---

**Article:**
Diong, J.Y., Xavier, P., Muhammad, F.A.A. et al. (Accepted: 2026) The Characteristics of Squall Lines in the Southeast Asia Region. Quarterly Journal of the Royal Meteorological Society. ISSN: 0035-9009 (In Press)

---

# The Characteristics of Squall Lines in the Southeast Asia Region.

Diong Jeong Yik[1] (diong@met.gov.my)

Prince Xavier[2] (prince.xavier@metoffice.gov.uk)

Muhammad Firdaus Ammar Abdullah[1] (kumar@met.gov.my)

Juliane Schwendike[3] (J.Schwendike@leeds.ac.uk)

[1]Research and Technical Development Division, Malaysian Meteorological Department, Petaling Jaya, Selangor, Malaysia

[2]Met Office, Exeter, UK

[3]University of Leeds, Leeds, UK

**Corresponding Author: Diong Jeong Yik,** diong@met.gov.my

The authors have no conflicts of interest to declare.

**Keywords**

Squall lines, Southeast Asia, rainfall features, squall line propagation, diurnal cycle.

**Code and Data Availability**

The squall lines database and detection algorithm are available upon request from the corresponding author.

**Funding Statement**

This work was funded by the Weather and Climate Science for Service Partnership (WCSSP) Southeast Asia. This publication for this research is funded through the Malaysian Industry-Government Group for High Technology (MIGHT Malaysia).

## Abstract

The Maritime Continent experiences active convective activity all year round, with strong diurnal rainfall variation. Squall lines are frequent in this region, which is surrounded by the warm sea, and coastlines mostly run perpendicular or parallel to the monsoon winds. This study investigates the characteristics of squall lines using Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (IMERG GPM) half-hourly precipitation data from 2001-2020 during the extended boreal summer monsoon. The squall lines are identified using the area-overlapping technique, and based on the size and rainfall thresholds. A total of 173,831 squall lines are identified, and classified into four regional clusters using K-means clustering. The clusters are Peninsular Malaysia (PM), Northwestern Borneo (NwB), Northern Borneo (NB) and South China Sea (SCS). The analysis reveals notable regional differences in propagation speed and direction, size, intensity and lifespan. Squall lines in the PM cluster are smaller and intense but have shorter lifespans, while those in the Borneo clusters are larger, of moderate intensity, and have longer lifespans. The squall line propagation speeds are only weakly linked to the background large-scale flow, indicating that the internal storm dynamics and local circulations play an important role. Composite case studies further reveal that nighttime squall lines are driven by the convergence of offshore winds and monsoonal winds. Along coastal regions, land-breeze fronts, which act as density currents, may also interact with storm-generated outflows. Overall, these findings provide important statistical insights into the structure and evolution of squall lines. The findings can support operational forecasters in anticipating squall lines' behaviour and serve as a foundation for developing regionally tailored nowcasting tools.

## 1. Introduction

The Maritime Continent has been known as the boiler box of the Tropics (Ramage 1968). The warm ocean surrounding this region and the complex distribution of islands with various orography have more convective activity than other tropical regions. Wind-terrain interaction in this region often leads to the formation of large-scale organised thunderstorms. In addition, local land-sea breeze regimes in the region also play a part in organising the thunderstorm activity. Therefore, the rainfall regime in the Maritime Continent is dominated by strong diurnal variation. The latent heat released from these large-scale thunderstorms is one of the main sources of energy driving global atmospheric circulation (Neale and Slingo 2003).

Various studies have been carried out to investigate the diurnal variation in the Maritime Continent. These studies found that convection exhibits a maximum over the large islands from late afternoon to early evening (e.g., Peatman et al. 2014). The convection maximum then shifts to the sea around the vicinity of the large islands in the early morning. Nesbitt and Zipser (2003) showed that the mesoscale convective systems (MCSs) are strongly tied to the time of the day, especially over the oceanic region. Their study found that the number of MCSs tends to increase during the nighttime in the oceanic region while the rainfall intensity for the MCSs over the land tends to peak during late evening through midnight. Houze et al. (1981) attributed the formation and propagation of the nocturnal convection in the sea in northwestern Borneo to the convergence of the land breeze and the prevailing monsoonal flow. The modelling study over the New Guinea region by Hassim et al. (2016) and the idealized modelling study on a generic island in the tropics by Coppin and Bellon (2019) further support this land breeze mechanism. However, Mapes et al. (2003) found that the land breeze is too weak to modulate the nocturnal convection in the west of Ecuador. Instead, they proposed that the offshore nocturnal convection is modulated by the gravity waves forced by the diurnal heating over the elevated land in the region. Some studies (e.g. Love et al. 2011; Yokoi et al. 2017; Peatman et al. 2023) suggested the propagating gravity waves generated by the previous convection on the land can trigger offshore convection.

A particular type of MCSs, which is the focus of this study, are squall lines. They can be defined as a narrow band of thunderstorms which last longer than a single-cell

thunderstorm, and it propagates in line. Additionally, the squall has a larger spatial extent, intense rainfall and often comes with gusting winds which could cause widespread damage to property and life compared to single-cell storms. The unique long coastlines of the large islands of the Maritime Continent, with terrains that run parallel, provide favourable conditions for the formation of squall lines. For instance, Wu et al. (2008) showed that the nocturnal convection along the coast of Western Sumatra is due to the topography and the local circulation. Bai et al. (2021) further show that this nocturnal convection results from the convergence between the offshore winds and the background low-level westerlies. Likewise, the late-night to early morning squall lines tend to form on the east coast of Sumatra or the Strait of Malacca and often propagate into Peninsular Malaysia and Singapore. These squall lines known as the “Sumatras” by the operational forecast offices in Malaysia and Singapore, can sometimes bring heavy rains to the coastal region of Peninsular Malaysia (Malaysian Meteorological Department) and Singapore (Singapore Meteorological Services). The semi-idealised numerical analysis of the Sumatras by Yi and Lim (2006) revealed that the squall in this region can be initiated by sea surface heating in the Strait of Malacca, without initial atmospheric thermal perturbations. However, the shallow convection induced by the sea surface heating can be amplified by the converging land breezes from Sumatra and Peninsular Malaysia. In contrast, Fujita et al. (2010) found that the morning precipitation peak over the Strait of Malacca is attributed to the convergence of two cold outflows produced by the previous daytime convection in Sumatra and Peninsular Malaysia. Lo and Orton (2016) studied the climatological features of the Sumatras by using 22 years (January 1988 to December 2009) of daily radar observations from Singapore. They found that the Sumatras commonly form after sunset in the Strait of Malacca and propagate toward Singapore just before midnight. The radar observations also showed that the Sumatras in Singapore tend to reach their maximum intensity at around 0300 LT (UTC+8). Fakaruddin et al. (2022) showed that the squall lines in Borneo are frequently initiated during midnight and predawn by using the 3-hourly TRMM dataset covering May to September from 1998-2018. The initiation of the squall lines is aided by the converging lower tropospheric strong westerly or southwesterly winds and the position of the trough closer to equator in the South China Sea. They further showed that the majority of the squall lines occurred in Borneo during tropical cyclone days in the northwestern Pacific due to the intensification of westerly winds in the region.

In the Malaysian region, squall lines often occur during the boreal summer monsoon from May to September, particularly in the Strait of Malacca and Borneo. The gusty winds from these squall lines can also make navigating the narrow and busy Strait of Malacca difficult. Therefore, they can increase the risk of maritime accidents such as collisions and oil spills in the strait. Additionally, when these squall lines move over land into low-lying or urban areas, the high-intensity rainfall can overwhelm the drainage system and cause flash floods. Similarly, in northeastern Borneo, the nocturnal squall can also cause damage to the coastal region. Despite the frequent occurrences and the damage caused by squall lines, there is a limited and incomplete understanding of squall lines in this region, highlighting the need for further research. Previous studies have mostly focused on individual cases of specific subregions (e.g. Fakaruddin et al. 2022; Yi and Lim 2006; Yulihastin et al. 2023), leaving the broader statistical behaviour of squall line activity across the region largely unexplored. Therefore, to address this gap, the main objective of this study is to document the statistical characteristics of the squall line activities in this region during the boreal summer monsoon. The squall lines in the region are identified using the method and data described in section 2. These squall lines are then regionalized using K-means clustering to investigate the characteristics of squall lines in each region. In section 3, the basic characteristics of the squall lines, including the average propagation speed, propagation direction, rain features, and the possible mechanisms of squall line formation are studied. Finally, the concluding remarks are provided in section 4.

## 2. Data and methods

The precipitation data from the Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (IMERG) mission dataset (Huffman et al. 2020) is used to identify the squall lines. The data is available at 30-minute intervals and have a 0.1° x 0.1° spatial resolution. The domain selected spans from 98°E to 120°E and 0° to 10°N covering the Malaysian region and the South China Sea. In this study, the results presented cover the boreal extended summer (May-October) for 20 years (2001-2020). We also use the 0.25° longitude-latitude grid hourly 10-m winds from the ERA5 reanalysis dataset (Hersbach et al. 2020) in this study. Although the ERA5 has a horizontal resolution of about 30 km, which may be insufficient to fully resolve the land-

sea breeze front, a study by Bai et al. (2021) shows that it can still capture the diurnal wind shifts effectively. Therefore, it will be used to study the large-scale mechanisms involved in the squall line formation. The orography data used in this study is obtained from Global Land Data Assimilation System (GLDAS) (Rodell et al. 2004), which is based on the Global 30 Arc Second Elevation Dataset (GTOPO30). The elevation dataset is averaged up to the GLDAS resolution at 0.25°.

To examine the evolution of the squall lines at the local solar time (LST), the local offset time, $t_\lambda$ is first calculated. This local time offset accounts for the difference between the Coordinated Universal Time (UTC) and the solar time at a given longitude ($\lambda$). Since the earth rotates about 15 degrees in an hour, the $t_\lambda$ at the given longitude is calculated as in Equation (1). The $t_\lambda$ is then added to the timestamp of the IMERG data to convert the data to LST (Equation (2)).

$$t_\lambda = \frac{\lambda}{15} \qquad (1)$$

$$LST = UTC + t_\lambda \qquad (2)$$

A previous study by Fakaruddin et al. (2022) showed that more than half of the squall lines occurred with the presence of tropical cyclones (TCs) in the Western Pacific. To investigate the effect of the TCs, we use the International Best Track Archive for Climate Stewardship data (IBTrACS; Knapp et al. 2010). In this study, instead of covering the storms in the entire Western Pacific region, we limit the storms in the Western Pacific region to an area bounded between 98.5°-130.5°E and between 0°-20°N.

### 2.1 Squall identification methods

The squall lines are identified and tracked using half-hourly data with the area overlapping technique described by Williams and Houze (1987). To determine a potential squall line, the rainfall is first filtered to meet a minimum threshold of 5 mm $h^{-1}$. Next, the size of the rainband must be at least 1000 $km^2$. Each rainband that satisfies the two criteria is then tagged as a cluster element, CE. The tracking algorithm uses the area overlapping technique to determine the group of cluster elements. The identified cluster element at time t, $CE_t$ that overlaps at least 50% with the cluster

element at time t+1($CE_{t+1}$) and t+2 ($CE_{t+2}$) is then assigned to the same identity (ID). To ensure that only long-lived, organised systems are retained, we further require each squall line to persist for at least two hours. This threshold excludes short-lived or transient rainbands and retains systems with sustained linear structure and focuses the analysis on systems that have developed sustained mesoscale organisation. From the CE IDs, the shape of each CE is determined by the eccentricity. Considering that the shape of CEs may evolve, the potential squalls are determined when the CEs must have the shape of an ellipse with an eccentricity greater than 0.8 for more than 50% of their tracked lifetime. In our study, a squall line is identified from the pool of potential squall lines that have reached the threshold area of 5000 $km^2$ for at least once in their tracked lifetime.

An example of an identified squall in the Strait of Malacca, which is detected using the method as prescribed above, is shown in Figure 1. This squall line occurred on 31 October 2020 and was tagged with a particular identity (ID) at 1900 UTC. The black outline shows the area of the squall. The blue X indicates the centroid for this squall line. At 1930 UTC, this tagged squall line split into two smaller CEs as shown by two outline regions with two separate centroids (red colour X). By 2000 UTC, rainfall in the southern CE weakens and is no longer detected as a squall line. Only the northern CE remains as a squall line with a new centroid (green colour X) located slightly north of its previous centroid.

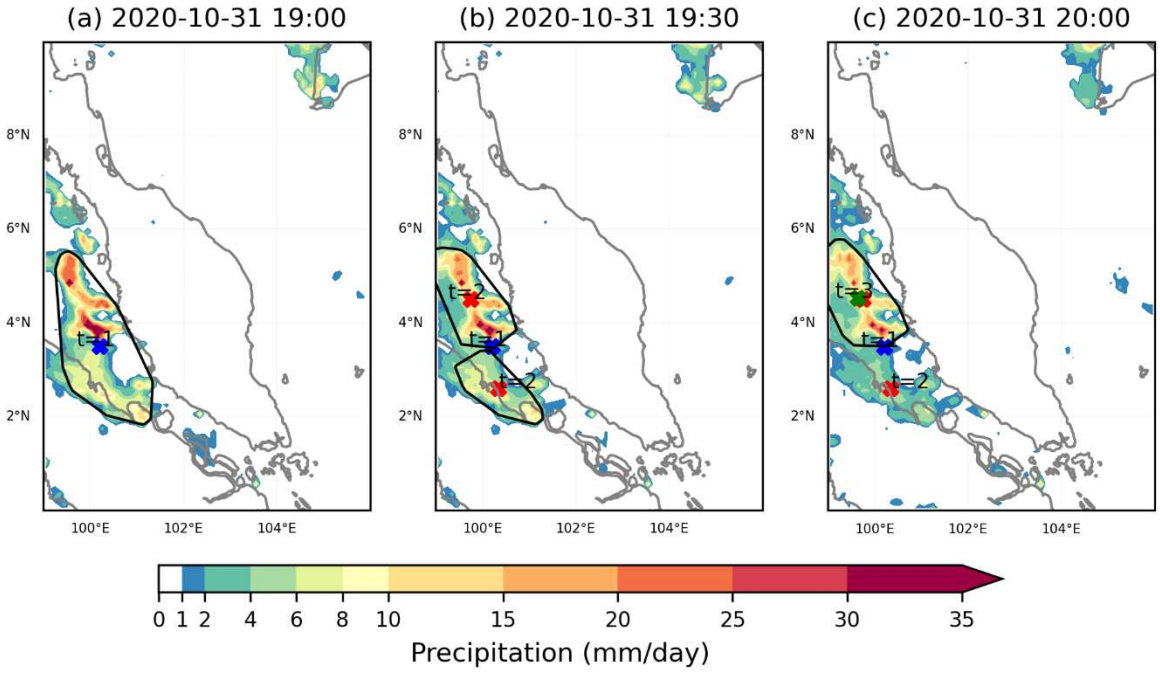

Figure 1: GPM rainfall and the black outline of an identified squall line tagged with the same ID from (a) 1900UTC to (c) 2000UTC on 31 October 2020 at half-hourly intervals. Centroids are given in the blue X (t=1), red X (t=2) and green X (t=3).

### 2.2 Clustering the squall lines

Using the squall line identification method described in Section 2.2, a total of 499,019 rainbands with straight line features with an area size of at least 1000 $km^2$ were detected from the 20 years of May to October GPM half-hourly data. Out of these, 173,831 were identified as squall lines, and 18,672 initial squall line centroids were identified. The initial squall line centroids are the centroids of each squall line when they are first detected and will be used to study the possible mechanisms for squall line formation. Figure 2 shows the heatmaps of the initial centroid (left column) and the mosaic plots of the convective areas of the initial squall lines (right column). The centroid frequency was computed on a 1° grid to highlight broad spatial patterns, whereas the rainfall areas mosaic frequency was on a 0.1° grid, which corresponds to the native resolution of the GPM dataset. Overall, the spatial distribution of the mosaic frequencies closely matches that of the initial centroids heatmap, but with approximately twice the magnitude. The larger frequency magnitude, together with a matching spatial pattern in the mosaic heatmap, reflects the typical size of the squall lines and supports the validity of the centroid-based detection. High spatial density of initial squall line centroids is found around the coastal region east of Sumatra, the Strait of Malacca, and Northern and Northwestern Borneo, which suggests the presence of underlying spatial clusters. We therefore perform a K-means clustering to identify and analyse these clusters. For all initial centroids, k=4 is determined using the elbow method. The centroids are clustered into Peninsular Malaysia (PM), Northern Borneo (NB), South China Sea (SCS) and Northwestern Borneo (NwB). This subdivision is not only driven by computational optimisation but also reflects meaningful meteorological contrasts across the region. In particular, NB is more frequently influenced by tropical cyclones and associated rainbands originating near the Philippines. In contrast, the NwB experiences frequent coastal squall lines driven by the southwesterly flow during the Southwest Monsoon (Fakaruddin et al. 2022). In addition, the northwest–southeast tilted convective structure associated with the

Boreal Summer Intraseasonal Oscillation (BSISO) affects these subregions differently, resulting in contrasting rainfall variability that would be difficult to resolve under a single Borneo domain.

The Peninsular Malaysia cluster accounts for about 26% of the initial squall lines detected in the region. It shows discernible evidence of Sumatra squall lines along the Strait of Malacca (Figure 2c & d). In Northern Borneo, the squall line activity is particularly high in the mountain ranges and along the coastal areas (Figure 2e & f). The squall lines in the Northern Borneo cluster constitute 20% of the total initial squall lines detected in the region. About 26% of the total initial squall lines detected in the region fall in the SCS cluster (Figure 2g). In this cluster, the spatial occurrence of squall lines is evenly distributed without the influence of the surrounding terrain and surface roughness. The spatial coverage frequency (Figure 2h), however, is slightly concentrated in the middle of the cluster and may reflect the large-scale monsoonal convergence in this region during the southwest monsoon. This pattern is consistent with the previous studies linking convective maxima in this region to monsoon trough activity and intraseasonal variability (e.g Xavier et al. 2024). In the NwB cluster (Figure 2i & j), distinct squall line activity can be observed near the coastal and interior regions in the NwB cluster. The concentration of the squall lines initiation along the coast is consistent with the findings of Fakaruddin et al. (2022). This NwB cluster contributed approximately 28% of the total initial squall lines.

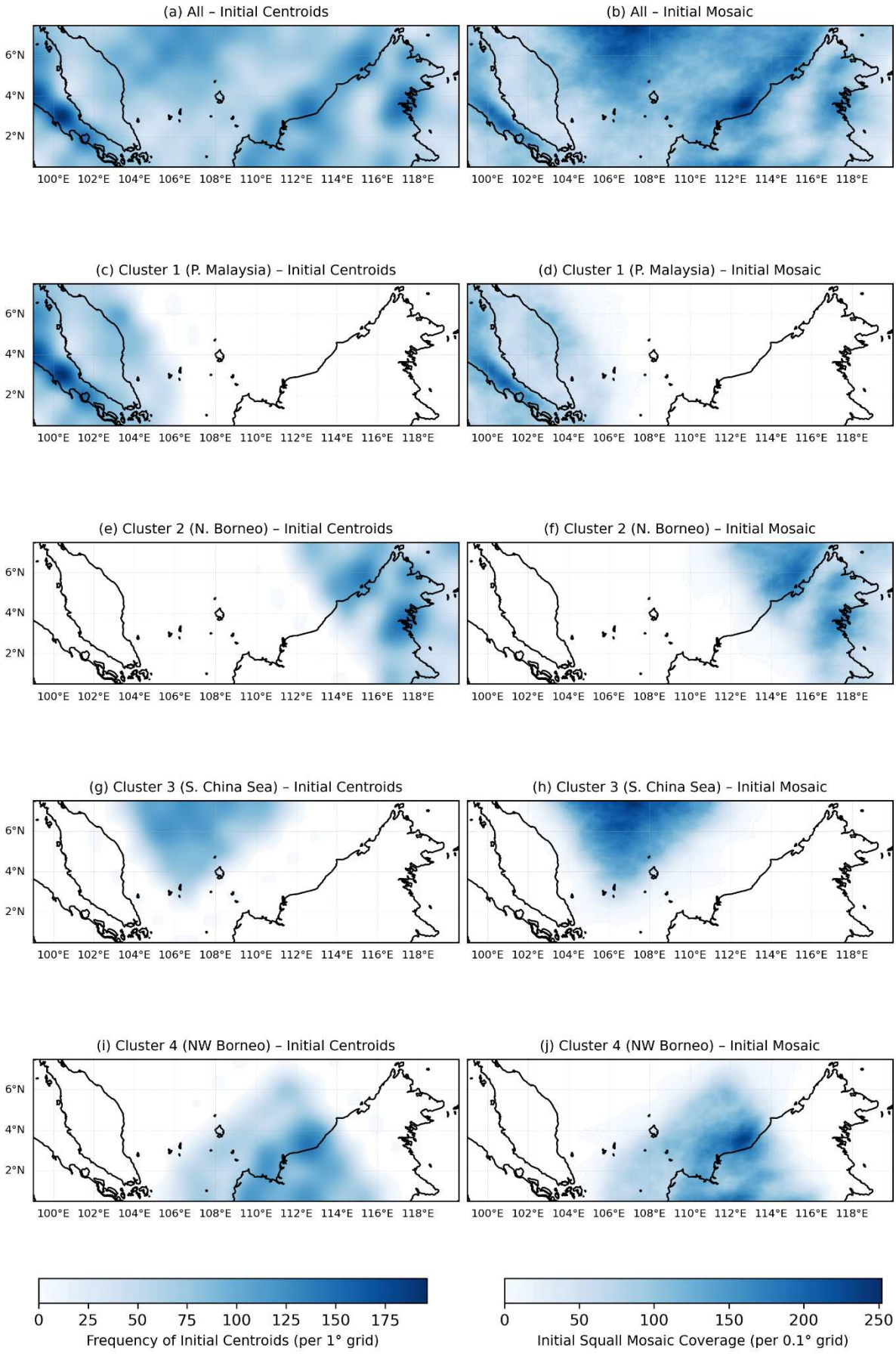
(a) All – Initial Centroids
(b) All – Initial Mosaic
(c) Cluster 1 (P. Malaysia) – Initial Centroids
(d) Cluster 1 (P. Malaysia) – Initial Mosaic
(e) Cluster 2 (N. Borneo) – Initial Centroids
(f) Cluster 2 (N. Borneo) – Initial Mosaic
(g) Cluster 3 (S. China Sea) – Initial Centroids
(h) Cluster 3 (S. China Sea) – Initial Mosaic
(i) Cluster 4 (NW Borneo) – Initial Centroids
(j) Cluster 4 (NW Borneo) – Initial Mosaic
6°N
4°N
2°N
100°E 102°E 104°E 106°E 108°E 110°E 112°E 114°E 116°E 118°E
0 25 50 75 100 125 150 175
Frequency of Initial Centroids (per 1° grid)
0 50 100 150 200 250
Initial Squall Mosaic Coverage (per 0.1° grid)

Figure 2: Initial squall line centroids and mosaics for all and in each clusters. Left column: heatmaps of initial centroids. Right column: initial squall line mosaics showing the spatial extent of rainfall at the time of formation. (a-b), All squall lines. (c-d) Cluster 1 (Peninsular Malaysia), (e-f) Cluster 2 (Northern Borneo), (g-h) Cluster 3 (South China Sea), (i-j) Cluster 4 (Northwestern Borneo).

## 3. Results and discussion

### 3.1 Average propagation speed and direction

Figure 3a shows the distributions of the average propagation speed of the squall lines in each cluster using the Gaussian kernel density estimation, with the bandwidth determined automatically following Scott's rule (Scott 1992). Before analysis, a small number of squall line cases (about 0.5% of the dataset) with propagation speed exceeding 21 m $s^{-1}$ (but less than 30 m $s^{-1}$) were removed to focus on the representative propagation range of tropical squall lines. After this filtering, the maximum speed reported in PM cluster is 19.4 m $s^{-1}$ while the rest of the clusters have a maximum speed of around 20.4 m $s^{-1}$. The average propagation speed of each squall line in the different regions is calculated by taking the difference between the final and initial positions of the centroid divided by the lifetime. Each cluster possesses a different distribution of propagation speed. From the cluster speed distributions, the SCS squall lines tend to propagate faster than squalls from other clusters, with most of them having a propagation speed of between 4 m $s^{-1}$ and 6 m $s^{-1}$ (green line). In contrast, the other clusters' speed distribution modes are slightly lower than the SCS and lie between 2 m $s^{-1}$ and 5 m $s^{-1}$. The majority of the squall lines in the Borneo clusters propagate at the speed of between 3 m $s^{-1}$ and 4 m $s^{-1}$ while the majority of the squall lines in the PM cluster propagate slightly faster at the rate of 4 m $s^{-1}$ to 5 m $s^{-1}$. The speed distribution mode for all clusters is almost similar to the convection propagation speed in Sumatra as noted in the previous studies using observational data (e.g., Wu et al. 2008; Yokoi et al. 2017; Bai et al. 2021) and the simulated data from the model (Love et al. 2011). Several explanations have been proposed for why tropical squall lines propagate more slowly than their mid-latitude counterparts. Wu et al. (2008) postulated that this slower-than-mid-latitude squall line speed may be due to the moist environment in the tropics, while Bai et al. (2021) argued that the

converging land breeze and monsoonal flow near the coast are responsible for the slow propagation.

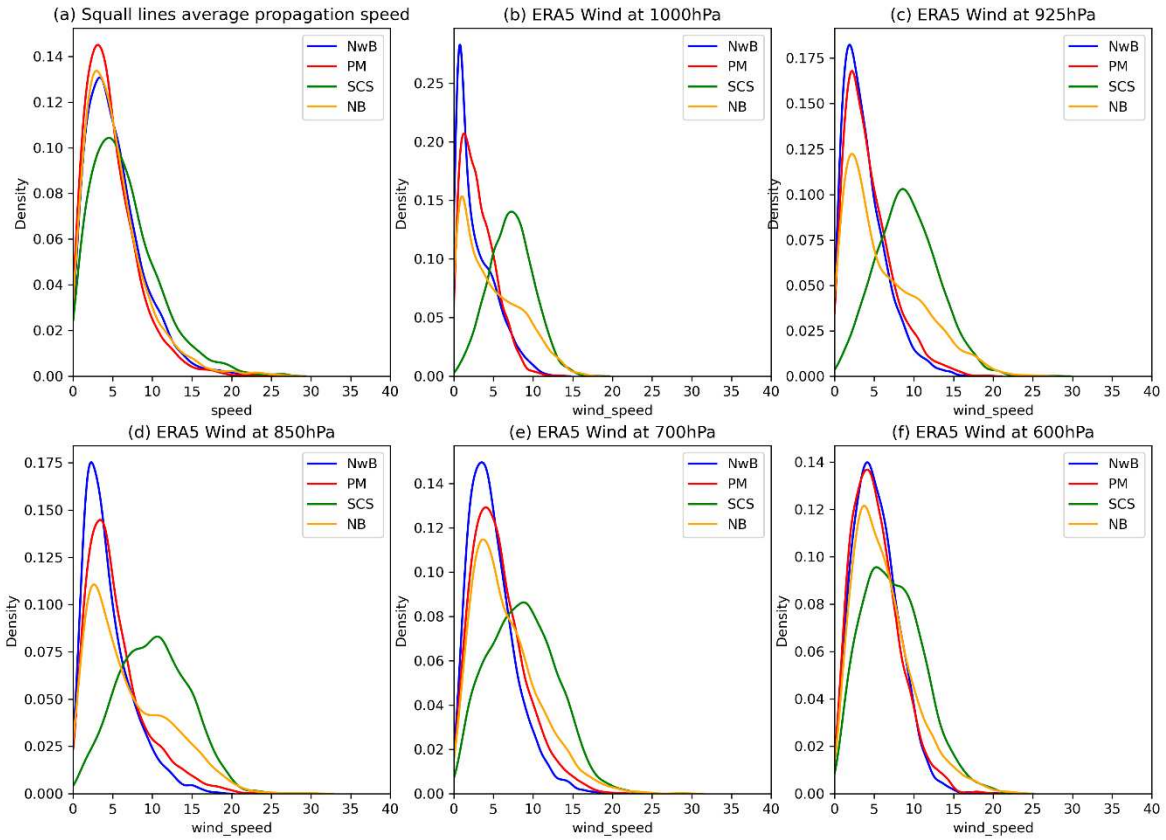


Figure 3: Distribution of average propagation speed of squall lines and ERA5 wind speeds at various levels for each cluster using kernel density estimation. (a) The distribution of average propagation speed. The distribution of ERA5 wind speeds at (b) 1000 hPa, (c) 925 hPa, (d) 850 hPa, (e) 700 hPa, and (f) 600 hPa.

Squall lines in the Maritime Continent can propagate in multiple directions. Table 1 summarizes the mean average propagation speeds of squall lines for various directions. Notably, zonally propagating squall lines are generally faster than the meridionally propagating squall lines across all clusters. The northward and southward moving squall lines display the smallest mean average propagation speeds, roughly around 4 m $s^{-1}$. The faster zonal propagation is consistent with stronger zonal low-level winds in the region. The results also indicate that the eastward-moving squall lines generally have higher mean average propagation speeds than westward-moving squall lines, suggesting monsoonal westerlies may contribute to this asymmetry.

Across all clusters, squall lines propagate eastward at speeds ranging from 5.4 to 7.8 m $s^{-1}$, the highest mean average propagation speed within each cluster. The only exception is the NwB cluster, where the westward-moving squall lines exhibit the highest mean average propagation speed. Among all the clusters, the eastward propagating squall lines in the SCS cluster display the highest mean average propagation speed, averaging 7.8 m $s^{-1}$. During the occurrence of TCs in the vicinity of the South China Sea and its surrounding, squall lines may be embedded within the outer spiral rainbands of the TCs. As a result, the squall lines' propagation speed may be influenced by the translation speed of the TCs. To examine the effect of TCs on the average propagation speeds, squall lines occurring during TC days located within the TC region of this study were identified. The influence is most pronounced in the SCS and Borneo clusters (see Supplementary Figure S1). Within these clusters, the mean average propagation speed increases during TC days for the northeastward, eastward and southeastward directions, with the largest increase observed for the SCS cluster. This increase in the northeastward and eastward direction mean average propagation speed could be attributed to the intensification of low-level westerly or southwesterly winds in the SCS region on the TC days. The propagation speed, however, decreases in the southward-moving squall lines in the Borneo clusters. In the PM cluster, the difference in mean average propagation speed between TC days and non-TC days is relatively small across all eastward propagation directions. In the NB cluster, the mean average propagation speeds for southeastward, northwestward and southward squall lines are slower during the TC days compared to non-TC days. These contrasting responses indicate that TC-induced changes in low-level winds can either enhance or inhibit squall line propagation depending on direction.

To investigate the frequency of the squall line propagation direction and the average propagation speed for each cluster, Figure 4 shows a wind rose chart where the length of each spoke represents the frequency, and each spoke indicates squall lines moving from that direction. For the PM cluster (Figure 4a), about 38% of the squall lines in this cluster propagate eastward (in this analysis, we are considering the propagation direction from northeast and southeast), with most of them propagating at a speed of between 3 and 6 m $s^{-1}$. However, a considerable number of the squall lines (about 25%) propagate westward, also with a similar propagation speed. In the NB cluster (Figure 4b), the squall line exhibits characteristics similar to those of the

PM cluster. In the SCS cluster (Figure 4c), squall lines frequently propagate eastward (about 56%), with most of them also propagating at speeds between 6 and 9 m $s^{-1}$. This cluster also shows a clear separation between the faster eastward propagation and the less frequent meridional directions, which further highlights the directional dominance shown in Table 1. The northward and southward propagating squall lines in the SCS cluster are infrequent. Meanwhile, for the NwB cluster (Figure 4d), most squall lines come from the east (moving westward). They account for about 40% of all the squall line propagation in this cluster. About 70% of these squall lines travel at a propagation speed of less than 15 m $s^{-1}$.

Table 1: Mean average propagation speed ($ms^{-1}$) of squall line for all clusters in different directions.

| | P. Malaysia | N. Borneo | S. China Sea | NW. Borneo |
|---|---|---|---|---|
| Direction | Mean propagation speed (m $s^{-1}$) | Mean propagation speed (m $s^{-1}$) | Mean propagation speed (m $s^{-1}$) | Mean propagation speed (m $s^{-1}$) |
| Northward | 3.9 | 4.2 | 4.5 | 4.1 |
| N'eastward | 4.6 | 4.4 | 5.7 | 4.5 |
| Eastward | 5.4 | 6.0 | 7.8 | 5.4 |
| S'eastward | 4.6 | 5.4 | 6.4 | 5.4 |
| Southward | 4.4 | 3.9 | 4.3 | 4.2 |
| S'westward | 4.6 | 4.2 | 4.8 | 4.6 |
| Westward | 4.9 | 5.4 | 5.9 | 6.2 |
| N'westward | 4.3 | 5.3 | 5.5 | 5.3 |

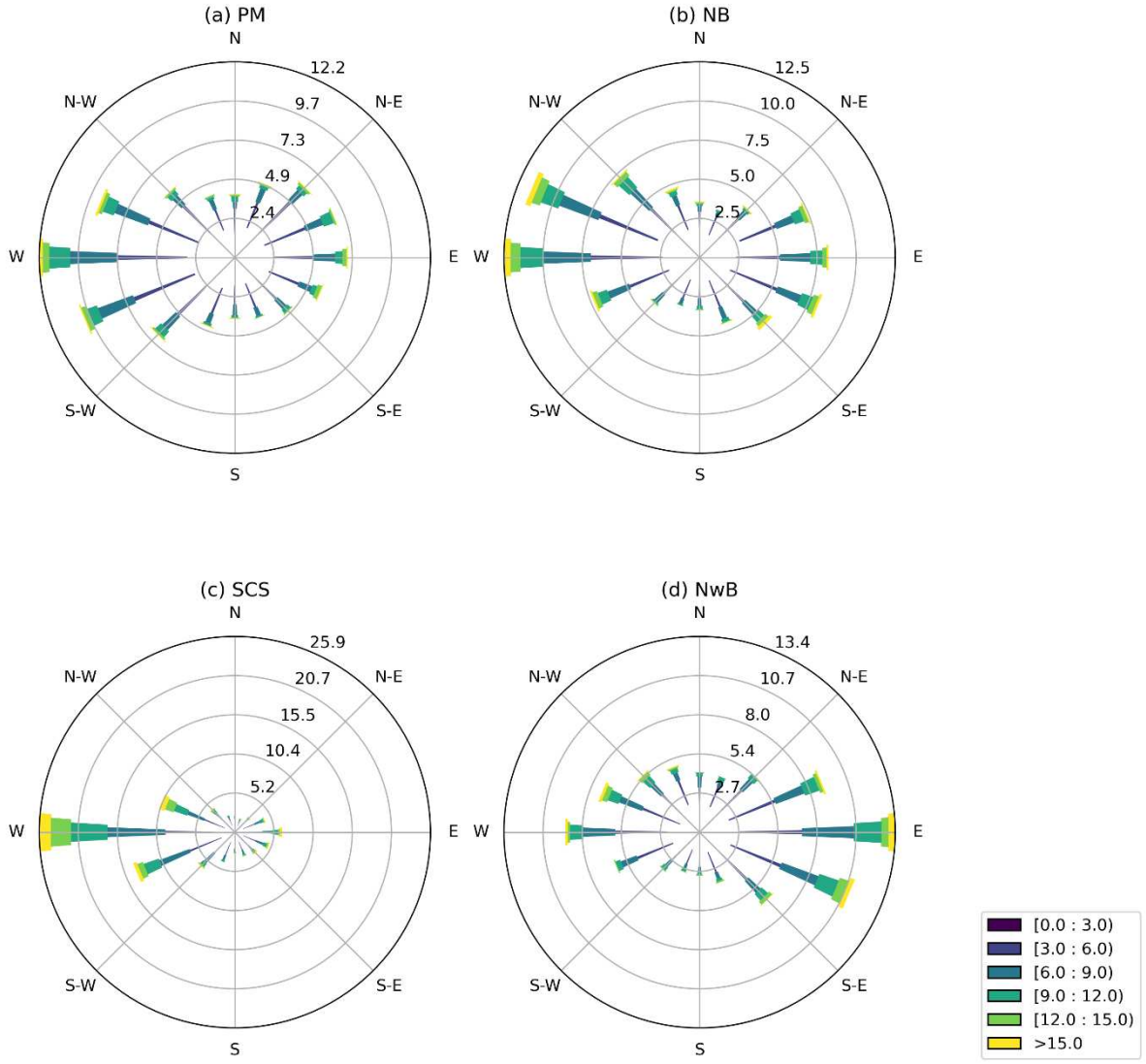


Figure 4: The 'wind rose' chart for squall lines in (a) Peninsular Malaysia (PM), (b) Northern Borneo (NB), (c) South China Sea (SCS) and (d) Northwestern Borneo (NwB) . The frequency of the propagation direction is represented by the length of the spoke and the average propagation speed is represented by colour shading. Note that the direction here indicates where the squall lines propagate from.

We investigate the large-scale circulation influence on the squall lines' average propagation speed, focusing on lower tropospheric wind speeds at 1000 hPa, 925 hPa, 850 hPa, 700 hPa and 600 hPa (see Figure 3b-f) from ERA5. To do so, we extract

ERA5 winds at each squall line's centroid. We also tested averaging the ERA5 winds over cluster-specific domains, but the results proved insensitive to the choice of the sampling method. Overall, the distribution of wind speeds shows that low-level winds are strongest in the SCS, with a modal value of around 6-11 m $s^{-1}$, but decrease above 700 hPa level. Across other clusters, the wind speed distributions are broadly similar and increasing with height, from the modal values of around 1-3 m $s^{-1}$ at 1000 hPa to aound 3-5 m $s^{-1}$. In these clusters, the slight increase in the wind speed at higher levels is expected, as the winds are no longer influenced by orography. Although these wind speed distributions resemble the squall lines' average propagation speed distribution, Spearman correlation coefficients between the wind speeds at each level and the average propagation speed are weak, despite being statistically significant (see Table 2). Thus, this suggests that the large-scale flow has little influence on the squall lines propagation speed. We next investigate whether the large-scale flow has any influence on the squall line propagation direction. To do so, we focus on the winds at 850 hPa and 700 hPa because the correlation with the propagation speeds is highest at these levels. The results are summarised in Table 3. Our analysis is restricted to just eastward and westward propagation since our previous result suggests that these are the predominant propagation directions. Here, the westerly winds include northwesterly, westerly and southwesterly flow, while easterly winds include the northeasterly, easterly and southeasterly flow. We find that most eastward propagating squall lines are embedded in the westerly monsoonal flow, while a smaller fraction propagates eastward under the influence of large-scale easterly winds. Given that the large-scale winds are predominantly westerly during the boreal summer months, many westward propagating squall lines are likewise embedded in westerly flow. However, the number of squall lines propagating westward under the easterly flows is higher compared to the westerly flows. The presence of a substantial fraction of squall lines propagating against the prevailing winds further indicates that the large-scale flows have limited influence on both their direction and speed. Instead, propagation exhibits distinct directional variability, including frequent cases of against-flow, and the correspondence with environmental flow appears to be confined to the lowest layers rather than extending uniformly throughout the monsoonal depth, as shown later using layer-dependent shear-propagation correlations.

399 Table 2: Spearman correlation coefficients between the low-level tropospheric winds
400 and the average propagation speed of squall lines in each cluster. * indicates
401 statistically significant at 99$^{th}$ percentile

| | P Malaysia | N. Borneo | S. China Sea | NW. Borneo |
|---|---|---|---|---|
| 1000 hPa | -0.006 | 0.109* | 0.263* | 0.072* |
| 925 hPa | 0.039 | 0.130* | 0.258* | 0.085* |
| 850 hPa | 0.083* | 0.145* | 0.300* | 0.100* |
| 700 hPa | 0.140* | 0.168* | 0.343* | 0.165* |
| 600 hPa | 0.082* | 0.151* | 0.281* | 0.150* |

402

403 Table 3: Percentage of squall lines propagating in different zonal directions under
404 varying ERA5 zonal wind components. The top portion shows the percentage under
405 different 850 hPa zonal wind components, while the bottom portion (in italic) shows
406 those at 700 hPa. The large-scale easterly component includes northeasterly, easterly
407 and southeasterly winds. Likewise, the large-scale westerly component includes
408 northwesterly, westerly and southwesterly winds.

| | Eastward propagation | | Westward propagation | |
|---|---|---|---|---|
| Cluster | Large-scale westerly component (%) | Large-scale easterly component (%) | Large-scale westerly component (%) | Large-scale easterly component (%) |
| P. Malaysia | 81<br>*83* | *7*<br>*8* | 55<br>*49* | 25<br>*37* |
| N. Borneo | 88<br>*85* | 6<br>*9* | 61<br>*43* | 28<br>*47* |
| S. China Sea | 97<br>*96* | 1<br>*3* | 63<br>*50* | 31<br>*43* |
| NW. Borneo | 79<br>*68* | 9<br>*21* | 55<br>*26* | 27<br>*60* |

409

The directional independence implies an important role for internal storm dynamics. To further understand this underlying cause, we examine the mechanism of squall line propagation using the Rotunno-Klemp-Weisman (RKW) theory (Rotunno et al. 1988). This theory explains how the balance between low-level (the lowest 2.5 km of the atmosphere) environmental shear and the cold pool influences the structure of the squall line. The cold pool acts as a gravity current whose leading edge continuously lifts warm, moist air ahead and thus drives the propagation. Since the cold pools are difficult to detect in coarse-resolution datasets such as ERA5, we use the directional correlation between environmental vertical shear and squall lines propagation as a statistical proxy to assess the RKW balance. Here, we emphasise that this approach assesses the directional implication of RKW theory rather than the full quantitative balance between cold pool strength and shear magnitude, which cannot be evaluated using ERA5 reanalysis data. Because both the vertical shear and squall lines propagation are directional data bounded between 0° to 360°, conventional linear correlation may not be appropriate. Instead, we use the circular-circular correlation (Jammalamadaka and Sarma 1988) to quantify their association while preserving the cyclic nature of the data. Taking into consideration the variability in the direction, a ± 30° tolerance window is applied to define 'directional agreement'. Data pairs with angular differences smaller than 30° are weighted proportionally to their proximity using the cosine function. In other words, pairs with small angular differences contribute more to the correlation, whereas pairs with angular differences greater than 30° are assigned zero weight. In this study, the vertical shear direction is defined as the direction from which the shear originates and the squall line propagation is defined as the direction towards which the system moves. Consequently, positive correlation indicates the squall lines are propagating upshear, consistent with the subcritical, cold pool-dominated regime prescribed by RKW theory. Since ERA-5 is provided on pressure levels rather than terrain-following coordinates, and given the complex topography in the region, Figure 5 shows the heatmap of correlation between the various environmental vertical shear layers and the squall line propagation directions. These various vertical layers allow, to some extent, a realistic representation of low-level shear in the region. The results indicate that squall line propagation is most strongly linked to near-surface shear, with correlation values exceeding +0.9 in the 1000–925 hPa and 925–850 hPa layers, while the correspondence weakens in deeper monsoonal layers (e.g. up to 700 hPa). This vertical preference, together with the

prevalence of upshear propagation, is qualitatively consistent with the RKW framework and suggests a subcritical regime in which cold-pool outflows exert a dominant influence on squall lines. Nonetheless, this interpretation is suggestive rather than conclusive, since ERA5 cannot explicitly resolve cold pool processes. A detailed analysis of this cold pool would require convection-permitting simulations (e.g. Peatman et al. 2023), but such simulations are computationally demanding and beyond the scope of this study.

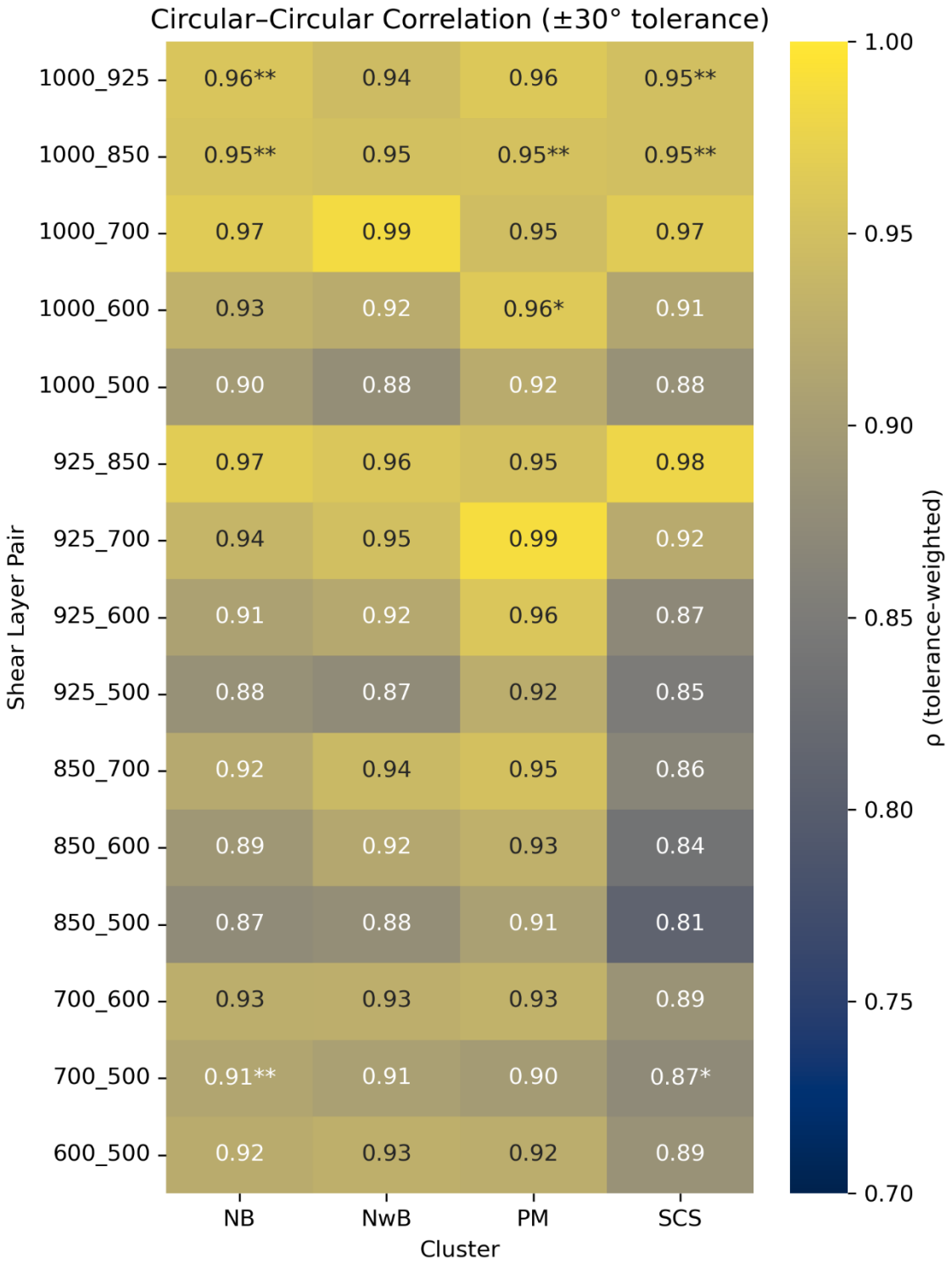


Figure 5: Circular–circular correlation (ρ) between squall line propagation and environmental low-level vertical shear for different shear-layer pairs and regional clusters, computed with a ±30° tolerance window. Asterisks denote statistical significance (*p < 0.05, **p < 0.01). Statistical significance is assessed using 10,000 permutation tests to derive empirical p-values.

### 3.2 Squall lines Rainfall Features

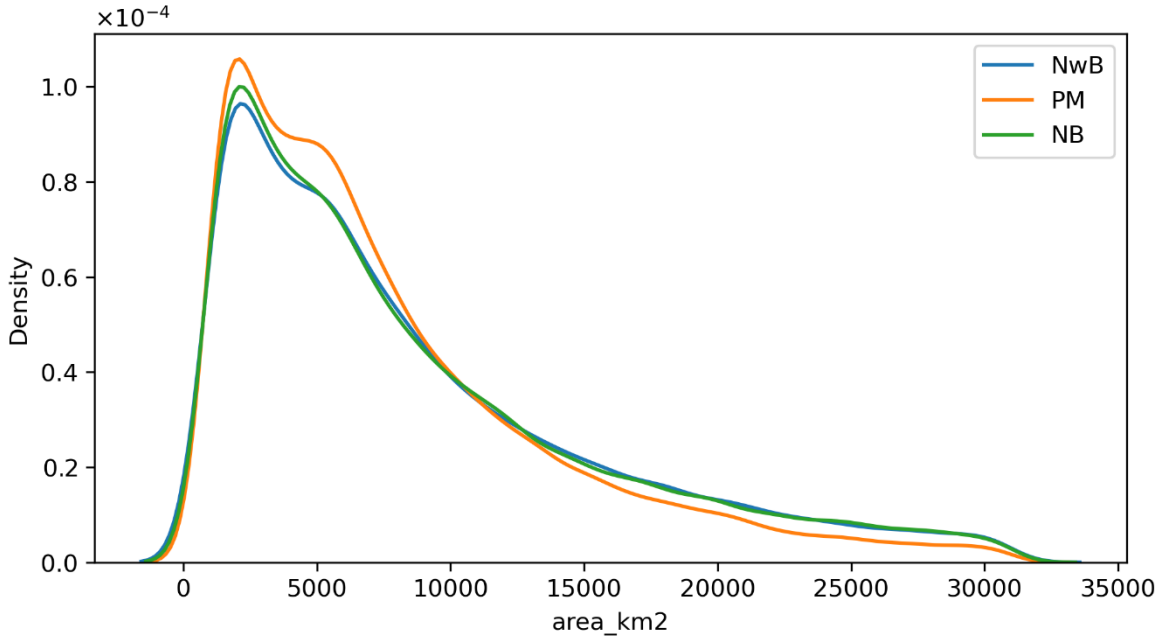


Figure 6: Distribution of the squall area for each cluster using kernel density estimation.

The distribution of the squall line sizes for each cluster is shown in Figure 6. In this analysis, we omitted the squall lines from the SCS cluster because our region does not encompass the entire SCS region and the squalls detected may be part of larger squalls. One obvious observation is that the distribution mode for the region's squall line sizes ranges between 2,000 and 5,000 km². The distributions also indicate that fewer squall lines in PM exceed an area of 10,000 $km^2$ compared to those in Borneo. To investigate the relationship between the size of the squall line and the rainfall features in the squall line, the area, mean and maximum rain rate inside each half-hourly squall line polygon is calculated. The analysis indicates that the joint probability distribution of mean rainfall and maximum rainfall with the squall line size for each cluster is shown in Figure 6. The squall lines of 1,000 - 6,000 $km^2$ are quite common in this region, indicating that they frequently occur in this size range (Figure 7a-c). These squall lines typically produce rainfall with a mean rate of 5 to 6 mm $h^{-1}$. Across all clusters, there is a general trend of increasing mean rain rate as the size of the squall line increases. However, the mean rain rate seldom exceeds 7 mm $h^{-1}$,

indicating a very low probability of encountering a large-sized squall line that produces mean rainfall exceeding this amount in this region. For squall lines with sizes ranging between 2,000-12,000 km², there is a high probability of observing squall lines with higher mean rain rates, though these values generally do not exceed 9 mm $h^{-1}$. This pattern shows that while larger squall lines are associated with higher mean rain rates, there are upper limits to which the intensity can be expected. In short, this shows that no matter how large the squall line is, the intensity does not increase in proportion to its size. While some patterns are consistent across the cluster, distinct regional differences exist in size-intensity relationships. In the PM cluster (Figure 7b), there is a higher likelihood of intense rainfall within a narrower squall line size range compared to squall lines in Borneo clusters (Figure 7a and c), and the intensity is highest among the three clusters. Additionally, in the NwB cluster (Figure 7a), rainfall intensity for a given squall line size is typically greater than that in the NB cluster (Figure 7c**)**. In contrast, the joint probability distribution of maximum rainfall with the squall line size for NB (Figure 7f**)** is comparable to NwB (Figure 7d). In both clusters, squall lines within the size range of 1,000 to 6,000 $km^2$ with maximum rainfall between 8 and 15 mm $h^{-1}$ are notably frequent. Conversely, the PM cluster (Figure 7e**)** typically features smaller squalls, ranging from 1,000 to 4,000 km², with maximum rainfall rates between 8 and 15 mm $h^{-1}$. When examining extreme rainfall events, specifically those exceeding 30 mm $h^{-1}$, a notable distinction emerges between the three clusters. The PM cluster shows the greatest likelihood of such extreme events in squall lines between 2,000 and 12,000 $km^2$, suggesting that squall lines in PM within this size are more capable of generating severe rainfall compared to the squalls in the Borneo clusters with a similar size.

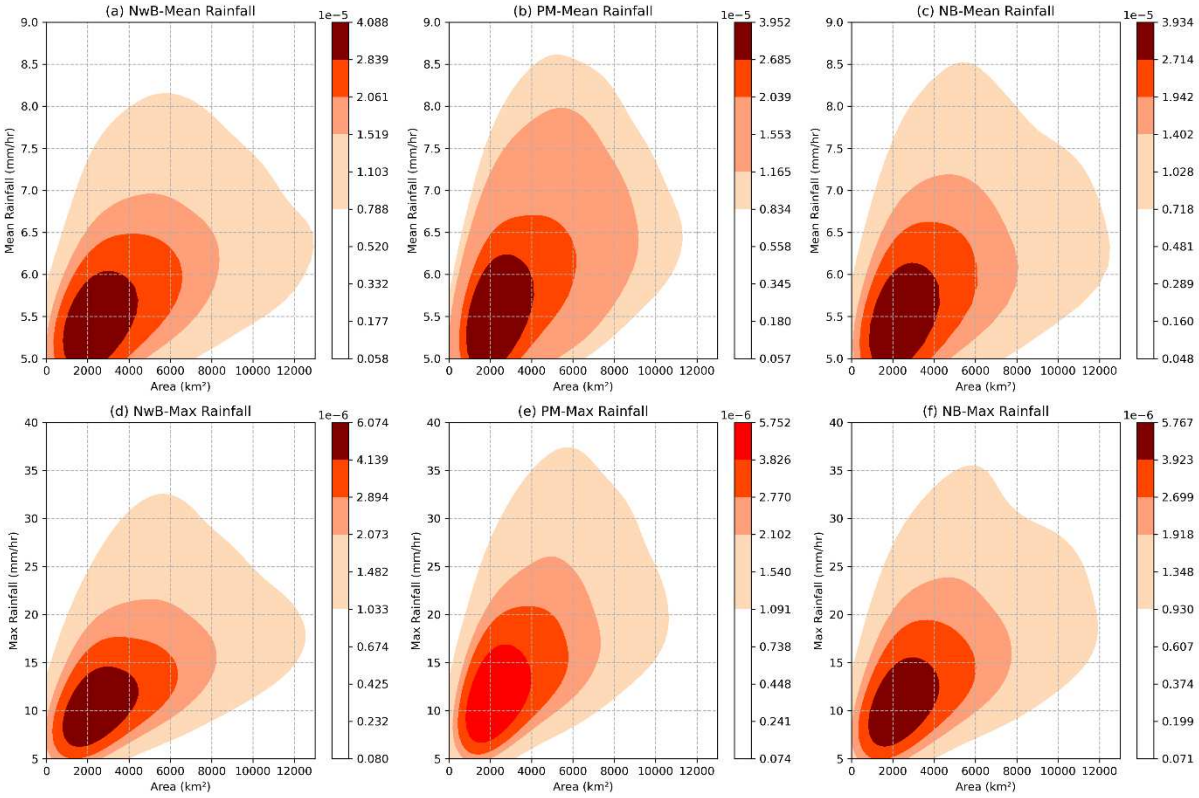


Figure 7: Joint distribution between the size of the squall line and mean rainfall (mm $h^{-1}$) and maximum rainfall (mm $h^{-1}$) using the kernel density estimation. (a,d) NwB, (b,e) PM, and (c,f) NB cluster.

The joint distribution for the mean rainfall and maximum rainfall as a function of the squall lines' lifetime is shown in Figure 8a-c. The squall line lifetime, in this case, refers to the elapsed time since the squall line was first detected and not the total life span from formation to dissipation. In the PM (Figure 8b) and NB clusters (Figure 8c), there is a high probability of finding 30- to 120-minute squall lines with a mean rainfall ranging from 5.3 to 6.5 mm $h^{-1}$. The joint distribution also shows that the mean rainfall for squall lines in NwB clusters (Figure 8a) at 30-120 minutes lifetime drops off more quickly compared to the squalls in the PM and NB clusters. This indicates that for the same lifetime, there is a lower likelihood of high mean rainfall for squall lines in the NwB cluster compared to the squall lines in the PM and NB clusters.

The joint distribution of maximum rainfall varies with the squall lines' lifetime in each cluster is shown in Figure 8d-f. The joint distribution for NwB (Figure 8d) and NB clusters (Figure 8f) are similar, with a high likelihood of maximum rainfall in the range

of 10-18 mm $h^{-1}$ for squall lines at 30- to 100-minute lifetime. However, for NB, the likelihood of maximum rainfall is slightly broader, with a range of 10-20 mm $h^{-1}$. Nonetheless, this range is still within the category of moderate rainfall. The joint distribution is different for the PM cluster (Figure 8e). It is common to find squall lines in this cluster with the maximum rainfall range of 15-25 mm $h^{-1}$ with a lifetime of 30 to 100 minutes. Thus, this suggests that given the same lifetime, the squall lines in PM are more likely to have higher maximum rainfall compared to the squalls in the Borneo clusters. All clusters show a peak in the maximum rainfall of around 45 mm $h^{-1}$ when the squall line's lifetime is around 60 minutes. The probability of this high maximum rainfall decreases for longer and shorter lifetimes.

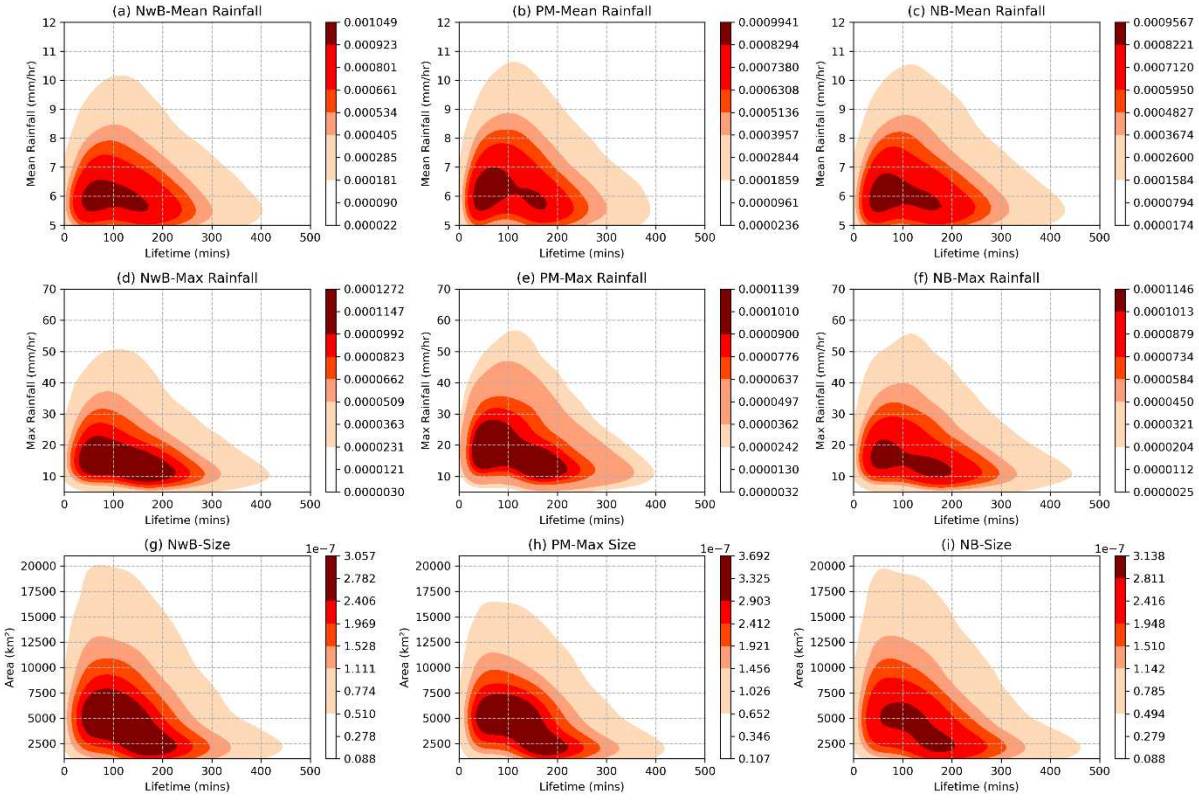


Figure 8: Joint distribution between lifetime and mean rainfall (mm $h^{-1}$), maximum rainfall (mm $h^{-1}$), and size ($km^2$) using the kernel density estimation. (a, d, g) NwB, (b,e,h) PM, and (c,f,i) NB cluster.

The squall line size and lifetime relationship are shown in Figure 8g-i. A consistent trend is observed in all three clusters. Squall lines of moderate size (3,000 to 8,000 km²) are the most common across all clusters, especially with lifetimes of 30

to 120 minutes, while the probability of larger squall lines decreases significantly as their lifetime exceeds 120 minutes. This suggests that larger squall lines struggle to sustain themselves for prolonged periods. However, upon inspecting individual clusters, it reveals that the NwB (Figure 8g) and NB clusters (Figure 8i) tend to have larger size ranges (up to 12,000 km²) for longer-lived squall lines (30 to 120 minutes). In contrast, the PM cluster (Figure 8h**)** shows a narrower range (between 3,500 and 7,000 km²) for shorter-lived squall lines (30–90 minutes).

Overall, the results above show that there are interlinks between the squall line size, rainfall intensity, and lifetime and they are unique in each cluster. In general, squall lines in Borneo (NB and NwB clusters) are large but have moderate rainfall intensity. In contrast, squall lines in PM are smaller than their counterparts in Borneo, short-lived but produce high-intensity rainfall and a greater likelihood of extreme rainfall. The weak dependence between size and intensity is consistent with the understanding of the tropical MCS rainfall, which is governed primarily by internal convective organisation and moisture structure, rather than by system extent alone (Houze Jr. 2004). These links provide information on how squall lines in this region might evolve and interact, as well as their roles in shaping regional rainfall patterns.

### 3.4 Squall initiation time

Figure 9a shows the spatial-temporal pattern of the squall line formation in the region. In the SCS cluster, the timing of the squall line formation is relatively uniform throughout the day. In contrast, squall lines commonly form at late night to pre-dawn in the oceanic region adjacent to the land or offshore (2100-0500 LST). A notable shift occurs by afternoon (1200-1500 LST), with the number of squall lines forming in the offshore region significantly decreasing. During this time, most squall lines form over land due to localized heating, which enhances convection. The interaction between the local sea breeze and local terrain can further induce these convective activities. As the day progresses into the mid-to-late afternoon (1500-1700 LST), the formations of land squall lines peak and extend further inland due to the intensification of the local sea breeze. The number of squall lines forming over the offshore region is minimal during this time. By early evening (1800-2000 LST), the squall line formations over the land decrease. In short, this pattern indicates that the majority of squall line formations

in the region are associated with the diurnal differential heating between the land and the adjacent ocean. The relative percentage distribution of squall line initiation locations over the land, ocean and coastal areas at each LST is shown in Figure 9b. It agrees with the previously discussed diurnal cycle, where most of the squall lines form offshore during late evening to early morning and inland during the early evening. A secondary peak between 0900 and 1200 LST also appears in Figure 9b, and this feature is primarily linked to squall lines in the SCS cluster. When these squall lines are excluded (Figure 9c), this secondary peak disappears, and the diurnal distribution becomes more straightforward. Thus, the difference shows that the secondary peak in the late morning is not a general feature of offshore squall line formation across all clusters, but rather specific to the SCS cluster. Further investigation is required to understand the mechanisms responsible for the secondary peak in SCS.

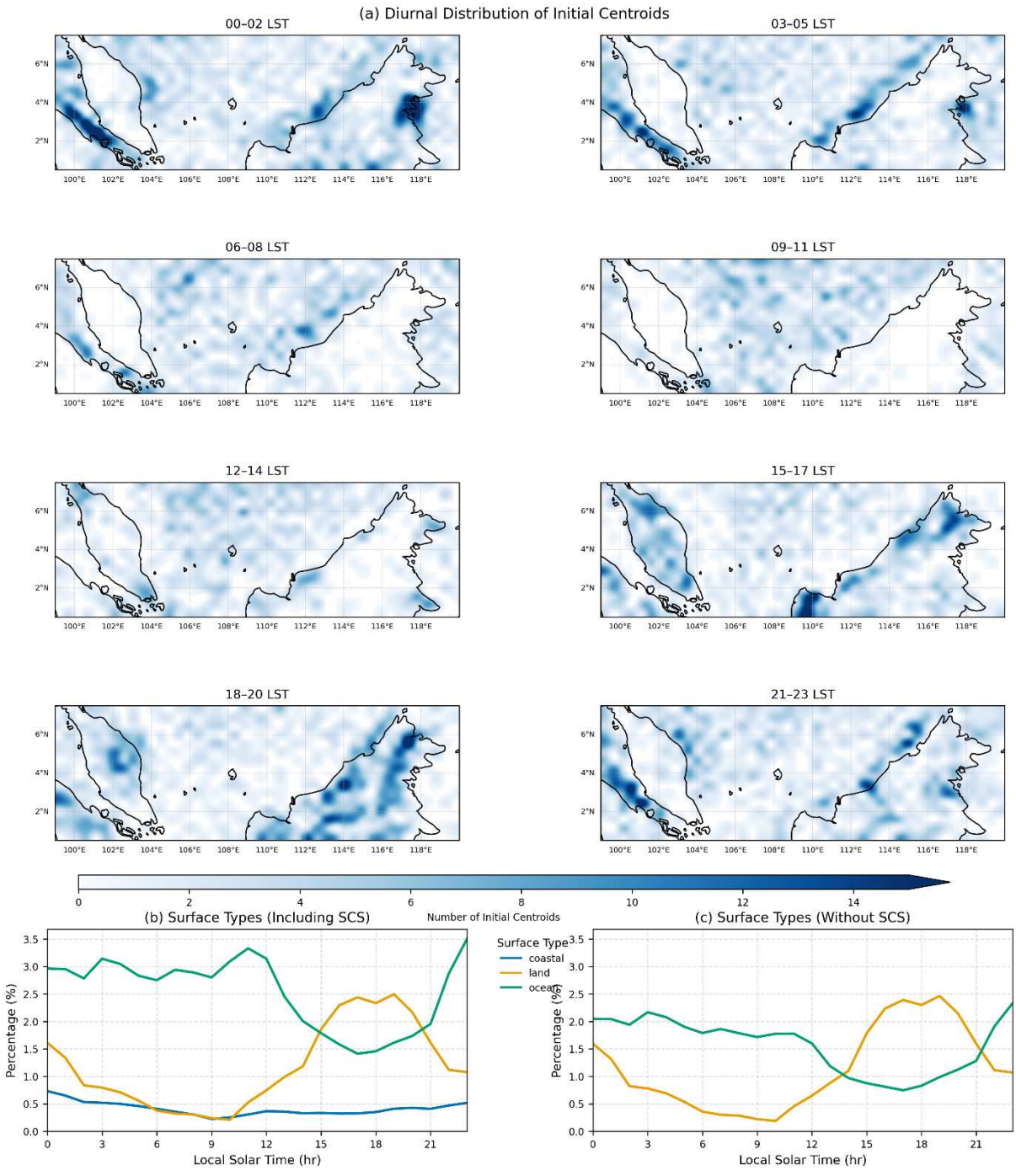


Figure 9: The spatial-temporal distribution of squall line formation. (a) The heat map of the squall line initiation locations is binned into 3-hourly. (b) The percentage distribution of squall line initiation locations over the land, ocean and coastal at each local solar time (LST). The land point is defined when it is located on a surface where less than 50% of the area is covered by water. In contrast, an ocean point is defined when it is entirely on the water surface, while a coastal point is defined when the

surface covers between 50% to 99% of water. (c) Same as (b) but squall lines from the SCS cluster are removed.

Figure 10 shows the time evolution composite mean of 300 observed squall lines in the oceanic region of the northern Strait of Malacca occurring between 0000-0200 LST (1600-1800 UTC) to illustrate further the role of the diurnal differential heating between the land and offshore. At 1800-2000 LST, six hours before (-6 hour) the squall line is detected offshore, northwesterly winds at 10-meter height are observed in the region and veer toward Sumatra as onshore winds (Figure 10a). Rain is observed predominantly along the foothills of the Barisan Mountains (indicated by brown contours) on the east coast of Sumatra and in the mountainous regions of Peninsular Malaysia. During this time, little rain is observed over the sea adjacent to the land. In the subsequent three hours, the onshore winds diminish from 2100 to 2300 LST and are replaced by offshore winds (Figure 10b). Along with this change, the rain begins to move offshore, while some rain in the land dissipates. Convergence is observed in the Strait of Malacca, where the offshore winds from Sumatra and Peninsular Malaysia meet. However, such convergence is not found on the east coast of Peninsular Malaysia. At 0000-0200 LST, the offshore winds intensify, causing the rain in this region to become well-organized and develop into a squall line (Figure 10c). Meanwhile, little rain is observed over the land at this time. A similar condition is observed at 0300-0600 LST, but the squall line intensified (Figure 10d). During the formation of the squall line, the line of convergence is observed on the eastern side of the squall. In this case, the rainband detected as a squall could be the trailing region of stratiform precipitation.

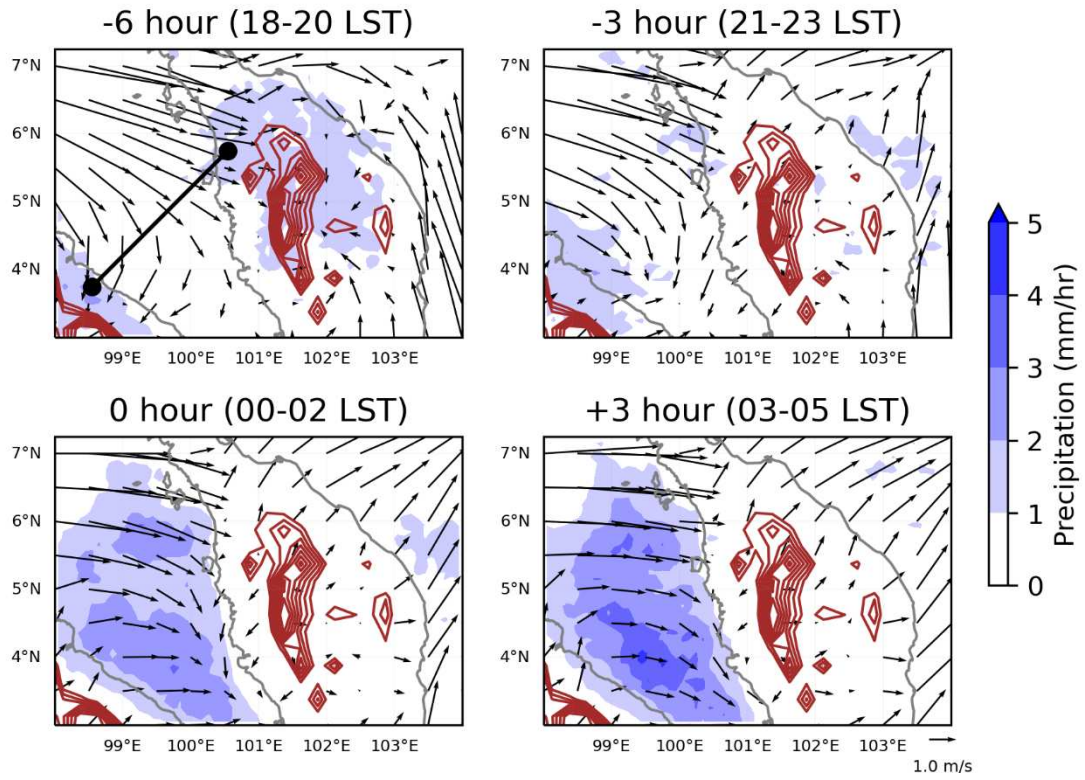


Figure 10: The time sequence composite of 300 squall lines observed in the northern Strait of Malacca occurring between 0000-0200 LST. 0 hour indicates the time when the squall is first detected in the region. The topography contours (in brown) start from 400m onward. The main range in Peninsular Malaysia is known as the Titiwangsa range. In Sumatra, the main mountain range is known as the Barisan range Shaded: precipitation rate; contour: topography from GLDAS; vector: 10-m wind from ERA5.

The composite mean shows the role of orography in the region. In Peninsular Malaysia, the onshore winds are rather weak or non-existent between 1800 and 2000 LST in certain parts of the east coast, except on the leeward side of the Titiwangsa range. This wind pattern occurs because the monsoonal winds during this period are westerly, opposing the east coast's onshore winds. In contrast, onshore winds in Sumatra are prominent on the leeward side of its Barisan mountain range. Between 2100 and 2300 LST, offshore winds develop on this side of the mountain range. Unlike in Sumatra, offshore winds in northwestern Peninsular Malaysia are weak and largely dominated by the monsoonal westerlies. The weak offshore winds can be attributed to the absence of significant orographic blockage in this region. As a result of weak offshore winds, rain in this region is unable to propagate seaward, whereas rain on the

east coast of Sumatra can propagate seaward along with the offshore winds. The Hovmöller diagram for the same composite shows that the centre of the intense rain propagates towards the northern Strait of Malacca from Sumatra at a mean speed of approximately 5 m $s^{-1}$ (Figure 11). As it propagates seawards, the rainband expands and organises into a squall line by 0000 LST. The role of the orography is further illustrated on the leeward side of the Titiwangsa range, on the east coast of Peninsular Malaysia. Here, offshore winds can be seen replacing the onshore winds starting 2100 LST. A rainband is observed in the coastal region at -3 hour, likely from the remnant of the rainband located in the highlands during the preceding hours (-6 hour). Unlike the condition in the northern Strait of Malacca, there is no low-level convergence between two land breezes in this region. As a result, the rainband could not be organised into a squall line but instead propagate weakly seaward, as observed at the 0 hour.

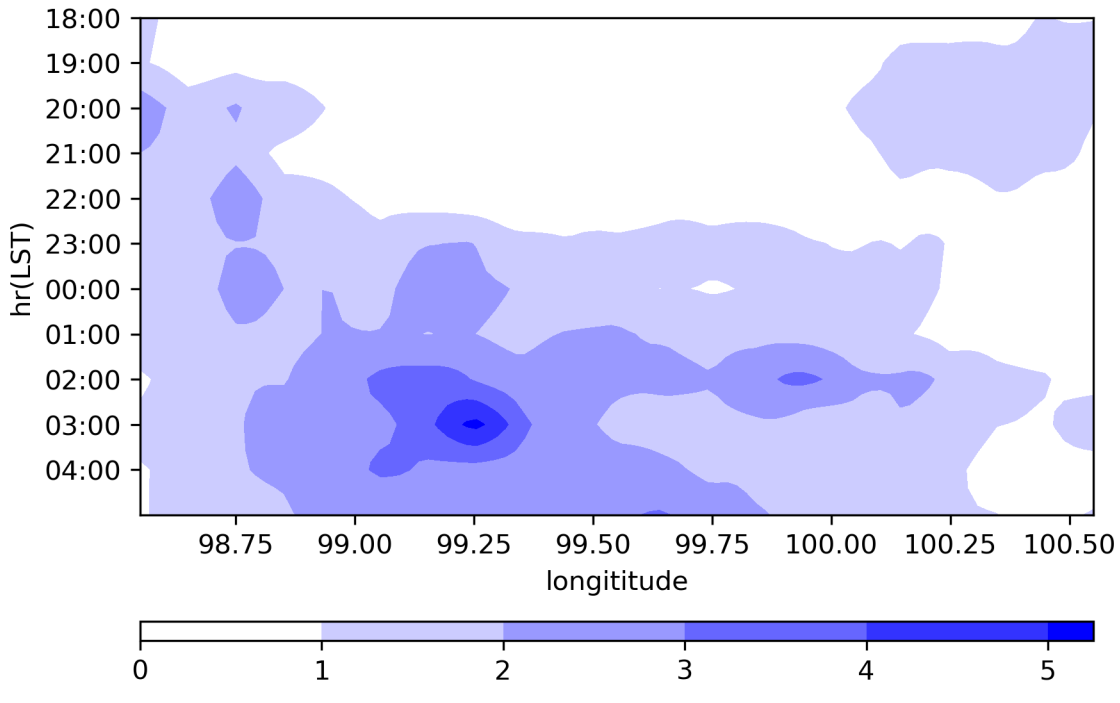


Figure 11: The Hovmöller plot of the composite mean rain rate for 300 identified squall lines form in the northern Strait of Malacca between 00-02 LST. The mean rain rate is taken along the black line in Figure 10.

The role of the offshore winds is further explored in the NwB cluster. In this cluster, similar to the east coast of Peninsular Malaysia, there is no convergence of two land breezes. Three hundred and six (306) squall lines formed over the oceanic

region at 0000 LST are composited (Figure 12) in a similar way to the northern Strait of Malacca case. During the 1800-2000 LST period (-6 hours), the winds are onshore, and rain forms mainly around the foothills of the mountain range (Figure 12a). Between 2100-2300 LST (-3 hours), the offshore winds begin to appear only on the northeastern part of the NwB (Figure 12b). This wind pattern suggests that the wind reversal in this part of the NwB is not solely attributed to the radiative cooling of the land but also the evaporative cooling by the rainfall located in the foothills during this time. This interpretation is consistent with previous studies showing that afternoon convection over western and northern Borneo is strongly tied to the terrain and often propagates offshore at night through shallow evening outflows and the onset of the land breeze (Ichikawa and Yasunari 2006; WU et al. 2008). Here, the offshore and monsoonal winds converge near the coast. Elsewhere, the offshore winds appear between 0000 and 0200 LST (0 hours), leading to the extension of the previous convergence line, which was initially confined to the northeastern side of the NwB to along the coast of NwB (Figure 12c). With this development, the squall line forms along the convergence line. Three hours later, between 0300-0500 LST (+3 hours), along with the intensification of the offshore wind, the squall line intensifies (Figure 12d).

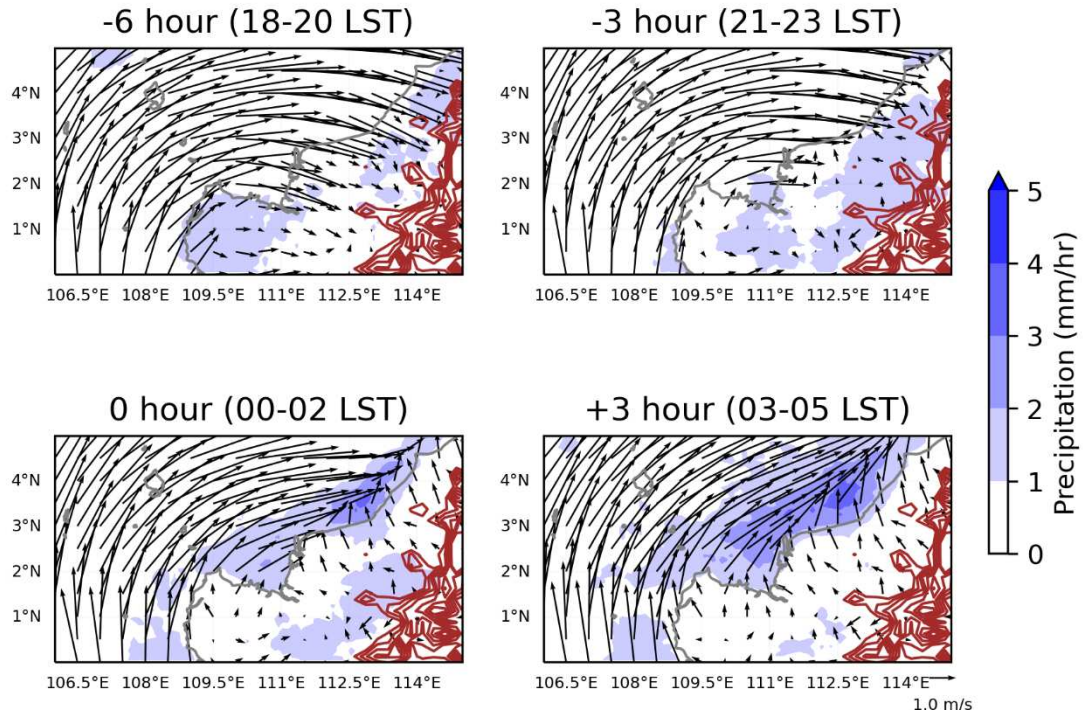


Figure 12: Same as Figure 10 but in the NwB region.

672 The two examples presented here illustrate the role of the offshore winds in the
673 formation of the squall lines in the region. In the Strait of Malacca case, the squall lines
674 form when the offshore winds from the west coast of Peninsular Malaysia converge
675 with those from the east coast of Sumatra. Additionally, the formation of the squall lines
676 is further aided by the propagation of rainfall remnants from the preceding land-based
677 convection over Sumatra. Therefore, without the accompanying propagation of rainfall
678 from previous land convection, low-level convergence alone may appear insufficient
679 to trigger squall line formation in the northern Strait of Malacca. The low-level winds
680 difference between the squall lines composite and the non-squall lines composite
681 indicates weaker northwesterly flow during the non-squall events (Supplementary
682 Figure S2). The weaker northwesterly winds may be insufficient to support the offshore
683 propagation of the rainfall. During the non-squall line events, low-level convergence
684 appears to be weaker. In the NwB cluster, although the squall lines formation is not
685 supported by the offshore propagation of rainfall, the absence of late afternoon land-
686 based convection appears to weaken the offshore winds (Supplementary Figure S3).
687 This weakening of offshore winds may be the result of a weaker land-sea thermal
688 contrast. The weaker southwesterly winds along the coast during the non-squall line
689 days are also consistent with the findings of Fakaruddin et al. (2022), who reported on
690 stronger southwesterly winds during the squall line days.

691

692 ### 3.5 Squall lines initiation in relation to BSISO Phases

693 The Boreal Summer Intraseasonal Oscillation (BSISO) represents the
694 dominating intraseasonal variability during the summer monsoon (Kikuchi et al. 2012;
695 Lee et al. 2013). Unlike the eastward propagating Madden Julian Oscillation (MJO),
696 which is the dominant intraseasonal oscillation during the winter monsoon, the
697 convection associated with the BSISO propagates north/northeastward. In this study,
698 we investigate the squall lines activity associated with the four phases of BSISO
699 defined by Xavier et al. (2024), which are based on the BSISO 1 (30-60 day) of Lee et
700 al. (2013). Table 4 summarizes the total number of initial squall lines during each active
701 phase (when the amplitude is greater than or equal 1). In the PM and NwB clusters,
702 approximately 58% and 61% of the total squall lines are observed during active phases
703 2 and 3 of the BSISO, which correspond to enhanced convection over the Indonesia-

Malaysia region. During the suppressed convection phase (phases 1 and 4) in the Indonesia-Malaysia region, the number of squall lines reduces drastically, particularly in phase 1. In contrast, a larger fraction of squall lines is observed during active phases 1 and 4 of the BSISO within the NB cluster and SCS clusters, when enhanced convection is found in the SCS, the Philippines and the Indochina region. Overall, these results indicate that the phases of BSISO modulate squall line activity in the region.

Table 4: Number of initial squall lines during each active BSISO phase for the clusters.

| | BSISO phase | | | |
|---|---|---|---|---|
| Cluster | 1 | 2 | 3 | 4 |
| P. Malaysia | 586 | 774 | 960 | 704 |
| N. Borneo | 971 | 496 | 627 | 1113 |
| S. China Sea | 1125 | 416 | 474 | 1203 |
| NW. Borneo | 334 | 553 | 826 | 517 |

## 4. Summary and concluding remarks

Convective activities are found all year round in the Maritime Continent. Surrounded by warm ocean and complex topography, this region experiences strong diurnal variation in the rainfall, peaking over land in the late afternoon and shifting offshore at night. Organised convection, including squall lines (elongated MCSs), is frequently observed near the long coastlines of the large islands like Sumatra and Borneo. Often, these squall lines often bring heavy rainfall, strong winds and significant maritime hazards, particularly during the boreal summer monsoon. The limited understanding underscores the need for further research to document their characteristics and possible impacts.

The squall lines in this study are detected from the IMERG GPM half-hourly final precipitation data spanning from 2001-2020, covering the extended boreal summer months. By applying the area overlapping technique for rainbands, the squall line is identified when the rainbands with a minimum rain rate of 5 mm $h^{-1}$ have a

minimum area of 1,000 $km^2$, and at least one instance exceeding 5,000 $km^2$. Using this method, a total of 191,312 ellipse-shaped rainbands are identified as squall lines. These squall lines are clustered into 4 clusters, namely the PM, NwB, NB and SCS clusters using the K-means clustering method. Our study reveals that the characteristics of the squall lines in the region vary across the clusters.

The average propagation speeds in this study align with previous studies that focus on individual observational case studies or modelling. Our results also reveal that the squall lines' average propagation speed in this region is linked to both large-scale systems, such as tropical storms and monsoonal flows, and the mesoscale circulation. The generally faster propagation speed in the SCS and Borneo clusters highlights the influence of TCs that occurred near the region. During TCs, the eastward and northeastward propagation speeds in these clusters increase, consistent with enhanced low-level westerlies or southwesterlies accompanying TCs. Nevertheless, these effects are non-uniform, suggesting that TC forcing does not fully control the propagation behaviour of the squall lines. Therefore, the weak positive correlation between the large-scale winds at the lower troposphere and the squall lines' average propagation speed is perhaps not unexpected, despite having similar distribution shapes. Since the study period covers the extended boreal summer months, it is unsurprising that the eastward and westward propagating squall lines occur under the monsoonal westerlies. However, when the prevailing westerly winds occasionally give way to easterly winds, westward propagations become more likely. However, our study also shows that there are squall lines that propagate against the background flow. Thus, this suggests the importance of internal storm dynamics, particularly the cold pool mechanism, which determines the squall line's propagation characteristics. Our results indicate that storm-generated dynamics also govern the propagation of squall lines in the region. These results underscore the need for modelling approaches capable of resolving cold pools, surface processes, and local terrain effects, as well as their combined effect on the squall lines' evolution.

Our analysis highlights that the squall lines in this region have sizes between 2,000 and 5,000 $km^2$. Across different clusters, large squall lines exceeding 10,000 $km^2$ are less frequent in the PM cluster than in the Borneo clusters. The formation of larger-sized squall lines in the PM cluster may be constrained either by the narrow Strait of Malacca or by the complex mesoscale circulation, which inhibits growth or

causes rapid decay. Despite the smaller size range, the squall lines in the PM cluster exhibit the highest mean maximum rain rates and the greatest likelihood of rainfall events exceeding 30 mm $h^{-1}$. These squall lines also tend to intensify over shorter periods and have shorter lifespans compared to those in the Borneo clusters. In other words, squall lines in Borneo clusters tend to have longer lifespans and distribute rainfall more broadly with moderate intensities.

The composite cases in the northern Strait of Malacca and NwB clusters indicate that the mechanism for the nighttime squall line formation differs across clusters. In the northern Strait of Malacca, the convergence of the offshore winds from Sumatra and Peninsular Malaysia intensified the previous evening rainband that propagated offshore from Sumatra and organised into a squall line in the region. In contrast, the offshore winds in the NwB converge with the large-scale southwesterly monsoon winds in the late evening, leading to the formation of squall lines along the NwB coast. On non-squall days, the offshore winds are weaker, primarily because the absence of late afternoon rainfall over land that limits the precipitation cooling over land, which in turn weakens the land-sea thermal contrast. This mechanism is consistent with previous findings that precipitation and cloudiness suppress the land-sea thermal contrast (Chen et al. 2016; Zhu et al. 2017). In the NwB, besides the weaker offshore winds, the southwesterly monsoonal winds are also weaker on the non-squall days. Therefore, the nighttime squall lines in the region are formed through the interaction between mesoscale offshore winds (including land-breeze) and the large-scale monsoonal flow, highlighting the multiscale interaction in the squall lines formation. Moreover, these multiscale interactions are further modulated at the intraseasonal timescale, as our results further show that the BSISO strongly modulates the squall line activity in the region. Squall lines in the PM and NwB clusters occur more frequently during BSISO phases 2 and 3, during the enhanced convection phase of the BSISO over the Indonesia-Malaysia region. In contrast, when the BSISO convective envelope shifts northeastward into the SCS region (BSISO phases 4-1), it thus favours the squall line development in the SCS and NB clusters. The extent to which BSISO phases modulate other squall-line characteristics, including size, propagation speed, and intensity, is beyond the scope of this study and requires a dedicated investigation in future work.

Our study provides statistical insight into the structure and evolution of squall lines without examining the environmental factors such as humidity and vertical wind shear profile, that may be important to elucidate the physical mechanisms of squall lines. Our findings also suggest that the propagation of the squall lines in the region is not solely steered by the large-scale low-level tropospheric winds, but also by complex mesoscale circulation such as cold-pool dynamics or the interactions with terrains. However, these processes cannot be investigated in detail using the coarse resolution ERA5 dataset. To improve understanding of squall lines, future work should incorporate detailed environmental diagnostics into high-resolution simulations capable of resolving storm-scale processes. Despite the limitations in the study, our findings offer valuable information to the forecasting community. Additionally, the squall lines dataset in this study, which contains detailed attributes of size, rainfall, propagation and duration, can be very useful for developing and validating nowcasting tools. This labelled dataset could be used to train supervised learning models to detect, classify and predict convective systems. Furthermore, the identified physical relationships and the regional clusters may provide support for the improvement of area-specific convective parameterisations in numerical weather prediction models.

**Acknowledgement:**

This research and the contributors were supported by the Weather and Climate Science for Services Partnership (WCSSP) Southeast Asia. The first and third authors would like to thank the Director-General of the Malaysian Meteorological Department for supporting this research. We would also like to thank the two anonymous reviewers for their comments that helped improve this manuscript.

**References:**

Bai, H., and Coauthors, 2021: Formation of Nocturnal Offshore Rainfall near the West Coast of Sumatra: Land Breeze or Gravity Wave? *Mon. Weather Rev.*, **149**, 715–731, https://doi.org/10.1175/MWR-D-20-0179.1.

Chen, X., F. Zhang, and K. Zhao, 2016: Diurnal Variations of the Land–Sea Breeze and Its Related Precipitation over South China. *J. Atmos. Sci.*, **73**, 4793–4815, https://doi.org/10.1175/JAS-D-16-0106.1.

Coppin, D., and G. Bellon, 2019: Physical Mechanisms Controlling the Offshore

827 Propagation of Convection in the Tropics: 2. Influence of Topography. *J. Adv.*
828 *Model. Earth Syst.*, **11**, 3251–3264,
829 https://doi.org/https://doi.org/10.1029/2019MS001794.

830 Fakaruddin, F. J., N. A. Nawai, M. Abllah, F. Tangang, and L. Juneng, 2022:
831 Climatological Features of Squall Line at the Borneo Coastline during Southwest
832 Monsoon. *Atmosphere (Basel).*, **13**, https://doi.org/10.3390/atmos13010116.

833 Fujita, M., F. Kimura, and M. Yoshizaki, 2010: Morning Precipitation Peak over the
834 Strait of Malacca under a Calm Condition. *Mon. Weather Rev.*, **138**, 1474–1486,
835 https://doi.org/10.1175/2009MWR3068.1.

836 Hassim, M. E. E., T. P. Lane, and W. W. Grabowski, 2016: The diurnal cycle of
837 rainfall over New Guinea in convection-permitting WRF simulations. *Atmos.*
838 *Chem. Phys.*, **16**, 161–175, https://doi.org/10.5194/acp-16-161-2016.

839 Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Q. J. R. Meteorol.*
840 *Soc.*, **146**, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803.

841 Houze Jr., R. A., 2004: Mesoscale convective systems. *Rev. Geophys.*, **42**,
842 https://doi.org/https://doi.org/10.1029/2004RG000150.

843 Houze, R. A., S. G. Geotis, F. D. Marks, and A. K. West, 1981: Winter Monsoon
844 Convection in the Vicinity of North Borneo. Part I: Structure and Time Variation
845 of the Clouds and Precipitation. *Mon. Weather Rev.*, **109**, 1595–1614,
846 https://doi.org/10.1175/1520-0493(1981)109<1595:WMCITV>2.0.CO;2.

847 Huffman, G. J., D. T. Bolvin, E. J. Nelkin, and T. J., 2020: Integrated multi-satellitE
848 retrievals for GPM (IMERG) technical documentation.

849 Ichikawa, H., and T. Yasunari, 2006: Time–Space Characteristics of Diurnal Rainfall
850 over Borneo and Surrounding Oceans as Observed by TRMM-PR. *J. Clim.*, **19**,
851 1238–1260, https://doi.org/10.1175/JCLI3714.1.

852 Jammalamadaka, S. R., and Y. R. Sarma, 1988: A correlation coefficient for angular
853 variables. *Stat. theory data Anal. II*, 349–364.

854 Kikuchi, K., B. Wang, and Y. Kajikawa, 2012: Bimodal representation of the tropical
855 intraseasonal oscillation. *Clim. Dyn.*, **38**, 1989–2000,
856 https://doi.org/10.1007/s00382-011-1159-1.

857 Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010:
858 The International Best Track Archive for Climate Stewardship (IBTrACS):
859 Unifying Tropical Cyclone Data. *Bull. Am. Meteorol. Soc.*, **91**, 363–376,
860 https://doi.org/https://doi.org/10.1175/2009BAMS2755.1.

861 Lee, J.-Y., B. Wang, M. C. Wheeler, X. Fu, D. E. Waliser, and I.-S. Kang, 2013:
862 Real-time multivariate indices for the boreal summer intraseasonal oscillation
863 over the Asian summer monsoon region. *Clim. Dyn.*, **40**, 493–509,
864 https://doi.org/10.1007/s00382-012-1544-4.

865 Lo, J. C.-F., and T. Orton, 2016: The general features of tropical Sumatra Squalls.
866 *Weather*, **71**, 175–178, https://doi.org/https://doi.org/10.1002/wea.2748.

867 Love, B. S., A. J. Matthews, and G. M. S. Lister, 2011: The diurnal cycle of
868 precipitation over the Maritime Continent in a high-resolution atmospheric

model. *Q. J. R. Meteorol. Soc.*, **137**, 934–947, https://doi.org/https://doi.org/10.1002/qj.809.

Malaysian Meteorological Department, Squall Line. https://www.met.gov.my/en/pendidikan/fenomena-cuaca/#Squall Line (Accessed March 27, 2025).

Mapes, B. E., T. T. Warner, and M. Xu, 2003: Diurnal Patterns of Rainfall in Northwestern South America. Part III: Diurnal Gravity Waves and Nocturnal Convection Offshore. *Mon. Weather Rev.*, **131**, 830–844, https://doi.org/10.1175/1520-0493(2003)131<0830:DPORIN>2.0.CO;2.

Neale, R., and J. Slingo, 2003: The Maritime Continent and Its Role in the Global Climate: A GCM Study. *J. Clim.*, **16**, 834–848, https://doi.org/10.1175/1520-0442(2003)016<0834:TMCAIR>2.0.CO;2.

Nesbitt, S. W., and E. J. Zipser, 2003: The Diurnal Cycle of Rainfall and Convective Intensity according to Three Years of TRMM Measurements. *J. Clim.*, **16**, 1456–1475, https://doi.org/10.1175/1520-0442(2003)016<1456:TDCORA>2.0.CO;2.

Peatman, S. C., A. J. Matthews, and D. P. Stevens, 2014: Propagation of the Madden–Julian Oscillation through the Maritime Continent and scale interaction with the diurnal cycle of precipitation. *Q. J. R. Meteorol. Soc.*, **140**, 814–825, https://doi.org/https://doi.org/10.1002/qj.2161.

——, C. E. Birch, J. Schwendike, J. H. Marsham, C. Dearden, S. Webster, R. R. Neely, and A. J. Matthews, 2023: The Role of Density Currents and Gravity Waves in the Offshore Propagation of Convection over Sumatra. *Mon. Weather Rev.*, **151**, 1757–1777, https://doi.org/10.1175/MWR-D-22-0322.1.

Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.*, **85**, 381–394, https://doi.org/10.1175/BAMS-85-3-381.

Rotunno, R., J. B. Klemp, and M. L. Weisman, 1988: A Theory for Strong, Long-Lived Squall Lines. *J. Atmos. Sci.*, **45**, 463–485, https://doi.org/10.1175/1520-0469(1988)045<0463:ATFSLL>2.0.CO;2.

Scott, D. W., 1992: *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley \& Sons,.

Singapore Meteorological Services, Weather Systems. https://www.weather.gov.sg/learn_weather_systems/ (Accessed April 23, 2025).

Wu, P., S. Mori, J.-I. Hamada, M. D. Yamanaka, J. Matsumoto, and F. Kimura, 2008: Diurnal Variation of Rainfall and Precipitable Water over Siberut Island off the Western Coast of Sumatra Island. *SOLA*, **4**, 125–128, https://doi.org/10.2151/sola.2008-032.

WU, P., D. Y. Manabu, and J. MATSUMOTO, 2008: The Formation of Nocturnal Rainfall Offshore from Convection over Western Kalimantan (Borneo) Island. *J. Meteorol. Soc. Japan. Ser. II*, **86A**, 187–203, https://doi.org/10.2151/jmsj.86A.187.

Xavier, P., J. Y. Diong, M. F. A. Bin Abdullah, D. Permana, A. Pura, and H. Lam, 2024: The influence of boreal summer intraseasonal oscillations on precipitation extremes and their characteristics in Southeast Asia. *npj Clim. Atmos. Sci.*, **7**,

112, https://doi.org/10.1038/s41612-024-00658-6.

Yi, L., and H. Lim, 2006: SEMI-IDEALIZED COAMPS® SIMULATIONS OF SUMATRA SQUALL LINES: THE ROLE OF BOUNDARY FORCING. *Advances in Geosciences*, 111–124.

Yokoi, S., S. Mori, M. Katsumata, B. Geng, K. Yasunaga, F. Syamsudin, Nurhayati, and K. Yoneyama, 2017: Diurnal Cycle of Precipitation Observed in the Western Coastal Area of Sumatra Island: Offshore Preconditioning by Gravity Waves. *Mon. Weather Rev.*, **145**, 3745–3761, https://doi.org/10.1175/MWR-D-16-0468.1.

Yulihastin, E., and Coauthors, 2023: Propagation of tropical squall line-induced storm coastal inundation episodes in Java-Bali, Indonesia. *Heliyon*, **9**, e19804, https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e19804.

Zhu, L., Z. Meng, F. Zhang, and P. M. Markowski, 2017: The influence of sea- and land-breeze circulations on the diurnal variability in precipitation over a~tropical island. *Atmos. Chem. Phys.*, **17**, 13213–13232, https://doi.org/10.5194/acp-17-13213-2017.

**Supplementary Material for The Characteristics of Squall Lines in the Southeast Asia Region.**

Diong Jeong Yik[1] (diong@met.gov.my)

Prince Xavier[2] (prince.xavier@metoffice.gov.uk)

Muhammad Firdaus Ammar Abdullah[1] (kumar@met.gov.my)

Juliane Schwendike[3] (J.Schwendike@leeds.ac.uk)

[1]Research and Technical Development Division, Malaysian Meteorological Department, Petaling Jaya, Selangor, Malaysia

[2]Met Office, Exeter, UK

[3]University of Leeds, Leeds, UK

**Corresponding Author: Diong Jeong Yik,** diong@met.gov.my

## Supplementary Figures:

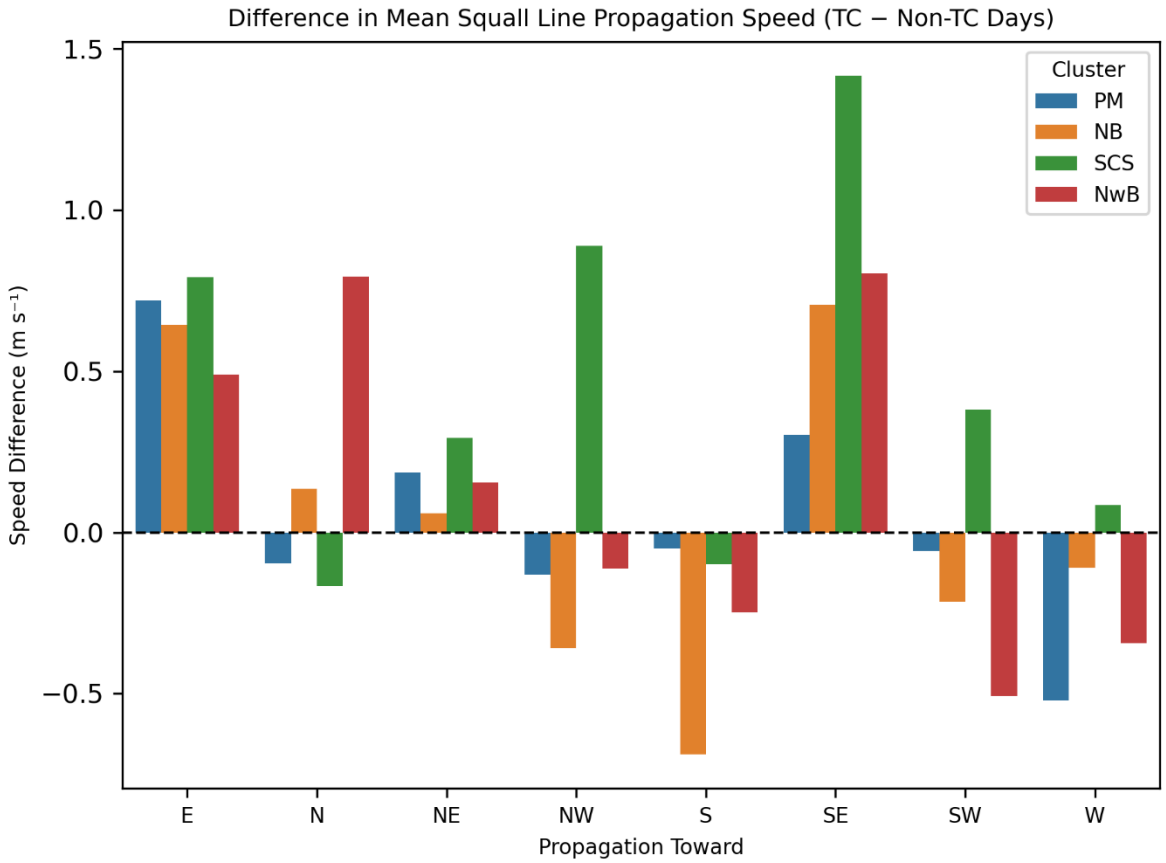


Figure S1: Difference in the mean propagation speed of squall lines between TC days and non-TC days. The direction is towards the squall line movement.

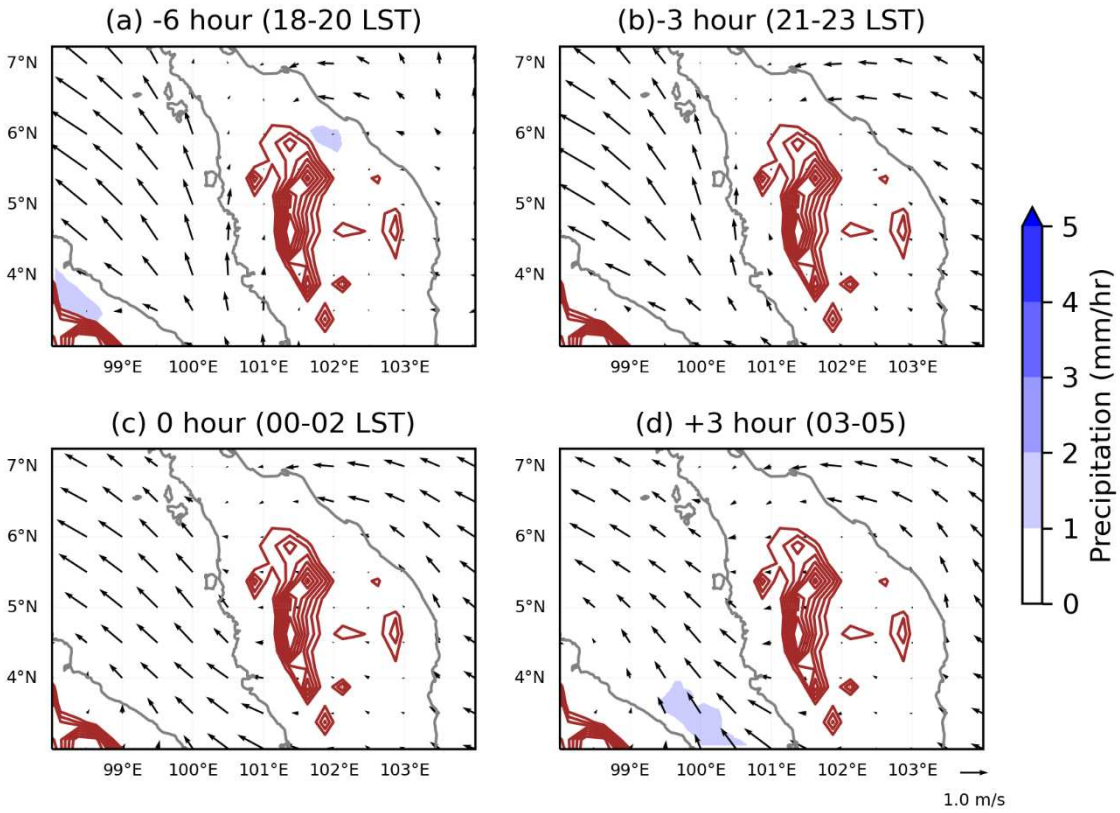


Figure S2: Composite difference of 10m wind field between squall line days and non-squall line days in the northern Strait of Malacca. Shaded contours indicate the rainfall on non-squall days.

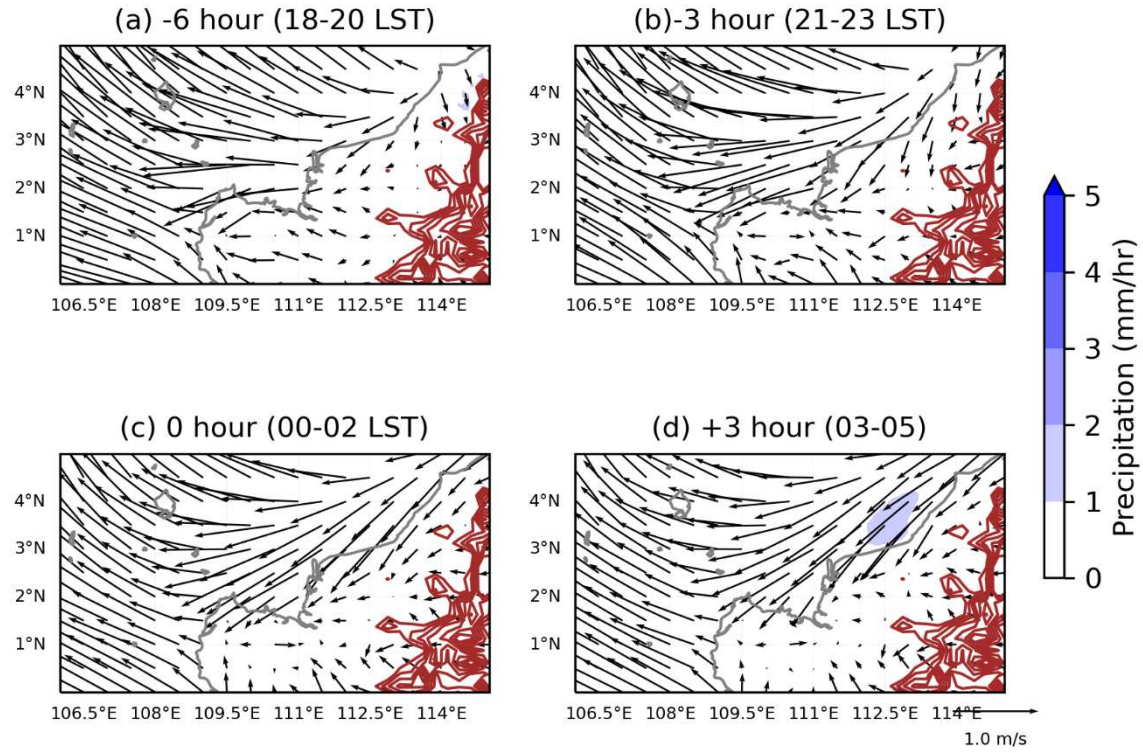


Figure S3: Same as Figure S2, but in NwB.