



KEMENTERIAN SUMBER ASLI DAN
KELESTARIAN ALAM
Ministry of Natural Resources and Environmental
Sustainability

**MALAYSIAN METEOROLOGICAL DEPARTMENT
MINISTRY OF NATURAL RESOURCES AND
ENVIRONMENTAL SUSTAINABILITY**

Research Publication No. 2/2024

**Relationship Between Rainfall Event
and Lifted Index in Sepang**

**Wan Fariza Mustafah, Diong Jeong Yik and
Nursalleh K Chang**

RESEARCH PUBLICATION NO. 2/2024

Relationship Between Rainfall Event
and Lifted Index in Sepang

By

Wan Fariza Mustafah, Diong Jeong Yik and Nursalleh K Chang

All rights reserved. No part of this publication may be reproduced in any form, stored in a retrieval system, or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher.

Perpustakaan Negara Malaysia

Data Pengkatalogan-dalam-Penerbitan



Cataloguing-in-Publication Data

Perpustakaan Negara Malaysia

A catalogue record for this book is available
from the National Library of Malaysia

ISBN 978-967-2327-20-2

Published and printed by:
Jabatan Meteorologi Malaysia
Jalan Sultan
46667 Petaling Jaya
Selangor Darul Ehsan
Malaysia

CONTENTS

NO.	SUBJECT	PAGE
1.0	INTRODUCTION	1
2.0	DATA	3
3.0	METHODOLOGY / ALGORITHM	6
	3.1 BAYES' THEOREM OF CONDITIONAL PROBABILITY	6
	3.2 K-MEAN CLUSTERING	8
4.0	RESULT AND DISCUSSIONS	10
	4.1 MONTHLY DISTRIBUTION OF THE LIFTED INDEX	10
	4.2 MONTHLY DISTRIBUTION OF ACCUMULATED RAINFALL	12
	4.3 PROBABILISTIC ANALYSIS	17
	4.4 K-MEAN CLUSTERING ANALYSIS	20
5.0	CONCLUSION	26
	REFERENCE	29
	APPENDIX	30

Relationship Between Rainfall Event and Lifted Index in Sepang

Wan Fariza Mustafah, Diong Jeong Yik and Nursalleh K Chang

Abstract

This study examines the effectiveness of the Lifted Index (LI) in predicting daytime atmospheric instability in tropical regions, in particular Sepang, Selangor, focusing on its relationship with rainfall. Statistical analysis reveals that using zero as a threshold to classify stable and unstable conditions is unreliable, except in January and February.

Accumulated rainfall displays seasonal patterns, morning to midday rainfall is minimal but prominent in July and August, largely driven by Sumatra squall events. The rainfall is more common in the late afternoon to early evening throughout the year (except Southwest Monsoon, SWM), with notable peaks in April and November that align with the inter-monsoon period and early Northeast Monsoon (NEM).

Bayesian analysis updates our understanding of rain forecasts by incorporating new information, refining predictions, and improving the accuracy of forecasting rain. This approach is particularly useful when integrating the LI during the NEM. However, it is less effective during the SWM, where K-means clustering identifies five distinct clusters, highlighting the limitations of LI. The study concludes that LI is more reliable during the NEM but requires region-specific thresholds for accurate tropical forecasting.

1.0 Introduction

Lifted Index (LI) is one of the popular indicators to determine atmospheric stability. Negative (positive) LI indicates an unstable (stable) atmosphere. Latent instability is instability caused by the release of latent heat. The more latent heat that is released, the more a parcel of air will warm. Latent instability refers to the atmosphere's potential to become unstable due to latent heat release, which occurs when water vapour in rising air condenses into liquid water (cloud formation). As an air parcel rises and cools, the condensation process releases latent heat, which warms the parcel, making it more buoyant. This additional warming can enhance the parcel's ability to rise further, leading to convective instability and potentially triggering thunderstorms.

Many studies have been carried out in the mid-latitude regions on the usage of LI, but such research in the tropical areas is still lacking. Haklander and Van Delden (2003) found that thunderstorms are most influenced by latent instability, followed by potential instability, and least by conditional instability in mid-latitude. The LI is a useful forecasting tool for latent instability forecasting (Galway 1956) and many studies show that the lowest 100 hPa LI has the best skill for thunderstorm forecasting (Haklander and Van Delden 2003) and (Kunz 2007). According to (Galway 1956) LI values between -3°C and -5°C indicate marginal instability, and LI is particularly useful for predicting latent instability, which often manifests as thunderstorms about six hours later in the afternoon.

This study aims to fill a knowledge gap about the use of LI in the tropics, where atmospheric conditions differ significantly from those in mid-latitudes. Specifically, the research aims to evaluate the effectiveness of LI as a forecasting index by examining how it performs under different seasonal conditions in the tropics. Through a detailed analysis

of LI values across different monsoon seasons, we seek to determine whether region-specific adjustments are necessary to improve the accuracy of weather forecasting in the tropics. In this study, morning LI data from the Sepang Meteorological Station is combined with hourly rainfall data from surrounding stations within a 50-kilometer radius of Sepang, providing a more comprehensive assessment of how well LI provide new information in predicting rainfall in the local area.

2.0 Data

The stability of the lower troposphere can be readily assessed using the LI. In this study, the LI is calculated by determining the temperature difference between the environmental temperature and the parcel temperature at 500 hPa. To determine the parcel temperature at 500 hPa, the surface temperature and moisture of the parcel are used, based on conditions in the low-level boundary layer. This involves calculating the mean values of temperature and mixing ratios from the lower atmospheric levels. The hypothetical parcel is then lifted dry adiabatically to the lifting condensation level (LCL) and pseudo-adiabatically to 500 hPa to obtain the parcel temperature at this level.

$$\text{Lifted Index} = T_{\text{environment at 500 hPa}} - T_{\text{parcel at 500 hPa}}$$

The LI used in this study is derived from radiosonde observations at the Sepang Meteorological Station (WMO Station Number: 48650; Coordinates: 2.73155°N, 101.70291°E; Elevation: 17 m a.s.l., Source: [UWYO Weather](#)), operated by the Malaysian Meteorological Department. Our analysis covers observation data from 08:00 LT from 1 January 2000 to 31 December 2021. A negative LI indicates that the air parcel is warmer than the surrounding air at 500 mb, making it buoyant and likely to continue rising, and cause thunderstorms. A positive LI indicates that the air parcel is cooler and thus stable, thus making thunderstorms unlikely to occur.

LI can be particularly useful in predicting afternoon thunderstorms, which are often driven by daytime heating. As the surface heats up during the day, air becomes more buoyant, and negative LI values may form in the late afternoon, making thunderstorms more likely. **Table 1** shows the threshold values for the LI that apply to the eastern two-thirds of the United States (U.S.). For regions at higher elevations, such as western

Canada and the western U.S., these threshold values should be adjusted upward (i.e., made less negative) to account for the differences in atmospheric conditions at altitude (Peppler 1988).

Lifted Index Value	Criteria
$LI > 6$	Very Stable Conditions
$1 < LI < 6$	Stable Conditions, Thunderstorms Not Likely
$-2 < LI < 0$	Slightly Unstable, Thunderstorms Possible, With Lifting Mechanism (i.e., cold front, daytime heating)
$-6 < LI < -2$	Unstable, Thunderstorms Likely, Some Severe with Lifting Mechanism
$LI < -6$	Very Unstable, Severe Thunderstorms Likely with Lifting Mechanism

Table 1: Classification of Atmospheric Stability Based on LI Values and Associated Thunderstorm Likelihood (Peppler 1988).

Wong et al. (2009) identified significant seasonal rainfall patterns in Malaysia, particularly noting an upward trend in both annual and monsoon rainfall in the west coast region. Their findings suggest that rainfall variability is largely driven by local and regional factors rather than global phenomena like ENSO events. Building on this understanding, our study examines how these rainfall patterns interact with the LI by analyzing hourly

rainfall data from 2000 to 2021 collected at three meteorological stations: Sepang, Subang, and Petaling Jaya. Rainfall was recorded hourly from 08:00 LT to 19:00 LT, over the entire period from January 1, 2000, to December 31, 2021. We summed the hourly rainfall data from all three stations and divided the accumulated rainfall into four time periods: the first 6 hours (08:00 – 13:00 LT), 9 hours (08:00 – 16:00 LT), 12 hours (08:00 – 19:00 LT), and a specific focus on the period from 14:00 to 19:00 LT, which is usually linked to afternoon thunderstorms.

3.0 Methodology / Algorithm

3.1 Bayes' Theorem of Conditional Probability

In this study, a rain day event is defined as any period during which the total rainfall recorded by any of the three meteorological stations exceeded 0.01 mm over 3-hour, 6-hour, 9-hour, and 12-hour intervals, with particular attention given to the 14:00 to 19:00 LT period. The probability of rain day events occurring during these periods is calculated by dividing the number of rain days by the total number of days in each month over the 22-year study period. Conversely, the probability of no rain day event is determined by counting the days of no rain day in each period and dividing by the total number of days in that month over the same 22-year period.

The probability of a rain day occurring with a negative LI is calculated by dividing the number of rain days in each period with a negative LI by the total number of rain day in that month. Conversely, the probability of a no rain day occurring with a negative LI is determined by dividing the number of no rain day in each period with a negative LI by the total number of no rain-days in that month.

Subsequently, we applied Bayes' Theorem of conditional probability to calculate the probability of a rain-day event given a negative LI, as shown below:

$$P(R|LI^-) = \frac{P(R) \times P(LI^-|R)}{(P(R) \times P(LI^-|R)) + (P(NR) \times P(LI^-|NR))}$$

Where:

$P(R)$ = probability of rain-day event

$P(NR)$ = probability of no rain-day event

$P(LI|R)$ = probability of LI negative given rain-day event

$P(LI|NR)$ = probability of LI negative given no rain-day event

Besides that, we also calculate the probability of rain event given LI is above and below the normalized value. The intention is to remove seasonality and find an appropriate threshold. The LI data can be normalized by dividing each anomaly data with the standard deviation (σ). Normalized is defined as:

$$y_{j,i} = \frac{x - \bar{x}_i}{\sigma_i}$$

$y_{j,i}$ = normalized data (j = day of the month, i = month)

x = lifted index

\bar{x}_i = lifted index monthly mean

$$\sigma_i = \sqrt{\frac{\sum_{i=1}^n (x - \bar{x}_i)^2}{n}}$$

anomaly data for each month is obtained by subtracting the data with the respective monthly mean. The anomaly data is then divided by month standard deviation to obtain normalized data.

3.2 K-mean Clustering

We applied a simple and widely used K-mean clustering algorithm (Dehariya et al. 2010) to understand LI and rainfall data based on their characteristics and similarities. K-means clustering is a recursive partition-based clustering method. It starts by first randomly assigning k centers (centroids) representing k clusters, then sets each data point to the nearest cluster center. For every new data point assigned to a cluster, the algorithm recalculates the mean (centroids) of the clusters. This process is repeated till all the data points are assigned to a cluster. Different initialization allocates initial centroids at different locations; thus, final clustering will differ every time. Each centroid is iterated until a minimal distance between the data points and respective cluster centroids is achieved and the centroids' location remains selectively stable. The distance is measured based on the Within-Cluster Sum of Square (WCSS), defined as the sum of the squared distance between each point and the respective centroids. The number of clusters, k , generally varies with WCSS, the sum of the squared distance between each point and the respective centroids.

$$WCSS = \sum_{k=1}^n \left(\sum_{d_i \in C_k}^m \text{distance}(d_i, C_k)^2 \right)$$

Given n is the total number of data points and m is the total number of input clusters. C_k are the cluster centroids, and d_i is the data point in each cluster.

For K-mean clustering, it is required to give number of cluster (k) in advance. One way to select optimal number clusters is the Elbow method. The elbow method is a graphical method for finding the optimal k value in a k-means clustering algorithm. The

elbow graph shows the within-cluster-sum-of-square (WCSS) values on the y-axis corresponding to the different values of k (on the x-axis).

4.0 Results and Discussion

4.1 Monthly Distribution of the LI

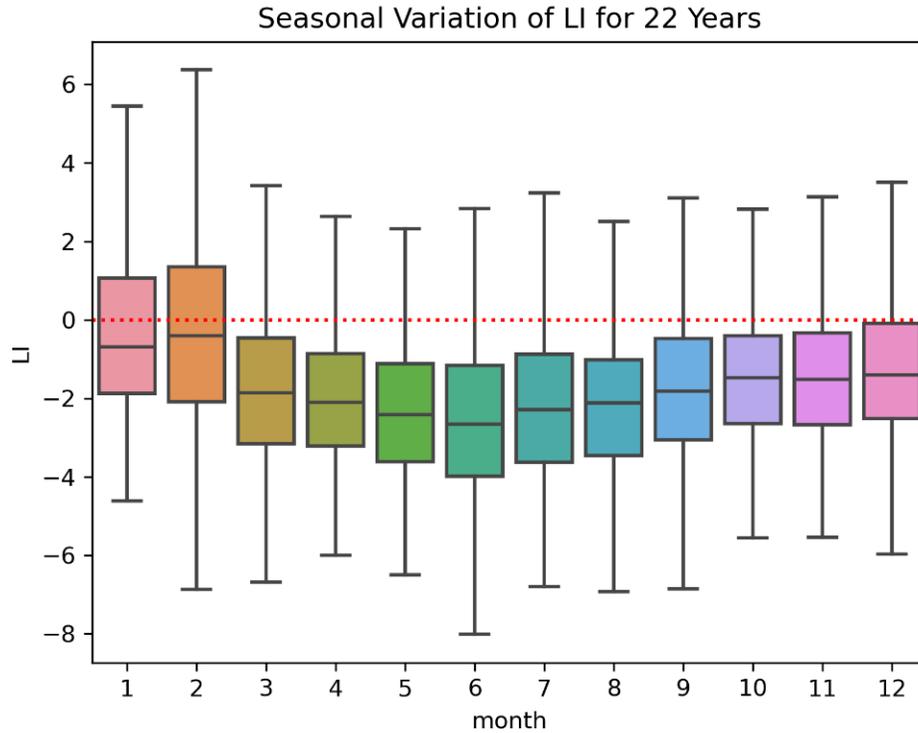


Figure 1: The box plots represent the monthly distribution of LI values from 2000 to 2021 at Sepang Meteorological Station, with the median shown as the central line in each box. The boxes show the interquartile range (IQR) from the 25th to the 75th percentile, indicating the middle 50% of LI values for each month. Whiskers extend to the maximum and minimum values, excluding outliers. The red dashed line at $LI = 0$ indicates the threshold between stable (positive LI) and unstable (negative LI) atmospheric conditions.

Figure 1 shows the monthly distribution of the LI from 2000 to 2021 at Sepang Meteorological Station, with the vertical Y-axis representing LI values and the horizontal X-axis representing the months. The median LI values are mostly negative throughout

the year, indicating a generally unstable atmosphere during most months in the research area. Only January and February have median LI values close to zero, at -0.55 and -0.34, respectively. A median LI value close to zero suggests a balance of stable and unstable atmospheric conditions during these months, with an even mix of negative and positive LI values.

From May to September, the median LI values consistently fall below -2, reflecting more unstable conditions, coinciding with the SWM, which may enhance instability and convective activity. The monthly distribution of the LI suggests that using zero as a threshold to differentiate between stable (positive LI) and unstable (negative LI) atmospheres may not be appropriate year-round, due to the uneven distribution of negative and positive LI values across the months, except in January and February.

4.2 Monthly Distribution of Accumulated Rainfall

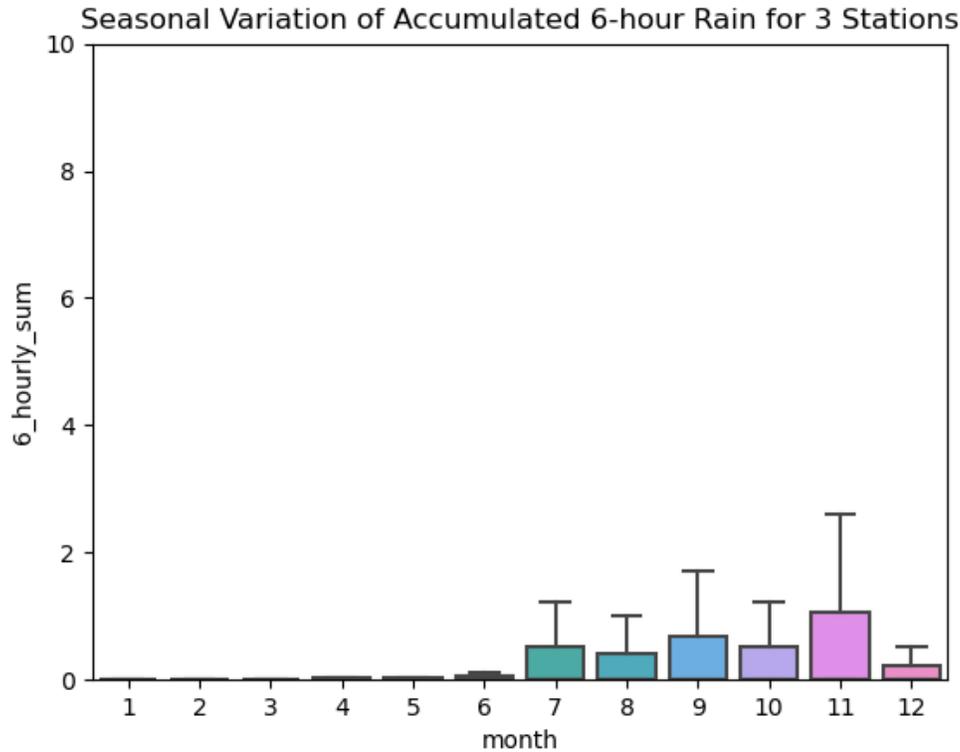


Figure 2: Monthly distribution of accumulated 6-hour rainfall starting at 08:00 LT for three stations from January to December. The box plots depict the distribution of rainfall sums, with the median as the central line and the interquartile range (IQR) representing the middle 50% of values. Whiskers extend to the minimum and maximum values, excluding outliers.

The monthly distribution of accumulated rain differs across three time periods: 6 hours (from 08:00 LT to 13:00 LT), 9 hours (from 08:00 LT to 16:00 LT), and 12 hours (08:00 LT to 19:00 LT). **Figure 2** illustrates the first 6 hours (08:00 LT – 13:00 LT) of accumulated rain at three stations. There is minimal accumulated rain from January to May, with median values close to zero, indicating little to no rainfall in the first 6 hours

across the study area during this period. Starting in June, there is a gradual increase in accumulated rain, which becomes more pronounced from July to October. Accumulated rain rises notably, peaking in November with the highest median and variability, reflecting a significant increase in rainfall. Accumulated rainfall decreases slightly in December but remains relatively high compared to earlier months. Despite the increasing trend in accumulated rain from June to December, the maximum accumulated rainfall remains low, consistently below 4 mm, showing that the research areas receive less rain from morning to midday.

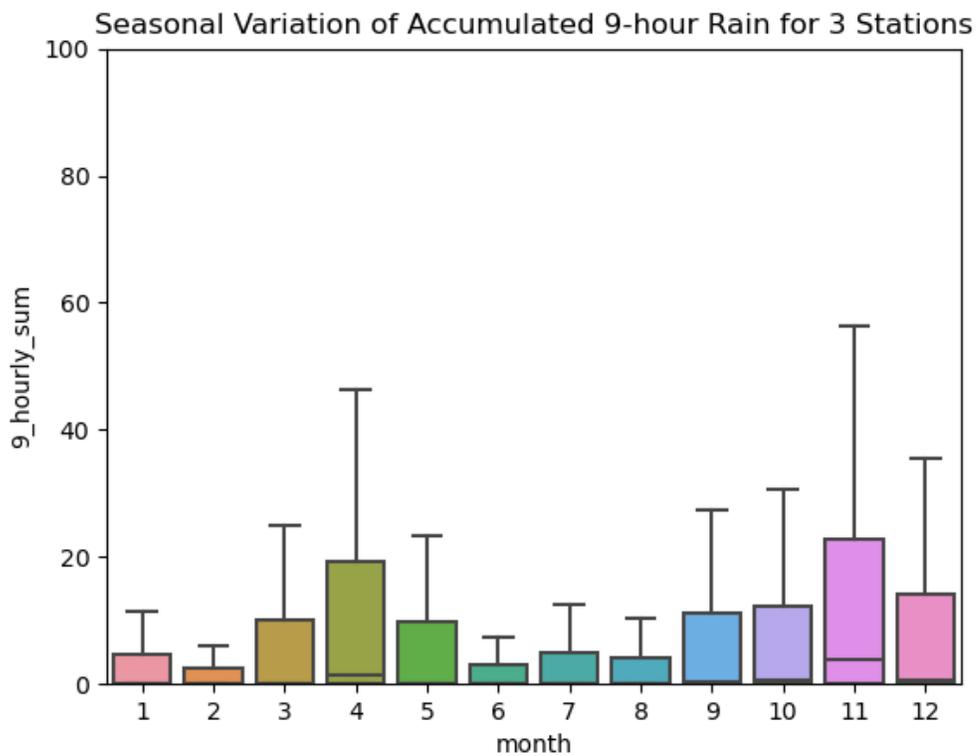


Figure 3: Same as **Figure 2**, except for accumulated 9-hour rainfall starting at 08:00 LT.

Figure 3 shows the monthly distribution of accumulated rain for the first 9 hours (08:00 LT – 16:00 LT). The accumulated rain is low from January to March, with median

values near zero. A noticeable increase starts in April, with a significant rise in both the median and the interquartile range (IQR). Between May and August, the median accumulated rain decreases from April. From September to December, the levels rise again, with November having the highest values, median, and IQR, indicating substantial rainfall and variability. In December, the accumulated rain slightly decreases but is still higher than in the early months.

In comparing the first 6-hour and 9-hour periods, June, July, and August stand out as months with rainfall more concentrated in the morning to midday (first 6 hours). This rain pattern is due to Sumatra squalls, which often occur in the morning to midday in the western part of Peninsular Malaysia (PM), where this study is conducted (Lo and Orton 2016). April shows the highest increase in accumulated 9-hour rainfall, with amounts exceeding 20 mm. This is due to the inter-monsoon period, which often brings late afternoon thunderstorms to the western part of PM. November stands out with the highest rain levels and variability in both 6-hour and 9-hour accumulated rainfall, as this month falls within the NEM season, commonly called the wet season.

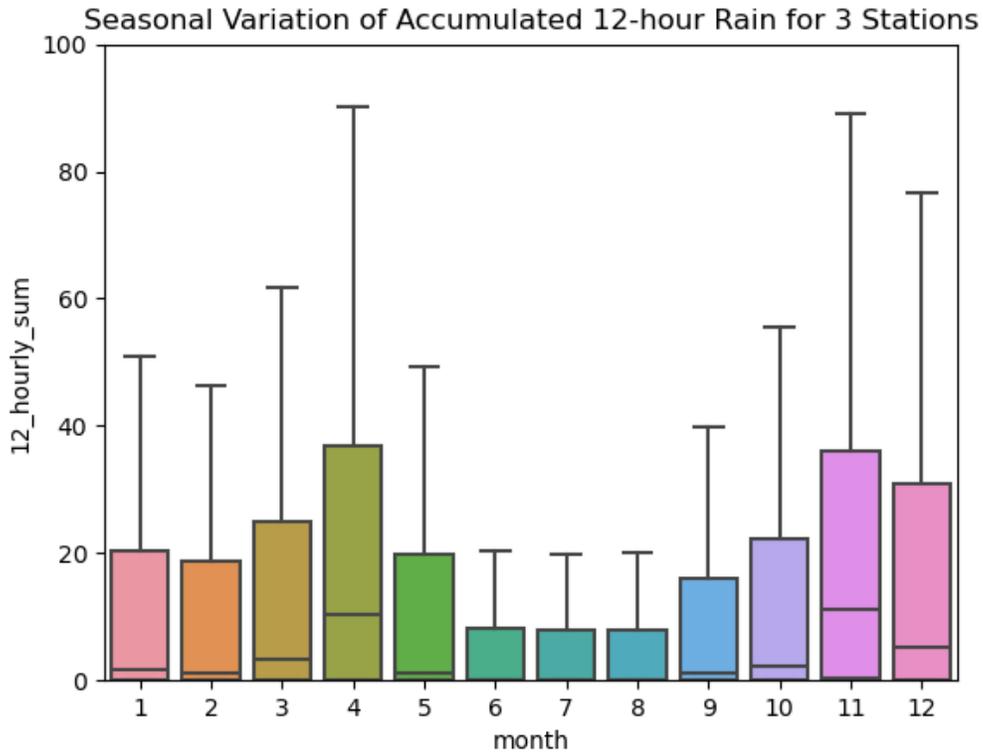


Figure 4: Same as **Figure 2**, except for accumulated 12-hour rainfall starting at 08:00 LT.

Figure 4 shows a similar seasonal trend for the first 9-hour and 12-hour (08:00 LT – 19:00 LT) periods, with an increase in rainfall across all months in the 12 hours. The data highlights two prominent peaks in accumulated rain, occurring in April and November (**Figures 3 and 4**). Median rainfall values are higher in April and November compared to other months. The consistent increases across all months suggest that the research area receives more rain in the afternoon or early evening. The accumulated rainfall is relatively low from June to September, with the median values significantly lower compared to the inter-monsoon peaks. This period typically corresponds to the SWM, which brings drier conditions to the research areas. Overall, the box plot highlights the bimodal nature of

rainfall in this tropical region, with two distinct peaks corresponding to the inter-monsoon and early NEM period and lower rainfall during the SWM.

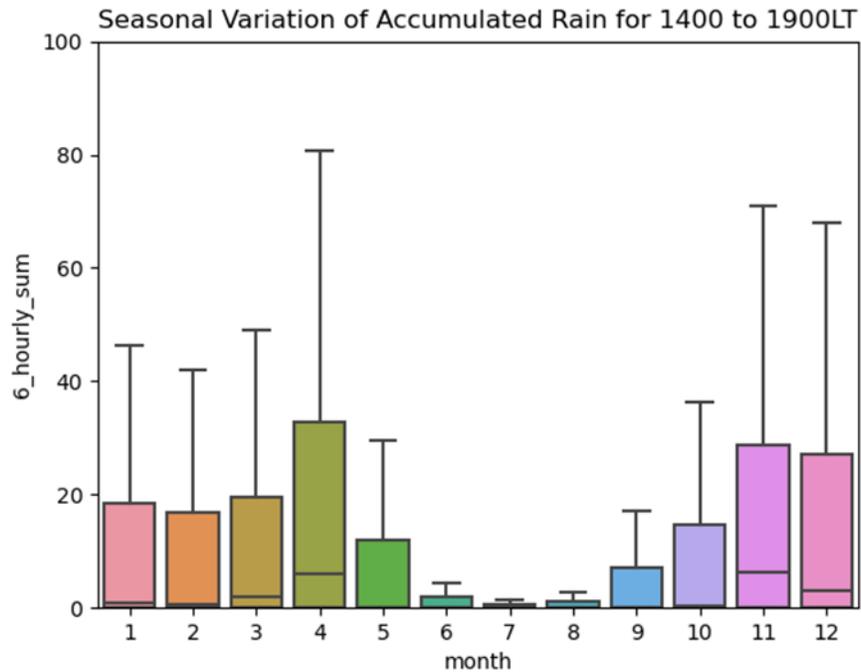
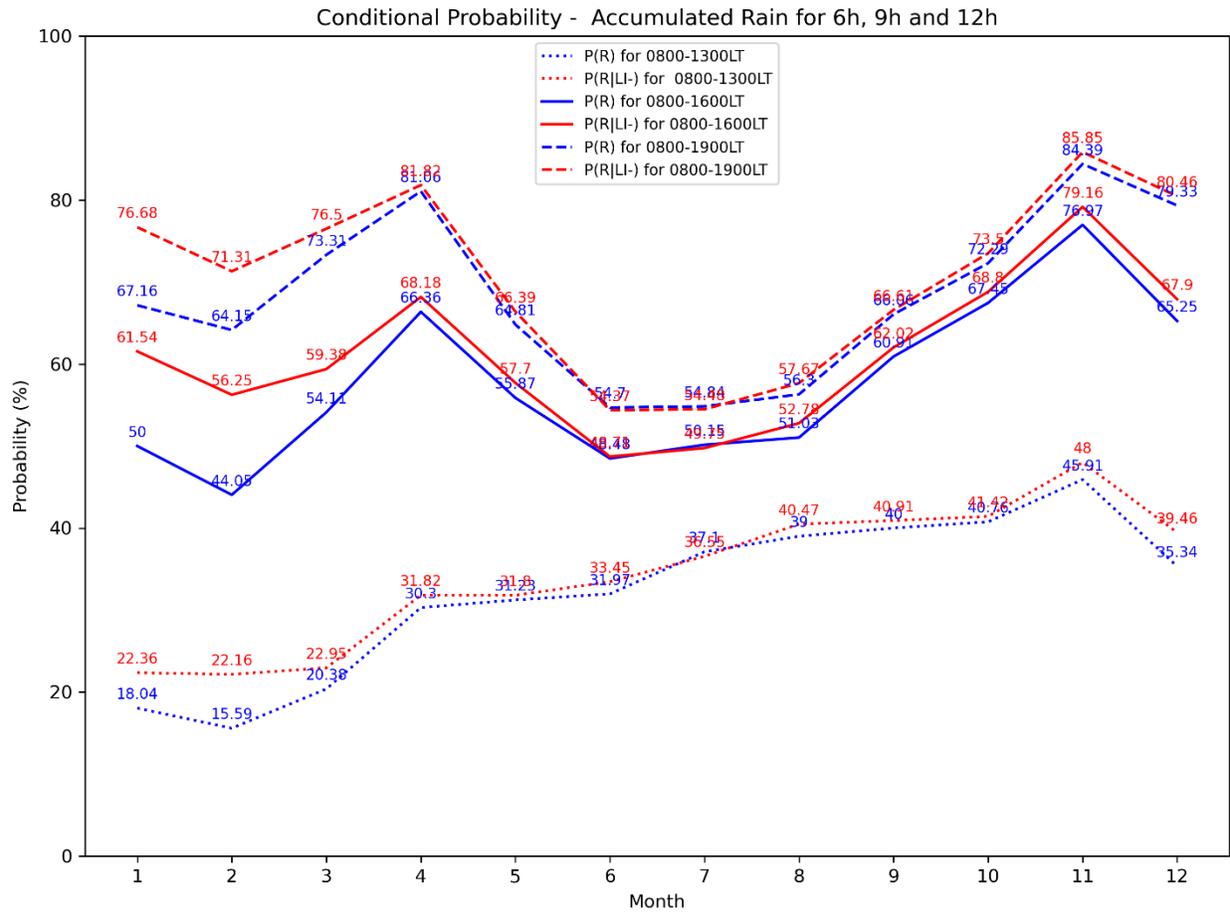


Figure 5: Same as **Figure 2**, except for accumulated rainfall from 14:00 LT to 19:00 LT at three stations from 2000 to 2021.

We analyzed the monthly distribution of accumulated rainfall between 14:00 and 19:00 LT to prove that the research areas experience more rain in the afternoon to early evening. The results confirm a significant increase in rainfall during these hours (**Figure 5**). The minimal increase in rainfall observed in July and August indicates that the research area predominantly experiences rain from morning to midday during the SWM, likely due to Sumatra squall events. The rises in accumulated rain in April and November across all periods can be attributed to the inter-monsoon season and the early phase of the NEM, respectively.

4.3 Probabilistic Analysis



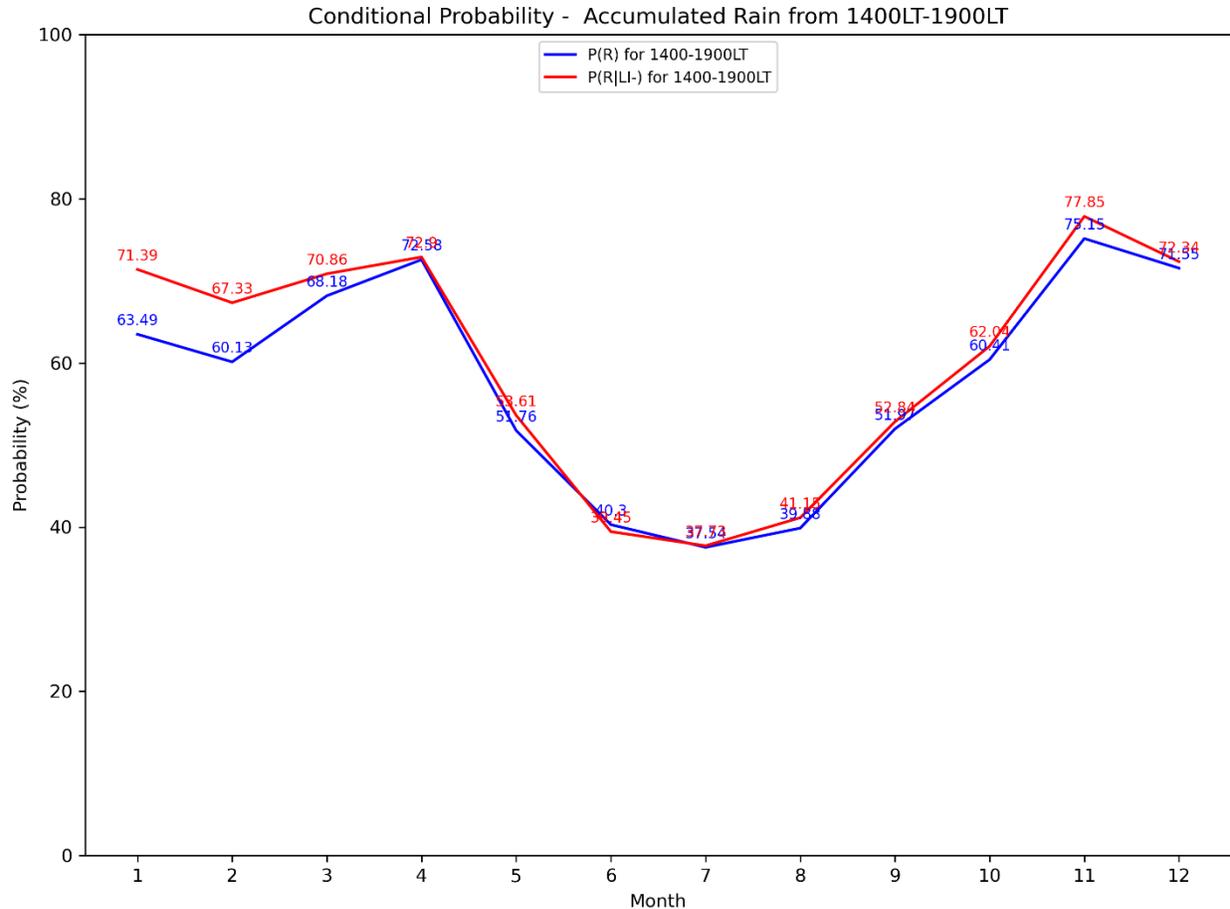


Figure 6 (a) and (b): The probability of rain (dotted/straight/dashed blue line) and the probability of rain given a negative LI (dotted/straight/dashed red line) for each month over 22 years, with monthly probabilities calculated for 6-hour (dotted line), 9-hour (straight line), 12-hour (dashed line), and specific time between 14:00 LT – 19:00 LT (straight line in Figure (b)) accumulated rain.

Figure 6(a) shows the probability of rain ($P(R)$) and the conditional probability of rain when information about negative LI is added ($P(R|LI-)$). The blue dotted line represents $P(R)$, while the red dotted line represents $P(R|LI-)$ for the first 6-hour period (0800-1300 LT). Both probabilities remain below 50% throughout the year, indicating a generally low probability of rain from morning to midday in the study area. This pattern is

particularly evident during the mid-to-late NEM season, where the addition of negative LI information does not significantly increase the likelihood of rain. However, the $P(R)$ and $P(R|LI-)$ increase significantly after 6 hours indicating more frequent rain in the late afternoon or early evening, this is supported by the accumulated rains boxplot and the analysis in **Figure 6(b)**. The probabilities peak in April and November for both the 9-hour and 12-hour periods, corresponding to the inter-monsoon period in April and early NEM in November. During SWM probabilities remain around 50%, suggesting an equally likely chance of rain in the research area. Across all periods, the probability of rain is lowest in February and highest in November, suggesting that February is typically the driest month, while November is the wettest month in the study area.

We use Bayes' theorem to determine how likely it is to rain when a negative LI is present. According to the results shown in **Figures 6 (a) and (b)**, the conditional probability of rain, when information about negative LI is added ($P(R|LI-)$), is generally higher than the overall probability of rain alone ($P(PR)$) in most months in all periods. This finding suggests that the LI provides valuable additional information by indicating atmospheric instability. Including LI in rain probability analysis enhances our understanding of atmospheric conditions, allowing us to identify periods when the environment is more favorable for rain. As a result, this approach helps refine predictions and increases the likelihood of forecasting rain. The only exceptions are June and July, where the probability of rain with negative LI is the same as the probability of rain alone.

The negative LI gives extra useful information, especially during the NEM season. In February, for example, including negative LI data helped increase the chance of rain by more than 12% during the 9-hour period. This indicates that using the LI is particularly

useful from January to March during the NEM, as it significantly increases the probability of rain, especially for the 9-hour accumulated rain period. In contrast, a negative LI does not impact the probability of rain in June and July, suggesting that the LI is not helpful during these months. Despite June and July having the most negative LI values, which usually indicate a more unstable atmosphere, they still have lower probabilities of rain compared to other months. This finding suggests that LI alone is not sufficient for predicting rain events, and further investigation is needed to understand the factors that influence rainfall.

4.4 K-mean Clustering

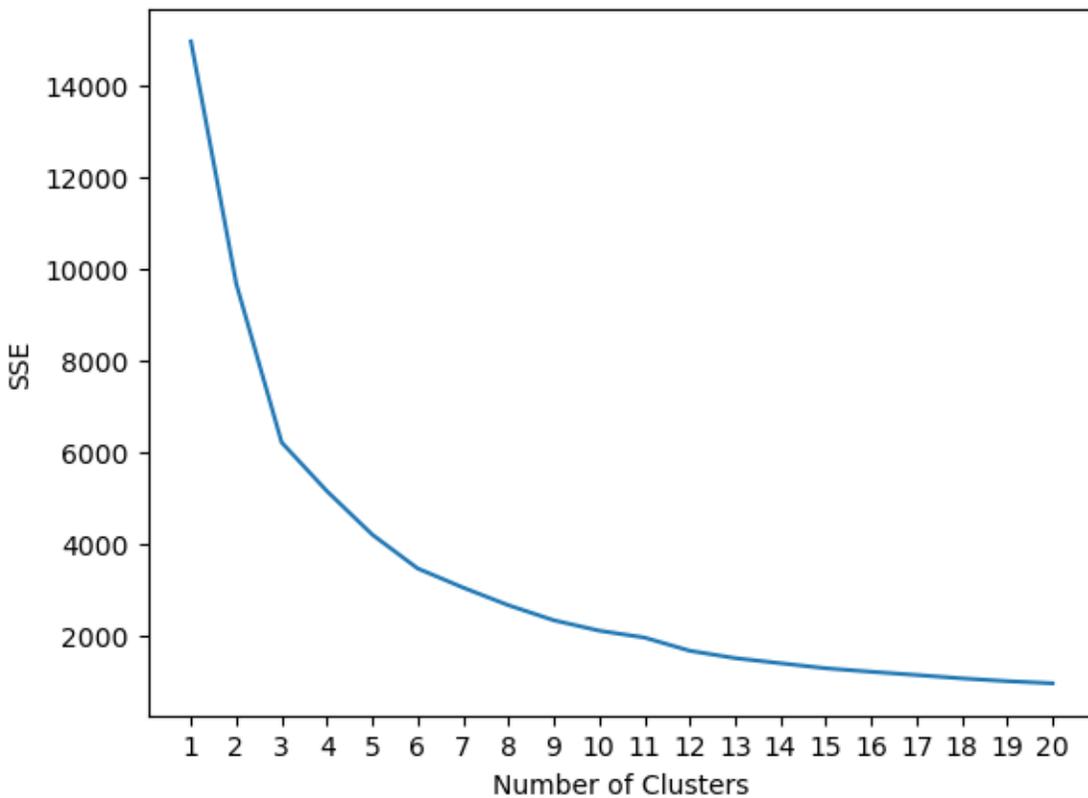


Figure 7: The elbow method with its pivotal point as an indicator for the optimal number of clusters corresponding to 5 clusters.

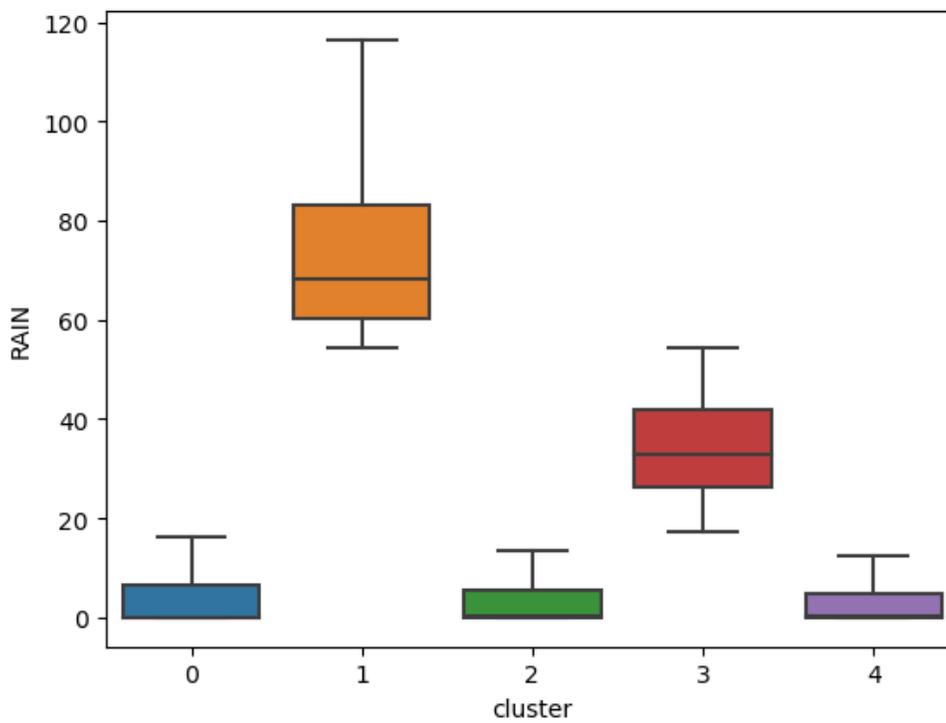
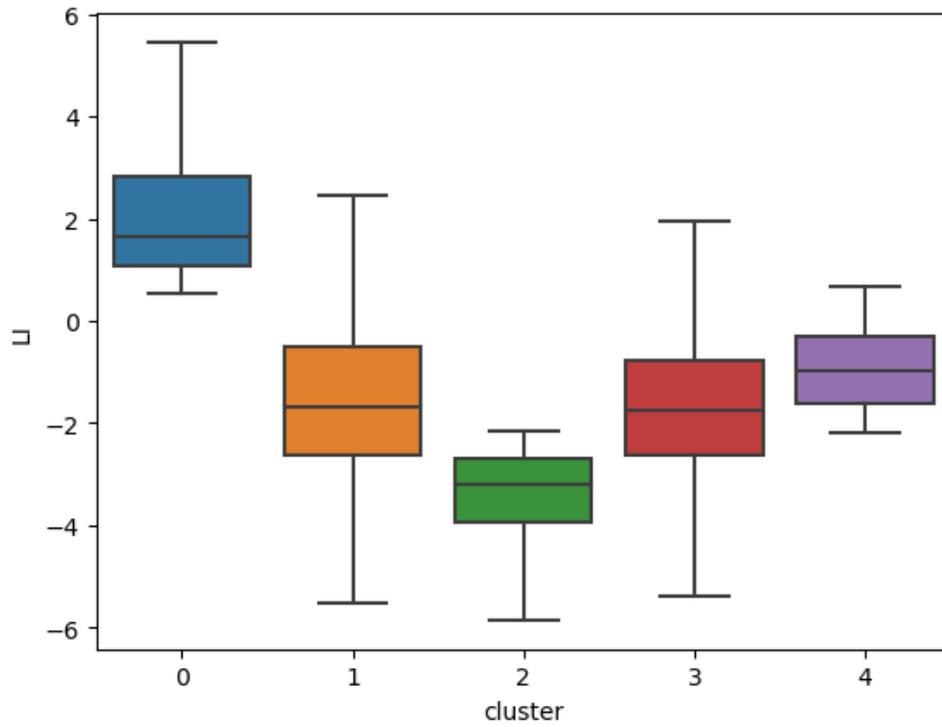


Figure 8: K-means clustering analysis for **(a)** LI and **(b)** maximum rainfall data

To investigate why the LI is significant during the NEM but not during the SWM, we used a **K-means clustering analysis** with 22 years of LI and maximum rainfall data. K-Means analysis requires the initialization of a number of clusters, k . Then each data point is assigned to the nearest cluster centres or centroids. Each centroid is iterated until a minimal distance between the data points and respective cluster centroids is achieved and the centroids' location remains selectively stable. The distance is measured based on the Within-Cluster Sum of Square (WCSS), defined as the sum of the squared distance between each point and the respective centroids. The number of clusters, k , generally varies with WCSS, the sum of the squared distance between each point and the respective centroids. We used the **elbow method** to find the best number of groups. The elbow method works by plotting how much the data varies within different numbers of groups and looking for the point where the variation starts to decrease more slowly. This point shows the ideal number of groups that best represent the data without being too complicated. In our case, the optimal k value is the point at which the graph forms an elbow (**Figure 7**).

This method allowed us to objectively group atmospheric conditions for both seasons, revealing why the LI is a reliable indicator during the NEM but less so during the SWM. The clustering analysis identified five distinct groups based on monthly LI and maximum rainfall from 2000 to 2021. **Figure 8 (a) and (b)** present box plots illustrating the characteristics of these clusters.

- **Cluster 0** has only positive LI values, with a median around +2, indicating stable atmospheric conditions. The maximum rainfall in this cluster is low, ranging from 0 to 20 mm, with a median close to 0 mm, suggesting little to no rainfall.

- **Clusters 1 and 3** have over 75% negative LI values, with a median around -2, indicating more unstable atmospheric conditions. Cluster 1 shows the highest rainfall, followed by Cluster 3.
- **Cluster 2** consists entirely of negative LI values, with a most negative median around -3, indicating the most unstable atmospheric conditions. However, despite this instability, the maximum rainfall is low, with a median near 0 mm, which contradicts the usual LI theory that suggests more instability should lead to more rain.
- **Cluster 4** has LI values ranging from about -2 to 1, with a median around -1, indicating slightly unstable conditions. This cluster also shows little to no rainfall.

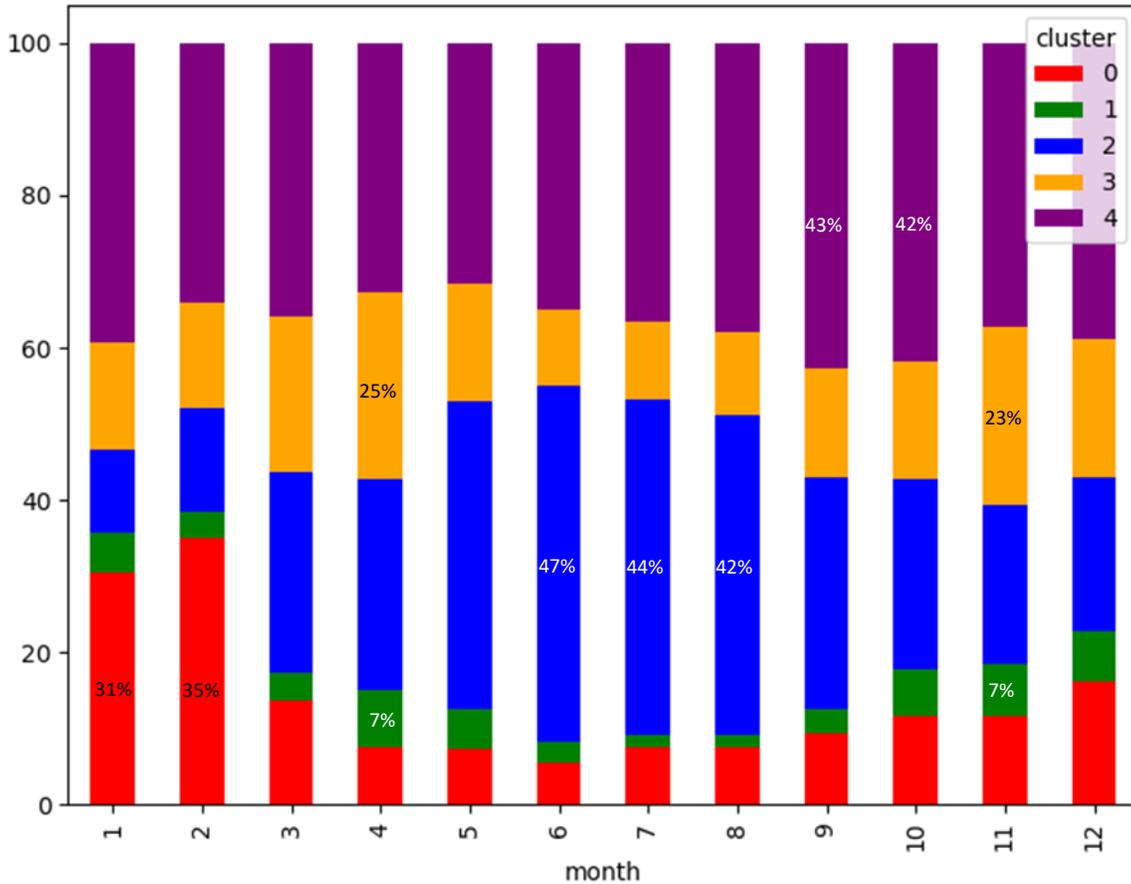


Figure 9: Bar chart showing the clustering results based on the monthly distribution of the LI and maximum rainfall at Sepang Meteorological Station from 2000-2021. The X-axis represents months, and the Y-axis shows the percentage of each cluster. Each bar is divided by colors representing clusters 0 to 4, as shown in the legend.

Clusters 0, 1, and 3 follow the fundamental principle of LI theory, which states that a positive LI indicates a stable atmosphere with less chance of rain, while a negative LI suggests the opposite. Cluster 0 has the highest percentage in January (31%) and February (35%), meaning these months have the most positive LI data compared to other months. This balance between positive and negative LI data allows for a more accurate assessment of atmospheric conditions. In January and February, clusters 0, 1, and 3

collectively represent 50% and 52% of the data, respectively, indicating that half of the sample during these months follows the fundamental principle of LI theory. This supports the finding that the probability of rain increases significantly with a negative LI during the NEM, especially in February. The analysis shows that the probability of rain is 44.05% without any extra information. However, when we add useful information from the LI—for example, when LI is negative, which suggests an unstable atmosphere—the chance of rain increases to 56.25%. This means that using the LI helps improve our ability to predict rain.. This suggests that the negative LI is a strong predictor of rainfall during the NEM season, especially in the typically drier month of February, highlighting the usefulness of LI in forecasting convective rainfall.

Cluster 2 contradicts the fundamental principle of LI theory, particularly during the SWM season. This cluster is notably prevalent in June, July, and August, with percentages of 47%, 44%, and 42%, respectively. This prevalence suggests that a significant portion of the atmospheric data during these months does not conform to the expected relationship between LI and rainfall, as indicated by the LI theory. Similarly, Cluster 4, despite having a less negative LI value range, also results in minimal rainfall. When the percentages for clusters 2 and 4 are combined, they exceed 80%, highlighting that a large majority of the data during the SWM does not adhere to LI theory. This suggests that other atmospheric factors are likely more important in influencing rainfall patterns during the SWM, making LI less useful in this situation. Overall, while the LI can help predict rainfall in some seasons, its effectiveness decreases during the SWM when other weather influences take over.

5.0 Conclusion

The monthly distribution of the LI and accumulated rainfall in tropical areas reveals complex interactions between atmospheric stability and rainfall patterns throughout the year. The use of zero as a threshold to distinguish between stable (positive LI) and unstable (negative LI) atmospheric conditions is found to be unreliable across most months due to the uneven distribution of LI values, except in January and February. This uneven distribution suggests that atmospheric conditions in tropical areas do not consistently align with the interpretation of the LI, making it less effective as a forecast index for rainfall outside of these two months.

Rainfall patterns across the research areas exhibit clear seasonal trends. From January to May, morning to midday rainfall is minimal, but a gradual increase begins in June. The morning to midday rainfall is most prominent in July and August, largely driven by Sumatra squall events. Except during the SWM, rainfall is more common in the late afternoon to early evening throughout the year, with notable peaks in April and November that align with the inter-monsoon period and early NEM. The analysis reveals that the probability of rain when the LI is negative is highest during the later part of NEM season (January and February). This suggests a stronger correlation between negative LI values and rainfall events during this period. However, the LI is found to be less effective during the SWM season, as the relationship between negative LI and rainfall weakens, reducing its utility as a forecasting index during these months.

The analysis of clusters based on the LI provides key insights into how atmospheric stability relates to rainfall in different seasons. Clusters 0, 1, and 3 align with the fundamental principle of LI theory, where positive LI suggests a stable atmosphere

with low rain probability, and negative LI indicates instability and higher rain chances. January and February have the most positive LI data, with clusters 0, 1, and 3 representing around half of the data in these months. This means that more than half of the data collected during these months supports the fundamental principle of LI theory, making it a reliable way to predict rainfall during this season. During the SWM season (June, July, and August), cluster 2—which does not align with the fundamental principles of LI theory—makes up a large portion of the data. This suggests that these months do not follow the expected relationship between LI and rainfall. Additionally, cluster 4 shows minimal rainfall, despite having slightly less negative LI values. When combining data from clusters 2 and 4, it becomes clear that over 80% of the information during the SWM season does not conform to LI theory, highlighting the limitations of LI as a predictive tool for rainfall in this period. This indicates that other atmospheric factors are more influential in determining rainfall patterns during the SWM, making LI less relevant in this context.

Overall, the clustering analysis helps us understand how the LI works differently in various seasons and regions. It shows that LI is a helpful tool for predicting rainfall during the NEM, however, we need to be careful and consider other factors when using it during the SWM season. The results indicate that the probability of rain across four time periods is low during the SWM, remaining below 50% compared to other seasons. The monthly distribution of accumulated rainfall also shows that during SWM, rain in the research area occurs mainly from morning to midday. Additionally, the unchanged probability of rain when negative LI is added suggests that the primary cause of rain during this period is likely due to Sumatra squall events, rather than heating or atmospheric instability.

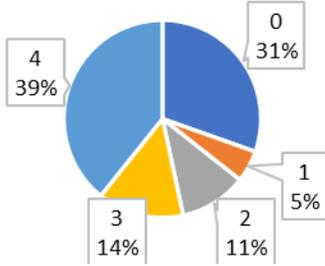
The use of the LI in tropical weather forecasting must be carefully adjusted, as the typical thresholds applied in mid-latitude regions do not directly transfer to tropical environments, where convection is influenced by different factors. To enhance the accuracy of LI-based forecasts in tropical regions, we should establish specific thresholds that capture the unique characteristics of tropical instability. This should include the use of other static stability indices alongside the LI to provide a more comprehensive understanding of atmospheric conditions.

REFERENCE

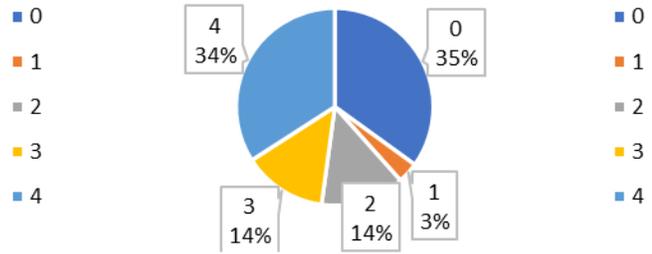
- Dehariya, V. K., Shrivastava, S. K., & Jain, R. C. (2010). Clustering of Image Data Set Using K-Means and Fuzzy K-Means Algorithms. In *2010 International Conference on Computational Intelligence and Communication Networks* (pp. 386–391). <https://doi.org/10.1109/CICN.2010.80>
- Galway, J. G. (1956). The Lifted Index as a Predictor of Latent Instability. *Bulletin of the American Meteorological Society*, *37*(10), 528–529. <https://doi.org/10.1175/1520-0477-37.10.528>
- Haklander, A. J., & Van Delden, A. (2003). Thunderstorm predictors and their forecast skill for the Netherlands. *Atmospheric Research*, *67–68*, 273–299. [https://doi.org/https://doi.org/10.1016/S0169-8095\(03\)00056-5](https://doi.org/https://doi.org/10.1016/S0169-8095(03)00056-5)
- Kunz, M. (2007). The skill of convective parameters and indices to predict isolated and severe thunderstorms. *Natural Hazards and Earth System Sciences*, *7*(2), 327–342. <https://doi.org/10.5194/nhess-7-327-2007>
- Lo, J., & Orton, T. (2016). The general features of tropical Sumatra Squalls. *Weather*, *71*, 175–178. <https://doi.org/10.1002/wea.2748>
- Peppler, R. A. (1988). A Review of Static Stability Indices and Related Thermodynamic Parameters. <https://hdl.handle.net/2142/48974>. Accessed 30 September 2024

APPENDIX

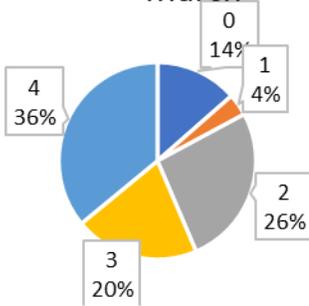
January



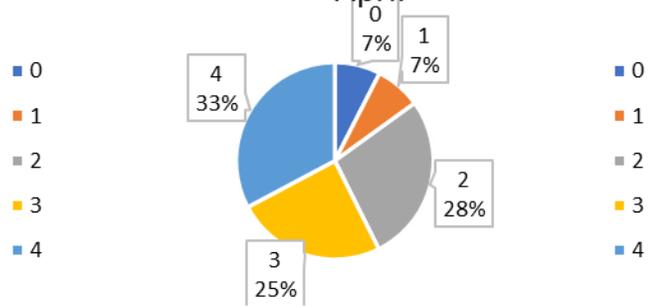
February



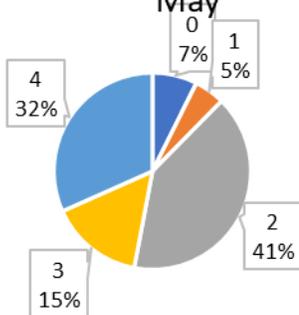
March



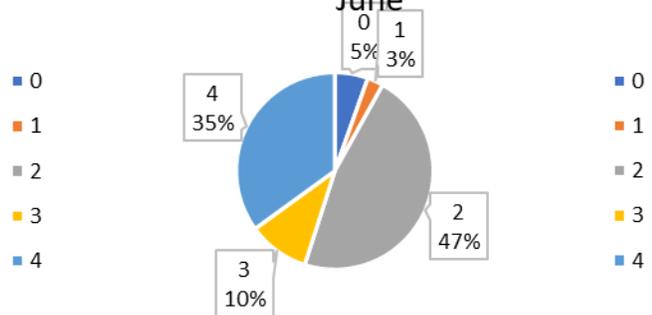
April



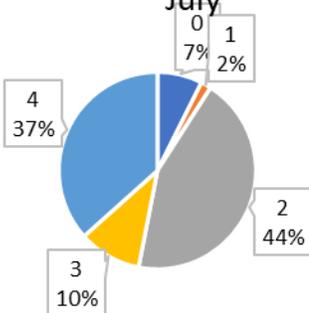
May



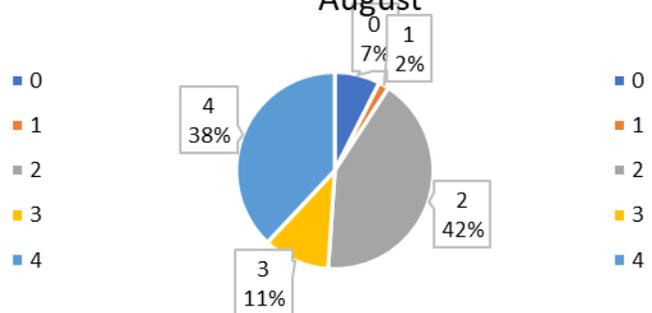
June



July



August



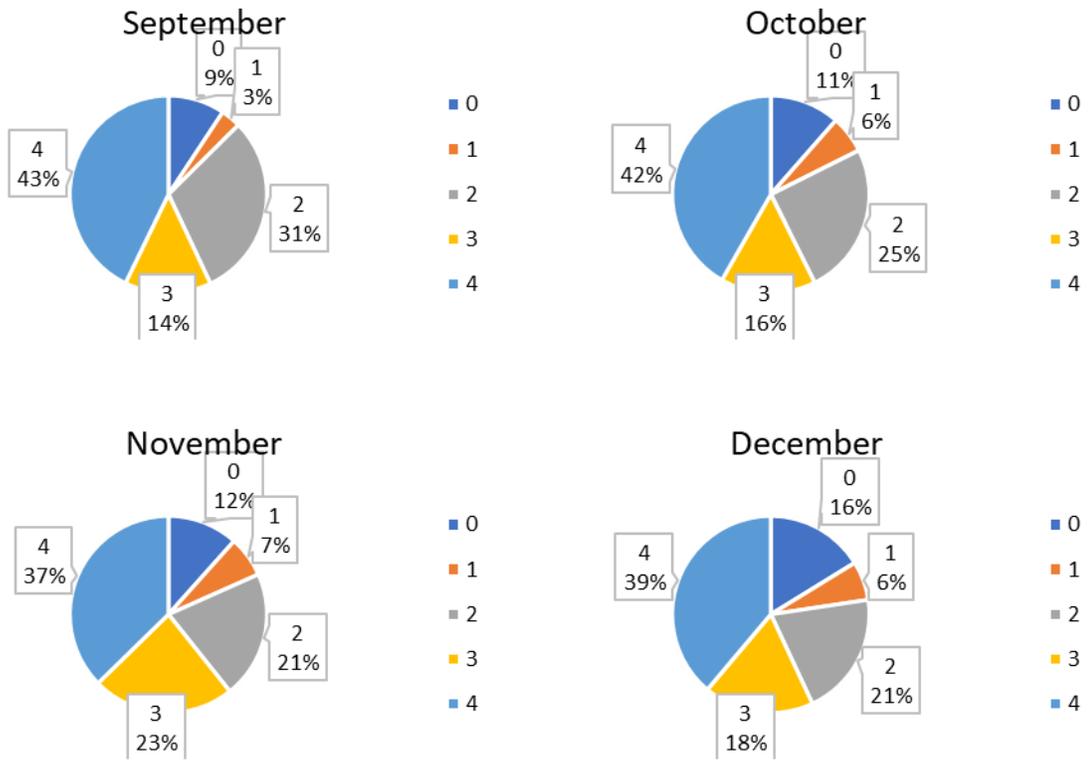


Figure A: Pie chart illustrating data distribution across five clusters based on the monthly LI and maximum rainfall. The colors indicate different clusters as shown in the legend.

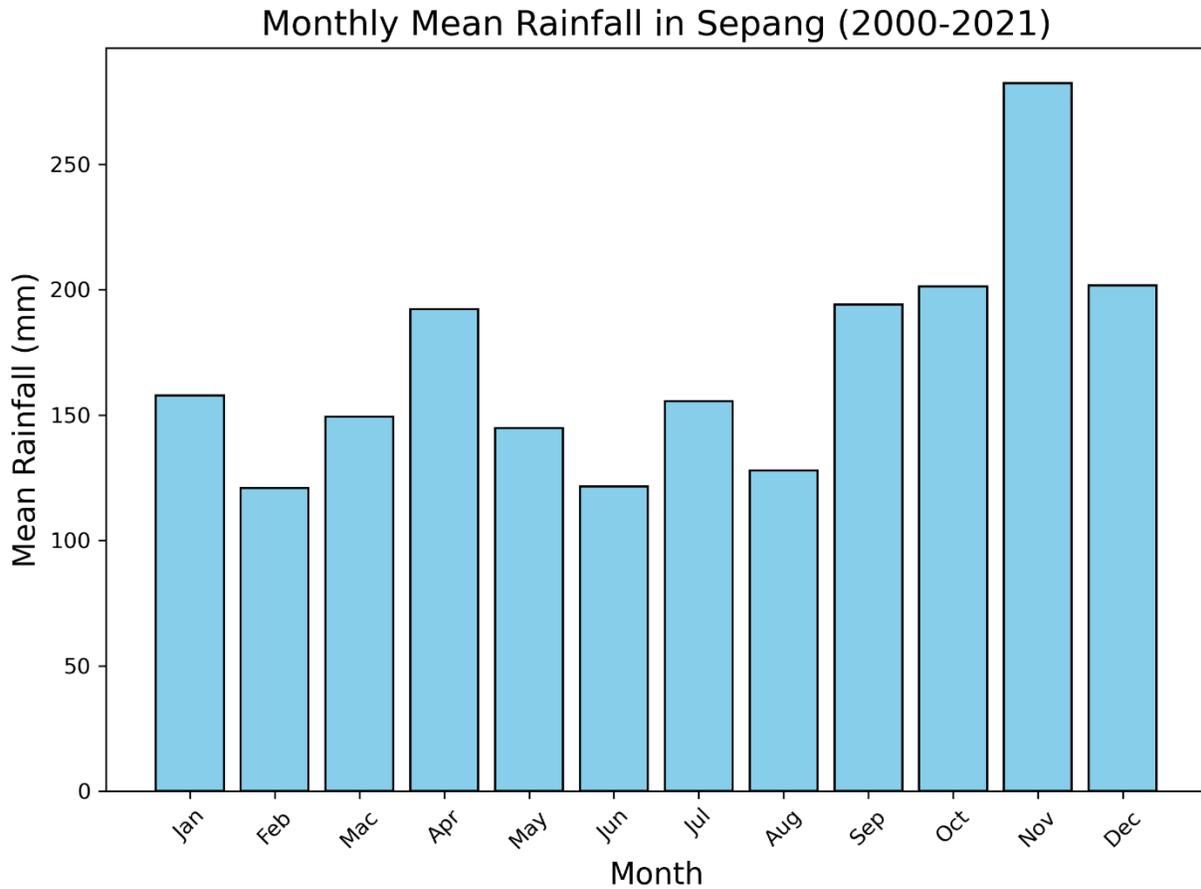


Figure B: Monthly Mean Rainfall in Sepang (2000–2021): Highlighting Seasonal Trends and Peak Rainfall Periods.

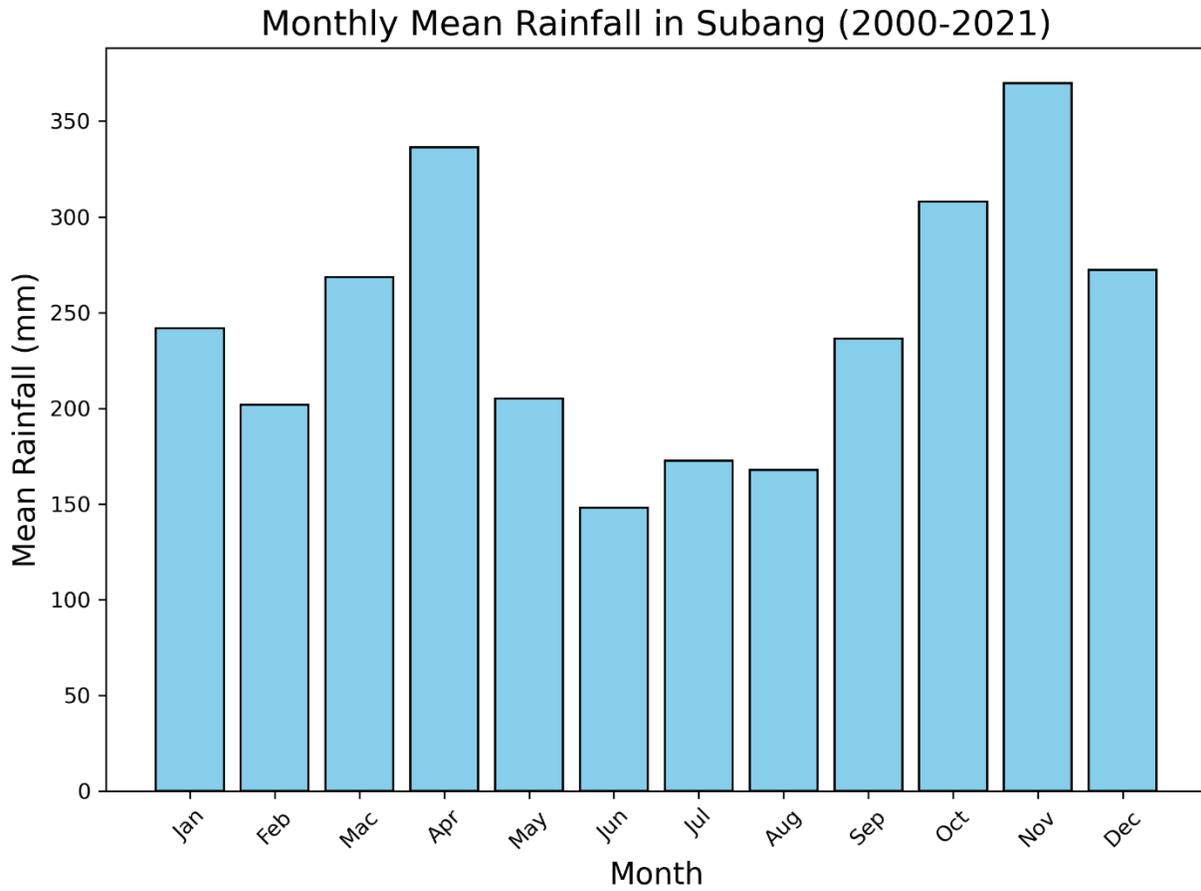


Figure C: Same as **Figure B**, except this is for Subang.

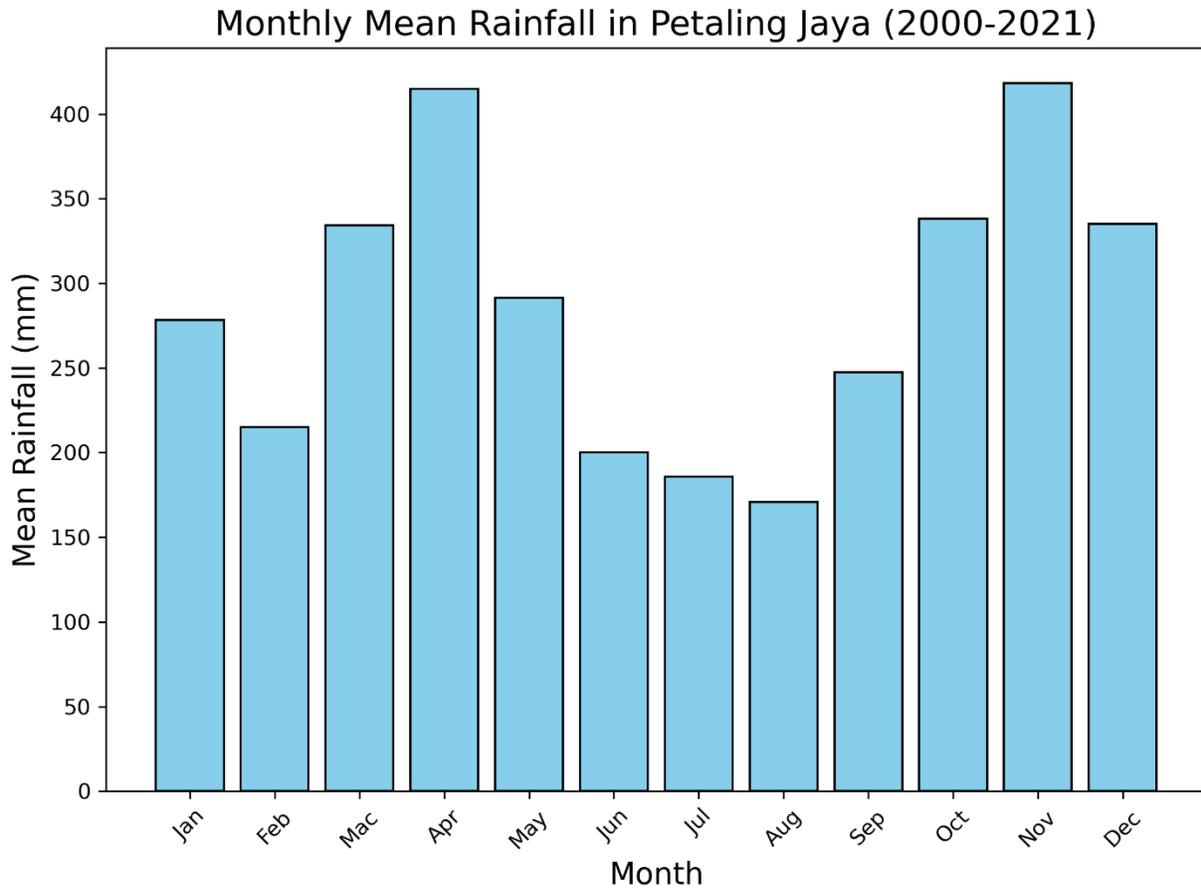


Figure D: Same as **Figure B**, except this is for Petaling Jaya.

MALAYSIAN METEOROLOGICAL DEPARTMENT

**JALAN SULTAN
46667 PETALING JAYA
SELANGOR DARUL EHSAN**

Tel : 603-79678000

Fax : 603-79550964

www.met.gov.my

ISBN 978-967-2327-20-2



9 789672 327202