Figure 3. Annual maximum sunspot numbers from *1849* to *2010 with minimum 20 and maximum 355*.

We fit a quadratic function of time $x$ to the series. The quadratic polynomial is $\{10.368+ 0.026x -0.0001\, x^2\}$. We remove the trend and get the residuals by subtracting the trend from the data. Then we fit the residuals to a stationary time series model. Its sample correlation and sample partial correlation functions are plotted in Figure 4a-b. The decaying behavior will imply that the series is stationary.

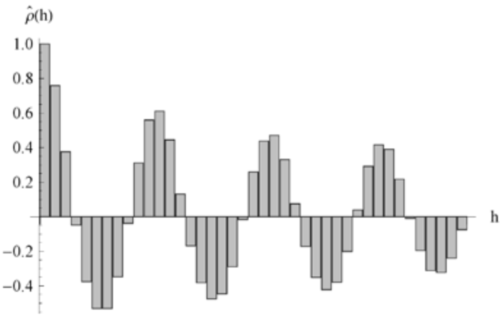


Figure 4a. The correlogram of the annual maximum sunspot numbers.

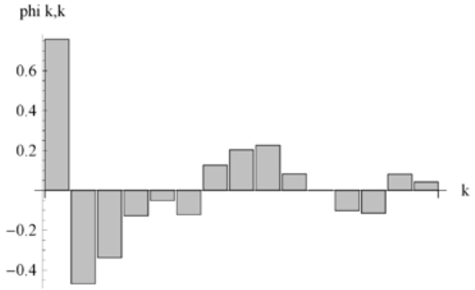


Figure 4b. The sample partial correlation function of the series.

We can model the series as observation from the AR model, and the smallest AIC model is AR(9). It is seen that the maximum likelihood estimates of the AR(9) parameters are given by

$$\{a_1,\ a_2,\ \ldots.,\ a_9\} = \{0.870, -0.076, -0.241, -0.035, -0.144, -0.192, 0.009, -0.042, 0.26\},$$

$$\sigma^2 = 2.40\,.$$

An *AR model* of order p is defined by

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \ldots. + a_p X_{t-p} + e_t,$$

where $\{e_t\}$ is white noise with mean zero and variance $\sigma^2$ (see, for example, [2], [4], [6]).

The parametric estimate of the spectral density based on the AR(9) model is shown as a solid curve together with a sample spectrum in Figure 5 .

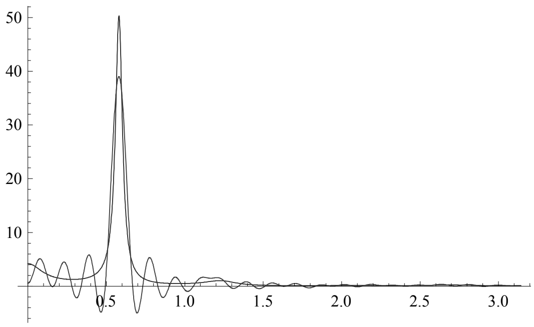


Figure 5. Sample spectrum and the spectral density based on the AR(9) model.

To test the adequacy of the model we calculate the residuals of the fitted AR(9) model and the correlogram of the residuals. The correlogram of the residuals shown in Figure 6 suggests that the residuals behave like white noise.

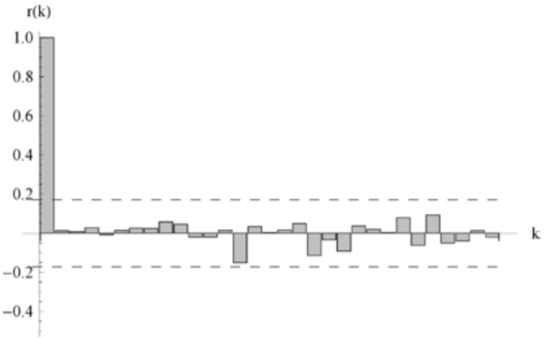


Figure 6. The Correlogram of the residuals and the 95% bounds.

In order to detect volatility (conditional heteroskedastic) we look at the correlogram of the squared residuals. From Figure 7 there is no evidence of serial correlation in the squared values since the correlation function falls within the *95*% bounds. The AR(9) model also passes the portmanteau test because the portmanteau statistic is 3*2.49* against the *95*% quantile of the Chi-square distribution with freedom *31* is *44.99*. Hence the AR(9) model is appropriate for the series.

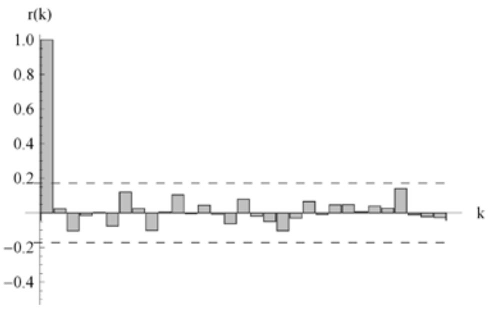


Figure 7. The Correlogram of the squared residuals and the 95% bounds.

Then we use the fitted model AR(9) to forecast the future values. The values of AR-parameters $a_2$, $a_4$, $a_7$, $a_8$ of the AR(9) model are very small, but we use them for the prediction. We calculate the best linear prediction up to *25* time steps ahead based on the series and the model AR(9) :

{*96, 159, 203, 204, 180, 136, 101, 68, 55, 61, 87, 129, 170, 196, 196, 175, 140, 105, 82, 75, 84, 106, 137, 165, 181*}.

The prediction of the annual maximum value for 2011 is *96*, but the true value is *136*.

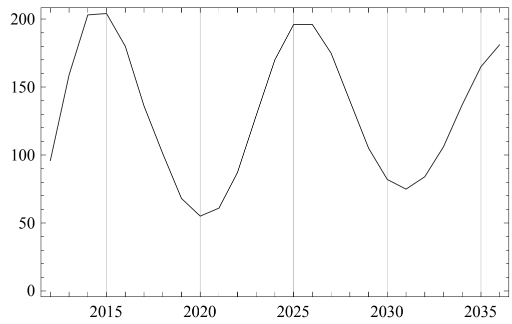


Figure 8. Time plot for 25 predicted values.

We plot the data along with the next *25* predicted values in Figure 9. From Figure 8 it is seen that the next peak of the annual maximum sunspot data is around in *2014* with a sunspot number *204*.

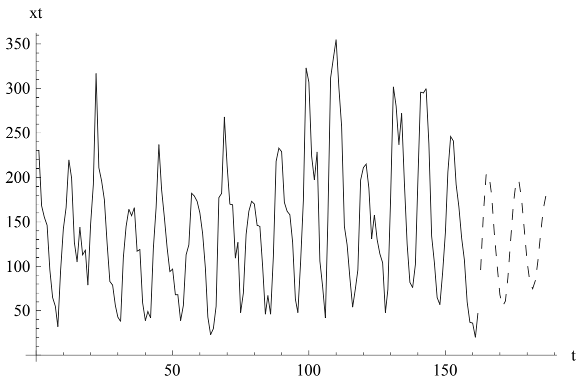


Figure 9. Time plot with last 25 years predicted values (dotted line)

Next we consider the estimation of a probability density function of the residuals. It is seen that the density function of the residuals is estimated by a *generalized extreme value* (GEV) distribution with parameters $\{\mu, \sigma, \xi\} = \{-0.637, 1.526, -0.197\}$ which has the form

$$G(x) = Exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\},$$

defined on the set $\{x : 1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0\}$, where $-\infty < \mu < \infty$ , $\sigma > 0$ and $-\infty < \xi < \infty$. (see [2], [4], [6])

Its density function is plotted in Figure 10 with a histogram of the residuals.


Figure 10. Histogram of the residuals of the AR(9) model and the density plot of the *GEV* distribution with $\{\alpha, \beta, \mu\} = \{-0.637, 1.526, -0.197\}$.

The corresponding density function estimate seems to be almost consistent with the histogram of the data. It is seen that the *GEV* distribution model is adequate for the data. This is confirmed by the standard diagnostic graphical check of the quantile plot shown in Figure 11 and also by *6* statistical goodness of fit tests in Table 1.

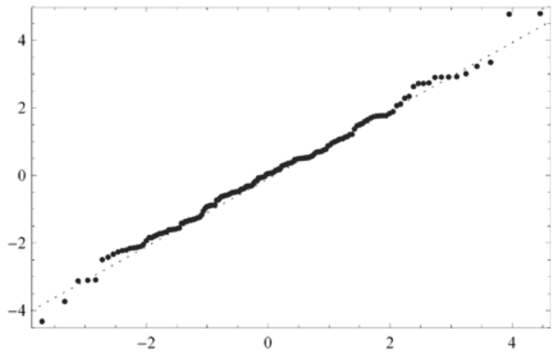


Figure 11. Quantile plot for *GEV* distribution function fitted to the residuals.

Table 1.

| | Statistic | P-Value |
|---|---|---|
| Anderson-Darling | *0.263* | *0.963* |
| Cramér-von Mises | *0.036* | *0.951* |
| Kolmogorov-Smirnov | *0.038* | *0.970* |
| Kuiper | *0.075* | *0.845* |
| Pearson $\chi^2$ | *17.95* | *0.265* |
| Watson $U^2$ | *0.036* | *0.859* |

## 3. Annual minimum data of the daily sunspot numbers

Similar to the annual *maximum* sunspot numbers in Sec.2, we also consider the modelling of an annual *minimum* sunspot numbers obtained from the daily sunspot numbers. The series is plotted in Figure 12 from *1849* to *2010* and the histogram is in Figure 13. It also seems to oscillate with the period of *11*. But we should note that the series has many zeros (about *66.7*%) and a maximum *86*. Then the graph is fluctuating sharply.

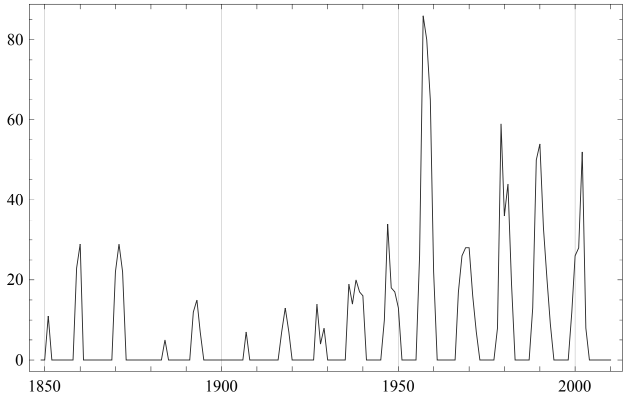


Figure 12. Annual minimum sunspot numbers from *1849* to *2010*.

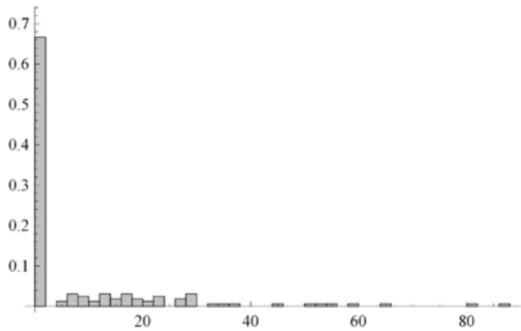


Figure 13. Histogram of Annual minimum sunspot numbers.

Its sample correlation and sample partial correlation functions are plotted in Figure 14a-b. The decaying behavior seems to imply that the series is also stationary.

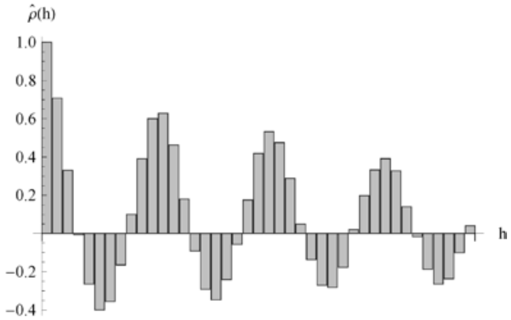


Figure 14a. The correlogram of the annual minimum sunspot numbers.

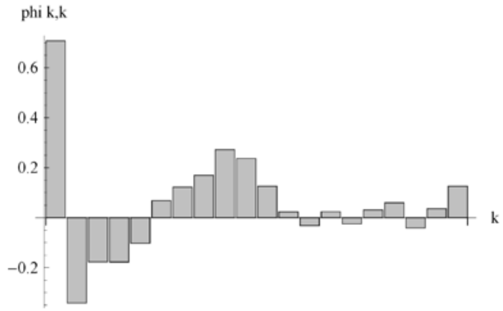


Figure 14b. The partial-correlogram of the annual minimum sunspot numbers.

We can model the series as observation from the AR model, and the smallest AIC model is AR(13). It is seen that the maximum likelihood estimates of the AR(13) parameters are given by

$$\{a_1,\ a_2,\ \ldots,\ a_{13}\} = \{0.674, -0.117, -0.083, 0.008, -0.085, 0.010,$$
$$0.089, -0.092, 0.085, 0.269, 0.054, 0.047, -0.027\},$$
$$\sigma^2 = 1.579\,.$$

The parametric estimate of the spectral density based on the AR(13) model is shown as a solid curve together with a sample spectrum in Figure 15 .

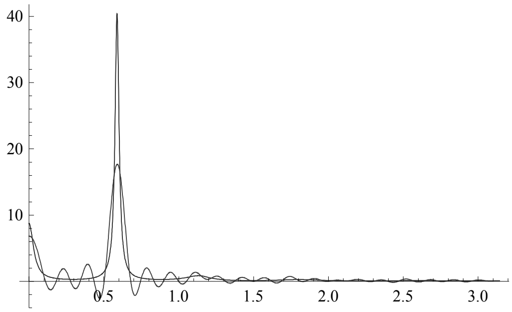


Figure 15. Sample spectrum and the spectral density based on the AR(13) model.

To test the adequacy of the model we calculate the residuals of the fitted AR(13) model and the correlogram of the residuals. The correlogram of the residuals shown in Figure 16 suggests that the residuals behave like white noise.

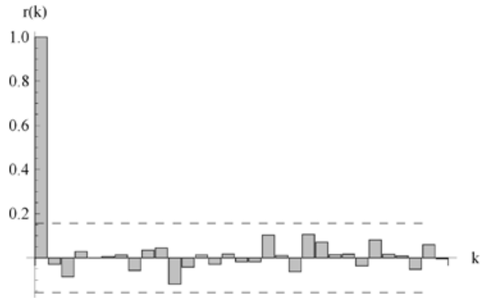


Figure 16. The Correlogram of the residuals and the 95% bounds.

In order to detect volatility we look at the correlogram of the squared residuals. From Figure 17 there is no evidence of serial correlation in the squared values since the correlation function almost falls within the *95*% bounds except at lag *11*. The AR(13) model also passes the portmanteau test because the portmanteau statistic is *22.22* against the *95*% quantile of the Chi-square distribution with freedom *27* is *40.11*. Hence the AR(13) model is appropriate for the series.

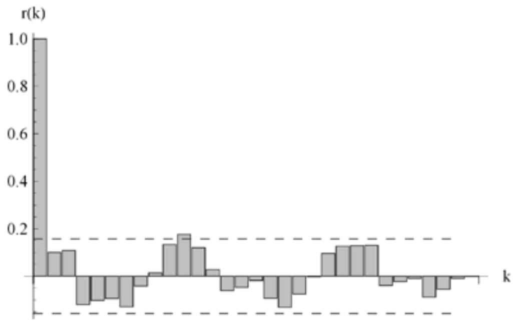


Figure 17. The Correlogram of the squared residuals and the 95% bounds.

Therefore we use the fitted model AR(13) to forecast the future values. The values of AR-parameters $a_4$, $a_6$ of the AR(13) model are very small, but we use them for the prediction. We calculate the best linear prediction up to 25 time steps ahead based on the series and the model AR(13) :

$$\{6, 21, 18, 8, 1, 0, 0, 0, 0, 0, 2, 8, 13, 11, 6, 1, 0, 0, 0, 0, 1, 4, 8, 10, 8\}.$$

The prediction of the annual minimum value for 2011 is *6*, but the true value is *0*.

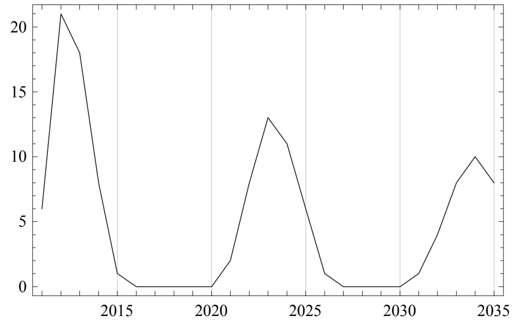


Figure 18. Time plot for 25 predicted values.

We plot the data along with the next 25 predicted values in Figure 19. From this result and Figure 18 the next peak of the annual minimum sunspot data will be in *2012* with a sunspot number *21*.

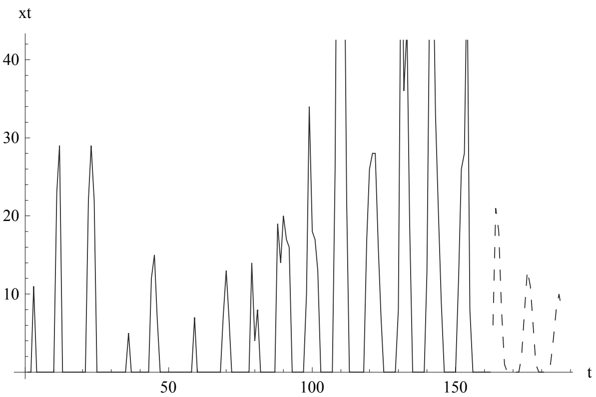


Figure 19. Time plot with last 25 years predicted values (dotted line)

Here we consider the estimation of a probability density function of the residuals. It is seen that the density function of the residuals is estimated by a *Laplace* distribution with parameters $\{\mu, \beta\}$ = {*-0.076, 0.981*} whose density function is

$$g(x) \;=\; \frac{1}{2\beta}\ \mathrm{Exp}\left\{-\frac{(x-\mu)\,\mathrm{Sign}[x-\mu]}{\beta}\right\}.$$

The function is plotted in Figure 20 with a histogram of the residuals.

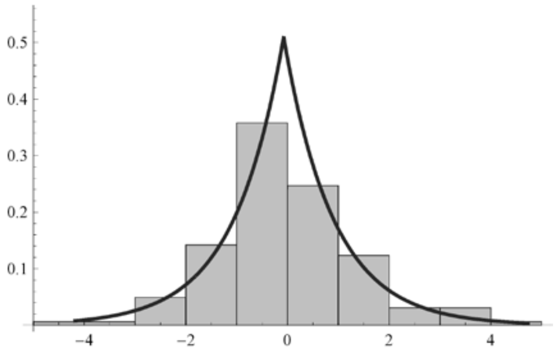


Figure 20. Histogram of the residuals of the AR(13) model and the density plot of the *Laplace* distribution with $\{\mu, \beta\} = \{-0.076, 0.981\}$.

The corresponding density estimate seems to be almost consistent with the histogram of the data. The standard diagnostic graphical check of the quantile plot is shown in Figure 21 and *6* statistical goodness of fit tests are given in Table 2. They do not show that the *Laplace* distribution model fits well to the data.

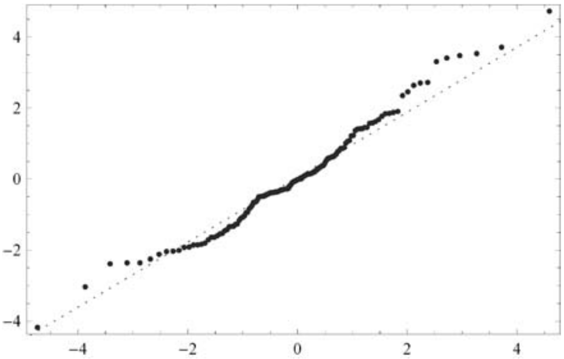


Figure 21. Quantile plot for the *Laplace* distribution function fitted to the residuals.

Table 2.

| | Statistic | P-Value |
|---|---|---|
| Anderson-Darling | *0.656* | *0.596* |
| Cramér-von Mises | *0.089* | *0.641* |
| Kolmogorov-Smirnov | *0.066* | *0.462* |
| Kuiper | *0.111* | *0.228* |
| Pearson $\chi^2$ | *30.20* | *0.011* * |
| Watson $U^2$ | *0.087* | *0.346* |

## Conclusions

We have considered a statistical time series modelling of the extreme values of a *daily* averages of the International relative sunspot number. We focused on two time series of annual *maximum* and *minimum* sunspot numbers, and individually we considered a fitting of an extreme value distribution function to the residuals derived from an AR model. The *AR(9)* model was fitted to the series, and it showed that the next peak of the solar cycle will be around in *2014* with a sunspot number of *204*. This result is almost consistent with the result for the yearly averaged sunspot number given in Tanaka[7]. Also we showed that the *GEV* (*generalized extreme value)* distribution function was fitted well to the residuals obtained from the *AR(9)* model.

On the series of annual minimum sunspot numbers, we have fitted the *AR(13)* model to the series. From the best linear prediction based on the model *AR(13)*, it is seen that the next peak will be around in *2013* with a sunspot number *21.* To the residuals obtained from the *AR(13)* model not a *GEV* distribution but a *Laplace* distribution was fitted better. But the result was not adequate because the model did not passed one of goodness of fit tests (*Pearson* $\chi^2$). Therefore finding a best distribution function for the annual minimum sunspot numbers must be a future work for us.

## Appendix

We here consider a question that there is a relationship between the series of the yearly averaged sunspot number (Figure A1) and the famous series of annual number of the *Lynx*, or not. The series of *Lynx* (Figure A2) has been also modeled by many researchers (see, for example, Moran[8], Pourahmadi[9]). The *Lynx* data seems to oscillate with a period of *10* (*9.82* year) although the sunspot numbers has a period of *11* (*11.42* year). In order to make the variance more uniform, taking the logarithm of the data and subtracting out a mean, we can get an AR(*12*) model for the *Lynx* data (see He[5], p.154). On the other hand, the series of the yearly sunspot data with the same time interval is fitted by an AR(*9*) model (see Tanaka[7]). Further we tried to calculate the sample cross-correlations between the two series, but we could not find clear relations.

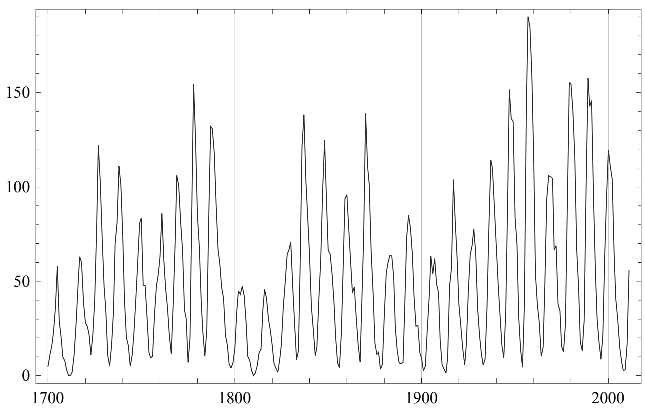


Figure A1. The yearly averaged sunspot numbers from 1700 to 2011.

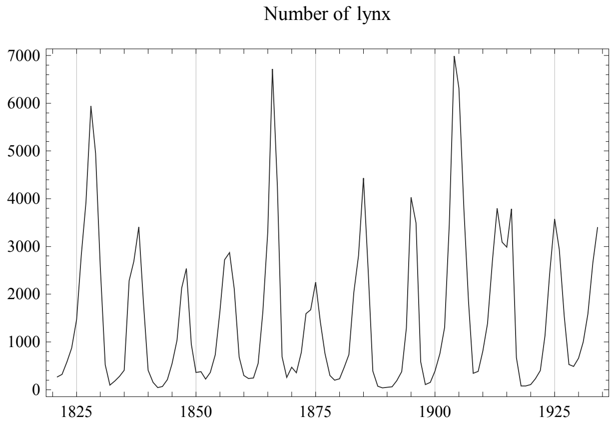


Figure A2. The number of *lynx* trapped annually in northwest Canada from 1821 to 1934.

# References

[1] Blockwell,P.J., Davis,R.A. (2002), *Introduction to Time Series and Forecasting*, Springer Verlag, New York.

[2] Coles, S.G. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer Verlag, New York.

[3] Cowpertwait.P.S.P., Metcalfe.A.V. (2009), *Introductory Time Series With R*, Springer Verlag, New York.

[4] Embrechts,P., Kluppelberg,G., & Mikosch,T. (1998), *Modelling Extremal Events for Insurance and Finance*, Springer Verlag, New York.

[5] He,Y. (1995), *Time Series Pack for Mathematica*, Wolfram Research.

[6] Leadbetter,M.R., Lindgren,C., & Rootzen,H. (1983), *Extremes and Related Properties of Random Sequences and Series*, Springer Verlag, New York.

[7] Tanaka, Minoru. (2010), *Time series modelling of Annual Maximum Sunspot Numbers*, Information Science and Applied Mathematics, Vol.18, 19-32.

[8] Moran, P. (1953), *The statistical analysis of the sunspot and lynx cycles*, J. Animal Ecol. 1, 163-173.

[9] Pourahmadi, H. (2000), *Foundations of Time Series Analysis and Prediction Theory*, Wile, New York.

[10] Thomas, J.H., Weiss, N.O. (2008), *Sunspots and Starspots*, Cambridge University Pree.