Estimation of the Hurst Exponent for the Daily Sunspot Number

Minoru Tanaka

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

Abstract. This paper deals with the estimation of the Hurst exponent of the International relative sunspot number data. We focus on three data sets of the yearly, monthly and daily sunspot numbers, and individually we consider the estimation of the Hurst exponent of them by use of three methods, (1) **Variance plot;** (2) **Rescaled range method (R/S);** and (3) **Partial sum of the absolute auto-covariances**. We compare these results.

Keywords: *the daily sunspot number, Hurst exponent, R/S statistics, autocovariance function*

1. Introduction

The Hurst exponent for a data set is known that it provides a kind of measure of whether the data is a short memory process or is a long memory process. The long memory process has the very long autocorrelations.

“There are a variety of techniques for calculating the Hurst exponent. The accuracy of the estimation can be a very complicated issue.” (see, Ian Kaplan, www.bearcave.com).

In Matsuba[6], the Hurst exponent of the yearly sunspot number (1700~2004) is estimated by $\hat{H} = 0.813$ by R/S statistics (given in Section2 Method (2) for this paper). Also for the data of monthly sunspot number (1749.1~2005.9), $\hat{H} = 0.745$ by using the partial sum of the autocorrelation function (Method (3) in Section2) (see also Fanchiotti, etc.[4]). From these results it is seen that the data set of the sunspots must be a long memory process.

In this paper our main object is to estimate the Hurst exponent of the daily sunspot number. Here we use a daily averages of the International Sunspot Number (published in Solar Influences Data Analysis Center (*SIDC*) in Belgium, Source: WDC-SILSO, Royal Observatory of Belgium, Brussels). It should be noted that daily values for years prior to *1849* are partly missing. The available data for this paper is for the period *1 January 1868* through *31 December 2016* plotted in Figure 1.3.

.

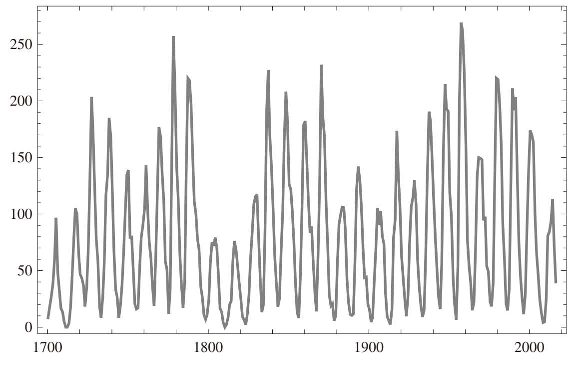

Figure 1.1. The yearly sunspot number from *1700* through *2016*.

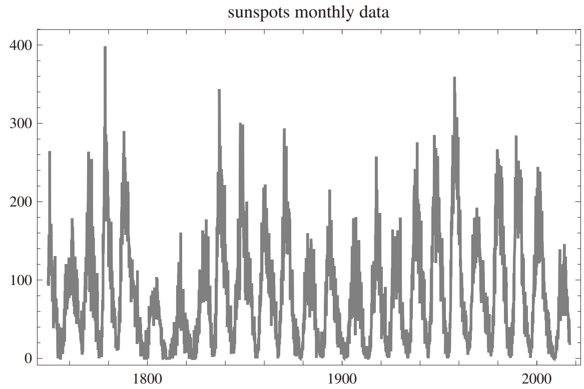

Figure 1.2. The monthly sunspot number from *1749.1* through *2016.12*.

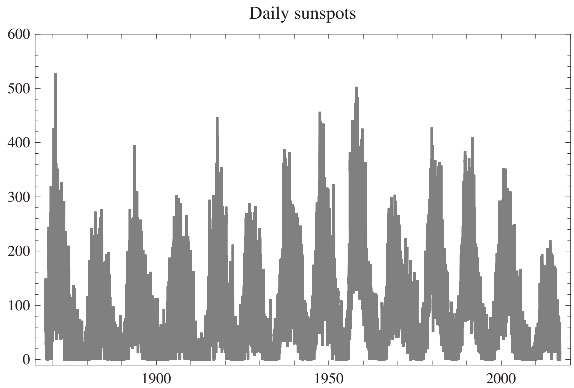

Figure 1.3. The daily sunspot number from *1 Jan. 1868* through *31 Dec. 2016*.

Similar to the yearly and also the monthly averaged sunspot numbers in Figure 1.1-2, the level of the daily sunspot number seems to oscillate with an approximate period of *11* (see for example, Cowpertwait [3] and Thomas [7]). But the series is fluctuating widely and sharply, and it has many zeros (see Figure 1.3) and the ratio of zero number is about *15%*. So it looks more difficult to get an appropriate estimator of the Hurst exponent directly than those of the yearly and the monthly series.

In Section 2 we introduce three heuristic methods for detecting and assessing the strength of long-memory, and define the Hurst exponent (coefficient, or parameter). In Section 3, we discuss data examples and consider the estimation of the Hurst exponent for the data sets; the yearly, monthly and daily sunspot numbers, independently.

This paper is supported by the computer software *Mathematica V11.0* and also its application *Time Series Pack for Mathematica* ([5]).

2. Methods for estimating the Hurst exponent parameter.

Let $\{X_t\}$ be a stationary linear process defined by

$$X_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j}, \qquad (2.1)$$

where ϵ_t ($t \in Z$) are independent identically distributed variables with variance σ_ϵ^2. The autocovariance function of X_t is then given as

$$\gamma_X(h) = \sigma_\epsilon{}^2 \sum_{j=0}^{\infty} a_j a_{j+h} \quad (h \in Z). \qquad (2.2)$$

If the L^2- linear process X_t has a condition such that, for sufficient large j,

$$a_j = O\left(j^{d-1}\right) \quad (0 < d < 1/2), \qquad (2.3)$$

then the process is called a long-memory process. In this case the autocovariance function has the condition, for sufficient large h,

$$\gamma_X(h) = O\left(h^{2d-1}\right) \quad (0 < d < 1/2) \qquad (2.4)$$

(see, for example, Beran [1], Blockwell [2] and Matsuba [6]).

The autocorrelation function $\rho_X(h) = \gamma_X(h)/\gamma_X(0)$ also has the condition (2.4).

It is notable that the definition of the long memory is an asymptotic matter, therefore it is often difficult to detect and quantify by use of finite samples.

The Hurst exponent (parameter) H is equal to $d + 1/2$. If the process is a long-memory process, then $1/2 < H < 1$. In the next section, we shall deal with the estimation of the Hurst exponent of the daily sunspot number plotted in Figure 3 from *1869.1.1* to *2016.12.31*.

There are many methods to estimate the Hurst exponent.

Following Beran [1], we consider the following three methods (these methods are mainly useful for descriptive purposes).

(1) **Variance plot**

Dividing the series $\{X_t\}$ into m non-overlapping, adjacent blocks of length k, where the length of the series is $n = [m\,k]$, then

$$\overline{X}_k(j) = \frac{1}{k} \sum_{t=1+(j-1)k}^{jk} X_t \quad (j = 1, 2, 3, \ .. m), \qquad (2.5)$$

$$S^2(k) = \frac{1}{m-1} \sum_{j=1}^{m} \left(\overline{X}_k(j) - \overline{x}\right)^2, \qquad (2.6)$$

where $\overline{x}$ is a overall mean. It is known that when k→∞,

$$\mathrm{Var}\left(\overline{X}_k(j)\right) \sim \mathrm{Const}\, k^{-\beta} \left(= \mathrm{Const}\, k^{2H-2}\right). \qquad (2.7)$$

To estimate the parameter H we calculate $S^2(k)$ for k = 2,3,..., [n/2], and plot log $S^2(k)$ against *log k*. Then the slope of the regression line will be an estimator of the $-\beta = 2H-2$. When k→∞,

$$\log S^2(k) \sim \mathrm{Const} + (2H - 2)\log(k). \qquad (2.8)$$

(2) **Rescaled range method (R/S)**

The rescaled range statistics was first introduced by Hurt (1951), and a simpler expression form is

$$R_n = \max_{1\le k\le n} \sum_{t=1}^{k} (X_t - \overline{x}_n) - \min_{1\le k\le n} \sum_{t=1}^{k} (X_t - \overline{x}_n) \quad (2.9)$$

$$S_n{}^2 = \frac{1}{n-1} \sum_{t=1}^{n} (X_t - \overline{x}_n)^2 \quad (2.10)$$

Limiting properties of the R/S statistics were investigated by Mandelbrot (1975) (see Beran [ ]). Under some conditions, it is seen that, as n→∞,

$$E\{R_n/S_n] \sim \text{Const} * n^H, \quad (2.11)$$

where H is known to Hurst exponent (parameter). Taking the logarithm on both side of (2.4), we have

$$\text{Log}\,(\, E[R_n/S_n]\,) \sim \text{Const} + H \log(n). \quad (2.12)$$

Therefore, the parameter H is interpreted as the slope of a regression line of log (Rn/Sn) against log n. It should be notable that the R/S method has a practical problem that it is not robust against departures from stationary of the series (see Beran [1]).

(3) **Partial sum of the absolute autocovariance function**

For the sample autocovariance function

$$\hat{\gamma}(j) = \frac{1}{n-j} \sum_{t=1}^{n-j} (X_t - \overline{X})(X_{t+j} - \overline{X}), \quad (2.13)$$

we denote the partial sum of the absolute autocovariance function

$$S_0(k) = \sum_{j=0}^{k} \hat{\gamma}(j) \quad (2.14)$$

$$S_3(k) = \sum_{j=0}^{k} \left| \hat{\gamma}(j) \right| \quad (2.15)$$

Then it is known that when k→∞,

$$S_3(k) \sim \text{Const}\, k^{2d} \left(= \text{Const}\, k^{2H-1} \right). \quad (2.16)$$

The sample autocorrelation function $\hat{\rho}_X(j) = \hat{\gamma}_X(j)/\hat{\gamma}_X(0)$ also has the condition (2.16).

When we plot log $S_3(k)$ against log k, we can get the slope of the regression line that will be an estimator of the *2H-1*.

$$\log S_3(k) \sim \text{Const.}+(2H-1)\log(k). \quad (2.17)$$

3. Data examples; the Hurst exponent of the Sunspot number data

We calculate the three estimators of the Hurst exponent defined in Section 2 for the yearly, monthly and daily sunspot numbers each .

▫ *3.1 Yearly sunspot number*

(1) **Variance Plot**

We calculate $S^2(k)$ for $k = 2,3,...,80$, and plot $S^2(k)$ against k in Figure 3.1.1.

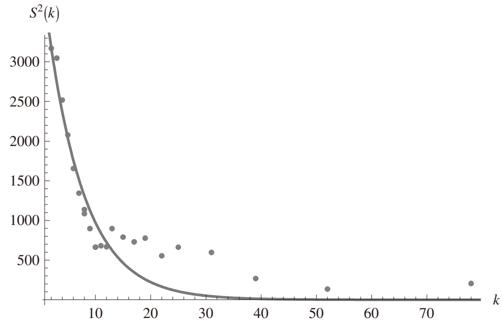

Figure 3.1.1 Plot of $S^2(k)$ against k and the fitted exponential curve.

The out-put from Mathematica program (NonlinearModelFit) is given as

```
fit = NonlinearModelFit[listv01, a*e^(-b*x), {a, b}, x]
       最適合非線形モデル
fit[{"BestFit", "RSquared", "ParameterTable"}]
fit["ParameterConfidenceIntervals"]
```

FittedModel[$4271.15\, e^{-0.14789\, x}$]

$\{4271.15\, e^{-0.14789\, x}$, 0.953854,

	Estimate	Standard Error	t-Statistic	P-Value
a	4271.15	393.095	10.8654	4.44279×10^{-10}
b	0.14789	0.0158655	9.32145	6.53476×10^{-9}

$\}$

```
{{3453.67, 5088.64}, {0.114896, 0.180884}}

Solve[2 H0 - 2 == -0.148, H0]
解く

{{H0 → 0.926}}
```

In the case when $k > 13$, the estimated exponential function is not fitted well, and thus the estimated exponent number may be large. Thus the estimate $\hat{H} = 0.926$ may not be good, it seems to be too large (too near *1.0*).

On the other hand, we transform the values $\{S^2(k)\}$ to the logarithm $\{log[S^2(k)]\}$, and fit a regression line to this data. Figure 3.1.2 shows that the regression line fits well.

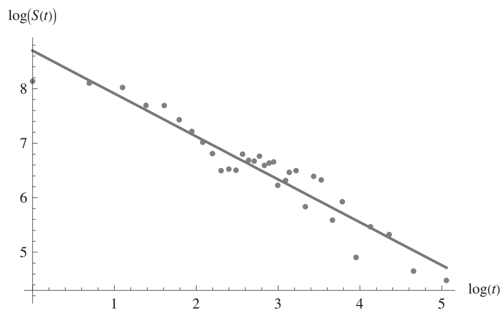

Figure 3.1.2 $log\, S^2(t)$ vs. $log\, t$ and the regression line.

The out-put from Mathematica program (LinearModelFit) is given as

```
fit = LinearModelFit[listr0, {x}, x]
      線形モデルフィット
fit[{"BestFit", "RSquared", "ParameterTable"}]
fit["ParameterConfidenceIntervals"]
```

FittedModel[8.69937 − 0.787918 x]

{8.69937 − 0.787918 x, 0.912876,

	Estimate	Standard Error	t-Statistic	P-Value
1	8.69937	0.130455	66.6851	4.77123×10^{-35}
x	−0.787918	0.0437183	−18.0226	5.53915×10^{-18}

}

{{8.43331, 8.96543}, {−0.877083, −0.698754}}

The linear function is *log(S(t)) = 8.70 - 0.79 log(t)* with R-squared *0.913*. The fitted slope is close to $\hat{\beta} = -0.788$ with the 95% confidence interval *(-0.877, -0.699)*, and this implies the Hurst exponent is estimated to *0.606* with the 95% confidence interval *(0.562, 0.651)*.

(2) **Rescaled range method (R/S)**

Let $Q_k = R_k / S_k$ in (2.4).

The case when $k = 317$, we have the 4 statistics

$\{j, Q_k, H0, R_k, S_k\}$

={1, 33.4097, 0.609291, 2071.36, 61.9988}.

Also when $k = 158$, we have two sets of statistics

{{1, 25.3397, 0.638481, 1492.96, 58.9178}, {2, 26.0503, 0.643944, 1680.07, 64.4932}}.

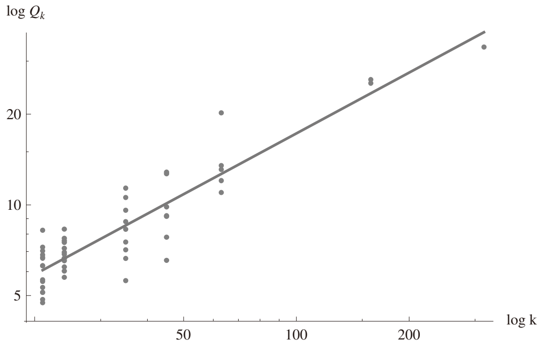

Figure 3.1.3 *log* Q_k vs. *log k* and the regression line.

The out-put from Mathematica program (LinearModelFit) is given as

FittedModel[−0.235883 + 0.669832 x]

{−0.235883 + 0.669832 x, 0.993265,

	Estimate	Standard Error	t-Statistic	P-Value
a	−0.235883	0.15434	−1.52833	0.132732
b	0.669832	0.0434609	15.4123	1.24987×10^{-20}

}

```
{{-0.545885, 0.0741185}, {0.582538, 0.757126}}
```

Then the slope of a regression line of *log* (Rn/Sn) against *log n* is *0.670* with *95%* confidence interval *(0.583, 0.757)*. This shows the estimator $\tilde{H}$ =*0.67* with 95% confidence interval *(0.583, 0.757).*

(3) **Partial sum of the absolute autocovariance function**

we consider the sample autocorrelation function of yearly sunspot number instead of the autocovariance function.

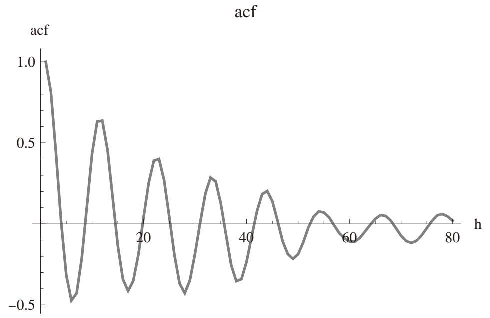

Figure 3.1.4 The sample autocorrelation function.

The sample autocorrelations $\hat{\rho}(k)$ in Figure 3.1.4 decay slowly with increasing *lag h*. This phenomenon indicates log memory, or long-range correlations. Also the sample power spectral density in Figure 3.1.5 has a frequency *0.58* (a period *10.81* year).

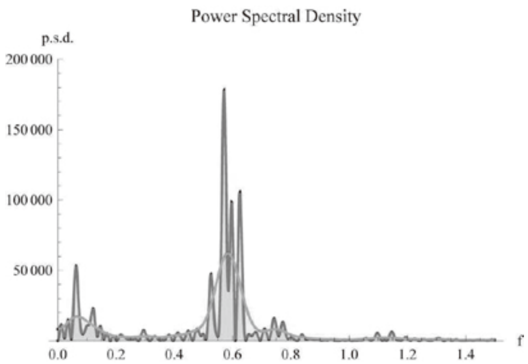

Figure 3.1.5 The sample power spectral density.

We plot *log* $S_0(k)$ against *log k*, for $k=1, 2, ,.., 80$, for autocorrelations. Figure 3.1.6 does not indicate the condition (2.15).

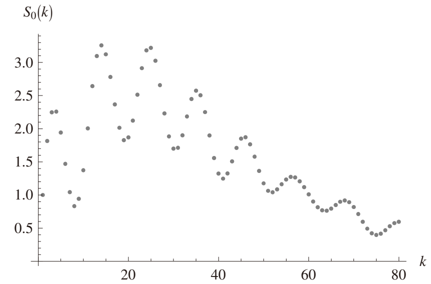

Figure 3.1.6 Plot of partial sums $S_0(k)$ for $k=1,2,3,..., 80$.

We then plot $S_3(k)$ against k, for $k=1, 2, ,.., 80$, for the absolute autocorrelations in Figure 3.1.7.

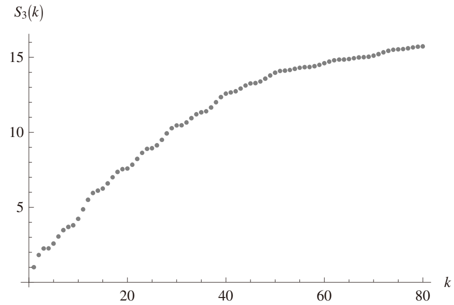

Figure 3.1.7 Plot of partial sums $S_3(k)$ for $k=1,2,3,..., 80$.

Next we plot $log\ S_3(k)$ against $log\ k$, for $k=1, 2, ,.., 80$, for the absolute autocorrelations. Figure 3.1.8 will indicate the condition (2.16).

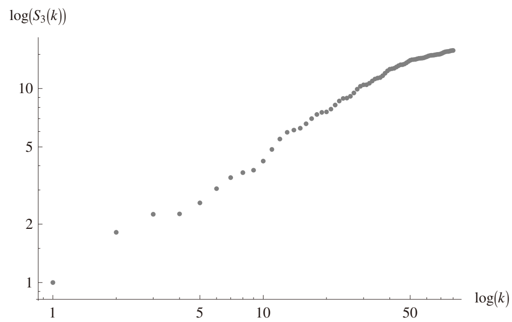

Figure 3.1.8 $\log S_3^2(k)$ vs. $log\ k$.

We then fit a linear function of time t to the data. The regression line is $log\ S_3(k) = 0.075 + 0.641\ log\ k$. It is plotted in Figure 3.1.9.

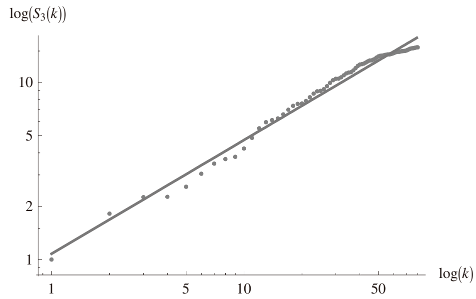

Figure 3.1.9 The log-log-plot with a regression line.

The out-put from Mathematica program (LinearModelFit) is given as

{0.0751236 + 0.641493 Log[t], 0.984106,

	Estimate	Standard Error	t-Statistic	P-Value
1	0.0751236	0.0326787	2.29885	0.0241909
Log[t]	0.641493	0.00923071	69.4955	6.3913×10^{-72}

}

{{0.0100652, 0.140182}, {0.623116, 0.65987}}

The fitted slope is close to $\hat{\beta} = 0.641$, and this implies the Hurst exponent is close to *0.821* with the 95% confidence interval *(0.812, 0.830)*.

▫ *3.2 Monthly sunspot number*

(1) **Variance Plot**

We fit a exponential function of k to the data S(k).

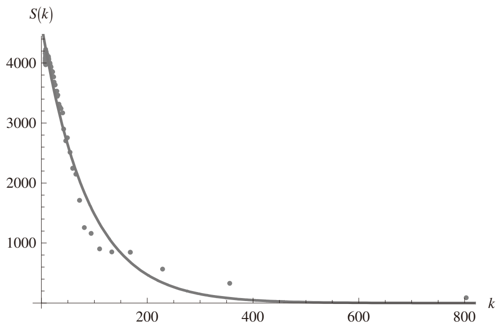

Figure 3.2.1 *S(k)* vs. *k* and the fitted exponential curve.

The out-put from Mathematica program (NonLinearModelFit) is given as

FittedModel[$4648.9\, e^{-0.011426\, x}$]

{$4648.9\, e^{-0.011426\, x}$, 0.997904,

	Estimate	Standard Error	t-Statistic	P-Value
a	4648.9	37.6669	123.421	3.40637×10^{-91}
b	0.011426	0.000379672	30.0943	1.103×10^{-44}

}

	Estimate	Standard Error	Confidence Interval
a	4648.9	37.6669	{4573.91, 4723.89}
b	0.011426	0.000379672	{0.0106701, 0.0121818}

```
Solve[2*H - 2 == -0.0114, H]
解く
0.5 {-0.012 + 2, -0.0107 + 2}
{{H → 0.9943}}
{0.994, 0.99465}
```

The fitted function is $S(k) = 4648.9\, e^{-0.011 x}$. It is plotted in Figure 3.2.1.

Then we have $\hat{H} = 0.994$ with the 95% confidence interval *(0.994, 0.9945)*.

Next, we transform the values $\{S^2(t)\}$ to the logarithm $\{\log S^2(t)\}$, and fit a regression line to this series shown in Figure 3.2.2.

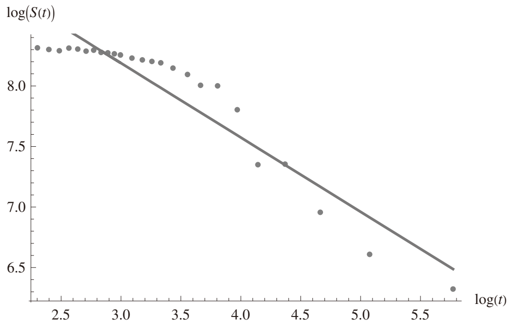

Figure 3.2.2 log-log-plot of $\log S^2(t)$ vs. $\log t$.

The out-put from Mathematica program (LinearModelFit) is given as

FittedModel[10.0297 − 0.613815 x]

{10.0297 − 0.613815 x, 0.892909,

	Estimate	Standard Error	t-Statistic	P-Value
1	10.0297	0.155167	64.6382	1.61679×10^{-27}
x	−0.613815	0.0443247	−13.8481	1.20514×10^{-12}

}

	Estimate	Standard Error	Confidence Interval
1	10.0297	0.155167	{9.70872, 10.3507}
x	−0.613815	0.0443247	{−0.705507, −0.522122}

The linear function is $\log S(t) = 10.030\text{-}0.614 \log t$. The fitted slope is close to $\hat{\beta} = -0.614$, and this implies the Hurst number is estimated to *0.693* with the 95% confidence interval *(0.647, 0.739)*.

(a) In the case for $k < 4000$, the regression line is $\log Q(k) = 8.848 - 0.208 \log k$ shown in Figure 3.2.3.

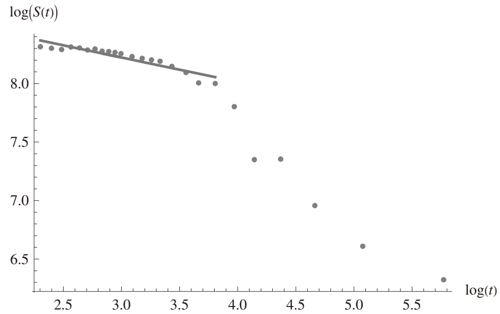

Figure 3.2.3 log-log-plot of *log* $S^2(t)$ vs. *log t*.

The out-put from Mathematica program (LinearModelFit) is given as

```
fit = NonlinearModelFit[listr0, a + b x, {a, b}, x]
  最適合非線形モデル

FittedModel[ 8.8478 - 0.208286 x ]

fit[{"BestFit", "RSquared", "ParameterTable"}]
```

{8.8478 − 0.208286 x, 0.999981,

	Estimate	Standard Error	t-Statistic	P-Value
a	8.8478	0.0615639	143.717	1.14449×10^{-27}
b	−0.208286	0.0203709	−10.2247	1.11676×10^{-8}

}

```
fit["ParameterConfidenceIntervalTable"]
```

	Estimate	Standard Error	Confidence Interval
a	8.8478	0.0615639	{8.71791, 8.97769}
b	−0.208286	0.0203709	{−0.251265, −0.165307}

```
Solve[2*H - 2 == -0.208, H]
解く
0.5 {-0.251 + 2, -0.165 + 2}
{{H → 0.896}}
{0.8745, 0.9175}
```

We have $\hat{H} = 0.896$ with the 95% confidence interval *(0.875, 0.918).*

(b) In the case for $4000 < k < 10000$, the regression line is $\log Q(k) = 11.143 - 0.867 \log k$ in Figure 3.2.4

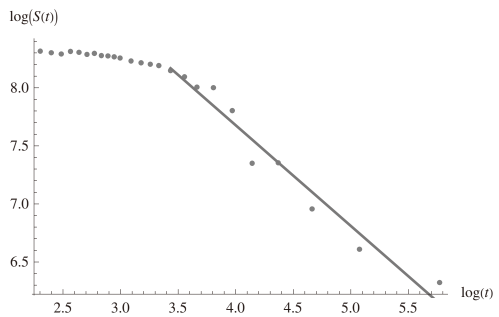

Figure 3.2.4 *log* $S^2(t)$ vs. *log t* and regression line.

The out-put from Mathematica program (LinearModelFit) is given as

```
fit = NonlinearModelFit[listr0, a + b x, {a, b}, x]
      最適合非線形モデル

FittedModel[ 11.1434 - 0.866603 x ]

fit15[{"BestFit", "RSquared", "ParameterTable"}]
```

{11.1434 − 0.866603 x, 0.961096,

	Estimate	Standard Error	t-Statistic	P-Value
1	11.1434	0.265268	42.0079	1.13632×10^{-10}
x	−0.866603	0.0616436	−14.0583	6.36355×10^{-7}

}

```
fit15["ParameterConfidenceIntervalTable"]
```

	Estimate	Standard Error	Confidence Interval
1	11.1434	0.265268	{10.5317, 11.7551}
x	−0.866603	0.0616436	{−1.00875, −0.724452}

```
Solve[2 * H - 2 == -0.867, H]
解く
0.5 {-1.00875 + 2, -0.724452 + 2}
{{H → 0.5665}}
{0.495625, 0.637774}
```

We have $\hat{H} = 0.567$ with the 95% confidence interval *(0.496, 0.638)*.

(2) **Rescaled range method (R/S)**

The regression line is $Q(k) = -0.105 + 0.773\ k$ in Figure 3.2.5.

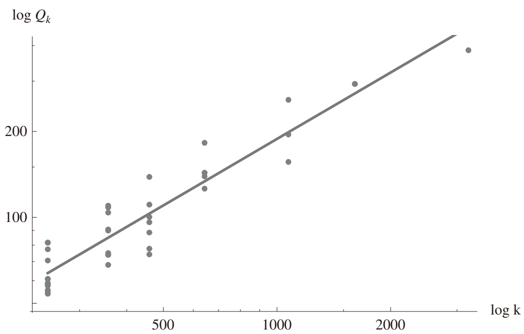

Figure 3.2.5 log-log-plot with a regression line.

The out-put from Mathematica program (LinearModelFit) is given as

```
Grid[Transpose[{#, fit[#]} &[{"AdjustedRSquared", "RSquared"}]], Alignment → Left]
格子  転置                                                    整列       左

AdjustedRSquared  0.998595
RSquared          0.998665
```

```
fit["ParameterConfidenceIntervalTable"]
```

	Estimate	Standard Error	Confidence Interval
a	−0.105072	0.272916	{−0.657562, 0.447419}
b	0.773379	0.0445996	{0.683092, 0.863666}

We have $\hat{H} = b = 0.773$ with the 95% confidence interval *(0.683, 0.864)*.

(3) Partial sum of the absolute autocorrelation function :

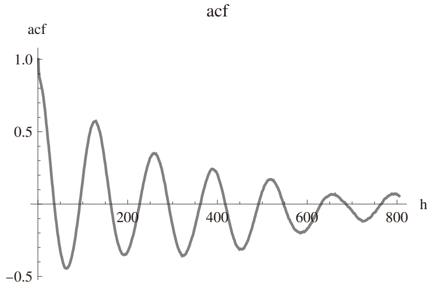

Figure 3.2.6 Sample acf of the monthly sunspot numbers.

The sample autocorrelations $\hat{\rho}(k)$ decay slowly with increasing *lag k*. This phenomenon indicates log memory, or long-range correlations.

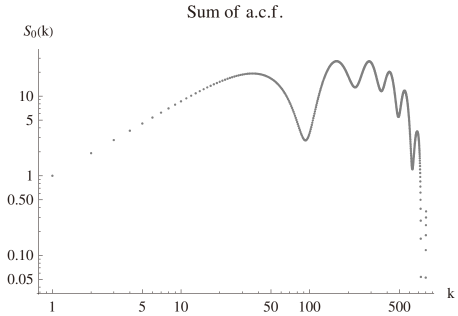

Figure 3.2.7 Partial sums $S_0(k)$ for $k=1,2,3,..., 804$.

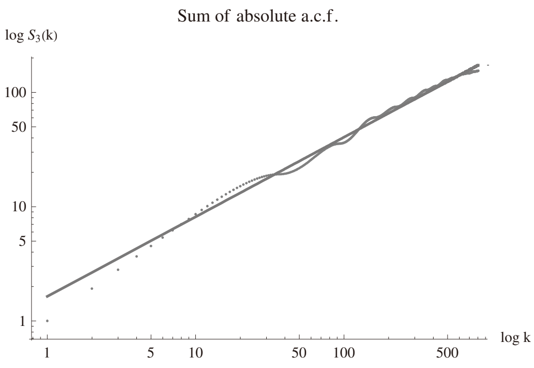

Figure 3.2.8 Regression line on $\log S_3(k)$ against $\log k$.

The out-put from Mathematica program (LinearModelFit) is given as

```
flog = LinearModelFit[Log[ac01], {1, Log[t]}, t]; Normal[flog]
       └線形モデルフィット └対数         └対数       └通常の式に変換
flog["RSquared"]
0.494386 + 0.69671 Log[t]
0.991336
flog["ParameterConfidenceIntervalTable"]
```

	Estimate	Standard Error	Confidence Interval
1	0.494386	0.013286	{0.468306, 0.520465}
Log[t]	0.69671	0.00229853	{0.692199, 0.701222}

```
flog["ParameterConfidenceIntervals"]
{{0.468306, 0.520465}, {0.692199, 0.701222}}
```

We fit a linear function of time k to the series. The linear function is $S_3(k)$ = *0.494+0.697 log k* and is plotted in Figure 3.2.8. "R-Squared" is *0.991*. The fitted slope is close to $\hat{\beta} = 0.697$, and this implies the Hurst number is close to *0.849* and the 95% confidence interval of the H will be *(0.846,0.850)*.

▫ *3.3 Daily sunspot number*

(1) **Variance Plot**

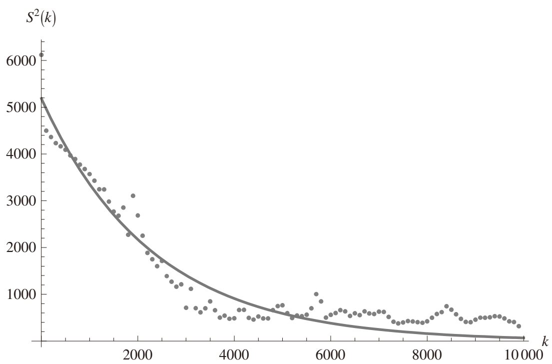

Figure 3.3.1 $S^2(k)$ vs. k and the fitted exponential curve.

The out-put from Mathematica program (LinearModelFit) is given as

```
fit = NonlinearModelFit[list000, a * e^(-b * x), {a, b}, x]
      └最適合非線形モデル
fit[{"BestFit", "RSquared", "ParameterTable"}]
fit["ParameterConfidenceIntervals"]

FittedModel[ 5186.71 e^(-0.000435469 x) ]
```

{5186.71 e^(−0.000435469 x), 0.964934,

	Estimate	Standard Error	t-Statistic	P-Value
a	5186.71	138.652	37.4082	7.91561×10^{-60}
b	0.000435469	0.0000168827	25.7939	1.78551×10^{-45}

}

```
{{4911.56, 5461.86}, {0.000401966, 0.000468973}}
Solve[2*H - 2 == -0.00044, H]
解く
{{H → 0.99978}}
```

It is seen that estimated Hurst exponent numbers $\hat{H} = 0.9998$ which seems to be too large.

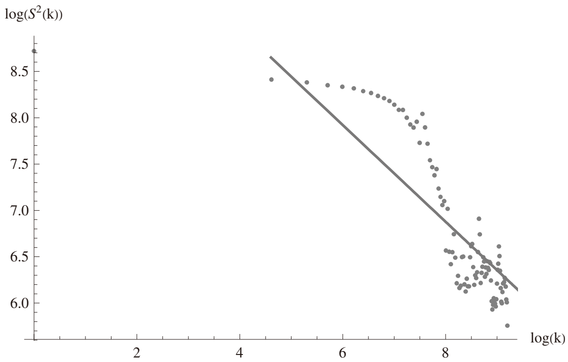

Figure 3.3.2 $log\ S^2(k)$ vs. $log\ k$ and the regression line.

The out-put from Mathematica program (LinearModelFit) is given as

```
fit = LinearModelFit[log20, {x}, x]
fit[{"BestFit", "RSquared", "ParameterTable"}]
fit["ParameterConfidenceIntervals"]
FittedModel[ 11.1253 - 0.531668 x ]
```

{11.1253 − 0.531668 x, 0.670453,

	Estimate	Standard Error	t-Statistic	P-Value
1	11.1253	0.310385	35.8437	3.92802×10^{-58}
x	−0.531668	0.0376532	−14.1201	2.3327×10^{-25}

}

```
{{10.5094, 11.7413}, {-0.60639, -0.456947}}
```

The linear regression line is $Log\ S^2(t) = 11.125 - 0.532\ Log(t)$. The estimate is $\hat{H} = 0.734$ with 95% confidence interval *(0.697, 0.772)*.

Next we fit the line in the case for $4000 < k < 10000$,

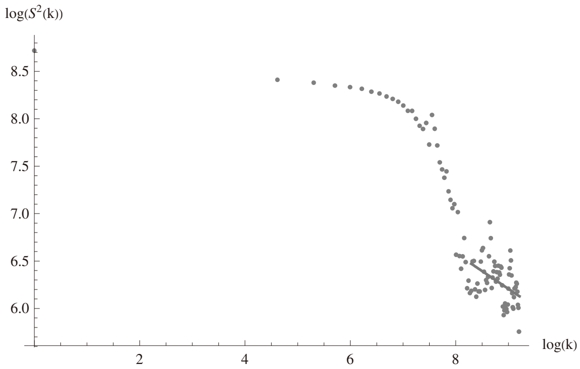

Figure 3.3.4 *log* $S^2(k)$ vs. *log k* and the regression line.

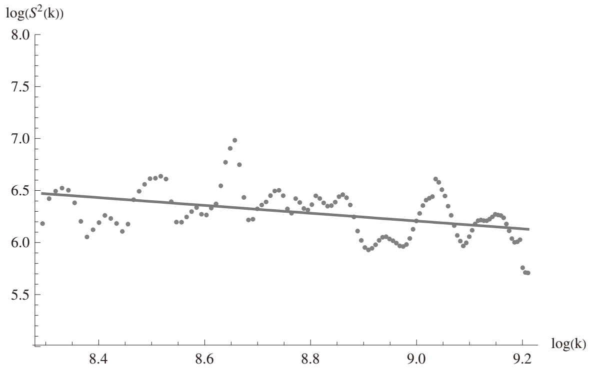

Figure 3.3.5 log-log-plot of *log* $S^2(k)$ against *log k.*

The out-put from Mathematica program (LinearModelFit) is given as

```
  fit = LinearModelFit[log20, {x}, x]
fit[{"BestFit", "RSquared", "ParameterTable"}]
fit["ParameterConfidenceIntervals"]
```

FittedModel[9.582 – 0.375218 x]

{9.582 − 0.375218 x, 0.190792,

	Estimate	Standard Error	t-Statistic	P-Value
1	9.582	0.625101	15.3287	6.05369×10^{-30}
x	−0.375218	0.070837	−5.29693	5.47585×10^{-7}

}

{{8.34423, 10.8198}, {−0.515483, −0.234954}}

```
Solve[2*H - 2 == -0.375, H]
解く
0.5 {2 - 0.515482846913814`, 2 - 0.23495408083363578`}
{{H → 0.8125}}
{0.742259, 0.882523}
```

In the case for *4000 < k <10000*, we can estimate $\hat{H}$ = *0.813* with 95% confidence interval *(0.742, 0.883)*. This estimate is bigger than that of the overall case ($\hat{H}$ = *0.734*).

(2) **Rescaled range method (R/S)**

(a) For the overall data, we have a regression line, *log Q(k) = 0.585 + 0.781 log k* .

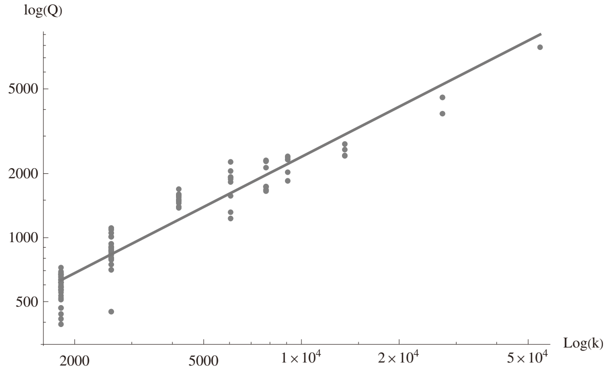

Figure 3.3.6 log-log-plot with a regression line.

The out-put from Mathematica program (LinearModelFit) is given as

```
fit = LinearModelFit[data00, {x}, x]
      線形モデルフィット
fit[{"BestFit", "RSquared", "ParameterTable"}]
fit["ParameterConfidenceIntervals"]

FittedModel[ 0.584722 + 0.781415 x ]
```

{0.584722 + 0.781415 x, 0.905163,

	Estimate	Standard Error	t-Statistic	P-Value
1	0.584722	0.218627	2.67452	0.00887239
x	0.781415	0.0265147	29.471	2.48221×10^{-48}

}

```
{{0.150446, 1.019}, {0.728747, 0.834083}}
```

Then we have the Hurst exponent estimator $\hat{H} = 0.781$ with the 95% confidence interval *(0.729, 0.834)*.

(b) For large *k* (*k* ⩾ *7776*), we have a regression line, *log Q(k) = 1.640 + 0.660 log k.*

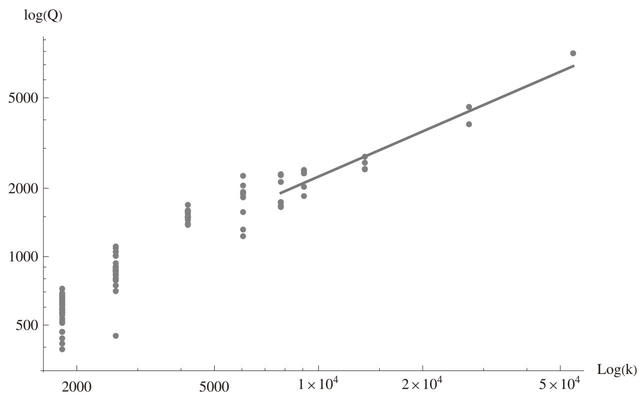

Figure 3.3.7 log-log-plot with a regression line.

The out-put from Mathematica program (LinearModelFit) is given as

```
fit7 = LinearModelFit[data007, {x}, x]
      └線形モデルフィット
fit7[{"BestFit", "RSquared", "ParameterTable"}]
fit7["ParameterConfidenceIntervals"]

FittedModel[ 1.64057 + 0.660101 x ]
```

{1.64057 + 0.660101 x, 0.894339,

	Estimate	Standard Error	t-Statistic	P-Value
1	1.64057	0.500243	3.27955	0.00416505
x	0.660101	0.0534788	12.3432	3.20066×10^{-10}

}

```
{{0.589602, 2.69154}, {0.547746, 0.772456}}
```

It is seen that $\hat{H}$ =*0.660* with the 95% confidence interval *(0.548, 0.772)*.

(3) **Partial sum of the absolute autocorrelation function**

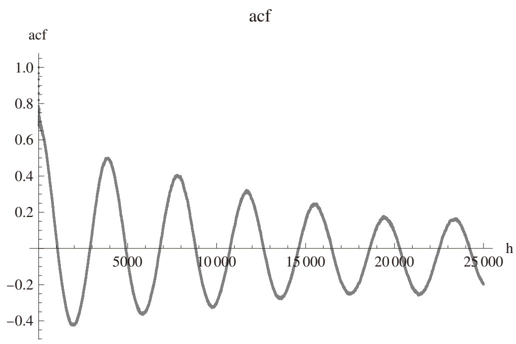

Figure 3.3.8 Sample acf of the daily sunspot numbers.

The sample autocorrelations $\hat{\rho}(k)$ decay slowly with increasing lag k. This phenomenon indicates log memory, or long-range correlations. It has a frequency *0.0016*, this implies that the series has a period *3927* (days), or *10.76* (years).

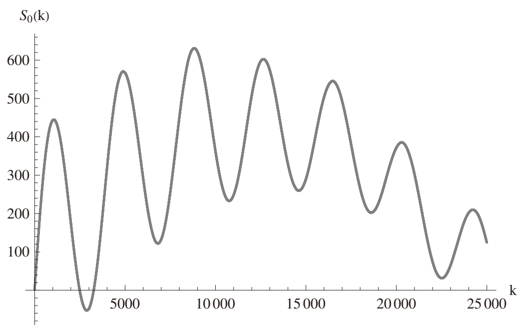

Figure 3.3.9 Plot of partial sums $S_0(k)$ for k=*1,2,3,..., 2500*.

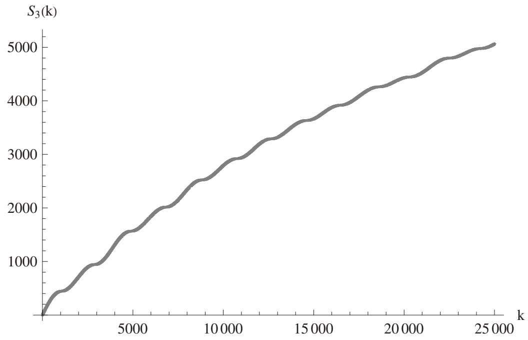

Figure 3.3.10 Plot of partial sums $S_3(k)$ for $k=1,2,3,..., 2500$.

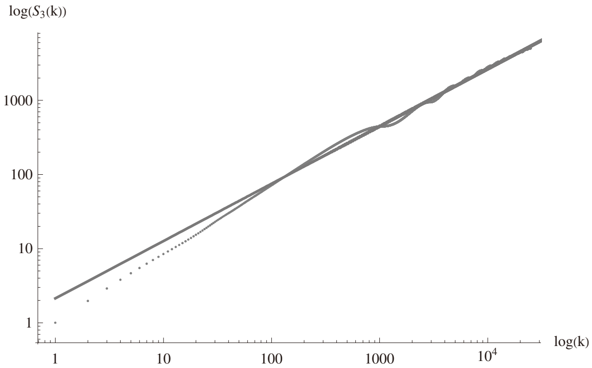

Figure 3.3.11 Regression line on $\log S_3(k)$ against $\log k$.

The out-put from Mathematica program (LinearModelFit) is given as

```
fit = LinearModelFit[Log[ac01], {1, Log[x]}, x]
      線形モデルフィット  対数        対数
fit[{"BestFit", "RSquared", "ParameterTable"}]
fit["ParameterConfidenceIntervals"]

FittedModel[ 0.756954 + 0.773614 Log[x] ]

{0.756954 + 0.773614 Log[x], 0.995346,
```

	Estimate	Standard Error	t-Statistic	P-Value
1	0.756954	0.00307194	246.409	$5.8577831991 \times 10^{-6692}$
Log[x]	0.773614	0.000334582	2312.18	$1.1302125854 \times 10^{-29153}$

```
}
{{0.750932, 0.762975}, {0.772958, 0.774269}}
```

We fit a linear function of $\log k$ to the series. The linear function is $\log S_3(k)= 0.757 + 0.774 \log k$. It is plotted in Figure 3.3.11. The estimated slope is close to $\hat{\beta} = 0.7736$, and this implies the Hurst number is close to *0.8868* with the 95% confident interval *(0.8865, 0.8871)*.

Conclusions

We have considered the estimation of the Hurst exponent of the International relative sunspot number data. We focused on three sunspot numbers, yearly, monthly and also daily sunspots. We individually estimated the Hurst exponent by use of three Methods (1) ~ (3) given in Section 2.

In Section 3 we identically estimated the Hurst exponent for each sunspots data. The results are given in Table 1 below. It is seen that as the sample size increases, the value of the Hurst exponent becomes large for each sunspots data.

The features of the three Methods are as follows:

Method (1) : This estimate will be least of these three methods for each data;

Method (2) : Sample size seems to be of no effect upon this estimator;

Method (3) : This estimate will be the most of three methods for each data.

Table 1. The estimated Hurst exponent values by Methods (1) ~ (3) for each sunspots data.

Method \ Data	Yearly	Monthly	Daily
Method (1) *	(*0.926*)	(*0.994*)	(*0.9998*)
Method (1)	*0.606*	*0.693*	*0.813*
Method (2)	*0.670*	*0.773*	*0.781*
Method (3)	*0.821*	*0.849*	*0.887*

The Method (1)* in Table 1 is the case when non-linear fitting: an exponential function is fitted directly to the series $S(k)$ each. This method may be not good, because the non-linear least squared estimation did not seem to work well, and all the estimated values of the Hurst exponent are too large (too near 1.0).

In Table 1, especially for Method (3), the Hurst exponent of the daily sunspot number is near *0.9* and this implies that the daily sunspot number must be exactly a long memory process.

It is known that there are many other methods for estimating the Hurst exponent (for example, KPSS statistic, Detrended Fluctuation Analysis and Temporal Aggregation, see Beran [1]). Statistical comparisons of these methods would be a future work for us.

References

[1] Beran, J., Feng,Y., Ghosh, S., & Kulik, R. (2013), *Long-Memory Processes*, Springer Verlag, New York.

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

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

[4] Fanchiotti, H., Sciutto,S.L., Canal,C.A.G., & Hojvat, C. (2004), Analysis of sunspot number fluctuations, Fractals, 12, 405-411.

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

[6] Matsuba, I., (2007), *Statistics for Long Memory Process - Theory and Method of Self-similar Time*, Kyoritsu Shuppan, Tokyo.

[7] Thomas, J.H., & Weiss, N.O. (2008), *Sunspots and Starspots*, Cambridge University Press.