Analysis on atmospheric pressure, temperature, and wind speed profiles during total solar eclipse 9 March 2016 using time series clustering

Lala Septem Riza\(^1\), Yaya Wihardi\(^1\), Enjang Ali Nurdin\(^1\), Nanang Dwi Ardi\(^2\), Cahyo Puji Asmoro\(^2\), Agus Fany Chandra Wijaya\(^2\), Judhistira Aria Utama\(^2\), and Asep Bayu Dani Nandiyanto\(^3\)

\(^1\) Department of Computer Science Education, Universitas Pendidikan Indonesia
\(^2\) Department of Physics Education, Universitas Pendidikan Indonesia
\(^3\) Department of Chemistry Education, Universitas Pendidikan Indonesia

E-mail: lala.s.riza@upi.edu

Abstract. Air temperature, pressure, and wind speed measurements on the surface taken during the Total Solar Eclipse (TSE) of March 9, 2016, are made. They were taken in Terentang Beach, Bangka Island, Indonesia. In this paper, we propose to analyze them by using time series clustering. The following steps are conducted: data collecting, splitting, smoothing, distance calculation, and clustering. The final results show cluster memberships of the three parameters on 3 time frames: one day before, the TSE day, and one day after. After doing some simulations, it can be seen that the profiles of temperature and pressure on the TSE day are on the same cluster while the wind-speed profile on the TSE day is the same as on the one day after.

1. Introduction

Analysis on temperature, pressure, and wind speed profiles during the solar eclipse has been done by some researchers, such as in [1–3]. In other words, it is an important task to do in astrophysics since solar eclipse, especially for TSE, provides rare opportunity for studying on the eclipse effects focused on the upper and lower atmosphere.

Basically, similar to the previous studies, this research is also aimed to analyze the profiles during TSE of March 9, 2016 in Bangka Island, Indonesia. However, this task is not done based on common approaches in the astrophysics field, but we will discover and analyze it from another perspective, which is based on machine-learning (ML) approaches for time series analysis. ML can be defined as the field of scientific study focusing on algorithms that are able to learn from data [4]. In this research, we use an essential technique in ML: clustering. Furthermore, to accomplish the experiment, we utilize the programming language R [5].

The remainder of this paper is structured as follows. Section 2 briefly introduces to R ecosystems and time series clustering. In Section 3 and 4, we describe the experimental design, its implementation on R, and results. Finally, Section 5 concludes the paper.

2. R ecosystem and time series clustering

R ecosystem provides a programming language and scientific-analysis environment that consists of more than 8000 packages for statistics, visualization, machine learning, etc. Furthermore,
interested readers can find some useful books introducing $R$, such as [6]. In $R$, there are two following repositories saving $R$ packages: the Comprehensive $R$ Archive Network (CRAN, http://cran.r-project.org/) and Bioconductor (http://www.bioconductor.org/). For example, in CRAN we can find following packages included in the Task View “Machine Learning:” frbs [7, 8] and RoughSets [9].

Time series clustering is an unsupervised learning aimed to separate a set of time series data according to their distances or similarities. In this case, we are using an $R$ package, called TSclust [10]. It contains four sets of dissimilarity measures: model-free (e.g., dynamic time warping (DTW)), model-based approaches (e.g., Maharaj distance), complexity-based approaches (e.g., compression-based dissimilarity measures), and prediction-based approaches. Moreover, interesting articles that discuss time series clustering can be found in the literature, such as in [11].

3. Experimental design

This research follows the experimental design illustrated in figure 1. Basically, it contains 5 steps as follows:

(i) Data acquisition: In this step, we take the data consisting of temperature (in °C), pressure (in milibar), and wind speed (in inchHg) at Terentang Beach (i.e., 2°26′38.77” South, 106°19′36.01” East and on the height of 2.5 m above sea level). The data, which contain 163893 points, are available from at 6:10:41 WIT (Western Indonesia Time, (UTC+8)) on March 8, 2016 until at 9:52:01 WIT on March 10, 2016.

(ii) Data splitting: This step is aimed to divide the data into 3 time frames corresponding to 3 days (i.e., Before, TSE, and After) with the time range: 6:10:41 and 9:52:02 WIT. This range already covers the TSE’s time. So, we will obtain 6 curves for all parameters.

(iii) Data filtering/smoothing: To make the profiles smooth, we use the linear filtering technique on time series data.

(iv) Dissimilarity calculation: We calculate distances among 3 time frames for each parameter. To ensure the results, we conduct 7 dissimilarity methods: Euclidean (EUCL), DTW, temporal correlation coefficient (CORT), integrated periodograms (INT.PER), invertible ARIMA (AR.PIC), invertible and stationary ARMA (AR.MAH), and complexity invariant (CID).

(v) Data clustering: In this research, we are using Hierarchical Clustering, which is a method of cluster analysis to build a hierarchy of clusters.

Figure 1. Experimental design using time series clustering.
4. Experiments and results
In this experiment, the R package TSclust is used to analyze our data. So, the following codes are run on the R environment. Furthermore, by considering 5 steps as illustrated in Section 3 we conduct the following experiments:

(i) Data acquisition: We assume that the data containing temperature, pressure, and wind speed are saved in datasetTSE.csv, and will be loaded into the R object data by the below code:

```
R> data <- read.csv("datasetTSE.csv", header=FALSE)
```

(ii) Splitting into 3 time frames: It can be done by the following code:

```
R> dayBefore <- data[start[1] : end[1], ]
```

It can be seen that we define starting and ending points, which are at 6:10:41 and 9:52:02 for each day.

(iii) Filtering the data: After obtaining the profiles of temperature, pressure, and wind speed for each day, we do smoothing by typing the below code:

```
R> dayBefore_FilterT <- filter(dayBefore[, 2], filter = rep(1/500, 500))
```

It should be noted that the above code is only used for filtering the temperature profile of the first time frame (one day before). For other parameters, we can do the similar code. We can also plot the profile, for example for the temperature curve (see figure 2):

```
R> plot(dataT, type="l")
```

![Figure 2. The profiles of temperature after splitting and smoothing.](image)

(iv) Calculating Dissimilarities: The distances among the time frames for each parameter will be determined by the following code:

```
R> T_comb <- cbind(T_Before, T_TSE, T_After)
R> T.eucl <- diss(T_comb, "EUCL")
```

It should be noted that because of the limited space, we just show the code for calculating dissimilarity of temperature by using Euclidean equation. The other parameters can be done by similar codes by following the ways on [10]. The result of this step can be seen in table 1.

(v) Clustering on each Parameter: After obtaining the distances, we can now determine the membership of clusters by typing the following code:

```
R> T.eucl.hclus <- cutree(hclust(T.eucl), k = 2)
```

For the other methods of dissimilarities can be done by executing similar codes. The final result can be summarized as illustrated in table 2. It can be seen that we can state that the temperature and pressure profiles on the TSE day are quite similar to the ones on one day before. On the other hand, the wind speed profile on the TSE day is closed to the one on one day after.
Table 1. The distances of temperature, pressure, and wind speed among 3 time frames using the Euclidean equation.

|                  | Distances of Temp. (T) | Distances of Press. (P) | Distances of Wind Speed (WS) |
|------------------|------------------------|------------------------|-----------------------------|
|                  | T_Before | T_TSE   | P_Before | P_TSE   | WS_Before | WS_TSE   |
| T_SE             | 111.392  |         |          |          |           |          |
| T_After          | 208.011  | 208.323 |          |          |           |          |
| P_TSE            | 19.067   |          |          |          |           |          |
| P_After          | 215.670  | 219.899 |          |          |           |          |

Table 2. The cluster membership of temperature, pressure, and wind speed for 3 time frames.

|                  | Cl. Membership of T Before | Cl. Membership of P Before | Cl. Membership of WS Before |
|------------------|----------------------------|---------------------------|----------------------------|
|                  | TSE | After | TSE | After | TSE | After |
| EUCL_hclust      | 1   | 2     | 1   | 2     | 1   | 2     |
| CORT_hclust      | 1   | 2     | 1   | 2     | 1   | 2     |
| DWT_hclust       | 1   | 2     | 1   | 2     | 1   | 2     |
| INT.PER_hclust   | 1   | 2     | 1   | 2     | 1   | 2     |
| AR.PIC_hclust    | 1   | 2     | 1   | 2     | 1   | 2     |
| AR.MAH_hclust    | 1   | 2     | 1   | 2     | 1   | 2     |
| CID_hclust       | 1   | 2     | 1   | 2     | 1   | 2     |

5. Conclusion and future work
The research can be summarized as follows: (i) a strategy to analysis temperature, pressure, and wind speed by using time series clustering has been presented; (ii) according to the experiments, the profiles of temperature and pressure on the TSE day are the same behavior, represented by the same cluster while the wind-speed profile on the TSE day is the same characteristic as on the one day after. As future work, we plan to extend the analysis on other areas of Indonesia influenced by partial solar eclipse and total solar eclipse, so that we will have a comprehensive understanding on the profile changes.

References
[1] Anderson R, Keefer D and Myers O 1972 Journal of the Atmospheric Sciences 29 583–587
[2] Fernández W, Castro V and Hidalgo H 1993 Earth, Moon, and Planets 63 133–147
[3] Wang K Y and Liu C H 2010 Geophysical Research Letters 37
[4] Kovahi R 1998 Machine Learning 30 271–274
[5] Ihaka R and Gentleman R 1996 Journal of computational and graphical statistics 5 299–314
[6] Team R D C 2008 An Introduction to R (Vienna, Austria: R Foundation for Statistical Computing)
[7] Riza L S, Bergmeir C, Herrera F and Benítez J M 2015 Journal of Statistical Software 65 1–30
[8] Riza L S, Bergmeir C, Herrera F and Benítez J M 2014 Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on (IEEE) pp 2149–2155
[9] Riza L S, Jannsz A, Bergmeir C, Cornelis C, Herrera F, Slezak D, Benítez J M et al. 2014 Information Sciences 287 68–89
[10] Montero P and Vilar J 2014 Journal of Statistical Software 62 1–43
[11] Warren Liao T 2005 Pattern Recognition 38 1857–1874