EMPIRICAL ORTHOGONAL FUNCTION COMPUTATION AND ANALYSIS OF AEROSOL OPTICAL DEPTH FROM MODIS DATA OVER NORTHERN INDIA

Manoj Kumar Singh
Email: manojks@iitb.ac.in
&
Parvatham Venkatachalam
Email: pvenk@iitb.ac.in

Centre of Studies In Resources Engineering
Indian Institute of Technology Bombay
Powai, Mumbai 400076, India

KEY WORDS: Aerosol, EOF, Filling missing values

ABSTRACT: In this paper, an iterative method was used to fill missing values in MODIS aerosol optical depth data. The iterative method used in the paper is particularly useful in extracting the empirical orthogonal functions (EOFs). EOF is a useful tool for extracting spatial and temporal pattern from a spatio-temporal process. Atmospheric aerosols distribution is spatio-temporal process. Aerosols pattern present in the south-east Asia in particular over the northern India is discussed using modes found by the iterative EOF method.

1 Introduction

Summer monsoon and water runoff from the Himalaya are two key parts in the livelihood of millions of people of northern India. By modulating heat budget and altering microscopic properties of clouds, atmospheric aerosols disrupt summer monsoon and Himalayan snow cover ([Gautam et al., 2009, 2011]). Clearer interaction of atmospheric aerosols with summer monsoon and the Himalayan climate need to be understood. Measurement of aerosols at synoptic spatial scale and at greater frequency in time domain is a key milestone in understanding above mentioned climate and aerosols relationship. Aerosol optical depth (AOD) is a measure of aerosols present in the atmosphere. Satellite sensors such as MODIS and MISR provide synoptic and frequent measurement of AOD. Different algorithms were in use to retrieve AOD data from different sensors. It is difficult to retrieve AOD data over areas with cloud cover, deserts and snow covered areas. Therefore, satellite AOD products have many missing values. Empirical orthogonal function (EOF) is a tool quite frequently used in climate data analysis. EOF analysis on data field with too many missing values may not give proper modes and principle components. Many times missing values overestimate the amplitude of EOF. Interpolation can be done to fill out missing values. But interpolation generally smoothen the spatial and temporal fields and thus reduces the signals in EOF modes.

In this paper we have used a method for filling missing values based on eigenvalue-eigenvector decomposition. The method involves Markov chains of eigenvalues and eigenvectors. The iterative EOF method used in the study is similar to that used by [Beckers and Rixen, 2003]. We have further analyzed the EOF modes and principle components obtained from this process to study the behavior of aerosols in the northern India. The spatio-temporal process which governs the aerosols over the northern India may not be independent from that of south-east Asia. Therefore EOF modes are derived for the region bounded by $40°E, 20°S, 120°E$ and $40°N$ graticule and different modes are discussed for northern India only.

In next section we describe data set used, section 3 describes the methodology used for EOF computation and handling missing values and section 4 describes the results and related discussion.

2 Data

Our study of aerosols was limited to the region bounded by $40°E, 20°S, 120°E$ and $40°N$ graticule. Study period was limited from January 2001 to December 2013. In this section MODIS AOD data set was described.

MODIS level 3 monthly terra data set collection 051 was used in this study. Data set has resolution of $1°$. MODIS has a swath width of 2330 km. With such a big swath width, it has a global coverage of almost 2 days. The AOD data sets in MODIS are available at 0.55 micrometer.

3 Methodology

To understand a spatio-temporal climate variable, knowledge of the physics of that variable is desirable. But climate system of earth is a complicated process. We have very little knowledge of the physics of many climate variables. In the absence of physical knowledge of the parameter, statistical methods can be used to extract the spatial and temporal pattern of a climate parameter. One such method, which is used intensively in climate science, is EOF. EOF method extracts independent and orthogonal pattern from the data of the process under consideration. These independent patterns are called modes in climate science literature and corresponding time series as principle components (PC). The method of finding the EOFs starts with calculation of variance covariance matrix.

Let S be the number of locations and T be the number of months at which measurement of the process under consideration is available. Let X be a matrix of size $S \times T$ such that the element at position (i, j) denotes the measurement at location index i and time epoch j. The variance covariance matrix M of size $S \times S$ is then given by,

$$M = \frac{1}{T-1} XX' \tag{1}$$

M is a positive definite matrix and its eigenvalues are real. First mode of spatial structure provided by the eigenvector corresponds to the largest eigenvalue. The magnitude the eigenvectors provide the measure of the variance at any mode. Time series corresponding to a particular mode E is given by

$$P = X'E \tag{2}$$

The time series E is the principle component corresponding to the mode E and provides the temporal structure.

In analysis of climate science data, it is customary to work with data anomaly rather than the data itself. In this study anomaly data is found by removing mean of the data for a particular month from the data of that month. i.e. if $Y = y(i, j)$ is a given data set (i is the location and j is the time),

$$m_i(\text{ jan }) = \text{mean of } y(i, j) \text{ for j in January month index} \tag{3}$$

where $m_i(\text{ jan })$ is mean of January month at location index i. The anomaly $x(i, j)$ is then given as

$$x(i, j) = y(i, j) - m_k, \text{ where } k = (j-1)(\text{ modulo } 12) + 1 \tag{4}$$

Satellite based data set have many missing values. For example, AOD data products obtained from MODIS have missing values due to cloud cover and reflection from snow and deserted land. Basis for EOF analysis is the variance covariance matrix in equation 1. If at location s, sufficient number of samples are not available then variance estimated at that location using available samples may not be a reliable estimate of the variance. Similarly covariance between two location may not be reliable. The method described in this paper fills out missing values in the satellite data using EOF method. The method presented in the paper is a variant of method the described in [Beckers and Rixen, 2003].

The algorithm to fill out missing value involves following steps.

1. Fill missing values with some initial value. In EOF method one of the popular methods to fill missing value in anomaly data, is to replace it with zeros and in absolute data with mean.

2. After filling missing values with initial values, we calculate eigenvalues and eigenvectors of the variance covariance matrix.

3. Reconstruct original data matrix $X1_k$ using first k eigenvectors for each $k = 1, 2, \ldots, T$.

4. For each k replace the value at missing point with the corresponding value of matrix $X1_k$.

5. Repeat steps 2 for each k, until mean of the difference between value at missing point of X and $X1_k$ is small enough. In other words the convergence is reached.

6. For computation of mth mode we looked at number $k = kmax$ for which mth eigenvector is maximum.

4 Results

In this study the empirical orthogonal function analysis has been performed on monthly anomaly data of MODIS AOD data. Method to find anomaly aod data was described in section 3. Method used to find EOF modes while simultaneously filling missing data was described in the previous section. Data has 156 independent and orthogonal eigenvectors. The number of eigenvectors to be selected for retrieving back the data is an interesting problem. In EOF analysis, our motivation for selecting modes are maximum signal extraction from a single mode. The mode corresponding to the largest eigenvalue gives maximum variation pattern. To select number i of eigenvectors to be used for extracting kth mode, we looked at the percent of variation of mode k, when $i = 1$ to 156.

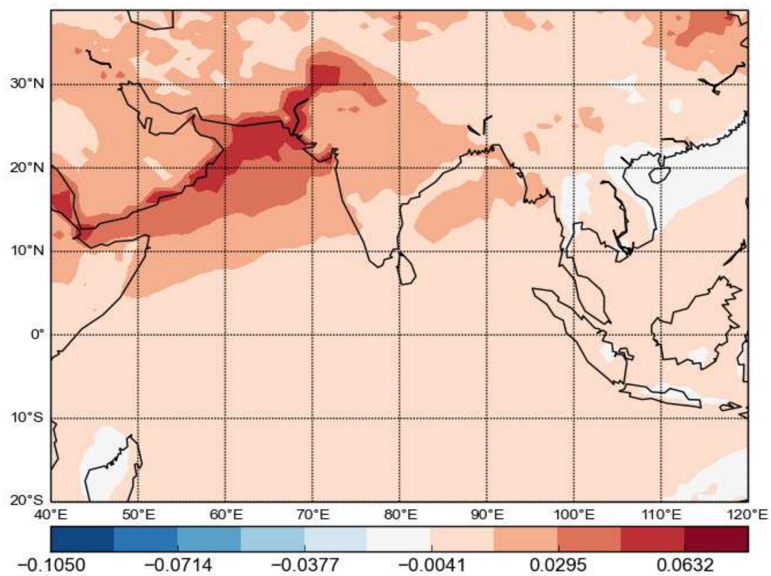

Figure 1: First mode of aod

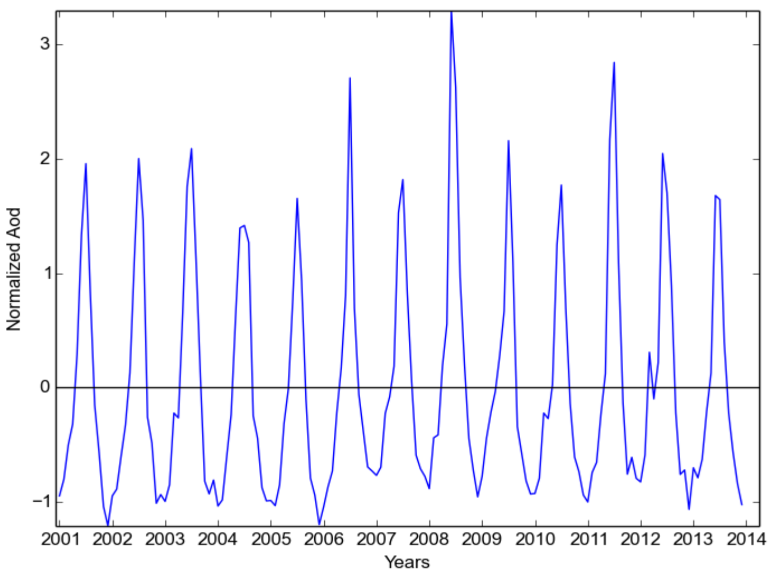

Figure 2: Time series of the first mode of aod

Figure 1 shows first mode of aod and corresponding principle component is shown in figure 2. We found that for first mode, first two eigenvectors give maximum variation, which is 43%. It takes 16 iteration to converge when iteration EOF method was used using only first two eigenvectors. First mode shows high aod values over the western India in particular over the that desert. Corresponding principle component shows seasonal pattern of aod over the norther India. With high aod in pre monsoon period and relatively low aerosols in winter time. The pre monsoon high aod is contributed mainly from the dust from the Thar and the Arabian peninsula.

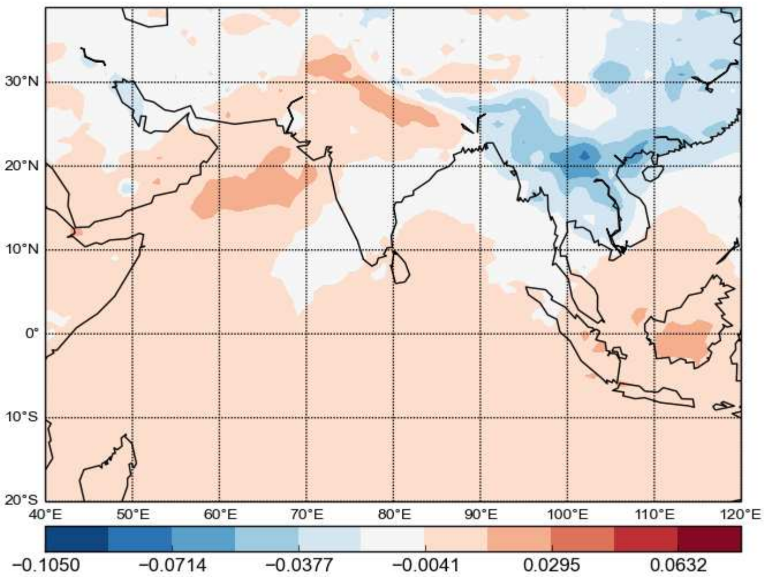

Figure 3: Second mode of aod

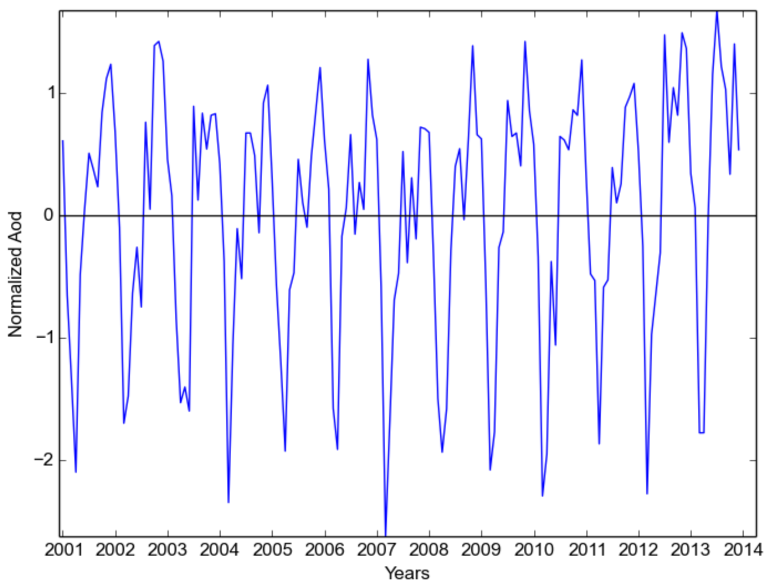

Figure 4: Time series of the second mode of aod

Figure 3 shows second mode of aod and corresponding principle component is shown in figure 4. We found that for second mode, first four eigenvectors gives maximum variation, which is 12%. It takes 19 iteration to converge when iteration EOF method was used using only first four eigenvectors. Second mode shows high aod values over the Indo-Gangetic plain. Corresponding principle component shows higher aod values after monsoon, which is mainly pollution generated and relatively low aerosols in late winter time.

Figure 5 shows third mode of aod and corresponding principle component is shown in figure 6. We found that for second mode, first five eigenvectors gives maximum variation, which is 9%. It takes 22 iteration to converge when iteration EOF method was used using only first five eigenvectors. Second mode shows low aod values over the Indo-Gangetic plain. Corresponding principle component shows two high aod peak, one pre monsoon and one for post monsoon. The time series also shows two low peaks, One during the monsoon and the other in the winter months. Fourth mode shows pattern similar to third mode but variation described by it is only 4%. It follows that pattern shown by mode three and four is distributed among them.

The iterative method described in this paper, is sensitive to initial values. In this study we have used mean values to provide initial values. When we have used, co-kriging interpolation method to fill missing values (in our other study) and initialize the iterative EOF system with interpolated aod values, the system convergence faster for low number of eigenvectors. The signal extracted from this system among first four modes also decreased. It may be

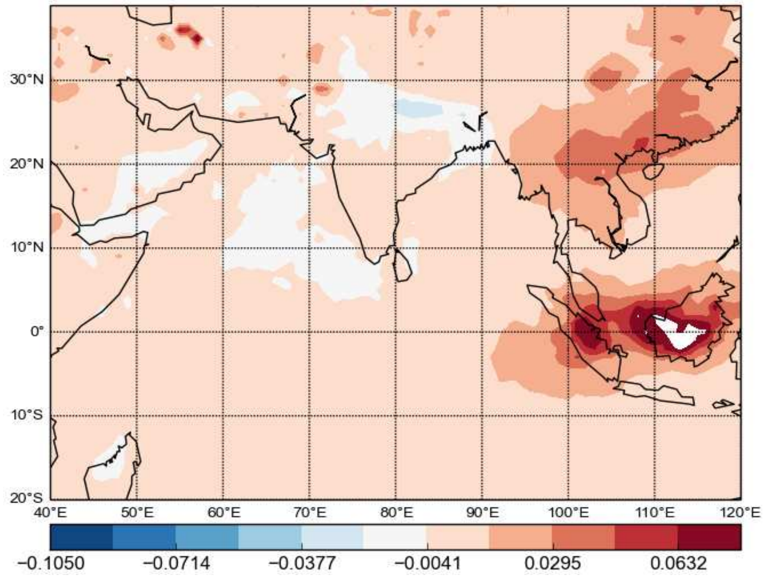

Figure 5: Third mode of aod

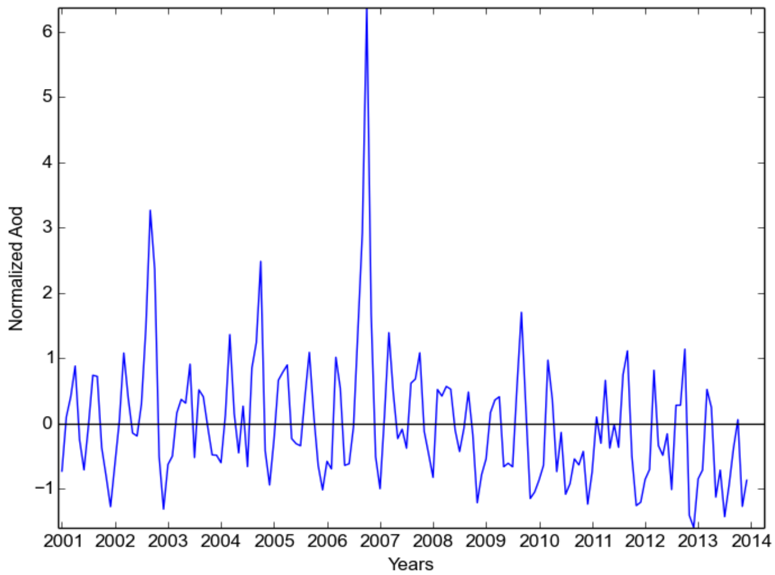

Figure 6: Time series of the third mode of aod

because of using interpolation, smoothing occurs in the spatial pattern. The dependence of iterative EOF system on initial values shows how unstable a aod process pattern can be obtained in the absence on missing values in a big cluster. Such areas are the Thar and the Himalaya, where MODIS aod product values are not available.

Acknowledgement: We thank MODIS team for providing, establishing and maintaining the AOD data set for research.

References

J. M. Beckers and M. Rixen. Eof calculations and data filling from incomplete oceanographic datasets*. *Journal of Atmospheric and Oceanic Technology*, 20(12):1839–1856, Dec 2003. ISSN 0739-0572. doi: 10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2. URL `http://dx.doi.org/10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2`.

R. Gautam, N. C. Hsu, K.-M. Lau, and M. Kafatos. Aerosol and rainfall variability over the indian monsoon region: distributions, trends and coupling. *Annales Geophysicae*, 27(9):3691–3703, 2009. doi: 10.5194/angeo-27-3691-2009. URL http://www.ann-geophys.net/27/3691/2009/.

R. Gautam, N. C. Hsu, S. C. Tsay, K. M. Lau, B. Holben, S. Bell, A. Smirnov, C. Li, R. Hansell, Q. Ji, S. Payra, D. Aryal, R. Kayastha, and K. M. Kim. Accumulation of aerosols over the indo-gangetic plains and southern slopes of the himalayas: distribution, properties and radiative effects during the 2009 pre-monsoon season. *Atmospheric Chemistry and Physics*, 11(24):12841–12863, 2011. doi: 10.5194/acp-11-12841-2011. URL http://www.atmos-chem-phys.net/11/12841/2011/.