Global and Planetary Change 129 (2015) 45–55

Contents lists available at ScienceDirect

Global and Planetary Change

journal homepage: www.elsevier.com/locate/gloplacha

Spatio-temporal trend analysis of air temperature in Europe and Western Asia using data-coupled clustering

Mihaela I. Chidean [a], Jesús Muñoz-Bulnes [b], Julio Ramiro-Bargueño [a], Antonio J. Caamaño [a,*], Sancho Salcedo-Sanz [b]

[a] Dept. of Signal Theory and Communications, Universidad Rey Juan Carlos, Madrid, Spain
[b] Dept. of Signal Processing and Communications, Universidad de Alcalá, Madrid, Spain

A R T I C L E   I N F O

Article history:
Received 6 December 2014
Received in revised form 27 February 2015
Accepted 18 March 2015
Available online 25 March 2015

Keywords:
Air temperature
Clustering techniques
Spatio-temporal trend
Climate change

A B S T R A C T

Over the last decades, different machine learning techniques have been used to detect climate change patterns, mostly using data from measuring stations located in different parts of the world. Some previous studies focus on temperature as primary variable of study, though there have been other works focused on precipitation or even wind speed as objective variable. In this paper, we use the self-organized Second Order Data Coupled Clustering (SODCC) algorithm to carry out a spatio-temporal analysis of temperature patterns in Europe. By applying the SODCC we identify three different regimes of spatio-temporal correlations based on their geographical extent: small, medium, and large-scale regimes. Based on these regimes, it is possible to detect a change in the spatio-temporal trend of air temperature, reflecting a shift in the extent of the correlations in stations in the Iberian Peninsula and Southern France. We also identify an oscillating spatio-temporal trend in the Western Asia region and a stable medium-scale regime affecting the British Isles. These results are found to be consistent with previous studies in climate change. The patterns obtained with the SODCC algorithm may represent a signal of climate change to be taken into account, and so the SODCC could be used as detection method.

1. Introduction

The state-of-art for detection and attribution (D&A) of climate change problems usually considers two types of models: General Circulation Models (GCMs) and statistic models. The GCMs are mainly used to understand the dynamics of the physical components of the atmosphere, that are related to the climatic change phenomenon. The goal of GCMs is to obtain spatio-temporal climatic change patterns (usually global patterns), also known in the literature as fingerprints of the climatic change, and they can also be used to make predictions and long-term projections of climatic variables (Cramer et al., 2013). In D&A problems, the GCMs have been used with different temporal scales, from seasonal to decadal time-horizons. At present, some of the most sophisticated GCMs used to study climate change are developed by the WCRP's Working Group on Coupled Modeling (WGCM, 2014), which provides a multi-model context for carrying out coordinated climate model experiments, especially well-suited for D&A problems (Solomon et al., 2009).

On the other hand, the use of alternative statistic models has grown in the last years, due to they are able to obtain clear evidence of climate change in a fraction of the computational time needed by the GCMs. Several statistical models have been applied to different D&A problems, i.e. methods developed for econometric series (Kaufmann & Stern, 1997), also known as co-integration methods (Kaufmann & Stern, 2002; Kaufmann et al., 2011), or methods based on regression type approximations to evaluate climate change patterns using temperature data (Douglass et al., 2004; Stone & Allen, 2005).

Air temperature is a key parameter for the detection of climate change in certain areas, the assessment of its impact in different ecosystems (Alva-Basurto & Arias-González, 2014; Gomiero & Viarengo, 2014). Air temperature it is also related to the evaluation of the human activity, such as agriculture (Smith et al., 2009; Cobaner et al., 2014), health-care (Garske et al., 2013; Xu et al., 2014), and energy (Paniagua-Tineo et al., 2011; Jaglom et al., 2014). Temperature analysis in climate change studies may also involve problems of prediction, reconstruction or spatio-temporal analysis of the results; we are interested in the latter, where there are important previous works in the literature, published in the last few years. In Carrera-Hernández and Gaskin (2007) a study of spatio-temporal analysis using minimum temperature and precipitation data was carried out for Mexico. Several regression algorithms based on Kriging were applied in this case to more than 200 stations located in the Basin of Mexico for the period 1978–1985. In Kousari et al. (2013) maximum temperature data in Iran over the 1960–2005 period were analyzed with implications for

* Corresponding author.
E-mail address: antonio.caamano@urjc.es (A.J. Caamaño).

http://dx.doi.org/10.1016/j.gloplacha.2015.03.006
0921-8181/

climate change detection in the area. The authors applied different techniques including statistical test, filtering, and hierarchical clustering to data from 32 synoptic stations that cover Iran, and showed the trends and variability of temperature at different scales (annual, seasonal, and monthly). Another recent work on temperature variability is Kloog et al. (2012), where satellite surface measurements are used to analyze the spatio-temporal trends of minimum temperature for Massachusetts (USA). The study applies regression techniques and spatial smoothing and incorporates alternative meteorological data. Somehow related to that work is Van De Kerchove et al. (2013), in which the authors present a study of spatio-temporal variability of temperature in a remote region of Russia. Signal processing methods such as Fast Fourier Transform and others such as multi-linear regression are applied in this case to carry out the study.

In this paper we focus on the study of spatio-temporal analysis of air temperature. We propose the application of a novel clustering algorithm to this end, with focus on European data. Specifically, we use the Second Order Data Coupled Clustering (SODCC) algorithm, that was initially developed for joint data codification and self-organization of wireless sensor networks (Chidean et al., 2013). The SODCC is a self-organized clustering approach, that uses characteristics of the measured data (air temperature in this case) and geographically groups the measuring stations. To achieve this end, the proposed algorithm uses second-order statistics to compute the minimum amount of linearly independent components in the cluster, i.e., the number of principal components that explain most of the variance in the data. Note that previous spatio-temporal clustering approaches such as Carrera-Hernández and Gaskin (2007), Horenko(2010), and Kousari et al. (2013) exploit the similarities among air temperature temporal series of different measuring stations to ascribe them to different clusters, which may or may not be spatially compact. In contrast, SODCC performs joint spatio-temporal clustering by: 1) minimizing the number of principal components needed to explain the variance of the data and 2) not allowing for spatially disjoint clusters to occur.

We apply the SODCC algorithm to temperature fields obtained from measuring stations throughout Europe (Klein Tank et al., 2002; Chimani et al., 2013) considering different initial points (at decadal time scale). From the obtained cluster distributions we are able to evaluate the relations between neighboring measuring stations, the probability of cluster formation, and their time evolution. This fact helps us identify a possible climate change (or climate variability) pattern in the Iberian Peninsula from the 1970s onwards.

The rest of this paper is organized as follows: Section 2 describes the SODCC algorithm and its implementation; in Section 3 we describe the experiments made using data from weather stations and we show a possible climate change pattern that can be identified from our results. Finally, Section 4 concludes the article.

2. Self-organized clustering based on second order statistics

In this section we describe the Second Order Data Coupled Clustering (SODCC) algorithm, which we use in this work to analyze air temperature data. SODCC is a self-organized clustering algorithm and was initially proposed for wireless sensor networks. By using characteristics of the measured data (e.g. temperature), SODCC groups the measuring stations by coupling the clusters to the measured data field. For it, the algorithm uses second-order statistics to compute the minimum amount of linearly independent components in the cluster (the dimension of the signal subspace or, similarly, the number of principal components that explain most of the variance in the data). In the subsequent stages, the clusters merge until a stopping criterion based on the dimension of signal subspace is reached. This approach ensures the non-singularity of the signal subspace of the data measured by each cluster (the covariance matrix of the data is well-posed).

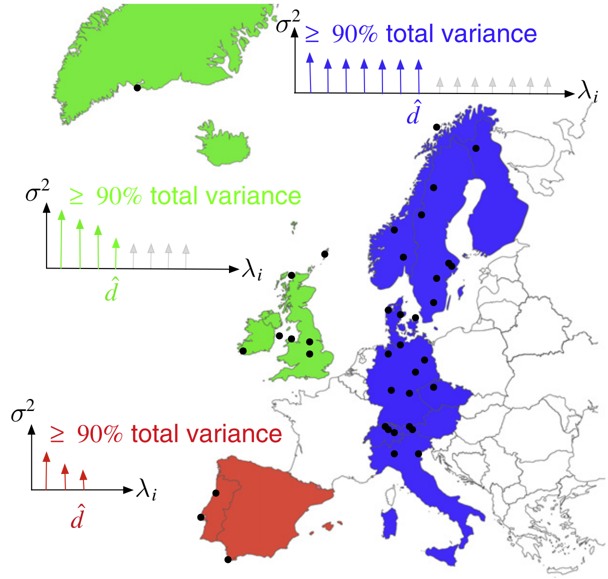

Fig. 1. The SODCC clustering of the measuring stations can be understood in terms of explained variance. The variance of the measurements from measuring stations in small clusters (red) can be explained with almost all of the (few) present principal components. As clusters enlarge, more principal components are needed to explain the variance of the measurements (green and blue).

By using the signal subspace dimension as a clustering criteria for the different measuring stations, we ensure that most of the variance in the data is captured within that cluster (see Fig. 1). It means that, if the cluster is small, the data from a small number of stations are needed to extract the eigenvalues of the largest principal components that explain their variance. Thus, the cross-correlation of the data among those stations is high. However, if the cluster is large, the data from many more stations are needed to explain the same percentage of the variance. Therefore, by using the dimension of the signal subspace we are ensuring that closely correlated temperature series are clustered together.

In the following subsections we describe the system model used, in which each measuring station is one of the nodes of the network. Next, we introduce the Fast Subspace Decomposition (FSD) algorithm (Xu & Kailath, 1994), the one used by SODCC to estimate the dimension of the signal subspace in each cluster. The third issue of this section is to carry out a detailed description of the SODCC algorithm. The main goal of SODCC is to group in the same cluster stations with high spatio-temporal correlation, linking the cluster configuration with the measured field. For this purpose, the output of FSD forms part of the criteria for deciding if a cluster has the minimum size required: when the signal subspace dimension is lower than the cluster size and the covariance matrix of the measured dataset is separable.

2.1. System model

Each measuring station considered is modeled as a node forming part of a network. Therefore, let $G = (V, E)$ be the graph modeling the network, with $|V| = N$ measuring stations and $|E| = C$ links or connections between the stations.

Each of the N stations takes a measurement each T_s time instants starting in T_I, obtaining a total of M measurements. Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_M]$, with $\mathbf{x}_m \in \mathbb{R}^N$, be the dataset measured by the entire network. The covariance matrix of $\mathbf{X}$ can be estimated as $\mathbf{\Sigma} = \mathbf{X}\mathbf{X}^\top/M$, where $(\cdot)^\top$ indicates the transpose operation.