Nonlinear Processes
in Geophysics

Trend analysis using non-stationary time series clustering based on the finite element method

M. Gorji Sefidmazgi[1], **M. Sayemuzzaman**[2], **A. Homaifar**[1], **M. K. Jha**[3], **and S. Liess**[4]

[1]North Carolina A&T State University, Dept. of Electrical Engineering, Greensboro, USA
[2]North Carolina A&T State University, Dept. of Energy and Environmental Systems, Greensboro, USA
[3]North Carolina A&T State University, Dept. of Civil, Architectural and Environmental Engineering, Greensboro, USA
[4]University of Minnesota, Department of Soil, Water and Climate, St. Paul, USA

Correspondence to: A. Homaifar (homaifar@ncat.edu)

Abstract. In order to analyze low-frequency variability of climate, it is useful to model the climatic time series with multiple linear trends and locate the times of significant changes. In this paper, we have used non-stationary time series clustering to find change points in the trends. Clustering in a multi-dimensional non-stationary time series is challenging, since the problem is mathematically ill-posed. Clustering based on the finite element method (FEM) is one of the methods that can analyze multidimensional time series. One important attribute of this method is that it is not dependent on any statistical assumption and does not need local stationarity in the time series. In this paper, it is shown how the FEM-clustering method can be used to locate change points in the trend of temperature time series from in situ observations. This method is applied to the temperature time series of North Carolina (NC) and the results represent region-specific climate variability despite higher frequency harmonics in climatic time series. Next, we investigated the relationship between the climatic indices with the clusters/trends detected based on this clustering method. It appears that the natural variability of climate change in NC during 1950–2009 can be explained mostly by AMO and solar activity.

1 Introduction

In recent years, new methods have been developed in the area of machine learning and computational statistics to analyze massive data. In the climate science area, it is important to find connections between results of such data-based analysis and physical phenomena. In this paper, we address the prob-

lem of climate variability and its relation to physical phenomena. The trend analysis is useful for a better understanding of climate change and variability. Furthermore, interpreting the results of trend detection is useful for an understanding of the physical relations behind both warming and cooling on inter-annual timescales and climate indices.

The linear trend is the most straightforward assessment of the long-term behavior of a time series (Wilks, 2011). During the last decade, numerous papers have discussed long-term linear tendencies of climate parameters such as precipitation, temperature and the North Atlantic Oscillation (NAO) index (Tabari and Hosseinzadeh Talaee, 2011; Sonali and Nagesh Kumar, 2013). Due to changes in the trends, real-world time series do not generally fit in a straight line. As outlined by the IPCC Fifth Assessment Report (IPCC, 2013), in the Northern Hemisphere, 1983–2012 was likely the warmest 30-year period of the last 1400 years and has a larger temperature trend. Also, a visual inspection of global temperature time series for the period 1880–1997 shows that the mean warming obtained by fitting a straight line did not occur in a consistent manner, except in two continuous periods, one beginning around 1910 and the other starting in the mid-1970s (Tomé and Miranda, 2004). It has been statistically shown that a linear trend does not adequately describe low-frequency behavior of temperature time series (Karl et al., 2000; Lyubchich et al., 2013). Seidel and Lanzante (2004) analyzed global surface temperature anomalies and proved using the Schwarz Bayesian information criterion (SIC) that the piecewise linear model is better than a single linear trend. Other analysis methods, such as Fourier analysis, cannot find change points that result from heterogeneity in the data and many of the features become

hidden. Therefore, it is useful to find local trends and detect the breakpoints in the time series.

Mudelsee (2010) indicated some parametric and non-parametric methods to find models for trend change. In parametric methods, a cost function is defined using ordinary least squares (OLS), and a brute-force search is performed over all candidate points to estimate the change points. Also, an algorithm is presented in Tomé and Miranda (2005) for fitting a continuous regression model with several break points to data and then it was applied in order to study changes in Azores temperature, the NAO index and Lisbon winter precipitation. In addition, Liu et al. (2010) used the same method to find partial trends of wind variability in the mesosphere and the lower thermosphere. In nonparametric smoothing, the points inside a sliding window of length L are averaged using a kernel function and the window runs along the time axis (Godtliebsen et al., 2012; Scherrer et al., 2013).

Methods based on statistical tests are also used in the literature to find changes in the linear trends. They consist of performing several kinds of tests to confirm or reject the hypothesis of a change occuring. Mann–Kendall is used to detect change points in trends of rainfall and stream flow (Modarres and Sarhadi, 2009; Ehsanzadeh et al., 2011). Toreti and Desiato (2008) analyzed mean temperature trends in Italy using two separated lines detected by the progressive Mann–Kendall method. Furthermore, Martínez et al. (2010) used four homogeneity tests to find one break point in the temperature trend of Spain and then related these changes to volcanic eruptions. The Bai–Perron test is also used in change point analysis of time series (Toreti et al., 2010a; Verbesselt et al., 2010). The former paper connected temperature trend change over Italy to NAO. Beaulieu et al. (2012a, b) assumed that the most likely position for a shift in model parameters of time series is one that minimizes the SIC and found best change points in linear trends of atmospheric CO_2. Kiely (1999) found change points in precipitation in Ireland using the Pettitt–Mann–Whitney method after smoothing using a 10 year moving average window.

Bayesian and Markovian methods are also used in the literature to find change points in climatic time series. In Bayesian methods, times of the change points and other model parameters are assumed to be random variables and their statistical distribution is estimated. Markov methods assume that change points in the time series are generated by a Markov process. Dose and Menzel (2004) and Rutishauser et al. (2007) analyzed phenological time series to find piecewise linear sections using a Bayes estimator and Monte Carlo and assessed the impacts of global change. Seidou and Ouarda (2007) proposed a Bayesian method to detect multiple change points in the trend of river stream flows of Quebec. Ruggieri (2013) used a combination of Bayesian analysis and dynamic programming to find change points in the trend of the $\delta^{18}O$ record of the Plio-Pleistocene. Kehagias and Fortin (2006) identified annual discharge phases in the

Senegal River during 1903–1982 using the Hidden Markov Model.

It is very important that the mean-shift change points caused by changes in observational procedures, instrumentation or station relocation are not discarded (Martínez et al., 2010). Many of these change point times are undocumented in station histories. Gradual changes may also be caused by trees growing taller. Different methods such as Hidden Markov Modeling (Toreti et al., 2012), minimum description length (Lu et al., 2010), higher-order moments (Della-Marta and Wanner, 2006; Toreti et al., 2010b), wavelet power spectrum (Li et al., 2013) and Bayesian analysis (Hannart and Naveau, 2009) are proposed to detect artificial changes in climatic time series. Finding these types of change points in the time series, which also have long-term trends, is very difficult (Gallagher et al., 2013).

All the aforementioned methods are applied only to one-dimensional time series. In addition, the methods based on brute-force are not suitable for longer time series because the search space becomes very large. In fact, time series containing N data points have approximately N^k distinct placements of k change points. Bayesian and Markov methods are based on some statistical assumptions that are not true in general. The precision of kernel methods depends highly on the length of the defined window. In practice, it is very difficult to discriminate between natural (climatic) and artificial (nonclimatic) variations in climatic time series. Conrad (1944) defined *relative homogeneity*, indicating that the variations in weather have similar tendencies over rather large regions. Therefore, neighboring stations should have the same pattern of temperature change. It is very unlikely that artificial change occurred at the same time for most of the sensors (Martínez et al., 2010). As a result, an algorithm to find breakpoints in climatic time series that processes several stations *in parallel* and also imposes minimal assumptions on the observed data is attracting ever-increasing attention in modern environmental studies.

In this paper, we study the change detection problem from the point of view of time series clustering. An important characteristic of climatic time series is nonstationarity where statistical properties of time series change under time translations. Assuming that the time series has a mathematical model with time-dependent parameters, a cluster (regime, phase or segment) is defined as a period such that in each of these periods, the model parameters are constant. The problem of finding these clusters and change points between them is called *non-stationary time series clustering* (Horenko, 2010a).

Time series clustering based on the finite element method (FEM) was recently introduced in Horenko (2010a), which can find the clusters without any statistical assumptions. Through this method, a multidimensional time series is processed to find clusters such that the change points between clusters are present in all dimensions at the same time. In FEM clustering, an optimization problem for clustering is

defined. Using Tikhonov regularization and FEM, this optimization is converted into a familiar *linear quadratic programming* (LQP) and a least mean square problem. Iteratively, these two sets of optimizations are solved. Assuming a different linear trend (Horenko, 2010b) and also a differential equation in each cluster (Horenko, 2010c), FEM clustering was used to analyze climate dynamics in ERA-40 reanalysis time series by finding fast switched clusters.

In this paper, FEM clustering is used to find long-term trends in sensors data in the form of time series. Section 2 describes basic definitions and the procedure of the FEM-clustering algorithm in detail. For the special case of trend detection, necessary equations are derived and it is explained how the optimal number of clusters and length of clusters can be chosen. In Sect. 3, the algorithm is tested for a generated time series. In Sect. 4, the method is applied to mean temperature time series of North Carolina (NC) consisting of 249 stations. The results represent a spatio–temporal pattern of change in the area of study. In Sect. 5 the relationship between simulation results and AMO and solar activity is presented, which shows the relationship between detected clusters and these indices.

2 Finite element method for time series clustering

Modeling is the process of finding an appropriate parametric model to explain the observed phenomenon accurately. Let x_t be an observed m-dimensional time series defined over period of time $[0, T]$. To describe this time series with partially linear trends, we assume a model in each cluster as $x_t = c_{0k} + c_{1k} \times t + \text{noise}$, where k is the index of cluster and $k = 1, \ldots, K$. Now define the distance from the observation x_t and the linear trend model.

$$d(x_t, c_k) = \|x_t - (c_{0k} + c_{1k} \times t)\|^2 \tag{1}$$

The problem of model approximation is treated as the optimization for c_k. Assuming the time series is available on $t = 1, \ldots, T$, the optimization problem can be rewritten as

$$\sum_{t=0}^{T} \sum_{k=1}^{K} \mu_k(t).d(x_t, c_k) \xrightarrow{M(t),C} \min \tag{2}$$

where $C = [c_1, \ldots, c_K]$ are the set of cluster parameters means the linear trend coefficients. $M(t) = [\mu_1(t), \ldots, \mu_K(t)]$ are the cluster membership functions that show how much the data in time t belong to cluster k, which fulfills these constraints (Horenko, 2010b):

$$\mu_k(t) \geq 0; \quad \sum_{k=1}^{K} \mu_k(t) = 1 \tag{3}$$

Solving the optimization in Eq. (2) is not possible, since the number of unknowns is more than the number of known variables. Different values of initial parameters may lead to different local minimums. Hence the problem is *ill-posed* in the sense of *Hadamard* and thus requires *regularization*. For example in Tikhonov regularization, the additional assumption (called the regularization term) is included in the minimization problem (Aster and Thurber, 2012). In FEM clustering, it is assumed that the cluster membership functions $\mu_k(t)$ are smooth and that their derivatives are bounded. For a given time series, this constraint limits the total number of switchings between the clusters and adds *persistency* to $\mu_k(t)$ (Horenko, 2010a). Since $\mu_k(t)$ is defined for all the times in $[0, T]$, it belongs to infinite-dimensional *Hilbert space*. Thus the smoothness assumption in Hilbert space is (Poznyak, 2008)

$$\left\| \frac{\partial \mu_k(t)}{\partial t} \right\|_{\mathbf{H}(0,T)} = \int_0^T (\frac{\partial \mu_k(t)}{\partial t})^2 dt < +\infty. \tag{4}$$

The new optimization problem using Tikhonov regularization can be defined as in Eq. (5). Tikhonov's regularized problem can be used as a trade-off among the objective function (first term) and the penalty function (second term), and this trade-off is controlled by $\delta > 0$. Here δ is the *regularization factor* that controls the persistency of the clusters (Horenko, 2010a).

$$\sum_{k=1}^{K} \sum_{t=0}^{T} \left[\mu_k(t).d(x_t, c_k) + \delta \left(\frac{\partial \mu_k(t)}{\partial t} \right)^2 \right] \xrightarrow{M(t),C} \min \tag{5}$$

Unfortunately there is no analytical solution for Eq. (5), but it can be solved using the subspace iteration procedure. The optimization can be solved iteratively with respect to two sets of unknown parameters ($M(t)$ and C) by solving one and substituting in the other one. In order to solve the optimization numerically with respect to $M(t)$, it should be converted from a continuous-time to a discrete-time domain. For this reason, in FEM clustering, the Galerkin method is utilized for discretization. Galerkin methods can convert a continuous operator problem (such as a differential equation) to a discrete problem. They are widely used in the FEM literature to solve differential equations (Zienkiewicz et al., 2013). Assuming $\mu_k(t)$ is defined over $[0, T]$, we define a set of N continuous functions (α_n) called FEM-basis over the support of $\mu_k(t)$. The FEM-basis functions are linearly independent in Hilbert space and have local support on $[0, T]$. Then mapping equations for $\mu_k(t)$ using FEM-basis functions can be written as Eqs. (6) and (7). In fact, $\mu_k(t)$ is mapped to a space spanned by the FEM-basis, where $\tilde{\mu}_k^{(n)}$ are scalars called *Galerkin coefficients*. In our problem, the function $\mu_k(t)$ is unknown (like in differential equations, etc.). By determining Galerkin coefficients and using FEM-basis functions, one can build an approximated version of functions $\mu_k(t)$. The FEM-basis functions here are defined in the form of N triangular functions each one called a *hat function* as in Eq. (8)

and Fig. 1 (Horenko, 2010a):

$$\mu_k(t) = \tilde{\mu}_k(t) + \text{error} = \frac{1}{\Delta}\sum_{n=1}^{N}\tilde{\mu}_k^{(n)}.\alpha_n(t) + \text{error} \quad (6)$$

$$\tilde{\mu}_k^{(n)} = \sum_{t=0}^{T}\mu_k(t)\cdot\alpha_n(t); \quad k=1,\ldots,K; \quad n=1,\ldots,N \quad (7)$$

$$\alpha_n(t) = \begin{cases} \frac{t_2-t}{\Delta} & n=1, t\in[t_1,t_2] \\ \frac{t-t_{n-1}}{\Delta} & 2\leq n\leq N-1, t\in[t_{n-1},t_n] \\ \frac{t_{n+1}-t}{\Delta} & 2\leq n\leq N-1, t\in[t_n,t_{n+1}] \\ \frac{t-t_{N-1}}{\Delta}, & n=N, t\in[t_{N-1},t_N] \end{cases} \quad (8)$$

Applying Eqs. (6), (7) and (8) in Eq. (5) and using the locality of FEM-basis functions support, one can find an optimization after some mathematical simplification in the form of Eqs. (9), (10), (11) and (12). **H** is a tri-diagonal matrix called *stiffness matrix*. Constraints on the cluster membership functions are converted to a set of constraints on Galerkin coefficients as Eq. (13) (Horenko, 2010a):

$$\tilde{L}^\varepsilon = \sum_{k=1}^{K}[\beta(c_k)^T\overline{\mu}_k + \delta\ \overline{\mu}_k^T\ \mathbf{H}\ \overline{\mu}_k] \xrightarrow{\overline{\mu}_k,C} \min \quad (9)$$

$$\beta(c_k) = [\sum_{t_1}^{t_2}\alpha_1(t).d(x_t,c_k)\ldots\sum_{t_{n-1}}^{t_{n+1}}\alpha_n(t).d(x_t,c_k) \quad (10)$$

$$\ldots\sum_{t_{N-1}}^{t_N}\alpha_N(t).d(x_t,c_k)]$$

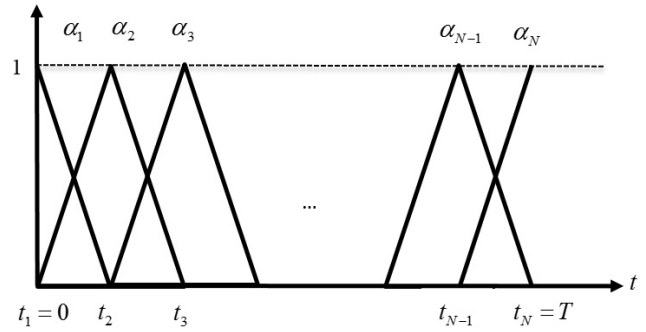$$\overline{\mu}_k = [\tilde{\mu}_k^{(1)},\ldots\tilde{\mu}_k^{(n)},\ldots,\tilde{\mu}_k^{(N)}] \quad (11)$$

$$\mathbf{H} = \begin{bmatrix} \frac{1}{\Delta} & \frac{-1}{\Delta} & 0 & \ldots & 0 & 0 \\ \frac{-1}{\Delta} & \frac{2}{\Delta} & \ddots & 0 & 0 & 0 \\ 0 & \ddots & \ddots & \frac{-1}{\Delta} & 0 & 0 \\ 0 & 0 & \frac{-1}{\Delta} & \frac{2}{\Delta} & \ddots & 0 \\ \vdots & 0 & 0 & \ddots & \ddots & \frac{-1}{\Delta} \\ 0 & 0 & 0 & 0 & \frac{-1}{\Delta} & \frac{1}{\Delta} \end{bmatrix} \quad (12)$$

$$\sum_{k=1}^{K}\tilde{\mu}_k^{(n)} = 1; \quad \tilde{\mu}_k^{(n)}\geq 0; \quad (13)$$

$$\forall n=1,\ldots,N; \quad k=1,\ldots,K$$

The minimization problem of Eq. (9) is in the form of a

Figure 1. Finite element basis functions are in the form of an N hat function. Increasing N makes the algorithm more computationally expensive, but helps to find shorter clusters.

summation of K linear quadratic programming (LQP) optimizations. To solve this optimization, all $\overline{\mu}_k$ are augmented in a vector λ to make one optimization in the form of

$$\frac{1}{2}\lambda^T G\lambda + B^T\lambda \xrightarrow{\lambda} \min \quad (14)$$

where

$$\lambda = [\overline{\mu}_1,\ldots,\overline{\mu}_k,\ldots\overline{\mu}_K]^T \quad (15)$$

$$B = [\beta(c_1),\ldots,\beta(c_k),\ldots\beta(c_K)] \quad (16)$$

$$G = 2\times\delta\begin{bmatrix} \mathbf{H} & 0 & \ldots & 0 \\ 0 & \mathbf{H} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \mathbf{H} \end{bmatrix} \quad (17)$$

As a result, the optimization constraints are converted to

$$\lambda_i \geq 0 \text{ for } i=1,\ldots,N\times K \quad (18)$$

$$F.\lambda = Q \quad (19)$$

$$F = [\ \underbrace{I_{N\times N} \quad I_{N\times N} \quad \ldots \quad I_{N\times N}}_{K\ \text{times}}\ ] \quad (20)$$

$$Q = [\ 1 \quad 1 \quad \ldots \quad 1\ ]_{N\times 1}^T \quad (21)$$

The resulting optimization problem is a large sparse quadratic programming, which includes constraints in the form of equalities and inequalities. In the past, different techniques such as the *interior point* method and the *active set* method have been applied to solve the LQP (Boyd and Vandenberghe, 2004). The large sparse problem in this study is solved by the toolbox developed in Gurobi (2014). This toolbox is specially designed to solve large sparse optimization problems and has options to choose among several solvers.

Assuming that $\mu_k(t)$ is known, one can solve the optimization with respect to C analytically using the least mean square. Note that the second term in Eq. (5) is not a function of C and thus can be eliminated in this step. Based on the

distance functional defined in Eq. (1),

$$
c_{0k} = \frac{\begin{vmatrix} \sum\limits_{t=0}^{T} x_t . \mu_k(t) & \sum\limits_{t=0}^{T} t . x_t . \mu_k(t) \\ \sum\limits_{t=0}^{T} t . x_t . \mu_k(t) & \sum\limits_{t=0}^{T} x_t^2 . \mu_k(t) \end{vmatrix}}{\begin{vmatrix} \sum\limits_{t=0}^{T} \mu_k(t) & \sum\limits_{t=0}^{T} t . \mu_k(t) \\ \sum\limits_{t=0}^{T} t . \mu_k(t) & \sum\limits_{t=0}^{T} t^2 . \mu_k(t) \end{vmatrix}} ; \tag{22}
$$

$$
c_{1k} = \frac{\begin{vmatrix} \sum\limits_{t=0}^{T} \mu_k(t) & \sum\limits_{t=0}^{T} x_t . \mu_k(t) \\ \sum\limits_{t=0}^{T} t . \mu_k(t) & \sum\limits_{t=0}^{T} t . x_t . \mu_k(t) \end{vmatrix}}{\begin{vmatrix} \sum\limits_{t=0}^{T} \mu_k(t) & \sum\limits_{t=0}^{T} t . \mu_k(t) \\ \sum\limits_{t=0}^{T} t . \mu_k(t) & \sum\limits_{t=0}^{T} t^2 . \mu_k(t) \end{vmatrix}}
$$

The general procedure for solving optimization problems is shown in Table 1. K, δ, and Δ are the numbers of clusters, the regularization factor, and finite discretization of time interval (width of the hat functions), respectively. It should be set at the beginning of the procedure. Changing Δ has an effect on the accuracy of clustering and the computational processing. To cluster a time series with K clusters and using N hat functions, the size of the vector to be minimized (λ) in Eq. (14) will be $N \times K$. Decreasing Δ will increase the number of FEM basis functions, and this leads to an increase in the size of LQP and, consequently, the volume of the computations. On the other hand, the accuracy of the algorithm will increase and help detect very short clusters. A more challenging problem arises when choosing K. A trivial solution exists when all the lines connecting every two points are in separate clusters. By increasing K, the value of the cost function always decreases and when $K = length\ of\ time\ series - 1$, this value approaches zero. As a result it is not possible to find the *optimal* number of clusters. Therefore, we assume an upper bound on the number of clusters. Trial and error and human judgment are used to select K subjectively. The criterion for choosing K is based on the assumption that clusters are deterministic. This means that each datum in the time series belongs to only one of the clusters at each time. As a result, the values of $\mu_k(t)$ should be about 1 or 0. When the $\mu_k(t)$ is 1, it means that the datum in time t completely belongs to cluster k. When it is 0, it means that this datum does not belong to cluster k. One important issue in the time series clustering problem is the length of the clusters. For a fixed number of clusters, switching may occur many times or just a few times between the clusters. If the number of switchings between clusters increases, the duration in which

Table 1. This procedure should be followed to solve clustering optimization and find two sets of known parameters: cluster membership functions and linear trend parameters in each cluster.

Step	Length
1	Set K the number of clusters, δ the regularization factor, Δ the width of hat functions.
2	Iteration number $L = 1$
3	Choose random initial λ^{initial} satisfying related constraints.
4	Solve the minimization problem with respect to C for fixed λ^{initial} to find C^{initial}.
5	Solve the minimization problem with respect to λ for fixed $C^{[L]}$ to find $\lambda^{[L+1]}$.
6	Solve the minimization problem with respect to C for a fixed $\lambda^{[L+1]}$ to find $C^{[L+1]}$.
7	Go to step 5 and continue until the pre-defined number of iterations.

the system stays in each of the clusters decreases. To investigate low-frequency variability in the system, small switching numbers are desirable. In contrast, for high-frequency variability we are interested in many transitions between clusters. It is shown that by increasing δ, clusters will become more *metastable* and the number of transitions between the clusters will be decreased. In fact, the observed process will stay for longer times in the identified metastable cluster before making a transition to the next identified cluster (Horenko, 2010a, b). Horenko (2010b) used the value of $\delta = 0$ to find less-metastable regimes in analyzing climate dynamics.

In summary, to assign clusters number K, we first assign a large $K \ll$ length of the time series and find clusters and $\mu_k(t)$ for different values of δ. If any of the cluster membership functions has values between 0 and 1, we decrease K and repeat the procedure until the cluster membership functions for all of the clusters becomes either about 0 or 1. We continue this procedure to find the largest K with a value of δ such that all of the clusters are deterministic.

3 Synthetic data testing

In order to test the algorithm and observe the effects of mean shift and outliers, we generated a two-dimensional time series with three clusters. Noise with mean value 0 and variance 4 is added to the time series. The models for the time series
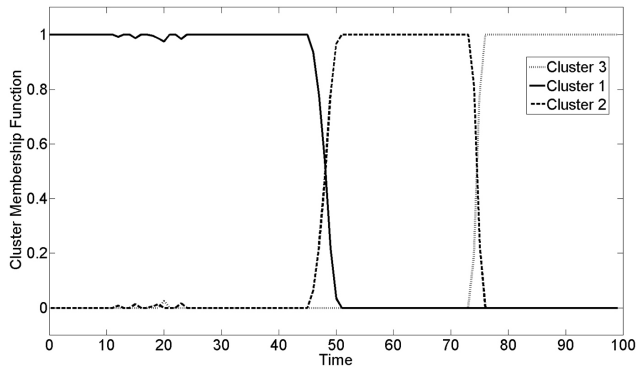
Figure 2. Cluster membership functions for synthetic data when $K = 3$ and $\delta = 1$. It can be seen that the clusters are found correctly.
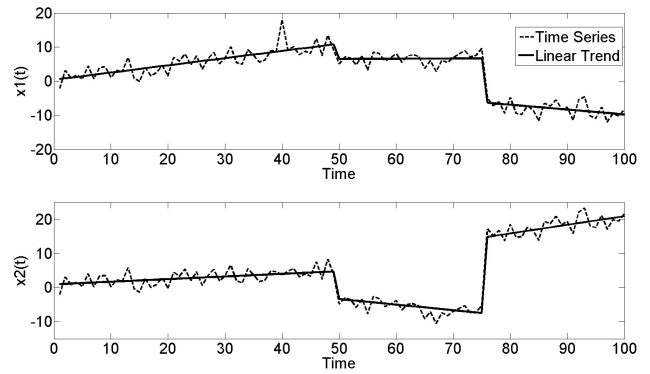
Figure 3. Detected clusters and linear trends for synthetic data when $K = 3$ and $\delta = 1$. Results show good clustering.

are

$$x_1(t) = \begin{cases} 0.2t + 0.1 & t \in [0, 50) \\ 0.1t - 0.1 & t \in [50, 75) \\ -0.1t + 0.2 & t \in [75, 100] \end{cases}$$

$$x_2(t) = \begin{cases} 0.3t + 0.4 & t \in [0, 50) \\ 0.2t + 0.1 & t \in [50, 75) \\ -0.1t - 0.4 & t \in [75, 100]. \end{cases} \quad (23)$$

Level shifts equal to 0.5 and 0.6 are added to $x_1(t)$ at $t = 20$ in the first cluster and to $x_2(t)$ in the third cluster at $t = 75$, respectively. In order to simulate an outlier the value of $x_1(t)$ at $t = 40$ is set to 18. The cluster membership function is shown in Fig. 2, where $\delta = 4$ and $K = 3$. Here, the change points are found correctly in the presence of a level shift and outlier since mean-shift changes and outliers are not at the same time in the two dimensions. Figure 3 shows detected trends in two dimensions of the original time series.

Now assume a case in which we set $K = 4$. The new values of $\mu_k(t)$ are shown in Fig. 4. In this figure, cluster membership functions for the first and fourth clusters are about either 0 or 1. Thus, these two clusters are found correctly, but the cluster membership functions for the second and third clusters are about 0.5 in the interval [50, 75] and 0 elsewhere. This means that clusters 2 and 3 are not *separable* and that K must be decreased from 4 to 3.

To observe the effect of the regularization factor δ on the clustering result, set $K = 3$ and increase the value of δ from 1 to larger numbers. The FEM clustering does not produce the deterministic clusters. It means that the values of $\mu_k(t)$ are always between 0 and 1. On the other hand, if we decrease the number of clusters and set $K = 2$, an acceptable result can be found when $\delta = 7$, as shown in Fig. 5. It is clear that the clusters [0, 50] and [50, 75] are merged together to create a new cluster. The second cluster is between [75, 100], which is found correctly. Thus, new clusters are found with a small number of switchings between them and with longer duration. In conclusion, in FEM clustering of time series, it is necessary to repeat clustering procedures with different values of K and δ until acceptable results are found.
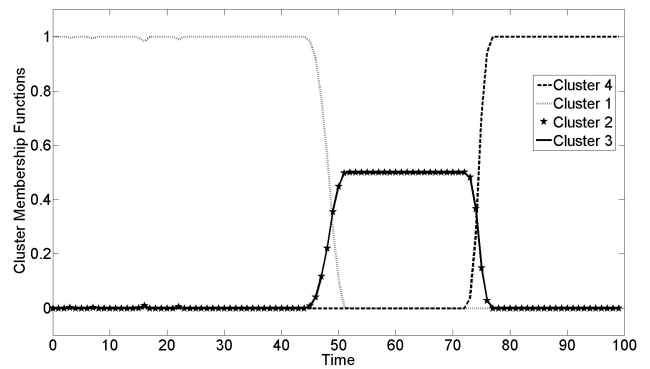

Figure 4. Cluster membership functions for synthetic data when $K = 4$; $\mu(t)$ for the first and fourth clusters are about 0 or 1 and thus these clusters are found correctly, but $\mu(t)$ for the second and third clusters are equal to about 0.5 in time domain [50, 75] and 0 elsewhere. This means that clusters two and three are not separable and that K must be decreased from 4 to 3.

4 Application: NC temperature trend analysis

In this paper, we study climate clusters/trends in NC. Boyles and Raman (2003) and Bililign et al. (2012) analyzed temperature trends in NC and showed the effect of global warming. The state of NC is located in the southeastern United States (75°30′–84°15′ W, 34°–36°21′ N). The study area covers approximately 52 664 square miles (136 399 km^2), encompassing 100 counties. A data set of average temperatures in 249 stations across NC is analyzed for the period beginning in 1950 until the end of October 2009. Figure 6 shows the locations of measuring stations. Daily temperatures are collected from the United States Department of Agriculture-Agriculture Research Service. This data set was facilitated by the National Oceanic and Atmospheric Administration (NOAA). These data contain quality control information from the Cooperative Observer Program (COOP) network and the Weather Bureau Army Navy (WBAN) and NOAA agencies.
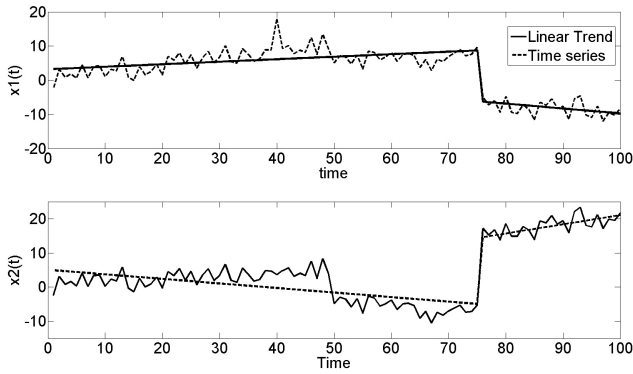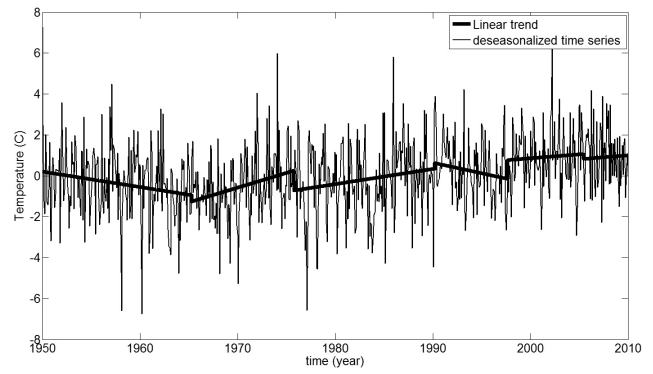
Figure 5. Result of clustering for synthetic data when $K = 2$ and $\delta = 7$. The first and second clusters are merged together. Increasing the δ leads to an answer with a smaller number of switching.
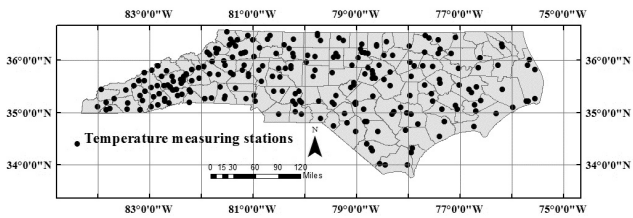
Figure 6. The state of NC is located in the southeastern United States. The study area covers an area of approximately $136\,399\,\text{km}^2$. A data set of average temperatures at 249 stations across NC is analyzed for the period including the beginning of 1950 till the end of October 2009.

Figure 7. Time series at one station and its clusters/trends are shown. Thin lines show deseasonalized temperature time series and thick lines are linear trends found by the FEM algorithm. All 249 dimensions of the time series are analyzed at the same time, but the figure only shows the result at one of these stations.

Table 2. By increasing δ, clusters will become more metastable and the number of transitions between the clusters will be decreased. In fact, the observed process will stay for longer and longer times in the identified metastable cluster before making a transition to the next identified cluster.

Regularization factor	0.01	1	10	25	60	80	100
Number of switching	50	40	24	18	6	5	4

The data are examined for consistency and missing information. About 0.1 % of the data is missing and filled with cubic interpolation. Daily data are averaged to derive monthly data to eliminate high-frequency harmonics and also reduce the complexity of calculations. Monthly average is generally preferred over daily data due to the fact that this time frame is longer than the period of most large-scale synoptic waves in the troposphere (Nigam, 2003). The result is a 249×718 dimension time series. In order to find the trends, the annual cycle has been removed by subtracting the multi-year monthly means.

In applying FEM clustering to the time series, the value of K is initially assumed to be 10 and the algorithm is executed for different values of the regularization parameter δ. The regularization parameter is a real value larger than 0. For each δ, the algorithm ran several times to avoid convergence to local minimums in the optimization problem. When we reach the largest K such that all the $\mu(t)$ are about 0 or 1, an optimal solution is found. The optimal solution for this particular time series is achieved when $K = 6$, $\delta = 80$ and $\Delta = 4$. This result represents six clusters with different trends varying in magnitude and direction (positive or negative) with different length in NC. Table 2 shows the effect of δ on the number of switchings between clusters. As expected, the number of switchings decreases by decreasing the δ, while the length of the clusters increases. Once the linear trends are obtained, the *Mann–Kendall* trend approach can be used to test the statistical significance of trends (Modarres and Sarhadi, 2009).

Figure 7 shows the linear trends detected for the deseasonalized monthly time series at one of the stations. Note that all of the 249 dimensions of the time series are analyzed in batch by applying the FEM-clustering algorithm, and this figure only shows results at one of these stations. Figure 8 represents the linear trends for all 249 stations. Table 3 shows the average temperature, the average change in temperature and the corresponding percentage change in temperature over 60 years. The average trend is calculated by averaging between trends at all 249 stations. The average temperature change is found by multiplying the average trend by the length of the cluster. Percentage change is calculated by average temperature change divided by the average temperature during that cluster. Figure 9a–f shows the map corresponding to the spatial distribution of linear trends in NC. Twelve colors are used in *ArcGIS 10.0* to sketch the spatial distribution of trends in maps; contours are generated using an inverse distance weighted (IDW) algorithm with a power of 2.0, 12 grid points and variable radius location (Tewolde et al., 2010).

A closer observation of annual linear trends of clusters at all of the stations (Fig. 8), the spatial interpolated annual
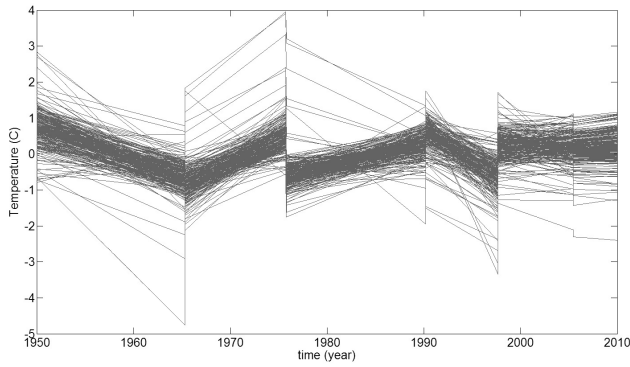
Figure 8. Linear trends and clusters at 249 stations during 60 years are shown. There are two notable decreasing trends between 1950–1964 and 1990–1998 and also a remarkable increase in 1964–1976. Other values of trends are smaller.

Table 3. Here are shown patterns of climate change in 60 years. They consist of the average temperature, the average change in temperature and the corresponding percentage change in temperature over 60 years. Average trend is calculated by averaging between trends at all 249 stations. Average temperature change is found by multiplying average trend by the length of the cluster. Percentage change is calculated by average temperature change divided by the average temperature during that cluster.

Clusters	Length	Average trend (°C per month)	Average temperature change (°C)	Percent change
Cluster 1	1950–1965	−0.0077	−1.42	−9.8
Cluster 2	1965–1976	0.0107	1.34	9.1
Cluster 3	1976–1990	0.0047	0.83	5.8
Cluster 4	1990–1998	−0.0133	−1.20	−8.0
Cluster 5	1998–2005	−0.0005	−0.05	−0.3
Cluster 6	2005–2009	0.0011	0.06	0.4

trends of NC (Fig. 9) and the corresponding Table 3 clearly shows two notable decreasing trends during 1965–1964 and 1990–1998, and a remarkable increase in temperature during 1964–1976. The longest and shortest periods of 15 and 5 years were found in cluster 1 and cluster 6, respectively. Also, clusters 2 and 4 show the warmest and coolest trends, respectively. Figure 9a corresponds to cluster 1, with the coolest trend statewide. After 1965, temperature in NC trends shifted to a positive trend (i.e., warmer). The coastal zone shows the highest warming trend in cluster 2 during the period 1950–1976, which indicates the greatest warming trends in 60 years (Fig. 9b). Cluster 3 (Fig. 9c) has a positive trend almost equally in all of the regions. Cluster 4 (Fig. 9d) has a negative trend in all areas except Piedmont, in which the amount of cooling is more significant. Cluster 5 (Fig. 9e) has a positive trend in the Piedmont area and a negative trend in the coastal and mountainous area. Cluster 6 shows the very mild warming and cooling trends statewide (Fig. 9f).
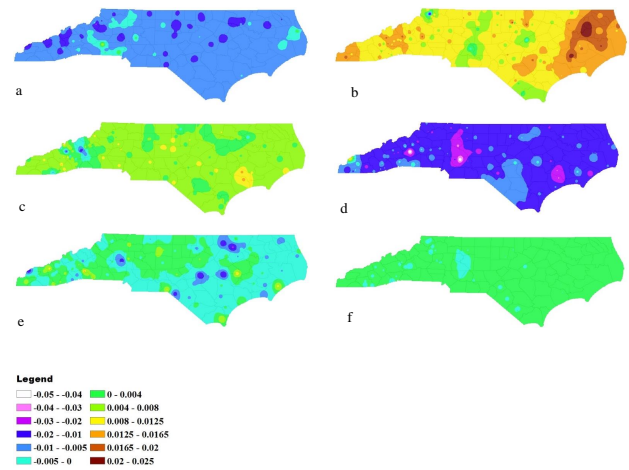

Figure 9. Six clusters with linear trends found in NC. **(a)** A notable negative trend between 1950 and 1965. **(b)** A positive trend after 1965. The coastal zone shows the highest warming trend. **(c)** Smaller positive trend all over the state between 1976 and 1990. **(d)** Negative trend in the 1990s with about 8 % decrease in temperature. **(e)** Positive trend in Piedmont and negative trend in other parts. **(f)** Most areas of the state have mild positive trends.

5 Comparison with climatic indices

In this section, the relationship between some climate indices and the results based on our proposed clustering method are discussed. Among the many available indices, we particularly analyzed the Atlantic Multi-decadal Oscillation (AMO) and the sunspot cycle. Investigation of these indices reveals that there exists some relation between temperature in NC and these indices.

The AMO is a pattern of multi-decadal variability in surface temperature centered in the North Atlantic Ocean. The index shows persistent warming and cooling phases that typically last a few decades. AMO has been linked with the occurrence of Sahel drought, variability in northeastern Brazilian rainfall, North American climate, river flows, and the frequency of Atlantic hurricanes (Knight et al., 2005). Figure 10 shows the AMO index time series (Schlesinger and Ramankutty, 1994; Enfield et al., 2001). FEM clustering was applied to the AMO time series using $\delta = 1$ and found four clusters and linear trends for the period spanning 1950–2010, which are shown in Fig. 10. The sunspot number is the sum of the number of individual sunspots plus ten times the number of sunspot groups. The number of sunspots visible on the sun, with waxes and wanes, has an approximate 11-year cycle (Frame and Gray, 2010). The total solar irradiance is measured from dark sunspots and bright faculae (Lean, 2009). Figure 11 shows the change of solar irradiance for the period of 1950–2010 (Lean, 2009; Hathaway, 2010). The annual climate trends in NC could be related to the AMO/solar activity.
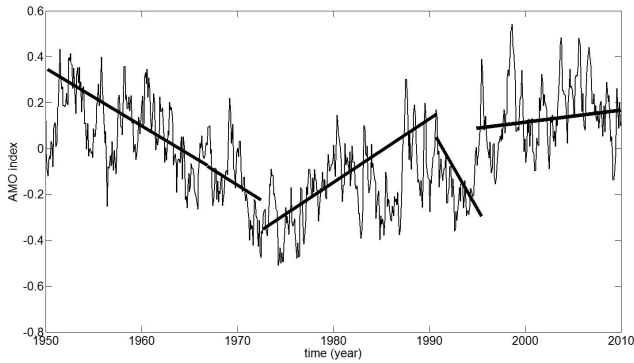
Figure 10. AMO is a pattern of multi-decadal variability in surface temperature centered in the North Atlantic Ocean. The index shows persistent warm and cool phases typically lasting a few decades. The figure shows the AMO index for the period of 1950–2010 and also its linear trends (thick lines).

– There is a cooling trend during 1950–1965, which coincides with the AMO dropping period in Fig. 10. As shown in Table 3, the ensemble average of trends in NC temperature is equal to -0.007 ($^{\circ}$C month^{-1}) and trends in the AMO index -0.001 ($^{\circ}$C month^{-1}).

– During 1965–1970 (cluster 2), the trend is positive. However, during this period, the AMO index continued to decrease, which is in contradiction with the positive trend in cluster 2. Nevertheless, comparing the trend in cluster 2 with solar activity, as in Fig. 11, shows that during this period solar activity increases, which coincides with the temperature increase in cluster 2. This comparison shows that the temperature increase could be due to an increase in solar activity. The ensemble average trend in NC temperature is 0.0107 ($^{\circ}$C month^{-1}) and the trend in AMO is -0.007 ($^{\circ}$C month^{-1}). After 1970, the AMO index has increased and at the same time the NC climate in cluster 2 has risen, too.

– During 1976–1990, NC climate is in cluster 3, with a smaller trend compared to cluster 2. We can observe that during the same period, the AMO index is negative but increasing with a positive trend. The AMO trend is 0.001 ($^{\circ}$C month^{-1}) and the ensemble average trend in NC temperature is 0.0047 ($^{\circ}$C month^{-1}).

– There is a negative trend in cluster 4 during 1990–1998. During this period the AMO index and solar activity decreases. Trends in AMO, solar irradiance and the NC temperature are -0.007 ($^{\circ}$C month^{-1}), -0.0012 (watt m^{-2} $\times$ month) and -0.0133 ($^{\circ}$C month^{-1}), respectively. Thus, the negative trend in temperature may be due to the negative trends in the AMO and the solar irradiation.
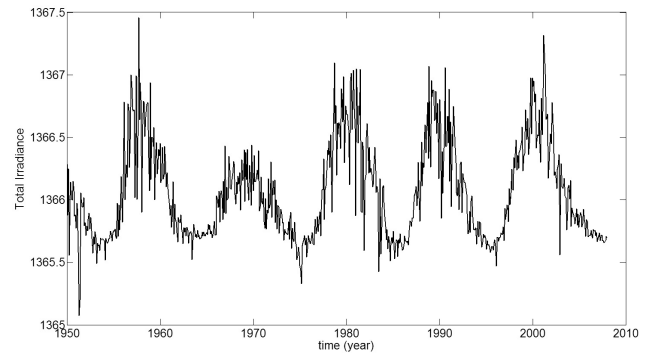
Figure 11. The total solar irradiance is measured from dark sunspots and bright faculae. This figure shows the change in solar irradiance for the period of 1950–2010 in Watts m^{-2}.

– Clusters 5 and 6 in NC have a very small negative and positive trend during 1998–2009. The AMO index oscillates around 0.2 ($^{\circ}$C month^{-1}) in this period and does not have a notable trend.

6 Conclusions

In this paper, multiple change points in the trend of multidimensional climatic time series are found using FEM-based clustering. Since the time series of all stations are processed at the same time, the result of clustering in this study only shows climate changes due to natural phenomena. FEM clustering is applied to the time series with different linear trends in each cluster to find change points. After setting up an appropriate cost function, FEM clustering utilized Tikhonov regularization and Galerkin discretization to convert the optimization problem into a linear quadratic and a least mean square problem. By trial and error, the optimal number and length of clusters can be determined in FEM clustering. A climatic time series of NC is analyzed by this method. The results represent a spatio–temporal pattern of climate change corresponding to the areas of study. Then, some relations between clusters and climatic indices were discussed. It appears that the natural variability of climate change in NC during 1950–2009 can be explained mostly by AMO and solar activity.

References

Aster, R. C. and Thurber, C. H.: Parameter estimation and inverse problems, Academic Press, Waltham, MA, 2012.

Beaulieu, C., Chen, J., and Sarmiento, J. L.: Change-point analysis as a tool to detect abrupt climate variations, Philos. Trans. R. Soc. Math. Phys. Eng. Sci., 370, 1228–1249, 2012a.

Beaulieu, C., Sarmiento, J. L., Mikaloff Fletcher, S. E., Chen, J., and Medvigy, D.: Identification and characterization of abrupt changes in the land uptake of carbon, Glob. Biogeochem. Cy., 26, GB1007, doi:10.1029/2010GB004024, 2012b.

Bililign, S., Lin, Y.-L., Davis, R., Ilias, S., Kurkalova, L., Kyei, Y., Rastigejev, Y., Uzochukwu, G., and Bae, S.: Effects of global warming on North Carolina, Int. J. Clim. Change Impacts Responses, 3, 53–70, 2012.

Boyd, S. P. and Vandenberghe, L.: Convex optimization, Cambridge University Press, Cambridge, UK, 2004.

Boyles, R. P. and Raman, S.: Analysis of climate trends in North Carolina (1949–1998), Environ. Int., 29, 263–275, 2003.

Conrad, V.: Methods in climatology, Harvard University Press, Cambridge, MA, 1944.

Della-Marta, P. M. and Wanner, H.: A method of homogenizing the extremes and mean of daily temperature measurements, J. Climate, 19, 4179–4197, 2006.

Dose, V. and Menzel, A.: Bayesian analysis of climate change impacts in phenology, Global Change Biol., 10, 259–272, 2004.

Ehsanzadeh, E., Ouarda, T. B. M. J., and Saley, H. M.: A simultaneous analysis of gradual and abrupt changes in Canadian low streamflows, Hydrol. Process, 25, 727–739, 2011.

Enfield, D. B., Mestas-Nuñez, A. M., and Trimble, P. J.: The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental US, Geophys. Res. Lett., 28, 2077–2080, 2001.

Frame, T. H. A. and Gray, L. J.: The 11-Yr solar cycle in ERA-40 data: An update to 2008, J. Climate, 23, 2213–2222, 2010.

Gallagher, C., Lund, R., and Robbins, M.: Changepoint detection in climate time series with long-term trends, J. Climate, 26, 4994–5006, 2013.

Godtliebsen, F., Holmström, L., Miettinen, A., Erästö, P., Divine, D. V., and Koc, N.: Pairwise scale space comparison of time series with application to climate research, J. Geophys. Res.-Ocean., 117, C03046, doi:10.1029/2011JC007546, 2012.

Gurobi Optimization: Gurobi optimizer reference manual, available at: http://www.gurobi.com (last access: 15 January 2014), 2014.

Hannart, A. and Naveau, P.: Bayesian multiple change points and segmentation: Application to homogenization of climatic series, Water Resour. Res., 45, W10444, doi:10.1029/2008WR007689, 2009.

Hathaway, D. H.: The solar cycle, Living Rev. Sol. Phys., 7, doi:10.12942/lrsp-2010-1, 2010.

Horenko, I.: Finite element approach to clustering of multidimensional time series, SIAM J. Sci. Comput., 32, 62–83, 2010a.

Horenko, I.: On clustering of non-stationary meteorological time series, Dynam. Atmos.-Ocean., 49, 164–187, 2010b.

Horenko, I.: On the Identification of Nonstationary Factor Models and Their Application to Atmospheric Data Analysis, J. Atmos. Sci., 67, 1559–1574, 2010c.

IPCC: Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 2013.

Karl, T. R., Knight, R. W., and Baker, B.: The record breaking global temperatures of 1997 and 1998: Evidence for an increase in the rate of global warming?, Geophys. Res. Lett., 27, 719–722, 2000.

Kehagias, Ath. and Fortin, V.: Time series segmentation with shifting means hidden markov models, Nonlin. Processes Geophys., 13, 339–352, doi:10.5194/npg-13-339-2006, 2006.

Kiely, G.: Climate change in Ireland from precipitation and streamflow observations, Adv. Water Resour., 23, 141–151, 1999.

Knight, J. R., Allan, R. J., Folland, C. K., Vellinga, M., and Mann, M. E.: A signature of persistent natural thermohaline circulation cycles in observed climate, Geophys. Res. Lett., 32, L20708, doi:10.1029/2005GL024233, 2005.

Lean, J.: Calculations of solar irradiance: monthly means from 1882 to 2008, available at: http://www.geo.fu-berlin.de/en/met/ag/strat/forschung/SOLARIS/Input_data/Calculations_of_Solar_Irradiance.pdf, 2009.

Li, Z., Yan, Z., Cao, L., and Jones, P.: Adjusting inhomogeneous daily temperature variability using wavelet analysis, Int. J. Climatol., 34, 1196–1207, doi:10.1002/joc.3756, 2013.

Liu, R. Q., Jacobi, C., Hoffmann, P., Stober, G., and Merzlyakov, E. G.: 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, D22105, doi:10.1029/2010JD014080, 2010.

Lu, Q., Lund, R., and Lee, T. C. M.: An MDL approach to the climate segmentation problem, Ann. Appl. Stat., 4, 299–319, 2010.

Lyubchich, V., Gel, Y. R., and El-Shaarawi, A.: On detecting non-monotonic trends in environmental time series: a fusion of local regression and bootstrap, Environmetrics, 24, 209–226, 2013.

Martínez, M. D., Serra, C., Burgueño, A., and Lana, X.: Time trends of daily maximum and minimum temperatures in Catalonia (ne Spain) for the period 1975–2004, Int. J. Climatol., 30, 267–290, 2010.

Modarres, R. and Sarhadi, A.: Rainfall trends analysis of Iran in the last half of the twentieth century, J. Geophys. Res.-Atmos., 114, D03101, doi:10.1029/2008JD010707, 2009.

Mudelsee, M.: Climate time series analysis classical statistical and bootstrap methods, Springer, Dordrecht, 2010.

Nigam, S.: Teleconnections, in: Encyclopedia of Atmospheric Sciences, 2243–2269, 2003.

Poznyak, A. S.: Advanced mathematical tools for automatic control engineers, Vol. 1, Elsevier, Amsterdam, 2008.

Ruggieri, E.: A Bayesian approach to detecting change points in climatic records, Int. J. Climatol., 33, 520–528, 2013.

Rutishauser, T., Luterbacher, J., Jeanneret, F., Pfister, C., and Wanner, H.: A phenology-based reconstruction of interannual changes in past spring seasons, J. Geophys. Res.-Biogeosci., 112, G04016, doi:10.1029/2006JG000382, 2007.

Scherrer, S. C., Wüthrich, C., Croci-Maspoli, M., Weingartner, R., and Appenzeller, C.: Snow variability in the Swiss Alps 1864–2009, Int. J. Climatol., 3162–3173, 2013.

Schlesinger, M. E. and Ramankutty, N.: An oscillation in the global climate system of period 65–70 years, Nature 367, 723–726, 1994.

Seidel, D. J. and Lanzante, J. R.: An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes, J. Geophys. Res.-Atmos., 109, D14108, doi:10.1029/2003JD004414, 2004.

Seidou, O. and Ouarda, T. B. M. J.: Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows, Water Resour. Res., 43, W07404, doi:10.1029/2006WR005021, 2007.

Sonali, P. and Nagesh Kumar, D.: Review of trend detection methods and their application to detect temperature changes in India, J. Hydrol., 476, 212–227, 2013.

Tabari, H. and Hosseinzadeh Talaee, P.: Analysis of trends in temperature data in arid and semi-arid regions of Iran, Global Planet. Change, 79, 1–10, 2011.

Tewolde, M. G., Beza, T. A., Costa, A. C., and Painho, M.: Comparison of different interpolation techniques to map temperature in the southern region of Eritrea, in: 13th AGILE International Conference on Geographic Information Science, 1–5, 2010.

Tomé, A. R. and Miranda, P. M. A.: Piecewise linear fitting and trend changing points of climate parameters, Geophys. Res. Lett., 31, L02207, doi:10.1029/2003GL019100, 2004.

Tomé, A. R. and Miranda, P. M. A.: Continuous partial trends and low-frequency oscillations of time series, Nonlin. Processes Geophys., 12, 451–460, doi:10.5194/npg-12-451-2005, 2005.

Toreti, A. and Desiato, F.: Temperature trend over Italy from 1961 to 2004, Theor. Appl. Climatol., 91, 51–58, 2008.

Toreti, A., Desiato, F., Fioravanti, G., and Perconti, W.: Seasonal temperatures over Italy and their relationship with low-frequency atmospheric circulation patterns, Clim. Change, 99, 211–227, 2010a.

Toreti, A., Kuglitsch, F. G., Xoplaki, E., Luterbacher, J., and Wanner, H.: A novel method for the homogenization of daily temperature series and its relevance for climate change analysis, J. Climate, 23, 5325–5331, 2010b.

Toreti, A., Kuglitsch, F. G., Xoplaki, E., and Luterbacher, J.: A novel approach for the detection of inhomogeneities affecting climate time series, J. Appl. Meteorol. Climatol., 51, 317–326, 2012.

Verbesselt, J., Hyndman, R., Newnham, G., and Culvenor, D.: Detecting trend and seasonal changes in satellite image time series, Remote Sens. Environ., 114, 106–115, 2010.

Wilks, D. S.: Statistical methods in the atmospheric sciences, Academic Press, Waltham, MA, 2011.

Zienkiewicz, O. C., Taylor, R. L., and Zhu, J. Z.: Finite Element Method: its basis and fundamentals, Butterworth-Heinemann, Amsterdam, 2013.