7.9 A COMPARISON OF METHODS FOR THE IDENTIFICATION OF STEPS IN TEMPERATURE SERIES: AN OVERVIEW

Lucie A.Vincent*
Climate Research Branch, Meteorological Service of Canada

Jean-Francois Ducré-Robitaille and Gilles Boulet
Ministère de l'Environnement du Québec

1. INTRODUCTION

Over the past years, considerable efforts were devoted in the development of techniques for the identification of discontinuities in climatological time series. The main objective was to identify steps caused by non-climatic factors such as changes in measurement practices, station relocation, and changes in the surrounding of a station. Many procedures have considered the use of a reference series, often created from surrounding stations, for explaining the natural variations in the climatological time series. The difference between the series of the candidate station and the reference series would be closely examined to identify abrupt jumps which could be caused by a non-climatic factor. Many techniques have been tested for their ability of detecting a single step. However, very few methods have been evaluated for the detection of multiple steps or to determine whether the series is homogeneous (the series is in agreement with the surrounding stations and no step is identified). In this study, we attempt to explore this issue by comparing the ability of seven methods for the identification of homogenous series, series with one step and series with a random number of steps using simulated annual temperature datasets. A detailed description of the methods and of the analysis is provided in Ducré-Robitaille et al., 2003.

2. SIMULATION OF TEMPERATURE SERIES

Three groups of datasets were created for this study: homogenous series, series with one step, and series with a random number of steps. To simulate homogenous series, random numbers normally distributed with mean zero and variance one were generated from an AR(1) model with an autocorrelation equal to 0.1. This process was repeated 1000 times to produce 1000 homogenous series of 100 values (or years). To simulate series with one discontinuity, a step was artificially introduced in the homogenous series. Its magnitude varied from 0.25°C to 2.0°C with an incremental increase of 0.25°C, and its position was set to 5, 10, 15, 20, 35 and 50. Using this approach, we have generated 1000 series for each combination of magnitude and position for a total of 48 000 series with a single step. We have also generated 25 000 series with a random number of steps. The magnitude of the step (ranging from 0.25°C to 2.0°C) followed a normal distribution whereas the time interval between two consecutives steps followed a truncated exponential. Overall, we have simulated 74000 candidate series: 1000 homogenous, 48 000 with one step and 25 000 with a random number of steps.

A reference series was also created for each candidate series. A homogenous series was first generated using the procedure described above. Then it was cross-correlated with the candidate series (before the introduction of the discontinuity) by multiplying the candidate values by 1.5 and adding them to the reference series. Each series were then re-standardized to have similar mean and variance of those of the candidate series. The correlation between the candidate and reference series is of the order of 0.8.

3. OVERVIEW OF THE METHODS

Seven methods were selected for the comparison study. They are based on a statistical test to determine the position and the magnitude of a step. In addition, all of them performed reasonably well for the detection of a single step. The various statistical tests were conducted at the 5% level. The methods are:

- Standard Normal Homogeneity Test (SNHT) without trend (Alexandersson, 1986)
- Standard Normal Homogeneity Test (SNHT) with trend (Alexandersson and Moberg, 1997)
- Multiple Linear Regression (MLR) (Vincent, 1998)
- Two-Phase Regression (TPR) (Easterling and Peterson, 1995)
- Wilcoxon Rank-Sum (WRS) (Karl and Williams, 1987)
- Sequential Testing for equality of means (ST) (Gullett et al., 1990)
- Bayesian Approach (Perreault et al., 2000).

Each method uses a different approach for detecting the most probable position of a step. The methods SNHT without and with trend, TPR, WRS and ST are applied directly on the difference between the candidate and reference series whereas the MLR and Bayesian approach use candidate and reference series separately. For SNHT without trend, the averages before and after a potential step are compared whereas for SNHT with trend, a trend is identified between two

* *Corresponding author address*: Lucie A. Vincent, Climate Research Branch, Meteorological Service of Canada, Downsview, ON, Canada, M3H 5T4 ; email: Lucie.Vincent@ec.gc.ca

potential changepoints. Regression models are used in MLR and TPR: in MLR, parameters are introduced in the model to describe trend and step, while two models are applied on the segments before and after a potential step in TPR. The SNHT with trend and MLR are the only two methods which can identify trends. The ST is similar to the SNHT without trend but the window to compute the average is much smaller. The WRS and Bayesian methods are very different from the methods above. The first one relies on a non-parametric test whereas the second is based on the density distribution of the data. Finally, to demonstrate the importance of using a reference series, the Bayesian approach was tested twice: once without the reference series and a second time with the reference series.

4. RESULTS

4.1 *Identification of homogenous series*

To assess the ability of each method to determine if a series is homogeneous, the methods were applied to 1000 homogeneous series and the frequency of the steps falsely detected was recorded (Table 1). The SNHT with trend, TPR and WRS are not as dependable as the others since their percentage of false detection reach 13.3%, 41.3% and 56.3%, respectively. The percentage of false detection represents the percentage of series for which the null hypothesis of no step is rejected when in fact this assumption is true and it is expected to be 5%. For ST and Bayesian without reference, the rate of false detection is small however the identified step is often greater than 0.6ºC. The SNHT no trend, MLR, and Bayesian with reference are found to be the most reliable techniques for the identification of homogeneous series since they have a relatively low rate of false detection and the detected steps tend to be very small.

4.2 *Identification of a single step*

To evaluate the performance of the methods to detect the position and magnitude of a single step, the methods were applied to the 48 000 series with a step. A step was identified when its position was within ±2 years and its magnitude within ±0.2°C. For both methods developed to identify trends, SNHT and MLR, a special process was planned. For MLR, the procedure stopped after detecting a significant trend and the method did not search further for a step. For SNHT with trend, any trend lasting more than 5 years was not considered as a step, and any trend lasting 5 years or less was a step, the middle point being its position and the slope its magnitude.

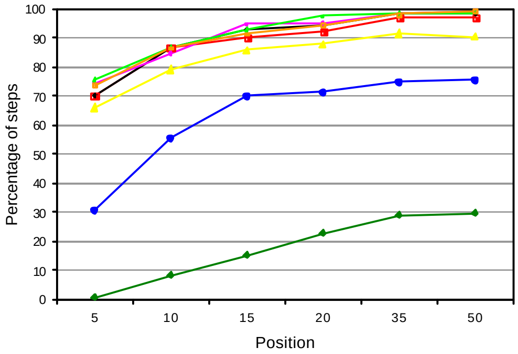
Figure 1 summarizes the results. For example, SHNT no trend identifies a single step of 0.5°C at position 15 about 70% of the time. Two methods are definitely less precise than the others: ST and Bayesian without reference. The ST uses a 5-years window to detect a step and the length of the window is probably insufficient to determine its statistical significance. As for the Bayesian approach without reference, the difficulty for detecting the step is due to the absence of reference series, which help explaining the variability inherent in the candidate series. The results also show that steps greater than 1.0ºC are identified by SNHT with and without trend, MLR, TPR, WRS and Bayesian with reference more than 80% of the time at position 20. For the steps less 1.0ºC, it is the SNHT without and with trend and the Bayesian with reference that seem to offer the best precision. It appears that both methods developed to identify trends, SNHT and MLR, are disadvantaged when they are used to detect small steps toward the middle of the series (the curves peak and decrease toward position 50 in Figure 1).

Table 1. Percentage of steps falsely detected by each method when applied to 1000 homogeneous series (%).

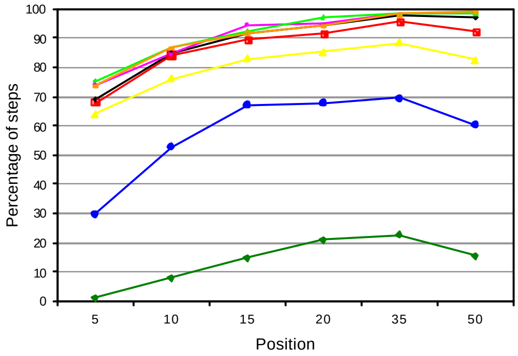
Magnitude (°C)	SNHT no trend	SNHT with trend	MLR	TPR	WRS	ST	Bayes no ref.	Bayes with ref.
0.0 - 0.1	0.2	0.0	0.2	12.9	6.6	0.0	0.0	0.5
0.1 - 0.2	1.8	3.0	1.0	15.0	22.0	0.0	0.0	2.5
0.2 - 0.3	4.0	4.4	1.4	8.2	16.7	0.0	0.0	2.1
0.3 - 0.4	1.4	2.8	0.6	3.9	8.1	0.3	0.0	0.9
0.4 - 0.5	0.9	1.9	0.3	1.0	1.7	0.3	0.5	0.4
0.5 - 0.6	0.3	1.1	0.1	0.2	1.0	0.3	1.5	0.4
> 0.6	0.0	0.1	0.0	0.1	0.2	3.5	3.8	0.0
Total	8.6	13.3	3.6	41.3	56.3	4.9	5.8	6.8
Total > 0.4	1.2	3.1	0.4	1.3	2.9	4.6	5.8	0.8

Figure 1. Percentage of steps identified when one artificial step is introduced. The magnitude is represented by a color (0.25-dark green, 0.5- blue, 0.75-yellow, 1.0-red, 1.25-black, 1.5-pink, 1.75-light green, 2.0-orange).

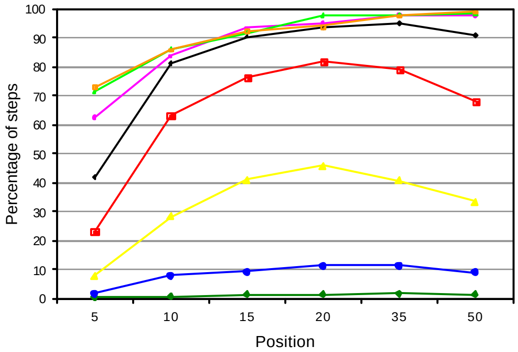
a) SHNT no trend



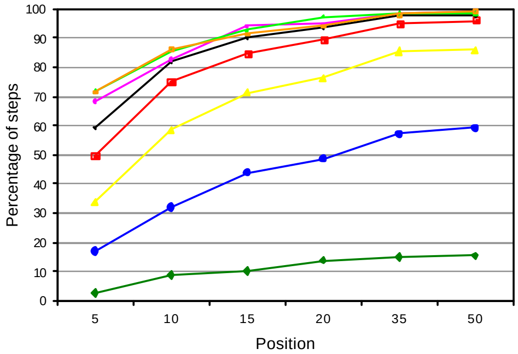
b) SNHT with trend



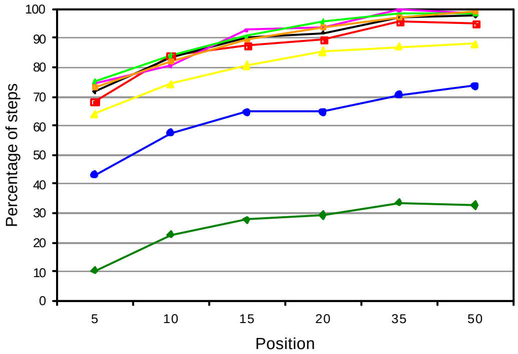
c) MLR



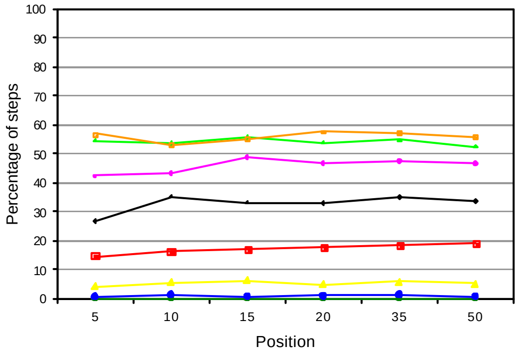
d) TPR



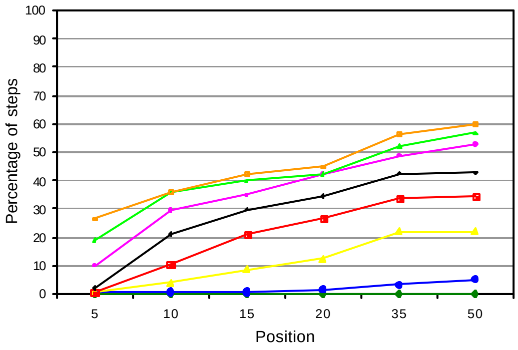
e) WRS



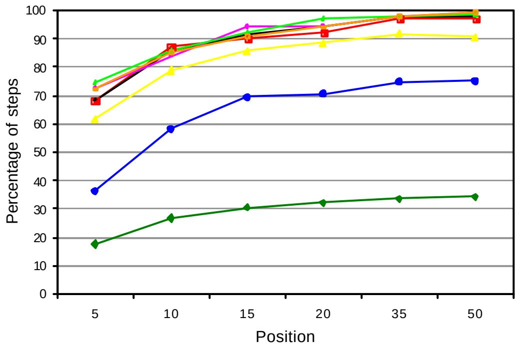
f) ST



g) Bayes no reference



h) Bayes with reference



4.3 *Identification of a random number of steps*

The methods were also tested to identify a random number of steps. After detecting the first step, the series is divided into two segments at the position of the step and each segment is re-tested separately. This process continues until each segment is assessed to be homogeneous or its length is less than 10. The percentage of series with the number of steps detected versus the number of steps artificially introduced in the candidate series was recorded. Of the 25 000 simulated series, only six series had seven steps, therefore, the results for seven steps are not very reliable. The SNHT without trend is definitely the best method for detecting the correct number of steps, followed by SNHT with trend and MLR. For example, 84.0 % of the time three steps are identified when three steps are introduced in the candidate series by SNHT without trend. For SNHT with trend and MLR, the percentage decreases to 69.3% and 63.9% respectively. A detailed description of this analysis is provided in Ducré-Robitaille et al., 2003.

5. CONCLUSION

This comparison study allows us to determine the advantages and weaknesses of each method. In application, it is probably favorable to apply several techniques to the same climatological time series and to verify if most methods reach consensus on the identified steps. In addition, it is also preferable to consult the station history file to identify the cause of the identified steps. In the field of climate data homogenization, more studies are needed. This includes the identification of steps in other time scale series such as in monthly and daily values. In addition, further work in the development of adjustment procedure would be greatly beneficial.

6. REFERENCES

Alexandersson, H,. 1986: A homogeneity test applied to precipitation data. *J. Climatol.*, **6**, 661-675.

Alexandersson, H., and A. Moberg, 1997: Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *Int. J. Climatol.*, **17**, 25-34.

Ducré-Robitaille, J.F., L.A. Vincent and Gilles Boulet, 2003: Comparison of techniques for detection of discontinuities in temperature series. *Int. J. Climatol.*, **23**, 1087-1101.

Easterling, D.R., T.C. Peterson, 1995: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.*, **15**, 369-377.

Gullet, D.W., L. Vincent and P.J.F. Sajecki, 1990: Testing homogeneity in temperature series at a Canadian climate stations. CCC report 90-4, Climate Research Branch, Meteorological Service of Canada, Ontario, Canada, 43 pp.

Karl TR, Williams CN Jr. 1987. 'An approach to adjusting climatological time series for discontinuous inhomogeneities', *J. Climate Appl. Meteorol.*, **26**, 1744-1763.

Perreault, L., J. Bernier, B. Bobée and E. Parent, 2000 : Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *J. Hydrology*, **235**, 221-241.

Vincent, L.A., 1998: A Technique for the identification of inhomogeneities in Canadian temperature series. *J. Clim.*, **11**, 1094-1104.