Change-Point Detection of Climate Time Series by Nonparametric Method

Naoki Itoh *and Jürgen Kurths

Abstract—**In one of the data mining techniques, change-point detection is of importance in evaluating time series measured in real world. For decades this technique has been developed as a nonlinear dynamics. We apply the method for detecting the change points, Singular Spectrum Transformation (SST), to the climate time series. To know where the structures of climate data sets change can reveal a climate background. In this paper we discuss the structures of precipitation data in Kenya and Wrangel Island (Arctic land) by using the SST.**

***Keywords:** Singular Spectrum Analysis, Principal Component Analysis, climate time series analysis, change-point detection*

1 Introduction

Data mining aims to extract any information that is nontrivial but useful. With the dramatic development of computer performance, we can manipulate large amounts of data and analyze the structured data dynamically. Change-point detection is defined as one of the techniques.

The work of change-point detection has been studied in many scientific fields to give a better interpretation of properties extracted from a complicated system. Singular Spectrum Transformation (SST)[5, 6, 7] that we introduce in this study is based on Principal Component Analysis (PCA). The methodology itself has been described as Singular Spectrum Analysis (SSA) by Golyandina et al. (2001)[2]. The basic idea of the change-point detection using the SSA is explained by a paper of Moskvia and Zhigljavsky et al. (2003)[4]. SST is still being developed and improved (e.g. Mohammad and Nishida[8]). An advantage of use of the SST is no requirements of knowledge about ad-hoc tuning and modification for time series. Therefore, we suppose that it is suitable for analyzing data in real world such that only a little a priori knowledge about the data is provided.

In this paper Kenyan towns (Nakuru, Naivasha, and Narok) of the equatorial region and Wrangel Island in Arctic land[3] significantly impacted by the global warming such as drought and ice melting will be analyzed by this method.

Section 2 describes the algorithm of SSA. Section 3 defines change-point detection by using results from the SSA. The application to the climate time series is discussed in section 4. Then finally the conclusion is described in section 5.

2 Singular Spectrum Analysis

SSA can be defined as a model-free technique, which aims to decompose measured time series into some useful and interpretable components such as global trends, harmonic terms, and noise. The tasks of SSA can basically be classified as follows: 1.finding trends of different resolution; 2.smoothing; 3.extraction of seasonality components; 4.simultaneous extraction of cycles with small and large periods; 5.extraction of periodicities with varying amplitudes; 6.simultaneous extraction of complex trends and periodicities; 7.finding structure in short time series; and 8.change-point detection. From these items we will discuss a change-point detection by using the idea of Idé et al[5, 6, 7].

2.1 Algorithm

On the first stage, we make a trajectory matrix from a single time series and then decompose it by using Singular Value Decomposition (SVD)

The trajectory matrix $[X_1, \cdots, X_K]$ consists of the short vectors, $X_j = (y_j, \cdots, y_{j+L-1})^T \in \mathrm{R^L}$ $(j = 1, \cdots, K)$, from the single time series $Y = (y_1, \cdots, y_N)$, which forms a Hankel matrix[9], where the parameters K and L can be described by $K = N-L+1$. The L is called a window length restricted by $2 \leq L \leq N/2$ as a single parameter for the SSA:

$$\begin{aligned} \mathrm{X} &= [X_1 : \cdots : X_K] = (x_{ij})_{i,j=1}^{L,K} \\ &= \begin{pmatrix} y_1 & y_2 & \cdots & y_K \\ y_2 & y_3 & \cdots & y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \cdots & y_N \end{pmatrix}. \end{aligned} \quad (1)$$

The trajectory matrix can be decomposed into submatri-

*Center for Dynamics of Complex Systems University of Potsdam, D-14476 Potsdam, Germany Email: naoki.itoh@gmail.com

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

ces by the Singular Value Decomposition (SVD):

$$\mathrm{X} = \mathrm{X}_1 + \cdots + \mathrm{X}_d, \quad (2)$$

where

$$\mathrm{X}_l = U_l \sqrt{\lambda_l} V_l^T, (l = 1, \cdots, d), \quad (3)$$

which is defined as a rank-one orthogonal elementary matrix. A set of these three notations consisting of singular value ($\sqrt{\lambda_l}$), empirical orthogonal functions (U_l), and principal components (V_l) is called the l-th eigentriple of the matrix X. The index of the elementary matrix should at most be taken until d which is the number of nonzero singular values.

3 Change-Point Detection

Change-point detection problem is defined as a quantitative estimate of structural changes behind time series. The topic has been discussed by using several methods, such as a method based on Autoregressive model. The parametric approach, however, does not always lead to a fine result from heterogeneous time series because it is not easy to prepare any sufficient knowledge for such time series.

Singular Spectrum Transformation (SST) is a nonparametric change-point detection method, in which feature extraction can basically be achieved by using SVD defined in section 2. Since this method does not need to assume a certain stochastic model and to focus on some local solutions, it is applicable to heterogeneous time series.

3.1 SST technique

Let us consider a time point t in time series. Then with reference to t we can define two kinds of time series of the past part, $\mathrm{Y}^{(\mathrm{p})}$, and future part, $\mathrm{Y}^{(\mathrm{f})}$, from original time series Y as follows:

$$\begin{aligned} \mathrm{Y}^{(\mathrm{p})} &= (y_1, \cdots, y_{t-1}), \\ \mathrm{Y}^{(\mathrm{f})} &= (y_t, \cdots, y_N). \end{aligned} \quad (4)$$

Then we make a trajectory matrix from each time series.

$$\begin{aligned} \mathrm{X}_t^{(\mathrm{p})} &= [X_{t-K} : \cdots : X_{t-1}], \\ \mathrm{X}_t^{(\mathrm{f})} &= [X_{t+\gamma} : \cdots : X_{t+\gamma+K-1}], \end{aligned} \quad (5)$$

where each vector consists of L elements of Y and the parameter γ which we can arbitrarily set is positive integer. Note that, here, vectors in the empirical orthogonal functions (EOF), $u_i^{(\mathrm{p})}$ $(i = 1, \cdots, L)$ obtained from the trajectory matrix of the past part, $\mathrm{X}^{(\mathrm{p})}$, are described in descending order of the singular values. From $\mathrm{X}^{(\mathrm{f})}$ defined here as *test matrix*, the EOF is described by $u_j^{(\mathrm{f})}$ $(j = 1, \cdots, L)$. In particular, the vector of EOF belonging to the largest singular value, $u_1^{(\mathrm{f})}$, will be defined as $\beta(t)$. The change-point detection can be performed by a comparison between their vectors of the both parts, which is called change-point score:

$$z = 1 - \sum_{i=1}^{l} K(i, \beta)^2, \quad (6)$$

where K is an inner product of the vectors:

$$K(i, \beta) := \beta^T u_i^{(\mathrm{p})}. \quad (7)$$

3.2 Demonstration

We will express the result of SST by using simple examples. The first example time series consists of three kinds of linear functions with the positive, zero, and negative slopes. The second one is generated by three kinds of sine functions with noise[5]. Fig. 1 shows each original time series and its result by SST. We can see that the results of both data sets show high scores at about $t = 150$ and at $t = 300$. This means that change points of both cases are found at the same positions.

4 Application

In our study the monthly precipitation at the three measuring stations in Kenya (Nakuru, Naivasha, and Narok) and in Wrangel Island are analyzed. Fig. 2 shows their positional relation and the atmospheres of places. Although we have the data with different time length, as shown in Fig. 3 the region between January 1950 and December 1985 corresponds to 432 data points overlapping at all the stations.

In order to compare their data structures during the overlapping time interval, we calculate the change-point scores of them by SST. For the parameter $L = 24$ (i.e. for two years) the change-point scores can be depicted as sharp curves (see Fig. 3).

Although it is not easy to discuss similarity and dissimilarity from these *noisy* original time series, the results from the application of SST provide those aspects. By comparing the four bottom panels in Fig. 3 we can especially see that the result of the three Kenyan towns around 1960 makes each high change-point score. This means that the precipitation structure at this time is different from the others. The result from Wrangel Island can be shown as a different kind of pattern than the results from the Kenyan towns.

5 Conclusion

In this study, as a data mining of the climate, we were able to demonstrate the capability of the SST method, which is defined here as a change-point detection based

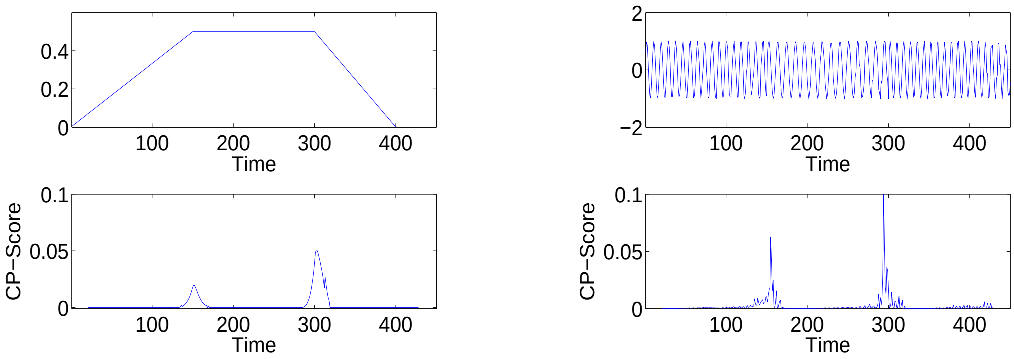

Figure 1: The results by SST; left panels: linear data set and right panels: harmonic data set

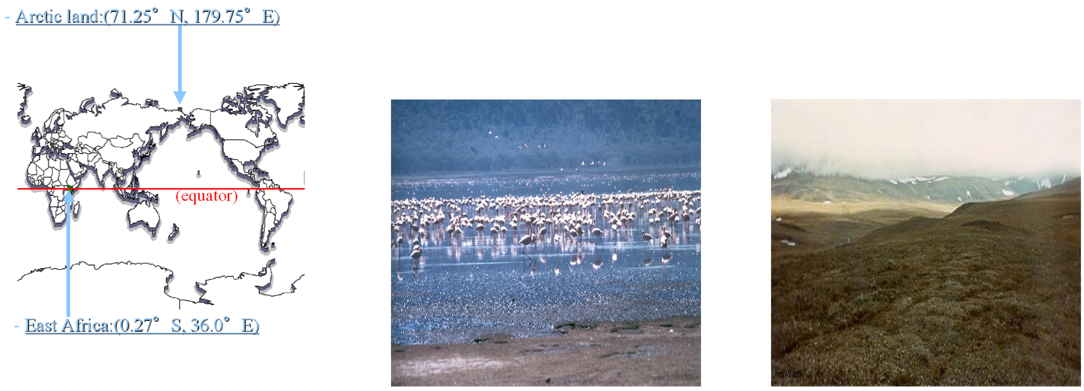

Figure 2: Measuring stations in Kenya (East Africa) and in Wrangel Island (Arctic land); left: their positional relation, middle: Lake-Nakuru in central Kenya, right: tundra on Wrangel Island.

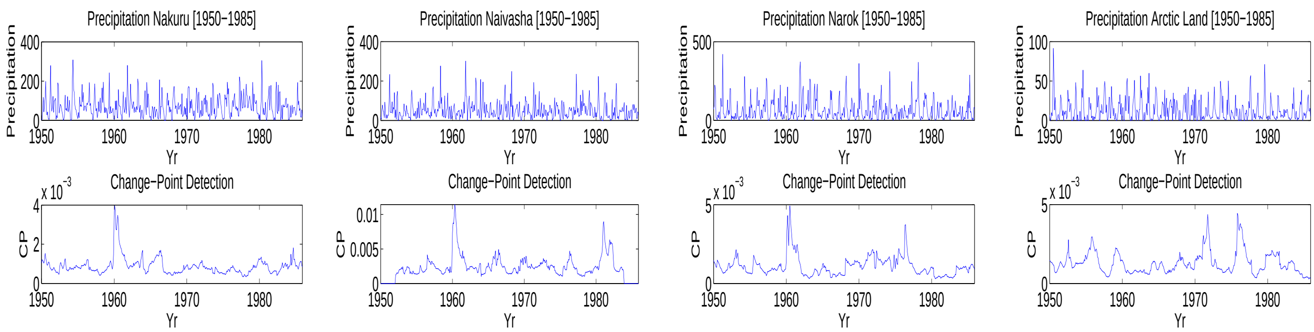

Figure 3: Top four panels: precipitation and bottom four panels: change-point score.

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

on singular spectrum analysis, to detect quantitative changes in the structure of heterogeneous data sets.

As shown in section 4 the change-point scores have displayed a characteristic expression according to location. From the measuring stations in Kenya the commonality of the change-point could be shown. On the other hand, Wrangel Island remote from the equatorial area has shown quite a different pattern of the change-point. Thus, applying SST to more widespread areas, we may have more precise insights into the climate period background.

Acknowledgment

The precipitation data from GHCN v2 database (`http://www.ncdc.noaa.gov/`) used in this study were provided by Norbert Marwan (PIK Potsdam). Udo Schwarz (Center for Dynamics of Complex Systems, University of Potsdam) and Tsuyoshi Idé (IBM Tokyo Research Laboratory) gave some quite valuable comments and suggestions. The authors would like to acknowledge the help of them.

References

[1] H. Björnsson, S.A. Venegas, "A Manual for EOF and SVD analyses of Climate Data," *Centre for Climate and Global Change Research*, Report No. 97.1, 1997

[2] N. Golyandina, V. Nekrutkin and A. Zhigljavsky, *Analysis of Time Series Structure, SSA and related techniques*, Chapman & Hall/CRC, 2001

[3] Kenji Matsuura and Cort J. Willmott, "Arctic Land-Surface Precipitation: 1930-2000 Gridded Monthly Time Series," *Center for Climatic Research*, Ver. 1.10, March, 2004

[4] V. Moskvina and A. Zhigljavsky, "An algorithm based on singular spectrum analysis for change-point detection," *Communications in Statistics–Simulation and Computation*, pp. 319–352, 03

[5] Tsuyoshi Idé, Keisuke Inoue, "Knowledge Discovery Heterogeneous Dynamic Systems using Change-Point Correlations," *2005 SIAM International Conference on Data Mining (SDM 05)*, 4/05

[6] Tsuyoshi Idé, Keisuke Inoue, "Knowledge Discovery from Time-series Data using Nonlinear Transformations," *Proceedings of the Fourth Data Mining Workshop (The Japan Society for Software Science and Technology, Tokyo, 2004)*, No.29, pp. 1–8, 2004 (Japanese)

[7] Tsuyoshi Idé, "Speeding up Change-Point Detection using Matrix Compression," *2006 Workshop on Information-Based Induction Science*, 10–11/06 (Japanese)

[8] Yasser Mohammad and Toyoaki Nishida, "Robust Singular Spectrum Transform," *the Twenty Second International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE 2009)*, Taiwan, 06/09

[9] J.L. Phillips, "The Triangular Decomposition of Hankel Matrices," *MATHEMATICS OF COMPUTATION*, V5, N115, 07/71

[10] M. Ghil, M.R. Allen, M.D. Dettinger, K. Ide, D. Kondrashov, M.E. Mann, A.W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou "ADVANCED SPECTRAL METHODS FOR CLIMATIC TIME SERIES," *The American Geophysical Union Review of Geophysics*, V40, pp. 1–41 9/01