Comparison of performance between rescaled range analysis and rescaled variance analysis in detecting abrupt dynamic change*

He Wen-Ping(何文平)[a)†], Liu Qun-Qun(刘群群)[b)], Jiang Yun-Di(姜允迪)[a)], and Lu Ying(卢 莹)[c)]

[a)] *National Climate Center, China Meteorological Administration, Beijing* 100081, *China*
[b)] *College of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing* 210044, *China*
[c)] *Yangzhou Meteorological Office, Yangzhou* 225003, *China*

(Received 12 September 2014; revised manuscript received 30 October 2014; published online 10 February 2015)

In the present paper, a comparison of the performance between moving cutting data-rescaled range analysis (MC-R/S) and moving cutting data-rescaled variance analysis (MC-V/S) is made. The results clearly indicate that the operating efficiency of the MC-R/S algorithm is higher than that of the MC-V/S algorithm. In our numerical test, the computer time consumed by MC-V/S is approximately 25 times that by MC-R/S for an identical window size in artificial data. Except for the difference in operating efficiency, there are no significant differences in performance between MC-R/S and MC-V/S for the abrupt dynamic change detection. MC-R/S and MC-V/S both display some degree of anti-noise ability. However, it is important to consider the influences of strong noise on the detection results of MC-R/S and MC-V/S in practical application processes.

Keywords: moving cutting data-rescaled range analysis, moving cutting data-rescaled variance analysis, abrupt dynamic change

PACS: 92.70.Aa, 95.10.Fh **DOI:** 10.1088/1674-1056/24/4/049205

1. Introduction

Hurst developed the rescaled range analysis (R/S),[1] which is a statistical method to analyze long records of natural phenomena. If we consider the same time series but increase the number of observations, the rescaled range will generally also increase. The rescaled range is calculated by dividing the range of values exhibited in a portion of the time series by the standard deviation of the values over the same portion of the time series. The increase in the rescaled range can be characterized by plotting the logarithm of R/S versus the logarithm of m (where m is the size of subsample data). The slope of this line gives the Hurst exponent, H. If the time series is generated by a random walk H has a value of 0.5, i.e., $H = 0.5$ (http://en.wikipedia.org/wiki/Rescaled range). The Hurst exponents of many types of time series are greater than 0.5.[2–16] For example, observations of the height of the Nile River measured annually over many years provided $H = 0.77$. The daily temperature records also have $H > 0.5$. It is now well known that a time series with a Hurst exponent greater than 0.5 exhibits a long-range correlation or long-term memory.

Chen *et al.*[17] and He *et al.*[18] found that the Hurst exponent of a correlated time series, which is generated by one dynamic system, does not change to a statistically significant degree when a segment is randomly cut from the correlated signals. Furthermore, the changes in the Hurst exponent when some data are removed are mainly caused by an insufficient sample size. However, if there is an abrupt change in the dynamic equation at a specific moment, the Hurst exponent of the correlated time series generated by the equation will change sharply. In view of these characteristics of the Hurst exponent, He *et al.* presented a novel measure, i.e., moving cutting data—R/S (MC-R/S),[18] for abrupt dynamic change detection in a correlated time series. Numerical tests demonstrate that MC-R/S performs well.

Based on the R/S algorithm, Giraitis *et al.* presented an amended method: rescaled variance analysis (V/S).[19] According to the V/S and MC-R/S algorithms, Sun *et al.* put forward a new method of detecting the abrupt dynamic change, which is called the moving cutting data—V/S (MC-V/S) method.[20] They claimed that MC-V/S performs better than MC-R/S in detecting an abrupt dynamic change in a correlated time series, but they did not compare the performances of the two methods by using either an artificial time series or the observational data. To facilitate the best application of the two methods, it is important to quantitatively compare the performances of MC-V/S and MC-R/S. To bridge this gap, the operational efficiency (namely CPU time consumed), accuracy of the detection result, and anti-noise ability of the two methods for observational data are comprehensively compared in this paper.

The rest of this paper is organized as follows. In Section 2, we briefly introduce our methods, including R/S, V/S, MC-R/S, and MC-V/S, and the model time series used in nu-

*Project supported by the National Basic Research Program of China (Grant No. 2012CB955902) and the National Natural Science Foundation of China (Grant Nos. 41275074, 41475073, and 41175084).

†Corresponding author. E-mail: wenping_he @163.com

© 2015 Chinese Physical Society and IOP Publishing Ltd

merical tests. In Section 3 our results are compared, particularly focusing on the influences of noise on the detection results of MC-V/S and MC-R/S. Finally, the conclusions and discussion are presented in Section 4.

2. Methods and model

2.1. Rescaled range analysis

Rescaled range analysis (R/S) is a statistical measure of the variability of a time series introduced by Hurst.[1] The purpose of R/S is to provide an assessment of how the apparent variability of a series changes with the length of the time-period being considered. For a time series $\{x_i,\ i = 1,2,\ldots, N\}$, R/S is calculated as follows:

(i) Consider an m-dimensional sample series $\{y_i,\ i = 1, 2,\ldots, m\}$, where $m = sN$, and $s \in (0,1)$;

(ii) Then, compute the mean of the subseries $\{y_i\}$,

$$\bar{y} = \frac{1}{m}\sum_{k=1}^{m} y_k;$$

(iii) Calculate the cumulative deviation $Z(k)$ of the series $\{y_i\}$,

$$Z(k) = \sum_{k=1}^{m} (y(k) - \bar{y}), \quad k == 1,2,\ldots,m;$$

(iv) Determine the range $R_m = \max\{Z(k)) - \min\{Z(k)\}$, and the rescaled range $(R/S)_m = R_m/S_m$, where the S_m is the standard deviation for the subseries $\{y_i,\ i = 1, 2,\ldots, m\}$, and is expressed as

$$S_m = \left[\frac{1}{m}\sum_{k=1}^{m} (y(k) - \bar{y})\right]^{1/2};$$

(v) Shift the subseries $\{y_i\}$ with a step size of m without changing the length of the subseries, namely, $\{y_i,\ i = 1 + m, 2 + m,\ \ldots,\ 2m\}$, and repeat steps (ii) to (iv);

(vi) Estimate the average rescaled range for each subseries;

(vii) Change the size m, and repeat the operation from steps (i) to (vi);

(viii) Create a double logarithm plot of the average of $(R/S)_s$ versus m.

These steps can be summarized in the following equation:

$$(R/S)_m = \frac{1}{S}\left[\max_{1\le i\le m}\sum_{k=1}^{i} (y(k) - \bar{y}) - \min_{1\le i\le m}\sum_{k=1}^{i} (y(k) - \bar{y})\right]. \quad (1)$$

Hurst found that the ratio R/S is very well described for a large number of natural phenomena by the following empirical relation:

$$(R/S)_m = am^H, \quad (2)$$

where a is a constant and H is the Hurst exponent. If the time series $\{x_i,\ i = 1,2,\ldots,N\}$ is generated by a random walk, H has a value of 0.5, i.e., $H = 0.5$. If the Hurst exponent is less than 0.5 and greater than 0, the time series is uncorrelated. If the Hurst exponent is greater than 0.5, the time series is characterized by long-range correlation. If the Hurst exponent is 1.0, the time series exhibits the behavior of $1/f$ noise.

2.2. Rescaled variance analysis

In the V/S algorithm, the rescaled variance statistic is described by the following equation[19]

$$(\mathrm{V/S})_m = \frac{1}{mS_s^2}\left[\sum_{L=1}^{m}\Big(\sum_{k=1}^{L}(y(k) - \bar{y})\Big)^2 - \frac{1}{m}\Big(\sum_{L=1}^{m}\sum_{k=1}^{L}(y(k) - \bar{y})\Big)^2\right], \quad (3)$$

where V/S refers to variance, S_m.

2.3. MC-R/S and MC-V/S algorithms

The detailed descriptions of the MC-R/S and MC-V/S algorithms are as follows.

Step 1 Choose a window size M;

Step 2 Continuously cut sections of data with a length of M from the i-th data to the $i + M$-1th, $i = 1,\ 1 + M,\ldots,$ $1 + (n-1)M$, $n = [N/M]$, where the symbol [ ] denotes the fetching integer and N is the total number of data. For example, [1000/30] = 33. Then, stitch the remaining parts together to obtain a new time series.

Step 3 Calculate the values of Hurst exponent H_i of the new time series (including N-M data) using R/S and V/S, respectively;

Step 4 Slide the window with a fixed size M in the original series, and repeat Steps 2 and 3 until reaching the end of the original series;

Step 5 Obtain a Hurst exponent series $\{H_i,\ i = 1,2,\ldots,n\}$;

Step 6 Calculate the variance contributions of the Hurst exponent series in Step 5, and obtain the time-instant of any abrupt dynamic changes.

2.4. Logistic map

The logistic map is a polynomial map and is often cited as an archetypal example of how complex, chaotic behavior can arise from very simple nonlinear dynamic equations. Mathematically, the logistic map is written as follows:[21]

$$x_{n+1} = ux_n(1 - x_n), \quad x \in [0,1]. \quad (4)$$

Here, x_n is a number between zero and one and it represents the ratio of existing population to the maximum possible population in year n. Thus, x_0 represents the initial ratio of the population in year 0. The u is a positive number and represents a combined rate of reproduction and starvation. The relative

simplicity of the logistic map makes it an excellent point of entry to the concept of chaos. A rough description of chaos is that a chaotic system exhibits tremendous sensitivity to the initial conditions, with most values of u falling in a range between approximately 3.57 and 4 on the logistic map. In this study, $x_0 = 0.8$, and $u = 3.8$.

3. Results

To compare the performances of the MC-R/S and MC-V/S, the artificial time series with two abrupt dynamic changes in Ref. [18] is adopted in this study. In the artificial series, an abrupt dynamic change case can be designed as follows. The evolution of a species can be described with a logistic map. A sudden natural disaster occurs in a certain time period and results in a change in the dynamic equation dominating the evolution of the species; specifically, the logistic map could be replaced with stochastic behavior from $n = 301$ to $n = 330$ (Fig. 1(c)). Two abrupt dynamic changes clearly occur at $n = 301$ and $n = 330$ in the artificial series.

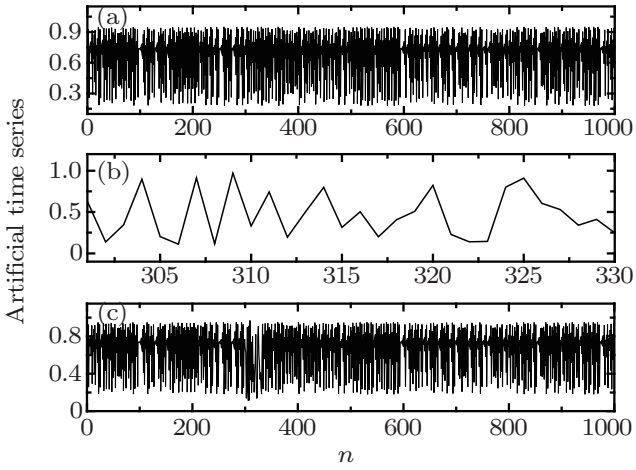


Fig. 1. Artificial time series with an abrupt dynamic change from $n = 301$ to $n = 330$. (a) A time series with length of 1000 generated by logistic map; (b) Gauss white noises with a length of 30; (c) the artificial time series with two abrupt dynamic changes, which can be obtained by replacing the data from $n = 301$ to $n = 330$ in the time series shown in Fig. 1(a) with the data shown in Fig. 1(b).[18]

3.1. Comparing the performances of MC-R/S and MC-V/S without noise

Figure 2 presents the MC-R/S and MC-V/S results for the artificial time series shown in Fig. 1(c). The abrupt change occurring from $n = 301$ to $n = 330$ can be identified equally using either MC-R/S or MC-V/S (Figs. 2(a) and 2(b)). The results shown in Figs. 2(a) and 2(b) indicate an abrupt decrease in the Hurst exponent series but fail to quantitatively indicate the specific time at which the abrupt change occurs. Thus, identification of the change point is left to the opinion of the analyst, based on a visual inspection of the Hurst exponent statistical graph.

In view of the sensitivity of the Hurst exponent to the data from different dynamic systems, He *et al.* presented a quantitative estimation method to identify the time-instants of abrupt changes based on the variance contribution of Hurst exponents.[18] It should be noted that the average variance contribution procedure was calculated using all Hurst exponents, and the threshold of the variance contribution was set to be triple the average. The variance contribution can distinguish between normal and abnormal fluctuation amplitudes. Normal fluctuations are primarily caused by the small sample size, and abnormal fluctuations are primarily caused by the sensitivity of the calculation method for the Hurst exponent to the data from different dynamic systems.

Based on the definition of the threshold of variance contribution, it is easy to find that the variance contributions in the period from $n = 301$ to $n = 330$ are clearly greater than those in other parts in the artificial series for different window sizes, such as $M = 2$, 5, 10, and 15. It can therefore be concluded that an abrupt change in Hurst exponent occurs from $n = 301$ to $n = 330$ in the artificial series and that the detection results are robust for different window sizes. However, it must be noted that false detections for abrupt change points are more or less likely to be obtained for both MC-R/S or MC-V/S when the window size M is relatively small, such as $M = 2$ or 5 (Figs. 2(b) and 2(c)). It can be found that the positions of these false detections depend on the window size (Fig. 2). In particular, the positions of the false detections vary with window size, but the true positions of the abrupt change are hardly affected. The false detections can be clearly identified with larger window sizes.

Under identical computational conditions (Dell precision T3400), the computer times for running the MC-R/S and MC-V/S detection programs are presented in Table 1. In general, the computer time consumed by MC-V/S is approximately 25 times that of MC-R/S for an identical window size. Therefore, the operating efficiency of the MC-R/S algorithm is clearly much higher than that of the MC-V/S algorithm.

3.2. Comparing the performances of MC-R/S and MC-V/S under different signal noise ratios

Noise is inevitable in observational data. In this subsection, the effects of noise on the performances of MC-R/S and MC-V/S are investigated. First, the performances of MC-R/S and MC-V/S are tested for detecting abrupt change in the artificial time series with a signal-to-noise ratio (SNR) of 30 dB. Similar to what has been done in Fig. 2, the abrupt change occurring from $n = 301$ to $n = 330$ can be identified equally well using either MC-R/S or MC-V/S (Figs. 2(a) and 2(b)) for different window sizes. When the window size is relatively small, e.g., $M = 2$ or $M = 5$, false detection results exist for both MC-R/S and MC-V/S. The false detection results disappear for MC-R/S after increasing the window size (Figs. 3(e)

and 3(f)). Most of the false detection results also disappear for MC-V/S after increasing the window size, but very few false detection results remain, such as $M = 10$ and $M = 15$ (Figs. 3(e) and 3(f)). When the window size M is 30, both MC-R/S and MC-V/S can exactly detect the abrupt Hurst exponent change (the relevant figures are omitted here). When the values of SNR are 25 dB and 20 dB, showing that both noises are stronger than SNR = 30, similar results can be obtained and are therefore not discussed here in detail (see Appendices A and B).

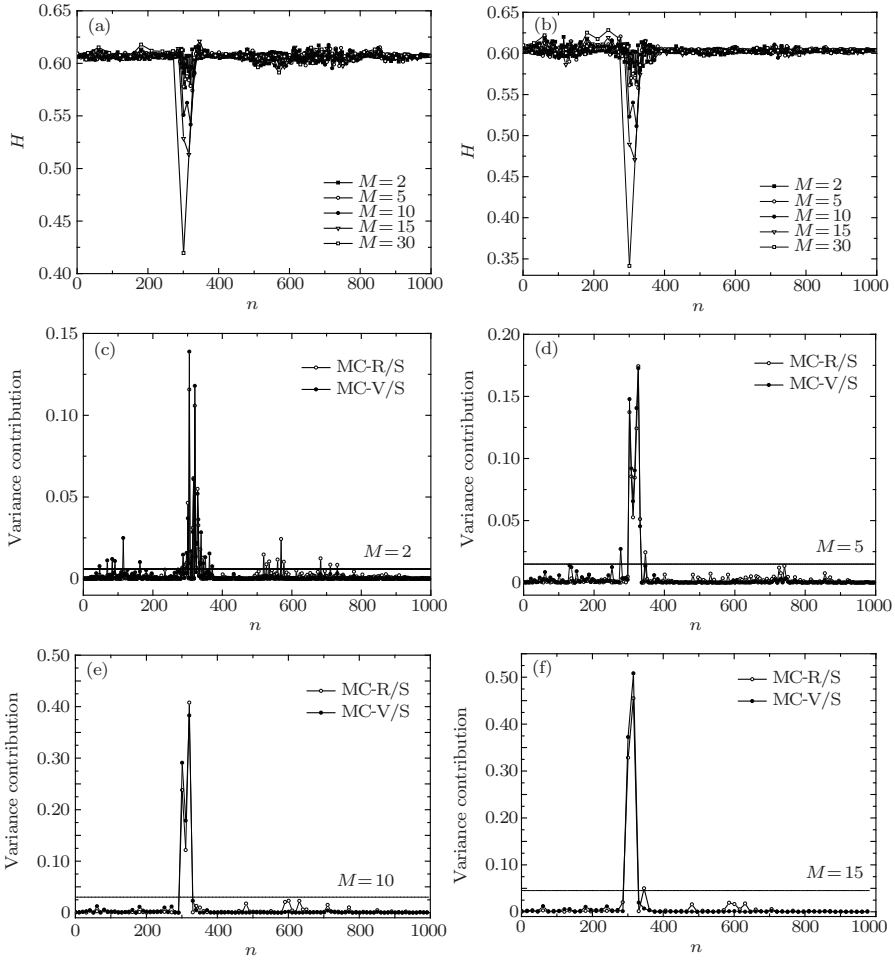


Fig. 2. MC-R/S and MC-V/S results for artificial time series and variance contributions of Hurst exponents. The results of (a) MC-R/S and (b) MC-V/S under different window scales, including $M = 2$, 5, 10, 15, and 30; variance contributions for detection results using MC-R/S (denoted by empty circles) and MC-V/S (denoted by solid circles), with window sizes $M = 2$ (c), 5 (d), 10 (e), and 15 (f). The dashed line represents the tripled average variance contribution of the Hurst exponent series.

Table 1. Computer running times of MC-R/S and MC-V/S for different window sizes (in units of s).

Window size	$M = 2$	$M = 5$	$M = 10$	$M = 15$	$M = 30$
MC-R/S	1.6703×10^{-3}	6.656×10^{-4}	3.313×10^{-4}	2.156×10^{-4}	1.047×10^{-4}
MC-V/S	4.13922×10^{-2}	1.611234×10^{-2}	8.0485×10^{-3}	5.275×10^{-3}	2.5172×10^{-3}

When SNR is 15 dB, it can be observed from Figs. 4(a) and 4(b) that the Hurst exponents abruptly decrease when the data from $n = 301$ to $n = 330$ are removed. This abrupt decrease is mainly due to the different influences of the data from different dynamics on the calculation of Hurst exponents. The largest variance contribution is still roughly located between $n = 301$ and $n = 330$ for window sizes $M = 2$, 5, 10, 15, and 30 (Fig. 4). In addition, an abrupt increase in the Hurst exponent can be observed when the window size is relatively large, such as $M = 10$, 15, and $M = 30$. Comparing these results with the detection results shown in Figs. 2(a) and 2(b), it is easy to conclude that the noise in the artificial series with SNR of 15 dB causes the abrupt increases in Hurst exponent in some time periods. Variation analyses of the MC-R/S and MC-V/S results indicate that some false detections still exist when the window size is increased, unlike in the samples without noise. Particularly for some large window sizes, the variance contribution of false detection could approximate that of the actual abrupt change period as observed in Fig. 4(g). These results demonstrate that the strong noise can result in false detections for both MC-R/S and MC-V/S.

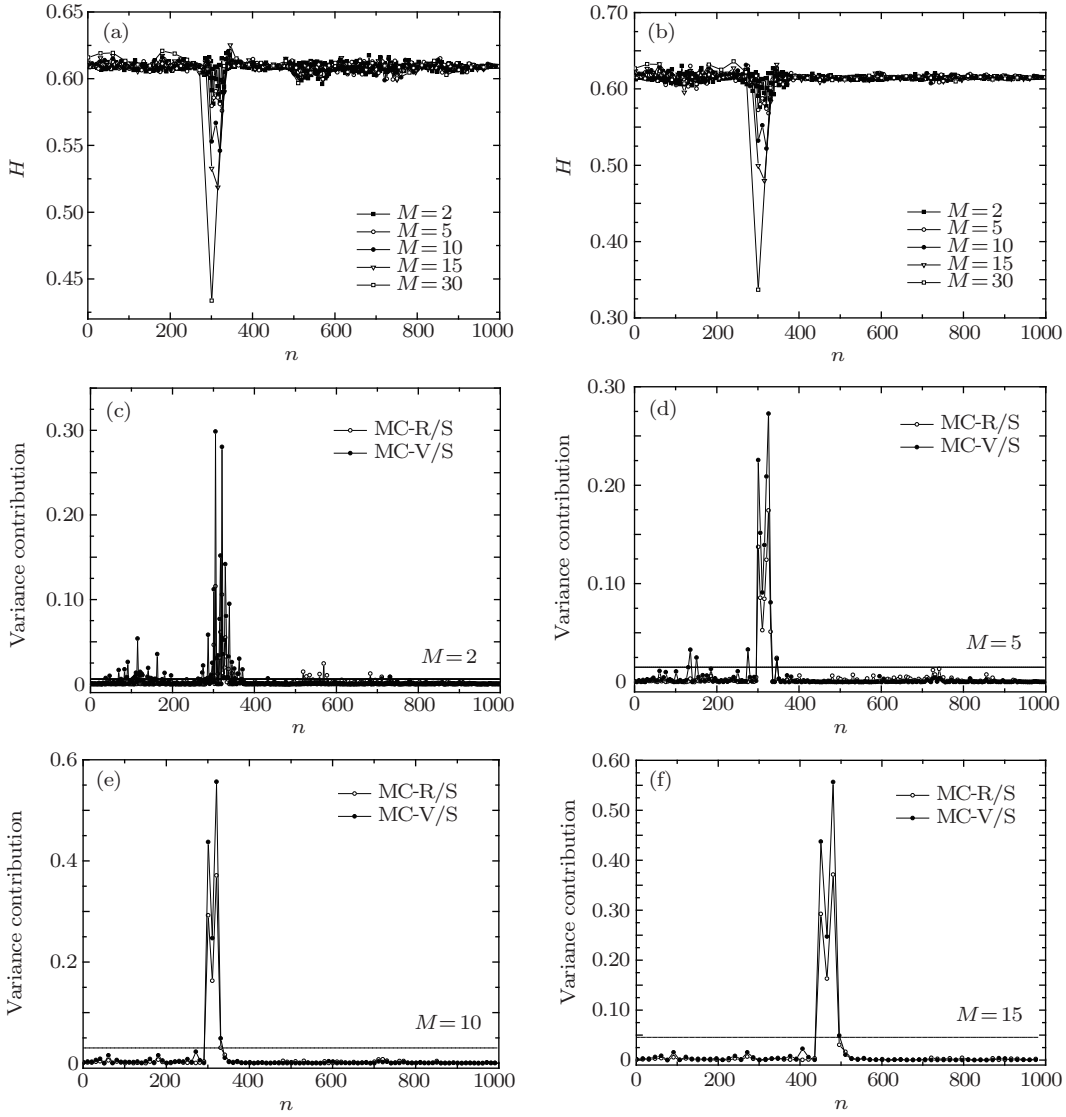
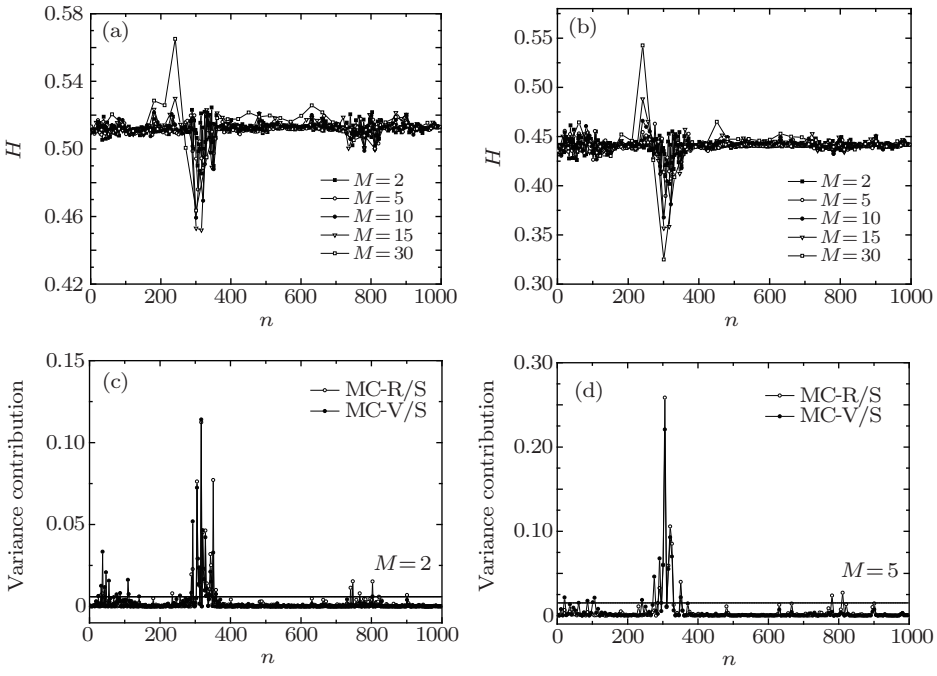


Fig. 3. The same as Fig. 2, but for the artificial time series with SNR = 30 dB.



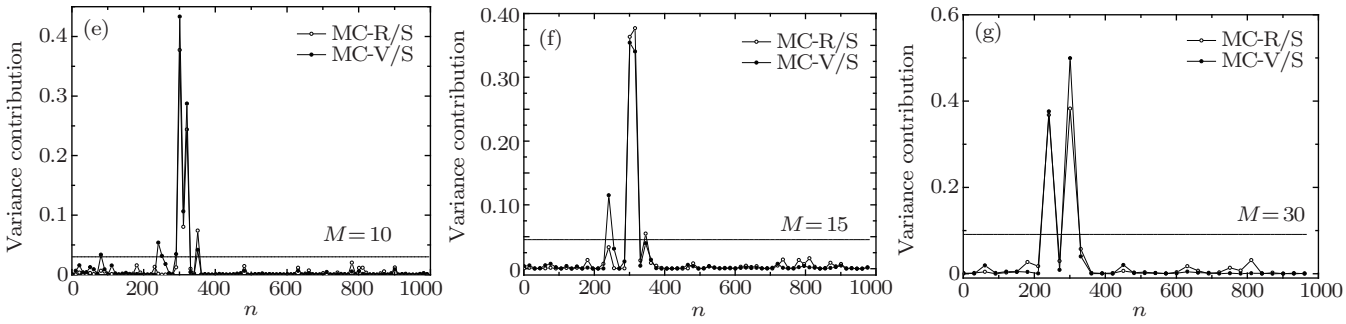


Fig. 4. Panels (a)–(f) being the same as Fig. 2, and panel (g) the same as Fig. 2(c) but for $M = 30$ and the artificial time series with SNR = 15 dB.

3.3. Comparing the performance of MC-R/S and MC-V/S for daily surface air pressure records

In addition to the above comparisons where an artificial time series is used, it is important to compare the performances of the MC-R/S and MC-V/S methods in detecting abrupt change in observational data. Here, the daily surface air pressure records are selected to further test the performances of the two methods. The records are from the Huma meteorological station in the Heilongjiang province, China. The detection results are shown in Fig. 5. It can be observed that the evolutionary trends of the Hurst exponents shown in Fig. 5(a) are similar to those in Fig. 5(b). They all reach a maximum value in 1964, which has significantly higher Hurst exponents compared with in the other years. To detect the exact moment of abrupt dynamic change in the air pressure records, the variance contributions of the Hurst exponents are analyzed. The evolutionary curves of the variance contributions in Figs. 5(c) and 5(d) are similar. The maximum values of the variance contributions are 41.2% and 41.4%, respectively, which are both far greater than the variance threshold (approximately 6.52%). Moreover, the variance contributions in 1992 and 1993 are slightly greater than the thresholds in both the MC-R/S and MC-V/S results. To ensure the accuracy of the detection results, we use the MC-detrended fluctuation analysis (MC-DFA)[22,23] to detect the abrupt change in the records and find that cutting data from any year other than 1964 has very little effect on the Hurst exponents (the relevant figures are omitted here). Observational daily surface air pressure records from nearby stations have also been tested using MC-R/S, MC-V/S and MC-DFA, and the detection results similarly indicate that there is an abrupt dynamic change in 1964 in Huma.

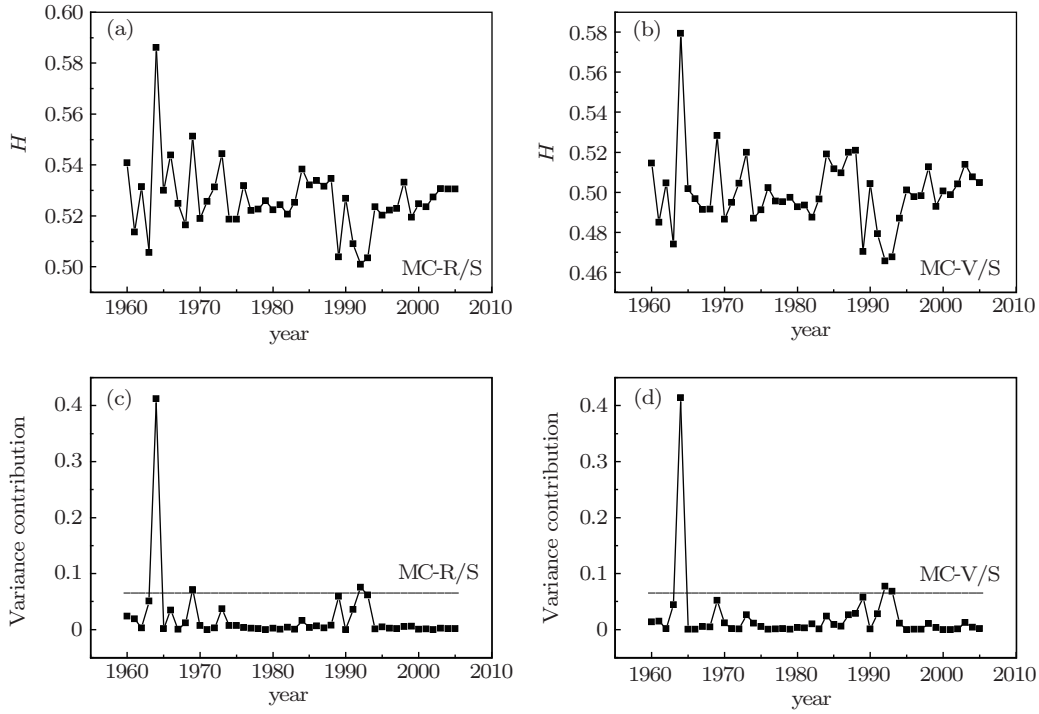


Fig. 5. The MC-R/S and MC-V/S results of Hurst exponents for daily surface air pressure records for the period from January 1, 1960 to December 31, 2006 at the Huma meteorological station in China, and the variance contribution analysis of the Hurst exponents with a one-year subseries length. Panel (a) shows the MC-R/S results, panel (b) the MC-V/S results, panel (c) the variance contribution analysis of the MC-R/S results, and panel (d) is the same as panel (c) but for the MC-V/S results. The dashed line represents the tripled average variance contribution of the Hurst exponent series.

4. Conclusions and discussion

A comparison of performance between MC-R/S and MC-V/S demonstrates that the operating efficiency of MC-R/S is clearly higher than that of MC-V/S. In general, the computer time consumed by MC-V/S is approximately 25 times greater than by MC-R/S. For the case without noise, false detection results occur for both MC-R/S and MC-V/S when the window size is relatively short. However, these false detection results disappear with an increase in the window size for both MC-R/S and MC-V/S. The influence of weak noise on the detection result is relatively small for each of the methods, but the influence of strong noise cannot be ignored, particularly for larger window sizes. To mitigate the influence of strong noise on the detection result, certain filter technologies can be applied prior to using MC-R/S and MC-V/S, such as the Vondrak Filter.[24]

The detection results of MC-R/S and MC-V/S for daily surface air pressure records show similar evolutionary trends of the Hurst exponents, and both indicate an abrupt dynamic change in 1964 at the Huma station. The time of the abrupt change is identical to in a previous study.[22,23] Notably, however, the computer times consumed by MC-R/S and MC-V/S are 8.172 min and 1267.164 min, respectively. Thus, the computer time consumed by MC-V/S is approximately 155 times greater than by MC-R/S.

In summary, MC-R/S and MC-V/S perform similarly for abrupt dynamic change detection, except for the difference in operating efficiency. Moreover, we find that the detection results of both MC-R/S and MC-V/S could depend on the definition of the variation threshold of the Hurst exponent to some extent. Particularly for the case in which the variation contribution of the Hurst exponent obtained using MC-R/S or MC-V/S is slightly greater than the threshold, it is very difficult to justify the authenticity of the abrupt change point. Therefore, further investigation is necessary to determine how to define a rational variation threshold in future studies.

Appendix A

MC-R/S and MC-V/S results for artificial time series and variance contributions of Hurst exponents with SNR = 25 dB.

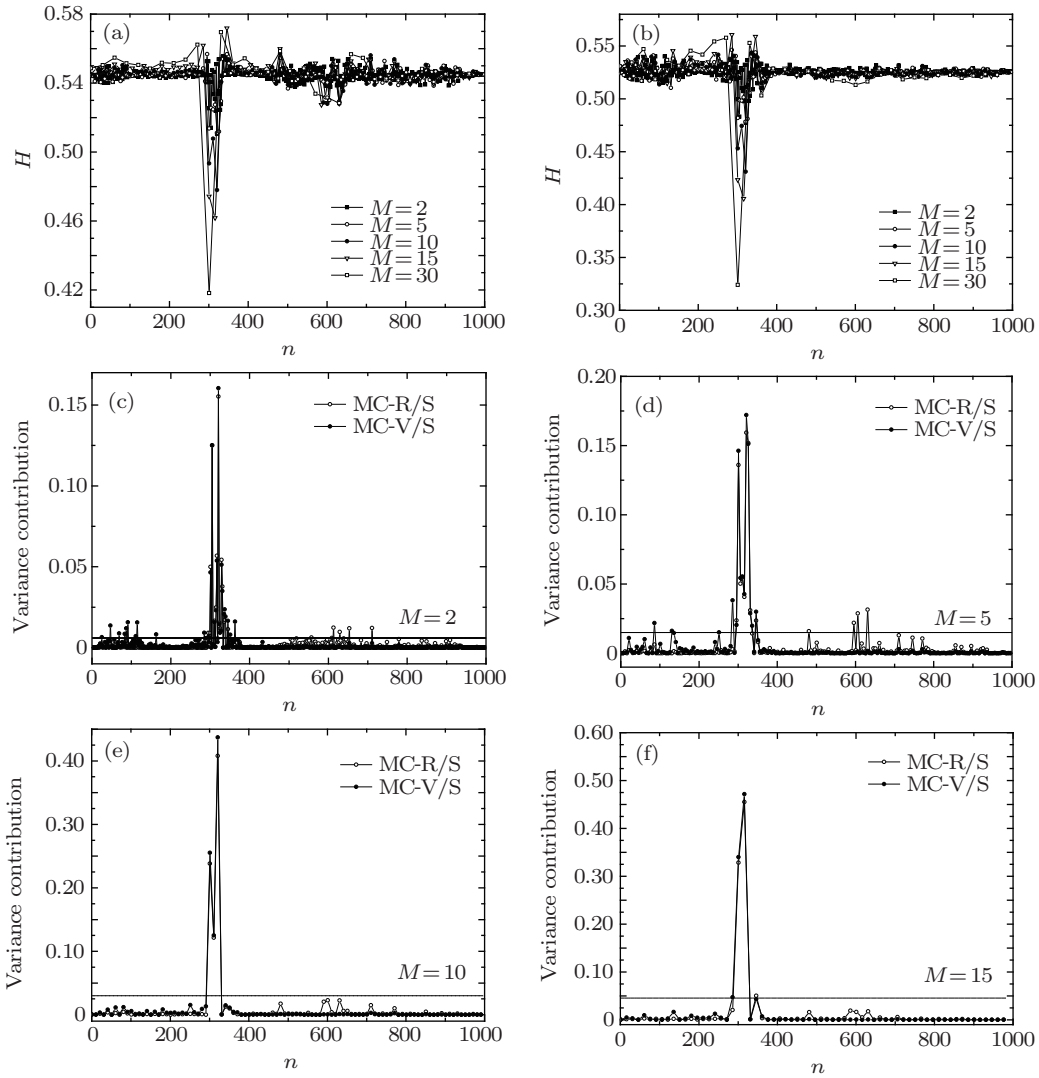


Fig. A1. The same as Fig. 2, but for the artificial time series with SNR = 25 dB.

Appendix B

MC-R/S and MC-V/S results for artificial time series and variance contributions of Hurst exponents for $M = 30$ and with SNR = 20 dB.

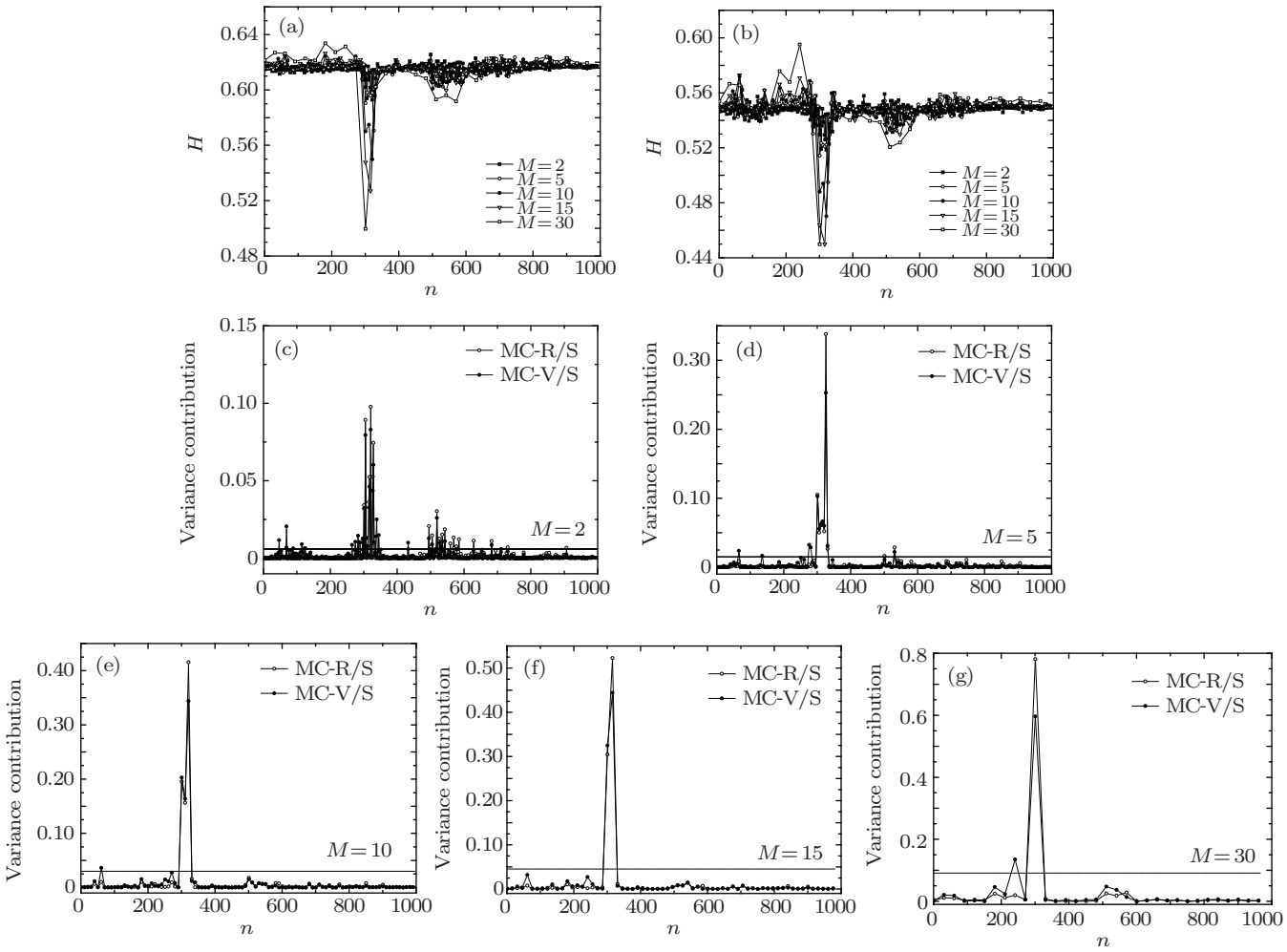


Fig. B1. Panels (a)–(f) being the same as Fig. 2, panel (g) being the same as Fig. 2(c) but for $M = 30$ and the artificial time series with SNR = 20 dB.

References

[1] Hurst H E 1951 *Trans. Am. Soc. Civ. Eng.* **116** 770

[2] Liu S D, Chen J and Liu S S 1999 *J. Appl. Meteor. Sci.* (Suppl.) **10** 9 (in Chinese)

[3] Liu S D, Rong P P and Chen J 2000 *Acta Meteor. Sin.* **58** 110 (in Chinese)

[4] Feng G L, Dai X G, Wang A H and Chou J F 2001 *Acta Phys. Sin.* **50** 606 (in Chinese)

[5] Shi N, Yi Y M, Gu J Q and Xia D D 2006 *Chin. Phys.* **15** 2180

[6] Abaimov S G, Turcotte D L, Shcherbakov R and Rundle J B 2007 *Nonlinear Proc. Geophys.* **14** 455

[7] Blender R, Fraedrich K and Sienz F 2008 *Nonlinear Proc. Geo.* **15** 557

[8] Bunde A, Havlin S, Kantelhardt J W, Penzel T T, Peter J H and Voigt K 2000 *Phys. Rev. Lett.* **85** 373641

[9] Bunde A, Eichner J F, Kantelhardt J W and Havlin S 2005 *Phys. Rev. Lett.* **94** 048701

[10] Eichner J F, Koscienly-Bunde E abd Bunde A 2003 *Phys. Rev. E* **8** 046133-1

[11] Feng G L, Dong W J and Jia X J 2002 *Acta Phys. Sin.* **51** 1181 (in Chinese)

[12] Feng G L, Dong W J, Gong Z Q, Hou W, Wan S Q and Zhi R 2006 *Nonlinear Theories and Methods on Spatial-Temporal Distribution of the Observational Data* (Beijing: Metrological Press) p. 227 (in Chinese)

[13] Fraedrich K and Blender R 2003 *Phys. Rev. Lett.* **90** 108501

[14] Lennartz S and Bunde A 2011 *Phys. Rev. E* **84** 021129

[15] Feng G L and Dong W J 2003 *Chin. Phys.* **12** 1076

[16] Yuan N and Fu Z 2014 *J. Climate* **27** 1742

[17] Chen Z, Ivanov P, Hu K and Stanley H E 2002 *Phys. Rev. E* **65** 041107

[18] He W P, Deng B S, Wu Q, Zhang W and Cheng H Y 2010 *Acta Phys. Sin.* **59** 8264 (in Chinese)

[19] Giraitis L, Kokoszka P, Leipus R and Teyssiere G 2003 *J. Econometrics* **112** 265

[20] Sun D Y, Zhang H B and Huang Q 2014 *Acta Phys. Sin.* **63** 209203 (in Chinese)

[21] May R 1976 *Nature* **261** 459

[22] He W P 2008 "The Research and Application of the Abrupt Detecting Methods in Dynamical Structure", Ph. D. Dissertation (Lanzhou: Lanzhou University) (in Chinese)

[23] He W P, Feng G L, Wu Q, He T, Wan S Q and Chou J F 2012 *Int. J. Climatol.* **32** 1604

[24] Vondrak J 1969 *Bull. Astron. Inst. Czech.* **20** 349