Non-stationary Time Series Clustering with Application to Climate Systems

Mohammad Gorji Sefidmazgi[1], Mohammad Sayemuzzaman[2], and Abdollah Homaifar[1]

[1] Department of Electrical Engineering,
Autonomous Control and Information Technology Center
[2] Department of Environmental Engineering
North Carolina A&T State University
{mgorjise,msayemuz}@aggies.ncat.edu, homaifar@ncat.edu

Abstract. In climate science, knowledge about the system mostly relies on measured time series. A common problem of highest interest is the analysis of high-dimensional time series having different phases. Clustering in a multi-dimensional non-stationary time series is challenging since the problem is ill-posed. In this paper, the Finite Element Method of non-stationary clustering is applied to find regimes and the long-term trends in a temperature time series. One of the important attributes of this method is that it does not depend on any statistical assumption and therefore local stationarity of time series is not necessary. Results represent low-frequency variability of temperature and spatiotemporal pattern of climate change in an area despite higher frequency harmonics in time series.

Keywords: Non-stationary time series, Time series clustering, spatiotemporal pattern.

1 Introduction

Complexity of climate change limits the knowledge about it and therefore decreases its predictability even over a few days. It is complex because many nonlinear variables within the Earth's atmosphere such as temperature, barometric pressure, wind velocity, humidity, clouds and precipitation are interacting. Analyzing climate system in longer timescales and larger areas and also other parameters which influence climate (Earth's surface, Sun, etc.) is complex too. All of climatic variables are observed in limited number of measurement stations and few times per day and limited accuracy, thus our knowledge is restricted to limited time series. Therefore, methods of time series analysis are important in climate [1]. An important characteristic of climatic time series is the non-stationarity. It means that their statistical properties are changing during time and a unique model cannot represent the time series. A common problem in this field is the analysis of high-dimensional time series containing different phases [2]. We assume that the time series has a dynamical model with some time dependent parameters. Then, define a phase (cluster, regime or segment)

M. Jamshidi et al. (eds.), *Advance Trends in Soft Computing WCSC 2013*,
Studies in Fuzziness and Soft Computing 312,
DOI: 10.1007/978-3-319-03674-8_6,

as a period of time such that during each of these phases, the model parameters are constant. In fact, temporal changes of model parameters take place at a much slower speed than the changes in the system variables themselves. The problem of non-stationary time series clustering is defined to find these regimes numerically [3].

There are many approaches such as Gaussian Mixture Model (GMM) and the Hidden Markov Model (HMM) in literature to detect the phases numerically [4], [5], [6]. In these approaches, one statistical model is assumed for each regime and then the best change points are found by minimizing a cost function in the form of Maximum Likelihood. The cost function is solved by the Expectation Maximization approach.

In this paper, we use a newly developed method based on Finite Elements to find regime changes [7]. The advantage of the FEM method is that it doesn't require any statistical assumption on time series (such as Markovian or Gaussian). It also has intrinsic flexibility to change the persistence of detected regimes. It means it can have longer or shorter regimes by changing some parameters in its procedure. An important characteristic of climatic time series is existence of a linear trend that shows whether variables (for example temperature) are rising or falling. In this work, those regimes are detected in the time series that have different linear trends. When data has a linear trend in each cluster, time series is not locally stationarity and other clustering methods can't solve this problem.

2 Finite Element Method for Clustering

FEM clustering is an approach developed to detect regimes in a non-stationary time series. The basic idea is to assume a model for the time series in each regime, and then find the best switching times and model parameters by solving an optimization problem. This is common in other clustering approaches too. The difference is that the model in each regime can be a non-statistical model. Including additional assumption to the cost function makes it possible to solve this problem using the finite element method. Finally, the minimization problem converted to a linear quadratic programming (LQP) which is solved iteratively to determine the parameters of interest that include the slope and intercept in each regime.

Let x_t be an observed *n*-dimensional time series defined over period of time $[0, T]$. Assuming that we want to fit a first degree polynomial in each regime in the form of $\theta_{0i} + \theta_{1i}.t$ (where *i* is the regime index), we can define *model distance functional*:

$$g(x_t, \theta_i) = \|x_t - (\theta_{0i} + \theta_{1i}.t\,)\|^2 \tag{1}$$

Since time series has *K* regimes with unknown switching times, the overall cost function can be defined as:

$$\sum_{i=1}^{K} \int_0^T \gamma_i(t).\, g(x_t, \theta_i)\, dt \xrightarrow[\Gamma(t),\Theta]{} min \tag{2}$$

where Θ is the time-independent set of unknown parameters and $\gamma_i(t)$ is the cluster affiliation function which are convex and positive. In (2), we have:

$$\Theta = [\theta_1, \dots, \theta_K] \qquad \sum_{i=1}^{K} \gamma_i(t) = 1 \tag{3}$$

$$\Gamma(t) = [\gamma_1(t), \dots, \gamma_K(t)] \qquad \gamma_i(t) \geq 0$$

If all of $\gamma_i(t)$ are 0 or 1 in different times, clusters are *deterministic*. On the other hand if it can have other values between 0 and 1, clusters are *fuzzy*. This can be defined based on application. As stated in [3], numerical solution for this optimization problem is difficult since the number of unknown can be much more than number of known parameters and also no information is available about function $\Gamma(\mathrm{t})$. Therefore, the problem is ill-posed in the sense of *Hadamard* and thus requires adding additional assumptions to solve the problem. This process is known as regularization [8]. For example in *Tikhonov regularization*, additional assumption (called regularization term) is included in the minimization problem. Here, it is assumed that the cluster affiliations functions $\gamma_i(t)$are smooth and their derivative are bounded.

$$\sum_{i=1}^{K} \int_0^T \left[\gamma_i(t).\, g(x_t, \theta_i) + \epsilon^2 \left(\frac{\partial \gamma_i(t)}{\partial t} \right)^2 \right] dt \xrightarrow[\Gamma(\mathrm{t}),\Theta]{} min \tag{4}$$

In the above equation, ϵ is called regularization term. To solve the above problem numerically, it must be converted from continuous time domain to discrete-time domain. For this reason, Galerkin discretization is utilized here. Galerkin methods can convert a continuous operator problem (such as a differential equation) to a discrete problem. It is widely used in FEM literature for solving differential equations [9]. In our problem, the FEM basis function defined in the form of *N* triangular functions which are called *hat* functions. A set of continuous functions is defined with the local support on $[0, T]$ as in Figure 1. Applying discretization procedure to $\Gamma(\mathrm{t})$ yields:

$$\gamma_i(t) = \tilde{\gamma}_i(t) + error = \sum_{k=1}^{N} \tilde{\gamma}_i{}^{(k)}.\, v_k(t) + error \tag{5}$$

$$\tilde{\gamma}_i{}^{(k)} = \int_0^T \gamma_i(t).\, v_k(t) dt \tag{6}$$

where $\tilde{\gamma}_i{}^{(k)}$ are scalars called Galerkin coefficient. After some mathematical simplification and using the locality of finite elements basis function support, one can find an optimization in the form of linear quadratic programming.

$$\sum_{i=1}^{K} \left[a(\theta_i)^T \bar{\gamma}_i + \epsilon^2 \bar{\gamma}_i{}^T H \bar{\gamma}_i \right] \xrightarrow[\bar{\gamma},\Theta]{} min \tag{7}$$

$$\bar{\gamma}_i = \left[\tilde{\gamma}_i{}^{(1)}, \dots, \tilde{\gamma}_i{}^{(k)}, \dots., \tilde{\gamma}_i{}^{(N)} \right] \tag{8}$$

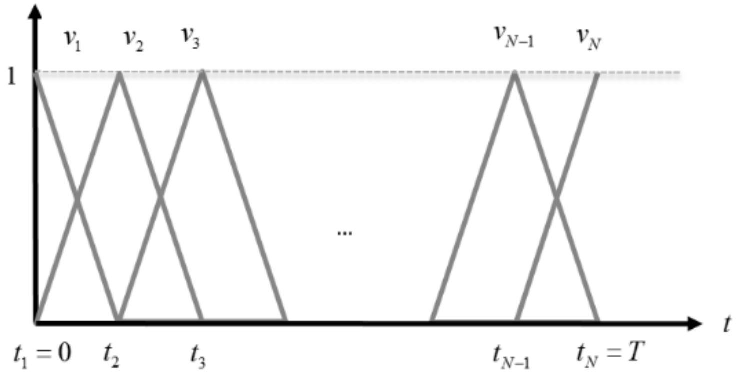

Fig. 1. Finite element basis functions in the form of hat function

$$a(\theta_i) = \left[\int_{t_1}^{t_2} v_1(t)g(x_t,\theta_i)dt, \dots, \int_{t_{k-1}}^{t_{k+1}} v_k(t)g(x_t,\theta_i)dt, \dots, \int_{t_{N-1}}^{t_{N+1}} v_N(t)g(x_t,\theta_i)dt\right] \tag{9}$$

$$H = \begin{pmatrix} \frac{1}{\Delta} & \frac{-1}{\Delta} & 0 & \cdots & 0 & 0 \\ \frac{-1}{\Delta} & \frac{2}{\Delta} & \ddots & 0 & 0 & 0 \\ 0 & \ddots & \ddots & \frac{-1}{\Delta} & 0 & \vdots \\ 0 & 0 & \frac{-1}{\Delta} & \frac{2}{\Delta} & \ddots & 0 \\ \vdots & 0 & 0 & \ddots & \ddots & \frac{-1}{\Delta} \\ 0 & 0 & \cdots & 0 & \frac{-1}{\Delta} & \frac{1}{\Delta} \end{pmatrix} \tag{10}$$

H is a tri-diagonal matrix called *stiffness matrix*. Convexity conditions on model distance functional are converted to constrain on Galerkin coefficients:

$$\sum_{i=1}^{K} \tilde{\gamma}_i^{(k)} = 1 \qquad k = 1, \dots, N \tag{11}$$

$$\tilde{\gamma}_i^{(k)} \geq 0 \qquad i = 1, \dots, K$$

The optimization problem above should be solved with respect to $\bar{\gamma}$ and θ iteratively. After finding Galerkin coefficient, we can build $\gamma_i(t)$ using FEM basis function and Eq. (5). Initially, some random initial ${\bar{\gamma}_i}^{initial}$ is assumed such that it fulfills the convexity conditions and then ${\theta_i}^{initial}$ is found. After that, the iterative procedure is ran for enough iteration numbers. First, the problem is solved with respect to $\bar{\gamma}_i$ for a fixed θ_i and second it is solved with respect to θ_i for a fixed $\overline{\gamma}_i$ and so on. The solution with respect to θ_i is found analytically based on the defined distance functional as:

$$\theta_{i0} = \frac{\begin{vmatrix} \sum_{t=0}^{T} x_t.\gamma_i(t) & \sum_{t=0}^{T} t.x_t.\gamma_i(t) \\ \sum_{t=0}^{T} t.x_t.\gamma_i(t) & \sum_{t=0}^{T} {x_t}^2.\gamma_i(t) \end{vmatrix}}{\begin{vmatrix} \sum_{t=0}^{T} \gamma_i(t) & \sum_{t=0}^{T} t.\gamma_i(t) \\ \sum_{t=0}^{T} t.\gamma_i(t) & \sum_{t=0}^{T} t^2.\gamma_i(t) \end{vmatrix}} \tag{12}$$

$$\theta_{i1} = \frac{\begin{vmatrix} \sum_{t=0}^{T} \gamma_i(t) & \sum_{t=0}^{T} x_t.\gamma_i(t) \\ \sum_{t=0}^{T} t.\gamma_i(t) & \sum_{t=0}^{T} t.x_t.\gamma_i(t) \end{vmatrix}}{\begin{vmatrix} \sum_{t=0}^{T} \gamma_i(t) & \sum_{t=0}^{T} t.\gamma_i(t) \\ \sum_{t=0}^{T} t.\gamma_i(t) & \sum_{t=0}^{T} t^2.\gamma_i(t) \end{vmatrix}} \tag{13}$$

For solving the optimization with respect to $\bar{\gamma}_i$, all the $\bar{\gamma}_i$ are augmented in a vector λ and the problem is converted to one linear quadratic programming.

$$\frac{1}{2}\lambda^T G\lambda + A^T\lambda \xrightarrow[\lambda]{} min \tag{14}$$

$$A = [a(\theta_1), \ldots, a(\theta_i),., a(\theta_K)] \tag{15}$$

$$G = \epsilon^2 \begin{bmatrix} H & 0 & \cdots & 0 \\ 0 & H & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & H \end{bmatrix} \tag{16}$$

and the new constraints become

$$\lambda_s \geq 0 \qquad\qquad \forall s = 1, \ldots, N \times K \tag{17}$$

$$F.\lambda = Q \tag{18}$$

$$F = \underbrace{[I_{N\times N} \quad I_{N\times N} \quad \ldots \quad I_{N\times N}]}_{K\ times} \tag{19}$$

$$Q = [1 \quad 1 \quad \ldots \quad 1]^T{}_{N\times 1} \tag{20}$$

There are three parameters that should be set at the start of the procedure: number of clusters K , regularization parameter ϵ and the number of hat functions N (or width of hat functions Δ). Decreasing N reduces the order of LQP and consequently complexity of calculations. On the other hand, it decreases accuracy of clustering and it may result in losing some very short regimes. A challenging problem arises when choosing K. A trivial solution exists when every data point is a cluster; and as a result, it is not possible to find the optimal number of clusters. By increasing K, the value of the cost function always decreases and when $K = number\ of\ data\ point$, this value approaches to zero. Since the number of clusters is unknown in advance, trial and error along with human judgment is used to select K subjectively. A criterion for choosing K is based on the value of the cluster affiliation functions when it becomes about 1 or 0. If we assume clusters are deterministic, when the cluster affiliation at time t for cluster i is 1, it means the datum at t completely belongs to that cluster (cluster affiliation equal to 0 means that data does not belong to cluster i at all). Increasing ϵ leads to an increase in the length of regimes. To find the optimal parameters, ϵ and K should be changed simultaneously. In the beginning, we set K equals to a sufficiently large number and then decrease K and run the algorithm for different ϵ to find acceptable results, this means the value of $\gamma_i(t)$ is about 0 or 1 in all of the time period [3]. After finding the trends, their statistical significance should be tested using Mann-Kendall approach [10].

3 Application in Climate Data Analysis

In this paper, temperature time series in North Carolina is studied as a case study. A data set of the average temperatures in 249 stations across NC are analyzed from the beginning of 1950 until the end of October 2009. The data is converted from daily to monthly in order to decrease the complexity of calculations. The dimension of resulting time series is 249 ×718. Temperature time series has a dominant harmonics with the period of one year which is called *seasonality*. This annual cycle has been removed by subtracting the multi-year monthly means. This is done, by subtracting the mean that is built over all values corresponding to the same month.

$$x_i^{new} = x_i - \bar{x}_i \tag{21}$$

where x_i^{new} is the deseasonalized value for month i (say for the month of January), x_i is original value for the same month (January) and $\bar{x}_i$ is the average monthly value in month i for the entire period of data (i.e. average of all January's data). Next, the FEM clustering applied to time series. Initially the value of K was assumed to be 10 and the algorithm was executed for different values of the regularization parameter (ε). For each ε, the algorithm ran several times to find best answer for constrained optimization problem. The regularization parameter is a real value between 0 and approximately 30. In this application, we are looking for deterministic clusters. When we reach a K where all the $\gamma_i(t)$ are about 0 or 1, an optimal solution is found. For the time series in this study, we found six regimes with different length and trends.

Figure 2 shows deseasonalized monthly time series in one of the dimensions and the linear trends detected by the FEM. In this figure, narrow lines show deseasonalized temperature time series in one of the stations and bold lines are linear trends

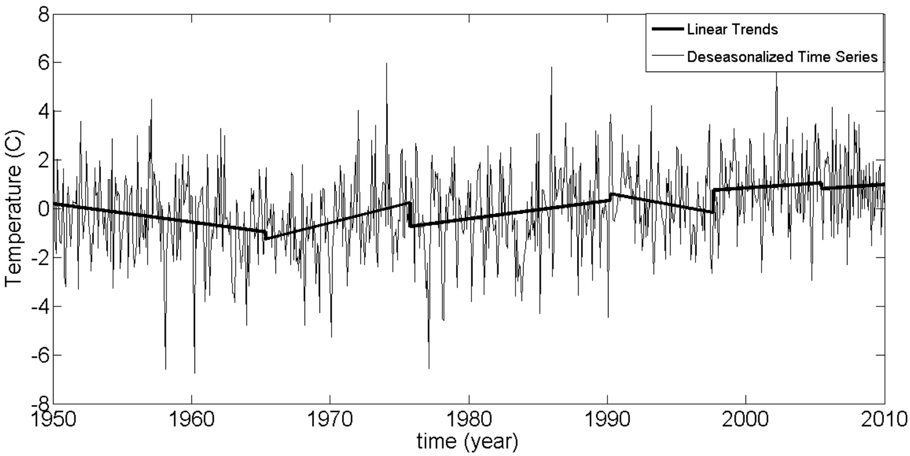

Fig. 2. Time series in one dimension and its regimes/trends

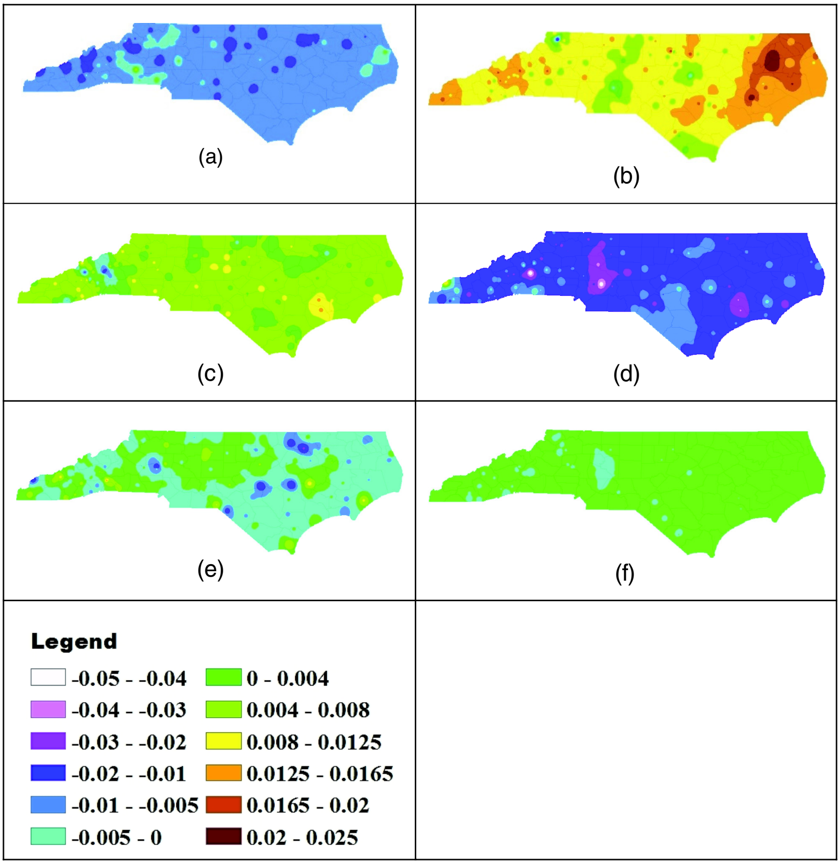

Fig. 3a-3f. Linear trend in NC for six regimes found by the FEM algorithm. Figure 3a-f shows the value of trend for six regimes in the North Carolina's map. Therefore, FEM clustering can reveal spatial in addition to temporal pattern of climate change. In Figure 3a-f, it is clear that there are two notable decreasing trends between 1965-1964 and 1990-1998. There is a remarkable increase in 1964-1976. Also, the 2^{nd} and the 4^{th} regimes show the warmest and coolest trend, respectively. Warming and cooling in eastern parts of the state in regime 2 and 5 are interesting. Different climatic phenomena may cause these patterns of change in NC, such as *El-Nino*, *Atlantic Multidecadal Oscillation* (AMO) and etc. [1]. We can compare these climatic indices with the results. For example comparison of these trends shows a correlation with AMO. Therefore we may infer that NC temperature is mostly affected by AMO.

4 Conclusion

In this paper, finite element method for clustering a multi-dimensional time series is used to find regimes in a climatic time series where each regime has a different linear trend. An appropriate cost function was defined and using Tikhonov regularization and Galerkin discretization, the cost function is converted to a familiar linear quadratic problem. There is a trade-of between number of Finite Elements Basis Function, volume of computation and consequently accuracy. Also, the regularization parameter can change the length of detected regimes. By trial and error, an optimal number of regimes can be estimated. A climatic time series of North Carolina is analyzed by this method. The results represent spatiotemporal pattern of climate change corresponding to areas of studies.

Acknowledgement. This work partially supported by the Expeditions in Computing by the National Science Foundation under Award Number: CCF-1029731. For additional information, contact corresponding author, Dr. Homaifar (homaifar@ncat.edu). With special thanks to Dr. Manoj Jha and Dr. Stefan Liess.

References

1. Mudelsee, M.: Climate Time Series Analysis. Classical Statistical and Bootstrap Methods. Springer (2010)
2. Kremer, H., Günnemann, S., Seidl, T.: Detecting Climate Change in Multivariate Time Series Data by Novel Clustering and Cluster Tracing Techniques. In: IEEE International Conference on Data Mining Workshops, San Jose (2010)
3. Horenko, I.: Finite element approach to clustering of multidimensional time series. SIAM Journal on Scientific Computing 32(1), 62–83 (2010)
4. Qi, Y., Paisely, J.W., Carin, L.: Music Analysis Using Hidden Markov Mixture Models. IEEE Transactions on Signal Processing 55(11), 5209–5224 (2007)
5. Meyer, F.G., Xilin, S.: Classification of fMRI Time Series in a Low-Dimensional Subspace with a Spatial Prior. IEEE Transactions on Medical Imaging 27(1), 87–98 (2007)

6. Kelly, J., Hedengren, J.: A Steady-State Detection (SSD) Algorithm to Detect NonStationary Drifts in Processes. Journal of Process Control 23(3), 326–331 (2013)
7. Metzner, P., Horenko, I.: Analysis of Persistent Non-stationary Time series and Applications. Communications in Applied Mathematics and Computational Science 7(2), 175–229 (2012)
8. Aster, R.: Parameter Estimation and Inverse Problems. Academic Press (2013)
9. Zienkiewicz, O.: The Finite Element Method Its Basis and Fundamentals. Elsevier (2005)
10. Modarres, R., Sarhadi, A.: Rainfall Trends Analysis of Iran in the Last Half of the Twentieth Century. Journal of Geophysical Research 114(D3) (2009)