DEVELOPMENT OF ADAPTIVE MOVING TWO-SIDED EXPONENTIAL SMOOTHING METHOD FOR RESTORING AND FORECASTING OF TIME SERIES

Olena Bratus
Department of Mathematical Methods of Systems Analysis
National Technical University of Ukraine
“Igor Sikorsky Kyiv Polytechnic Institute”
37 Peremohy ave., Kyiv, Ukraine, 03056
LenaBratus@i.ua

Abstract

Two algorithms for restoring of missing values of time series with using of adaptive moving two-sided exponential smoothing method with different initial conditions are developed in the article. Adaptive moving two-sided exponential smoothing method for restoring of true regularities and forecasting of time series is developed. The integral criterion of model adequacy and the proximity criterion for using for restoring of the true regularities of time series evolution are suggested. Practical researches with restoring of true regularities of Wolf numbers and solar radio fluxes at a wavelength of 10.7 cm, restoring of missing values and forecasting of solar radio fluxes at a wavelength of 10.7 cm are performed. Comparisons of created method with traditional methods are performed for all experiments. Developed adaptive moving two-sided exponential smoothing method is shown superiority in comparison with all traditional methods in the restoring of true regularities, missing values and forecasting of solar data.

Keywords: adaptive moving two-sided exponential smoothing method, restoring of true regularities of time series, restoring of missing time series values, integral criterion of model adequacy, proximity criterion, Wolf numbers, solar radio fluxes.

DOI: 10.21303/2461-4262.2017.00400

© Olena Bratus

1. Introduction

An urgent modern task of data processing is analysis, mathematical modeling and forecasting of nonlinear non-stationary processes, in particular data, which describes solar cycles. One of the well-known works, in which the model of the 11-year solar cycle is described, is the work of Hathaway and others [1]. These authors in a later work [2] proposed the method for calculating of 13-month moving average for processing of solar activity data. This method is widespread nowadays. Researches regarding the dynamics of the solar cycle are performed in the works [3–5]. In the work [6] it is shown that the Waldmeier effect is specific to Wolf numbers only. The historical development of the research of sunspot numbers, transition from Wolf numbers to the International index of sunspot numbers are shown in the works [7–10]. Solving the problem of uncertainty and finding accurate estimates for the sunspot numbers are also important, these researches are performed in the works [11–14]. The mathematical model for the Wolf number cycle using the Ian Wilson’s Tidal Torque theory is proposed by Salvador in the work [15].

Problem of solar data forecasting is actual nowadays. Hathaway and others are shown in the work [1] that proposed equation for describing of the 11-year solar cycle can be used for forecasting. Researches of using data about dynamics of solar radio fluxes for forecasting of solar cycles are described in the works [16–18]. Detailed review of many existing methods of solar data forecasting is performed in the work [2]. A Bayesian approach for forecasting solar cycles using a Fokker-Planck equation is proposed by Noble and Wheatland in the article [19]. It is shown in the work [15] that proposed model can be used both for describing of regularities and for forecasting of Wolf numbers. The state of the research regarding forecasting of solar cycles nowadays is shown and approach for increasing the accuracy of forecast is proposed in the work [20].

Researches regarding regularities and missing values of solar data restoring are relevant, because it is important to find more accurate estimates for their using in the further processing and forecasting of solar data. Forecasting of solar data is also actual problem nowadays, because finding of more accurate forecasting estimates is important for research of solar data.
2. Materials and Methods

Moving two-sided exponential smoothing method is proposed in the work [21]. In the work [21] this method is used on interval 13 to work with data in the same range as for the 13-month moving average. Moving two-sided exponential smoothing method is development of quasioptimal smoothing procedure, which is described in the work [22]. Moving two-sided exponential smoothing method, which is proposed in the work [21], is used on the 13 months interval, selection of this interval is performed on the basis of expert data. However, on practice the value of the optimal interval for this method may differ from the value of 13. Therefore in this paper adaptive moving two-sided exponential smoothing method is proposed, in which the variable interval and the smoothing factor are determined.

The first stage of adaptive moving two-sided exponential smoothing method is to obtain estimates using the traditional exponential smoothing method. The initial approximation is chosen as equal to the first value of time series from the appropriate interval on each moving interval. This procedure is described by the following formula:

\[ S_k = \alpha \times y(k) + (1 - \alpha) \times S_{k-1}, \]  

where \( S_k \) – value of exponential average in the \( k \)-th moment of time; \( \alpha \) – smoothing factor; \( y(k) \) – value of time series in the \( k \)-th moment of time; \( S_{k-1} \) – value of exponential average in the \( (k-1) \)-th moment of time [8], periods of time \( k = 1, \ldots, M, k = 2, \ldots, M + 1, \ldots, k = N - M + 1, \ldots, N \); \( M \) – value of moving interval.

In this method, it is suggested to choose moving interval among the following values: 3, 5, 7, 9, 11, 13. To determine the optimal value of the smoothing factor and the optimal moving interval, it is considered the following statistical characteristics, which are used for estimating of quality of constructed model in the whole: the coefficient of determination \( R^2 \), sum of squared errors of model \( SSE \), Durbin-Watson statistic \( DW \).

The coefficient of determination \( R^2 \) is calculated by the following formula:

\[ R^2 = \frac{\hat{\text{Var}}[y(k)]}{\text{Var}[y(k)]}, \]  

where \( \hat{\text{Var}}[y(k)] \) – dispersion of the main variable \( y(k) \), estimated by the model; \( \text{Var}[y(k)] \) – actual dispersion of the main variable \( y(k) \), calculated based on data sample [8].

Durbin-Watson statistic \( DW \) is calculated by the following formula:

\[ DW = 2 - 2 \times \rho, \]  

where \( \rho \) – coefficient of autocorrelation for \( e(k) \) and \( e(k-1) \), namely when offset is \( S = 1 \) [8]. Coefficient of autocorrelation \( \rho \) for \( e(k) \) and \( e(k-1) \) in the formula (3) is calculated as follows:

\[ \rho = \frac{\sum_{k=2}^{N} e(k) \times e(k-1)}{\sum_{k=1}^{N} (e(k))^2}, \]  

where \( e(k) \) – model error in the \( k \)-th moment of time; \( e(k-1) \) – model error in the \( (k-1) \)-th moment of time; \( N \) – number of values in the data sample [8].

Sum of squared errors of model \( SSE \) is calculated as follows [8]:

\[ SSE = \sum_{k=1}^{N} (y(k) - y(k))^2. \]
In this paper integral criterion of model adequacy is proposed:

$$I = |1 - R^2| + |2 - DW|,$$  \hspace{1cm} (5)

where $R^2$ – the coefficient of determination, which is calculated by the formula (2); DW – Durbin-Watson statistic, which is calculated by the formula (3).

The criterion of choosing optimal value of the smoothing factor is the minimum of the values of sum of squared errors of model $SSE$ and the integral criterion $I$, which are calculated, respectively, by the formulas (4) and (5). The criteria of choosing the optimal moving interval are similar to criteria of choosing of optimal value of the smoothing factor. Simultaneous estimations of the smoothing factor and moving interval are performed.

The second stage of adaptive moving two-sided exponential smoothing method is to apply to the obtained estimates the procedure of smoothing in reverse time at the each interval $k = N - 1, ..., N - M + 1, k = N - 2, ..., N - M, ..., k = M - 1, ..., 1$ [6]. Each refined value $S'_k$ is calculated by the following formula [6]:

$$S'_k = \alpha \times S_k + (1 - \alpha) \times S'_{k+1},$$  \hspace{1cm} (6)

where $S'_k$ – value of exponential average in the reverse time in the $k$-th moment of time; $\alpha$ – smoothing factor; $S_k$ – value of exponential average in the $k$-th moment of time; $S'_{k+1}$ – value of exponential average in the reverse time in the $(k+1)$-th moment of time, where $k = N - 1, ..., N - M + 1, k = M - 1, ..., 1$.

Initial values $S_1 = S_N, ..., S_M = S_M$.

The criterion of optimal smoothing factor choosing is the minimum of the values of sum of squared errors of model $SSE$ and the integral criterion $I$, calculated by the formulas (4) and (5), respectively, but $S_k$ values are used as initial values during calculation of these characteristics. When smoothing is performed in the reverse time, the moving interval is choosing the same, as when smoothing is performed in the direct time. Weights of $S_k$ estimates arrived to the input in reverse time, when refined $S'_k$ estimations are calculated.

Then, as an estimate of the value of the data sample, the resulting value is used in the middle point of each interval, that is, value $S'_{M+1}$ is used on the interval $k = 1, ..., M$, value $S'_{M+1}$ is used on the interval $k = 2, ..., M + 1$ and so on. Estimates for the first $\frac{M + 1}{2}$ and for the last $\frac{M + 1}{2}$ values of data sample are used as appropriate values of exponential average in the reverse time on the first and last estimation intervals respectively, namely $S'_1, ..., S'_M$ on the interval $k = 1, ..., M$ and $S'_{2N-M+2}, ..., S'_N$ on the interval $k = N - M + 1, ..., N$.

It is proposed to use adaptive moving two-sided exponential smoothing method for the time series forecasting in the following way: it is necessary to use the full described procedure and the calculations are performed by the formulas (1) and (6), with the value of the smoothing factor and the moving interval used the same as for all known values, the last known value is used as the initial value.

It is important to restore missing values of the time series to obtain the most accurate values instead of the number of missed ones for further processing, for example, for forecasting. Two algorithms with using of adaptive moving two-sided exponential smoothing method are proposed for restoring of missing time series values in this work.

**Algorithm No. 1**

If before the missing value there are from 1 to $\frac{M + 1}{2}$ values of time series, then for finding of smoothing factor and moving interval on the each stage of adaptive moving two-sided exponential smoothing method by the criterion of minimization of values, described by the equation (4) and (5), it is proposed to take the values of the time series, which are situated after the missing value. It is proposed to use $\frac{M + 1}{2}$ as the border value, because during applying of adaptive moving two-sid-
ed exponential smoothing method the estimation \( \frac{M+1}{2} \) is chosen on the first interval. Then, with the value of the initial approximation, equal to the arithmetic mean of the all known values, and using last known value before the missing value as known value it is proposed to apply procedure of adaptive moving two-sided exponential smoothing method, which is described by equations (1) and (6).

**Algorithm No. 2**

If before the missing value there are more than \( \frac{M+1}{2} \) values of time series, then for finding of smoothing factor and moving interval on the each stage of adaptive moving two-sided exponential smoothing method by the criterion of minimization of values, described by the equation (4) and (5), it is proposed to take the values of the time series, which are situated before the missing value. Then, with the value of the initial approximation, equal to the arithmetic mean of the all known values, and using last known value before the missing value as known value it is proposed to apply procedure of adaptive moving two-sided exponential smoothing method, which is described by the equations (1) and (6).

Thus, adaptive moving two-sided exponential smoothing method can be effectively applied to restore missing values of time series.

3. Experimental Procedures

On the basis of the data sample of monthly values of Wolf numbers from January 2013 to August 2016 (44 values), according to the official site of the National Oceanic and Atmospheric Administration of the USA [23], the true values of the time series are restored. Two methods are applied to this data: adaptive moving two-sided exponential smoothing method, which is described by the formulas (1), (4)–(6); 13-month moving average method, which is described by the following formula [2]:

\[
\hat{y}(k) = \frac{1}{24} \sum_{i=-6}^{5} y(k+i) + \frac{1}{24} \sum_{i=-3}^{6} y(k+i),
\]

where \( y(k) \) – monthly average value for the \( k \)-th month, \( \hat{y}(k) \) – 13-month moving average with the center in the \( k \)-th month.

To compare the obtained results, the proximity criterion, which consists of finding the minimum for the expression (4), is used.

On the basis of the data sample of monthly values of solar radio fluxes at a wavelength of 10.7 cm (hereinafter referred to as index F10.7) in the units of measurement SFU from January 2013 to August 2016 (44 values), according to the official site of the National Oceanic and Atmospheric Administration of the USA [23], the true values of the time series are restored. Two methods are applied to this data: adaptive moving two-sided exponential smoothing method, which is described by the formulas (1), (4)–(6); 13-month moving average method, which is described by the formula (7). To compare the obtained results, the proximity criterion, which consists of finding the minimum for the expression (4), is used.

On the basis of the same data sample of the monthly values of solar radio fluxes at a wavelength of 10.7 cm, according to the official site of the National Oceanic and Atmospheric Administration of the USA [23], research is performed for the artificial creation of missing values and restoring of missing values.

Restoring of missing values of the index F10.7 is performed using the following methods: adaptive moving two-sided exponential smoothing method, which is described by the formulas (1), (4)–(6), with using algorithms No. 1 and No. 2; exponential smoothing method, which is described by the formulas (1), (4), (5).

In order to estimate the quality of the constructed models as a whole, when restoring of missing values is performed, the statistical characteristics are calculated by the formulas (2)–(4). It is found that the value of the moving interval is 11 for all cases of restoring.
Case No. 1
In the data sample of the index F10.7 the artificial missing of three values with numbers N = 6, 7, 8 is performed and restoring using the described above methods is performed.

Case No. 2
In the data sample of the index F10.7 the artificial missing of three values with numbers N = 28, 29, 30 is performed and restoring using the described above methods is performed.

On the basis of the data sample of monthly values of solar radio fluxes at a wavelength of 10.7 cm in the units of measurement SFU from January 2013 to April 2016 (41 values) [23] forecasting of the values for the next 3 months is performed. Forecasting is performed using adaptive moving two-sided exponential smoothing method, which is described by the formulas (1), (4)–(6), and method, which is described by the following formulas [1]:

\[
f(i) = \frac{a \times (i-i_0)^1}{\exp((i-i_0)^2 / b^2) - c},
\]

\[b(a) = 27.12 + 25.15 / (a \times 10^3)^{1/4},\]

\[a(n) = 9.2 \times 10^{-4} + 8.0 \times 10^{-5} n,\]

where \(f(i)\) – function that described phases of growth and fall of solar data values; \(a\) – amplitude of cycle; \(i\) – moment of time, which measures in the months; \(b\) – time in the months from minimum to maximum; \(c\) – asymmetry of the cycle; \(i_0\) – the initial moment of time; \(n\) – number of cycle.

4. Results
The measured values of Wolf numbers and obtained results for each model for restoring of true values of time series are shown in the Fig. 1.

![Fig. 1. Values \(y(i)\) and \(\hat{y}(i)\) for the data sample of Wolf numbers, where -- measured values \(y(i)\); --- estimates \(\hat{y}(i)\), obtained by adaptive moving two-sided exponential smoothing method; ----- estimates \(\hat{y}(i)\), obtained by 13-month moving average method](image-url)
The value of the proximity criterion, which is calculated by the formula (4), is shown in the Table 1. Models are indicated as follows in the Table 1: adaptive moving two-sided exponential smoothing – AMTES; 13-month moving average – TMA.

<table>
<thead>
<tr>
<th>Criterion name</th>
<th>AMTES</th>
<th>TMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>6857,8</td>
<td>13459,1</td>
</tr>
</tbody>
</table>

The measured values of index F10.7 and the results, obtained for each of the models when the true values of the time series are restored, are shown in the Fig. 2.

![Fig. 2. Values \( y(i) \) and \( \hat{y}(i) \) for the data sample of index F10.7, where -- measured values \( y(i) \); --- estimates \( \hat{y}(i) \), obtained by adaptive moving two-sided exponential smoothing method; ... estimates \( \hat{y}(i) \), obtained by 13-month moving average method](image)

The value of the proximity criterion, which is calculated by the formula (4), is shown in the Table 2. Models are indicated in the same way in the Table 2, as in the Table 1.

<table>
<thead>
<tr>
<th>Criterion name</th>
<th>AMTES</th>
<th>TMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>2593,4</td>
<td>6528,3</td>
</tr>
</tbody>
</table>

The statistical characteristics, which show the quality of the constructed models as a whole for the case No. 1 at the restoration of the index F10.7 values, calculated by the formulas (2), (3), (4) on the interval \( N = 1, \ldots, 44 \), are shown in the Table 3. Models are indicated as follows in the Table 3: exponential smoothing – ES; adaptive moving two-sided exponential smoothing – AMTES. The statistical characteristics of the models are shown in the Table 4 for the case No. 2 at the restoration of index F10.7 in the same way, as in the Table 3.
### Table 3
Statistical characteristics of the models

<table>
<thead>
<tr>
<th>Statistical characteristic</th>
<th>ES</th>
<th>AMTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0,9162</td>
<td>0,9241</td>
</tr>
<tr>
<td>SSE</td>
<td>95,3291</td>
<td>91,7328</td>
</tr>
<tr>
<td>DW</td>
<td>1,8552</td>
<td>1,9171</td>
</tr>
</tbody>
</table>

### Table 4
Statistical characteristics of the models

<table>
<thead>
<tr>
<th>Statistical characteristic</th>
<th>ES</th>
<th>AMTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0,9271</td>
<td>0,9423</td>
</tr>
<tr>
<td>SSE</td>
<td>94,5172</td>
<td>88,1214</td>
</tr>
<tr>
<td>DW</td>
<td>1,8632</td>
<td>1,9364</td>
</tr>
</tbody>
</table>

The values of the following forecasting characteristics for the constructed models are calculated by the formulas, which are shown in [24]: root mean squared error RMSE, mean absolute percentage error MAPE and Theil index $U$. Forecasting characteristics, which are calculated on the interval $N = 42, \ldots, 44$, are shown in the Table 5. Models are indicated as follows in the Table 5: exponential approach, which is described by the formulas (8)–(10) – EA; adaptive moving two-sided exponential smoothing – AMTES.

### Table 5
Forecasting characteristics of the models

<table>
<thead>
<tr>
<th>Forecasting characteristic</th>
<th>EA</th>
<th>AMTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>8,0462</td>
<td>6,1278</td>
</tr>
<tr>
<td>MAPE</td>
<td>8,2261</td>
<td>6,5432</td>
</tr>
<tr>
<td>$U$</td>
<td>0,0047</td>
<td>0,0045</td>
</tr>
</tbody>
</table>

Thus, experiments of restoring of missing values of Wolf numbers and index F10.7, restoring of the true regularities and forecasting of index F10.7 are described. Also tables with values of proximity criterion, statistical and forecasting characteristics of the constructed models are shown in this section.

### 5. Discussion
Restoration of true regularities, restoration of missing values of time series and forecasting of time series are performed in the article. Comparison with known methods, which are used for solar data, is performed on each stage.

In the work [25] it is examined the influence of the window width on the change of the accuracy characteristics of the moving average and is shown that moving average could significantly distort the process, reduce short-term fluctuations, while fluctuations with period, which significantly exceeds the window width, are restored almost without distortions. It is shown in the Fig. 1 that on the time intervals $[5, 10]$ and $[15, 20]$ there is a distortion of the process for a model constructed by the method of 13-month moving average: there is rotation of the fluctuations of the Wolf numbers. It is shown in the Fig. 2 that on the time interval $[10, 25]$ for the model, which is constructed using 13-month moving average method, there is significant distortion of the process: fluctuations are zeroed on the time intervals $[10, 15]$ and $[20, 25]$, there is rotation of the fluctuations of the index F10.7 on the time interval $[15, 20]$. At the same time, adaptive moving two-sided
exponential smoothing does not distort the fluctuations during the restoration of the true regularities of the index F10.7.

Applying the proximity criterion to the values, which are shown in the Table 1, 2, the conclusion is done that adaptive moving two-sided exponential smoothing method is shown superiority in comparison with 13-month moving average method when restoring of true regularities of Wolf numbers and index F10.7 are performed.

By the values of the statistical characteristics, which are shown in the Table 3, 4, it can be concluded that adaptive moving two-sided exponential smoothing method is shown superiority in comparison with exponential smoothing method when restoring of missing values of index F10.7 is performed. Therefore, it can be concluded that the proposed method is effective for restoring of missing values of time series for different initial conditions.

According to the values of the forecasting characteristics in the Table 5, it can be concluded that adaptive moving two-sided exponential smoothing method is shown superiority in comparison with the traditional exponential approach for forecasting.

6. Conclusions

Adaptive moving two-sided exponential smoothing method, in which variable moving interval is used, is developed in the article. Using the variable moving interval allows you to adapt to the features of each separate time series. An integral criterion is proposed for estimation of adequacy of constructed models when restoring of true regularities of time series is performed. Also the proximity criterion is proposed to evaluate the efficiency of restoring the true regularities of time series. Adaptive moving two-sided exponential smoothing method is used for restoring of the true regularities of the values of Wolf numbers and index F10.7, comparison with 13-month moving average method is performed. It is shown that adaptive moving two-sided exponential smoothing method does not distort the fluctuations during the restoration of true regularities of the time series. It can be concluded by applying the proximity criterion to the obtained values, that adaptive moving two-sided exponential smoothing method is shown superiority in comparison with 13-month moving average method for restoring of the true regularities of the Wolf numbers and the index F10.7.

Two algorithms for restoring of missing values of time series for different initial conditions are developed using adaptive moving two-sided exponential smoothing method. These algorithms are used for restoring of values of the index F10.7, comparison with exponential smoothing method is performed. The obtained results are shown superiority of the developed method in comparison with exponential smoothing method.

Adaptive moving two-sided exponential smoothing method is used for forecasting of values of index F10.7 and comparison with traditional exponential approach is performed. It is concluded on the basis of the values of forecasting characteristics that adaptive moving two-sided exponential smoothing method is shown superiority in comparison with traditional exponential approach for forecasting.

Thus, a new method for analysis and forecasting of solar data has been developed in the article. Perspectives of further research are creation of new advanced methods for analysis and forecasting of solar data, as well as creation of modern effective decision making support systems using the developed methods to provide faster and more accurate processing of time series.

References


