



HAL
open science

Detecting gradual and abrupt changes in hydrodological records

C. Rougé, Y. Ge, X. Cai

► **To cite this version:**

C. Rougé, Y. Ge, X. Cai. Detecting gradual and abrupt changes in hydrodological records. *Advances in Water Resources*, 2013, 53, p. 33 - p. 44. 10.1016/j.advwatres.2012.09.008 . hal-00913336

HAL Id: hal-00913336

<https://hal.science/hal-00913336>

Submitted on 3 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Detecting Gradual and Abrupt Changes in Hydrological Records

Charles Rougé^{1,2}, Yan Ge¹ and Ximing Cai¹

September 2012

Abstract

Hydrological time series increasingly exhibit non-stationarity, e.g., variables such as precipitation and streamflow values do not maintain a consistent mean over long periods, due to natural and anthropogenic changes. Detecting whether such shifts are gradual or abrupt is a growing concern for water resources planning and management. This paper shows that conventional trend and change-point tests do not adequately enable these two types of change to be distinguished. We propose a method for combining the rank correlations of the Mann-Kendall and Pettitt statistics to extract an indicator whose value determines whether a shift observed in a given time series is gradual or abrupt. This method allows the success rate to be independent of the length of record, and it is validated with Monte-Carlo experiments. The limitations caused by the short and noisy nature of hydroclimatic time series are discussed. As an application, the proposed method provides useful insights on changes in hydroclimatic variables in the United States during 1910–2009 using time series from 1217 stations in the United States Hydroclimatic Data Network (USHCN).

Keywords: Hydrological statistics; Non-stationarity; Trend test; Duration of a change; Change-point test; Climate change

¹Ven-Te Chow Hydrosystems Laboratory, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 N. Mathews Avenue, Urbana, IL 61801, United States

²Irstea, UR LISIC Laboratoire d'ingénierie des systèmes complexes, 9 avenue Blaise Pascal - CS 20085, 63178 Aubière, France

1 Introduction

Changes in climate, land use and water use have caused statistical shifts in the long-term means of hydrological variables. In some cases, these evolutions could lead to a water management crisis, especially since their uncertainty makes it challenging for water managers to adapt to the changes [33]. Detecting causes of these shifts can be complex, as many factors, some of which exhibit nonlinear or chaotic behavior [37, 38], affect hydrological processes jointly and reinforce one another (e.g. [17, 18]). Comprehending their complexity is crucial in river basins facing serious environmental problems, such as the Aral sea in Central Asia [11, 3, 4], and the Yellow River basin in China [5, 10]. For example, the long-term gradual trends [10, 6] and step changes [54] identified in the headwaters of the Yellow River demonstrate the impacts of human interferences on streamflow since climatic variations alone could not explain the shifts.

Thus, descriptive statistics and change detection methods can be used to understand the causes of historic changes in hydrologic time series. This paper focuses on a particular issue of growing importance – determining whether a change in a hydrological time series is gradual or abrupt [32, 46, 43, 44]. We particularly focus on changes in the mean, which are monotonous throughout the time series, and we define them as “shifts”. The term abrupt is here used for the time series in which the shift takes place in a single short period of time, i.e., a step change. We define gradual as the opposite of abrupt: the gradual shift or trend occurs on a large portion of the time series, or all of it. According to these specifications, a shift can be either gradual or abrupt. Indeed, as noted by Villarini et al. [43], the length of most hydrological time series makes it difficult to retrieve information on more than one change pattern (gradual changes and step changes) at once. In particular, if a trend and a step change are contained in the same time series, it is notoriously difficult to separate them [46]. We therefore propose to label the resulting shift as gradual, even if it is the result of the superposition of several shifts of smaller amplitude. Identifying this characterization can improve the understanding of the causal mechanisms (see e.g., [43, 44]).

There is an extensive literature on hydrological time series analysis, which deals with the issues of the timing and significance of a change. In hydrological time series analysis, many available statistical tests assess the significance of a change [25], e.g., [29, 30, 53, 24, 41] for streamflow; [20, 36] for rainfall; [28, 13, 14] for temperature. However, these cited trend detection methods are not designed to detect the duration or timing of changes. To obtain information on the timing of a shift, change-point tests have been performed in association with a trend test, using the various methods, such as Bayesian inference [46, 52], a moving t-test [54], and the nonparametric Pettitt test [42]. Usually, the trend test is performed first to diagnose whether there is a statistically significant change in the entire time series. Following that, the change-point test is applied to identifying the date(s) at which the major change occurs. The identified dates show a shift in the mean where they can be related to other contemporaneous events, such as a concomitant shift in other climatic variables or major human interventions (such as the building of a dam for streamflow regulation). Change-point tests are also performed separately from trend tests to provide information on both the significance and timing of change (e.g. in [26, 2, 34, 51, 50]). However, there is no mean for those studies to differentiate between abrupt (a step change) and gradual (a trend) changes in a rigorous and consistent way, unless the nature of the shift is visually obvious.

Using these statistical tests requires a prior knowledge of the duration of a change. When dealing with shifts, Kundzewicz and Robson [25] recommended the use of a trend test to analyze a gradual change or a change-point test to analyze a step change. The efficiency of trend tests is commonly tested on linear trend series (e.g., [47]), while that of change-point test is for time series featuring an abrupt change (e.g. [51]). Even cutting-edge time series modeling techniques, such as GAMLSS (Generalized Additive Models for Location, Scale and Shape), require the modeler to decide whether to examine step changes prior to modeling [43] or to assume that all changes are the result of gradual processes [45].

Furthermore, McCabe and Wolock [32] warned that the outcome of a test could lead to hasty conclusions about the duration of change, stated as follows “*previous research has not demonstrated, however, whether commonly used statistical tests for trends can distinguish a gradual monotonic change from an abrupt ‘step’ change*”. By performing a trend test on a moving temporal window, they detected a step change in streamflow in the continental United States around 1970. Another challenge associated with gradual changes is determining whether such changes should be extrapolated beyond the end of a time series [32, 43]. Furthermore, one cannot easily validate the selection of a particular extrapolation model [22].

A recent study by Villarini et al. [43] further questioned the assumption that a statistically significant trend test is an indication of the existence of a gradual change, and proposed a novel method that enables one to discriminate between gradual and abrupt changes in short time series (under a hundred years) featuring high variability. Paradoxically, this method still fundamentally relies on the skillful interpretation of trend and change-point test results. The present work, however, contends that tests used to detect change are not adequate to perform a different yet specific task to decide whether this change is gradual or abrupt. It also contends that there is to date no proper statistical test that performs that task. Instead, it proposes a novel method that derives an indicator whose value allows for the shift to be classified as gradual or abrupt using a criterion that is unrelated to the length of the record. This indicator is extracted directly from the rank correlations used by the Mann-Kendall trend test [31, 21], which is widely recognized as one of the most popular trend tests in the field [25, 43].

The rest of the paper is organized as follows: Section 2 introduces the MK and Pettitt tests that have been used previously to characterize change in hydrologic time series. It justifies that they alone can only determine the statistical significance of a change but cannot distinguish between step changes and trends. Section 3 shows that using rank correlations from the statistics of both tests can retrieve valuable information that the direct outcomes of the tests do not convey. This leads to a method that can be used to determine if change in a given time series is gradual or abrupt. Next, Section 4 tests this method and further discusses some practical aspects surrounding the use of these findings. Following this theoretical analysis, a case-study application is presented in Section 5. Finally, Section 6 discusses and summarizes the conclusions from this research.

2 Challenges in identifying patterns using statistical tests

This section introduces the statistical methods used in the present paper, and then discusses how a trend test and a change-point test can be compared when investigating the type of change in a time

series. For the purpose of the comparison, we use the two statistical methods on both a monotonic linear trend of constant slope and a step change. All the tests in this study are conducted through Monte-Carlo experiments, featuring over 10,000 runs (unless stated otherwise) and an independent, identically distributed standard normal noise, using time series with a length of $n = 100$.

2.1 Statistical tests

The tests used in this paper are non-parametric, and therefore do not require normally distributed data. These tests are well-suited for hydrology, a field in which many time series are skewed or contain outliers. Throughout this paper, two types of tests are used: a trend test (Mann-Kendall) and a change-point test (Pettitt). They both measure change through rank correlations. For a more thorough description of these tests, one can refer to Kundzewicz and Robson [25] and Villarini et al. [43].

2.1.1 The Mann-Kendall (MK) trend test

The MK test [31, 21] tests the null hypothesis H_0 that there is no change in the median of the independent observations. For a time-series of n observations, the MK statistics uses the sign (sgn) of the difference of two distinct observations (i and j), which is defined as:

$$\text{sgn}(x_j - x_i) = \begin{cases} 1 & \text{if } x_j > x_i \\ 0 & \text{if } x_j = x_i \\ -1 & \text{if } x_j < x_i \end{cases} \quad (1)$$

The MK statistic (S) is a measure of change over the whole time series:

$$S = \sum_{1 \leq i < j \leq n} \text{sgn}(x_j - x_i) \quad (2)$$

For $n \geq 8$, if H_0 holds, then S is approximately normally distributed with the following mean and variance:

$$E(S) = 0 \quad (3)$$

$$V(S) = \frac{1}{18} \left(n(n-1)(2n+5) - \sum_m t_m m(m-1)(2m+5) \right) \quad (4)$$

where t_m is the number of ties of extent m (there is a tie when $x_i = x_j$, and the number of realizations that are equal give the extent of the tie). More specifically, the following variable Z follows a standard normal distribution:

$$Z = \frac{S - \text{sgn}(S)}{V(S)} \quad (5)$$

Thus, H_0 is rejected at the level of significance α when $|Z| > z_{1-\alpha/2}$, the $(1 - \alpha/2)$ quantile of the standard normal distribution.

2.1.2 The Pettitt change point test

The Pettitt test [35] tests the null hypothesis H_0 that, when arbitrarily splitting the sample in two, there is no change in the median. It produces a rank-based comparison between the observations situated before and after a date τ through the so-called Pettitt statistics $k(\tau)$, which can be computed as follows:

$$k(\tau) = \sum_{i=1}^{\tau} \sum_{j=\tau+1}^n \text{sgn}(x_j - x_i) \quad (6)$$

We are interested in the time for which $k(\tau)$ has the greatest absolute value, since it is a measure of the change that takes place at date τ . Therefore, let us introduce the two following quantities:

$$T = \arg \max_{1 \leq \tau \leq n} (|k(\tau)|) \quad (7)$$

$$K = \max_{1 \leq \tau \leq n} (|k(\tau)|) \quad (8)$$

K is the final Pettitt statistics, and the date T is the date of change. The significance probability associated with the rejection of H_0 is approximated by:

$$p \approx 2 \exp\left(\frac{-6K^2}{n^3 + n^2}\right) \quad (9)$$

with an accuracy within 1% for $p < 0.5$. However, Eq. (9) does not give any confidence that T is the date of a shift, as it merely reports the greatest likelihood of a change in the median.

2.2 The difficulty of pattern change recognition

To understand the limits of using significance tests to detect change patterns, we examine the respective capabilities of the Pettitt change-point test and the MK trend test on a linear trend, a centered step (change at $\tau = 50$) and a non-centered step (change at $\tau = 33$). For each time series, upon which a standard normal white noise ε is superimposed, the mean shifts from 0 to 1:

$$f_1(t) = \frac{1}{99}(t - 1) + p\varepsilon(t) \quad (10)$$

$$f_2(t) = p\varepsilon(t) + \begin{cases} 0 & \text{if } t \leq 50 \\ 1 & \text{if } t > 50 \end{cases} \quad (11)$$

$$f_3(t) = p\varepsilon(t) + \begin{cases} 0 & \text{if } t \leq 33 \\ 1 & \text{if } t > 33 \end{cases} \quad (12)$$

Thus, p corresponds to the standard deviation of the noise, and also happens to be the coefficient of variation (CV) of the time series. Tests are conducted using a Monte-Carlo simulation for each value of p , starting at 0.1 and increasing by a step of 0.1 until the total detection power of Pettitt and MK taken together goes below 0.45 (for $p = 1.6$). That value of the CV will serve as an upper boundary for subsequent simulations, because the question whether a shift is gradual or abrupt is only relevant

when this shift can be consistently detected in the first place. Following [51] and [47], we use the term power to quantify the rejection rate of the null hypothesis of a test when the alternative hypothesis is true.

The results with a 5% significance level (Figure 1) show that the MK and Pettitt tests both detect change in over 80% of the cases for a CV of 1 and even 60% of the cases for CV= 1.6, regardless of the gradual or abrupt nature of the shift. In other words, this example shows that the trend test detects changes other than trends while the change-point test detects change points even when the change is a linear trend. In addition, both tests are more powerful in detecting a step change, even when it is not centered, than a linear trend. This not only poses a terminology problem but, more importantly, confirms a methodological problem (suspected by McCabe and Wolock [32] or Villarini et al. [43]), because it contradicts the assumed association between test type and duration of change.

Even recent developments that are aware of this conceptual problem do not really surmount it. Villarini et al. [43] contended that the presence of a change point was determined by a change-point test. Then, gradual changes were determined by detecting those with a trend test, applied to 1) whole time series if the change-point test did not detect change, or 2) separately the parts before and after the identified change-point. Thus, this method still relies on the assumption that the detection of change by a change-point test implies the existence of a step, and posits a trend if a trend test and a change point test yield contradictory results regarding the rejection of the null hypothesis of no change (case 1).

This paper proposes to develop a specific method of surmounting this conceptual problem by using the tests' rank correlations, which contain information on the time series, instead of the test results themselves, which use only a portion of that information.

3 A method to detect abrupt and gradual changes

This section explains the derivations that lead to our proposed method. First, we provide a way to relate the MK and Pettitt statistics, and then introduce a generalized Pettitt “period” statistic, before describing a novel method for distinguishing gradual and abrupt changes.

3.1 Common information of the various change patterns

To develop a method to distinguish gradual trends from abrupt step changes through the combination of the MK and Pettitt statistics, we examine the MK statistics (S) not only for the purpose of determining if there is a change in the mean of the time series, but also to extract information about the change from the rank correlation pairs $(x_j - x_i)$. Based on the definition of S in Eq. (2), we have $\forall \tau, 1 \leq \tau \leq n$ (where n is the length of the time-series), the following decompositions of S :

$$\begin{aligned}
S &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n \operatorname{sgn}(x_j - x_i) \\
&= \sum_{i=1}^{\tau} \sum_{j=i+1}^n \operatorname{sgn}(x_j - x_i) + \sum_{i=\tau+1}^{n-1} \sum_{j=i+1}^n \operatorname{sgn}(x_j - x_i) \\
&= \sum_{i=1}^{\tau-1} \sum_{j=i+1}^{\tau} \operatorname{sgn}(x_j - x_i) + \sum_{i=1}^{\tau} \sum_{j=\tau+1}^n \operatorname{sgn}(x_j - x_i) + \sum_{i=\tau+1}^{n-1} \sum_{j=i+1}^n \operatorname{sgn}(x_j - x_i)
\end{aligned} \tag{13}$$

where the middle item can be recognized as the quantity introduced by Eq. (6), taking place on date τ . This directly leads to the following relationship between the MK and Pettitt statistics:

$$S = \sum_{1 \leq i < j \leq \tau} \operatorname{sgn}(x_j - x_i) + k(\tau) + \sum_{\tau+1 \leq i < j \leq n} \operatorname{sgn}(x_j - x_i) \tag{14}$$

Introducing the MK statistics as computed over a subset of a time series of length n , we have $\forall(p, q)$, $1 \leq p < q \leq n$:

$$S(p, q) = \sum_{p \leq i < j \leq q} \operatorname{sgn}(x_j - x_i) \tag{15}$$

This allows Eq. (14) to be written as:

$$S = S(1, \tau) + k(\tau) + S(\tau + 1, n) \tag{16}$$

Abrupt and gradual changes can be characterized by expressing Eq. (16) in a matrix form. Defining the elements of the square matrix A of size $n \times n$ by:

$$a_{ij} = \begin{cases} \operatorname{sgn}(x_j - x_i) & \text{if } j > i \\ 0 & \text{if } j \leq i \end{cases} \tag{17}$$

A is an upper triangular matrix that has zeroes on the diagonal. Using Eq. (2) and (17), the MK statistics is merely computed as the sum of the terms in the upper triangle. The decomposition by Eqs. (16) can be graphically demonstrated by Figure 2.a, which illustrates how the difference between a gradual trend and a step change can be conveyed by the MK and Pettitt statistics.

For a step change, $S(1, \tau)$ and $S(\tau + 1, n)$ are not supposed to imply any relevant change if $\tau = T$, because there is no change in the first τ points nor in the last $n - \tau$ points. Thus we have:

$$E[S(1, n) - k(\tau)] = 0 \tag{18}$$

However, for a trend, there is still change to be measured in the first τ points and in the last $n - \tau$ points. We have:

$$E[S(1, n) - k(\tau)] = a \times \operatorname{sgn}(S) \tag{19}$$

Where $a > 0$, and it depends on three parameters: the length of the time series, the amplitude of the change, and that of the noise in the time series. Thus, Eqs. (18) and (19) provide a means of comparing between abrupt and gradual changes.

3.2 Generalizing the Pettitt statistics

Since the comparison between a trend and step change can rely on the comparison between $S(1, n)$ and $k(\tau)$, in this section we modify and extend the Pettitt statistic, now noted $k(\tau, d)$ so that it compares the median of the values after and before an interval of time of length d . Then, we may have $E[S - k] = 0$ when considering the actual period of change (even if it does not behave as the functions f_1 or f_2 introduced in Section 2). Thus, Pettitt period statistics can be introduced as follows and graphically represented on Figure 2.b, again using the matrix A :

$$\begin{aligned}
 k(\tau, d) &= \sum_{i=1}^{\tau+d} \sum_{j=\tau+1}^n a_{ij} \\
 &= \sum_{i=1}^{\tau} \sum_{j=\tau+d+1}^n \text{sgn}(x_j - x_i) + \sum_{i=\tau+1}^{\tau+d-1} \sum_{j=i+1}^{\tau+d} \text{sgn}(x_j - x_i) \\
 &\quad + \sum_{i=1}^{\tau} \sum_{j=\tau+1}^{\tau+d} \text{sgn}(x_j - x_i) + \sum_{i=\tau+1}^{\tau+d} \sum_{j=\tau+d+1}^n \text{sgn}(x_j - x_i)
 \end{aligned} \tag{20}$$

The first item of the decomposition (designated by (i) on Figure 2.b) of $k(\tau, d)$ compares a period *before* τ (from 1 to τ) to a period *after* $\tau + d$ (from $\tau + d + 1$ to n). In addition, based on equation Eq. (15), the second item of Eq. (20) can be understood as being $S(\tau + 1, \tau + d)$ (ii). It is a rank-based measure of the increase (or the decrease) during the considered period, during which the measurements are ranked in an increasing (or decreasing) order for an increase (or decrease) change. Finally the last two items, respectively, compare the periods *before* (iii) and *after* (iv) to the period under consideration (from $\tau + 1$ to n). For any duration d , we can use $k(\tau, d)$ in the same way as $k(\tau)$ defined by Eq. (6) to introduce a modified Pettitt statistic $K(d) = \max_{\tau}(|k(\tau, d)|)$. This is analogous to Eq. (8), except that it measures change based on a period of length d rather than a single point. For any d , the associated date of change T_c is then given as:

$$T_c = \arg \max\{|k(\tau, d)|\} \tag{21}$$

In this new configuration, the generalized Pettitt statistic is still related to the MK statistic by:

$$S = S(1, T_c) + k(T_c, d) + S(T_c + d + 1, n) \tag{22}$$

3.3 Change duration detection method

Now that we have derived the Pettitt period statistic, let us suppose that, for a given time series, a change in the median occurs between the dates T_c and $T_c + d$. Then, since the expected values of $S(1, T_c)$ and $S(T_c + d + 1, n)$ are both zero, we can write a formula that is analogous to Eq. (19):

$$E[S(1, n) - k(T_c, d)] = 0 \tag{23}$$

The pair (T_c, d) cannot be used to identify the exact timing and duration of change since Eq. (23) is only true on average. In fact, as with the variance of S , that of $[S(1, n) - k(T_c, d)]$ is roughly proportional to $n^{3/2}$. Let us formally define d_c as the minimum duration for which $|S| \leq |k(T_c, d)|$:

$$d_c = \arg \min_d \{d/|k(T_c, d)| > |S|\} \quad (24)$$

by which d_c cannot be interpreted as a meaningful physical duration, such as the exact duration of change. Rather, it can only be used as an indication of whether the change is gradual or abrupt, because Eq. (23) shows that, on average, d_c increases when the actual duration of change increases. Thus, a relevant criterion to distinguish between abrupt and gradual changes can be to compare d_c to a threshold duration D . If $d_c > D$, the change is considered gradual, otherwise it is considered abrupt. D can be arbitrarily chosen, and has no physical meaning. Yet, scaling D with the length n of the time series under study guarantees that the outcome of the proposed method will not be dependent on n .

The algorithm depicted in Figure 3 diagrams the process for discriminating between gradual and abrupt changes. d is incremented, starting with value $d = 1$ and stops when the condition given by Eq. (24) is met. This ensures that the first d for which $|k(T_c, d)| > |S|$ is d_c as defined in Eq. (24). The comparison between this d_c and D allows for the change to be characterized as gradual or abrupt. One should note that, in the absence of noise, the d_c and T_c values found with the algorithm would, in fact, be the exact dates at which a shift begins and ends, because we have $S(1, t) = S(t + d + 1, n) = 0$.

4 Discussion

4.1 Validation: comparison with existing methods

The method presented in the above section cannot compute any confidence level for d_c being larger, or smaller, than D for a gradual change. For a method that aims at distinguishing gradual from abrupt, the first step is to recognize a shift where the mean changes at every date in the time series as gradual, and a shift occurring only at one point as abrupt. A linear trend and a step change are common respective examples of such shifts. Hence, validation is conducted through Monte-Carlo simulations using time series of type f_1 for gradual changes and f_2 and f_3 for abrupt changes, as defined in Section 2.1. A large number (10,000) runs are used for each value of the CV (comprised between 0.1 and 1.6), where the noise is described by two types of probability distribution: the standard normal distribution and the gamma distribution (also known as Pearson type III distribution, to assess the impact of skewness), with unit parameters for both scale and shape so as to keep a unit standard deviation. The latter is an example of a skewed distribution with extreme values. Based on the validation results, we compare our method with the method developed by Villarini et al. [43], which, to our knowledge, is the only other one that attempts to systematically identify abrupt and gradual changes in a short and noisy time series.

Under the convention we chose to describe a shift as abrupt or gradual, we will consider that the method that Villarini et al. proposed leads to label a shift as abrupt when the Pettitt test detects a change point, but the MK test applied to the portions of the time series before and after this change point does not. In all other cases, the change will be treated as gradual. The most relevant range for the results is when the CV is comprised between 0.5 and 1. Below 0.5, the median values at the beginning and at the end of the time series are separated by more than two standard deviations of

the noise. Under this condition, one can assume the nature of the shift to be detectable. With the values above 1, the standard deviation outweighs the amplitude of the shift and detection of change itself becomes an issue.

Results for normal and gamma-distributed noise are displayed in Figures 4 and 5, respectively, with the values of D/n ranging from 0.1 to 0.4. Success rates are defined by the number of times a change is correctly detected by the statistical tests. Thus, the complementary value of the success rate (1–success rate) is the chance of mistaking a trend for a step, and vice versa. Comparisons of our method with that from Villarini et al. for $n = 40, 60, 80$ and 120 with the same values of D/n yields similar results for both methods, which justifies the scaling of D with n . The choice of D/n matters, as a bigger D/n results in more changes being characterized as abrupt. This means that the higher the value of D/n , the better chance to detect an abrupt change when it is present, but the greater the risk of false detection when a gradual change is present. Thus one has to consider this trade-off, illustrated by the results from Figures 4 and 5, when using this method.

Results show that for both methods, noise has far less impact on the detection of gradual changes when $\varepsilon \sim \Gamma(1; 1)$. The method from Villarini et al. [43] detects gradual changes with low levels of noise, but its performance decreases sharply when $CV > 0.3$ (normal) or $CV > 0.5$ (gamma). The method detects step changes well but incorrectly identifies over two thirds of trends as step changes when $CV > 0.4$ (normal) or $CV > 0.7$ (gamma). When the CV is smaller than one, the risk of false detection of gradual changes does not disappear, making it risky to use this method as a robust indicator of the presence of trends. The issue of detecting existing trends results from the short length of record and the level of noise, which are two factors well known for compromising the effectiveness of both parametric and non-parametric trend tests [47]. In contrast, our proposed method detects gradual changes better since the use of the scaled value (D/n) largely eliminates the dependence of the success rate on the length of record. Even though the success rate of detecting gradual changes decreases when the CV increases, simulation results show that for $D/n = 0.3$ and lower, the method constitutes a considerable improvement in the most relevant range of the CV. The success rate of the method by Villarini et al. slightly increases for gradual changes when $CV > 1.2$ (after a sharp decline when $0.4 < CV < 1.2$). However, this is only an artifact of the specific statistical tests being used, as the MK test becomes more powerful than the Pettitt test when the CV exceeds 1.2 (Figure 1).

As for abrupt changes, both methods produce good results and have similar success rates, which fall around 90% for $D/n = 0.3$, and the distribution type seems to have little effect on the detection of abrupt changes. The location of the shift is shown to slightly affect the proposed method, but not that of Villarini et al. [43]. Thus, the comparison to the only existing method suggests that the choice of D/n around 0.3 is optimal because the large improvement in gradual change detection generally outweighs a slightly lower rate of abrupt change detection (if any exists). Very low values of D/n , such as 0.1, perform poorly on step changes, and the same happens, to a lesser extent, for gradual changes when D/n is over 0.3. Indeed the power of detection of gradual changes with $D/n = 0.3$ is larger than $D/n = 0.4$ by over a 10% margin, for $0.3 < CV < 1.2$ with a normally distributed noise and for $CV > 0.5$ with a gamma-distributed noise.

4.2 Remaining challenges in the exploration of change duration

Even though the proposed method was shown to be the best available for the time series given above, results show that it does not give perfect results, especially when it comes to detect gradual changes. This demonstrates the difficulty of distinguishing gradual shifts from abrupt ones, and it shows two important practical consequences. Firstly, according to the results, if the whole time-series is considered, a shift that occurs over a fraction $0 < \lambda < 1$ of a time series is detected as gradual when $d_c > D$. While if the λn points over which the shift occurs are considered in isolation, then we only need $d_c > \lambda D$. This is due to the fact that the outcome hinges on the scaled value D/n , and suggests that the best way to assess whether there is a gradual change on a given part of a time series (for instance, its first half) is to apply the proposed method to the shortened record. However, the development of more specific methods to deal with the existence of gradual changes on a given part of a time series is beyond the scope of this study. Yet, this remark also shows that even though D has no physical meaning, the idea of using the scaled value D/n fosters the idea that the notions of gradual and abrupt depend on the timescales one is looking at. In a dramatic fashion, "abrupt" can refer to millenia if the record spans 800,000 years [1].

Secondly, this difficulty justifies the choice to only determine whether changes in the mean occur over a long period of time or over a short one, considering for instance both linear trends and some succeeding step changes to be gradual changes. Tests have been carried out for a change pattern with both a trend and a step, under the form of $f_1 + f_2$ (see Figure 6). Using the method proposed in this paper, the results are remarkably similar to those obtained with f_1 alone. The detection of gradual changes is improved by using Villarini's method when the CV is below 1, but the success rate decreases if the CV is above 1. Indeed, the artifact that causes the success rate to increase for a time series in the form of f_1 disappears if the total amplitude of the shift increases. This is the case that is observed here.

Another noteworthy issue is spatial and temporal correlation. The latter, also called serial correlation, has been widely debated in the analysis of hydrological data since Lettenmaier et al. [27]. Serial correlation interferes with the statistical significance of patterns discovered by the proposed method. Yet, Yue et al. [48] and Yue and Wang [49] showed how, in the case of a linear trend, the removal of the change pattern prior to the computation of the serial correlation coefficients improves the accuracy of the first-rank correlation estimate. We provide an upgrade to the removal of a linear trend before applying the MK test to any given time series [48]. In our method, if the change is gradual, one then removes a linear slope similar as what the previous studies did; while if it is abrupt, one can locate the step change (e.g. with the Pettit test) and remove it. Once the residuals are obtained, a variety of modifications to statistical tests, such as MK's [16, 48, 49], can be made to account for serial correlation when it becomes significant. Yet, the removal of the change pattern should not be used to assess how the CV impacts the results. Indeed, the value of the CV hinges on the outcome of the proposed method, so that it cannot be used as a parameter that influences the result.

Spatial correlation in the testing of changes has received some attention, e.g. [7, 15]). Data from close geographic locations may contain considerable common information due to the spatial correlation, although it has been argued that a common timing of shifts, rather than identical year-to-year variations, could sometimes account for a large fraction of spatial correlation [36]. Combining

the method proposed here with a way of separating spatial correlation from concomitant shifts in a regional database could be a natural follow-up to this paper, though this is out of the scope of the present work.

Finally, as mentioned before, the robustness of a signal detected cannot be inferred from the examination of a single time series. For this reason, we recommend the methodology presented in this paper as an exploratory tool for large datasets. Preliminary data exploration through statistical analysis can provide insights that guide model development (e.g., [23, 45]). Furthermore, discovering temporal patterns at different locations can thus lead to useful insights into the spatial repartition of different change patterns, as illustrated in the following case study.

5 Application

We use the United States Historical Climatology Network (USHCN) Version 2 Serial Monthly Dataset for the continental United States (U.S.), the most recent update to a dataset designed to understand contemporary past changes in hydroclimatology (e.g., [19, 9]), which has been extensively used in recent years (e.g., [8, 12, 13, 14, 20, 28, 39]). Annual total precipitation and annual average temperature time series for all 1217 stations with no missing data were prepared from the monthly precipitation and temperature data, respectively, for each of the stations. We chose 1910 – 2009 as the 100-year study period.

Both the method described in Section 3 and the one advocated by Villarini et al. are applied. The MK and Pettitt tests are used with a significance level of 5%. We also use the results from the method to detrend the time series before computing their autocorrelation structure. Serial correlations are insignificant for most of the time series (Table 1). When they are significant, we use the MK test from Yue et al. [48] if there is only a lag-1 autocorrelation, while we employ the one by Hamed and Rao [16] for all other cases. As explained in Section 4.2, we do not recommend to use the residuals as a parameter to evaluate how the values of the CV for different time series influence the results. To better illustrate the impact of the choice of the threshold value D/n , we pick both $D/n = 0.2$ and $D/n = 0.3$ to discriminate between abrupt and gradual change. For each of these two values, we thus provide results for each site and for both precipitation and temperature with regard to the following questions: whether there is change in the median; if there is, whether it is upward or downward, and whether it is gradual or abrupt. Figures 7 and 8 show the spatial distribution of annual total precipitation and average temperature trends, respectively, while Table 2 summarizes them and compares them with the results from the method of Villarini et al. [43].

Let us first explore the difference between the two methods, with an emphasis on how they address the detection of gradual changes. For precipitation, the number of gradual changes detected through Villarini’s method is close to the result from the method introduced here. However, in 92 out of 101 cases, the detection of gradual changes by Villarini’s method is due to the MK test rejecting the null hypothesis of no change (Table 3), while the Pettitt test cannot reject this same hypothesis. Thus, most of these detected gradual changes can be linked to what can be seen as a theoretical weakness of that method, as it relies on two statistical tests drawing opposite conclusions on the validity of the null hypothesis of no change. Figure 9 illustrates this for a time series where only the MK test detects

change, and for which the abrupt change reported by our framework is a perfectly valid model. It also displays a case where both MK and Pettitt tests detect change, and where only our method (even for $D/n = 0.3$) detects the apparent gradual nature of this change.

For temperatures however, the results from Table 2 are widely different between both methods. Much like in Figures 4 to 6, our method detects 147 (for $D/n = 0.3$) or 175 ($D/n = 0.2$) more gradual changes than the one by Villarini et al., which labels these shifts as abrupt. An example of this is provided by 10, where the gradual change is apparent, as only the framework proposed in this paper detects it, even for $D/n = 0.3$. These results are in line with the results from Figures 4 and 5, which point to a more effective detection of gradual changes with our method. In the end, results suggest a wide difference in the detection of gradual changes between the two methods. In the following we focus on the results provided by our method.

This study corroborates the presence of increasing precipitation that many other studies have detected across large portions of the continental U.S. during the twentieth century (e.g., [8, 12, 13, 14, 20, 28, 39]). In addition, we find that abrupt increases appear more widely than gradual ones, although both are present over large areas in the eastern U.S., especially the Mississippi River basin and the northeastern U.S. Around 40% of gradual increases detected with $D/n = 0.2$ are not robust to the rise of the value of D/n to 0.3, which corroborates this predominance of abrupt increases. This proves that some metrics, such as the average rate of change, which one would strongly associate with linear, uniform increases, can be misleading when describing precipitation increases in the U.S. Moreover, any abrupt changes in precipitation can be dated around 1970, which corresponds to the date for which McCabe and Wolock [32] identified a step increase in the mean annual flow in the eastern U.S.. However, about one-third of the increases identified East of the Rocky Mountains are gradual, especially in the Great Plains, the Midwest (especially in Nebraska, Minnesota and northern Illinois and Indiana) and New England. These results suggest a more complex reality than previously thought.

As for temperature (Figure 8 and Table 2), we find that almost two-thirds of the reported increases are gradual, and in most cases this diagnosis is robust to an increase in the value of D/n . Most regions of the country, with the exception of the Southeast, exhibit increasing annual average temperature trends. In most of these areas, especially in a belt in the northern U.S. from the state of Wisconsin to the state of Washington, these gradual increases are intertwined with step increases, which temporally occurred around the 1970s. This pattern of increases reflects the general evolution of global temperature during the 20th century [40]. However, as reported in the literature [8, 14, 28], the annual average temperature had decreased over 1910 – 2009 throughout the Southeastern U.S. The method presented in this paper, in addition, characterizes this downward change as an abrupt change that took place in the 1950s.

6 Conclusions

This study attempts to combine two rank-based statistical tests, the MK trend test and the Pettitt change-point test, to extract information that can be used to decide whether a shift observed in a given time series is gradual or abrupt. The method goes beyond ordinary uses of these tests as tools

for detecting significant changes in a time series. We show how the statistics from both the trend and change-point tests could be related and articulated together via a decomposition of the Mann-Kendall test statistic and the introduction of a modified Pettitt statistic to retrieve an indicator, which provides a tool for determining whether a change in a short and noisy time series is gradual or abrupt. It proposes a threshold that can be scaled by the length of the time series, so that only the amplitudes of the change and the noise can affect its outcome. The method is shown to outperform the recent pioneering development by Villarini et al. [43].

However, extracting more effective information from short and noisy time series, such as hydroclimatic ones, still remains a challenge, as highlighted by the limitations of the proposed method in this study. This paper has addressed the issue by removing the conceptual hurdle of using the same method for detecting a shift and describing it once it has been detected; however, more sophisticated statistical tools, possibly new tests, have yet to be developed to fully resolve the issue.

An application of this method to real data shows the importance that a broad detection of the duration of change can have in a spatiotemporal analysis of change. Observations based on the application results have clarified some hypotheses regarding climate change and variability in the continental U.S. For instance, there is a gradual increase in annual average temperature in most of the country in the twentieth century, except for the southeastern U.S. where there is a step decrease in the 1950s. It is crucial to advance statistical change detection and characterization methods, such as the one presented in this paper, in an era in which many regions of the world are experiencing climate change [33].

Acknowledgments

This study was financially supported by U.S. National Science Foundation (NSF) projects, CMMI 0825654 and CBET 0747276. Special thanks to Jory Hecht and Mohamad Hejazi for their insightful reviews and comments on earlier versions of this manuscript.

References

- [1] S. Barker, C. Knorr, R. L. Edwards, F. Parrenin, A. E. Putnam, L. C. Skinner, E. Wolff, and M. Ziegler. 800,000 years of abrupt climate variability. *Science*, 334:347–351, 2011.
- [2] T. A. Buishand. Some methods for testing the homogeneity of rainfall records. *J. Hydrol.*, 58:11–27, 1982.
- [3] X. Cai, D. McKinney, and L. S. Lasdon. A framework for sustainability analysis in water resources management and application to the syrdarya basin. *Water Resour. Res.*, 38(6), 2002.
- [4] X. Cai, D. McKinney, and L. S. Lasdon. Integrated hydrologic-agronomic-economic model for river basin management. *J. Water Resour. Plann. Manage.*, 129(1):4–17, 2003.
- [5] X. Cai and M. W. Rosegrant. Optional water development strategies for the yellow river basin: Balancing agricultural and ecological water demands. *Water Resour. Res.*, 40, 2004.
- [6] Z. Cong, D. Yang, B. Gao, H. Yang, and H. Hu. Hydrological trend analysis in the yellow river basin using a distributed hydrological model. *Water Resour. Res.*, 45, 2009.
- [7] E. M. Douglas, R. M. Vogel, and C. N. Kroll. Trends in floods and low flows in the united states: impact of spatial correlation. *Journal of Hydrology*, 240:90–105, 2000.
- [8] D. R. Easterling. Observed climate change and transportation. Paper presented at the Federal Research Partnership Workshop, Brookings Institution, Washington, DC, October 2002.
- [9] D. R. Easterling, T. R. Karl, E.H. Mason, P. Y. Hughes, and D. P. Bowman. United states historical climatology network (u.s. hcn) monthly temperature and precipitation data. ornl/cdiac-87. Technical report, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, 1996.
- [10] G. Fu, S. Chen, C. Liu, and D. Shepard. Hydro-climatic trends of the yellow river basin for the last 50 years. *Climatic Change*, 65:149–178, 2004.
- [11] M. H. Glantz, editor. *Creeping Environmental Problems and Sustainable Development in the Aral Sea Basin*. Cambridge University Press, Cambridge, 1999.
- [12] P. Y. Groisman and D. R. Easterling. Variability and trends of precipitation and snowfall over the united states and canada. *J. Climate*, 7:184–205, 1994.
- [13] P. Y. Groisman, R. W. Knight, and T. R. Karl. Heavy precipitation and high streamflow in the contiguous united states: Trends in the twentieth century. *Bull. Am. Meteorol. Soc.*, 82(2):219–246, 2001.
- [14] P. Y. Groisman, R. W. Knight, T. R. Karl, D. R. Easterling, B. Sun, and J. H. Lawrimore. Contemporary changes of the hydrological cycle over the contiguous united states: Trends derived from in situ observations. *J. Hydrometeorol.*, 5:64–85, 2004.

- [15] Hamed. The distribution of kendall's tau for testing the significance of cross-correlation in persistent data. *Hydrolog. Sci. J.*, 56(5):841–853, 2011.
- [16] K. H. Hamed and A. R. Rao. A modified mann-kendall trend test for autocorrelated data. *J. Hydrol.*, 204:182–196, 1998.
- [17] M. I. Hejazi and G. E. Moglen. Regression-based approach to low flow prediction in the maryland piedmont region under joint climate and land use change. *Hydrol. Processes*, 21:1793–1801, 2007.
- [18] M. I. Hejazi and G. E. Moglen. The effect of climate and land use change on flow duration in the maryland piedmont region. *Hydrol. Processes*, 22:4710–4722, 2008.
- [19] T. R. Karl, Jr. C. N. Williams, and F. T. Quinlan. United states historical climatology network (hcn) serial temperature and precipitation data. ornl/cdiac-30, ndp-019/r1. Technical report, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, 1990.
- [20] T. R. Karl and R. W. Knight. Secular trends of precipitation amount, frequency, and intensity in the united states. *Bull. Am. Meteorol. Soc.*, 79(2):231–241, 1998.
- [21] M. G. Kendall, editor. *Rank correlation methods*. C. Griffin, London, 2nd edition, 1975.
- [22] D. Koutsoyiannis. Nonstationarity versus scaling in hydrology. *J. Hydrol.*, 324:239–254, 2006.
- [23] S. Kumar, V. Merwade, J. Kam, and K. Thurner. Streamflow trends in indiana: Effects of long term persistence, precipitation and subsurface drains. *J. Hydrol.*, 374:171–183, 2009.
- [24] Z. W. Kundzewicz, D. Graczyk, T. Maurer, I. Pińskwar, M. Radziejewski, C. Svensson, and M. Szwed. Trend detection in river flow series: 1. annual maximum flow. *Hydrol. Sci. J.*, 50(5):797–810, 2005.
- [25] Z. W. Kundzewicz and A. J. Robson. Change detection in hydrological records: a review of the methodology. *Hydrol. Sci. J.*, 49(1):7–20, 2004.
- [26] T. R. Lazaro. Nonparametric statistical analysis of annual peak flow data from a recently urbanized watershed. *Water Resour. Bull.*, 12(1):101–107, 1976.
- [27] D. P. Lettenmaier. Detection of trends in water quality data from records with dependent observations. *Water Resour. Res.*, 12(5):1037–1046, 1976.
- [28] D. P. Lettenmaier, E. F. Wood, and J. R. Walls. Hydro-climatological trends in the continental united states, 1948-1988. *J. Climate*, 7:586–607, 1994.
- [29] H. F. Lins and J. R. Slack. Streamflow trends in the united states. *Geophys. Res. Lett.*, 26(2):227–230, 1999.
- [30] H. F. Lins and J. R. Slack. Seasonal and regional characteristics of u.s. streamflow trends in the united states from 1940 to 1999. *Phys. Geogr.*, 26:489–501, 2005.

- [31] H.B. Mann. Non-parametric tests against trend. *Econometrica*, 13:245–259, 1945.
- [32] G.J. McCabe and D.M. Wolock. A step increase in streamflow in the conterminous united states. *Geophys. Res. Lett.*, 29(24), 2002.
- [33] P. C. D. Milly, J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer. Stationarity is dead: Whither water management? *Science*, 319(5863):573–574, 2008.
- [34] L. Perreault, M. Haché, M. Slivitzky, and B. Bobée. Detection of changes in precipitation and runoff over eastern canada and u.s. using a bayesian approach. *Stoch. Env. Res. Risk A.*, 13:201–216, 1999.
- [35] A. N. Pettitt. A non-parametric approach to the change point problem. *Appl. Stat.-J. Roy. St. C*, 28(2):126–135, 1979.
- [36] S. C. Pryor, J. A. Howe, and K. E. Kunkel. How spatially coherent and statistically robust are temporal changes in extreme precipitation in the contiguous usa? *Int. J. Climatol.*, 29:31–45, 2009.
- [37] M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, and B. Walker. Catastrophic shifts in ecosystems. *Nature*, 413(6856):591–596, 2001.
- [38] B. Sivakumar. Nonlinear dynamics and chaos in hydrologic systems: latest developments and a look forward. *Stoch. Env. Res. Risk A.*, 23(7):1027–1036, 2008.
- [39] D. Small, S. Islam, and R. M. Vogel. Trends in precipitation and streamflow in the eastern u.s.: Paradox or perception? *Geophys. Res. Lett.*, 33, 2006.
- [40] S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, editors. *IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.
- [41] C. Svensson, Z. W. Kundzewicz, and T. Maurer. Trend detection in river flow series: 2. flood and low-flow index series. *Hydrol. Sci. J.*, 50(5):811–824, 2005.
- [42] R. Tomozeiu, A. Busuioc, V. Marletto, F. Zinoni, and C. Cacciamani. Detection of changes in the summer precipitation time series of the region emilia-romagna, italy. *Theor. Appl. Climatol.*, 67:193–200, 2000.
- [43] G. Villarini, F. Serinaldi, J. A. Smith, and W. F. Krajewski. On the stationarity of annual flood peaks in the continental united states during the 20th century. *Water Resour. Res.*, 45, 2009.
- [44] G. Villarini and J. A. Smith. Flood peak distributions for the eastern united states. *Water Resour. Res.*, 46, 2010.

- [45] G. Villarini, J. A. Smith, F. Serinaldi, J. Bales, P. D. Bates, and W. F. Krajewski. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Adv. Water Resour.*, 32:1255–1266, 2009.
- [46] L. Xiong and S. Guo. Trend test and change-point detection for the annual discharge series of the yangtze river at the yichang hydrological station. *Hydrol. Sci. J.*, 49(1):99–114, 2004.
- [47] S. Yue, P. Pilon, and G. Cavadias. Power of the mann-kendall and spearman’s rho tests for detecting monotonic trends in hydrological series. *J. Hydrol.*, 259:254–271, 2002.
- [48] S. Yue, P. Pilon, B. Phinney, and G. Cavadias. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol. Processes*, 16:1807–1829, 2002.
- [49] S. Yue and C. Y. Wang. Applicability of prewhitening to eliminate the influence of serial correlation on the mann-kendall test. *Water Resour. Res.*, 38(6), 2002.
- [50] S. Yue and C. Y. Wang. The influence of autocorrelation on the mann-whitney test for detecting a shift in median. *Adv. Water Resour.*, 25:325–333, 2002.
- [51] S. Yue and C. Y. Wang. Power of the mann-whitney test for detecting a shift in median or mean of hydro-meteorological data. *Stoch. Env. Res. Risk A.*, 16:307–323, 2002.
- [52] Q. Zhang, C.-Y. Xu, S. Becker, Z.X. Zhang, Y.D. Chen, and M. Coulibaly. Trends and abrupt changes in precipitation maxima in the pearl river basin, china. *Atmos. Sci. Lett.*, 10:132–144, 2009.
- [53] X. Zhang, D. Harvey, W. D. Hogg, and T. R. Yuzyk. Trends in canadian streamflow. *Water Resour. Res.*, 37(4):987–998, 2001.
- [54] H. Zheng, L. Zhang, C. Liu, Q. Shao, and Y. Fukushima. Changes in stream flow regime in headwater catchments of the yellow river basin since the 1950s. *Hydrol. Processes*, 21:886–893, 2007.

List of Tables

1	Impact of serial correlation depending on the detrending method (significance level of 5%). “V” stands for the method from Villarini et al.	20
2	Repartition of the results by category (significance level of 5%). “V” stands for the method from Villarini et al.	20
3	Repartition of the gradual changes detected by the method from Villarini et al.	21

<i>Correlation structure</i>	<i>Precipitation</i>			<i>Temperature</i>		
	<i>D/n = 0.2</i>	<i>D/n = 0.3</i>	<i>V</i>	<i>D/n = 0.2</i>	<i>D/n = 0.3</i>	<i>V</i>
No serial correlation	1084	1084	1067	933	937	915
Lag-1 correlation	68	67	70	187	182	207
Presence of longer lags	65	66	80	97	98	95

Table 1: Impact of serial correlation depending on the detrending method (significance level of 5%). “V” stands for the method from Villarini et al.

<i>Change pattern</i>	<i>Precipitation</i>			<i>Temperature</i>		
	<i>D/n = 0.2</i>	<i>D/n = 0.3</i>	<i>V</i>	<i>D/n = 0.2</i>	<i>D/n = 0.3</i>	<i>V</i>
No significant change	737	737	740	363	362	362
Abrupt increase	316	363	347	270	299	446
Gradual increase	124	77	91	439	411	264
Abrupt decrease	34	36	29	145	145	141
Gradual decrease	6	4	10	0	0	4

Table 2: Repartition of the results by category (significance level of 5%). “V” stands for the method from Villarini et al..

<i>Cause for gradual change detection</i>	<i>Precipitation</i>	<i>Temperature</i>
MK positive before change point	3	39
MK positive after change point	6	118
MK positive both before and after	92	75
MK positive but not Pettitt	0	36
Totals	101	268

Table 3: Repartition of the gradual changes detected by the method from Villarini et al..

List of Figures

1	Compared power of different detection methods on trend (f_1) and step (f_2 and f_3), for different values of the coefficient of variation CV	23
2	(a) Matrix representation of the MK and Pettitt statistic. The former is represented by the big triangle, and the latter by the rectangle. (b) Matrix representation of S and the Pettitt period statistic, $k(\tau, d)$	23
3	Algorithm for deciding whether a change is gradual or abrupt, depending on the value of d	24
4	Compared success rates of the method proposed in this paper and that by [43] in the case of normally distributed noise, with $n = 100$ and for different values of D/n . a) Success rates for the detection of gradual changes (when change is detected), and b) and c) for the detection of abrupt changes (when change is detected).	24
5	Same as in Figure 4, but for gamma-distributed noise ($\varepsilon \sim \Gamma(1; 1)$).	25
6	Success rate of the detection of gradual changes when both a trend and a step are present, for a normally distributed noise.	26
7	Changes in annual precipitation over the continental United States during 1910 – 2009, with $D/n = 0.2$ (upper panel) and $D/n = 0.3$ (lower panel).	27
8	Changes in annual average temperature over the continental United States during 1910 – 2009, with $D/n = 0.2$ (upper panel) and $D/n = 0.3$ (lower panel).	28
9	Patterns found both for $D/n = 0.2$ and $D/n = 0.3$: step change model for annual precipitation totals at Prescott, AR, and linear trend at Goshen, IN.	29
10	A pattern found both for $D/n = 0.2$ and $D/n = 0.3$: linear trend model for annual average temperatures at Saint Johnsbury, VT.	29

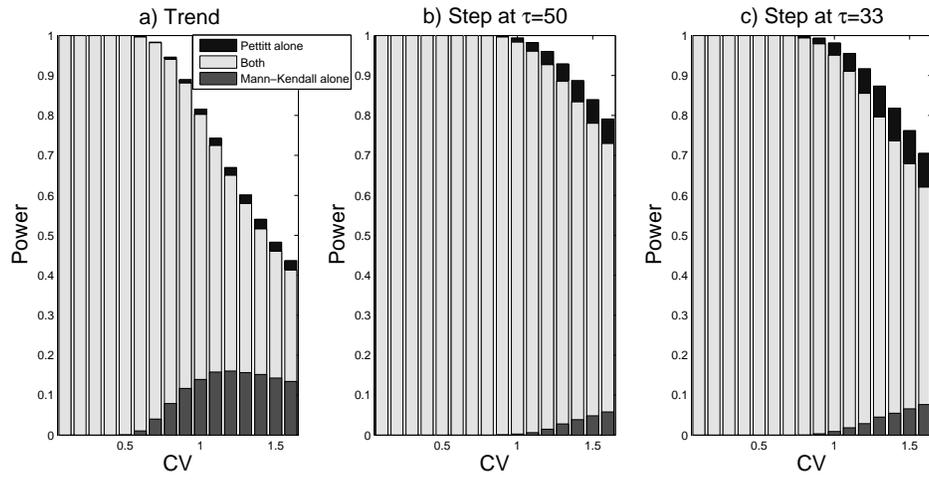


Figure 1: Compared power of different detection methods on trend (f_1) and step (f_2 and f_3), for different values of the coefficient of variation CV

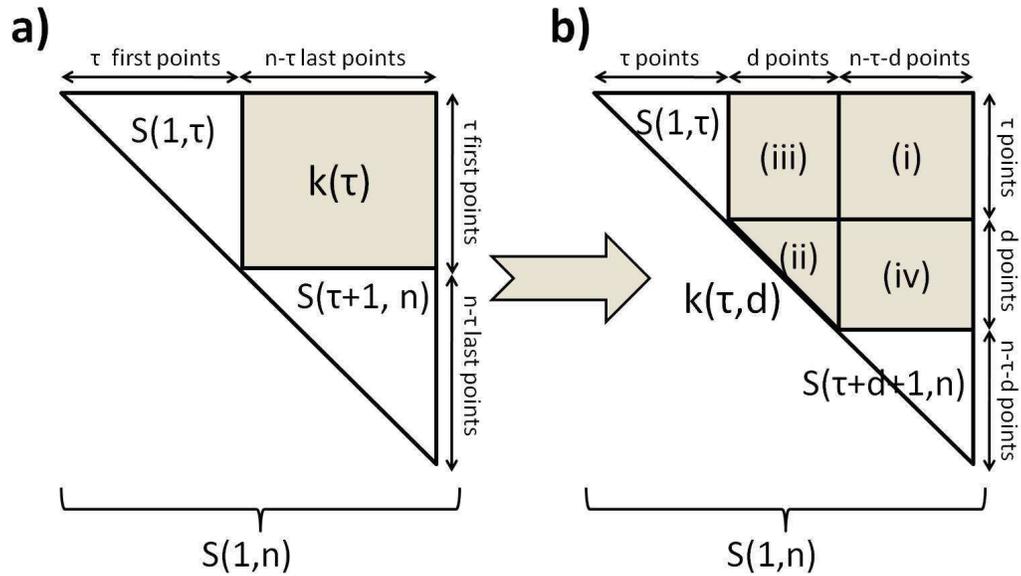


Figure 2: (a) Matrix representation of the MK and Pettitt statistic. The former is represented by the big triangle, and the latter by the rectangle. (b) Matrix representation of S and the Pettitt period statistic, $k(\tau, d)$.

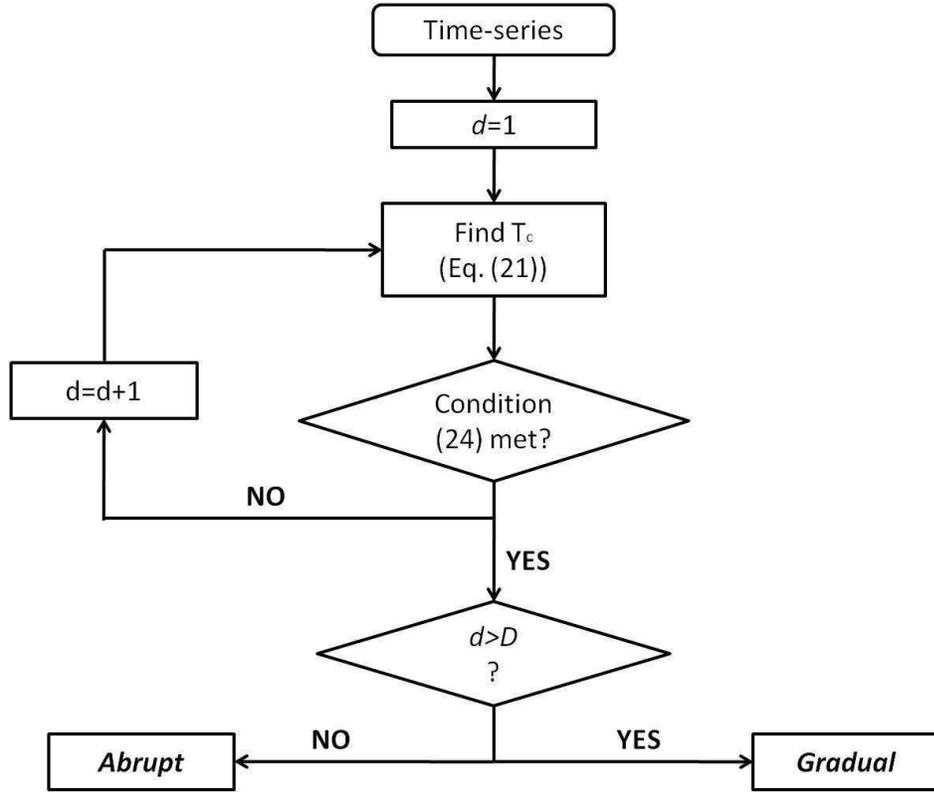


Figure 3: Algorithm for deciding whether a change is gradual or abrupt, depending on the value of d .

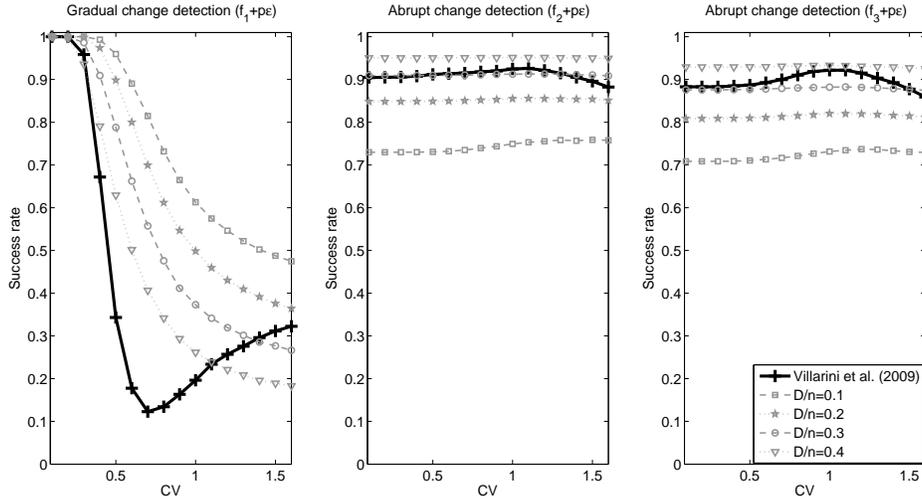


Figure 4: Compared success rates of the method proposed in this paper and that by [43] in the case of normally distributed noise, with $n = 100$ and for different values of D/n . a) Success rates for the detection of gradual changes (when change is detected), and b) and c) for the detection of abrupt changes (when change is detected).

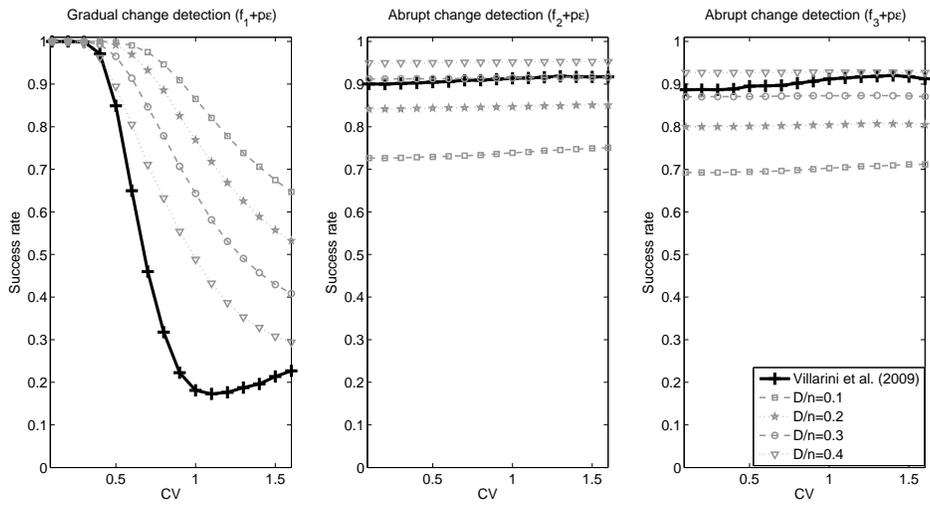


Figure 5: Same as in Figure 4, but for gamma-distributed noise ($\varepsilon \sim \Gamma(1; 1)$).

Gradual change detection ($f_1 + f_2 + p\varepsilon$)

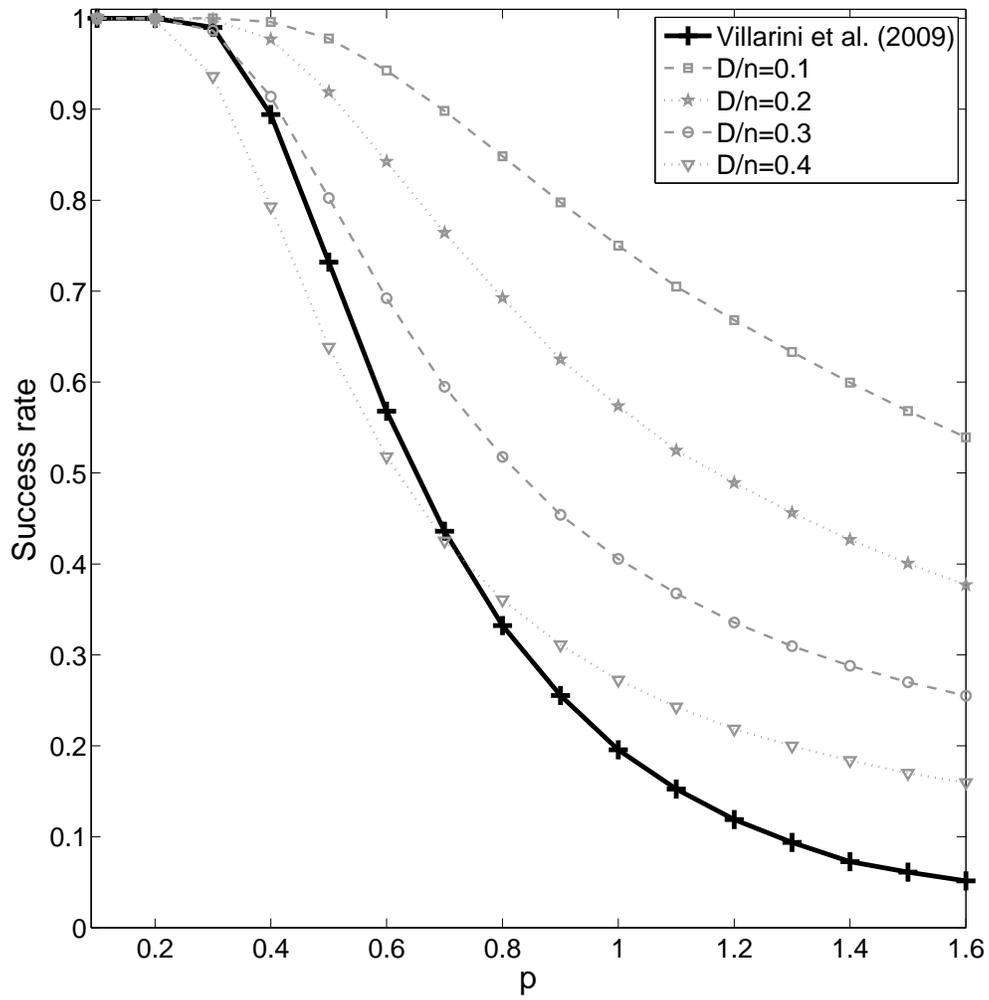


Figure 6: Success rate of the detection of gradual changes when both a trend and a step are present, for a normally distributed noise.

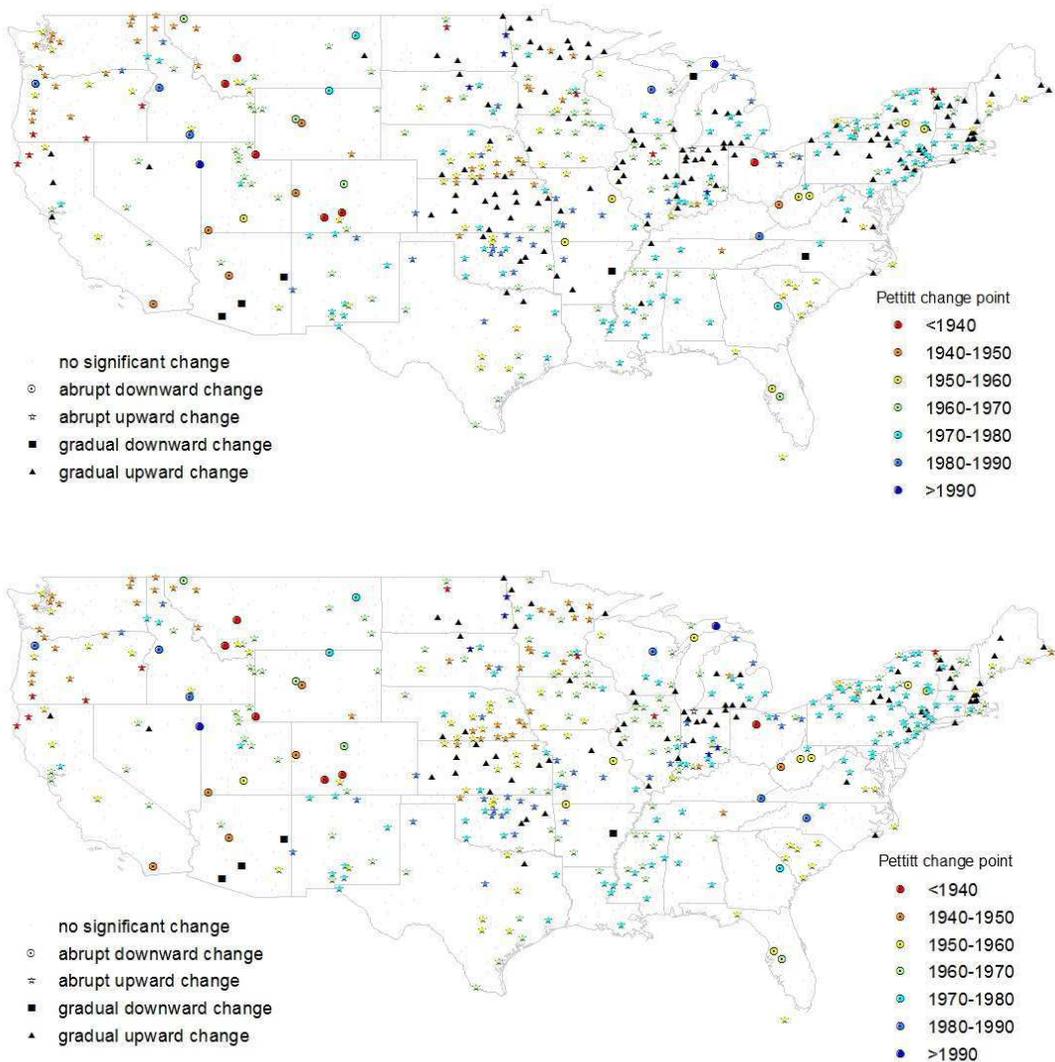


Figure 7: Changes in annual precipitation over the continental United States during 1910 – 2009, with $D/n = 0.2$ (upper panel) and $D/n = 0.3$ (lower panel).

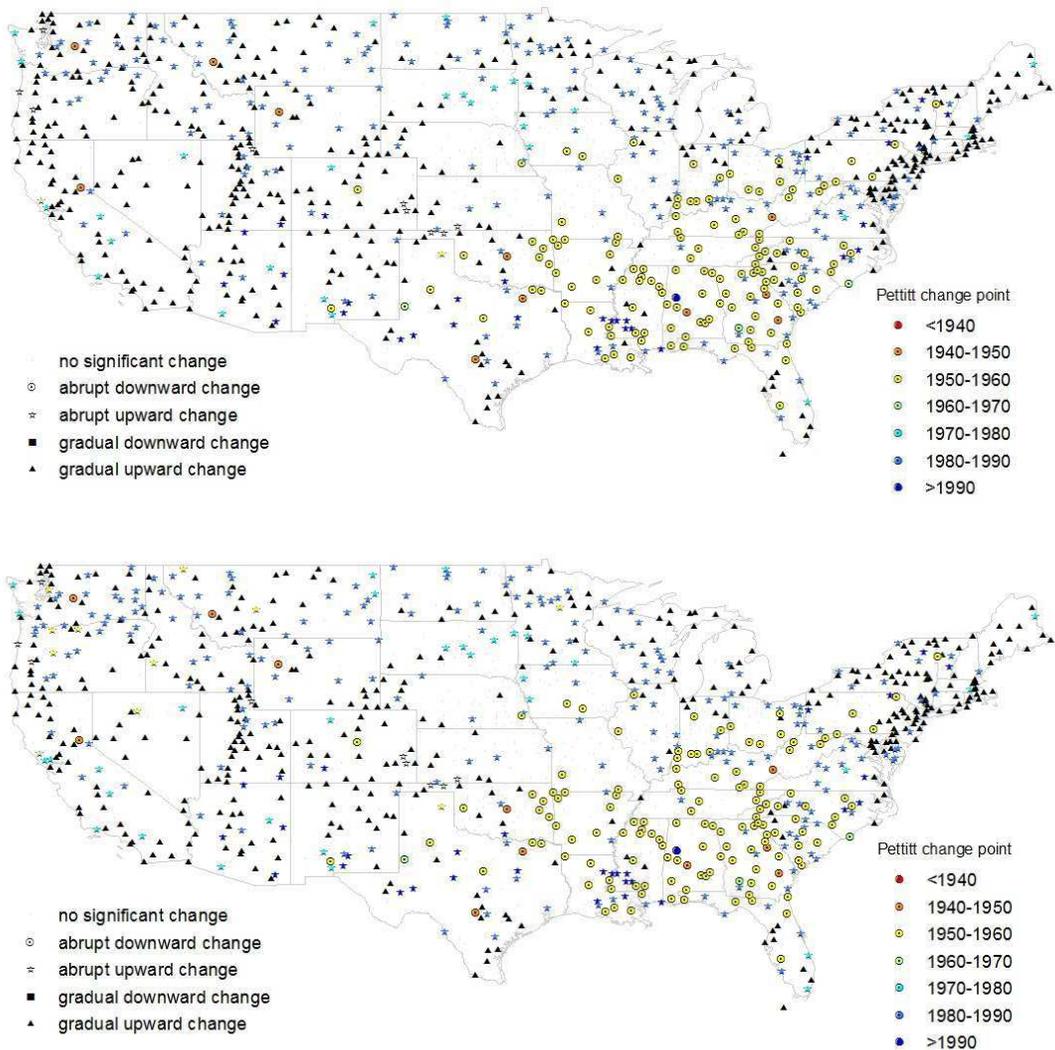


Figure 8: Changes in annual average temperature over the continental United States during 1910–2009, with $D/n = 0.2$ (upper panel) and $D/n = 0.3$ (lower panel).

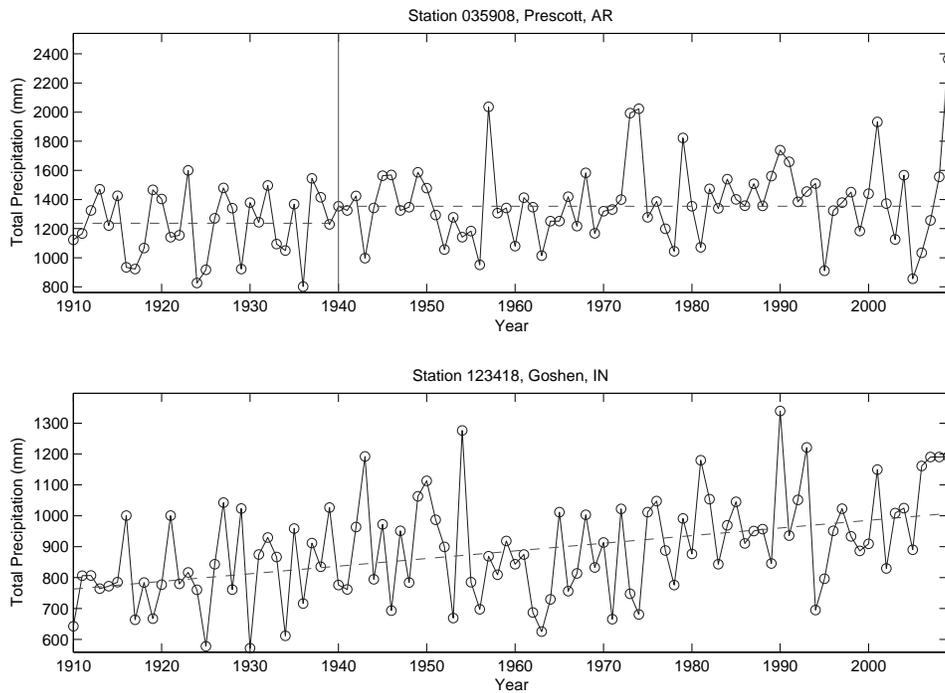


Figure 9: Patterns found both for $D/n = 0.2$ and $D/n = 0.3$: step change model for annual precipitation totals at Prescott, AR, and linear trend at Goshen, IN.

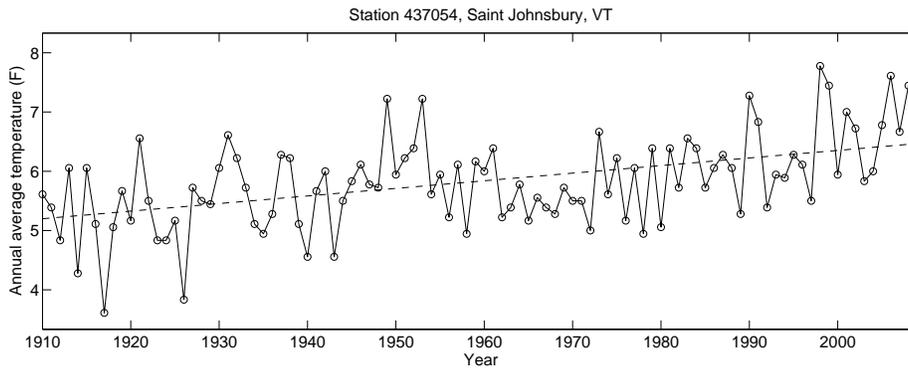


Figure 10: A pattern found both for $D/n = 0.2$ and $D/n = 0.3$: linear trend model for annual average temperatures at Saint Johnsbury, VT.