Jurnal Varian
Vol. 8, No. 1, October 2024, pp 25–38
e-ISSN: 2581-2017

Analyzing Lightning Strike Susceptibility Using the Elliptical Fitting Method with a Principal Component Analysis Approach

Helmalia A. Lovytaji[1], Rozikan Rozikan[2], Djati C. Kuncoro[2], Ria D. L. N. Karisma[1]
[1]UIN Maulana Malik Ibrahim, Malang, Indonesia
[2]Badan Meteorologi, Klimatologi, dan Geofisika, Pasuruan, Indonesia

Article Info

Article history:

Received : 07-20-2023
Revised : 07-20-2024
Accepted : 10-29-2024

Keywords:

Cloud to ground lightning;
Ellipse fitting method;
Lightning strike;
Principal component analysis.

ABSTRACT

Lightning is a high-current discharge that occurs in Cumulonimbus clouds, with CG (Cloud to Ground) lightning strikes posing significant dangers, especially to human life. Pasuruan, located in the highlands between mountains and the ocean in Indonesia, is particularly vulnerable to such strikes. This study aims to mitigate the impact of lightning strikes, particularly in industrial areas like Pasuruan, by delineating lightning-prone areas using a sophisticated methodological approach. Our research employs a robust Ellipse Fitting Method, parameterized with Principal Component Analysis (PCA), to accurately define the boundaries of these high-risk zones. The Ellipse Fitting Method, which involves forming an ellipse from the intersection of a plane and a cone, uses five key parameters: a center point, two vertex points, and two focus points. PCA is then applied to these parameters to determine the ellipse's configuration, with the center point derived from the mean of all data points. The major and minor axes are defined by the first and second eigenvalues of the principal components, respectively. The size of the ellipse correlates with the confidence level, with higher confidence resulting in a larger ellipse. The result of integrating these advanced techniques is the generation of two PCA models from data collected across 28 sub-districts in Pasuruan, with findings indicating a high level of vulnerability in Lumbang District and a moderate level of risk in Gempol District. This methodological framework not only enhances the precision in identifying lightning-prone areas but also provides a scalable approach for similar studies in other regions. Suggestion for the further research are to overcome extreme points or extreme points in the PCA confidence ellipse such as MVEE.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v8i1.3183

Corresponding Author:

Ria D. L. N. Karisma,
Department of Mathematics, UIN Maulana Malik Ibrahim Malang,
Email: riadhea@uin-malang.ac.id

How to Cite:
Lovytaji, H. A., Rozikan, R., Kuncoro, D. C., & Karisma, R. D. L. N. (2024). Analyzing Lightning Strike Susceptibility Using the Elliptical Fitting Method with a Principal Component Analysis Approach. *Jurnal Varian*, *8*(1), 25-38.

A. INTRODUCTION

Lightning is a high-current electric discharge whose path length is measured in kilometers. Lightning occurs due to the release of electric charges that are separated in cumulonimbus clouds (Wahjudi, 2014). Lightning discharges based on the type of strike are divided into four types, namely cloud-to-ground (CG) discharge, cloud-to-cloud (CC), inter-cloud discharge (CA), and single cloud discharge (IC). Lightning from the cloud to the ground (CG) is the most dangerous lightning because it can directly affect human activities (Shindo, 2018).

Rapid population growth is then supported by regional development causing the region to become vulnerable to lightning strikes. Pasuruan is one of the areas in East Java which located in the highlands between the mountains and the vast ocean. Cumulonimbus clouds easily form, then turn into rain which is sometimes accompanied by lightning. Lightning strikes can occur during the rainy season. In this season, the air holds more air. The insulation capacity is reduced, and the electric charge is easier to move (Narut et al., 2018). Pasuruan recorded a large population growth rate for Pasuruan from 2010 to 2020 of 1.08% annually (Badan Pusat Statistik Kota Pasuruan, 2020). Furthermore, Pasuruan is known as an industrial city which is an area with a high level of development (Badan Pusat Statistik, 2022). Hence, Pasuruan has some areas that have potential lightning strikes. It has categories of extreme lightning strikes areas which is 68% (Faizatin, 2014). During 2022, Pasuruan is the area most affected by lightning strikes, namely around 700 thousand compared to areas around East Java (Syarief, 2023). Pasuruan is ranked 5th in the region that has the largest industrial area in East Java (Badan Pusat Statistik, 2022). One of the impacts of lightning strikes, especially in industrial areas, can cause fires and large explosions (Mora García et al., 2015).

Principle Component Analysis as known as PCA was discovered in 1901 by Karl Peason. The method use to determine the main components of a data set that explain the largest amount of total variance in the data (Sanderson, 2019). The method changes data dimension by reduced a number of correlated variables to form uncorrelated variables or main components (Greenacre et al., 2022). Therefore, dimension in data become smaller but it can explain the data as a well as before reducing. Research on PCA in the Ellipse Fitting Method has been reviewed in several research sources. Nurunnabi et al. (2019) investigated the new idea of forming an ellipse. In this study, the application of PCA is represent an object, such as a fruit that represented as elliptical parameters. Based on Zhao et al. (2021) highlighted practical algorithms for PCA-based shape fitting. PCA in reducing data dimensionality, filtering noise, and providing robust shape-fitting solutions such as ellipse fitting. It offers a simple, efficient, and interpretable approach to ellipse fitting, making it a valuable tool across various applications.

Despite extensive research on PCA and its application in ellipse fitting, there is a lack of studies specifically addressing the use of PCA and ellipse fitting to analyze lightning strike patterns and vulnerabilities in regions like Pasuruan. This gap highlights the need for research that integrates PCA-based ellipse fitting methods to assess and predict lightning strike risks, particularly in rapidly developing industrial areas with high population growth. Addressing this deficiency could lead to more accurate identification of lightning-prone areas and enhance preventive measures.

In addition, PCA was used to examine the distance of areas that are prone and safe to lightning strikes from the resulting elliptical boundary (Sanderson, 2019). The method reduces the dimensionality of high-dimensional data, making elliptical fitting computationally more efficient. The study aims to discuss the relationship with lightning-prone areas in the Pasuruan area. Areas prone to lightning strikes can be used as a map of lightning-prone areas as input, especially in the private and non-private industrial sectors.

This study aims to classify areas prone to lightning strikes in Pasuruan using ellipse fitting method with PCA as level of confidence to define lightning-prone area. Lightning strikes pose a serious threat to human life and property. By carrying out disaster mitigation such as mapping lightning strike areas, it can help make land use planning decisions. In addition, helps avoid placing important infrastructure in high-risk zones. Furthermore, the novelty of the PCA process is carried out with the help of the elliptical fitting method. Especially in the industrial area, Pasuruan, which is one of the largest industrial areas in East Java (Faunia, 2022). It could be an input for mapping lightning-prone areas, especially in the industrial sector, both private and public, in the future to avoid the danger of lightning strikes.

B. RESEARCH METHOD

1. Data Source

The data used in this study is Cloud to Ground (CG) lightning density data in Pasuruan. The research locations have 24 sub-districts in the Pasuruan district and four sub-districts in Pasuruan City. Lightning density data were obtained from the Meteorology, Climatology and Geophysics Agency (BMKG) Geophysics Station Class II – Tretes, Pasuruan. The lightning density data is used during the rainy season from 2017 to 2021. The instrument or tool used to assist researchers is a record of lightning events for five years obtained from lightning data processing at the Meteorology, Climatology, and Geophysics Agency (BMKG) Class II Geophysics Station - Tretes, Pasuruan with the help of a Lightning Detector.

Lightning is the withdrawal of temporary high-current electric charges, and the measurement length is measured in kilometres (Peterson et al., 2022). Lightning occurs due to the withdrawal of electric charges in cumulonimbus clouds and is followed by a flash of light in the sky and a thunderous sound (Sterpka et al., 2021). However, lightning occurs due to the presence of cumulonimbus clouds which are vertical clouds that are tall, dense and cause thunderstorms and cold weather. Cumulonimbus clouds form from atmospheric instability and form in clusters or along squall lines. Then it forms into cumulus clouds which later become supercells that cause large thunderstorms (Sulistiyono et al., 2023). One lightning strike can release electrical energy reaching millions of volts, which can destroy objects that are struck (Davis et al., 2014). This type of CG lightning is directly related to human activities because it can strike buildings, cut off electricity, and more. Moreover, it is considered to be able to cause material loss or even death. Indonesia is prone to lightning strikes due to its high rate of incidence. An area is said to be prone to lightning strike activity based on the high number of incidents, the high potential for victims or losses, and the density of buildings in an area (Gunawan & Pandiangan, 2014).

This study uses the PCA application on the ellipse by determining the centre point of the ellipse so that a lightning strike-prone area is obtained. Three areas are prone to lightning strikes: low, medium and high. The lightning strike area is determined based on the extreme points or extreme points that are outside the elliptical boundary. Susanto (2018) explains that areas prone to lightning strikes are determined based on predetermined threat level intervals. The following is the definition of the threat level.

$$I_{threat} = \Delta d3$$

When
I_{threat} : The level interval for each lightning strike threat
Δd : The range of data point distances on the ellipse

Based on the type of strike in this study use Cloud-to-ground (CG) strike. CG strike is a strike between clouds. CC lightning is formed due to the discharge of opposite electrical charges in different clouds. Suppose a negatively charged cloud releases its electrons to a positively charged cloud and vice versa.

2. Ellipse Fitting Method

The study used ellipse fitting method with the PCA as level of confidence to define lightning-prone area in Pasuruan. Ellipse is defined as the set of all points (x, y) in a plane that the distance from two different fixed points is constant (Sanderson, 2019). The two distinct fixed points are called focal points or foci $F_1(c, 0)$ and $F_2(-c, 0)$. Mashadi (2018) defines an ellipse as one of the results of the intersection between a plane and a cone. An ellipse has five parameters, namely the center point, two vertices, and two foci. In a plane, ellipse delineation shows in Figure 1.

The general form the elliptic equation for any set (x,y) defined as Thurnhofer-Hemsi et al. (2021).

$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$

Whereas, $B^2 - 4AC < 0$. Based on those the ellips devided by two which is $A < B$ then it horizontal. Whether $A > B$ then it vertical.

Regarding on the center point the elliptical equation defined as:

a. Ellipse equation centered at O (0,0)
The ellipse that has main axis in lies on or parallel to the X axis is $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ then, $a \geq b$. However, the ellipse that has main main axis in lies on or parallel to the Y axis is $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ then, $a \geq b$.

b. The elliptical equation centred at $P(\alpha, \beta)$
Whether (Figure 1), the ellipse is shifted by a distance (α, β) then obtained ellipse that centered in $P(\alpha, \beta)$. Hence, the main axis in lies on or parallel to the X axis in ellipse is $\frac{(x-\alpha)^2}{a^2} + \frac{(y-\beta)^2}{b^2} = 1$, then $a > b$., the main axis in lies on or parallel to the Y axis in ellipse is $\frac{(x-\alpha)^2}{b^2} + \frac{(y-\beta)^2}{a^2} = 1$, then $a < b$.

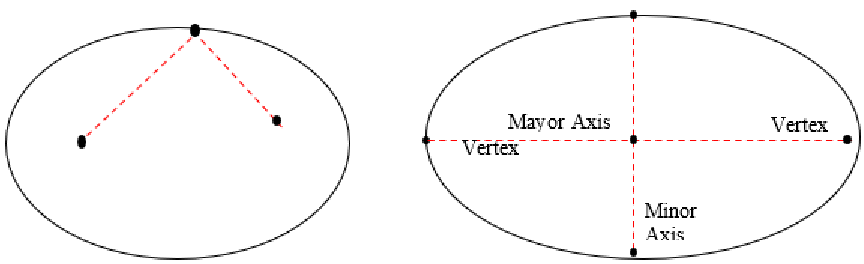

Figure 1. Ellipses and Parameterization

3. PCA application in ellipse

Principle Component Analysis (PCA) is an unsupervised learning technique algorithm that used for dimensionally reduction in machine learning method (Forkman et al., 2019). PCA in multivariate statistics use to develop an ellipse. The method established principal component that explained the number of variance in data. The first principal component contained of the largest number of variance in data (Sun & Wang, 2018). Therefore, the second principal component described the second largest number of variance that is not reduction by the first principal component. Hence, the subsequent principal component explained the largest amount of variance that is orthogonal to the previous principal component (Kherif & Latypova, 2020).

The PCA application in ellipse is formed by steps (Reiml et al., 2016).

a. Find center point of ellipse

The ellipse center determined by finding the average value at all data point. The mean vector is $c = [c_1 c_2]$. Suppose data set X has vector n in m dimension then $x(k) = [x_1(k) \cdots x_m(k)]$.

$$X = [x(1) \vdots x(n)] \tag{1}$$

While, $\hat{X}$ in $\hat{X} = X - c$ that illustrated data X after mean c has been removed.

b. Calculating variance covariance matrix

In data that has 2 dimensions, then $m = 2$. Thus, matrix size is $n \times 2$ and it contained all point coordinates that will be used to form an ellipse. Variance covariance matrix S in vectors m used to measure correlation in each component then,

$$S = \frac{1}{n-1} \hat{X}^T \cdot \hat{X} \tag{2}$$

Where S is positive definite symmetric matrix and $\hat{X}$ is not a zero matrix.

Variance covariance matrix that has $m = 2$ then $S = [\sigma_{11}\ \sigma_{12}\ \sigma_{21}\ \sigma_{22}\ ]$ where,

$$\begin{aligned} &\sigma_{ii} = \tfrac{1}{n-1} \textstyle\sum_{k=1}^{n} [x_i\,(k)]^2\,, \quad i = 1, 2 \text{ and} \\ &\sigma_{12} = \sigma_{21} = \tfrac{1}{n-1} \textstyle\sum_{k=1}^{n} x_1\,(k) \cdot\ x_2(k) \end{aligned} \tag{3}$$

c. Calculating eigen vector and eigen value

Based on variance covariance matrix that formed then eigen vector of v an eigen value λ is

$$S \cdot v = \lambda \cdot v \tag{4}$$

Where S is positive definite symmetric matrix. Hence eigen value in variance covariance S is real. Equation (2.19) can be written

$$S \cdot V = V \cdot \Lambda \; or \; S = V \cdot \Lambda \cdot V^{-1} \tag{5}$$

Where Λ is diagonal matrix in eigen value and V is matrix column contained eigen vector. Then, it will be ortoghonal which is eigen vector equal orthogonal matrix. Hence, it has inverse equal to transpose value as in the Equation 5.

$$S = V \cdot \Lambda \cdot V^T \tag{6}$$

d. Estimation of ellipse formation

Based on process then it has component formed ellipse. The coordinates of the center point of the ellipse are obtained from the average (mean) of all data points. Furthemore, the major axis is obtained from the first eigenvalue of the first principal component. For the minor axis is obtained from the second eigenvalue on the second principal component. The length of the major axis and minor axis from eigen value is determined λ_1 and λ_2 by significance level. Thus, the higher the confidence level then the ellipse will be larger.

The length of the two axes (a_i) is

$$a_i = \sqrt{z \cdot \lambda_i} \qquad i = 1,\ 2 \tag{7}$$

Where z is $-z$ value for a level of α with m, then degree of freedom $n - m$.

4. Research Procedure

The research flow is explained through the flowchart according to Figure 2, and explained in detail through the following steps:

a. Exploring data on Cloud-to-ground (CG) strike by Station Class II – Tretes, Pasuruan
   1. Collecting lightning density data (.txt) during the 2017 to 2021 rainy season
   2. Mapping monthly density data by sub-district with the help of ArcGIS 10.8 software.

b. Perform PCA to data on lightning strike area in Pasuruan using Equation 1 to 6

c. Perform an ellipse with a confidence level of 95% or $\alpha = 0,05$

d. Group the vulnerability level to lightning strikes for each sub-district in the Pasuruan area into three levels.

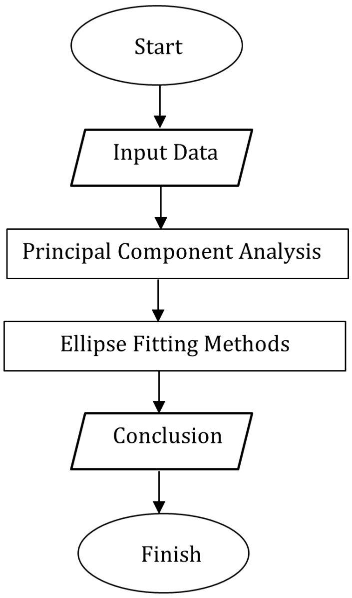

Figure 2. Flowchart of research analysis steps

C. RESULT AND DISCUSSION

1. Description of Lightning Strike Intensity

Figure 3 shows that the lightning strike density in the Pasuruan area varies greatly. Several sub-districts have the highest lightning strikes because there are more than 500 strikes/km2. These areas are Gempol District, Pandaan District, and Lumbang District, as well as several small areas in Sukorejo District. Meanwhile, Prigen and Beji Districts are areas with low lightning strike density. The lightning density data is then interpolated to determine the number of strikes in each sub-district from the rainy season of 2017 to 2021; then there are 30 variables, namely $X_1, X_2, \ldots, X_{30}$.

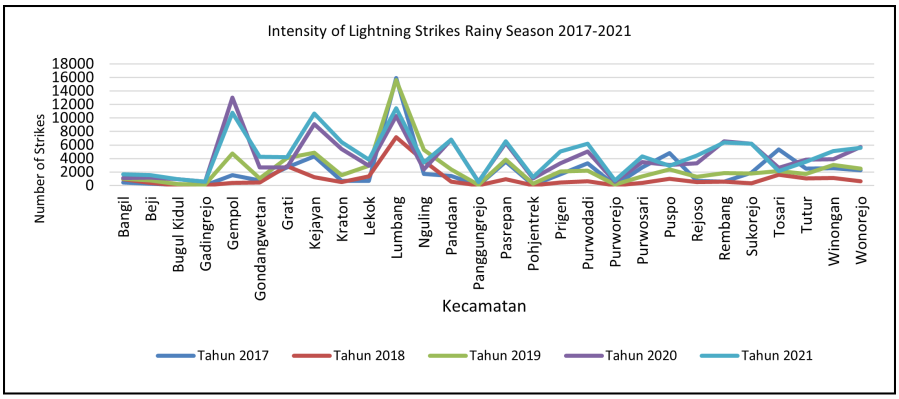

Figure 3. District Lightning Strikes in Pasuruan 2017-2021

DOI: https://doi.org/10.30812/varian.v8i1.3183

The highest intensity of lightning strikes occurred in 2017 in the Lumbang sub-district, with 15919 strikes (Figure 3). Meanwhile, the lowest intensity of lightning strikes occurred in 2018 in the Gadingrejo sub-district, with ten strikes. For five years, the Lumbang sub-district had the highest total lightning strikes, with 60402 strikes, with an average of around 12080 lightning strikes occurring yearly. However, the Panggungrejo sub-district has the lowest total lightning strikes, with 562 strikes, where there are around 250 strikes each year on average. In addition, in 2021, the intensity of lightning strikes in most districts will experience its highest peak in five years. Because the intensity of lightning strikes in 2021 is the highest compared to previous years (Figure 3).

PCA is used in forming ellipses to see areas prone to lightning strikes. The study used R software to help with PCA analysis. The first step is to calculate the variance matrix. The variance matrix is needed to get the eigenvalues and eigenvectors in data units with the same unit. The eigenvalues obtained from the matrix of variances are sorted from the largest to the smallest (Jolliffe & Cadima, 2016). Table 1 shows the corresponding eigen value and eigen vector.

Table 1. Corresponding Eigen Value and Eigen Vector

var	$\mathbf{a_1}$	$\mathbf{a_2}$	$\mathbf{a_3}$	...	$\mathbf{a_{28}}$
X_1	-0,2100	-0,1531	-0,0093	...	-0,1128
X_2	-0,2031	0,0503	-0,2410	...	-0,3146
X_3	-0,2082	-0,1417	-0,0379	...	-0,1734
X_4	-0,1912	-0,1671	-0,1467	...	0,3469
X_5	-0,2043	-0,1563	-0,1189	...	-0,3282
X_6	-0,2069	-0,1562	-0,0716	...	-0,0285
X_7	-0,1935	-0,1536	0,2244	...	-0,0468
X_8	-0,1916	-0,1368	0,2695	...	-0,1324
X_9	-0,1834	-0,1573	0,2080	...	-0,2520
X_{10}	-0,2055	-0,1510	0,0099	...	0,1676
...	...	...	...	...	...
X_{30}	-0,1395	0,2595	0,0478	...	0,0021

Table 1 illustrates that the k main components whose eigenvalues are more than one ($\lambda_i \geq 11$) are the four main components. When an eigenvector has negative components, it means that in a certain direction of the data space, there is a negative influence from the original variables that constitute the data. For instance, an increase in some original variables decreases the principal component represented by that eigenvector. Thus, this vector can be used to measure the correlation of each component, which will be used to form an ellipse. PCA has the proportion of total variance, which explains the contribution of the main components formed and the diversity of the data (Elhaik, 2022).

Table 2. Eigenvalues, Proportion of Variance, and Cumulative Value of Proportion of Variance

PC	**Principle Component Analysis**		
	Eigen Value	**Proportion of Diversity**	**Cumulative Proportion**
1	17,980	0,599	0,599
2	6,944	0,231	0,831*
3	2,145	0,071	0,902
4	1,005	0,034	0,936
5	0,745	0,025	0,961
6	0,310	0,010	0,971
7	0,258	0,009	0,980
8	0,187	0,006	0,986
9	0,121	0,004	0,990
10	0,090	0,003	0,993
...	...	...	...
28	-0,000	-0,000	1,000

Table 2 shows that the cumulative value of the proportion of total diversity, which can explain 80% of the original data, is the two main components. Hair et al. (2019) explained that the selection of the main components used is by taking k main components whose eigenvalues are more than one ($\lambda_i \geq 1$). In Table 2 reveals that the k main components are the four main components. In contrast, Johnson & Wichern (2015) explained that the proportion of diversity considered to represent the total diversity of the data is the k main component with a cumulative value of the proportion of diversity of at least 80%. Whereas,

DOI: https://doi.org/10.30812/varian.v8i1.3183

it denotes that the cumulative value of the proportion of total diversity. It includes at least 80% of the original data, are the two main components.

The cumulative value of the total variance proportion is obtained from the cumulative sum of the variance proportions of each main component formed. Two principal component equations are formed based on the coefficient (loading factor) obtained from the eigenvectors of each principal component.

$$Y_1 = -0,2100\,x_1 - 0,2031x_2 - 0,2082\,x_3 + \ldots - 0,1395\,x_{30} \quad (8)$$

$$Y_2 = -0,1531\,x_1 + 0,0503x_2 - 0,1417\,x_3 + \ldots + 0,2595\,x_{30} \quad (9)$$

The scree plot diagram in Figure 4 shows the bend at $i = 3$. However, in Table 2, represent that the eigenvalues after λ_2 are relatively small and have nearly the same value. Based on these three methods, the two main components most effectively explain the data.

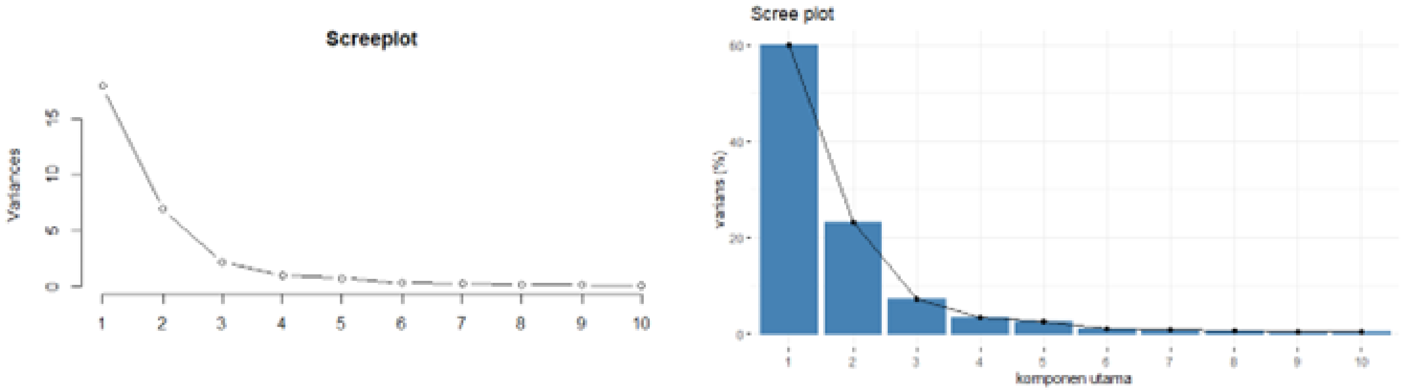

Figure 4. Scree Plot PCA

Figure 4 shows that ellipse has four quadrants in the Cartesian coordinate plane. Based on the variable coefficient values on PC1 and PC2, areas prone to lightning strikes are areas with points in quadrant I, quadrant II, and quadrant III in the Cartesian coordinate plane. Otherwise, there is an area where the point is outside the elliptical boundary, and the location of the point is in quadrant IV, then that area is an area that is not prone to lightning strikes. The grouping of areas prone to lightning strikes is divided into three levels: low, medium and high. The hazard of lightning strikes in an area is determined based on the interval of the threat level of lightning strikes (Susanto, 2018). The distance between the centre points of the ellipse and the data points has been distributed in Table 6. The results of determining the level of vulnerability can be seen in Table 6.

2. Ellipse Fitting Method

In this study, PCA use to form the ellipse fitting method. Based on Nurunnabi et al. (2019), the steps in applying PCA to form an ellipse are as follows:

a. The center point of the ellipse

The average value of the two principal components forms coordinate (x, y). The x-coordinate is the average of the first principal component, and the average value of the second principal component formed the y-coordinate. The average of the first principal component:

$$\underline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2,8789 + 3,3381 + 4,0538 + \ldots - 0,0288}{28} = \frac{-2 \times 10^{-08}}{28} = -7,14 \times 10^{-10}$$

Average of the second principal component:

DOI: https://doi.org/10.30812/varian.v8i1.3183

$$\underline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{-1,7200 - 1,4403 - 1,5740 + \ldots + 1,6452}{28} = \frac{1 \times 10^{-08}}{28} = 3,57 \times 10^{-10}$$

Thus, the coordinates of the center of the ellipse are $(x, y) = (-7,14 \times 10^{-10}, 3,57 \times 10^{-10})$.

1. Calculating the variance matrix
   Variance matrix is obtained as follows

$$S = \begin{bmatrix} 17,3382 & -6,76 \times 10^{-10} & -6,76 \times 10^{-10} & 6,6955 \end{bmatrix}$$

   While the inverse of the matrix S is

$$S' = \begin{bmatrix} 0,0577 & 5,83 \times 10^{-12} & 5,83 \times 10^{-12} & 0,1494 \end{bmatrix}$$

2. Calculating eigenvectors and eigenvalues
   The eigenvalues of the variance matrix S are:

$$S - \lambda I = \begin{vmatrix} 17,3382 - \lambda & -6,76 \times 10^{-10} & -6,76 \times 10^{-10} & 6,6955 - \lambda \end{vmatrix}$$
$$0 = (17,3382 - \lambda)(6,6955 - \lambda) - \left(-6,76 \times 10^{-10}\right)\left(6,76 \times 10^{-10}\right)$$
$$0 = \lambda^2 - 24,0337\,\lambda + 116,0886$$

   then the eigenvalues are $\lambda_1 = 17,3382$ and $\lambda_2 = 6,6955$. The first eigenvalue λ_1 use as the major axis. However, the second eigenvalue λ_2 use as the minor axis.

3. Estimating the formation of an ellipse
   The size of the ellipse is based on the level of confidence. The greater the level of confidence, the larger the resulting ellipse will be. The size of the ellipse is based on the level of confidence. The greater the confidence level, the larger the resulting ellipse will be. In this study, the 95% confidence level will be used or equal to the degree of freedom 2. Hence, the $\chi_2^2(.05) = 5,99$. The first focus point $(F_1(c, 0))$:

$$F_1(c, 0) = \sqrt{\lambda_1 \chi_2^2} = \sqrt{17,3382 \times 5,99} = 10,1922$$

   The coordinates of the focus point $(F_1(c, 0))$ are $(F_1(10,1922, 0))$
   Second focus point $(F_2(-c, 0))$:

$$F_1(-c, 0) = \sqrt{\lambda_1 \chi_2^2}$$
$$= \sqrt{17,3382 \times 5,99} = 10,1922$$

   The coordinates of the focus point $(F_2(-c, 0))$ are $(F_2(-10,1922, 0))$.

Table 3. Determination of Ellipse's Points

Observation	X-bar	Y-bar	Place Point	Observation	X-bar	Y-bar	Place Point
1	2.8789	-1.7200	inside	15	-2.0117	1.1527	inside
2	3.3381	-1.4403	inside	16	3.9062	-1.3889	inside
3	4.0538	-1.5740	inside	17	0.6115	0.6424	inside
4	4.2628	-1.8696	inside	18	-0.8652	1.4407	inside
5	-3.0893	8.0195	outside	19	4,1982	-1.7792	inside
6	1.9476	0.2241	inside	20	0.6421	0.0401	inside
7	-1.2630	-2.1687	inside	21	-0.6872	-2.3335	inside
8	-3.9423	4.3799	inside	22	1.6555	0.3901	inside
9	0.6381	2.1419	inside	23	0.3940	2.6106	inside
10	0.9049	-0.8659	inside	24	0.0231	2.4349	inside
11	-18,3209	-4.0930	outside	25	-1.2563	-3.4626	inside
12	-1.4328	-3.0574	inside	26	-0.2840	-0.7916	inside
13	-0.1748	3.2464	inside	27	-0.3827	0.0883	inside
14	4.2842	-1.9124	inside	28	-0.0288	1.6452	inside

Table 3 illustrates that the slope of the ellipse is obtained from the arc tangent value (x,y) where the -axis is the PC2 variance of the matrix and the axis is the subtraction of the eigenvalues with the PC1 variance of the matrix, then the slope of the ellipse is $-3,1415\pi$ or is -180 (see Table 3). Determine which points are inside or outside the ellipse. The variable coefficient value at PC1 gets smaller, the area will be more vulnerable. It also the most miniature principal component score in PC1 is one of the areas prone to lightning strikes. In addition, if the variable coefficient value at PC2 is more significant, the area will be more vulnerable. It represents the most significant principal component score in PC2, one of the areas prone to lightning strikes.

The ellipse formed in Figure 5 shows 26 of the 28 points inside the ellipse boundary line. In contrast, the other two points are outside the Lumbang District, and Gempol District is outside the elliptical boundary line. It represents that there are deviations or outliers in the data. Sanderson (2019) defines points outside the ellipses or deviations in the data as extreme points. Extreme points are probability theories that explain a random variable's minimum and maximum values.

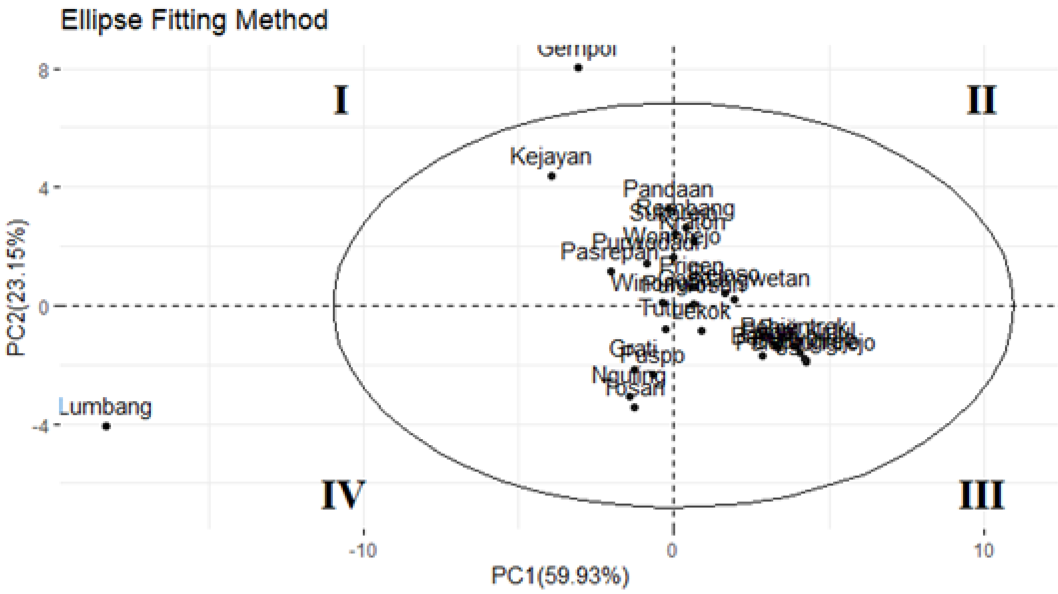

Figure 5. Ellipse Fitting Method

Table 4 shows that Lumbang District has the highest (maximum) total strikes, namely 60402 strikes during the 2017-2021 rainy season, with an average of 2013.40 strikes occurring each month. The lightning strike case has made Lumbang Sub-District known as the maximum extreme point because the number is much higher than the lightning strike cases in other sub-districts. Furthermore, Gempol District, which has 30,351 strikes and an average of 1,011.70 strikes per month, is also the maximum extreme point in the data. Meanwhile, Panggungrejo District has the lowest (minimum) total strikes, with 1253 strikes during the 2017-2021 rainy season, where an average of 41.77 strikes occur each month. Hence, the case of a lightning strike in the Panggungrejo District is the minimum extreme point in the data.

Table 4. Distances on Ellipses

Subdistrict	PC1	PC2	Distance	Total	Mean
Bangil	2,8789	-1,7200	3,3535	5125	170,83
Beji	3,3381	-1,4403	3,6356	4117	137,23
Bugul Kidul	4,0538	-1,5740	4,3487	2155	71,83
Gadingrejo	4,2628	-1,8696	4,6548	1344	44,80
Gempol	-3,0893	8,0195	8,5940	30351	1011,70
Gondangwetan	1,9476	0,2241	1,9605	9181	306,03
Grati	-1,2630	-2,1687	2,5096	16474	549,13
Kejayan	-3,9423	4,3799	5,8928	30097	1003,23
Kraton	0,6381	2,1419	2,2349	14564	485,47
Lekok	0,9049	-0,8659	1,2524	11770	392,33
Lumbang	-18,3209	-4,0930	18,7726	60402	2013,40

Subdistrict	PC1	PC2	Distance	Total	Mean
Nguling	-1,4328	-3,0574	3,3765	16416	547,20
Pandaan	-0,1748	3,2464	3,2511	17910	597,00
Panggungrejo	4,2842	-1,9124	4,6916	1253	41,77
Pasrepan	-2,0117	1,1527	2,3186	21197	706,57
Pohjentrek	3,9062	-1,3889	4,1458	2674	89,13
Prigen	0,6115	0,6424	0,8869	12538	417,93
Purwodadi	-0,8652	1,4407	1,6806	17423	580,77
Purworejo	4,1982	-1,7792	4,5596	1518	50,60
Purwosari	0,6421	0,0401	0,6434	12257	408,57
Puspo	-0,6872	-2,3335	2,4326	14187	472,90
Rejoso	1,6555	0,3901	1,7008	10288	342,93
Rembang	0,3940	2,6106	2,6401	15936	531,20
Sukorejo	0,0231	2,4349	2,4350	16361	545,37
Tosari	-1,2563	-3,4626	3,6835	13856	461,87
Tutur	-0,2840	-0,7916	0,8410	12653	421,77
Winongan	-0,3827	0,0883	0,3927	15680	522,67
Wonorejo	-0,0288	1,6452	1,6455	16661	555,37
Total				13479,60	31016

b. Lightning Strike Prone Grouping

The prone areas lightning strikes are divided into three levels, namely low, medium, and high. Lightning strike susceptibility in an area is determined based on the lightning strike threat level interval (Susanto, 2018). The distance between the center of the ellipse and the data points has been distributed in Table 4. The results of determining the level of vulnerability can be seen in Table 5.

Table 5. Lightning Strike Vulnerability Per District

Vulnerability Level	Subdistrict	d
Low	Bangil, Beji, Bugul Kidul, Gadingrejo, Gondangwetan, Grati, Kejayan, Kraton, Lekok, Nguling, Pandaan, Panggungrejo, Pasrepan, Pohjentrek, Prigen, Purwodadi, Purworejo, Purwosari, Puspo, Rejoso, Rembang, Sukorejo, Tosari, Speech, Winongan, Wonorejo	$d \leq 6,1266$
Currently	Gempol	$6,1266 < d \leq 12,2532$
Tall	Porcupine	$d > 12,2532$

The study reveals that the majority (26) of Pasuruan's sub-districts exhibited low vulnerability, while one displayed moderate vulnerability and another high vulnerability (see Table 5). Lumbang District, the maximum extreme point on the ellipse, is the only area with a high vulnerability to lightning strikes in Pasuruan. Meanwhile, Panggungrejo District, the minimum extreme point on the ellipse, is one of the areas with a low vulnerability to lightning strikes. Lumbang District is the area with the largest area in Pasuruan Disrict. Most areas in Lumbang District are agricultural, animal husbandry, and mining areas. Furthermore, Gempol District is a densely populated residential area and one of the industrial areas in Pasuruan. Lightning strikes are dangerous in open places such as paddy fields and fields and are prone to striking tall buildings (Shindo, 2018). The research highlights a significant disparity in lightning strike vulnerability across Pasuruan's sub-districts, with most areas having low vulnerability but specific hotspots identified. Lumbang District, characterized by its large size, agricultural dominance, and extreme geographical position, emerges as a high-risk area. The contrast between Lumbang and Gempol districts suggests that population density and land use patterns are crucial factors in lightning strike vulnerability. In essence, the research highlights the need for targeted lightning protection measures in specific areas of Pasuruan, particularly Lumbang and Gempol, to reduce the risk of human casualties, property damage, and economic losses.

Harits et al. (2023) researched the threat, vulnerability, and susceptibility to lightning strikes in Mataram City and its surroundings. The level of vulnerability to lightning strikes is processed using simple additive weighting. Furthermore, Anggraini et al. (2024) did the same thing to find the distribution pattern of lightning strikes in Bandar Lampung and South Lampung. The novelty of ellipse fitting lies in its excellence at analyzing spatial patterns compared to simple additive weighting, which is better suited for ranking choices based on various factors. While simple additive weighting has limitations like subjectivity and difficulty handling complex relationships, ellipse fitting offers a more objective and spatially focused approach.

D. CONCLUSION AND SUGGESTION

Lightning strike-prone areas are grouped into high, medium, and low. Areas with a high level of vulnerability are Lumbang District, with a total of 60.402 strikes and the farthest distance from the predetermined elliptical boundary. Areas with a moderate level of vulnerability are Gempol and Kejayan Districts. Meanwhile, 25 other areas are prone to low-level lightning strikes.

Regions with high, medium, and low levels of lightning vulnerability have different impacts. Lightning will be more dangerous in areas with medium and high vulnerability levels and can even cause death. Nevertheless, the positive impact is that the air is fresher after rain because lightning can kill germs and bacteria in clouds and make the soil more fertile. Energy from lightning can also be developed into Lightning Power Generation and used as alternative electrical energy. In contrast, areas with a low level of vulnerability will have a low level of danger. However, developing lightning energy into alternative electrical energy takes more works. Suggestion for the further research is to overcome extreme points or extreme points in the PCA confidence ellipse such as MVEE.

ACKNOWLEDGEMENT

The title for the thank you to the institution or the person who has contributed during the research and references is not numbered.The Acknowledgments section is optional. Research sources can be included in this section.

DECLARATIONS

AUTHOR CONTIBUTION

Conceptualization, H. A. L; and R. D. L. N. K; methodology, R. D. L. N. K, H. A. L, R. R, and D. C. K; software: H. A. L; validation, R. D. L. N. K, D. C. K, and R. R; formal analysis, H. A. L; and R. D. L. N. K, investigation, R. R; and D. C. K, resources, R. R; data curation, D. C. K, R. R; and H. A. L; writing H. A. L; and R. D. L. N. K; writing-review and editing, R. D. L. N. K; visualization, H. A. L; project administration, H. A. L; funding acquisition R. D. L. N. K. All authors have read and agreed to the published version of the manuscript.

FUNDING STATEMENT

This research was funded by Ministry of Religion, Indonesia and LP2M UIN Maulana Malik Ibrahim Malang.

COMPETING INTEREST

The authors declare no conflict of interest.

REFERENCES

Anggraini, W., Dwi Gustia Rasyidi, R., Dian Yusandika, A., Sodikin, Rawuni, L., & Amalia, P. (2024). CG (*Cloud-To-Ground*) Lightning Strike Distribution Patterns In Bandar Lampung and South Lampung in 2022 (A. Saregar, R. Umam, & A. Rahmat, Eds.). *E3S Web of Conferences*, *482*, 03001. https://doi.org/10.1051/e3sconf/202448203001

Badan Pusat Statistik. (2022). *Direktori Perusahaan Industri Besar dan Sedang Provinsi Jawa Timur 2022*.

Badan Pusat Statistik Kota Pasuruan. (2020, January). Jumlah Penduduk Kota Pasuruan Hasil Sensus Penduduk 2020 (SP2020). https://pasuruankota.bps.go.id

Davis, C., Engeln, A., Johnson, E. L., McIntosh, S. E., Zafren, K., Islas, A. A., McStay, C., Smith, W. R., & Cushing, T. (2014). Wilderness Medical Society Practice Guidelines for the Prevention and Treatment of Lightning Injuries: 2014 Update. *Wilderness & Environmental Medicine*, *25*(4_suppl), S86–S95. https://doi.org/10.1016/j.wem.2014.08.011

Elhaik, E. (2022). Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports*, *12*(1), 14683. https://doi.org/10.1038/s41598-022-14395-4

Faizatin, T. U. (2014). Pemetaan Daerah Rawan Petir Cloud to Ground Positif Wilayah Pasuruan Tahun 2012 Menggunakan Metode Inverse Distance Weighted. *Inovasi Fisika Indonesia*, *3*(3), 6–10. https://doi.org/10.26740/ifi.v3n3.p%p

Faunia. (2022). Nomor 1 Bukan Gresik Atau Surabaya, Ini 5 Kabupaten Kota dengan Kawasan Industri Terluas di Jawa Timur – The Industrial News. https://industrialnews.id/nomor-1-bukan-gresik-atau-surabaya-ini-5-kabupaten-kota-dengan-kawasan-industri-terluas-di-jawa-timur/

Forkman, J., Josse, J., & Piepho, H.-P. (2019). Hypothesis Tests for Principal Component Analysis When Variables are Standardized. *Journal of Agricultural, Biological and Environmental Statistics*, *24*(2), 289–308. https://doi.org/10.1007/s13253-019-00355-5

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, *2*(1), 100. https://doi.org/10.1038/s43586-022-00184-w

Gunawan, T., & Pandiangan, L. N. L. (2014). Analisis Tingkat Kerawanan Bahaya Sambaran Petir dengan Metode Simple Additive Weighting di Provinsi Bali. *Jurnal Meteorologi dan Geofisika*, *15*(3). https://doi.org/10.31172/jmg.v15i3.221

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis*. Cengage.

Harits, M. A., Suardi, I., Marsono, A., Ali, M., & Prayitno Adi, S. (2023). Analysis of lightning vulnerability level in the City of Mataram and surrounding area using the simple additive weighting method (A. Hakam, Fauzan, S. Subbarayan, & A. Safuan A. Rashid, Eds.). *E3S Web of Conferences*, *464*, 01014. https://doi.org/10.1051/e3sconf/202346401014

Johnson, R. A., & Wichern, D. W. (2015). *Applied Multivariate Statistical Analysis* (6th Edition). Pearson Prentice Hall.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kherif, F., & Latypova, A. (2020). Principal component analysis. In *Machine Learning* (pp. 209–225). Elsevier. https://doi.org/10.1016/B978-0-12-815739-8.00012-2

Mashadi. (2018). *Geometri Edisi Kedua*. https://mashadi.staff.unri.ac.id/files/2018/10/BUKU-GEOMETRI-EDISI-2.pdf

Mora García, M., Riesco Martín, J., Rivas Soriano, L., & De Pablo Dávila, F. (2015). Observed impact of land uses and soil types on cloud-to-ground lightning in Castilla-Leon (Spain). *Atmospheric Research*, *166*, 233–238. https://doi.org/10.1016/j.atmosres.2015.07.009

Narut, F., Wahid, A., & Sumawan, S. (2018). Karakterisasi Peristiwa Petir di Wilayah Kota Kupang Serta Keterkaitannya dengan Curah Hujan. *Jurnal Fisika : Fisika Sains dan Aplikasinya*, *3*(2), 110–116. https://doi.org/10.35508/fisa.v3i2.611

Nurunnabi, A., Sadahiro, Y., Lindenbergh, R., & Belton, D. (2019). Robust cylinder fitting in laser scanning point cloud data. *Measurement*, *138*, 632–651. https://doi.org/10.1016/j.measurement.2019.01.095

Peterson, M. J., Lang, T. J., Logan, T., Wee Kiong, C., Gijben, M., Holle, R., Kolmasova, I., Marisaldi, M., Montanya, J., Pawar, S. D., Zhang, D., Brunet, M., & Cerveny, R. S. (2022). New WMO Certified Megaflash Lightning Extremes for Flash Distance and Duration Recorded from Space. *Bulletin of the American Meteorological Society*, *103*(4), 257–261. https://doi.org/10.1175/BAMS-D-21-0254.1

Reiml, S., Toth, D., Panayiotou, M., Fahn, B., Karim, R., Behar, J. M., Rinaldi, C. A., Razavi, R., Rhode, K. S., Brost, A., & Mountney, P. (2016, March). Interactive visualization for scar transmurality in cardiac resynchronization therapy. In R. J. Webster & Z. R. Yaniv (Eds.). https://doi.org/10.1117/12.2214737

Sanderson, D. (2019). Modeling the Distribution of Lightning Strike Distances Outside a Preexisting Lightning Area. *Theses and Dissertations*. https://scholar.afit.edu/etd/2194

Shindo, T. (2018). Lightning striking characteristics to tall structures. *IEEJ Transactions on Electrical and Electronic Engineering*, *13*(7), 938–947. https://doi.org/10.1002/tee.22649

Sterpka, C., Dwyer, J., Liu, N., Hare, B. M., Scholten, O., Buitink, S., Veen, S. T., & Nelles, A. (2021). The Spontaneous Nature of Lightning Initiation Revealed. *Geophysical Research Letters*, *48*(23), e2021GL095511. https://doi.org/10.1029/2021GL095511

Sulistiyono, W., Ramadhan, R. S., & Donni, Y. (2023). Kajian Kondisi Atmosfer Saat Kejadian Hujan Lebat di Kota Surakarta Menggunakan Analisis Skala Meteorologi. *OPTIKA: Jurnal Pendidikan Fisika*, *7*(1), 32–45. https://doi.org/10.37478/optika.v7i1.2626

Sun, S., & Wang, H. (2018). Principal component analysis-based features generation combined with ellipse models-based classification criterion for a ventricular septal defect diagnosis system. *Australasian Physical & Engineering Sciences in Medicine*, *41*(4), 821–836. https://doi.org/10.1007/s13246-018-0676-1

Susanto, E. (2018). Penentuan Daerah Rawan Bencana Sambaran Petir di Wilayah Kabupaten dan Kota Bandung Jawa Barat. *Jurnal Ilmu dan Inovasi Fisika*, *2*(2), 137–144. https://doi.org/10.24198/jiif.v2i2.19728

Syarief, I. S. (2023, January). Paling Banyak Se-Jatim, 700 Ribu Sambaran Petir Terjadi di Pasuruan Selama 2022. https://www.suarasurabaya.net/Kelanakota/2023/Paling-Banyak-Se-Jatim-700-Ribu-Sambaran-Petir-Terjadi-Di-Pasuruan-Selama-2022/

Thurnhofer-Hemsi, K., López-Rubio, E., Blázquez-Parra, E. B., Ladrón-de-Guevara-Muñoz, M. C., & de-Cózar-Macías, Ó. D. (2021). Ensemble ellipse fitting by spatial median consensus. *Information Sciences*, *579*, 310–324. https://doi.org/10.1016/j.ins.2021.08.011

Wahjudi, D. (2014). Mengatasi Bahaya Petir dan Proteksi Petir Gedung Bertingkat. *Teodolita: Media Komunkasi Ilmiah di Bidang Teknik*, *15*(2), 57–70. https://doi.org/10.53810/jt.v15i2.140

Zhao, C., Fu, C., Dolan, J., & Wang, J. (2021). L-Shape Fitting-Based Vehicle Pose Estimation and Tracking Using 3D-LiDAR. *IEEE Transactions on Intelligent Vehicles*, *6*(4), 787–798. https://doi.org/10.1109/TIV.2021.3078619

[This page intentionally left blank.]