Multivariate analysis on Indonesian forest fire using combined empirical orthogonal function and covariance matrices

S Nurdiati¹, A Sopaheluwakan², A Agustina¹ and P Septiawan¹
¹Department of Mathematics, IPB University, Bogor 16680, Indonesia
²Research and Development Center, Meteorology, Climatology and Geophysics Agency, Jakarta, 10720, Indonesia
*E-mail: nurdiati@ipb.ac.id, s.nurdiati@gmail.com

Abstract. Forest and land fires are annual events that occur very frequently during the dry season in several regions in Indonesia. One of the main triggers is low rainfall and extreme anomalies in those regions. The objective of the research was to find the correlation between rainfall and Forest fire based on Tropical Rainfall Measuring Mission (TRMM) rainfall data and Global Fire Emission Database (GFED) forest fire data. Both of these data are big data, this research utilized singular value decomposition to generate information on the correlation of fire patterns and rainfall spatially and temporally. Empirical Orthogonal Function (EOF) based on Singular value decomposition (SVD) was used to reduce large data, without removing main information from the data. SVD approach in this research was conducted using the covariance matrix and combined EOF. EOF mode contribution is able to explain the joint variability of the data set to the phenomenon of fire and its relation to rainfall in Indonesia. Spatial patterns in this analysis show that Kalimantan, Sumatra, and Papua emerged as endemic areas for fires in 1998, 2002, 2006, 2009 and 2015. The year of 2004 and 2014 was also recorded as years of severe forest and land fires but was relatively not as severe as mentioned earlier. The temporal pattern shows that the characteristics of Forest fire in Indonesia are dominant in the final months of the year, from June to October.

1. Introduction
A forest fire has become an interesting issue globally due to El Nino Southern Oscillation (ENSO) in 1997/1998 that caused 25 million ha land burned. In some decades, Endrawati said that the biggest forest fire in Indonesia happened in 1982, 1983, 1991, 1994, 1997, 1998, 2006 and 2015. In 2015, 80% of Sumatra, Kalimantan, and Papua were covered by the smoke caused by Forest fire [2]. The economic loss estimated by Word Bank Group was more than 221 billion rupiah or similar to 1.9% of Gross Domestic Product Indonesia in 2015 [12]. This loss has caused 2.61 million ha forest that was burned during 2015. Forest fire in Indonesia occurred very frequently during the dry season in Indonesia. So, rainfall and anomalies have become one of the main factors that caused Forest fire. In the dry season, low rainfall with extreme anomalies and high intensity of sunlight caused dry land especially in peatland which leads to a forest fire. This condition will be getting worse if there is ENSO that characterized by increasing sea surface temperature is unnaturally in equatorial Pacific.

Research about forest fire and factor that caused it was very important to increase the ability to predict Forest fire. Can also be used as a source of policy on how to handle forest fire in the future [13]. Global
Forest Watch said that historical data of Forest fire, lead prediction to Forest fire in the future [5]. One of the data that saves historical data of Forest fire was the Global Fire Emissions Database (GFED). GFED was monthly data with 0.25° x 0.25° spatial resolution. GFED is not only save burned area data but also consists of some variable such as emissions, biosphere and so on. Besides Forest fire data, we also use rainfall data that is Tropical Rainfall Measuring Mission (TRMM) produced by Multi-satellite Precipitation Analysis (TMPA) having same spatial and temporal resolution such as those of GFED data.

Both of data matrices have big dimension spatially and temporally, therefore they need a big memory and more complex computation. This research used singular value decomposition (SVD) which can reduce large data, without removing important information from the data. SVD approach in this research was conducted using the covariance matrix and combined empirical orthogonal function (EOF). EOF mode is able to explain the joint variability of the data set to the phenomenon of fire and its relation to rainfall in Indonesia.

The objective of the research was to find the correlation between rainfall and forest fire based on TRMM and GFED data using multivariate analysis of EOF based on SVD. It is also meant, to analyze spatial and temporal pattern obtained from those variables. This research is expected to get a correlation pattern between forest fire and rainfall, that can be used to predict Forest fire in the future.

2. Materials
Data used in this research was the Global Fire Emissions Database version 4 (GFED4) and monthly Tropical Rainfall Measuring Mission (TRMM 3B43) version 7. Both of those data are satellite data with HDF format. GFED4 data was produced by Oak Ridge National Laboratory (ORNL) and distributed by Distributed Active Archive Centre (DAAC) that can be downloaded freely from http://daac.ornl.gov/. This is monthly data has 0.25° x 0.25° spatial resolution. TRMM 3B43 has the same spatial and temporal resolution as GFED4. TRMM was produced by National Aeronautics and Spaces Administration (NASA) Goddard Earth Science Data Information and Services Centre (GES DISC) and Japan Aerospace Exploration Agency (JAXA). TRMM data can be downloaded freely from http://disc.gesdisc.eodis.nasa.gov/. We used data from January 1998 up to December 2016.

3. Methods
Singualr Value Decomposition with Covariance Matrices Principle
Defined matrix S and P as two data that having same size resolution m x n, with m was spatial dimension and n was time dimension. Covariances matrices between those data were $R_{SP} = S \times P^T$ as follows:

$$R_{SP} = \begin{bmatrix}
\langle S_1 P_1 \rangle & \langle S_1 P_2 \rangle & \ldots & \langle S_1 P_{M_P} \rangle \\
\langle S_2 P_1 \rangle & \langle S_2 P_2 \rangle & \ldots & \langle S_2 P_{M_P} \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle S_{M_S} P_1 \rangle & \langle S_{M_S} P_2 \rangle & \ldots & \langle S_{M_S} P_{M_P} \rangle 
\end{bmatrix}$$

with each of element $\langle S_i P_j \rangle$ was spatially cross covariance between $S_i$ and $P_j$ at location i and j. Using singular value decomposition to that matrix we obtained $R_{SP} = U \sum V^T$, where $U$ and $V$ was orthogonal matrices and $\sum$ was diagonal ma atrix that contain singular vthe alue of $R_{SP}$. Every column of $U$ was singular va ector of $S$ that called as left patte the and every column of $V$ was singular va ector of $P$ that called as right patta ern. Expansion coefficient of $S$ and $P$ was obtained by projecting every matrix to its singular vector. $A = U^T \times S$ and $B = V^T \times P$, where every row of $A$ and $B$ was expansion coefficients or principal components from $S$ and $P$ [11].

Singular Value Decomposition with Combined EOF Principle
Combined EOF is a generalization of EOF which can be considered as an EOF with combined parameter. Combined EOF is a method that allows simultaneous consideration of modes of different variable variations. One way is to combine several different data by equating the variables owned by...
both data. Defined a new data set by combining variables in data. The data matrix is then arranged in such a way that the bounded fields at the same time are arranged into the same column.

The SVD-based EOF method is used to reduce the data matrix into computer programs using the software. Suppose \( X \) is any matrix measuring \( n \times p \) with rank \((X) = r\). Singular value decomposition (SVD) or the singular value decomposition of \( X \) is a factorization in the form of:

\[
X = U \Sigma V^T.
\]

\( U \) and \( V \) are left and right singular vector \( X \). The variant of \( i \) main component \( i = 1,2, ..., r \) can be calculated by:

\[
\mu_i = \frac{\sigma_i^2}{\sum_{i=1}^{r} \sigma_i^2}
\]

with \( \mu_i, i = 1,2, ..., r \) is the singular value of matrix \( X \). In practice, \( k \) mode of EOF1 or the first major component with \( k \ll r \) explains the proportion of the largest variant. EOF2 mode is a linear combination of all observed variables that are orthogonal to EOF1 mode and has the second largest variant and so on. Therefore, the \( k \) EOF mode has a maximum variance of \( k \) and does not correlate to the previous EOF mode [6].

4. Result and Discussions
The first result obtained from using covariant matrices principle. In this analysis, we used two types of standardization. There are detrend that removes linear trend in the data by reducing each of data with a mean of each column or time domain and zscore standardization that transforms the mean of data into 0 with standard deviation 1. The weakness of using zscore standardization was could not determine which one is the strong event or weak event due to variability loss of the data. In this research, we compare the result obtained from original, detrend, and zscore data. The result of the SCF percentage of biggest EOF mode with these treatments can be seen in Figure 1.

![Figure 1. SCF of the biggest EOF mode in covariance matrices using (a) original data, (b) detrend, and (c) zscore standardization.](image)

Figure 1 shows the first mode that obtained by analyzing original data represents the joint variability of GFED and TRMM 3B43 data to forest fire event by 91.75% and the second mode by 6.51%. First treatment using detrend standardization obtained 2 dominant modes, that each of them can explain the joint variability by 71.8% and 20.57%. In line with this, zscore standardization also gives 3 dominant modes with 43.96%, 29.5%, and 14.17%.
The first analysis using original data obtained one very dominant mode that can explain joint variability of GFED and TRMM 3B43 data by 91.75%. Figure 2 shows the spatial and temporal pattern of joint variability that was created by the first mode. We can conclude that forest fire happened in Riau, Palembang, South Kalimantan, East Kalimantan, and Merauke (Papua). All of those regions have a large burned area in 1998 until 2016 that was shown by dark red in the spatial pattern. Besides that, there was a small forest fire event that happened in Nusa Tenggara and Maluku. The temporal pattern shows that forest fire happened almost every year in Indonesia at half the end of the year. The biggest forest fire happened in 2002, 2006, 2009, and 2019 with above of 6 million ha area was burned in those years respectively. Furthermore, above 4 million ha burned areas in 1998, 2004 and 2014 but still below 6 million ha burned areas. By analyzing original data with covariance matrices, principle SVD obtained one very dominant pattern that showed all of the regions that have forest fire event in 1998 until 2016 and how much area that has burned in every event. A second analysis was used the detrend standardization. As mentioned before, from this treatment we obtained two dominant modes which have cumulative SCF 92.37%. Spatial pattern in both modes give the same area as those of first analysis, but some area appear in the first mode and the others appear in the second mode. The spatial and temporal pattern of mode 1 and 2 using detrend standardization was shown in Figure 3.

The first mode in Figure 3 showed that big forest fire has ever happened in Palembang, South Kalimantan, and South Papua. This event happened in 2002, 2006, 2009, and 2015 which confirmed to the result of analysis from original data. Unusual result occurred at temporal of the first mode in 1998 that showed a negative number of burned area. Actually, that is not a negative number, It explained further by looking at the second mode. Spatial pattern in the second mode shown the very large burned area in East Kalimantan which happened only in 1998. We can conclude that in 1998 there was also forest fire in Palembang, South Kalimantan, and Merauke (Papua), but in this first mode was dominated by an event that happened in East Kalimantan. Furthermore, the temporal pattern of the first mode in 1998 shown a negative number.
The first of two modes using detrend standardization gives an area that is similar to the first analysis in the original data. Detrend standardization can separate two patterns, there is a region that almost every year has forest fire and region that only has forest fire in 1998. But, in detrend analysis, spatial pattern in the first and second mode did not show big forest fire activity happened in Riau. In fact, Riau was a region that also almost every year suffers forest fire which same as shown by the spatial pattern in the original data analysis.

The last treatment in covariant principle used zscore standardization. Zscore standardization will transform the mean of data into 0 and standard deviation into 1 on every time domain. With this treatment we will get data around -1 until 1 and loss the variability in the magnitude of the data. As shown in Figure 1, zscore standardization treatment obtained three dominant modes with total SCF is 86.73%. If compared to other treatments, zscore standardization obtained more dominant modes with the other treatment. The first mode in the first and second treatments explained the joint variability of GFED and TRMM datasets above 70%. The first mode in zscore only can explain the joint variability of about 43.06%.

Figure 4. Spatial and temporal pattern of mode 1 (top), mode 2 (middle) and mode 3 (bottom).

Spatial pattern in the first mode using zscore standardization showed that forest fire happened in Riau, Palembang, South Kalimantan, and Merauke. In the second mode, a forest fire happened in Riau and East Kalimantan and the third mode showed that forest fire only happened in Riau. The temporal pattern in all modes in this treatment showed that in 1998 until 2016 forest fire happened with the same frequency in every year. This is caused by zscore standardization that makes all data have the same variability in all time.

Loss of the variability of the data made the temporal pattern hard to determine a big or small event. Instead of we can only determine when forest fire has happened without knowing the scale of the forest fire. However, we still can get the pattern dominated by another pattern shown in the spatial pattern of mode 2 and 3. Focusing on analyzing the third mode, we can conclude that Riau was also had forest fire at the beginning of the year, different from the temporal pattern shown in the first mode. Furthermore, if we combined the first and third mode, we can conclude that Riau had 2 event forest fire in one year, happened at the end of the year, and the beginning of the year.
The second result obtained by using the coupled SVD principle. This result was similar to using covariance matrices principle. With this principle, we obtained 2 dominant modes from original data and detrend standardization data. The zscore standardization also obtained 3 dominant modes the same with the first result.

![Pie charts showing dominant modes](image)

**Figure 5.** SCF of biggest EOF mode in coupled SVD using (a) original data, (b) detrend, and (c) zscore standardization.

The dominant modes from the original data shown in Figure 6. The first mode showed that forest fire only happened in east of Kalimantan and only happened in 1998. The second mode showed that forest fire happened in Palembang, South Kalimantan, and Merauke (Papua). The second mode pattern happened almost every year, with the big event happened in 2002, 2006, 2009, and 2015. We conclude that without standardization, the data was dominated by a forest fire that happened in East Kalimantan. However, we don’t get the pattern of forest fire in Riau that dominated by the first mode.

![Spatial and temporal pattern](image)

**Figure 6.** Spatial and temporal pattern of mode 1 and 2 from original data analysis.
Figure 7. Spatial and temporal pattern of mode 1 and 2 from detrend analysis.

Figure 6 and Figure 7 showed two identical patterns in the first and the second mode. Detrend standardization did not give a significant difference with the original data. From this fact, we conclude that there is no significant linear pattern in the original data constructed by coupled SVD principle. Because of detrending standardization was processed to remove the linear pattern, there is no difference pattern that appears in the dominant pattern.

Figure 8. Spatial and temporal pattern of mode 1, 2, and 3 from zscore analysis.

Figure 8 showed that zscore standardization decreased the domination of data and produced many patterns than those of original and detrend standardization. Figure 8 showed that the spatial pattern of mode 1, 2, and 3 had a similar pattern with those of forest fire happened in Riau, Palembang, South Kalimantan, and Merauke. East Kalimantan that appeared in the original data and detrend standardization had disappeared from this pattern. This result caused by an event in East Kalimantan that only happened in 1998, so there is no pattern in that event when using zscore standardization.
The conclusion of this research is there was an annual pattern of forest fire in Indonesia happened at half end of the year. This pattern happened during the dry season in Indonesia happened around August. This pattern caused by a forest fire in Riau, Palembang, South Kalimantan, and Merauke. Then, there was a pattern caused by a forest fire in Riau that happened annually at the beginning of the year. Consequently, Riau has 2 patterns of forest fire annually, which are in the beginning and the end of the year. Furthermore, we can confirm that this event was corresponding to rainfall period in Riau that has a twice dry season in one year.

Acknowledgments
This research was financed by Direktorat Riset dan Pengabdian Masyarakat, Direktorat Jenderal Penguatan Riset dan Pengembangan, Kementerian Riset, Teknologi, dan Pendidikan Tinggi, as national postgraduate research team contract 2017. No: 011/SP2H/LT/DRPM/IV/2017, April 20th, 2017.

5. References
[1] Chen CC, Lin HW, Yu JY, Lo MH 2016 The 2015 Borneo fires: What have we learned from the 1997 and 2006 El Niños? IOP Publishing Environ. Res. Lett 11 104003
[2] Endrawati 2016 Analisa Data Titik Panas (Hotspot) dan Areal Kebakaran Hutan dan Lahan (Jakarta: Direktorat Inventarisasi dan Pemantauan Sumber Daya Hutan, Direktorat Jenderal Planologi Kehutanan dan Tata Lingkungan, KLKH)
[3] Fanin T, Van der Werf GR 2017 Precipitation–fire linkages in Indonesia (1997-2015) Biogeosciences, 14, 3995–4008
[4] Fernandes K, Verchot L, Baethgen W, Gutierrez-Veles V, Pinedo-Vasquez M, Martius C 2017 Heightened fire probability in Indonesia in non-drought conditions: the effect of increasing temperatures IOP Publishing Environment. Res. Lett. 12 054002
[5] [GFW] Global Forest Watch 2018 Riwayat kebakaran di Indonesia untuk mencegah kebakaran di masa depan Available on: http://blog.globalforestwatch.org/
[6] Hannachi A 2004 A Primer for EOF Analysis of Climate Data (Reading: Department of Meteorology, University of Reading)
[7] Harris N, Minnemeyer S, Stolle F, Aris OP 2015 Indonesia’s Fire Outbreaks Producing More Daily Emissions than Entire US Economy World Resources Institute
[8] Navarra A, Simoncini V 2010 A Guide to Empirical Orthogonal Function for Climate Data Analysis (New York: Springer)
[9] Nicholson WK 2001 Elementary Linear Algebra (Singapore: McGraw-Hill)
[10] Van dur Werf GR, Randerson JT, Giglio L, Collatz GJ, Mu M, Kasibhatla PS, Morton DC, DeFries RS, Jin Y, Van Leeuwen TT 2010 Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009) Atmospheric Chemistry and Physics, 10 11707-11735
[11] Venegas SA 2001 Statistical Methods for Signal Detection in Climate (Denmark: University of Copenhagen)
[12] [World Bank Group] 2016 The Cost of Fire: An Economic Analysis of Indonesia’s 2015 Fire Crisis (Jakarta: World Bank)
[13] Yang J, H Tian, B Tao, W Ren, J Kush, Y Liu, and Y Wang 2014 Spatial and temporal patterns of global burned area in response to anthropogenic and environmental factors: reconstructing global fire history for the 20th and early 21st centuries. J Geophys Res Biogeosci. 119:249–263. doi:10.1002/ 2013JG002532