Remote Sensing of Environment 148 (2014) 84–96

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Novel approaches in Extended Principal Component Analysis to compare spatio-temporal patterns among multiple image time series

Neeti Neeti [*,1], J. Ronald Eastman

Graduate School of Geography, Clark University, 950 Main Street, Worcester, MA 01610, USA

ARTICLE INFO

Article history:
Received 7 April 2013
Received in revised form 13 February 2014
Accepted 16 March 2014
Available online 15 April 2014

Keywords:
Extended PCA
Time series analysis
S-mode PCA
T-mode PCA

ABSTRACT

Extended Principal Component Analysis (EPCA) aims to examine the patterns of variability shared among multiple datasets. In image time series analysis, this is conventionally done by virtually extending the spatial dimension of the time series by spatially concatenating the different time series and then performing *S*-mode PCA. In *S*-mode analysis, samples in space are the statistical variables and samples in time are the statistical observations. This paper introduces the concept of temporal concatenation of multiple image time series to perform EPCA. EPCA can also be done with a *T*-mode orientation in which samples in time are the statistical variables and samples in space are the statistical observations. This leads to a total of four orientations in which EPCA can be carried out. This research explores these four orientations and their implications in investigating spatio-temporal relationships among multiple time series. This research demonstrates that EPCA carried out with temporal concatenation of the multiple time series with *T*-mode (*tT*) is able to identify similar spatial patterns among multiple time series. The conventional *S*-mode EPCA with spatial concatenation (*sS*) identifies similar temporal patterns among multiple time series. The other two modes, namely *T*-mode with spatial concatenation (*sT*) and *S*-mode with temporal concatenation (*tS*), are able to identify patterns which share consistent temporal phase relationships and consistent spatial phase relationships with each other, respectively. In a case study using three sets of precipitation time series data from GPCP, CMAP and NCEP-DOE, the results show that examination of all four modes provides an effective basis for comparison of the series.

1. Introduction

The availability of Earth observation system (EOS) based global time series data have made it possible to explore the functioning of the Earth as a whole and to examine its main sub-systems: climate, biogeochemical cycles, the global hydrological cycle and linkages between them (Asrar & Dozier, 1994; Steffen, 2010; Steffen et al., 2004). The EOS global time series datasets have widely been used for understanding coupled systems in ESS using spectral decomposition techniques (Ambaum & Hoskins, 2001; Batstone, Matthews, & Stevens, 2005; Benestad, 2004; Deser & Blackmon, 1993; Singh & Kripalani, 1986; Weare & Nasstrom, 1982). Spectral decomposition refers to a group of mathematical transformations that are effective in understanding underlying relationships between groups of variables and have proven to be effective in geographic time series analysis (Azzali & Menenti, 2000; De Beurs & Henebry, 2010; Eastman & Fulk, 1993; Elsner & Tsonis, 1996; Fensholt, Rasmussen, Nielsen, & Mbow, 2009; Hall-Beyer, 2003).

Extended Principal Component Analysis (EPCA) is a form of spectral decomposition that aims to examine the patterns of variability that are shared among multiple time series where each of these image time series can be represented as a space-time cube (Fig. 1a). EPCA identifies the shared patterns of variability among these time series by virtually extending the spatial dimension by concatenating the multiple space-time cubes spatially (Fig. 1b) and then performing PCA on the extended series. EPCA is also known as Joint PCA (Navarra & Simoncini, 2010; Preisendorfer & Mobley, 1988) or Multivariate EOF (Empirical Orthogonal Function) (Cressie & Wikle, 2011). However, EPCA can also be done by virtually extending the temporal dimension of the series by temporal concatenation (Fig. 1c) and then performing PCA on the extended series.

In the literature, the conventional (spatial concatenation) EPCA is carried out primarily using *S*-mode orientation (Jolliffe, 2002; Kutzbach, 1967; Singh & Kripalani, 1986; Van den Dool, 2007; Wheeler & Hendon, 2004). Orientation here represents the way the image time series is organized for the analysis. In *S*-mode orientation, the statistical variables are samples in space and statistical observations are samples in time (Richman, 1986). Also, in addition to the *S*-mode orientation, in image time series analysis there is another orientation called *T*-mode orientation (Cattell, 1973; Richman, 1986). In *T*-mode orientation, the statistical variables are samples in time and the

* Corresponding author at: Graduate School of Geography, Clark University, 950 Main Street, Worcester, MA, USA.
E-mail address: neeti.squared@gmail.com (N. Neeti).
[1] Present address: Department of Earth & Environment, Boston University, Boston 02115, USA.

http://dx.doi.org/10.1016/j.rse.2014.03.015
0034-4257/

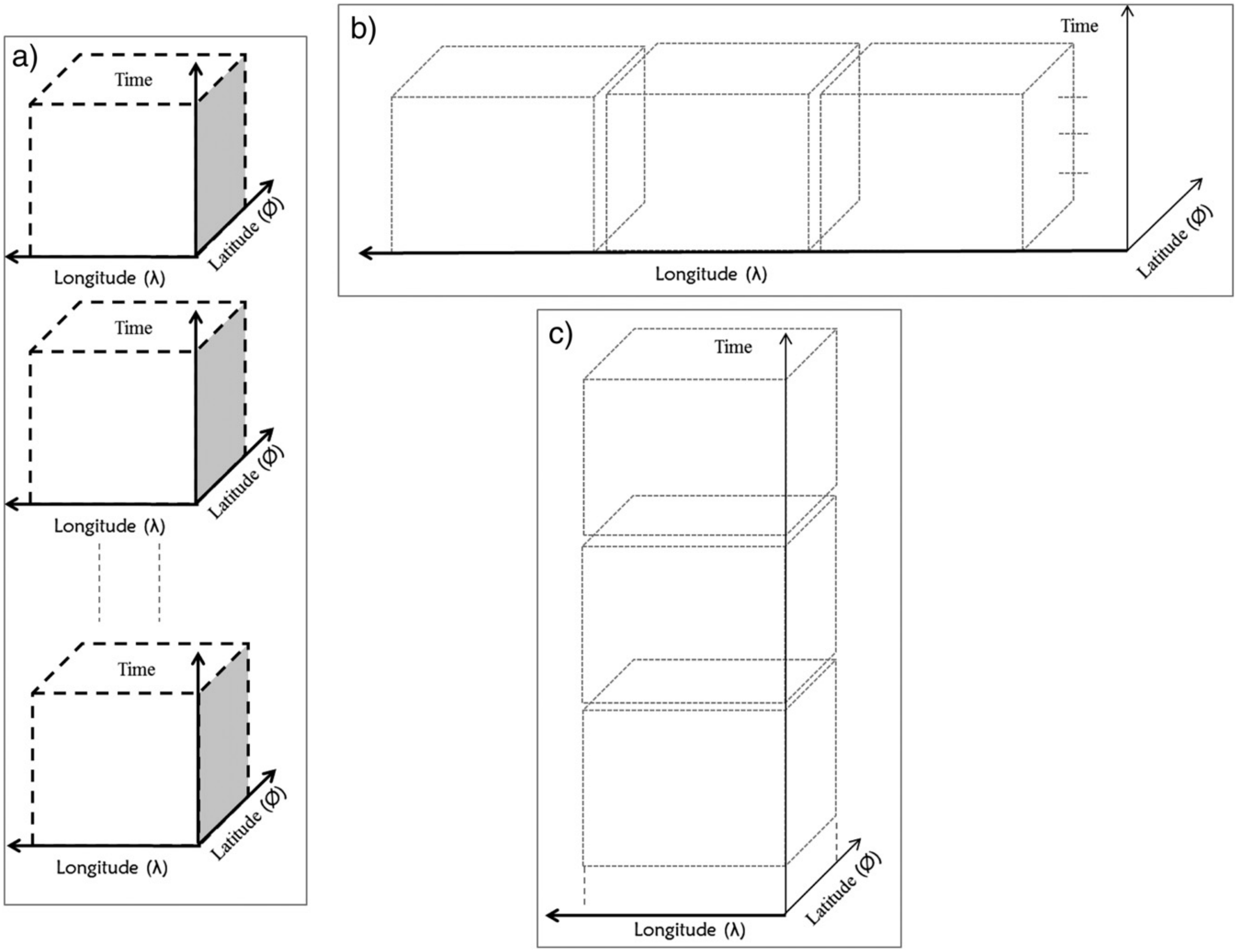

Fig. 1. Conceptual framework for EPCA: a) multiple image time series, b) spatial concatenation, and c) temporal concatenation.

statistical observations are samples in space. This results in a total of four modes of analysis (*S*- and *T*- with spatial and temporal concatenation) for EPCA. As will be discussed below, *S*- and *T*-mode analyses provide different insights about the nature of the series investigated. In this paper we explore the implications of adding these two different concatenation procedures, yielding in combination, four orientation modes for investigating relationships among multiple image time series.

2. *S*- and *T*-mode orientation

In the Factor Analysis/Principal Components Analysis literature, there have been six different ways observed for decomposing multidimensional data, namely *O*, *P*, *Q*, *R*, *S* and *T* (Cattell, 1973; Richman, 1986). Only the *S*- and *T*- modes involve time and are applicable to geographical data if space is considered as one dimensional. The focus of *T*-mode analysis is to find the relationship between samples in time over space (Compagnucci, Salles, & Canziani, 2001) and thus identifies dominant recurring spatial patterns over time (Machado-Machado, Neeti, Eastman, & Chen, 2011). In the case of *S*-mode, the primary focus is to find the relationship between samples in space over time and thus identifies recurrent temporal patterns over space (Machado-Machado et al., 2011). For example, Machado-Machado et al. (2011) found that a *T*-mode PCA on a time series (monthly, 1982–2007) of Microwave Sounding Unit (MSU) lower tropospheric temperature imagery shows the pattern of the North Atlantic Oscillation (NAO) in the first component whereas the *S*-mode PCA results in a first component showing a pan-tropical atmospheric bridge response to the El-Nino Southern Oscillation (ENSO) — a pattern which is never revealed in the *T*-mode analysis. With the *S*- and *T*-mode orientations and spatial (*s*) and temporal (*t*) concatenations, the four modes of EPCA explored here are a) *sS*-mode b) *sT*-mode c) *tS*-mode and d) *tT*-mode where *sS*-mode is the conventional way of analyzing EPCA.

3. Extension through spatial and temporal concatenation

S-mode based spectral decomposition identifies dominant temporal patterns (Machado-Machado et al., 2011); therefore, its extension by either spatial or temporal concatenation (i.e., *sS*-mode and *tS*-mode) will also extract temporal patterns. Since in *sS*-mode, the statistical variables are samples in space over time, the total number of variables is the number of samples in space over the multiple datasets combined. In *sS*-mode EPCA, the focus is to find similar temporal patterns among the concatenated time series of datasets. Although it is possible that for any given component it may find a temporal pattern in only one of the series, it is far more likely to preference patterns that exist in multiple series since a greater portion of variance will be explained.

In *tS*-mode, the statistical variables are again samples in space, but the number of variables equals the number of samples in one dataset since the concatenation is across time rather than space. In this case, the focus is again to look for similar temporal patterns, but this