

# Chapter 17

## Change Detection in Climate Time Series Based on Bounded-Variation Clustering

Mohammad Gorji Sefidmazgi, Mina Moradi Kordmahalleh,  
Abdollah Homaifar, and Stefan Liess

**Abstract** Climate time series are generally nonstationary which means that their statistical properties change with time. Analysis of nonstationary time series requires detecting of change points between a set of clusters, where model of time series in each cluster has different statistical parameters. Common change detection methods are based on assumptions that may not be valid generally. Bounded-variation clustering can solve the change detection problem with minimum restrictive assumptions. In this paper, this method is employed to detect the pattern of changes in the Pacific Decadal Oscillation and the piecewise linear trend of US temperature. An optimal number of the change points are found with the Bayesian information criterion.

**Keywords** Time series • Non-stationary • Change detection • Abrupt climate change • Autocorrelation

### 17.1 Introduction

Studying climate time series such as temperature and precipitation requires modeling with statistical techniques. Although the earth's climate has changed gradually in response to both natural and human-induced processes, it is known that climate may have abrupt change, i.e., a large shift may happen in climate that persists for

---

M. Gorji Sefidmazgi • M. Moradi Kordmahalleh  
Department of Electrical Engineering, North Carolina A&T State University,  
Greensboro, NC, USA  
e-mail: [mgorjise@aggies.ncat.edu](mailto:mgorjise@aggies.ncat.edu); [mmoradik@aggies.ncat.edu](mailto:mmoradik@aggies.ncat.edu)

A. Homaifar (✉)  
Department of Electrical and Computer Engineering, North Carolina Agricultural and Technical  
State University, 1601 East Market Street, Greensboro, NC 27411, USA  
e-mail: [Homaifar@ncat.edu](mailto:Homaifar@ncat.edu)

S. Liess  
Department of Soil, Water, and Climate, University of Minnesota, Minneapolis, MN, USA  
e-mail: [liess@umn.edu](mailto:liess@umn.edu)

years or longer. Example of these changes includes the changes in average temperature, patterns of storms, floods, or droughts over a widespread area (Lohmann 2009). Climatic records show that large and widespread abrupt changes have occurred repeatedly throughout the geological records (Alley et al. 2003). Many studies have analyzed climate time series in the stationary framework; i.e., the statistical parameters are assumed to be constant over time. However, stationary assumption of climate time series is invalid considering various internal dynamics and external forcings (Milly et al. 2008). Thus, statistical techniques based on stationary assumption should be modified to reveal the characteristics of the abrupt climate change. Nonstationary time series have a set of *clusters* (regime, phase, or segment), while the model of each cluster is stationary. These clusters are separated in time by some change points (breaks). The analysis of nonstationary time series, including finding the change points between the clusters, is an ongoing research area in the climate data analysis.

Several approaches were proposed in the literature for the change detection in climate time series. Brute-force search was performed over all candidate points to find the best change points (Liu et al. 2010). However, this method is not applicable for longer time series with high number of change points due to huge volume of computations. The change points were estimated by Bayesian inference, where the change points and other model parameters were assumed as random variables (Ruggieri 2013). Kehagias and Fortin (2006) used a method based on hidden Markov models, assuming that the time series was generated by a Markov process. Then, the unknown parameters were determined by the maximum likelihood. However, statistical assumptions on the data and the change points in Bayesian and Markov methods may not be true in general. Several statistical tests were used in atmospheric studies such as sequential Mann–Kendall, Bai–Perron, and Pettitt–Mann–Whitney to find the change points. However, the results of these tests are valid only if the data are not serially correlated (Lyubchich et al. 2013). For the correlated time series with only one change point, proper hypothesis tests were introduced (Robbins et al. 2011).

The bounded-variation (BV) clustering (Metzner et al. 2012) is another technique which finds the change points in nonstationary time series. In this method, instead of statistical assumptions on the data or the change points, a reasonable assumption is made such that the total number of the change points between the clusters is bounded. The BV clustering is computationally efficient and is also applicable for the serially correlated time series. Assuming a different linear trend (Horenko 2010a) and a vector autoregressive (Horenko 2010b) in each cluster, the BV clustering was used to analyze the climate dynamics in the ERA-40 reanalysis data. The BV clustering was used in Gorji Sefidmazgi et al. (2014c) for analyzing the climate variability of North Carolina. The effect of covariates on nonstationary time series was analyzed by the BV clustering (Gorji Sefidmazgi et al. 2014a; Horenko 2010b; Kaiser and Horenko 2014).

In this paper, we have shown the applicability of the BV clustering by two numerical examples: the pattern of changes in the Pacific Decadal Oscillation (PDO)

and the US land surface temperature. Moreover, the Bayesian information criterion (BIC) is applied to find the optimal number of the change points and the number of clusters.

## 17.2 Dataset

The bias-adjusted monthly average temperature of the US continental stations is derived from the US Historical Climatology Network database (<http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html>). The period 1900 until 2013 is selected, and the stations with continuous missing data for more than 4 months are eliminated. Then, the missing data in the remaining 1,189 stations are filled by interpolation, and also the mean cycle of time series is removed to eliminate the effect of seasonality.

Annual time series of the PDO for 1900–2013 is from NOAA database (<http://www.esrl.noaa.gov/psd/data/climateindices/list>).

## 17.3 Method

The BV clustering might be applied in two cases, where the model of the time series in each cluster is in the form of a mathematical function (such as polynomial or differential equation) or a statistical distribution (such as Gaussian, Gamma, etc.). The change points and the parameters of each cluster are determined by solving a least square/maximum likelihood (LS/ML) and a constrained optimization.

Let  $x(t)$  be a nonstationary time series with  $M$  clusters. The first case is when the model in each cluster is a function of time (and other covariates  $u(t)$  if exist), i.e.,  $x(t) = f(x(t-1), \dots, x(t-p), u(t-1), \dots, u(t-n), t, \alpha_m)$ . Here,  $f$  is the model of the time series,  $\alpha_m$  is the set of parameters in the  $m$ th cluster where  $m \in \{1, \dots, M\}$ . Also,  $p$  and  $n$  are the order of the lagged outputs and the covariates, respectively. In the second case,  $f$  is a probability density function, i.e.,  $P(X = x(t)) = f(x(t)|u(t), t, \alpha_m)$ . For these cases, *model distance function*  $d_m(x(t))$  is the distance between the time series at time  $t \in \{1, \dots, T\}$  and the model of the  $m$ th cluster, which can be defined by the Euclidean distance or the likelihood function:

$$d_m(x(t)) = \|x(t) - f(x(t-1), \dots, x(t-p), u(t-1), \dots, u(t-n), t, \alpha_m)\|^2 \quad (17.1)$$

$$d_m(x(t)) = \ell\left(f\left(x(t)|u(t), t, \alpha_m\right)\right) \quad (17.2)$$

where  $\|\cdot\|$  and  $\ell(\cdot)$  are the  $L_2$ -norm and the negative log-likelihood operators, respectively. Now, the change detection problem can be defined as a minimization:

$$\min_{\mu, \alpha} \sum_{t=1}^T \sum_{m=1}^M \mu_m(t) \cdot d_m(x(t)) \tag{17.3}$$

In (17.3),  $\mu_m(t) \in \{0, 1\}$  is the *cluster membership function* indicating whether the datum at time  $t$  belongs to the  $m$ th cluster or not. The change points are the times when the values of  $\mu_m(t)$  are changed. For example, if  $\mu_2(50) = 0$  and  $\mu_2(51) = 1$ , then the cluster 2 is started at the change point  $t = 51$ . Clearly, the datum at each time belongs to only one of the clusters, and hence,

$$\sum_{m=1}^M \mu_m(t) = 1 \quad t = \{1, \dots, T\} \tag{17.4}$$

Now, we augmented the model distance functions in  $D_m = [d_m(x(1)), d_m(x(2)), \dots, d_m(x(T))]$ , and also the cluster membership functions in  $U_m = [\mu_m(1), \mu_m(2), \dots, \mu_m(T)]$  for  $m = \{1, \dots, M\}$ . The optimization in (17.3) is rewritten as below (Metzner et al. 2012):

$$\min_{U_m, D_m} \sum_{m=1}^M D_m \cdot U_m^T \tag{17.5}$$

The BV clustering solves the optimization (17.5) in two iterative steps using the coordinate-descent algorithm. In the first step, it assumes that cluster membership function  $\mu_m(t)$  is known and the cluster parameters  $\alpha_m$  are found. In the second step,  $\alpha_m$  are fixed and  $\mu_m(t)$  is determined.

In the first step, assume that  $\mu_m(t)$  and the change points are known. The data belong to each cluster are separated, and the parameters  $\alpha_m$  are found by LS/ML. Using the estimated parameters  $\alpha_m$ , the model distance function  $d_m(x(t))$  is determined using (17.1) or (17.2). In the next step, the cluster membership function  $\mu_m(t)$  should be found. The assumption that the number of the change points is bounded should be added to the problem formulation with imposing constraints on  $\mu_m(t)$ . First, a counter  $q_m(t) \in \{0, 1\}$  is defined which is increased by one unit when  $\mu_m(t)$  changes from 0 to 1 or vice versa (i.e., on the change points) (Metzner et al. 2012):

$$|\mu_m(t+1) - \mu_m(t)| \leq q_m(t) \rightarrow \begin{cases} \mu_m(t+1) - \mu_m(t) - q_m(t) \leq 0 & m = 1, \dots, M \\ -\mu_m(t+1) + \mu_m(t) - q_m(t) \leq 0 & t = 1, \dots, T-1 \end{cases} \tag{17.6}$$

In order to limit the total number of the change points to a constant  $Q$ , a constraint on  $q_m(t)$  is added in the following form:

$$\sum_{t=1}^{T-1} \sum_{m=1}^M q_m(t) = 2Q \tag{17.7}$$

Now,  $\bar{q}_m = [q_m(1), q_m(2), \dots, q_m(T-1)]$  is defined to record  $q_m(t)$  over time and is added to the set of unknown parameters. By defining (17.8) and (17.9), (17.5) is rewritten in (17.10) to find  $\mu_m(t)$  (Metzner et al. 2012):

$$\tilde{\mathbf{D}} = \begin{bmatrix} \underbrace{D_1 \ D_2 \ \dots \ D_M}_{T \times M} & \underbrace{0 \ \dots \ 0}_{(T-1) \times M} \end{bmatrix} \quad (17.8)$$

$$\tilde{\mathbf{U}} = \begin{bmatrix} \underbrace{U_1 \ U_2 \ \dots \ U_M}_{T \times M} & \underbrace{\bar{q}_1 \ \bar{q}_2 \ \dots \ \bar{q}_M}_{(T-1) \times M} \end{bmatrix} \quad (17.9)$$

$$\min_{\tilde{\mathbf{U}}} \tilde{\mathbf{U}} \cdot \tilde{\mathbf{D}}^T \quad (17.10)$$

All the elements of unknown vector  $\tilde{\mathbf{U}}$  in (17.10) are either 0 or 1. Thus, this optimization is a constrained optimization in the form of a *binary integer programming*. The set of linear constraints are (17.4), (17.6), and (17.7) which include  $2M \times (T-1)$  inequality and  $T+1$  equality constraints. There are standard methods for solving constrained optimization using some toolboxes in Matlab or R (Gurobi 2014). Once  $\tilde{\mathbf{U}}$  is found,  $\mu_m(t)$  for all of the clusters and the change points can be determined.

In conclusion, the BV-clustering algorithm includes the following steps: first, a random initial  $\mu_m(t)$  is selected such that it satisfies (17.4). Then, the parameters of each cluster are calculated by the LS/ML, and the model distance function  $d_m(x(t))$  is determined by (17.1) or (17.2). Then, the optimization problem in (17.10) is constructed and solved with the constraints of (17.4), (17.6), and (17.7). The LS/ML and the constrained optimization steps are repeated for some predefined number of iterations (usually five). This procedure converges to at least a local solution of the optimization in (17.3). For finding the global solution, the algorithm should be started with different initial random  $\mu_m(t)$ .

## 17.4 Model Selection

The number of clusters  $M$  and the change points  $Q$  should be set in the BV clustering. In the time series literature, the number of change points is usually found by information theory methods such as the BIC (Jandhyala et al. 2013). The BIC is a well-known approach to perform a trade-off between the goodness of fit and the complexity of models and to prevent over-fitting/under-fitting.

The index for the detected cluster at time  $t$  is determined by  $m^*(t) = \arg \max_m (\mu_m(t))$  for  $t = \{1, \dots, T\}$ . By obtaining the cluster parameters using the maximum likelihood, the minimized value of the negative log-likelihood  $V$  and the BIC are determined by

$$V = \sum_{t=1}^T \ell \left( x(t) \middle| u(t), t, \alpha_{m^*(t)} \right) \quad (17.11)$$

$$BIC(M, Q) = 2V + \ln(T) \times (\text{number of estimated parameters}) \quad (17.12)$$

Otherwise, by obtaining the clusters parameters by the least square method, the residual  $w(t)$  is determined by

$$w(t) = x(t) - f \left( x(t-1), \dots, x(t-p), u(t-1), \dots, u(t-n), t, \alpha_{m^*(t)} \right) \quad (17.13)$$

Adding the assumption that the residual follows a normal distribution with a constant variance, the loss function  $V$  and the BIC are found (Hastie et al. 2009):

$$V = \ln \left( \frac{\left( \sum_t w^2(t) \right)}{T} \right) \quad (17.14)$$

$$BIC(M, Q) = T \times V + \ln(T) \times (\text{number of estimated parameters}) \quad (17.15)$$

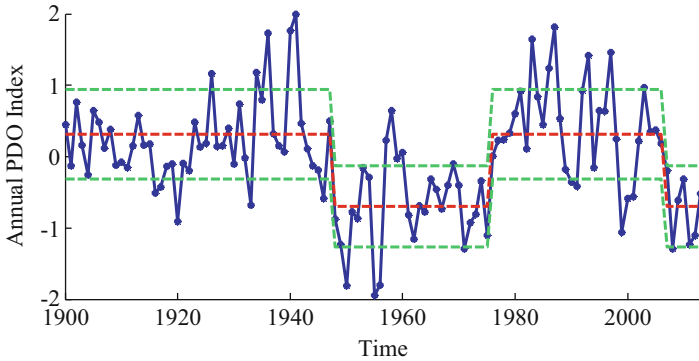
Assume that the number of parameters for each cluster is  $\omega$ . For example, the set of parameters for a time series with Gaussian distribution in each cluster includes the mean and the variance, and thus  $\omega = 2$ . Hence, in addition to  $Q$  change points, we need to estimate  $\omega$  parameters for each of the  $M$  clusters:

$$\text{number of estimated parameters} = M \times \omega + Q \quad (17.16)$$

The BV clustering should be applied to the data with various possible values of  $M$  and  $Q$ . Finally, the model with the smallest BIC is chosen.

## 17.5 Results and Conclusion

In this section, we applied the BV-clustering method on two time series, the PDO and the US surface temperature. It is well known that the PDO has some regimes with different mean values (Rodionov 2006). Assume that the model of each cluster is a normal distribution  $N(\rho_m, \sigma_m^2)$ . The distance function is defined as  $d_m(x(t)) = -1/2 [\ln(2\pi) + \ln(\sigma_m^2) + (x(t) - \rho_m)/\sigma_m^2]$  which is equivalent to the negative log-likelihood of the normal distribution. Using the maximum likelihood, the mean and the variance of each cluster are found similar to the parameters of the mixture models (Hastie et al. 2009):



**Fig. 17.1** Pattern of change in the PDO time series. The *red lines* show the averages of time series in each cluster and *green lines* are average  $\pm$  standard deviation

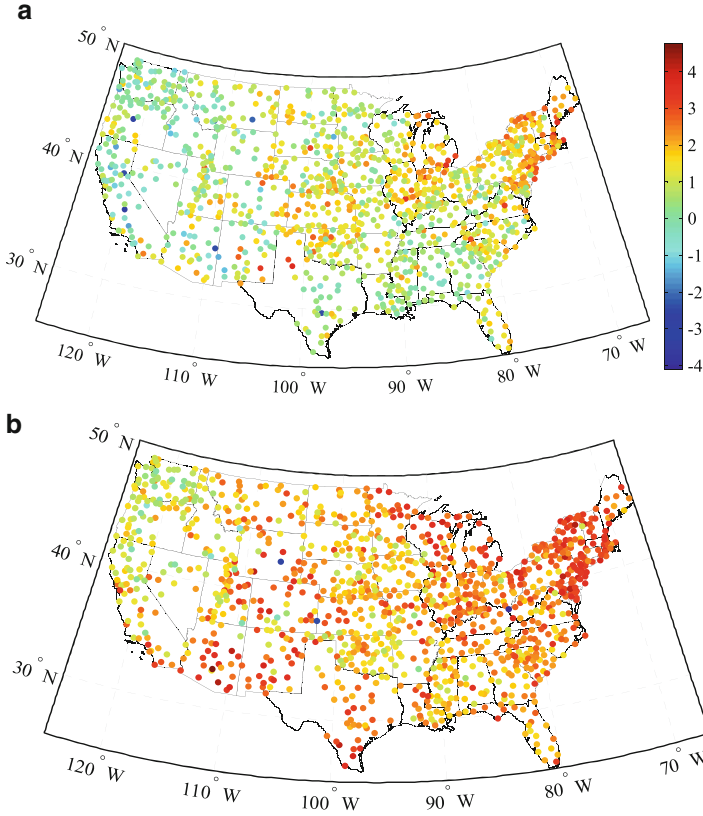
$$\rho_m = \frac{\left( \sum_t \mu_m(t) \cdot x(t) \right)}{T}; \quad \sigma_m^2 = \frac{\left( \sum_t \mu_m(t) \cdot (x(t) - \rho_m)^2 \right)}{T} \quad (17.17)$$

The result of the change detection is shown in Fig. 17.1, where three change points in 1948, 1976, and 2007 are found among the three clusters. The models of the time series in [1948–1976] and [2007–2013] are the same. These results are similar to the change points found in (Rodionov 2006), while no prior knowledge about the minimum length of the clusters is necessary in the proposed approach.

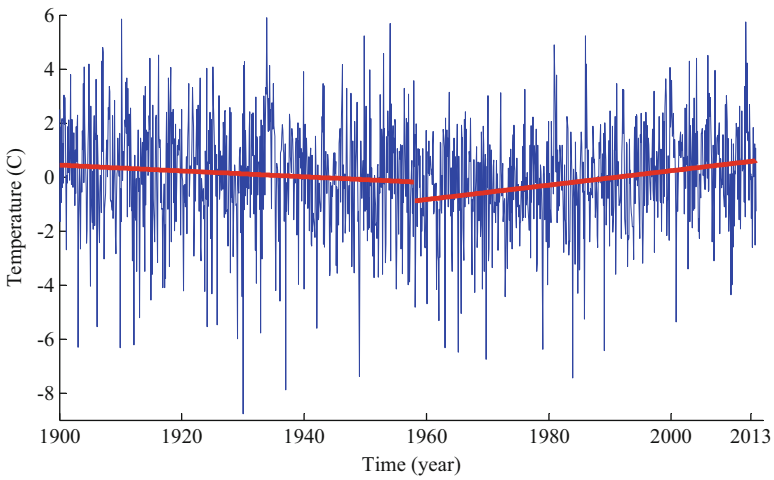
The second example is the analysis of the US average temperature where the data are serially correlated. It is known that piecewise linear trend with the first order autoregressive (AR(1)) residuals is better than the single linear trend for the surface temperature in the sense of the BIC. Moreover, the trend is not necessarily continuous at the break points (Seidel and Lanzante 2004). Assume that the model of each cluster is a linear trend plus AR(1) noise. Thus,  $x(t) = \beta_{0m} + \beta_{1m} \cdot t + \varepsilon(t)$  and  $\varepsilon(t) = \rho_m \cdot \varepsilon(t-1) + w(t)$ , where  $w(t)$  is the white noise. The model distance function is defined as

$$d_m(x(t)) = \|[x(t) - (\beta_{0m} + \beta_{1m}t)] - \rho_m [x(t-1) - (\beta_{0m} + \beta_{1m}(t-1))]\|^2 \quad (17.18)$$

If  $\rho_m = 0$ , then the linear model parameters  $\beta_0$  and  $\beta_1$  can be found in a closed form using ordinary linear square (Gorji Sefidmazgi et al. 2014b). However, in the case of  $\rho_m \neq 0$ , no closed form solution exists and these parameters should be determined by the feasible generalized least square. Results of the BV clustering for average temperature show that there are  $M = 2$  clusters with one change point in 1958. Figure 17.2 shows the spatiotemporal pattern of the linear trends in each of the clusters. Figure 17.3 shows the piecewise linear trend in one of the stations. It can



**Fig. 17.2** (a) Linear trend of temperature in 1,189 stations in cluster 1 during 1900–1958. (b) Linear trend of temperature in cluster 2 during 1958–2013



**Fig. 17.3** Anomaly of average temperature in Boulder, CO, and its piecewise linear trend. The linear trend increased after the change point of 1958



be seen that the linear trends increased after 1958 in most of the areas, especially over the eastern and central sections of the USA. Finding relations between this change point and existing physical phenomena is difficult, since there are many anthropogenic and natural factors contributing to the climate variability. However, common breaks in the trend of the global temperature and the anthropogenic forcings are reported in the early 1960s (Estrada et al. 2013). This fact can establish a direct relationship between the human effects on altering the long-term trend of the temperature.

**Acknowledgments** This work is supported by the Expeditions in Computing by the National Science Foundation under Award Number: CCF-1029731: Expedition in Computing: Understanding Climate Change.

## References

- Alley RB, Marotzke J, Nordhaus WD, Overpeck JT, Peteet DM, Pielke RA, Pierrehumbert RT, Rhines PB, Stocker TF, Talley LD, Wallace JM (2003) Abrupt climate change. *Science* 299(5615):2005–2010. doi:[10.1126/science.1081056](https://doi.org/10.1126/science.1081056)
- Estrada F, Perron P, Martinez-Lopez B (2013) Statistically derived contributions of diverse human influences to twentieth-century temperature changes. *Nat Geosci* 6(12):1050–1055. doi:[10.1038/ngeo1999](https://doi.org/10.1038/ngeo1999)
- Gorji Sefidmazgi M, Moradi Kordmahalleh M, Homaifar A, Karimoddini A (2014a) A finite element based method for identification of switched linear systems. In: American Control Conference (ACC). IEEE, Portland, USA, pp 2644–2649. doi:[10.1109/ACC.2014.6858898](https://doi.org/10.1109/ACC.2014.6858898)
- Gorji Sefidmazgi M, Sayemuzzaman M, Homaifar A (2014b) Non-stationary time series clustering with application to climate systems. In: Jamshidi M, Kreinovich V, Kacprzyk J (eds) *Advance trends in soft computing*, vol 312. Studies in fuzziness and soft computing. Springer International Publishing, Switzerland, pp 55–63. doi:[10.1007/978-3-319-03674-8\\_6](https://doi.org/10.1007/978-3-319-03674-8_6)
- Gorji Sefidmazgi M, Sayemuzzaman M, Homaifar A, Jha M, Liess S (2014c) Trend analysis using non-stationary time series clustering based on the finite element method. *Nonlinear Processes Geophys* 21(3):605–615. doi:[10.5194/npg-21-605-2014](https://doi.org/10.5194/npg-21-605-2014)
- Gurobi (2014) Gurobi optimizer reference manual, Houston, USA
- Hastie T, Tibshirani R, Friedman JH (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York
- Horenko I (2010a) On clustering of non-stationary meteorological time series. *Dyn Atmos Ocean* 49(2–3):164–187. doi:<http://dx.doi.org/10.1016/j.dynatmoce.2009.04.003>
- Horenko I (2010b) On the identification of nonstationary factor models and their application to atmospheric data analysis. *J Atmos Sci* 67(5):1559–1574. doi:[10.1175/2010JAS3271.1](https://doi.org/10.1175/2010JAS3271.1)
- Jandhyala V, Fotopoulos S, MacNeill I, Liu P (2013) Inference for single and multiple change-points in time series. *J Time Ser Anal*. doi:[10.1111/jtsa.12035](https://doi.org/10.1111/jtsa.12035)
- Kaiser O, Horenko I (2014) On inference of statistical regression models for extreme events based on incomplete observation data. *Commun Appl Math Comput Sci* 9(1):143–174. doi:[10.2140/camcos.2014.9.143](https://doi.org/10.2140/camcos.2014.9.143)
- Kehagias A, Fortin V (2006) Time series segmentation with shifting means hidden markov models. *Nonlin Processes Geophys* 13(3):339–352. doi:[10.5194/npg-13-339-2006](https://doi.org/10.5194/npg-13-339-2006)
- Liu RQ, Jacobi C, Hoffmann P, Stober G, Merzlyakov EG (2010) A piecewise linear model for detecting climatic trends and their structural changes with application to mesosphere/lower thermosphere winds over Collm, Germany. *J Geophys Res Atmos* 115(D22), D22105. doi:[10.1029/2010JD014080](https://doi.org/10.1029/2010JD014080)

- Lohmann G (2009) Abrupt climate change modeling. In: Meyers RA (ed) Encyclopedia of complexity and systems science. Springer, New York, pp 1–21. doi:[10.1007/978-0-387-30440-3\\_1](https://doi.org/10.1007/978-0-387-30440-3_1)
- Lyubchich V, Gel YR, El-Shaarawi A (2013) On detecting non-monotonic trends in environmental time series: a fusion of local regression and bootstrap. *Environmetrics* 24(4):209–226. doi:[10.1002/env.2212](https://doi.org/10.1002/env.2212)
- Metzner P, Putzig L, Horenko I (2012) Analysis of persistent nonstationary time series and applications. *Commun Appl Math Comput Sci* 7(2):175–229. doi:[10.2140/camcos.2012.7.175](https://doi.org/10.2140/camcos.2012.7.175)
- Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ (2008) Stationarity is dead: whither water management? *Science* 319(5863):573–574. doi:[10.1126/science.1151915](https://doi.org/10.1126/science.1151915)
- Robbins M, Gallagher C, Lund R, Aue A (2011) Mean shift testing in correlated data. *J Time Ser Anal* 32(5):498–511. doi:[10.1111/j.1467-9892.2010.00707.x](https://doi.org/10.1111/j.1467-9892.2010.00707.x)
- Rodionov SN (2006) Use of prewhitening in climate regime shift detection. *Geophys Res Lett* 33(12), L12707. doi:[10.1029/2006GL025904](https://doi.org/10.1029/2006GL025904)
- Ruggieri E (2013) A Bayesian approach to detecting change points in climatic records. *Int J Climatol* 33(2):520–528. doi:[10.1002/joc.3447](https://doi.org/10.1002/joc.3447)
- Seidel DJ, Lanzante JR (2004) An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes. *J Geophys Res Atmos* 109(D14), D14108. doi:[10.1029/2003JD004414](https://doi.org/10.1029/2003JD004414)