Earth and Space Science

Robustness Assessment of the RSD *t*-Test for Detecting Trend Turning in a Time Series

Bin Zuo[1], Zhaolu Hou[1], Fei Zheng[2], Lifang Sheng[1] [iD], Yang Gao[3,4] [iD], and Jianping Li[1,4] [iD]

[1]Key Laboratory of Physical Oceanography/Institute for Advanced Ocean Studies/College of Oceanic and Atmospheric Sciences, Ocean University of China, Qingdao, China, [2]State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, [3]Key Laboratory of Marine Environment and Ecology, Ministry of Education/Institute for Advanced Ocean Study, Ocean University of China, Qingdao, China, [4]Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

Abstract Trend turning (or trend change) is a type of structural change that is common in climate data, and methods for detecting it in time series with multiple turning-points need to be developed. A recently developed method for this, the running slope difference (RSD) t-test, examines trend differences in sub-series of the sample time series to identify the trend turning-points. In this paper, we use Monte Carlo simulation to evaluate this method's detection ability. Evaluation results show the method to be an effective tool for detecting trend turning time series and identify three major advantages of the RSD t-test: ability to detect multiple turning-points, capacity to detect all three types of trend turning, and great performance of reducing false alarm rate.

1. Introduction

Detecting structural changes in time series is an important yet complicated task in climate-related research (Karl et al., 2000; Liu et al., 2010; Tomé & Miranda, 2004; Vigo et al., 2005; Xiao & Li, 2007). The forms of structural change include changes of mean (Pettitt, 1979; Sen & Srivastava, 1975; Srivastava & Worsley, 1986), changes of variance (Abraham & Wei, 1984; Hsu et al., 1974; Lee & Park, 2001; Zhao et al., 2010), and changes of trend (Andrews, 1993; Chu & White, 1992; Miao, 1988; Thanasis et al., 2011; Toms & Lesperance, 2003). The structural change of the trend is also known as trend turning or trend change. In 1988, Miao gave the following model of trend turning:

$$x(t) = f(t) + \varepsilon_t,\ 0 < t \leq 1, \tag{1}$$

where $f(t)$ is a nonrandom function of the form

$$f(t) = \begin{cases} \mu + \beta_1(t - t_0),\ 0 < t \leq t_0 \\ \mu + \beta_2(t - t_0),\ t_0 < t \leq 1 \end{cases}, \tag{2}$$

where t_0 is the trend turning-point (initially the slope change point in Miao's paper) under the condition $\beta_1 \neq \beta_2$ (Miao, 1988).

Trend turning is a commonly observed phenomenon in climate data; it can be found in parameters such as Surface Air Temperature (SAT, Xing et al., 2017; Yao et al., 2017; Yu & Lin, 2018), precipitation (Alexander et al., 2006), tropical cyclone frequency (Cheung et al., 2015), and haze days (Zhao et al., 2016). Some studies in ecoclimate have also reported trend turning (Liang et al., 2015; Peng et al., 2011; Piao et al., 2011). Detection of climate trend turning is beneficial to reveal the mechanism of climate change and the internal relationship between different climate variables. A well-known example of this is the trend turning in the global SAT, which is characterized by a multidecadal warming-warming slow down-warming pattern since 20th century (Li et al., 2013; Xing et al., 2017; Yao et al., 2017; Yu & Lin, 2018). Studies have shown that the North Atlantic Oscillation (NAO) has strong impact on the SAT multidecadal variability (Li et al., 2013; Li et al., 2019; Sun et al., 2015; Wang et al., 2017). The positive NAO forces the enhancing of the Atlantic meridional overturning circulation (AMOC), resulting in a positive phase of Atlantic Multidecadal Oscillation (AMO), which can affect the SAT not only in pan-Atlantic regions but also in the northern hemisphere

Table 1
The Detection Abilities of the RSD t-Test Assessed in Each Monte Carlo (MC) Simulation Experiment

| Experiment	Error types in detecting the trend turning-point	Possible influencing factors
MC1	False acceptance	Size of the sample, noise amplitude
MC2	False rejection	Trend turning type, noise amplitude
MC3	False rejection	Trend turning amplitude
MC4	False rejection	Size of the sample
MC5	False rejection	Interference from nearby turning-points in multiple turning-point cases

and even globally (Kravtsov & Spannagle, 2008; Sun et al., 2015; Sun et al., 2017; Sun et al., 2019a; Sun et al., 2019b; Tung & Zhou, 2013; Wyatt & Curry, 2014; Wyatt & Peters, 2012).

Currently, methods for detecting trend turning require further development, as existing methods are often restricted to a single turning-point (Andrews, 1993; Chu & White, 1992; Perron & Zhu, 2005; Ploberger & Kramer, 1996; Toms & Lesperance, 2003), and methods for multiple turning-points are relatively few and inadequate. Optimal piecewise linear regression (OPLR; Liu et al., 2010; Tomé & Miranda, 2004) and running trend test (RTT; Meehl et al., 2011; Thanasis et al., 2011) are the two commonly used methods for detecting multiple turning-points time series. However, both OPLR and RTT methods struggle to detect same sign but different trend magnitude (TD) type of turning (Zuo et al., 2019).

Trend turning can be divided into three main types according to the different combinations of the trends before and after turning: Trend Reversal (TR) type is turning from a significant positive trend to a significant negative trend (or vice versa), Trend/No-trend (TN) type involves turning from a significant linear trend to no significant trend (or vice versa), and same sign but Different Trend magnitude (TD) type turning occurs between two significant trends (both either positive or negative) of different magnitude. TD type of turning is as important as the TN and TR turning in climate research; it is quite common in observational data which closely related to climate change (e.g., global atmospheric CO_2 concentration, Zuo et al., 2019; and arctic seasonal sea ice extent, Meehl et al., 2018). However, the existing OPLR and RTT method cannot effectively detect TD type of trend turning. OPLR method computes the best piecewise regression combination of the sample time series that minimize the residual sum under the condition that two nearby sub-series trends have opposite sign (Tomé & Miranda, 2004). This condition of opposite sign will reject trend turnings in which both sub-series have the same sign (i.e., TD type of trend turning). RTT method detects trend turnings making use of trend significant test on running windows along the sample time series, if the running test results (there are three possible outcomes: significant positive trend, significant negative trend, and no significant trend) changed after a certain point, the RTT will return this point as the trend turning-point (Thanasis et al., 2011). Since both sides of TD trend turning are significant positive trend (or significant negative trend), RTT cannot effectively detect this type of trend turning as well.

We recently proposed the running slope difference (RSD) *t*-test as a novel method for detecting trend turning (Zuo et al., 2019). It can detect multiple turning-points from all three turning types using a principle different from either OPLR or RTT. It conducts statistical *t*-test of slope differences (of two nearby sub-series) to identify the turning-points, as the sub-series slope difference between both sides of the turning-point is the basic characteristic of all three turning types.

In the previous study which based on an idealized time series case and several climate time series cases, we reported that the RSD *t*-test performs better than either OPLR or RTT in identifying TD trend turning, and its probability of triggering false alarm is also lower compared to the two existing methods (Zuo et al., 2019). However, there is a lack of robustness assessment of the RSD *t*-test for detecting trend turning because the former research achievements are based on case studies. In addition, the possible factors that influence the detection ability of the RSD *t*-test have not been discussed either. In this article, we will assess the detection ability of the RSD *t*-test and study various factors that influence its performance in several Monte Carlo simulation experiments (Tomé & Miranda, 2005; Yue et al., 2002). These influencing factors include trend turning amplitude, noise amplitude, and interference from nearby turning-points in multiple
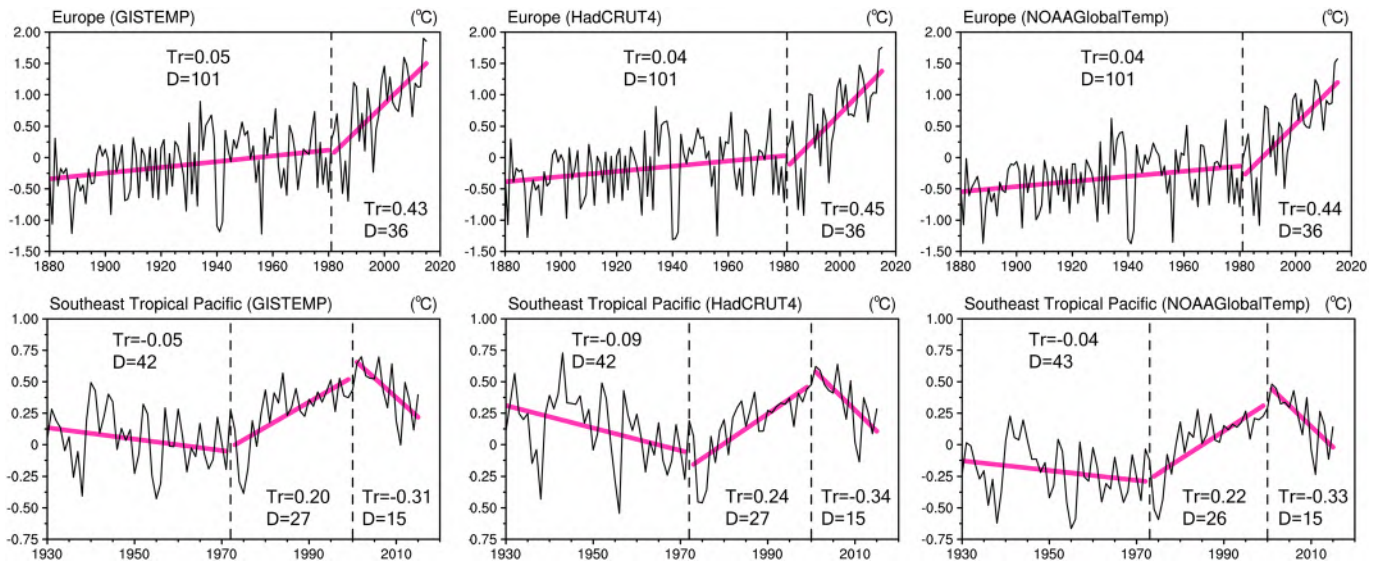
Figure 1. Trend turning detection results (purple line) of two regional surface air temperature (SAT) time series (solid, black line) using the RSD *t*-test method (detection parameter timescale as 15 years, confidence level at 95%). The two regions are Europe (40–60°N, 0–40°E) and southeast tropical Pacific (10–30°S, 100–150°W). "Tr" and "D" represent the least squares trend value and duration of each sub-series, respectively. Note that the trend value unit here is °C per decade and the duration unit is year.

turning-points cases (details please see Table 1). This paper will provide researchers meaningful references for choosing trend turning detection methods.

The remainder of this article is structured as follows. Section 2 describes the RSD *t*-test method and the evaluation of its detection ability. Its ability to avoid false acceptance and false rejection errors is discussed in sections 3 and 4, respectively. Section 5 discusses a simple and useful pretreatment for small sample series. Finally, the conclusions will be presented in section 6.

## 2. Method

2.1. The RSD *t*-Test Method

The statistical model of trend turning by Miao (1988) suggests that its basic feature is the significant trend difference between two sub-series around the turning-point ($\beta_1 \neq \beta_2$). Given this, the RSD *t*-test is designed using a statistical *t*-test of slope differences to identify the turning-points.

Let $Y : \{y_i = \beta_Y i + \alpha_Y + \varepsilon_i | 1 \leq i \leq n\}$ and $Z : \{z_j = \beta_Z j + \alpha_Z + \varepsilon_j | 1 \leq j \leq m\}$ be two sample sub-series. Assume that the error terms ε_i and ε_j are normally distributed independent random variables with zero mean and same variance σ^2. The form of this statistical *t*-test for slope differences is as follows:

$$t_{slope} = \frac{\sqrt{C}\left(\widehat{\beta}_Y - \widehat{\beta}_Z\right)}{\sqrt{\frac{\left(\sum\limits_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2 + \sum\limits_{j=1}^{m}\left(z_j - \widehat{z_j}\right)^2\right)}{n + m - 4}}}, \quad C = \frac{NM}{N + M}, \tag{3}$$

where $\widehat{\beta}_Y$ and $\widehat{\beta}_Z$ are the least squares linear slope of Y and Z, respectively, $\widehat{y_i}$ and $\widehat{z_j}$ are the least squares regressions of y_i and z_j, respectively, and $N = \sum\limits_{i=1}^{n}\left(i - \frac{n+1}{2}\right)^2$ and $M = \sum\limits_{j=1}^{m}\left(j - \frac{m+1}{2}\right)^2$ are two known constants of n and m, respectively. The null hypothesis of no slope difference between Y and Z is rejected at the significance level α if $|t_{slope}| \geq t_{f,1-\alpha/2}$, where $f = n + m - 4$ is the degrees of freedom of the statistic t_{slope}.

If the above Y and Z are two adjacent sub-series of the sample time series, then we can use the statistic t_{slope} to detect whether there is a trend turn between Y and Z. Let L be the sample time series and τ be the chosen
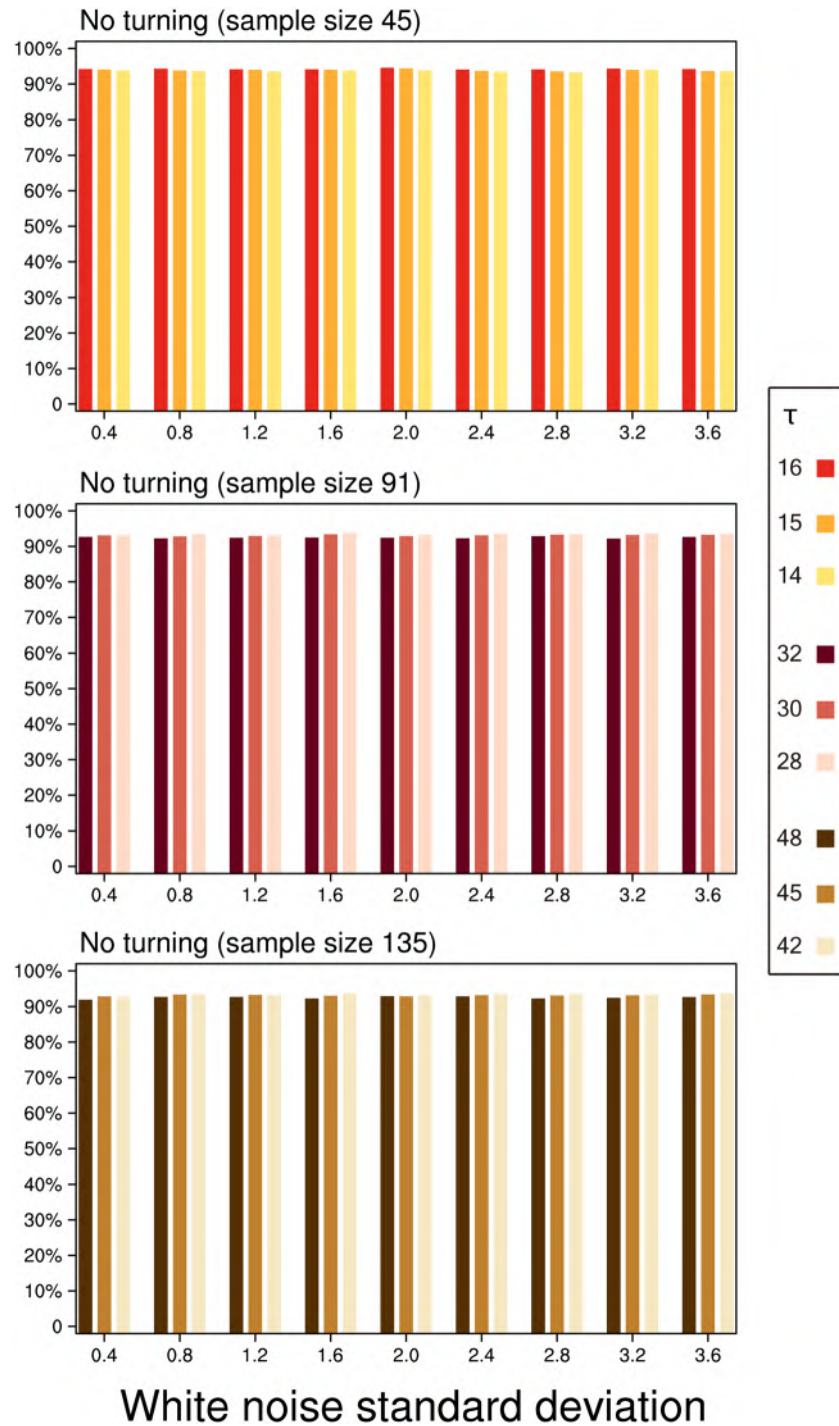
Figure 2. Ability of the RSD *t*-test to avoid false acceptance errors (A_{fa}) in MC1 as a function of the standard deviation of superposed white noise using different sizes of sliding window, τ. Top, middle, and bottom charts are for SS series (synthetic sample series) sample sizes of 45, 91, and 135, respectively.

sliding window. The trend turning-point of L is identified from the slope difference between the sub-series (for a length of sliding window τ) before and after each point using the t_{slope} test.

Figure 1 is an example of using the RSD *t*-test detecting trend turning in regional SAT time series (temperature data from GISTEMP, Hansen et al., 2010; HadCRUT4, Morice et al., 2012; and NOAAGlobalTemp, Vose et al., 2012). For more case studies and other details of the RSD *t*-test method, please refer to Zuo et al. (2019).
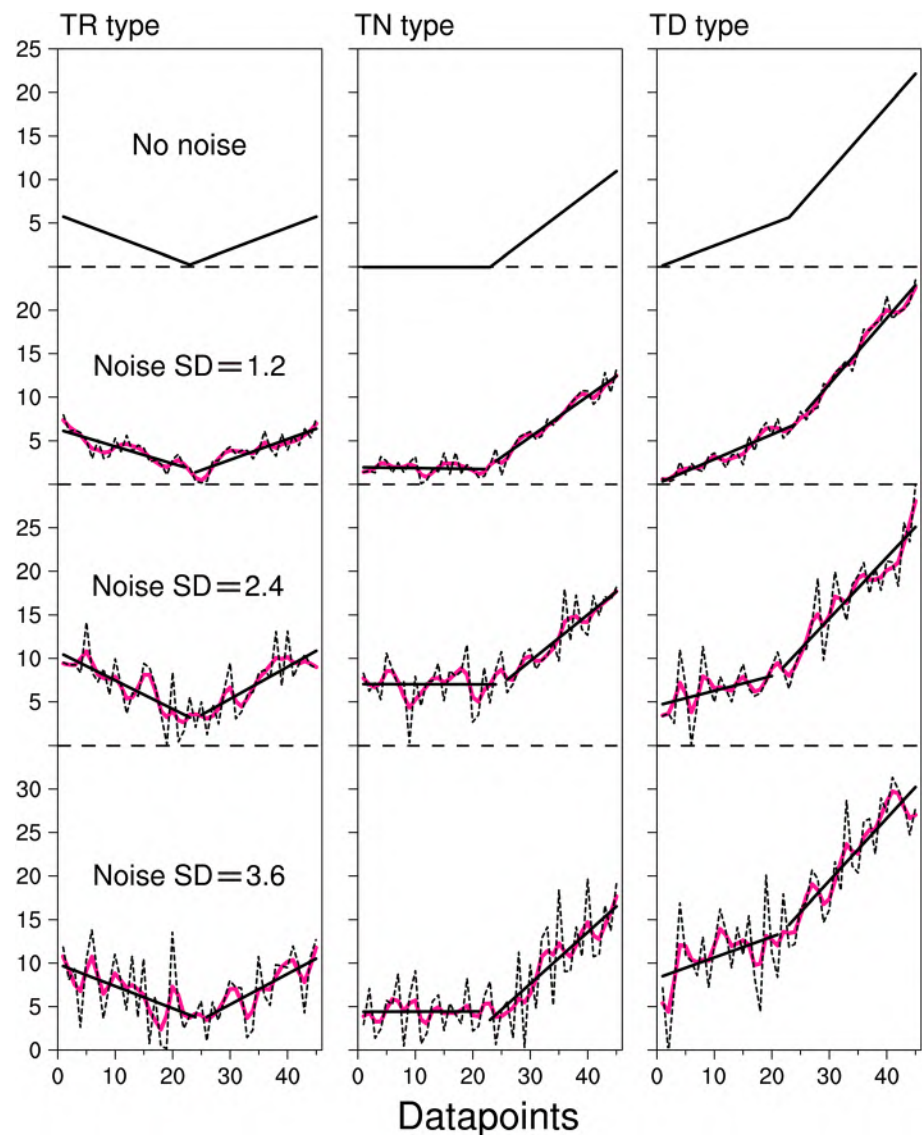
Figure 3. SS series (dashed lines) of the three types of trend turning used in MC2 with their detection results (solid, black, two-section line; $\tau = 15$). "No noise" represents the trend turning true value series, and "noise SD" is the standard deviation of the superposed white noise. Purple lines are the SS series after de-noising pretreatment using a 5-point Gaussian filter.

2.2. Evaluation of Detection Ability

Monte Carlo simulation experiments are used to evaluate the detection ability of the RSD t-test method. A large number of sample series are first generated by Monte Carlo simulations (these sample series are here-inafter called SS series, Synthetic Sample series); they are pre-designed with or without trend turning-points. All these SS series are then analyzed by the RSD t-test, and the detection ability of the test method can then be evaluated given its detection results for identifying the pre-designed turning-points or returned false turning.

To evaluate the method's ability to avoid false turning (i.e., to avoid the false acceptance type of statistical error), SS series are generated in the simulation experiments as independent, normally distributed series with zero mean (white noise series). A white noise series has neither significant overall trend nor trend turn-ing; therefore, any turning-point returned by the detection method will be a false positive. The ability of the method to avoid false acceptance is defined by
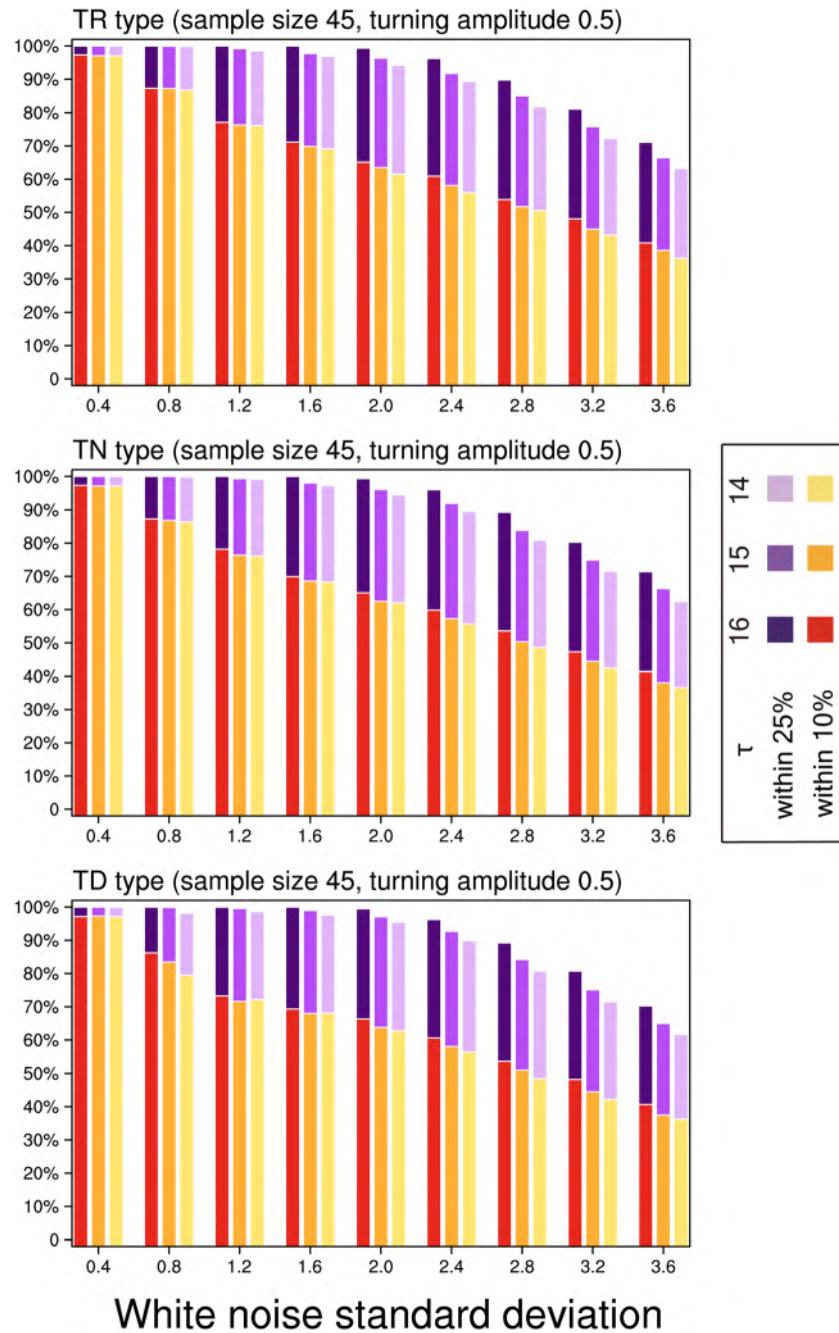
Figure 4. Ability of the RSD *t*-test to avoid false rejection errors (A_{fr1} and A_{fr2}) in MC2 as a function of the standard deviation of the superposed white noise using different sizes of sliding window, τ. A_{fr1} represents the ability of the method correctly to locate the turning-point (with an error between detection result and its true value of less than 10% of the length of the two sub-series beside the turning-point, only bottom bar). A_{fr2} represents the ability of the method correctly to judge the existence of a turning-point, but not necessarily accurately locate it (error less than 25% of the length of two sub-series, bottom bar plus top bar).

$$A_{fa} = \frac{N_{fa}}{N}, \tag{4}$$

where N is the total number of SS series in the simulation experiments and N_{fa} is the number of correctly detected SS series (i.e., those not identified as have false turning).
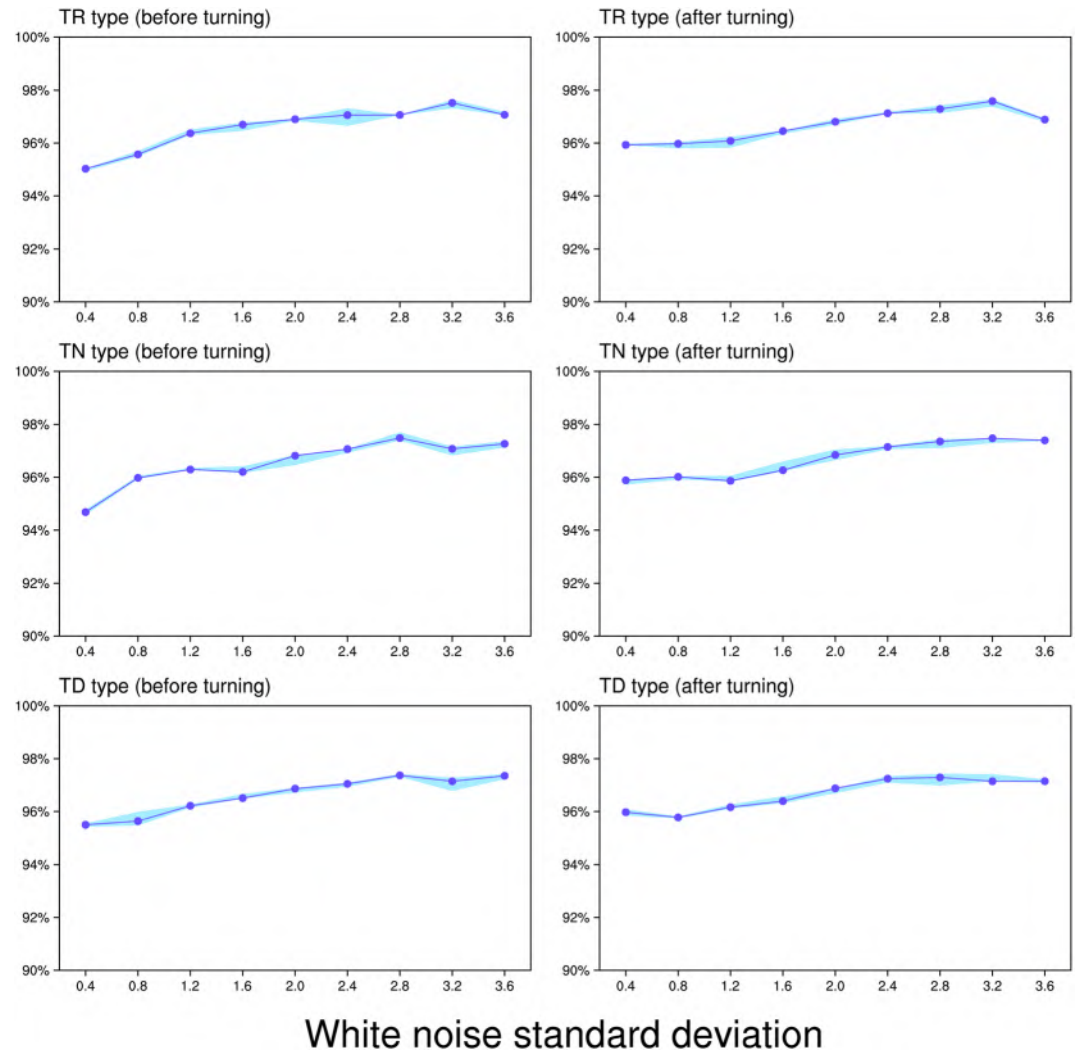
Figure 5. Ability of the RSD *t*-test to estimate the partial trend of a sub-series (A_{slope}) in MC2 as a function of the standard deviation of the superposed white noise using different sizes of sliding window, τ. The lines represent the middle value of three results obtained using different sliding windows τ, and shading indicates the range of change.

To evaluate the method's ability to identify turning-points (i.e., to avoid the false rejection type of statistical error), SS series are generated as combinations of two parts: One part is a true value series, and the other is a superimposed white noise series. The true value series has two linear sections either side of a pre-designed trend turning-point, and the superimposed white noise series is an independent normally distributed series with zero mean and has the same sample size as the true value series. Detection ability here is evaluated through whether the method identified the pre-designed turning-points. As the statistical properties of the true value series may have changed due to the superimposed noise, small errors between the detection results and the pre-designed true value should be allowed. Here, let l_1 and l_2 be the lengths of the two sub-series beside the turning-point. If the error between the detection result and its true value is less than 10% of min (l_1, l_2), the detection result can be considered correct. Correct judgment of the existence of turning-points is also important (allowing for possible incorrect location of the turning-point); therefore, we apply another index that allows an error no more than 25% of min(l_1, l_2). The ability of the RSD *t*-test to avoid false rejection is defined by both

$$A_{fr1} = \frac{N_{fr1}}{N} \quad \text{and} \tag{5}$$

$$A_{fr2} = \frac{N_{fr2}}{N}, \tag{6}$$
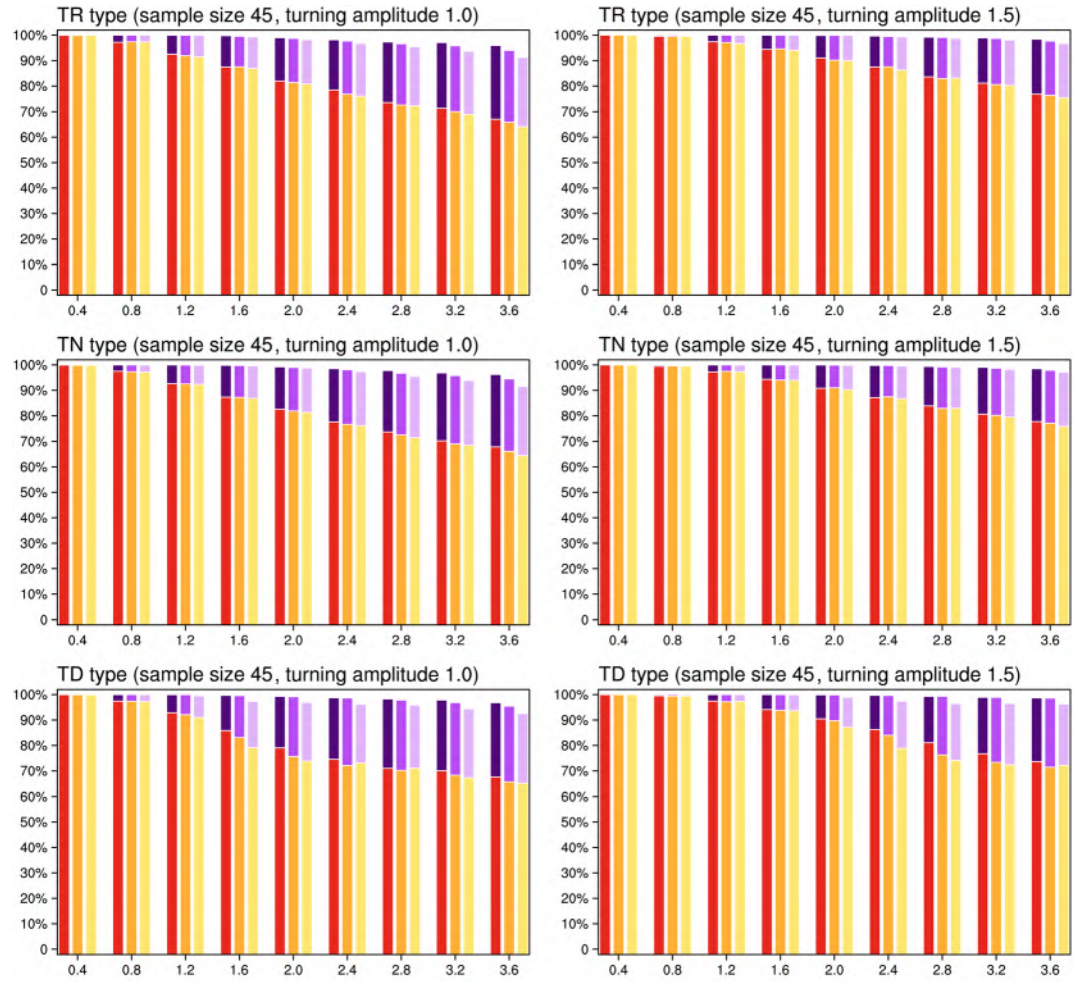
Figure 6. Abilities A_{fr1} and A_{fr2} in MC3 as a function of the standard deviation of the superposed white noise using different sizes of sliding window, τ. The conditions of MC3 are the same as those for MC2, except that the trend turning amplitude is 1.0 (double that in MC2, left) and 1.5 (triple that in MC2, right).

where N_{fr1} is the number of SS series whose error between detection result and true value is less than 10% of $\min(l_1, l_2)$ and N_{fr2} is the number with error less than 25%. In summary, A_{fr1} is the proportion that the method correctly located the turning-points in the Carlo simulation experiment, and A_{fr2} ($A_{fr2} \geq A_{fr1}$) is the proportion that the method correctly determined the existence of turning-points (allowing incorrectly locating of the turning-point).

For sample series containing trend turning, estimating the trend of each partial segment is as important as locating the turning-point. Another parameter, A_{slope}, for evaluating the ability of the method to estimate the partial trend of each sub-series is defined by

$$A_{slope} = \frac{N_{slope}}{N_{fr1}}, \tag{7}$$

where N_{slope} is the number of SS series whose pre-designed partial trends have no significant difference (at the 99% confidence level) from those estimated. As accurate estimation of the partial trend of each sub-series relies on locating the turning-point, only SS series with correctly identified turning-points are included.
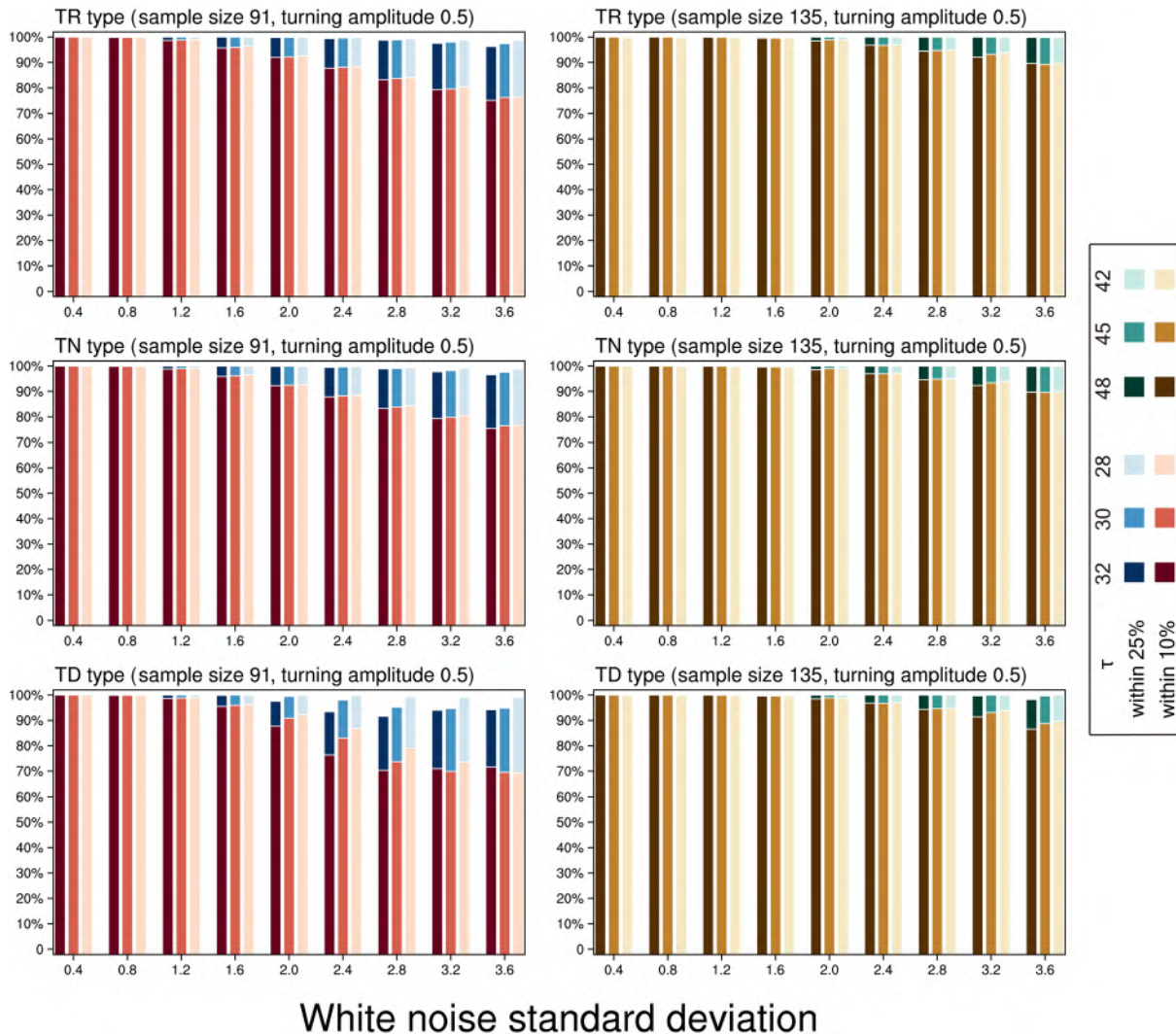
Figure 7. Abilities A_{fr1} and A_{fr2} in MC4 as a function of the standard deviation of the superposed white noise using different sizes of sliding window, τ. The conditions of MC4 are the same as those in MC2, except the sample sizes are 91 (double that in MC2, left) and 135 (triple that in MC2, right), and τ is similarly doubled and tripled.

3. False Acceptance

Possible factors affecting the detection ability regarding false acceptance are the size of the sample series and the magnitude of white noise (SD_{noise}). The first Monte Carlo simulation experiment (MC1) here generates 10,000 independent white noise series for each sample size (45, 91, and 135), with nine different levels of SD_{noise} from 0.4 to 3.6 (in 0.4 increments).

Two detection parameters used in the RSD t-test (being confidence level, P, and sliding window, τ) are set before trend turning detection. The confidence level, P, is the threshold of the slope difference test during trend turning detection: Commonly used values are 90%, 95%, and 99%. A large P value (near 100%) will increase the method's ability to avoid false acceptance but increase the probability of false rejection. In order to evaluate both types of statistical error appropriately, P is selected as 95% here. The other parameter, sliding window, τ, should be determined in practice according to the timescale of the specific study. Here, we aim to show the detection ability of the RSD t-test in an idealized situation with appropriate parameters; therefore, for SS series of sample size 45, 91, and 135, τ is selected to be around 15, 30, and 45, respectively. As this study also intends to show the influence of small changes of the sliding window on the detection ability, each SS series in the experiment is detected three times using three close, but distinct, τ values.
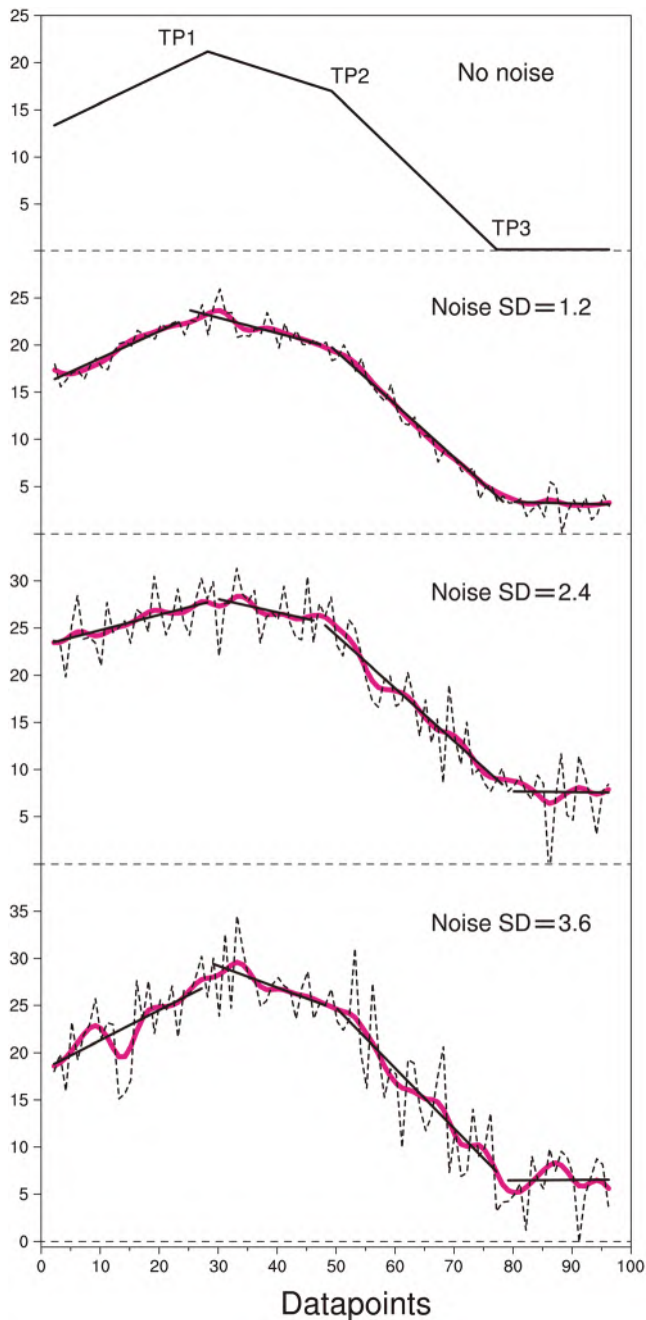
Figure 8. SS series (dash line) used in MC5 with their detection results (black solid linear sections; $\tau = 15$). "No noise" represents the true value series (with three turning-points TP1, TP2, and TP3), and "noise SD" is the standard deviation of the superposed white noise. Purple lines are the SS series after de-noising pretreatment using a 5-point Gaussian filter.

Figure 2 shows the results of the detection ability A_{fa} in experiment MC1. The top figure is the result for SS series of sample size 45: In each case A_{fa} fluctuates around 94%, and neither varying the white noise level nor slightly changing τ greatly affects it (impact less than $\pm 0.50\%$). These results indicate that the RSD t-test has a great performance of avoiding false acceptance, which is a major advantage of this method.

The detection ability A_{fa} for SS series of sample sizes 91 and 135 is given respectively in the middle and bottom figures of Figure 2. The A_{fa} values are within a narrow range between 92% and 94%, thus showing that the sample size of the SS series has no notable effect on detection ability. The next section assesses the significant influence of SS series sample size on the abilities A_{fr1} and A_{fr2} and discusses the reasons for the dissimilarity to the results seen here for ability A_{fa}.

4. False Rejection

There are more possible factors affecting false rejection besides the size of the sample series and the magnitude of white noise, for example, trend turning amplitude $|\beta_1 - \beta_2|$ and interference from nearby turning-points (in cases with multiple turning-points).

This section first studies the influence of the white noise amplitude, SD_{noise}, the trend turning amplitude, $|\beta_1 - \beta_2|$, and the SS series sample size on the ability of the RSD t-test. The experiments (MC2–MC4) involve several single turning-point simulations. The following experiment (MC5) examines interference from nearby turning-points in multiple turning-point simulations.

4.1. Single Turning-Point

The false rejection tests consider all three main types of turning-point, with each series having a single turning-point of type TR, TN, or TD. The SS series in Monte Carlo simulation experiments MC2–MC4 are generated by superimposing white noise series on the true value series. Take the sample in MC2 (Figure 3) as an example. Its true value series is a two-section line with sample size 45; it contains one pre-designed trend turning-point at the 23rd data point (Figure 3, "no noise"). The trend turning amplitude $|\beta_1 - \beta_2|$ for all three types of turning-point is set as 0.5, and each section has a set trend value: TR ($\beta_1 = -0.25, \beta_2 = 0.25$), TN ($\beta_1 = 0, \beta_2 = 0.50$), and TD ($\beta_1 = 0.25, \beta_2 = 0.75$). The superimposed white noise series in MC2 is of equal sample size to the true value series; it is an independent normally distributed series with zero mean and standard deviation varying from 0.4 to 3.6 (in nine level at 0.4 increments). For each different turning type and noise amplitude level, MC2 generates 10,000 independent white noise series, similar to in MC1.

In experiment MC2, the RSD t-test is applied to detect trend turning in each series using the same confidence level and sliding window as in the MC1 cases with sample size 45 (i.e., P as 95% and τ is selected to be around 15).

Figure 4 shows results for detection abilities A_{fr1} and A_{fr2} in experiment MC2. As expected, the RSD t-test detects TD type turning with similar ability to types TR and TN, indicating its capacity to detect all three types. Both detection abilities, A_{fr1} and A_{fr2}, for all three turning types approach 100% at low noise but decrease with increasing noise. Even at the highest level of white noise considered here, the method shows a 65% rate of correctly judging the presence of a turning-point, and a 40% rate of correctly locating it. The
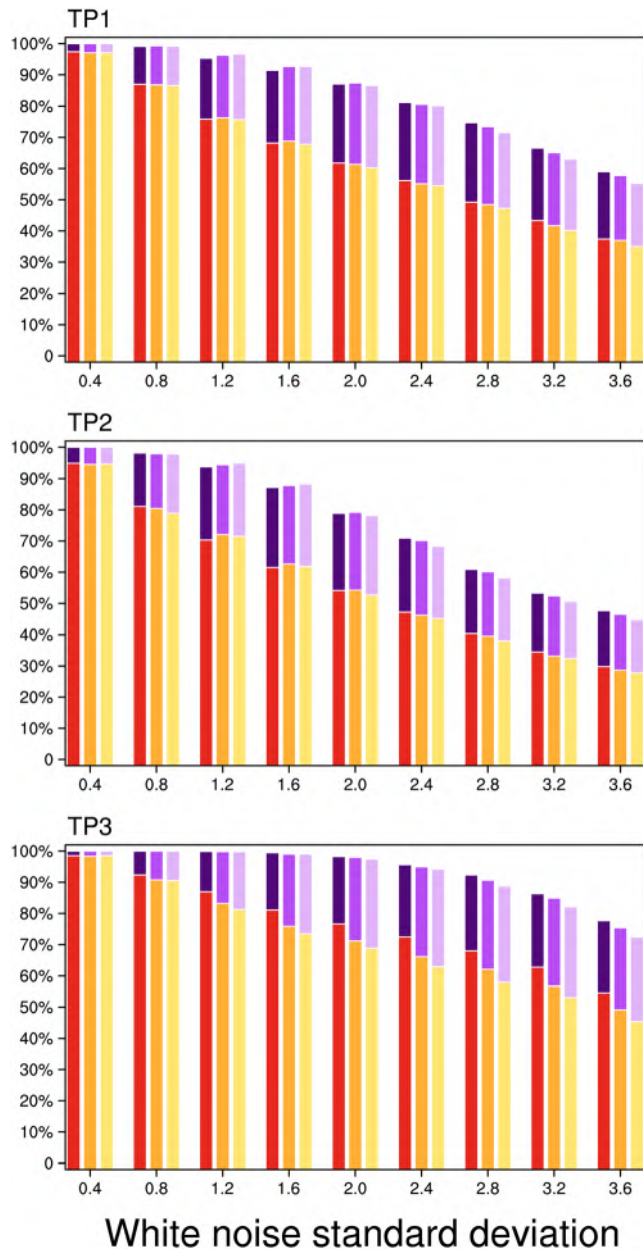
Figure 9. Abilities A_{fr1} and A_{fr2} in MC5 as a function of the standard deviation of the superposed white noise using different sizes of sliding window, τ.

minor variations of sliding window size, τ, have little effect on detection capability: for A_{fr1} the influence ranges from approximately 0.20% to 2.5% as the noise level rises.

For sample series with trend turning-point, if the detection method can correctly locate the turning-point, it will have a high probability of correctly estimating the trend value of each sub-series. Figure 5 shows results for sub-series trend estimation, A_{slope}, in MC2. All estimations of partial trends are above 90% (most great than 95%), indicating that the trend of a sub-series can be well estimated if the turning-point has been correctly located. These results further emphasize the importance of correctly identifying the turning-point during trend turning detection.

Simulation experiment MC3 (Figure 6) considers the influence of different trend turning amplitudes on the detection ability. Its experiment design is similar to that in MC2, except that the trend turning amplitude is 1.0 or 1.5 (compared with 0.5 in MC2). This increase improves the detection ability: Doubling the amplitude to 1.0 increases A_{fr1} by about 25% at a high level of white noise, and tripling it to 1.5 sees an increase of about 35%. Increasing the trend turning amplitude also reduces the influence of small changes of τ, from about 2.5% to 1.5% at a high level of noise for A_{fr1}.

Simulation experiment MC4 (Figure 7) is also based on MC2 and considers the influence of the SS series sample size on detection ability. Series lengths are 91 and 135 (compared with 45 in MC2), and the turning-point is moved to the 46th and 68th data points, respectively. However, the trend values β_1 and β_2 before and after the turning-point are both the same as those in MC2. The sliding window, τ, in MC4 doubled and tripled with the series size. Detection results of MC4 (Figure 7) show that the detection abilities A_{fr1} and A_{fr2} increase to nearly 75% and 97%, respectively, when the sample series size is doubled, and to about 90% and 100%, respectively, when it is tripled. Changing the SS series sample size clearly influences the abilities A_{fr1} and A_{fr2}.

In comparison with a small sample size, a larger sample size will have more points included during t_{slope} testing. In this case, if the test sample contains a trend turning-point, the difference in slopes either side of the turning-point will be more significant than for a small sample size. However, if the test sample contains no trend turning, changes in sample size will have only limited influence on the detection results. This feature can also be seen from the formula of statistic t_{slope}. From formula (3), the numerator of statistic t_{slope} is defined by the square root of C multiplied by the slope difference $\widehat{\beta}_Y - \widehat{\beta}_Z$. C is the constant related to the sample size of the two sample sub-series Y and Z. A larger sample size will increase the statistic t_{slope} when the slope difference is not near zero and will eventually improve the RSD t-test when identifying the turning-point. This is why changing of SS series sample size has only limited influence on the detection ability A_{fa} but notably affects the abilities A_{fr1} and A_{fr2}.

4.2. Multiple Turning-Points

This section discusses samples with multiple turning-points (experiment MC5). Like the previous experiments, the Monte Carlo simulation experiments use SS series generated by superimposing white noise series on the true value series. The data series comprise four linear sections (containing three turning-points) of the true value series of sample size 95 (Figure 8, "no noise"). To make the experiments comprehensive, each of the three turning-point types is included: The first turning-point (TP1, at data point 27) is type TR; the

Table 2
Percentage Identification of Turning-Points TP2 and TP3 in MC5 Under the Condition of Correct (or Incorrect) Identification of TP1

Noise SD	$\frac{N(TP2,TP1)}{N(TP1)}$	$\frac{N(\overline{TP2},TP1)}{N(TP1)}$	$\frac{N(TP3,TP1)}{N(TP1)}$	$\frac{N(TP3,\overline{TP1})}{N(\overline{TP1})}$
0.4	98.65%	93.22%	99.87%	98.31%
0.8	85.21%	70.91%	95.11%	95.51%
1.2	78.69%	61.38%	87.84%	87.71%
1.6	72.85%	50.31%	80.14%	78.88%
2.0	64.96%	42.79%	75.49%	74.78%
2.4	57.77%	37.46%	70.23%	69.69%
2.8	47.56%	34.30%	64.61%	64.82%
3.2	36.94%	32.14%	59.75%	58.54%
3.6	27.10%	31.19%	52.16%	52.41%

Note. $N()$ is the number of SS series (synthetic sample series) that meet the conditions in the bracket, with TP1, TP2, and TP3 representing correct identification of each turning-point, and an overbar representing incorrect identification. "Correct" here means that the error between the returned turning-point and its true value is less than 10% of the length of the two sub-series beside the turning-point.

second turning-point (TP2, at data point 48) is type TD; and the third turning-point (TP3, at data point 76) is type TN. The trend values of the four sub-series are set as 0.3, −0.2, −0.6, and 0. The superimposed white noise series are generated independently with nine different standard deviation levels from 0.4 to 3.6 (each with 10,000 independent samples). The detection parameters P and τ in MC5 are the same as those in MC2.

Figure 9 shows results for A_{fr1} and A_{fr2} for all three turning-points in MC5. Comparison with Figure 4 shows that the detection ability of each turning-point in series with multiple turning-points is similar to that for series with only one. In MC5, the RSD t-test demonstrates the ability to detect all three types of trend turning, with both detection abilities A_{fr1} and A_{fr2} for all three turning-points being near 95–100% at a low white noise and decreasing as the noise level rises.

The detection ability for each turning-point in a multiple turning-point sample is not greatly affected by the other nearby turning-points, as shown more clearly in Table 2, which lists the detection ability A_{fr1} for TP2 and TP3 under the condition of TP1 being or not being correctly located ($\tau = 15$). Whether TP1 is correctly located barely influences the detection of TP3. For the nearer turning-point TP2, although the detection of TP1 does influence its detection, failure to detect TP1 does not necessarily lead to failed detection of TP2.

As the detection of each turning-point in a multiple turning-point sample is largely independent, the RSD t-test shows similar abilities to detect single or multiple turning-points. The results in Figure 10 for the detection ability A_{slope} for each sub-series in experiment MC5 show that the trend estimation for a sub-series will likely be close to the true value if the turning-point has been correctly located.
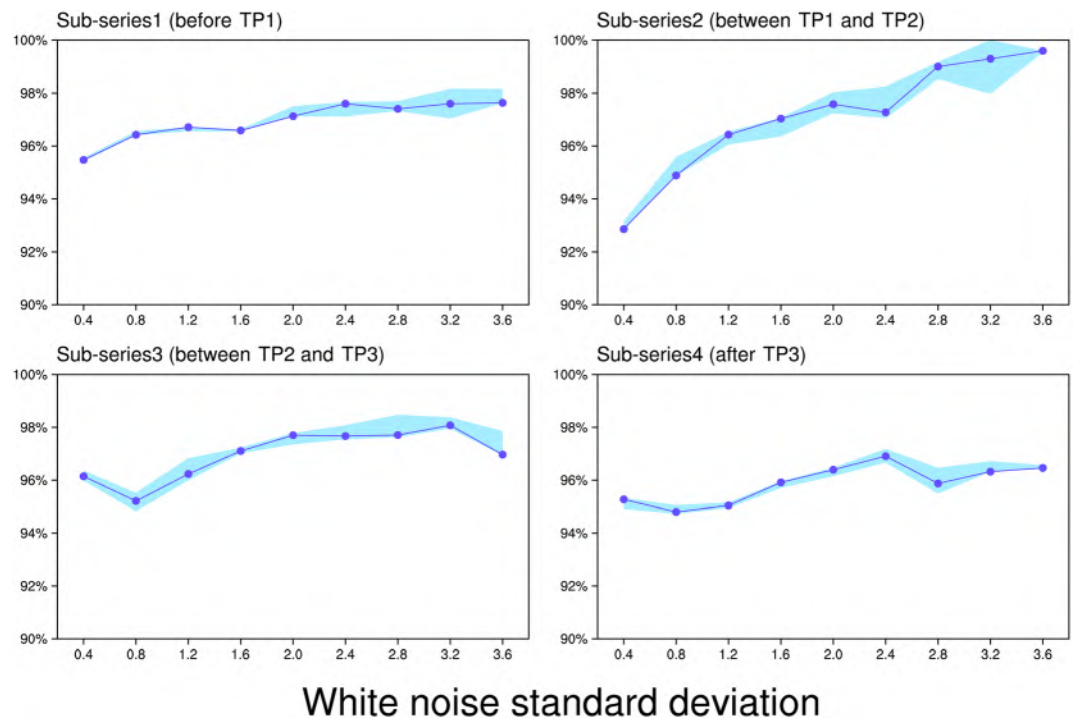
Figure 10. Ability A_{slope} in MC5 as a function of the standard deviation of the superposed white noise using different sizes of sliding window, τ. The lines represent the middle value of three results obtained using different sliding windows τ, and shading indicates the range of change.
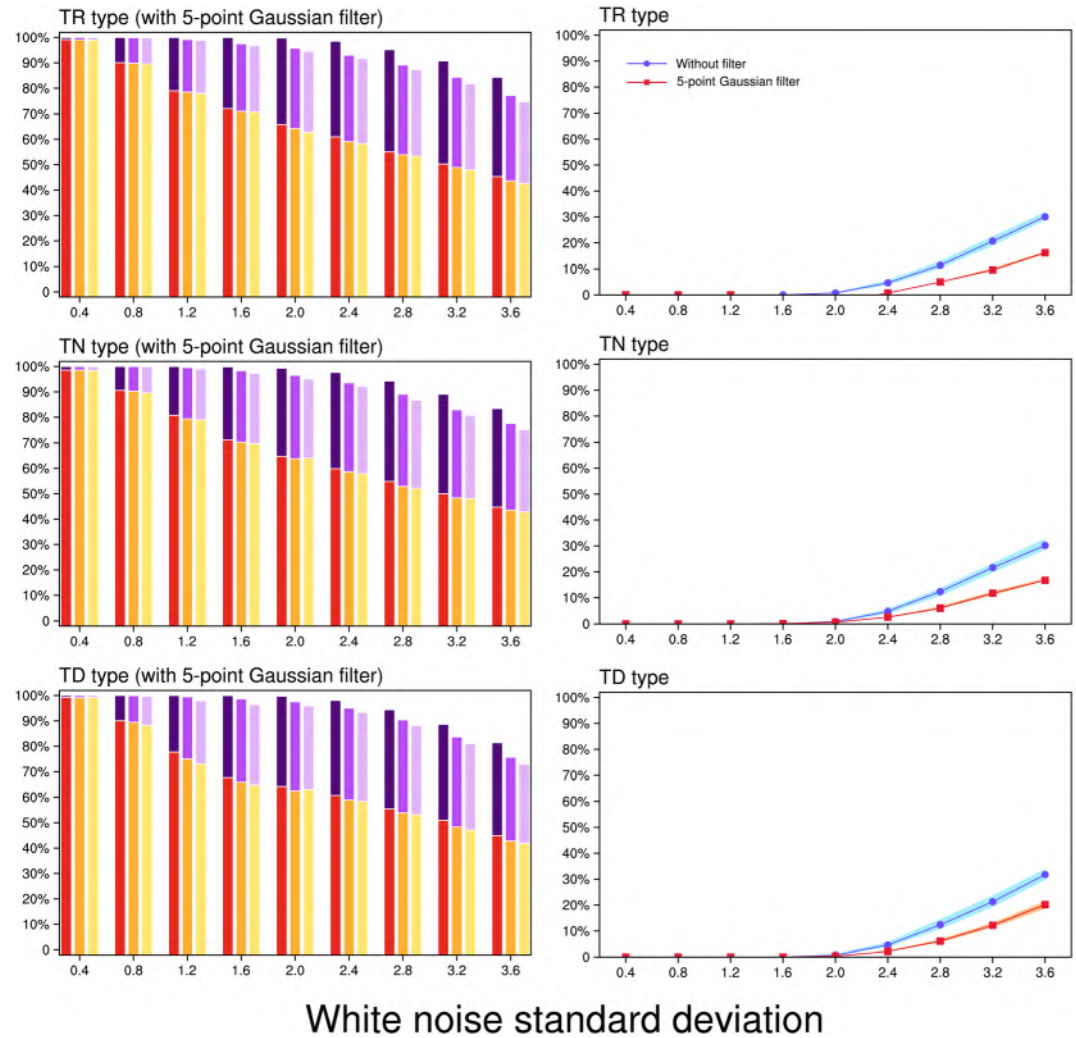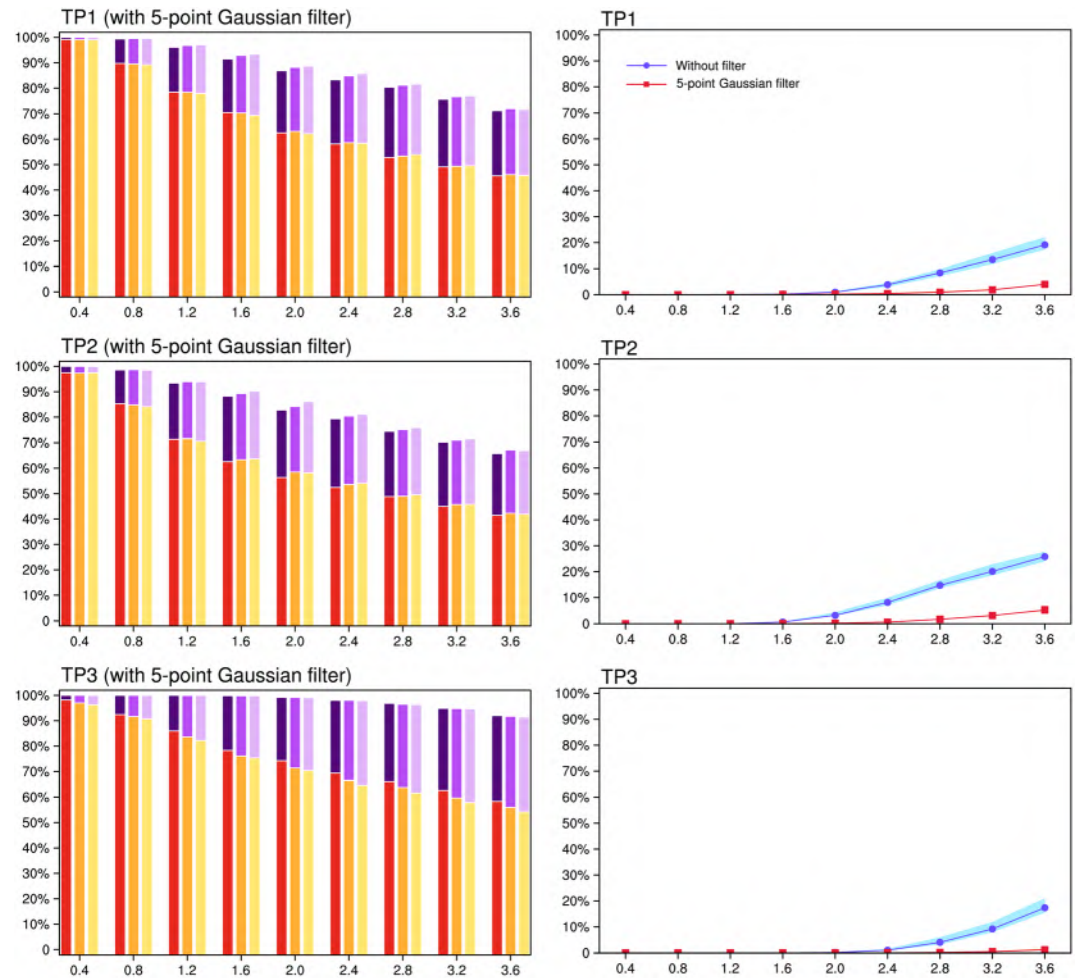
Figure 11. (left) Abilities A_{fr1} and A_{fr2} in MC2, as in Figure 4 after de-noising pretreatment using a 5-point Gaussian filter. (right) Percentage of series samples their turning-points not detected (i.e., failed detection) in MC2. (right) Percentage of series samples for which turning-points were not detected (i.e., failed detection) in MC2.

5. De-noising Pretreatment

As discussed in the previous section, a large sample size will lead to better detection results. However, small samples (such as MC2) are unavoidable in practice. Therefore, a de-noising pretreatment is introduced here to improve the detection ability of the RSD *t*-test for small sample sizes.

Climate variability is frequently characterized by the multi-timescales (Ji et al., 2014; Wang et al., 2009); small timescale disturbance is a kind of noise to the long timescale trend. Using filter to de-noise the sample time series, which is a very common technique during trend analysis in climatology research (Xie et al., 2019; Xing et al., 2017), can better highlight the trend of sample time series. As for the detection of trend turning, the results in the previous experiments also showed that the excessive random noise is usually the reason to the failed trend turning detection. Thus, a small filter may be able to improve the detection ability of the RSD *t*-test. In addition, a recent RTT detection method proposed by Thanasis et al. (2011) which used a small filter to remove glitches while retaining the signature of the data demonstrates the effectiveness of using filter in practice as well. Therefore, a pretreatment step using a 5-point Gaussian filter was added to each SS series in MC2 (single turning-point) and MC5 (multiple turning-points). The left part of Figure 11 shows results for the detection abilities A_{fr1} and A_{fr2} in MC2 after the pretreatment step. Pretreatment improved both A_{fr1}

Figure 12. (left) Abilities A_{fr1} and A_{fr2} in MC5, as in Figure 9, but after de-noising pretreatment using a 5-point Gaussian filter. (right) Percentage of series samples their turning-points not detected (i.e., failed detection) in MC5. (right) Percentage of series samples for which turning-points were not detected around 15 data points of TP1, TP2, or TP3 in MC5.

Table 3
Percentage Identification of Turning-Points as in Table 2, But After De-noising Pretreatment Using a 5-Point Gaussian Filter

Noise SD	$\frac{N(TP2,TP1)}{N(TP1)}$	$\frac{N(TP2, TP1)}{N(TP1)}$	$\frac{N(TP3,TP1)}{N(TP1)}$	$\frac{N(TP3, TP1)}{N(TP1)}$
0.4	99.50%	100%	98.99%	100%
0.8	90.85%	78.84%	96.53%	97.18%
1.2	79.67%	60.43%	90.01%	88.72%
1.6	73.52%	48.76%	81.26%	79.64%
2.0	69.26%	44.42%	75.16%	74.77%
2.4	66.50%	40.25%	69.79%	69.89%
2.8	63.03%	37.40%	65.62%	66.25%
3.2	59.65%	34.25%	61.29%	61.78%
3.6	56.23%	33.46%	58.73%	57.62%

(by approximately 5%) and A_{fr2} (by 10%) at a high white noise level. The right part of Figure 11 shows that the percentage of no turning-point detection (i.e., failure of trend turning detection) in this experiment clearly drops after de-noising. These results indicate that adding the de-noise pretreatment step can effectively improve the detection ability of the RSD *t*-test.

Figure 12 shows results for A_{fr1} and A_{fr2} in MC5 after pretreatment. The situation for experiment MC5 is similar to that for MC2 and further supports the effectiveness of de-noising pretreatment. An important question is whether this pretreatment will invalidate the relative independence of the ability to detect each turning-point individually in series with multiple turning-points. This is assessed by considering A_{fr1} for TP2 and TP3 under the condition of TP1 being or not being correctly located by the method after de-noising pretreatment. The results in Table 3 show that

de-noising does not significantly affect the independence of detecting each point in samples with multiple turning-points.

6. Conclusions

This paper evaluates the detection ability of the RSD t-test in several Monte Carlo simulation experiments. The tested ability criteria are avoiding false acceptance (A_{fa}), correctly locating a turning-point (A_{fr1}), correctly judging the presence of a turning-point (A_{fr2}), and estimating the trend of each partial segment (A_{slope}).

The results show that the RSD t-test has a great performance of reducing false acceptances regardless of the sample series size or magnitude of white noise in the sample series. It also demonstrates the ability to detect all three types of trend turning. The method's detection ability increases when the trend turning amplitude increases and decreases when the white noise magnitude increases. The method's ability to detect each turning-point in the samples including multiple ones is relatively independent from whether other nearby turning-points are detected; this means the method applies to the detection of single turning-point the similar way from that of multiple turning-point samples. Its ability to estimate trends in sub-series is generally accurate if the turning-point has been correctly located.

Overall, the RSD t-test is an effective tool for detecting trend turning series, showing three major advantages: It has a high probability of avoiding false acceptances, competently detects all three types of trend turnings, and is applicable to samples with multiple turning-points.

While detection using a large sample size will lead to better detection results, working with small samples is unavoidable in practice. Therefore, we propose a de-noising pretreatment based on Gauss filtering for detection cases using small samples. This relatively simple technique can improve the detection ability for small sample detection.

Finally, this paper discusses the most common random perturbation (i.e., the normal distribution). Although stochastic perturbation based on the normal distribution has wide applicability and is worth examining, it is not necessarily universally applicable. When using the RSD t-test to detecting trend turning series, researchers should also consider the actual stochastic perturbations in their specific data.

References

Abraham, B., & Wei, W. W. S. (1984). Inferences about the parameters of a time series model with changing variance. *Metrika*, *31*(1), 183–194. https://doi.org/10.1007/BF01915199

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Tank, A. M. G. K., et al. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research*, *111*, D05109. https://doi.org/10.1029/2005jd0062 90

Andrews, W. K. D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*, 821–856. https://doi.org/10.2307/2951764

Cheung, K. K. W., Jiang, N. B., Liu, K. S., & Chang, L. T. C. (2015). Interdecadal shift of intense tropical cyclone activity in the Southern Hemisphere. *International Journal of Climatology*, *35*, 1519–1533. https://doi.org/10.1002/joc.4073

Chu, C., & White, H. (1992). A direct test for changing trend. *Journal of Business & Economic Statistics*, *10*(3), 289–299. https://doi.org/10.1080/07350015.1992.10509906

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. *Reviews of Geophysics*, *48*(4), RG4004. https://doi.org/10.1029/2010RG000345

Hsu, D. A., Miller, R. B., & Wichern, D. W. (1974). On the stable Paretian behavior of stock market prices. *Journal of the American Statistical Association*, *69*, 108–113. https://doi.org/10.2307/2285507

Ji, F., Wu, Z., Huang, J., & Chassignet, E. (2014). Evolution of land surface air temperature trend. *Nature Climate Change*, *4*(6), 462–466. https://doi.org/10.1038/nclimate2223

Karl, T. R., Knight, R. W., & Baker, B. (2000). The record breaking global temperatures of 1997 and 1998: Evidence for an increase in the rate of global warming? *Geophysical Research Letters*, *27*(5), 719–722. https://doi.org/10.1029/1999gl010877

Kravtsov, S., & Spannagle, C. (2008). Multidecadal climate variability in observed and modeled surface temperatures. *Journal of Climate*, *21*(5), 1104–1121. https://doi.org/10.1175/2007jcli1874.1

Lee, S., & Park, S. (2001). The cusum of squares test for scale changes in infinite order moving average process. *Scandinavian Journal of Statistics*, *28*, 625–644. https://doi.org/10.1111/1467-9469.00259

Li, J. P., Sun, C., & Jin, F. F. (2013). NAO implicated as a predictor of Northern Hemisphere mean temperature multidecadal variability. *Geophysical Research Letters*, *40*, 5497–5502. https://doi.org/10.1002/2013GL057877

Li, J. P., Zheng, F., Sun, C., Feng, J., & Wang, J. (2019). Pathways of influence of the Northern Hemisphere mid-high latitudes on East Asian climate: A review. *Advances in Atmospheric Sciences*, *36*(9), 902–921. https://doi.org/10.1007/s00376-019-8236-5

Liang, W., Yang, Y. T., Fan, D. M., Guan, H. D., Zhang, T., Long, D., et al. (2015). Analysis of spatial and temporal patterns of net primary production and their climate controls in China from 1982 to 2010. *Agricultural and Forest Meteorology*, *204*, 22–36. https://doi.org/10.1016/j.agrformet.2015.01.015

Liu, R. Q., Jacobi, C., Hoffmann, P., Stober, G., & Merzlyakov, E. G. (2010). A piecewise linear model for detecting climatic trends and their structural changes with application to mesosphere/lower thermosphere winds over Collm, Germany. *Journal of Geophysical Research: Atmospheres*, *115*(D22), 105. https://doi.org/10.1029/2010JD014080

Meehl, G. A., Arblaster, J. M., Fasullo, J. T., Hu, A. X., & Trenberth, K. E. (2011). Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nature Climate Change*, *1*, 360–364. https://doi.org/10.1038/nclimate1229

Meehl, G. A., Chung, C. T. Y., Arblaster, J. M., Holland, M. M., & Bitz, C. M. (2018). Tropical decadal variability and the rate of Arctic sea ice decrease. *Geophysical Research Letters*, *45*(11) 333, 326–311. https://doi.org/10.1029/2018GL079989

Maio, B. Q. (1988). Inference in a model with at most one slope-change point. *Journal of Multivariate Analysis*, *27*(2), 375–391. https://doi.org/10.1016/0047-259X(88)90136-4

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research*, *117*(D8), D08101. https://doi.org/10.1029/2011jd017187

Peng, S. S., Chen, A. P., Xu, L., Cao, C. X., Fang, J. Y., Myneni, R. B., et al. (2011). Recent change of vegetation growth trend in China. *Environmental Research Letters*, *6*(4). https://doi.org/10.1088/1748-9326/6/4/044027

Perron, P., & Zhu, X. (2005). Structural breaks with deterministic and stochastic trends. *Journal of Econometrics*, *129*, 65–119. https://doi.org/10.1016/j.jeconom.2004.09.004

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *J. R. Stat. Soc. C*, *28*(2), 126–135. https://doi.org/10.2307/2346729

Piao, S. L., Wang, X. H., Ciais, P., Zhu, B., Wang, T., & Liu, J. (2011). Changes in satellite-derived vegetation growth trend in temperate and boreal Eurasia from 1982 to 2006. *Global Change Biology*, *17*(10), 3228–3239. https://doi.org/10.1111/j.1365-2486.2011.02419.x

Ploberger, W., & Kramer, W. A. (1996). A trend resistant test for structural change based on the OLS residuals. *Journal of Econometrics*, *70*, 175–186. https://doi.org/10.1016/0304-4076(94)01688-7

Sen, A., & Srivastava, M. S. (1975). On tests for detecting change in mean. *The Annals of Statistics*, *3*, 96–103. Retrieved from. www.jstor.org/stable/2958081

Srivastava, M. S., & Worsley, K. J. (1986). Likelihood ratio tests for a change in the multivariate normal mean. *Journal of the American Statistical Association*, *81*(393), 199–204. https://doi.org/10.2307/2287990

Sun, C., Kucharski, F., Li, J. P., Jin, F. F., Kang, I. S., & Ding, R. Q. (2017). Western tropical Pacific multidecadal variability forced by the Atlantic multidecadal oscillation. *Nature Communications*, *8*(1), 15998. https://doi.org/10.1038/ncomms15998

Sun, C., Kucharski, F., Li, J. P., Wang, K. C., Kang, I. S., Lian, T., et al. (2019a). Spring Aleutian Low weakening and surface cooling trend in northwest North America during recent decades. *Journal of Geophysical Research-Atmospheres*, *124*(22), 12078–12092. https://doi.org/10.1029/2019JD031405

Sun, C., Li, J. P., & Jin, F. F. (2015). A delayed oscillator model for the quasi periodic multidecadal variability of the NAO. *Climate Dynamics*, *45*, 2083–2099. https://doi.org/10.1007/s00382-014-2459-z

Sun, C., Li, J. P., Kucharski, F., Kang, I. S., Jin, F. F., Wang, K. C., et al. (2019b). Recent acceleration of Arabian Sea warming induced by the Atlantic-western Pacific trans-basin multidecadal variability. *Geophysical Research Letters*, *46*, 1662–1671. https://doi.org/10.1029/2018GL081175

Thanasis, V., Efthimia, B. S., & Dimitris, K. (2011). Estimation of linear trend onset in time series. *Simulation Modelling Practice and Theory*, *19*, 1384–1398. https://doi.org/10.1016/j.simpat.2011.02.006

Tomé, A. R., & Miranda, P. M. A. (2004). Piecewise linear fitting and trend changing points of climate parameters. *Geophysical Research Letters*, *31*(2), 207. https://doi.org/10.1029/2003gl019100

Tomé, A. R., & Miranda, P. M. A. (2005). Continuous partial trends and low-frequency oscillations of time series. *Nonlinear Processes in Geophysics, European Geosciences Union (EGU), 2005*, *12*(4), 451–460.

Toms, J. D., & Lesperance, M. L. (2003). Piecewise regression: A tool for identifying ecological thresholds. *Ecology*, *84*(8), 2034–2041. https://doi.org/10.2307/3450028

Tung, K. K., & Zhou, J. S. (2013). Using data to attribute episodes of warming and cooling in instrumental records. *Proceedings of the National Academy of Sciences*, *110*(6), 2058–2063. https://doi.org/10.1073/pnas.1212471110

Vigo, I., Garcia, D., & Chao, B. F. (2005). Change of sea level trend in the Mediterranean and Black seas. *Journal of Marine Research*, *63*(6), 1085–1100. https://doi.org/10.1357/002224005775247607

Vose, R. S., Arndt, D., Banzon, V. F., Easterling, D. R., Gleason, B., Huang, B. Y., et al. (2012). NOAA's merged land–ocean surface temperature analysis. *B. Am. Meteorol. Soc.*, *93*, 1677–1685. https://doi.org/10.1175/BAMS-D-11-00241.1

Wang, B., Huang, F., Wu, Z., Yang, J., Fu, X., & Kikuchi, K. (2009). Multi-scale climate variability of the South China Sea monsoon: A review. *Dynamics of Atmospheres and Oceans*, *47*(1–3), 0–37. https://doi.org/10.1016/j.dynatmoce.2008.09.004

Wang, X., Li, J. P., Sun, C., & Liu, T. (2017). NAO and its relationship with the Northern Hemisphere mean surface temperature in CMIP5 simulations. *Journal of Geophysical Research-Atmospheres*, *122*, 4202–4227. https://doi.org/10.1002/2016JD025979

Wyatt, M. G., & Curry, J. A. (2014). Role for Eurasian Arctic shelf sea ice in a secularly varying hemispheric climate signal during the 20th century. *Climate Dynamics*, *42*(9–10), 2763–2782. https://doi.org/10.1007/s00382-013-1950-2

Wyatt, M. G., & Peters, J. (2012). A secularly varying hemispheric climate-signal propagation previously detected in instrumental and proxy data not detected in CMIP3 data base. *Springerplus*, *1*(1), 1, 68–25. https://doi.org/10.1186/2193-1801-1-68

Xiao, D., & Li, J. P. (2007). Spatial and temporal characteristics of the decadal abrupt changes of global atmosphere-ocean system in the 1970s. *Journal of Geophysical Research*, *112*(D24), D24S22. https://doi.org/10.1029/2007jd008956

Xie, T., Li, J., Sun, C., Ding, R. Q., Wang, K. C., Zhao, C. F., & Feng, J. (2019). NAO implicated as a predictor of the surface air temperature multidecadal variability over East Asia. *Climate Dynamics*, *53*(1-2), 895–905. https://doi.org/10.1007/s00382-019-04624-4

Xing, N., Li, J. P., & Wang, L. N. (2017). Multidecadal trends in large-scale annual mean SATa based on CMIP5 historical simulations and future projections. *Engineering*, *3*(1), 136–143. https://doi.org/10.1016/j.eng.2016.04.011

Yao, S. L., Luo, J. J., Huang, G., & Wang, P. F. (2017). Distinct global warming rates tied to multiple ocean surface temperature changes. *Nature Climate Change*, *7*(7), 486–491. https://doi.org/10.1038/Nclimate3304

Yu, B., & Lin, H. (2018). Coherent changes of wintertime surface air temperatures over North Asia and North America. *Scientific Reports*, *8*(1), 5384. https://doi.org/10.1038/s41598-018-23750-3

Yue, S., Pilon, P., & Cavadias, G. (2002). Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology*, *259*(1–4), 254–271. https://doi.org/10.1016/S0022-1694(01)00594-7

Zhao, S., Li, J. P., & Sun, C. (2016). Decadal variability in the occurrence of wintertime haze in central eastern China tied to the Pacific Decadal Oscillation. *Scientific Reports*, *6*, 27,424. https://doi.org/10.1038/srep27424

Zhao, W., Tian, Z., & Xia, Z. (2010). Ratio test for variance change point in linear process with long memory. *Statistical Papers*, *51*(2), 397–407. https://doi.org/10.1007/s00362-009-0202-3

Zuo, B., Li, J. P., Sun, C., & Zhou, X. (2019). A new statistical method for detecting trend turning. *Theoretical and Applied Climatology*, *138*(1-2), 201–213. https://doi.org/10.1007/s00704-019-02817-9