IJIE

UTHM Penerbit

INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

Vol. 18 No. 1 (2026) 254-272
https://publisher.uthm.edu.my/ojs/index.php/ijie

GIS-Based Investigation of Lightning Activity with Its Meteorological and Topographical Correlates on a 500 kV Transmission Line in Peninsular Malaysia

N. Roslan[1]*, U. A. U. Amirulddin[1], M. Z. A. Ab-Kadir[2], N. Abdullah[3]

[1] *Institute of Power Engineering (IPE), Universiti Tenaga Nasional, 43000, Kajang, MALAYSIA*
[2] *Advanced Lightning, Power and Energy Research Centre (ALPER), Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, MALAYSIA*
[3] *Lightning & Earthing, TNB Research Sdn Bhd, 43000, Selangor, MALAYSIA*

*Corresponding Author: nurzanariah@uniten.edu.my
DOI: https://doi.org/10.30880/ijie.2026.18.01.019

Article Info

Received: 1 January 2026
Accepted: 30 March 2026
Available online: 20 April 2026

Keywords

Correlation, GIS, lightning, meteorological variables, PCA, topography, transmission line

Abstract

Climate change has caused a sudden surge in meteorological hazards in recent years, such as lightning strikes that have been a considerable challenge to transmission lines, causing flashovers, equipment damage, and outages. This study proposes a new method utilized Geographic Information System (GIS) integrated with Principal Component Analysis (PCA) to investigate the correlation of lightning, meteorology, and topography on the 500 kV transmission line A-B in Peninsular Malaysia from 2015 to 2017. This study utilized data on lightning, weather, topography and transmission lines from various sources. PCA was used to perform the analysis of the relation between lightning activity, elevation, and meteorological variables including temperature, rainfall, wind speed, and pressure. The study results show that lightning Ground Flash Density (GFD) exhibits positive correlations with elevation (r= 0.16) and rainfall (r= 0.18) and negative correlation with temperature (r= -0.13), wind speed (r= -0.34) and pressure (r= -0.49). PCA further reveals that lightning GFD is most strongly associated with elevation while showing not largely correlated with warm, windy and elevated pressure conditions along the selected transmission line. These findings suggest that lightning activity along transmission line A-B is more strongly influenced by elevation than other meteorological variables. Identifying the factors that govern lightning occurrences is crucial in characterizing lightning distribution across Malaysia's complex terrain. Understanding these correlations helps researchers and engineers better identify high-risk zones, forecast lightning-prone areas, and integrate these insights into spatial planning and grid design. This approach can be applied to other transmission lines, potentially improving the reliability of power grids through effective protection systems.

1. Introduction

Rapid and intensifying climate change has led to a notable increase in the frequency and severity of extreme weather events worldwide. Among these, lightning occurrences have become more prevalent and unpredictable, posing significant challenges to infrastructure and human safety. Malaysia, renowned for its tropical climate, complex terrain, and unique geographical location, experiences pronounced effects of these climatic transformations, especially in the form of heightened lightning activity [1]-[4]. Due to its position in a region highly susceptible to convective thunderstorms and seasonal monsoons, lightning strikes in Malaysia would frequently disrupt power supply, especially on overhead transmission lines (132/275/500 kV system) [5]. Consequently, these incidents result in high overvoltage, electrical discharges, equipment damage, energy losses, and maintenance costs, particularly for 500 kV transmission lines [6].

Lightning occurrences are governed by a myriad of environmental variables, including topography, elevation, meteorological factors such as humidity, precipitation, convective potential energy, temperature, wind, and pressure. These variables would influence the distribution and intensity of lightning, shaping spatial and temporal lightning patterns across vast regions. Furthermore, the effect of these variables can have both direct and indirect impacts on critical infrastructures, specifically power lines and the electrical grid. Thus, it is crucial to understand the trends and relationships between lightning events and their causes to enhance grid resilience and protect electrical assets [7].

Numerous studies conducted over the past decades have aimed to correlate trends in lightning strikes with various meteorological and topographical factors [8]-[18]. These factors could possibly influence lightning distribution in terms of flash density, which also affects transmission line performance. However, a comprehensive understanding of the complex relationships between lightning occurrence, weather and topographic variables remains underdeveloped. In addition, the increasing interest in this research area, observed at both regional and global levels, highlights a growing recognition of the need for approaches. This encompasses climatological, geographical, and engineering perspectives to better comprehend and manage lightning-related risks.

Therefore, this study proposes a novel GIS-based approach that integrates Principal Component Analysis (PCA) techniques to investigate the correlation between lightning activity, weather events, and elevation along the selected 500 kV transmission line. To date, this is the first study to explore and address the relationship between lightning, weather variables, and elevation at a selected 500 kV transmission line in Peninsular Malaysia using an integrated GIS-PCA framework. The findings of this study can help enhance the resilience, reliability, and sustainability of energy infrastructure in lightning prone regions, such as Malaysia. Assessing how meteorological conditions and topographical features can influence lightning strikes is essential to improve transmission line durability. Due to the vulnerability of high-voltage transmission lines towards atmospheric discharges and failures, widespread outages, costly equipment damage, and compromised safety can occur at any time.

These advancements can directly contribute to Sustainable Development Goals (SDG) 7 (Target 7.1), and reliable and secure energy services are a top priority [19], [20]. In addition, increased in-line loss and emergency procedures may weaken grid performance, resulting in lightning-related outages. This paper will apply Principal Component Analysis (PCA) to reveal the associations between lightning and its correlations. The paper will also utilize Geographic Information System (GIS) software, which makes it possible to visualize both spatial and temporal trends, which helps to implement proactive interventions that prevent server downtime and energy wastage. In that sense, this act enhances the fight under Target 7.3 of SDG 7, which focuses on doubling the global measure of energy efficiency improvement by 2030 [19], [21].

2. Review of Related Studies

In putting in a global context and regional literature, there is various research that provides essential insight into the interaction between lightning events, topographic, as well as meteorological factors. As an example, Zhang et al. [13] found strong positive relationships between lightning and meteorological factors (humidity, temperature, wind components) to enhance the existing active lightning warning system within the province of Fujian, China. The research relied on the use of ERA5 atmospheric reanalysis data and Lightning Location System (LLS) data, together with spatial interpolation through the Kriging method. Based on the methods applied, inconsistent parameters were ruled out, areas where the parameters have a larger possibility of lightning were determined, and the high-risk lightning locations confirmed. Subsequently, Shankar et al. [14] distinguished the locations that are prone to lightning in the state of Jharkhand, India, by estimating the covariance between the activity of lightning and the topography and the monsoon conditions in the state. It was found that there is a positive correlation between the lightning flash density and elevation.

Furthermore, Kumar et al. [15] analyzed the lightning distribution in the western region using Tropical Rainfall Measuring Mission - Lightning Imaging Sensor (TRMM-LIS) datasets and PCA. They found a strong correlation between lightning flash rate density and surface thermodynamic parameters such as sensible heat flux, Bowen ratio, and convective rainfall. They identified Convective Available Potential Energy (CAPE), humidity, and surface temperature as dominant meteorological drivers, offering a foundation for integrating the PCA technique into regional lightning modeling. In addition, Hazmi et al. [16] explored the correlation between lightning occurrence and monsoon weather parameters in Padang, Indonesia. The results indicated that the strongest positive correlation exists between lightning and rainfall. Meanwhile, moderate and weak positive correlations were observed for humidity and wind speed, as well as air temperature. The study emphasized that rainfall acts as a proxy for convective intensity and is thus the most influential factor for the occurrence of lightning during the monsoon season in Padang.

In parallel, a growing body of research has emerged focusing on the Malaysian region, where numerous studies have been conducted to examine the intricate connections between lightning activity and its factors, such as environmental variables. Two similar studies, by Yusop et al. [17] and Bahari et al. [18], investigated the correlation between lightning and rainfall rates in the southeast of Peninsular Malaysia. According to both studies, rainfall rate demonstrated a strong positive correlation with lightning. Building on this, radar reflectivity and electric field measurements were used as a part of the process to obtain and validate the outcome in [17], while the Pearson correlation coefficient method was employed in [18]. To highlight the research gap addressed by this study, Table 1 summarizes a comparative overview of previous research on lightning and environmental variable correlation including their analyzed parameter, their methodological approaches, advantages and limitations.

Table 1 *Comparative overview of previous studies on lightning activity and environmental variable correlation*

Authors [Ref]	Study Area	Parameters Analyzed	Correlation Method Used	Advantages	Limitations
Yusop et al. [17]	Malacca, Malaysia	Lightning and rainfall rate.	No specific statistical method is mentioned.	Regional lightning pattern analysis	No consideration of topography or wind and no transmission line focus.
Tinmaker et al. [11]	India	Lightning activity, maximum air. temperature, Bowen ratio and cloud ice content.	Pearson correlation coefficient	Used high-resolution data	No transmission line focus.
Hazmi et al. [16]	Padang, Indonesia	Lightning, rainfall, humidity, air. Temperature and wind speed.	Pearson correlation coefficient	First detailed study in Indonesia using Pearson.	No transmission line focus.
Bahari et al. [18]	Universiti Teknologi Johor, Malaysia	Lightning, rainfall and flash flood.	Pearson correlation coefficient	Integrated flooding with lightning data.	Focused on flash flood events, not power lines.
Kumar et al. [15]	Uttarakhand, India	Lightning, CAPE, temperature, humidity, rain rate and cloud base height.	PCA	Identified multiple environmental variables.	Limited to the Indian region; no transmission line focus.
Shankar et al. [14]	Jharkhand, India	Lightning	No statistical correlation was conducted.	Mapping and statistical analysis using GIS.	No correlation analysis of lightning and environmental variables.

Although the connection between lightning activity and different meteorological and topographical variables has been explored by a variety of studies in China, India, and Indonesia, research in Peninsular Malaysia lacks both extensiveness and depth in methodology. Interestingly, most authors in the existing Malaysian literature have focused on the trend between lightning phenomena and environmental factors, especially rainfall. Yet, other factors that can also contribute to lightning occurrence, such as elevation and atmospheric pressure, are usually left out in these studies. In addition, those multivariate methods, which have been very successful in unravelling complex interactions in parameter combinations in other situations, have not been used extensively with lightning studied in Malaysia. Moreover, GIS-based mapping of lightning hotspots based on combining meteorological and elevation data is non-existent in Malaysian studies. Consequently, this paper contributes an innovative methodology that integrates a Geographical Information System based framework and Principal Components Analysis in investigating the spatial-temporal distribution of lightning occurrence within a 500 kV transmission line of the Peninsular of Malaysia and offers a complete system to improve the predictive accuracy and enable regional lightning hazard mitigation practices.

3. Materials and Methods

3.1 Study Area and Data Sources

This study was conducted in Peninsular Malaysia, which is located in Southeast Asia between the latitudes of 1° to 7° North and longitudes of 99° to 105° East. To facilitate the analysis, a specific transmission line, namely the 500 kV transmission line A-B located in Negeri Sembilan was selected due to its reported number of outages. The study area illustrates in Fig. 1 encompasses a 5 km buffer zone around the 124 transmission towers along with the locations of the weather stations and an elevation map within the buffer zone.

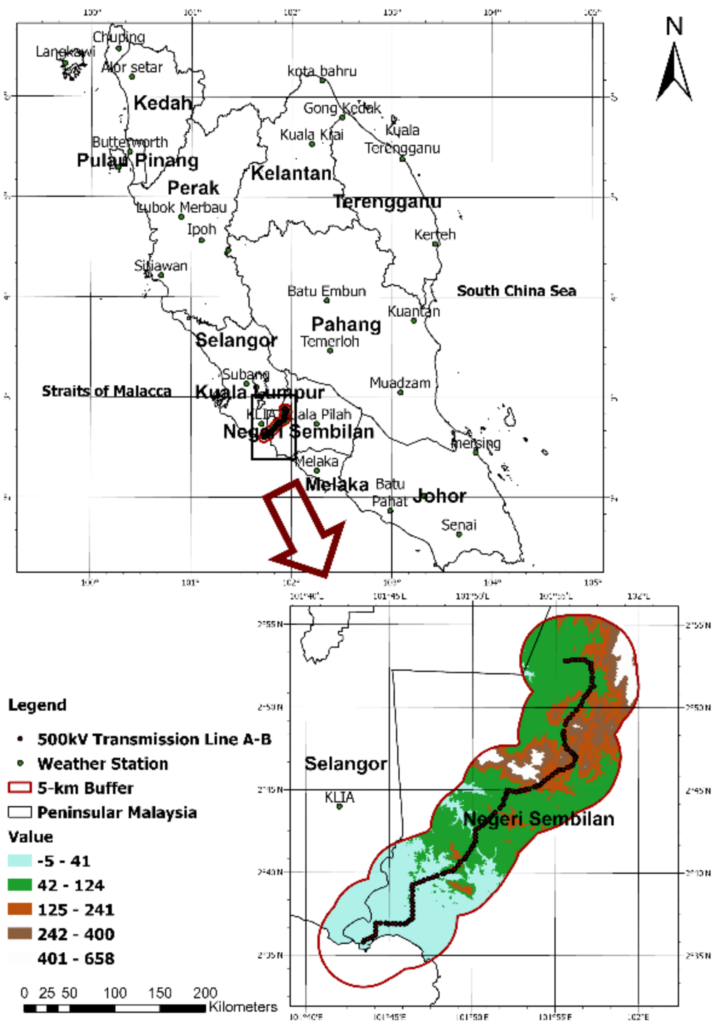

Fig. 1 *Map of the study area*

The data analyzed in this study are detailed in Table 2. The research is based on the lightning flash data collected from January 2015 to December 2017 by the LLS, also known as the Lightning Detection Network (LDN). The lightning data available up to the year 2017, as data from 2018 onwards were inaccessible for this study, as provided by the LLS [4]. The LLS employs a combination of Magnetic Direction Findings (MDF) and Time of Arrival (TOA) techniques, which have a detection efficiency of 95% and location accuracy of 500 m [22], [23]. In particular, the MDF technique uses orthogonal loop antennas to detect the direction of the magnetic field generated by a lightning discharge. For azimuth detection in MDF, each sensor determines the angle between true north and the direction of the lightning strike. This allows for the location of the strike to be estimated. Meanwhile, the function of TOA is to measure the precise time it takes for a lightning signal to reach each sensor. The time differences between the sensors define the hyperbolas intersection, which locates the strike location. Thus, using a combination of these two techniques enables the reception of redundant and complementary lightning flash data [24]-[26].

For this study, the lightning flash data from the three years were separated into two datasets: negative flashes and positive flashes. However, the analysis excluded all lightning flashes with positive polarity due to the greater occurrence of negative polarity lightning. This decision was made to ensure that the statistical reliability of the analysis remains high, as negative polarity flashes represent the dominant form of cloud-to-ground lightning, as it is more hazardous than positive polarity lightning in tropical regions such as Malaysia [6]. Therefore, by focusing on the more frequent negative flashes, the study enhances its ability to derive meaningful insights into lightning behavior and its implications.

In addition, meteorological variables such as rainfall, mean temperature, mean wind speed, and pressure were retrieved from all 27 weather stations located throughout Peninsular Malaysia for the period spanning from 2015 to 2017. This corresponds with the timeframe of the lightning flash data acquisition. A 20 m resolution digital elevation model of Peninsular Malaysia was also employed. Additionally, the transmission line data detailing the tower locations was also extracted.

Table 2 *Source and data types*

Data Type	Source
Lightning data	TNB Research Sdn. Bhd. (TNBR)
Digital topographic map and Digital Elevation Model of Peninsular Malaysia	Department of Survey and Mapping Malaysia (JUPEM)
Meteorological data	Malaysia Meteorological Department (MMD)
Transmission line data	TNB Grid Maintenance

3.2 Methodology

This study presents the first application of GIS integrated PCA frameworks, which is a novel approach to assess the relationships between lightning activity, meteorological variables as well as topography directly along the 500 kV transmission line in Malaysia. While previous research has focused on regional climatological patterns, our method enables spatially targeted identification of high-risk zones, supporting more accurate prediction of lightning-prone areas and facilitating the integration of these findings into transmission line planning and grid resilience design.

All available meteorological datasets were collected and utilized in this study, encompassing lightning flash, rainfall, mean temperature, pressure, and mean wind speed. These datasets were sourced from a range of reputable providers. Following the data collection, all input layers were preprocessed and integrated into ArcGIS Pro version 3.2.2 software. This platform served as the primary database for managing, processing and visualizing all spatial data. ArcGIS facilitated the preparation of spatially aligned and extracted data within a defined area for use in correlation analysis between lightning activity, weather events, and topography. The implementation procedure of the overall methodological framework adopted in this paper is visually summarized in Fig. 2, outlining the steps involved in data acquisition, transformation, integration, spatial extraction in ArcGIS and correlation analysis in R Studio.

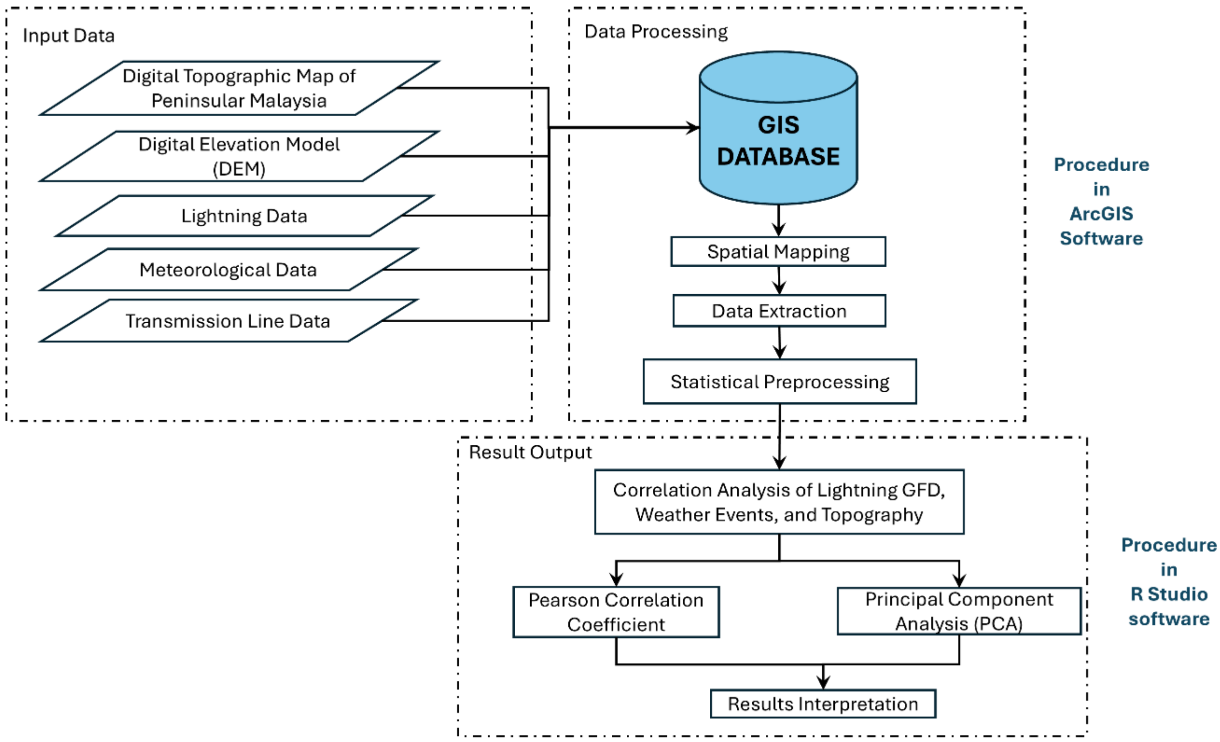

Fig. 2 *Procedure for assessing the relationship between lightning, weather events, and topography*

To assess the lightning performance of transmission lines in Peninsular Malaysia, the negative lightning Ground Flash Density (GFD) map was adapted from [4], where it was originally generated using the Point Density to calculate the lightning flash density per square kilometer across Peninsular Malaysia. The map represents the number of lightning flashes per 1 km^2 grid cells specifically by counting the number of lightning flash events occurring within each cell. This approach provides a localized representation of lightning exposure along the transmission corridors. Concurrently, weather maps including rainfall, temperature, wind speed and pressure were generated using the Inverse Distance Weighting (IDW) interpolation technique, one of the GIS analysis techniques to assess meteorological data over Peninsular Malaysia based on the proximity and influence of surrounding data points. Additionally, the outcome of this integrated mapping process includes a negative lightning GFD map, weather maps, and an elevation map, all based on a 1 km^2 grid cell size to ensure consistency in spatial analyses. The resulting raster maps including negative lightning GFD, weather and elevation were stored in the GIS database and overlaid with the selected 500 kV Transmission Line A-B, as illustrated in Fig. 3.

To further explore the correlation between lightning, meteorological variables, and elevation, data for lightning GFD, weather variables (rainfall, temperature, wind speed and pressure), and elevation were extracted within a 5 km buffer radius around the centerline of the 500 kV transmission line A-B using the Extract by Mask tool. A 1 km^2 fishnet grid was then overlaid and used to extract the data. The resulting dataset was exported as a CSV file and imported into R Studio for the correlation study. R Studio is a powerful tool for examining multivariate data. In this study, a Pearson correlation coefficient was first calculated to determine pairwise linear relationships between the negative lightning GFD and each environmental variable. Then, the PCA method was applied for further analysis to investigate the relationship and identify the dominant variables influencing the lightning distribution. The mathematical equations for both methods are presented in Sections 3.3 and 3.4, respectively. As shown in Fig. 3, this workflow links spatial data preparation in ArcGIS with correlation analysis in R Studio providing a geographically targeted insight into the resilience of transmission lines.

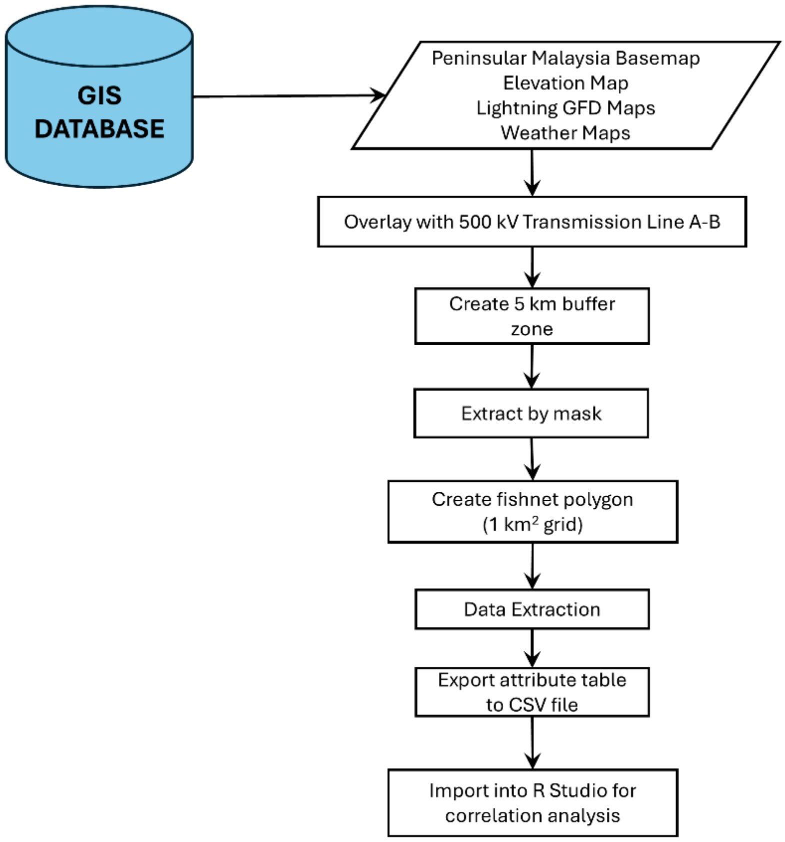

Fig. 3 *Workflow of data processing in ArcGIS software*

3.3 Correlation

A correlation matrix based on the Pearson correlation coefficient method is used to analyze the relationships and interdependencies among pairs of variables, facilitating an understanding of their interrelations. This statistical tool provides valuable insight into how one variable behaves in relation to another, particularly in complex environmental systems where multiple factors contribute to a given phenomenon [27]. Thus, Eq. (1) used to calculate the Pearson correlation coefficient, denoted as r, is expressed as follows [28]:

$$r = \frac{\sum(X-\bar{X})\sum(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2}\,.\sqrt{\sum(Y-\bar{Y})^2}} \tag{1}$$

where X and Y are the variables, $\bar{X}$ and $\bar{Y}$ are the means of X and Y variables.

This study focuses on the linear correlation between negative lightning GFD and five environmental variables, including elevation, rainfall, mean temperature, mean wind speed, and pressure. In Eq. (1), the variable X represents the negative lightning GFD, and variable Y corresponds to one environmental variable at a time. The resulting r values range from -1 to +1, which indicates the strength and direction of the linear relationships. Prior to dimensionality reduction via PCA, this matrix provides insight into which variables exhibit significant correlation with negative lightning GFD. Therefore, it helps in indicating the selection of components and improving data interpretability.

Variables such as elevation, rainfall, mean temperature, mean wind speed, and pressure align with existing literature on convective storm dynamics and terrain-induced lightning distribution, particularly in tropical regions like Malaysia. In brief, the Pearson correlation method offers a straightforward and statistically sound way to quantify how lightning activity is related to topographical and meteorological factors such as elevation, rainfall, mean temperature, mean wind speed, and pressure [27], [29]. By calculating the degree of linear association between these variables, the environmental conditions with the most influence on lightning occurrence can be identified. This method is especially useful in preprocessing steps for multivariate analyses, such as PCA, where understanding variable interdependence helps reduce dimensionality and uncover latent patterns in lightning distribution across complex terrains [13].

3.4 Principal Component Analysis (PCA)

PCA is a powerful dimensionality reduction technique utilized for analyzing multivariate datasets, which are used in this study, that exhibit complex interdependencies among their variables [30]. The primary purpose of PCA is to investigate the relationship among all six variables by reducing the number of variables [30], [31]. This reduction is achieved by transforming and representing the data into a new set of orthogonal variables known as Principal Components (PCs). Moreover, this reduction is capable of improving interpretability and computational efficiency [30]. These PCs are derived as linear combinations of the original variables as defined in Eq. (2)[32]:

$$PC_i = I_{1i}X_1 + I_{2i}X_2 + \cdots . + I_{ni}X_n \tag{2}$$

where PC_i represents the i^{th} PC and I_{ji} is the factor loading of the observed variable X_j.

When employing PCA for data exploration, two parameters are explained variance and factor loadings. Explained variance is based on eigenvalues of the covariance matrix. It reveals how much of the total variability in the data is captured by each PC. This is achieved by decomposing the covariance matrix into its eigenvalues and eigenvectors, where each eigenvalue corresponds to the variance explained by a specific PC. The proportion of explained variance is commonly used to determine the optimal number of components to retain, ensuring that the reduced dataset maintains its informational integrity [33], [34].

On the other hand, factor loadings indicate the strength and direction of correlation between each variable and the PCs. The factor loadings are derived from the eigenvectors and serve as coefficients in the linear combinations that define each PC. Note that high absolute values of loadings suggest strong associations between variables and components, while the sign of the loading reflects the direction of the relationship [33]. Accordingly, it involves computing the covariance of datasets using Eq. (3) and determining eigenvalues by solving the covariance matrix as per Eq. (4) [8].

$$cov(x,y) = \frac{\sum_{i=0}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)} \tag{3}$$

$$cov\,[dim_x \quad dim_y] - \begin{bmatrix} i & 0 \\ 0 & i \end{bmatrix} \tag{4}$$

where n denotes the number of samples, $\bar{x}$ = mean value of x variables, $\bar{y}$ = mean value of dimension y, x_i and y_i represent the values of variables at point i.

To recapitulate, PCA facilitates dimensionality reduction, allowing high-dimensional datasets to be condensed into a smaller set of uncorrelated PCs while retaining the majority of the original variance [34]. Additionally, it effectively mitigates multicollinearity by transforming correlated variables such as elevation, rainfall, and pressure into orthogonal components. Thus, improving the stability and interpretability of statistical analysis of the datasets [35]. Furthermore, the PCA has the ability to highlight the dominant patterns within the complex datasets. It can identify the most influential combination of meteorological and topographical factors that drive lightning occurrence at the transmission line [36]. This is considered useful for tropical regions such as Malaysia, where terrain-induced convection interacts in nonlinear ways.

4. Results and Discussion

4.1 Lightning Occurrence, Meteorological Variables and Elevation along the 500 kV Transmission Line A-B

To gain a deeper insight into the spatial and temporal patterns of lightning activity and its polarity characteristics in proximity to high-voltage infrastructure, cloud-to-ground lightning data from three years were systematically examined across the study area. Hence, an overview of lightning flashes recorded from 2015 until 2017 is displayed in Fig. 4. Based on the figure, both negative and positive lightning flashes exhibit upward trends, with 2017 recording the highest number of lightning flashes recorded annually. Specifically, the most striking jump is for positive lightning flashes, which were from 437 in 2016 to 1,441 in 2017. This drastic increase could be due to the intensified storm dynamics or shifts in convection systems. Correspondingly, stronger thunderstorms with elevated cloud tops produce more positive flashes, which are often linked to high CAPE values and deep convective systems [37], [38].

However, the predominance of negative lightning flashes over positive lightning flashes is the main highlight along the 500 kV transmission line A-B, as negative lightning flashes significantly outnumber positive ones each year. Since Malaysia's equatorial location fosters intense convective thunderstorms, these storms would develop strong mid-level negative charge centers that are ideal for initiating negative cloud-to-ground flashes [39]. Furthermore, the warm and moist air of the region encourages vertical cloud development. Consequently, it enhances the separation of charges with negative charges accumulating closer to the ground and increases the likelihood of negative stepped leaders forming [40], [41].

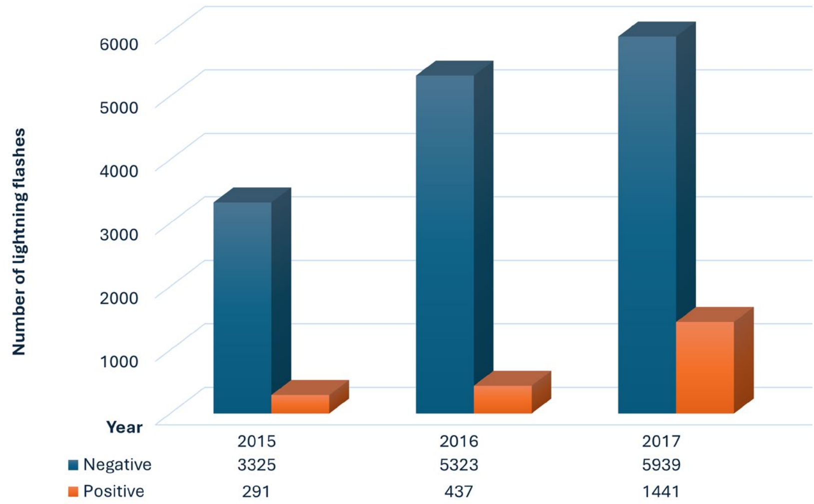

Fig. 4 *Annual distribution of lightning flashes based on polarity along the 500 kV line A-B (2015-2017)*

This study investigates negative lightning GFD in relation to weather variables and elevation between the years 2015 and 2017. To aid the investigation, spatial data were extracted from an ArcGIS-based map along the 500 kV transmission line A-B. Three lightning GFD maps along the 500 kV transmission line A-B from 2015 to 2017 are illustrated in Fig. 5. Meanwhile, Fig. 6 displays the weather maps for 2017, which include the rainfall, temperature, wind speed, and pressure maps. All of these maps were generated using the ArcGIS platform due to its ability to provide spatial precision, real-time data, and advanced visualization tools, which cater to the needs of this study.

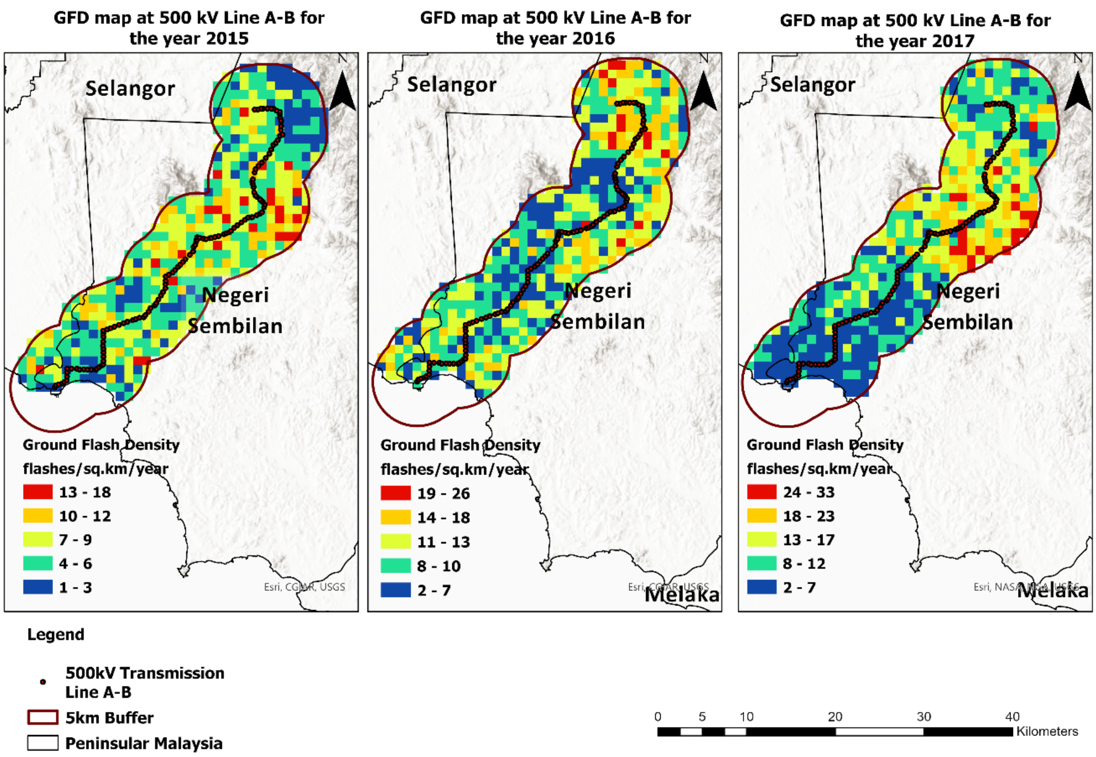

Fig. 5 *Lightning GFD map at the 500 kV Transmission Line A-B for the years 2015, 2016, and 2017*

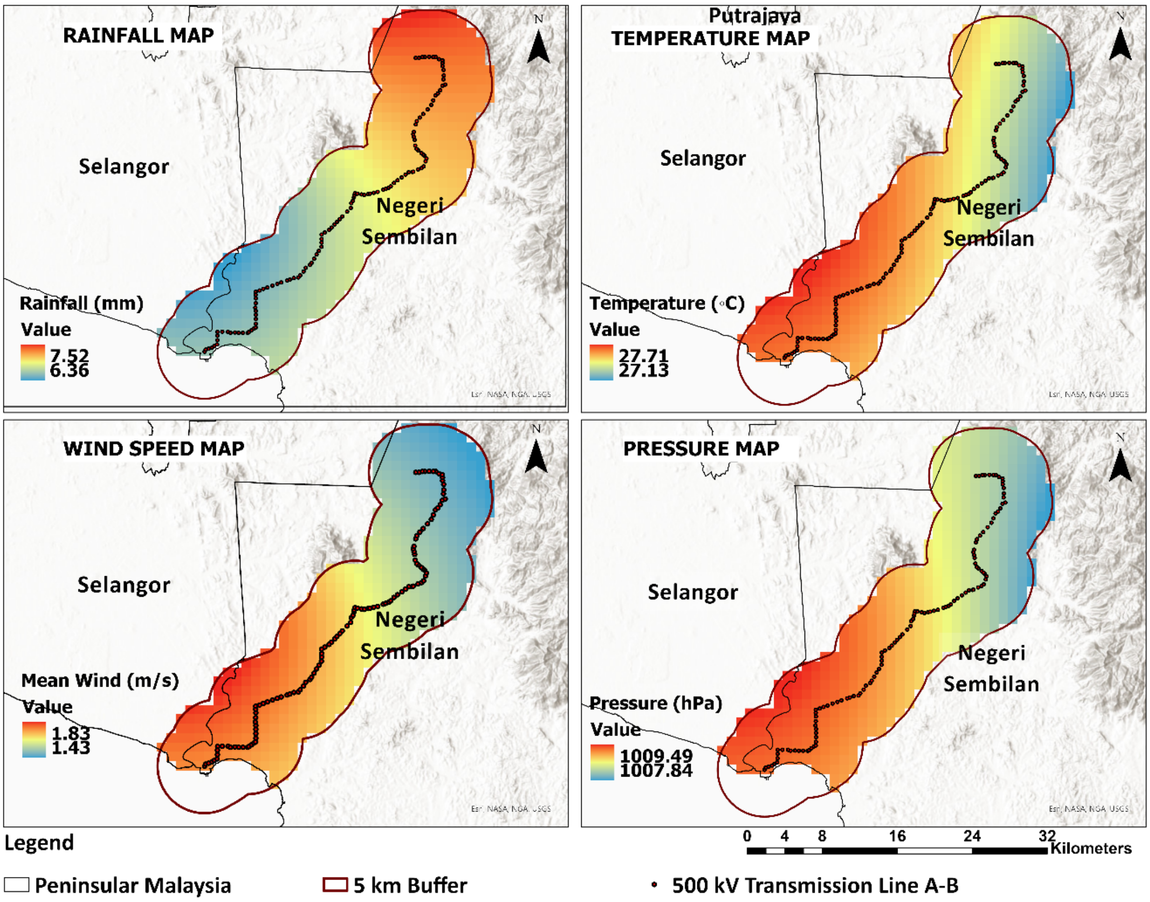

Fig. 6 *Weather maps at the 500 kV Transmission Line A-B for the year 2017*

Furthermore, the descriptive statistics of lightning GFD, weather parameters, and elevations along the 500 kV transmission line A-B from 2015 to 2017 are presented in Table 3. From the table, a comprehensive overview of the studied variables of interest can be determined. Across the three years, the negative lightning GFD value indicated a seemingly gradual increase in lightning occurrence, with a mean rising from 6.51 flashes/km^2/year in 2015 to 11.59 flashes/km^2/year. This result indicates that lightning activity has increased over time, which may potentially be influenced by the climatic shifts that occurred within those three years.

In terms of topographical data, particularly elevation, is not considered to change over time. Over the study period, the mean values for elevation were found to be consistent, ranging from 108 to 110 m. Nonetheless, the large standard deviation, which is approximately 120 m, reveals substantial altitude diversity along the studied transmission line, which can be a factor in localized lightning intensification through orographic lifting effects. On the other hand, the mean values for meteorological parameters such as rainfall, mean temperature, mean wind speed, and pressure demonstrated no drastic increase. According to the statistics, the year 2017 had the lowest mean values for mean temperature, mean wind speed, and pressure, while having the highest mean value for rainfall.

Table 3 *Descriptive statistics of lightning GFD, weather parameters, and elevation from 2015 to 2017 along 500 kV transmission line A-B*

Year	Variable	Unit	Min	Max	Mean	SD
2015	Negative GFD	flashes/km2/year	0	18	6.51	3.31
	Elevation	m	-1.44	590.2	108.57	120.49
	Rainfall	mm	5.7	6.92	6.14	0.26
	Mean Temperature	°C	27.3	27.84	27.61	0.14
	Mean Wind speed	m/s	1.51	2.02	1.75	0.15
	Pressure	hPa	1008.95	1010.33	1009.76	0.38
2016	Negative GFD	flashes/km2/year	1	26	10.34	3.97
	Elevation	m	-1.44	590.2	110.06	120.65
	Rainfall	mm	4.89	6.12	5.39	0.28
	Mean Temperature	°C	27.81	28.08	27.96	0.05
	Mean Wind speed	m/s	1.53	1.91	1.71	0.11
	Pressure	hPa	1008.17	1009.7	1009.05	0.42
2017	Negative GFD	flashes/km2/year	1	33	11.59	6.17
	Elevation	m	-1.44	590.2	108.79	120.51
	Rainfall	mm	6.36	7.52	6.93	0.32
	Mean Temperature	°C	27.13	27.71	27.47	0.15
	Mean Wind speed	m/s	1.43	1.83	1.62	0.12
	Pressure	hPa	1007.84	1009.49	1008.78	0.46

Note: explanation of abbreviations: Min- minimum, Max- maximum, SD -standard deviation.

4.2 Correlation of Negative Lightning GFD, Meteorological Variables and Elevation

The correlation matrix between variables was calculated and visualized as a correlation heat map, illustrated in Fig. 7. The colour intensity in the map represents the strength and direction of the correlation, with darker colours indicating stronger correlations, as outlined in Table 4.

To further validate the observed relationships, a comparative analysis was conducted across the three years dataset. The annual correlation analysis, presented in Fig. 7 (a)-(c), revealed distinct patterns across the study period. For 2016 and 2017, consistent correlation patterns were observed. Negative lightning GFD exhibited a positive correlation with both rainfall (r=0.24 and 0.36, respectively) and elevation (r=0.04 and 0.36, respectively). This indicates that areas with higher rainfall and greater elevation tend to experience more frequent negative lightning flashes. Conversely, negative lightning GFD showed negative correlations with mean temperature, mean wind speed, and pressure for the years 2016 and 2017, indicating that higher values of these meteorological variables are associated with less lightning occurrence.

However, in 2015, the correlation patterns deviated from those observed in 2016 and 2017. Specifically, the correlations were either absent or significantly weaker compared to 2016 and 2017. This inconsistency is possibly due to enhanced lightning data quality resulting from a significant upgrade to the LLS system after 2015 [4]. The consistency between the 2016 and 2017 results, therefore, supports the reliability of the findings.

To identify the most significant relationship across the entire study period and to establish consistency, the datasets from 2015, 2016 and 2017 were combined into a single dataset and analyzed together. Fig. 7(d) presents the correlation heat map for the combined 2015-2017 dataset illustrating the overall relationship between all variables. The results show that negative lightning GFD exhibits a positive correlation with both elevation (r=0.16) and rainfall (r=0.18) with the correlation slightly stronger for rainfall. This suggests that intense and frequent rainfall consistently accompanies convective thunderstorms, which are responsible for lightning, underscoring the role of moisture availability in fuelling electrically active storm systems. In contrast, it exhibits negative correlations with mean temperature (r=-0.13), mean wind speed (r=-0.34), and pressure (r=-0.49). Based on the heatmap presented, negative lightning GFD has the strongest negative correlation with pressure, while mean temperature has the weakest negative correlation. This suggests that pressure has a higher influence on negative lightning GFD than mean temperature and wind speed.

Table 4 *Indicator of the correlation [42]*

Pearson correlation coefficient (r)	Type of correlation	Interpretation
0 to 1	Positive	Both variables change in the same direction
0	None	No significant relationship between the variables
0 to -1	Negative	Variables change in opposite directions

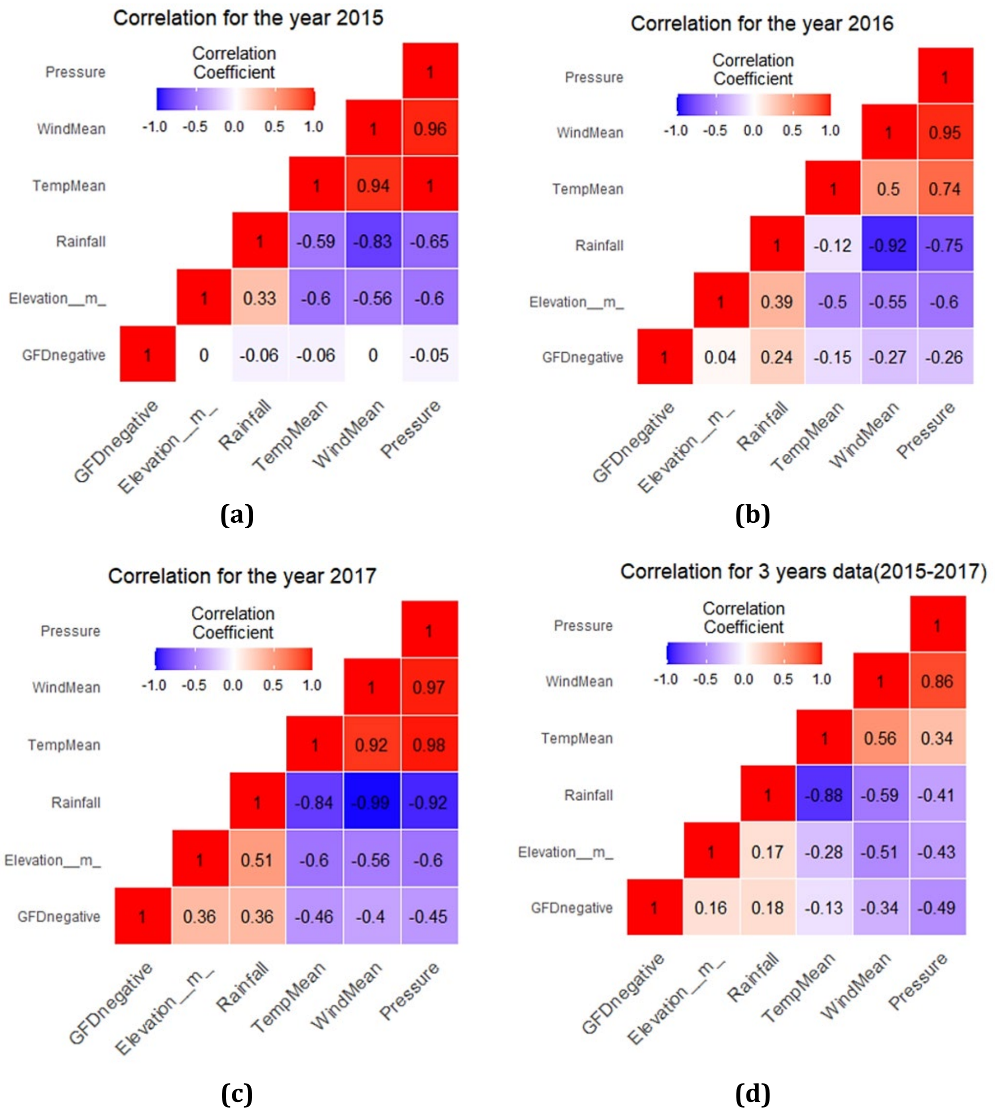

Fig. 7 *Correlation heat map between variables, (a) 2015; (b) 2016; (c) 2017; (d) Combined data 2015-2017*

The findings of this study have been compared with previous studies on the relationship between lightning and other meteorological factors in other regions to validate the results. Table 5 summarizes this comparison.

Table 5 *Summary of comparison data with previous studies on the relationship between lightning and meteorological factors*

Study	Location	Relationship finding
Hamzi et al. [16]	Padang, Indonesia	Strong positive correlation between lightning and rainfall only during the wet season. Weak and negative correlation between lightning and temperature and wind.
Bahari et al. [18]	Southeast Peninsular Malaysia	Lightning and rainfall exhibit a strong relationship during flash-flood events.
Y. Deng et al. [12]	Yunnan Province, China	Lightning is influenced by terrain features such as altitude and slope.
Tinmaker et al. [11]	Indian Region	Positive correlation between lightning and temperature during the dry season. Negative relationship during the wet season.

As shown in Table 5, the positive correlation between lightning activity and rainfall observed in this study is consistent with findings from Padang, Indonesia [16] and Southeast Peninsular Malaysia [18]. Similarly, the influence of elevation on lightning activity observed in this study aligns with findings from Yunnan Province, China [12], where terrain features such as altitude and slope significantly affect lightning occurrence.

Regarding temperature, the negative correlation observed in this study is consistent with Hamzi et al. [16] in Indonesia. However, Tinmaker et al. [11] observed a positive correlation during the dry season and a negative correlation during the wet season in the Indian region. suggesting that lightning activity can also be influenced by seasonal variation.

Overall, the present findings align with previous studies from other regions, thereby providing validation for the results identified along the 500 kV transmission line in Peninsular Malaysia. Furthermore, the analysis of the combined 2015-2017 dataset provides a generalized view of the relationships, smooths out annual inconsistencies and highlights the dominant trends influencing lightning occurrence.

4.3 Principal Component Analysis Interpretation

In the analysis of PCA, six PCs denoted as PC1 to PC6, along with their corresponding eigenvalues and variance percentages, were computed for the years 2015, 2016, and 2017, as summarized in Table 6. Each PC explains a percentage of the total variation in the dataset. According to Kaiser's rules, only PCs with an eigenvalue greater than 1.0 were considered [43]. Correspondingly, analysis of eigenvalues revealed that the two main PCs, PC1 and PC2, could explain more than 75% of the total variance within the datasets. It is worth noting that cumulative variance between 70% and 90% is generally considered reasonably good [44]. In each year, the percentage of cumulative variances varies slightly, with 81.81% for 2015, 76.92% for 2016, and 87.46% for 2017.

Based on the eigenvalues, the first PC, PC1, stands out as the most significant in explaining the amount of variance, followed by PC2. In the year 2015, PC1 accounts for 64.9% of the variance, signifying its explanation of 64.9% of the data. Subsequently, PC2 contributes to a cumulative variance of 81.81%. Therefore, considering both PC1 and PC2 provides a thorough grasp of the data relationship, as they collectively explain 81.81% of the variance. To interpret the findings meaningfully, the discussion focuses solely on PC1 and PC2 to explore the correlation between negative lightning GFD, rainfall, mean temperature, mean wind speed, pressure, and elevation.

In the aggregated PCA, which includes datasets from 2015 to 2017, only two PCs are considered and retained for interpretation, reflecting the multidimensional structure of lightning-related meteorological and topographical variables across temporal scales. In particular, PC1 emerged as the most dominant among the other PCs, accounting for 53.77% of the cumulative variance, while PC2 contributed 20.16% to achieve a cumulative variance of 73.93% when combined with the percentage of PC1. Based on the values from the table, having high cumulative variance suggests a well-consolidated data structure that facilitates efficient dimensionality reduction without any significant loss of information [44].

Through combining datasets over several years, the PCA captures underlying patterns that persist beyond annual fluctuations. It offers a more stable representation of environmental and geographical interactions that influence negative lightning GFD. The dominance of PC1 across the merged dataset from the three-year period is likely linked to key weather factors as well as topographic features. At the same time, PC2 and beyond seem to capture more localized or seasonal changes, including year-to-year shifts in wind patterns or pressure. Overall, the combined analysis confirms the robustness of the PCs in representing the lightning-associated viability across the years studied. In addition, this approach enhances model generalizability and supports the identification of persistent drivers of lightning activity across climatological periods [44], [45].

Table 6 *Eigenvalues and percentage of variance of each principal component*

	PC1	PC2	PC3	PC4	PC5	PC6
2015						
Eigenvalue	3.89	1.01	0.71	0.38	1.80E-03	4.30E-04
Proportion Variance (%)	64.9	16.91	11.76	6.39	0.03	0.01
Cumulative Variance (%)	64.9	81.81	93.57	99.96	99.99	100
2016						
Eigenvalue	3.56	1.06	0.89	0.49	2.10E-03	7.60E-04
Proportion Variance (%)	59.27	17.66	14.8	8.23	0.03	0.01
Cumulative Variance (%)	59.27	76.92	91.72	99.95	99.99	100
2017						
Eigenvalue	4.46	0.79	0.58	0.17	1.90E-03	5.60E-05
Proportion Variance (%)	74.28	13.17	9.71	2.8	0.03	9.40E-04
Cumulative Variance (%)	74.28	87.46	97.16	99.97	100	100
Combined data of 2015-2017						
Eigenvalue	3.23	1.21	0.86	0.51	0.10	0.09
Proportion Variance (%)	53.77	20.16	14.31	14.31	1.72	1.47
Cumulative Variance (%)	53.77	73.93	88.24	96.81	98.53	100

The interpretation of PCA results, as depicted in Table 7, relies on the factor loadings of the variables with respect to the two PCs. Factor loadings reveal how variables relate to the PCs. If a factor loading exceeds 0.5, it indicates that the variable significantly influences the variation of the PC [32]. In 2015, the PC1 distinguished elevation, rainfall, mean temperature, mean wind speed, and pressure. PC1 demonstrates strong positive correlations with mean wind speed (0.99), pressure (0.97), and mean temperature (0.95), while elevation and rainfall appear to be negatively correlated with PC1. Conversely, PC2 is associated with negative lightning GFD (0.99).

In 2016, PC1 differentiates elevation, rainfall, mean temperature, mean wind speed, and pressure, suggesting positive correlations with pressure (0.98), mean wind speed (0.96), and mean temperature (0.66) as well as substantial negative correlations with rainfall (-0.80) and elevation (-0.69). The results also reveal that PC2 is associated with negative lightning GFD (0.61) and mean temperature (0.54). In 2017, PC1 distinguishes all six variables highlighting strong positive correlations with pressure (0.98), mean wind speed (0.97) and mean temperature (0.96) and significant negative correlations with rainfall (-0.93), elevation (-0.69) and negative lightning GFD (-0.54) and PC2 is associated with negative lightning GFD (0.80).

Throughout the three-year period, a trend emerged in which the factor loading on PC1 became increasingly positively correlated with negative lightning GFD, elevation, and rainfall. In comparison, negative lightning GFD presented a negative correlation with mean temperature, mean wind speed, and pressure. These consistent patterns were observed during the years 2016 and 2017, indicating a notable trend in the data analysis.

When the data from the three-year period is combined, PC1 maintains strong positive correlations with several meteorological variables such as mean temperature, mean wind speed, and pressure. Among these, the strongest positive correlation is observed with mean wind speed (0.91), indicating that this variable has the most pronounced influence on PC1. For negative correlation, PC1 demonstrates the strongest correlation with rainfall (-0.78). However, the correlation with elevation and negative lightning GFD is still regarded as moderate (-0.55) and (-0.50), respectively. Conversely, PC2 exhibits a strong positive correlation with negative lightning GFD (0.53), implying a significant connection between this component and lightning activity. Additionally, PC2 suggests a moderate positive correlation with mean temperature (0.58), further reinforcing its link to thermal conditions [13], [46]. It also maintains a moderate negative correlation with rainfall (-0.56), implying an inverse relationship that could reflect seasonal or convective influences inherent to tropical lightning formation.

The variable correlation plot in Fig. 8 provides a simultaneous visualization of observations and variables in the PCA, thereby enabling a better understanding of the relationships between them. As the correlation patterns remained relatively similar across all individual years, 2015, 2016, and 2017, the combined data from 2015 to 2017 were used for the overall interpretation. Note that the direction and length of the arrows on the plot are the key to understanding the influence of each variable on the PCs. Furthermore, examining this plot and observing the patterns for the variables and relating them to the context of the negative lightning GFD, PC1, which represents the general atmospheric conditions, reveals that pressure, mean wind speed, and mean temperature are all closely aligned along the positive axis. This implies that these variables have a strong and positive relationship among them and hence constitute the terms warm, windy, and high pressure component. In contrast, rainfall and elevation are in the negative PC1 axis, which shows that there is a great negative correlation between these variables and this component. As it can be deduced based on the results, negative lightning GFDs are strongly correlated by time of less warm, less windy, as well as lower pressure weather, and related to the high elevation.

The results of the analysis presented by the variable correlation diagram and the value factor loading of the PCA make it easy to understand the study area. The analysis demonstrates that negative lightning GFD forms a unique weather occurrence strongly unrelated to the warm, windy as well as high-pressure conditions. It is, in fact, related only to a higher altitude. In addition, such a finding replicates the findings in the Yunnan, Tibetan Plateau and Arctic regions, where topography favouring lightning is typical of lightning activity at the high altitudes of the region, promoting atmospheric instabilities leading to the occurrence of thunderstorms and lightning [12][47].

Table 7 *Factor loading of the variables with PC1 and PC2*

Variables	2015		2016		2017		2015-2017	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Negative GFD	-0.02	0.99	-0.33	0.61	-0.54	0.80	-0.50	0.53
Elevation	-0.68	0.03	-0.69	-0.42	-0.69	0.20	-0.55	0.31
Rainfall	-0.77	-0.15	-0.80	0.44	-0.93	-0.23	-0.78	-0.56
Mean Temperature	0.95	-0.06	0.66	0.54	0.96	0.07	0.76	0.58
Mean Wind Speed	0.99	0.03	0.96	-0.17	0.97	0.19	0.91	-0.14
Pressure	0.97	-0.05	0.98	0.06	0.98	0.11	0.82	-0.40

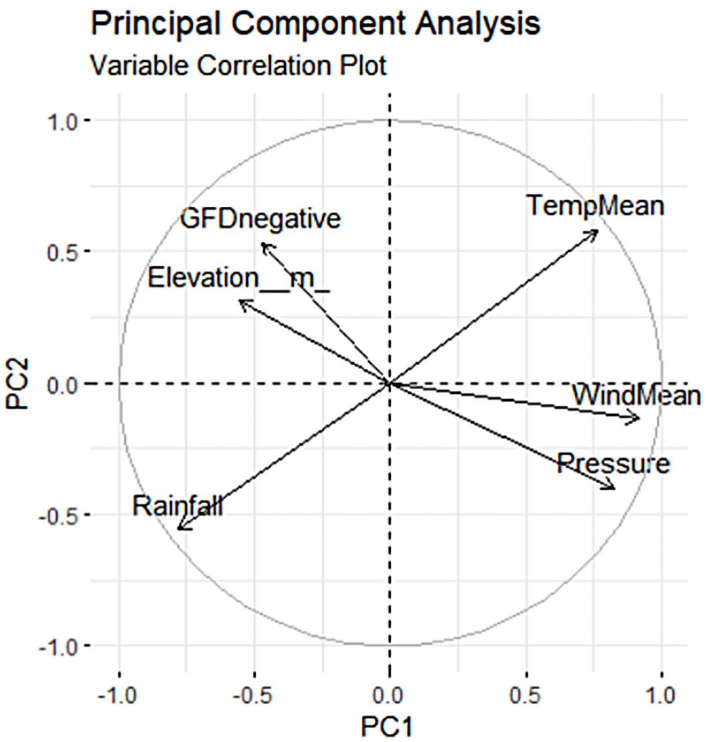

Fig. 8 *Variable correlation plot for the combined data from 2015 to 2017*

5. Conclusion

In conclusion, this study highlights the effectiveness of employing a GIS-based approach to investigate the correlation in lightning events. By integrating GIS technology, the complex relationships between lightning activity, weather variables, and topography along the 500 kV transmission line A-B in Peninsular Malaysia were thoroughly investigated.

The analysis revealed that negative lightning GFD represents a distinct weather event that is largely uncorrelated with warm, windy, and high-pressure conditions. Instead, it is specifically associated with higher elevation. However, since the exact relationship varies by region, with local topography and climate playing important roles, further investigation is required. Overall, the results provide valuable insights into understanding lightning distribution, which aids in predicting and preparing for severe weather events, especially in regions such as Peninsular Malaysia. Furthermore, the results of this study deepen the understanding of meteorological phenomena and develop a foundation and GIS framework for future research.

A more detailed temporal analysis, including an investigation of how seasonal variations affect lightning activity is required. Additionally, this established GIS framework can be extended to provide a crucial first insight into the data and directly support a comprehensive transmission line assessment. This approach would enable the identification of specific line segments or tower locations that are highly susceptible to lightning strikes due to their topographical and meteorological exposure. In the future, this approach will be extended to other utility assets and geographic regions with different topographies and weather patterns to validate the findings. Ultimately, this research will enhance the ability of researchers and engineers to predict and mitigate the impacts of lightning strikes in vulnerable regions, thereby improving the resilience of power grids.

Acknowledgement

This work was supported by Universiti Tenaga Nasional Malaysia [J510051043], Tenaga Nasional Berhad, and UNITEN R&D Sdn. Bhd [U-TS-RD-19-30] and Advanced Lightning, Power and Energy Research Centre (ALPER), Universiti Putra Malaysia (UPM) [GP-GPB/2022/9708800]. The kind appreciation is extended to TNB Research Sdn. Bhd. for lightning data, Malaysia Meteorological Department for weather data, TNB Grid Maintenance for transmission line data, and the Department of Survey and Mapping Malaysia (JUPEM) for the Peninsular Malaysia basemap and topography map.

Conflict of Interest

Authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *N. Roslan U. A. U. Amirulddin, M. Z. A. Ab-Kadir;* ***data collection:*** *N. Roslan, U. A. U. Amirulddin, M. Z. A. Ab-Kadir, N. Abdullah;* ***analysis and interpretation of results****: N. Roslan, U. A. U. Amirulddin;* ***draft manuscript preparation:*** *N. Roslan. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Rawi, I. M., Kadir, M. Z. A. A., & Izadi, M. (2017). Seasonal variation of transmission line outages in peninsular Malaysia, *Pertanika J. Sci. Technol*, *25*(5), 213-220, Accessed: Feb. 13, 2023. [Online]. Available: http://psasir.upm.edu.my/id/eprint/55832/1/24-JTS%28S%29-0100-2016-4thProof.pdf

[2] Ab-Kadir, M. Z. A. (2016). Lightning severity in Malaysia and some parameters of interest for engineering applications, *Thermal Science*, *20*(suppl. 2), 437-450, https://doi.org/10.2298/TSCI151026028A

[3] Islam, M. A., Chan, A., Khaw, S. K. K., Salleh, S. A., Zakaria, N. H., Isa, N. A., ... & Azari, M. (2019). Importance of lightning detection network development in Peninsular Malaysia, *IOP Conference Series: Earth and Environmental Science* (Vol. 385, No. 1, p. 012036), https://doi.org/10.1088/1755-1315/385/1/012036

[4] Roslan, N., Amirulddin, U. A. U., Ab Kadir, M. Z. A., & Abdullah, N. (2024). Development of GIS-based Ground Flash Density and its Statistical Analysis for Lightning Performance Evaluation of Transmission Lines in Peninsular Malaysia, *Pertanika Journal of Science & Technology*, *32*(1), https://doi.org/10.47836/pjst.32.1.21

[5] Rawi, I. M., & Kadir, M. Z. A. A. (2016). A case study and observation on cause of transmission line outages in Malaysia. In *Proc. CIGRE Session* (Vol. 46, pp. 1-7).

[6] Roslan, N., Ab-Kadir, M. A., Amirulddin, U. U., Shafri, H. M., & Abdullah, N. (2023). GIS-Based Development and Visualization of Ground Flash Density Maps during the Monsoon Seasons in Peninsular Malaysia. In *2023 12th Asia-Pacific International Conference on Lightning (APL)* (pp. 1-6). IEEE. https://doi.org/10.1109/APL57308.2023.10181634

[7] Qie, K., Qie, X., & Tian, W. (2021). Increasing trend of lightning activity in the South Asia region. *Science Bulletin*, *66*(1), 78-84, https://doi.org/10.1016/j.scib.2020.08.033

[8] Yadava, P. K., Sharma, A., Payra, S., Mall, R. K., & Verma, S. (2023). Influence of meteorological parameters on lightning flashes over Indian region. *Journal of Earth System Science*, *132*(4), 179, https://doi.org/10.1007/s12040-023-02188-w

[9] Singh, R., Singh, V., Gautam, A. S., Kumar, S., Singh, K., Soni, P. S., ... & Gautam, S. (2025). Temporal and Spatial Variations in Lightning Activity and Meteorological Parameters Across the Indian Himalayan Region and Indo-Gangetic Plains. *Asia-Pacific Journal of Atmospheric Sciences*, *61*(2), 8, https://doi.org/10.1007/s13143-025-00391-x

[10] Damase, N. P., Banik, T., Paul, B., Saha, K., Sharma, S., De, B. K., & Guha, A. (2021). Comparative study of lightning climatology and the role of meteorological parameters over the Himalayan region. *Journal of Atmospheric and Solar-Terrestrial Physics*, *219*, 105527, https://doi.org/10.1016/j.jastp.2020.105527

[11] Tinmaker, M. I. R., Ghude, S. D., Dwivedi, A. K., Islam, S., Kulkarni, S. H., Khare, M., & Chate, D. M. (2022). Relationships among lightning, rainfall, and meteorological parameters over oceanic and land regions of India. Meteorology and Atmospheric Physics, 134(1), 5, https://doi.org/10.1007/s00703-021-00841-x

[12] Deng, Y., Li, M., Wang, Y., He, X., Wen, X., Lan, L., ... & Pan, H. (2023). Relationship between lightning activity and terrain in high-altitude mountainous areas. *IEEE Transactions on Power Delivery*, *38*(5), 3561-3570, https://doi.org/10.1109/TPWRD.2023.3283606

[13] Zhang, H., Deng, Y., Wang, Y., Lan, L., Wen, X., Fang, C., & Xu, J. (2024). Extraction of factors strongly correlated with lightning activity based on remote sensing information. *Remote Sensing*, *16*(11), 1921, https://doi.org/10.3390/rs16111921

[14] Shankar, A., Anand, A., Kumar, A., & Singh, S. P. (2025). Mapping the lightning hotspot: reducing deaths in Jharkhand, India, using spatiotemporal and statistical insights. *MAUSAM*, *76*(3), 778-795, https://doi.org/10.54302/mausam.v76i3.6763

[15] Kumar, S., Gautam, A. S., Kamra, A. K., Singh, K., Potdar, S. S., & Siingh, D. (2024). Lightning distribution and its association with surface thermodynamics parameters over a western Himalayan region. *Journal of Earth System Science*, *133*(4), 234, https://doi.org/10.1007/s12040-024-02406-z

[16] Hazmi, A., Hamid, M. I., Fernandez, R., Andre, H., Pratama, R. W., & Emeraldi, P. (2023). The correlation between lightning and various weather parameters in the Padang monsoon system. *Indonesian Journal of Electrical Engineering and Computer Science*, *31*(1), 1-9, https://doi.org/10.11591/ijeecs.v31.i1.pp1-9

[17] Yusop, N., Ahmad, M. R., Ching, T. S., Baharin, S. A. S., Esa, M. R. M., & Sidik, M. A. B. (2022). Correlation analysis between lightning flashes and rainfall rate during a flash flood thunderstorm. *Indonesian Journal of Electrical Engineering and Computer Science*, *28*(3), 1322, https://doi.org/ 10.11591/ijeecs.v28.i3.pp1322-1329

[18] Bahari, N., Esa, M. R. M., & Wahab, M. A. (2023, June). Correlation Analysis of Lightning and Flash Flood Events using Pearson Model in Southeast Peninsular Malaysia. In *2023 12th Asia-Pacific International Conference on Lightning (APL)* (pp. 1-6). IEEE, https://doi.org/ 10.1109/APL57308.2023.10181418

[19] Yu, M., Kubiczek, J., Ding, K., Jahanzeb, A., & Iqbal, N. (2022). Revisiting SDG-7 under energy efficiency vision 2050: the role of new economic models and mass digitalization in OECD. *Energy Efficiency*, *15*(1), 2, https://doi.org/10.1007/s12053-021-10010-z

[20] Ramasubramanian, B., & Ramakrishna, S. (2023). What's next for the Sustainable Development Goals? Synergy and trade-offs in affordable and clean energy (SDG 7). *Sustainable earth reviews*, *6*(1), 17, https://doi.org/10.1186/s42055-023-00069-0.

[21] Oshiobugie A. and Adeel P. (2024) Building Resilient and Sustainable Energy Infrastructure: Enhancing Renewable Integration and Emergency Power Management, *International Journal of Energy and Environmental Research*, 12 (3), 20-39, https://doi.org/10.37745/ijeer.13/vol12n32039

[22] Abdullah, N., & Hatta, N. M. (2012). Cloud-to-ground lightning occurrences in Peninsular Malaysia and its use in improvement of distribution line lightning performances. In *2012 IEEE International Conference on Power and Energy (PECon)* (pp. 819-822). IEEE, https://doi.org/10.1109/PECon.2012.6450330

[23] Wooi, C. L., Abdul-Malek, Z., Ahmad, N. A., Mokhtari, M., & Khavari, A. H. (2016). Cloud-to-ground lightning in Malaysia: A review study. *Applied Mechanics and Materials*, *818*, 140-145, https://doi.org/10.4028/www.scientific.net/AMM.818.140

[24] Nicora, M., Mestriner, D., Brignone, M., Bernardi, M., Procopio, R., Fiori, E., ... & Rachidi, F. (2022). Assessment of the lightning performance of overhead distribution lines based on lightning location systems data. *International Journal of Electrical Power & Energy Systems*, *142*, 108230, https://doi.org/10.1016/j.ijepes.2022.108230

[25] Herrera, J., Younes, C., & Porras, L. (2018). Cloud-to-ground lightning activity in Colombia: A 14-year study using lightning location system data. *Atmospheric Research*, *203*, 164-174, https://doi.org/10.1016/j.atmosres.2017.12.009

[26] Hunt, H. G. P., Nixon, K. J., & Naudé, J. A. (2017). Using lightning location system stroke reports to evaluate the probability that an area of interest was struck by lightning. *Electric Power Systems Research*, *153*, 32-37, https://doi.org/10.1016/j.epsr.2016.12.010

[27] Etaga Harrison, O., & Ifeanyichukwu, O. (2021). Methods of estimating correlation coefficients in the presence of influential outlier (s). *Methods*, *4*(3), 157-185, https://doi.org/10.52589/ajmss-llnzxuoz

[28] P. Schober and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesth Analg*, vol. 126, no. 5, 2018, https://doi.org/10.1213/ANE.0000000000002864

[29] Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia*, *126*(5), 1763-1768, https://doi.org/10.1093/ckj/sfab085

[30] Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), 433-459, https://doi.org/10.1002/wics.101

[31] Morales, J. A., Orduna, E., Rehtanz, C., Cabral, R. J., & Bretas, A. S. (2015). Comparison between principal component analysis and wavelet transform 'filtering methods for lightning stroke classification on transmission lines. *Electric power systems research*, *118*, 37-46, https://doi.org/10.1016/j.epsr.2014.05.018

[32] Hashim, N. I. M., Noor, N. M., & Annas, S. (2018, December). Influence of meteorological factors on variations of particulate matter (PM10) concentration during haze episodes in Malaysia. In *AIP Conference Proceedings* (Vol. 2045, No. 1, p. 020103). AIP Publishing LLC, https://doi.org/10.1063/1.5080916

[33] I. Si-ahmed, L. Hamdad, C. J. Agonkoui, Y. Kande, and S. Dabo-Niang, "Principal component analysis of multivariate spatial functional data," *Big Data Research*, vol. 39, Feb. 2025, https://doi.org/10.1016/j.bdr.2024.100504

[34] Morales, J. A., Orduna, E., Rehtanz, C., Cabral, R. J., & Bretas, A. S. (2015). Comparison between principal component analysis and wavelet transform 'filtering methods for lightning stroke classification on transmission lines. *Electric power systems research, 118,* 37-46, https://doi.org/10.1016/j.epsr.2014.05.018

[35] Chen, D. (2024). High frequency principal component analysis based on correlation matrix that is robust to jumps, microstructure noise and asynchronous observation times. *Journal of Econometrics*, *240*(1), 105701, https://doi.org/10.1016/j.jeconom.2024.105701

[36] 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, https://doi.org/10.30812/varian.v8i1.3183

[37] Ravaglio, M. A., Küster, K. K., Santos, S. L. F., Toledo, L. F. R. B., Piantini, A., Lazzaretti, A. E., ... & da Silva Pinto, C. L. (2019). Evaluation of lightning-related faults that lead to distribution network outages: An experimental case study. *Electric Power Systems Research*, *174*, 105848, https://doi.org/10.1016/j.epsr.2019.04.026

[38] Souto, L., Taylor, P. C., & Wilkinson, J. (2023). Probabilistic impact assessment of lightning strikes on power systems incorporating lightning protection design and asset condition. *International Journal of Electrical Power & Energy Systems*, *148*, 108974, https://doi.org/10.1016/j.ijepes.2023.108974

[39] Wooi, C. L., Abdul-Malek, Z., Salimi, B., Ahmad, N. A., Mehranzamir, K., & Vahabi-Mashak, S. (2015). A comparative study on the positive lightning return stroke electric fields in different meteorological conditions. *Advances in Meteorology*, *2015*(1), 307424, https://doi.org/10.1155/2015/307424

[40] Zhu, B., Lin, B., Liu, B., & Liu, H. (2025). Analysis of temporal and spatial distribution of CG lightning activity and its relationship with terrain features around Jiuxianshan mountain. *Journal of Atmospheric and Solar-Terrestrial Physics*, 106546, https://doi.org/10.1016/j.jastp.2025.106546

[41] Cai, L., Chen, M., Han, T., Zhou, M., Cao, J., Wang, J., & Fan, Y. (2025). The Relationship between Cloud-to-ground Lightning Spatial Distribution and Topography in Guangzhou. *Journal of Atmospheric and Solar-Terrestrial Physics*, 106554, https://doi.org/10.1016/j.jastp.2025.106554

[42] Weisburd, D., Britt, C., Wilson, D. B., & Wooditch, A. (2021). Measuring association for scaled data: Pearson's correlation coefficient. In *Basic statistics in criminology and criminal justice* (pp. 479-530). Cham: Springer International Publishing, https://doi.org/10.1007/978-3-030-47967-1_14

[43] Everitt, B., & Dunn, G. (2001). *Applied multivariate data analysis* (Vol. 2). London: Arnold. https://doi.org/10.1002/9781118887486

[44] Ayesha, S., Hanif, M. K., & Talib, R. (2020). Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, *59*, 44-58, https://doi.org/10.1016/j.inffus.2020.01.005

[45] Migenda, N., Möller, R., & Schenck, W. (2021). Adaptive dimensionality reduction for neural network-based online principal component analysis. *PloS one*, *16*(3), e0248896, https://doi.org/10.1371/journal.pone.0248896

[46] Nandhulal, K., Varikoden, H., & Vishnu, R. (2025). A regional study on convective lightning activity and its interaction with sea surface temperature and thermodynamic indices. *Discover Atmosphere*, *3*(1), 5, https://doi.org/10.1007/s44292-025-00032-y

[47] Qie, X., Yair, Y., Di, S., Huang, Z., & Jiang, R. (2024). Lightning response to temperature and aerosols. *Environmental Research Letters*, *19*(8), 083003, https://doi.org/10.1088/1748-9326/ad63bf