Scientific
Research
Publishing

Criteria for Weighted Moving-Mean Method

Shuo Jiang, Jinliang Wang

Research Institute for ESMD Method and Its Applications, College of Science, Qingdao University of Technology, Qingdao, China
Email: 652478221@qq.com, wangjinliang0811@126.com

Abstract

The moving-mean method is one of the conventional approaches for trend-extraction from a data set. It is usually applied in an empirical way. The smoothing degree of the trend depends on the selections of window length and weighted coefficients, which are associated with the change pattern of the data. Are there any uniform criteria for determining them? The present article is a reaction to this fundamental problem. By investigating many kinds of data, the results show that: 1) Within a certain range, the more points which participate in moving-mean, the better the trend function. However, in case the window length is too long, the trend function may tend to the ordinary global mean. 2) For a given window length, what matters is the choice of weighted coefficients. As the five-point case concerned, the local-midpoint, local-mean and global-mean criteria hold. Among these three criteria, the local-mean one has the strongest adaptability, which is suggested for your usage.

Keywords

Weighted Moving-Mean, Least Square Method, Extreme-Point Symmetric Mode Decomposition Method, Auto Regressive Moving-Mean, Data Analysis Methods

1. Introduction

The moving-mean method is commonly used for removing data noise or for finding trend curve of a given data [1] [2]. The documented improvements on this method lied in weighting the smoothing coefficients in various ways, such as the exponential moving-mean [3], auto regressive moving-mean [4] and mean filtering [5]. When the value of the time series is affected by cyclical and random fluctuations, particularly for the case with large fluctuations, it is difficult to grasp the development trend of the event. The moving-mean method can meet this request. It eliminates the influence of these factors and draws out the trend

line of this event. In turn, with the help of this line one can analyze the future variation of the concerned time series.

The moving-mean method has certain subjectivity and arbitrariness. Because its application effect largely depends on the selection of algorithm parameters. In the previous study on the moving-mean, the parameters are usually selected according to the mechanism of the dynamic testing process itself and the actual change state of the actual test data. There is no uniform standard for this experience, and there is no definite standard for the selection of parameters. If the study can find a uniform or regular criterion to select the window length and weight coefficient for different data, this will bring great convenience and feasibility to the application of the moving-mean.

The present study, regarding the window length, for m points moving-mean, uses the smoothing coefficient of $1/m$ for exploration. Note that the given data is $X = \{x_k\}_{k=1}^{n}$, then the moving-mean formula for the k-th point is:

$$r_k = \frac{1}{m} \sum_{i=k-(m-1)/2}^{k+(m-1)/2} x_i ,\tag{1.1}$$

where $k \in ((m+1)/2, n-(m-1)/2)$, n is the length of data points.

Regarding the weight coefficient, this study takes a five-point moving-mean as an example. For the convenience of calculation, the moving coefficient is set to be symmetrical. The data after the moving-mean is:

$$r_k = cx_{k-2} + bx_{k-1} + ax_k + bx_{k+1} + cx_{k+2} ,\tag{1.2}$$

where $k \in (3, n-2)$, $a + 2b + 2c = 1$, thereby $c = (1-a-2b)/2$. The variance of the smoothed data and the original data is recorded as

$$\varepsilon^2 = \sum_{k=1}^{n} (r_k - x_k)^2 .\tag{1.3}$$

For the data boundary processing, the boundaries are supplemented as follows. The supplementary data before r_1 is $e = (r_1 + r_2)/2$, and the supplementary data before e is $f = (r_1 + e)/2$. Then there is $r_2 = ce + bx_1 + ax_2 + bx_3 + cx_4$ and $r_1 = cf + be + ax_1 + bx_2 + cx_3$. The post-boundary processing method is the same as above, assuming $g = (r_1 + r_2)/2$ and $h = (r_1 + f)/2$. Then there is $r_{n-1} = cx_{n-3} + bx_{n-2} + ax_{n-1} + bx_n + cg$ and $r_n = cx_{n-2} + bx_{n-1} + ax_n + bf + ch$. When the number of smoothing points is large, in order to avoid the influence of the boundary data on the smoothing result, measures for cutting off the boundary point data are taken to reduce the error when smoothing.

2. Research on Window Length of Moving-Mean Method

By investigating many kinds of data, the results show that: Within a certain range, the more points which participate in moving-mean, the better the trend function. However, in case the window length is too long, the trend function may tend to the ordinary global mean, the range of the trend line obtained after the moving-mean will become smaller, and it will not reflect the fluctuation of the data (see Figure 1(a)). This situation is particularly evident in the periodic
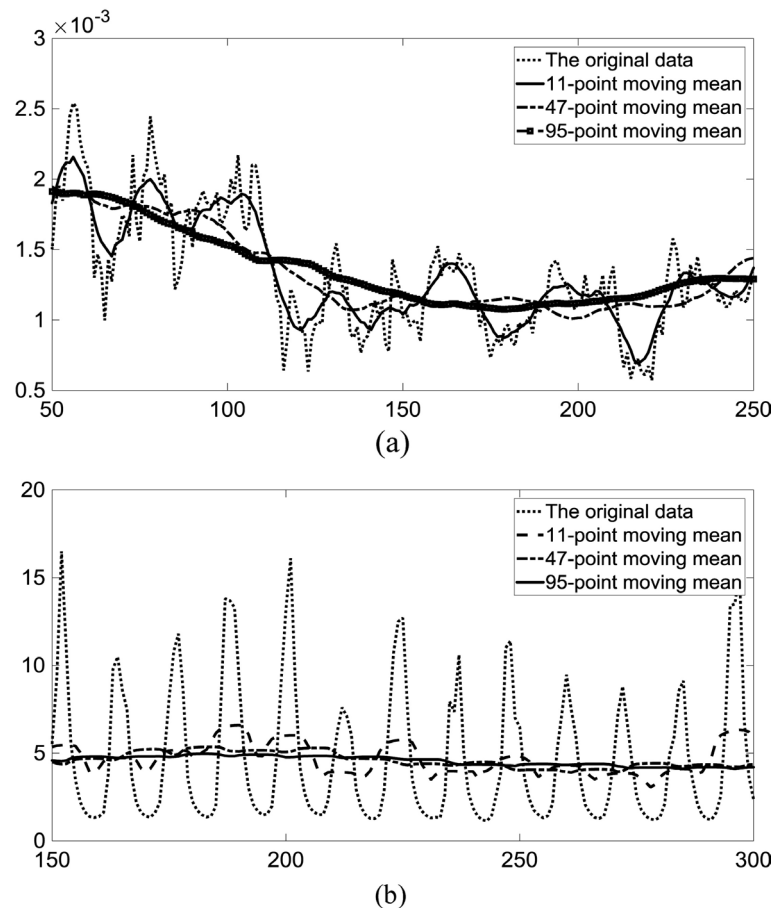
Figure 1. (a) Multiple-point moving-mean of wind speed data at 20 Hz sampling frequency; (b) Multiple-point moving-mean of annual and seasonal surface runoff data.

function (see **Figure 1(b)**). It can also be seen from the figure that when the number of smoothing points is too large, the local fluctuation direction of the trend line does not match the original data, and is greatly affected by the surrounding data.

For the periodic change data, after multiple-point smoothing, it is found that the best number of moving points is the number of points close to the period but not exceeding the period. For example, the data with a period of 12 points has the best effect when the number of smooth points is 11 points (see **Figure 2**). In the figure, the surface runoff varies from year to year. The local fluctuations are large when 3 points and 7 points are smoothed, and the trend line fluctuations obtained when 11 points are smoothed are the smallest, and the overall trend of the data can be seen. When the number of smoothing points is greater than the period, the local fluctuation direction of the trend line may be opposite or offset from the local fluctuation direction of the given data. The fluctuation direction after the 13-point smoothing is opposite to the given data fluctuation direction.

3. Research on Weight Coefficient of Moving-Mean Method

In the sense of least squares, the variance between the trend line and the data is
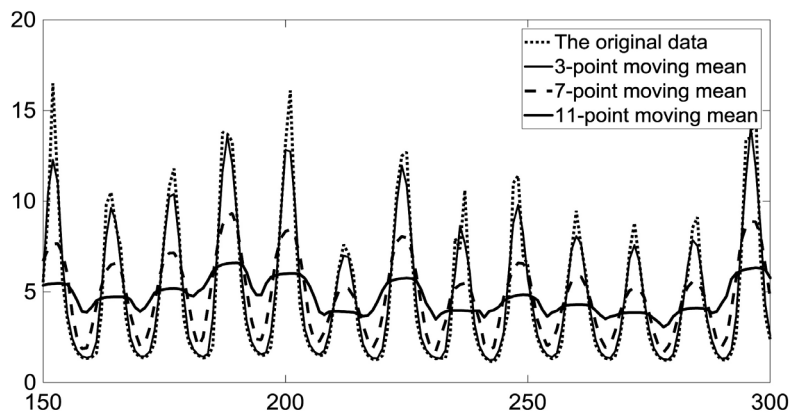
Figure 2. Comparison of multiple-point moving-mean of surface runoff data.

required to be minimal. In this study, multiple sets of data were used for the experiment. The final results show that the coefficients corresponding to the smoothed data and the given data have the smallest variance $a = 0$, $b = 0$, $c = 0.5$. It can be seen that the variance is the smallest when the smoothed data is coincident with the original data. Therefore, it is explained that the moving average cannot solve the variance of the smoothed data and the original data in the sense of least squares.

In this study, we propose the following three criteria for the selection of the moving-mean weight coefficient:

1) local-midpoint criterion.
2) local-mean criterion.
3) global-mean criterion.

3.1. Local-Midpoint Criterion

Between a maximum point and two adjacent minimum points, or between a minimum point and two adjacent maximum points, this is a period. That is to say, there is half a period between two adjacent extreme points. The pole symmetry refers to the midpoint of the line segment between adjacent maximum points and minimum points. All symmetrical midpoints are on the zero value line, where the midpoint is called the local midpoint.

By performing cubic spline interpolation on the obtained local midpoints, a median curve $m(k)$ can be obtained. Using this curve as a criterion, the variance ε^2 between the data after the multiple-point weighted moving-mean and the data corresponding to the interpolation line of the local midpoint is obtained by using formula (1.3). And solve the corresponding weight coefficient when the variance is minimum.

The second chapter says that the moving-mean corresponds to the window length should be close to but not more than the number of periodic points. This study focuses on the 5-point weighted moving-mean as in formula (1.2). It is a bit less when it is for 12 points of data for the season change cycle. The calculated results are $a = 0$, $b = 0$, $c = 0.5$. It is suitable for wind speed data with many

varying frequencies. Therefore, this study solves the corresponding smoothing coefficients for the 15 groups of wind speed data with the smallest variance. When the calculation accuracy is 0.1, 11 groups of data have a result of $a = 0.3$, $b = 0.2$, $c = 0.15$; 3 groups of data have a result of $a = 0.4$, $b = 0.2$, $c = 0.1$, and 1 group of data has a result of $a = 0.3$, $b = 0.3$, $c = 0.05$. The line after the moving-mean is smoother. **Figure 3** is the local midpoint interpolation line and the smooth graph of corresponding coefficients of one set of wind speed data. For the convenience of observation, the same set of data was used for the following study.
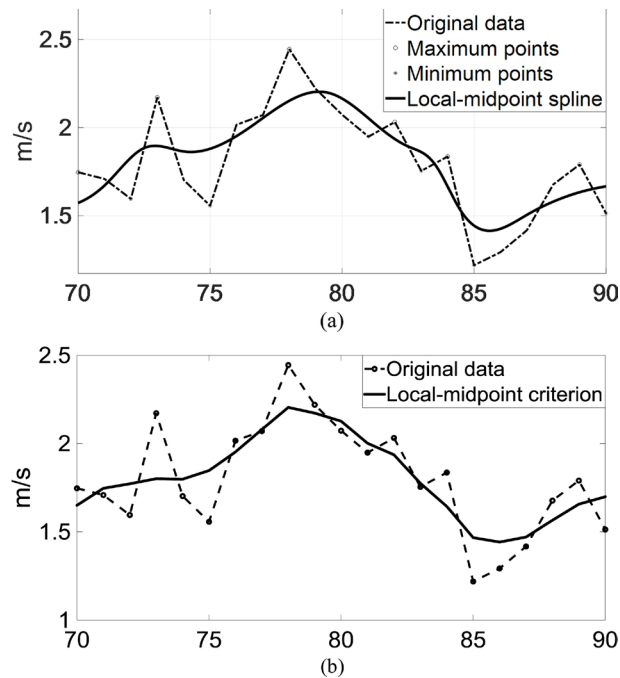

Figure 3. (a) The cubic spline at the local-midpoint; (b) The result of smoothing with the corresponding coefficient of minimum variance.

Referring to the two-line interpolation in the ESMD method [6] [7], the local midpoints here is divided into the odd-order midpoint interpolation line and the even-order midpoint interpolation line according to the coordinates, and the boundary processing is the same as the ESMD method. Then take the mean of the two interpolation lines. The smoothing coefficient with the smallest variance is solved by using this mean line as a standard. In the end, most of the data is obtained as $a = 0.2$, $b = 0.2$, $c = 0.2$, which is the same as the traditional 1/5 moving effect.

3.2. Local-Mean Criterion

Different from Smith's local mean method [8], the data in each half cycle of this study is all added and divided by the corresponding number, which is the local mean. Cubic spline interpolation is performed on the obtained local mean to obtain a mean curve $n(k)$. Using this curve as a criterion, the variance ε^2 between the data after multiple-point weighted moving mean and the data corres-

ponding to the interpolation of the local mean is obtained by using formula (1.3). And solve the corresponding weight coefficient when the variance is minimum.

When the calculation accuracy is 0.1, 7 groups of data have a result of $a = 0.3$, $b = 0.2$, $c = 0.15$; 8 groups of data have a result of $a = 0.4$, $b = 0.2$, $c = 0.1$. The line after the moving-mean is smoother. **Figure 4** is the local mean interpolation line and the smooth graph of corresponding coefficients of one set of wind speed data.

Most of the results are basically similar to the local-midpoint results, but there are 5 groups of data that are different from its results. When the calculation accuracy of the five groups of data is 0.01, it is found that the weight coefficients obtained by the local-mean and the local-midpoint are very similar (see **Table 1**), which indicates that the calculation accuracy is limited.
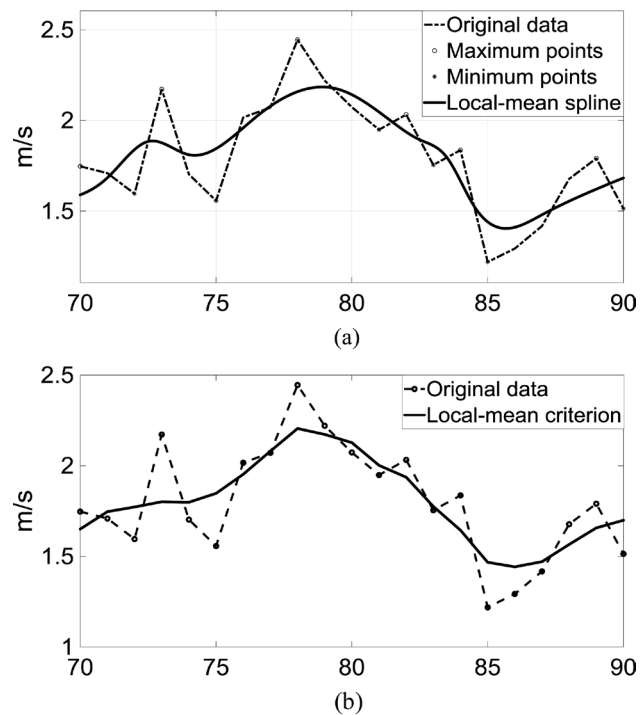

Figure 4. (a) The cubic spline of local-mean; (b) The result of smoothing with the corresponding coefficient of minimum variance.

Table 1. Weight coefficient corresponding to the minimum variance of each criterion.

	Local-mean (accuracy 0.1)	Local-mean (accuracy 0.01)	Local-midpoint (accuracy 0.1)	Local-midpoint (accuracy 0.01)
Data 1	$a = 0.4, b = 0.2, c = 0.1$	$a = 0.32, b = 0.24, c = 0.10$	$a = 0.3, b = 0.2, c = 0.15$	$a = 0.31, b = 0.23, c = 0.115$
Data 2	$a = 0.4, b = 0.2, c = 0.1$	$a = 0.32, b = 0.24, c = 0.10$	$a = 0.3, b = 0.2, c = 0.15$	$a = 0.29, b = 0.23, c = 0.125$
Data 3	$a = 0.4, b = 0.2, c = 0.1$	$a = 0.35, b = 0.25, c = 0.075$	$a = 0.3, b = 0.2, c = 0.15$	$a = 0.33, b = 0.25, c = 0.085$
Data 4	$a = 0.4, b = 0.2, c = 0.1$	$a = 0.33, b = 0.24, c = 0.095$	$a = 0.3, b = 0.2, c = 0.15$	$a = 0.30, b = 0.24, c = 0.110$
Data 5	$a = 0.4, b = 0.2, c = 0.1$	$a = 0.32, b = 0.24, c = 0.10$	$a = 0.3, b = 0.2, c = 0.15$	$a = 0.30, b = 0.23, c = 0.120$

3.3. Global-Mean Criterion

It has been explained above that in the case of least squares, the case where the

variance of the moving-mean and the original data is the smallest is consistent with the original data. If the global-mean is used as the standard, the calculated result will be very different from the original data. Therefore, considering the original data and the global-mean as the constraint of the coefficient solution, the global-mean criterion is proposed.

The global-mean criterion refers to first finding the variance between the smoothed data and the original data, and then finding the variance between the smoothed data and the global-mean of the original data, and finally summing the two variances. The weight coefficient corresponding to the minimum sum is the best coefficient.

When the calculation accuracy is 0.1, 9 groups of data have a result of $a = 0.4$, $b = 0.1$, $c = 0.2$; 2 groups of data have a result of $a = 0.4$, $b = 0.2$, $c = 0.1$, 2 groups of data have a result of $a = 0.3$, $b = 0.2$, $c = 0.15$; 1 groups of data has a result of $a = 0.5$, $b = 0.1$, $c = 0.15$, 1 groups of data has a result of $a = 0.5$, $b = 0$, $c = 0.25$. The line is not smooth after the moving-mean processing. Figure 5 is a smooth graph of corresponding coefficients of wind speed data.
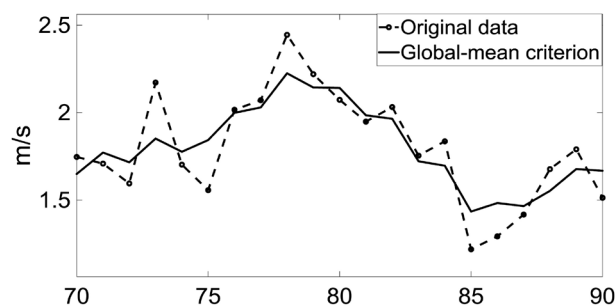

Figure 5. The result of smoothing with the corresponding coefficient of minimum variance.

4. Comprehensive Comparison of Criteria

The study [9] has shown that in the processing of asymmetric data, the local mean is more representative than the local midpoint. However, for general data, the two methods have little difference, which is why the weight coefficients obtained by the local-mean criterion and the local-midpoint criterion are relatively close. Therefore, in the comprehensive comparison here, the local-midpoint and the local-mean are classified into one category.

The local-mean criterion, the global-mean criterion, the traditional criterion are comprehensively compared. The comparison content includes the variance of the smoothed data and the original data, the weight coefficient corresponding to the minimum variance, and the characteristics of the data line after the smoothing process. The unit of variance is the square of the original data unit. The unit of standard deviation is the same as the unit of the original data. It is relatively intuitive to help the actual physical meaning. Moreover, the value of the variance is small, and it is difficult to observe the relationship between the data, so the difference between the data and the original data after 15 groups of data is smoothed is compared by the standard deviation (see Table 2).

It can be seen from Table 2 that the variance and corresponding weight coefficients of the local-midpoint and the local-mean criterion are similar, but the variance of the local-mean is generally smaller. This is because for a uniformly distributed data, the position of the local-midpoint is similar to the local-mean, but the local-midpoint is susceptible to extreme points, so the local-mean can better reflect the trend of the data than the local-midpoint.

Figure 6 is a graph showing the characteristics of the corresponding data lines after the smoothing of the coefficients obtained by the three criteria. As can be seen from the figure: 1) The smoothed data line under the traditional criterion is opposite to the original data direction, and there is no retention of data fluctuations and peaks. 2) Although the smooth line of the global-mean criterion retains the peak value, it is greatly affected by the weight c, so the data still fluctuates

Table 2. The minimum standard deviation of the smoothed data and the original data.

|	Local-mean criterion	Local-midpoint criterion	Global-mean criterion	Traditional criterion (1/5)
Data 1	0.0418	0.0618	0.7074	0.1248
Data 2	0.0400	0.0518	0.6919	0.1248
Data 3	0.0312	0.0425	0.6591	0.1176
Data 4	0.0122	0.0186	0.6693	0.0423
Data 5	0.4836	0.6135	0.6396	2.1382
Data 6	0.5109	0.6658	0.6322	2.1421
Data 7	0.4226	0.6821	0.6292	1.9654
Data 8	0.3337	0.4697	0.6224	1.2005
Data 9	0.3361	0.4125	0.6385	1.0656
Data 10	0.2127	0.2718	0.6052	0.9359
Data 11	0.2647	0.3939	0.6119	1.1828
Data 12	0.2430	0.3231	0.6047	1.0543
Data 13	0.2000	0.2619	0.5863	0.9118
Data 14	0.1090	0.1512	0.6703	0.3493
Data 15	0.1364	0.1744	0.6538	0.4031
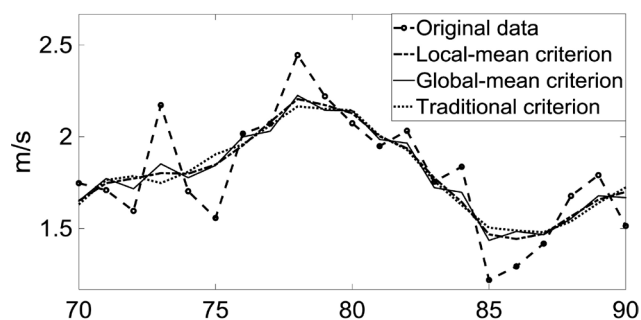
Figure 6. The result of smoothing with the corresponding coefficient of minimum variance of each criterion.

Table 3. Comprehensive comparison of criteria.

	Variance	Weight coefficient (accuracy 0.1)	Smoothed data line
Traditional criterion (1/5)	Relatively centered numerical value	$a = 0.2$, $b = 0.2$, $c = 0.2$	Non-smooth, can't keep peaks well, smoothing faster
Local-midpoint criterion	Relatively small numerical value	$a = 0.3$, $b = 0.2$, $c = 0.15$	Smoother, better noise reduction, smoothing speed in general
Local-mean criterion	Relatively minimum numerical value	$a = 0.3$, $b = 0.2$, $c = 0.15$	Smoother, better noise reduction, smoothing speed in general
Global-mean criterion	Relatively large numerical value	$a = 0.4$, $b = 0.1$, $c = 0.2$	Non-smooth, can retain peaks, and the degree of noise reduction is generally

greatly after smoothing. 3) The smooth line under the local-mean criterion is in the whole, and the smoothed data fluctuation is small, which achieves better noise removal effect and can better observe the overall trend of data.

As can be seen from the comprehensive comparison in **Table 3**, in general, the variance of the local-mean criterion smoothed data and the original data is the smallest, and the smoothed data of the local-mean criterion is smoother than the data obtained by other criteria. However, for the data with more data in the cycle, its Smoothing speed is slower than the other two criteria.

5. Conclusions

This paper draws some uniform conclusions through the study of the weighted moving-mean method. The extent to which the data is smoothed depends on the choice of the moving-mean window length and the weight coefficients. The selection of these two parameters needs to be selected according to the change pattern of the original data.

Regarding the smoothed window length, within a certain range, the more points which participate in moving-mean, the better the trend function. However, in case the window length is too long, the trend function may tend to the ordinary global mean, the range of the trend line obtained after the moving-mean will become smaller, and it will not reflect the fluctuation of the data. This situation is particularly evident in the periodic function. When there are too many smoothing points, the local fluctuation arc of the trend line does not match the original data, and is greatly affected by the surrounding data. For periodically varying data, the best number of moving points is the number of points that is close to the period and does not exceed the period.

Regarding the choice of smoothing weight coefficients, this study compares the three newly established criteria with traditional ones. 1) If the smoothed data is to achieve the noise removal effect and the change is smooth, the local-midpoint criterion and the local-mean criterion can be used to solve the weight coefficients. 2) If you want the data to be smoothed quickly without being affected by intermediate data, you can use the traditional criterion to solve the weight coeffi-

cients. 3) If you want the data to retain the peak of the detail and the smoothness is greater, you can use the global-mean criterion to solve the weight coefficients.

By observing the experimental data, it is found that the weight coefficient obtained by the five-point weighted moving-mean under the local-midpoint and local-mean criteria is similar to the coefficient of the binomial. Therefore, it is assumed that the n-point weighted moving-mean coefficient is similar to the (n − 1)-th binomial coefficient (that is, the n-th column coefficient of the Yanghui triangle). Whether there is a relationship between the two? It is also necessary to research a large number of data in a multiple-point smoothing experiment to make a specific judgment.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Leon-Castro, E., Aviles-Ochoa, E. and Merigo, J.M. (2018) Induced Heavy Moving Averages. *International Journal of Intelligent Systems*, **33**, 1823-1839. https://doi.org/10.1002/int.21916

[2] Huang, K.M. (2006) Research on the Parameters of Sliding Averaging for Digital Filtering. *Journal of Jimei University* (*Natural Science*), **4**, 381-384. (In Chinese)

[3] Abujiya, M.R., Lee, M.H. and Riaz, M. (2014) Improving the Performance of Exponentially Weighted Moving Average Control Charts. *Quality & Reliability Engineering International*, **4**, 571-590. https://doi.org/10.1002/qre.1509

[4] Sun, H.X. (2007) Stochastic Processes. China Machine Press, Beijing. (In Chinese)

[5] Zhu, S.H. and You, C.X. (2013) A Modified Average Filtering Algorithm. *Computer Applications and Software*, **12**, 97-99. (In Chinese)

[6] Wang, J.L. and Li, Z.J. (2013) Extreme-Point Symmetric Mode Decomposition Method for Data Analysis. *Advances in Adaptive Data Analysis*, **5**, Article ID: 1350015. https://doi.org/10.1142/S1793536913500155 http://arxiv.org/abs/1303.6540

[7] Wang, J.L. and Li, Z.J. (2015) Extreme-Point Symmetric Mode Decomposition Method: A New Approach for Data Analysis and Science Exploration. Higher Education Press, Beijing. (In Chinese)

[8] Smith, J.S. (2005) The Local Mean Decomposition and Its Application to EEG Perception Data. *Journal of the Royal Society Interface*, **2**, 443-454. https://doi.org/10.1098/rsif.2005.0058

[9] Jiang, S. and Wang, J.L. (2018) Improvement of ESMD Algorithm for Asymmetric Data. *Advances in Applied Mathematics*, **7**, 1500-1505. (In Chinese) https://doi.org/10.12677/AAM.2018.712174