Weather patterns in Southeast Asia: Relationship with tropical variability and heavy precipitation

Emma Howard, Simon Thomas, Thomas H.A. Frame, Paula L.M. Gonzalez, John Methven, Oscar Martínez-Alvarado, Steven Woolnough

Received: 26 May 2021 | Revised: 24 November 2021 | Accepted: 6 December 2021 | Published on: 24 December 2021

DOI: 10.1002/qj.4227

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Weather patterns in Southeast Asia: Relationship with tropical variability and heavy precipitation

Emma Howard[1] | **Simon Thomas**[1,2] | **Thomas H.A. Frame**[3] | **Paula L.M. Gonzalez**[1,4] | **John Methven**[3] | **Oscar Martínez-Alvarado**[1] | **Steven J. Woolnough**[1]

[1]National Centre for Atmospheric Science, Department of Meteorology, University of Reading, UK

[2]L'Institut de Recherche en Astrophysique et Planétologie, Toulouse, France

[3]Department of Meteorology, University of Reading, Reading, UK

[4]International Research Institute for Climate and Society, The Earth Institute, Columbia University, Palisades, New York USA

Correspondence
E. Howard, NCAS/Department of Meteorology, University of Reading, Reading RG6 6BG, UK.
Email: e.r.howard@reading.ac.uk

Funding information
Met Office Weather and Climate Science for Service Partnership (WCSSP) Southeast Asia as part of the Newton Fund the National Centre for Atmospheric Science ODA national capability program ACREW NERC and the GCRF, Grant/Award Number: NE/R000034/1

Abstract

Two sets of weather patterns describing variability in 850 hPa winds in Southeast Asia are presented and compared. Patterns are calculated using EOF/*k*-means clustering with and without imposing a separation between planetary-scale and regional-scale circulation features. The former are labelled as tiered patterns while the latter are referred to as flat. The ability of the patterns to distinguish between known modes of tropical circulation variability is examined. This includes climate modes such as the seasonal monsoons, the El Niño Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) as well as sub-seasonal modes including cold surges, phases of the MJO and Boreal summer intraseasonal oscillation (BSISO), tropical cyclones, Borneo vortices and equatorial waves. All these modes are well captured by the weather patterns except for the equatorial waves and the IOD. The tiered patterns are shown to better describe large-scale modes of variability, while the flat patterns better describe the synoptic variability. Both sets of weather patterns are then used to study the likelihood of heavy precipitation depending on synoptic circulation by considering the regime-conditioned probability of high-percentile precipitation using the satellite-derived Global Precipitation Measurement (GPM) dataset. It is shown that the pattern centroids explain up to 10% of the seasonally anomalous precipitation over land, and that a perfect weather pattern forecast would outperform a perfect MJO forecast. These weather patterns show promising potential in extending the useful forecast range for the risk of heavy precipitation, dependent on their forecastability.

KEYWORDS

Weather patterns, Southeast Asia, synoptic variability, intraseasonal precipitation variability

1 | INTRODUCTION

High-Impact Weather (HIW) is a key economic and social risk in Southeast Asia. The Philippines, Vietnam and Thailand all rank in the ten countries most affected by extreme weather from 1999–2018 (Eckstein *et al.*, 2019). HIW events in this region largely take the form of floods and landslides associated with heavy precipitation due to both tropical storms and other convective events. Improving prediction of heavy precipitation events across the

sub-seasonal to seasonal (S2S) forecast ranges would bring major societal and economic benefits by allowing timely decisions to be made to protect lives and livelihoods. However, accurately forecasting precipitation is notoriously difficult owing to the strong dependence of precipitation on processes across a wide range of length-scales. In Southeast Asia, this problem is exacerbated by the complex geography of islands, mountains and seas, which generate complex interactions between atmospheric circulation and precipitation that are difficult for general circulation models to capture (Love *et al.*, 2011; Birch *et al.*, 2016; Ferrett *et al.*, 2020). Large-scale circulation itself is generally considered more predictable than precipitation (e.g., Boer, 2003; Hohenegger and Schäar, 2007; Zhang *et al.*, 2007; Selz and Craig, 2015; Ying and Zhang, 2017), however forecast models struggle to simulate the dependence of precipitation on atmospheric circulation. The increased predictability of circulation motivates the application of *regional-scale circulation types*, or *weather patterns*, together with their conditional relationships with precipitation, for the generation of probabilistic heavy precipitation forecasts. This methodology allows the poorly resolved dynamical relationship between precipitation and circulation to be replaced by a more robust statistical relationship derived from observations.

The use of weather patterns has become widespread since the early work by Baur *et al.* (1944) and Lamb (1972) (also James, 2007). Weather patterns have been defined using different techniques (reviews by Huth *et al.*, 2008; Philipp *et al.*, 2010; Ghil *et al.*, 2018). Two such techniques are Empirical Orthogonal Function (EOF) analysis (North *et al.*, 1982) and cluster analysis. EOF analysis is used in atmospheric science to extract sets of patterns (the EOFs) which explain most of the variance in a given system (e.g., Philipp *et al.*, 2010). The EOFs can then be used as a basis to express the original time series as a linear combination of EOFs multiplied by coefficients called principal components (PCs). The EOF patterns are orthogonal in space and their PCs are orthogonal in time (a property known as bi-orthogonality). Cluster analysis consists of seeking groups within multi-dimensional data with the aim of minimising within-cluster variances, also known as *inertia*. While EOF analysis produces patterns that are in principle present throughout the time series, only changing in intensity, cluster analysis, as applied in this paper, assigns each point in the dataset to a given group or cluster. These groups are characterised by patterns which can be understood as centroids, or composite means, of the data within the cluster. Cluster analysis has been used to define weather patterns in a variety of regions around the globe including (but not limited to) the North Atlantic (e.g., Michelangeli *et al.*, 1995; Cassou, 2008), the United Kingdom and Europe (e.g., Neal *et al.*, 2016; Grams *et al.*, 2017), India (Neal *et al.*, 2020), Central America (Sáenz and Durán-Quesada, 2015) and Mexico (Thomas *et al.*, 2020). Another methodology which has been applied to identify rainfall states in subtropical and tropical regions is the use of Hidden Markov Models, for example in Asia (Holsclaw *et al.*, 2016), India (Pal *et al.*, 2015) and New Caledonia (Moron *et al.*, 2016). This approach derives states directly from precipitation data, rather than considering atmospheric circulation.

In Southeast Asia and its surrounds, cluster analysis has been applied to a range of variables to analyse variability on sub-seasonal to interannual time-scales. Hassim and Timbal (2019) applied clustering to radiosonde soundings over Singapore to generate eight clusters, which they found partitioned the seasonal passage of the ITCZ across the Equator. Moron *et al.* (2015) limited their focus to the Austral monsoon season and clustered on 850 hPa winds, for which they identified six clusters. In both these studies, interannual precipitation variability is reflected in variations in the timings and frequencies of the separate clusters. Despite using very different input datasets, these studies achieved qualitatively similar results, suggesting that a degree of robustness to the choice of input data is present. This was explored by Moron *et al.* (2019) in a study of Northern Australia which overlapped Southeast Asia, who found that three of the five Darwin sounding-based weather regimes of Pope *et al.* (2009) had a strong correspondence with one of their six horizontal wind-based weather types.

The distinction between weather regimes and weather patterns in this paper is based on the approach of Neal *et al.* (2016, 2020). Under their methodology, both weather regimes and patterns describe circulation types within a regional domain, however weather patterns typically vary on a daily basis, whereas weather regimes are longer lasting. They describe weather regimes as persistent states of the atmosphere, which are of larger length-scale and fewer in number than weather patterns. Neal *et al.* (2016) first identified the 30 European weather patterns and secondly objectively grouped these patterns into eight regimes characterising the larger-scale flow. These eight larger-scale circulation regimes were identified with the positive and negative phases of the North Atlantic Oscillation and other key circulation types over Europe. Similarly, Neal *et al.* (2020) classified 30 Indian weather patterns *a posteriori* according to their seasonality and circulation and precipitation characteristics into seven large-scale weather regimes, which can be identified with known circulation features such as the different phases of the monsoon.

Generally across the all studies of low-latitude tropical and subtropical domains described above, planetary variability associated with the annual cycle and the El Niño Southern Oscillation (ENSO) dominate the

large-scale regimes. This stands in contrast to the midlatitude domains, where sub-seasonal variability such as the North Atlantic Oscillation dominates.

This paper addresses the question: is it possible to achieve a classification of the modes of variability that influence the regional circulation over Southeast Asia directly from gridded meteorological data? To answer this question, a two-tiered approach motivated by the weather pattern/weather regime split was taken. The flow is first classified into regimes within a planetary-scale domain (tier-1) followed by a secondary classification into weather patterns, conditioned on the tier-1 regimes, within a smaller, regional-scale domain (tier-2). We then compare the tiered set to another set of weather patterns computed directly on the smaller Southeast Asia domain. This methodology differs from Neal *et al.* (2016, 2020) as it identifies the weather regimes first and then uses the regimes to generate the patterns, rather than aggregating weather patterns to generate weather regimes.

Southeast Asia is influenced by a varied set of weather phenomena including equatorial waves, the Madden–Julian Oscillation (MJO) and Boreal Summer Intraseasonal Oscillation (BSISO), tropical cyclones, Borneo vortices and cold surges associated with the Asian winter northeast monsoon. Therefore, the weather patterns must encode a large amount of information and may struggle to capture all the potential synoptic drivers. A key result of this paper is the characterisation of how these weather phenomena are manifest in the weather patterns.

This paper describes the methodology used to generate the two sets of weather patterns, the degree to which the weather patterns capture the variability in atmospheric circulation, and the conditionality of HIW on the assigned patterns. We contrast the ability of the two pattern sets to:

- Describe planetary-scale modes of variability, including ENSO, the MJO and the seasonal monsoons;
- Separate different synoptic weather phenomena (cold surges, Borneo vortices, equatorial waves, BSISO and tropical cyclones) into different patterns; and
- Explain variability in the temporal distribution of precipitation.

A companion paper, (Personal Communication, P. Gonzalez 2021, hereafter G21), considers the predictability of these weather patterns and the resulting skill of a pattern-conditioned precipitation forecast based on the Met Office Global Seasonal Forecasting System version 5 (GloSEA5) ensemble forecasts.

The rest of this paper proceeds as follows. Section 2 provides details on the data used and the methodology applied in the definition and evaluation of the tiered and flat weather patterns. Section 3 describes the general characteristics of the resultant pattern sets, including their seasonal cycles, persistence and the Euclidean distances between the pattern centroids. Section 4 considers the degree to which each set of patterns describes planetary- and synoptic-scale atmospheric variability. The relationship between the patterns, precipitation and HIW is then studied in Section 5. Finally, Section 6 provides a summary of the work and a discussion on the potential usefulness of each approach towards the aim of enhancing the probabilistic forecast of HIW over Southeast Asia at sub-seasonal to seasonal forecast lead times.

2 | DATA AND METHODS

2.1 | Data

The EOF and cluster analysis was performed on 850 hPa horizontal wind components, which have been shown to well describe the monsoonal circulation in the Tropics (e.g., Moron *et al.*, 2010), and have been used as the basis for the definition of weather patterns in previous studies (e.g., Moron *et al.*, 2015; Neal *et al.*, 2020). The data were taken from the reanalysis ERA5 (Hersbach *et al.*, 2020) produced by the European Centre for Medium-range Weather Forecasts. The ERA5 high-resolution realisation (referred to as HRES) 850 hPa hourly horizontal wind components were retrieved at the ERA5 native resolution (0.28°), equivalent to 31 km longitudinal grid spacing at the Equator. Daily means were then derived from the hourly wind data. ERA5 HRES surface pressure was also retrieved and used to mask out any grid point for which the 850 hPa pressure level was beneath the ground at any time during the ERA5 period (1979–2018).

The fields were retrieved in a domain defined by the latitudinal band between 35°S and 35°N and a longitudinal sector between 60° and 180°E. This domain is referred to as the *planetary-scale domain*. Analysis was performed on both this planetary-scale domain and a spatial subset occupying approximately one quarter of the planetary-scale domain area. This subset is referred to as the *regional-scale domain* and defined by the latitudinal band between 15°S and 25°N and a longitudinal sector between 90° and 140°E. The data was retrieved for the period from 01 January 1979 to 31 December 2018 from the Copernicus Climate Change Service Climate Data Store (CDS) through the CDS API.

Global Precipitation Measurement (GPM) daily data (Huffman *et al.*, 2015) was used to quantify the precipitation over both the planetary-scale domain and the regional-scale domain. The GPM data were retrieved from June 2000 until the end of 2018 and interpolated from its native $0.1° \times 0.1°$ resolution onto a $1.5° \times 1.5°$ grid. This interpolated resolution was selected for ease of comparison with S2S forecast precipitation in G21.

2.2 | Methodology

To derive the weather pattern sets, we applied EOF analysis to compute the leading patterns of variability within the data. Cluster analysis was then performed on the reduced phase space defined by the leading principal components. This methodology is conceptually similar to that followed by other previous studies (e.g., Cassou, 2008; Sáenz and Durán-Quesada, 2015; Neal *et al.*, 2016; 2020; Thomas *et al.*, 2020). This methodology has been applied ten times to construct three sets of patterns:

- First to the full 38-year dataset across the planetary-scale domain to obtain the eight tier-1 weather patterns;
- Then separately to time subsets classified into each of the eight tier-1 weather patterns restricted to the regional-scale domain in order to obtain the 51 tier-2 weather patterns; and
- Finally to the full dataset across the regional-scale domain to obtain a corresponding set of 51 flat weather patterns.

The planetary-scale (or tier-1) weather patterns were defined using daily 850 hPa horizontal wind components with annual cycle components retained on the planetary-scale domain (defined in Section 2.1). Forty EOFs were retained to explain 95% of the total variance (proportional to kinetic energy). *k*-means cluster analysis was used on the PCs corresponding to the 40 retained EOFs. The Gap method of Tibshirani *et al.* (2001), with reference distribution defined using method b (section 4 of Tibshirani *et al.*, 2001), was used to estimate the optimal number of clusters. This was found to be nine. Two of these nine patterns were found to be dominated by the two in-quadrature phases of a propagating Southern Hemisphere subtropical wave: that is, they have the same spatial waveform shifted by $\pi/2$ in the zonal direction but were otherwise very similar. This waveform was located outside of the inner domain described below. These two clusters were therefore merged into a single pattern. The resulting eight centroid patterns are consistent with the eight clusters found on an analysis of weather regimes over Singapore (Hassim and Timbal, 2019), although the latter were defined taking a very different approach by clustering vertical soundings at a point location and compositing the large-scale circulation around the region. Centroid maps are provided in Figure 3 below.

The second tier of weather patterns were conditioned on membership of one of the eight planetary-scale patterns, and calculated using the regional-scale domain only. For each of the eight tier-1 regimes, EOF analysis was applied separately to horizontal wind components on each of the eight data subsets. As a consequence of this design, the tier-1 means were removed from the resultant EOFs, but all remaining seasonality not captured in tier-1 was retained. In each case, sufficient PCs were retained to explain 95% of the total variance. *k*-means cluster analysis was then applied to the retained PCs in order to identify the second-tier regional weather patterns. The number of patterns in each sub-set, determined using the Gap method (Tibshirani *et al.*, 2001), varied between five and eight, to produce 51 regional weather patterns in total in the second tier. Centroid maps are provided as a part of the Supporting Information in Figure S1.

The flat weather patterns were calculated using the 38-year 850 hPa horizontal wind dataset with annual cycle components retained, limited to the regional-scale domain. EOF analysis was used to reduce the original data dimensionality to 45 EOFs, explaining 95% of the variance in the first 100 EOFs. As before, *k*-means cluster analysis was applied to the retained PCs. However, in this case, rather than using the Gap method to determine the optimal number of clusters, the number of patterns was chosen as 51 to construct a set of regional weather patterns comparable to the output of the tiered analysis. Centroid maps are provided in Figure S2.

Elbow plots of inertia were also examined to test the sensitivity of the explanatory power to the number of patterns mandated. Figure 1 shows elbow plot for the flat patterns (black), and an adaptation of elbow plots for the tiered patterns (grey). The tiered adaption shows a pointed oval indicating the inertia of all combinations of numbers of tier-2 patterns between five and eight, for the fixed set of tier-1 regimes. The vertical extent of this ellipse indicates the sensitivity of the tier-2 pattern inertia to the proportion of patterns allocated to each tier-1 cluster when the total number of patterns is fixed. Wind and precipitation inertia are respectively calculated as

$$I_w = \frac{\sum_{t,x,y}\left[\left\{u-\widetilde{u}(r_t)\right\}^2+\left\{v-\widetilde{v}(r_t)\right\}^2\right]}{\sum_{t,x,y}\left\{(u-\overline{u})^2+(v-\overline{v})^2\right\}} \tag{1a}$$

$$\text{and } I_P = \frac{\sum_{t,x,y}\left\{P-\widetilde{P}(r_t)\right\}^2}{\sum_{t,x,y}\left(P-\overline{P}\right)^2}, \tag{1a}$$

where u and v are the 850 hPa eastward and northward winds respectively, P is precipitation, r_t is the regime or pattern which is current at time t, $\widetilde{v}(r)$ is the centroid of variable v during regime/pattern r and $\overline{v}$ is the time-mean of variable v. It is clear that the flat methodology is able to achieve a lower inertia with the same number of patterns than the tiered methodology when both winds and precipitation are considered (Figure 1). The same wind

inertia as the tiered patterns can be achieved by 32 flat patterns, and the same precipitation inertia is achieved by 39 flat patterns. This is due to redundancy present in the tiered patterns, originating from similar weather patterns occurring across different tier-1 regimes. For the tiered methodology, some weather types need to be represented by a separate tier-2 pattern for multiple tier-1 regimes, whereas in the flat methodology, one pattern will suffice.

All sets of patterns have been ordered according to their climatological seasonal cycle by computing the mean day of year (modulo 365 days) for each pattern and sorting by this value in ascending order.

For the purposes of this paper, heavy precipitation has been defined as occurring when GPM precipitation interpolated to a $1.5 \times 1.5°$ grid exceeds its seasonal 90th percentile. This P90 percentile value is calculated separately for each grid location and each day of the year using a 60-day rolling window. By definition, there is a 10% probability of heavy precipitation occurring on any given day at each location. In order to determine whether the pattern sets are useful for describing the risk of heavy precipitation, we wish to demonstrate whether the pattern-conditioned exceedance likelihood of the seasonal P90 is statistically different from 10%. Using the binomial distribution, we can construct a 5% one-sided confidence bound such that there is a <5% probability that the number of days with exceedance in an independent selection of 130 days (the average number of days in each pattern during the GPM period) exceeds this bound. By computation we find this bound to be 20 days, equivalent to a 15% exceedance rate. Therefore, when we present the pattern-conditioned exceedance of the seasonal P90 in Section 4, we set the contour interval levels to 10%, with the 5–15% exceedance rates unshaded.

3 | DESCRIPTION OF PATTERNS

This section describes the seasonal cycle, longevity and transitional behaviour of the three sets of weather regimes and patterns.

3.1 | Seasonal cycle

The eight planetary-scale regimes of the first tier discriminate primarily between phases of the seasonal cycle, and between the different monsoon circulations present in and around Southeast Asia. The timing of each tier-1 regime is shown in Figure 2a, while the associated ERA5 850 hPa wind and GPM precipitation regime centroids are shown in Figure 3. Figure 2b shows the probability of tier-1 regimes conditioned on the Niño 3.4 SST index. Regimes 1 and 2 are dominated by the austral monsoon with westerly winds and high precipitation in Indonesia, concurrent with the northeast monsoon in the Northern Hemisphere. These two regimes show the strongest relationship with ENSO in Figure 2, where Regime 2 dominates during La Niña years and Regime 1 coincides with El Niño years. These westerly winds are stronger over the equatorial Indian Ocean and Sumatra in Regime 2 than in Regime 1 (Figure 3a, b). Regime 3 represents the transition season as the Inter-Tropical Convergence Zone (ITCZ) passes north over the Equator. Following this, the boreal monsoon takes hold with the onset of Regime 4, with westerly winds extending from India through to Vietnam in the regime centroids (Figure 3d). Regime 4 is the longest-lasting tier-1 regime identified, typically occurring between late May and early September. However, from July onward the regime frequently switches

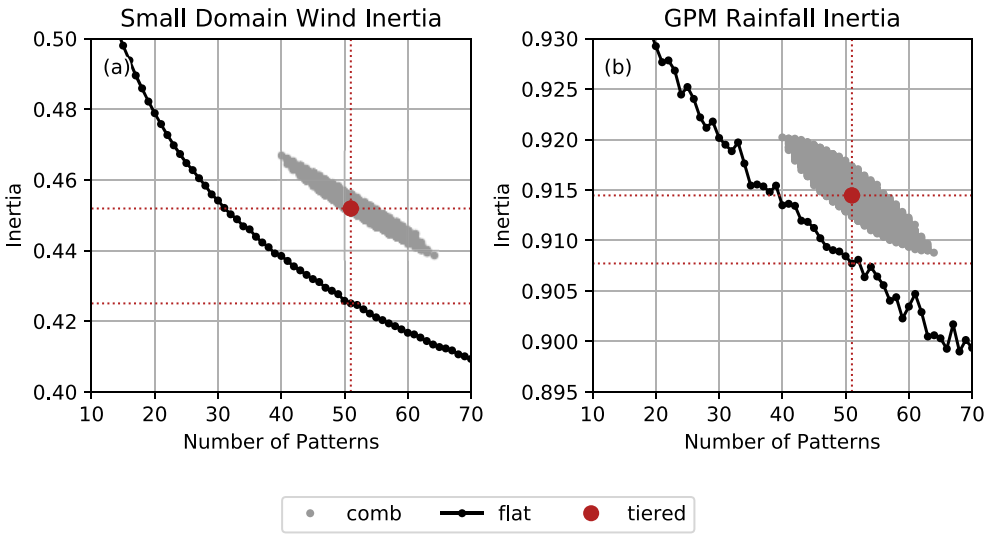

FIGURE 1 Inertia of various pattern sets calculated using (a) small-domain daily 850 hPa wind vectors and (b) GPM precipitation and plotted against total number of patterns. Black line: Inertia of 15–70 flat patterns, calculated using the flat pattern methodology. Grey dots: Inertia of all possible combinations with 5–8 sub-patterns in each tier-1 pattern. The same tier-1 pattern decomposition has been used in each case. Red dot: Inertia of the selected set of tiered regimes [Colour figure can be viewed at wileyonlinelibrary.com]

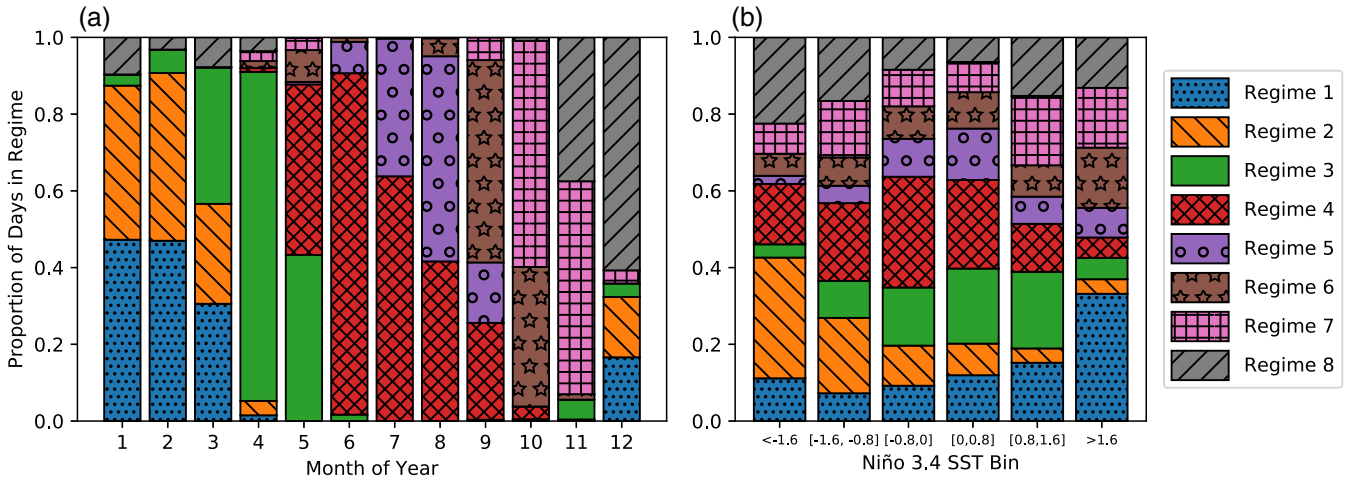

FIGURE 2 Climatological proportion of days in (a) each month and (b) Niño 3.4 SST bins assigned to each tier-1 regime [Colour figure can be viewed at wileyonlinelibrary.com]

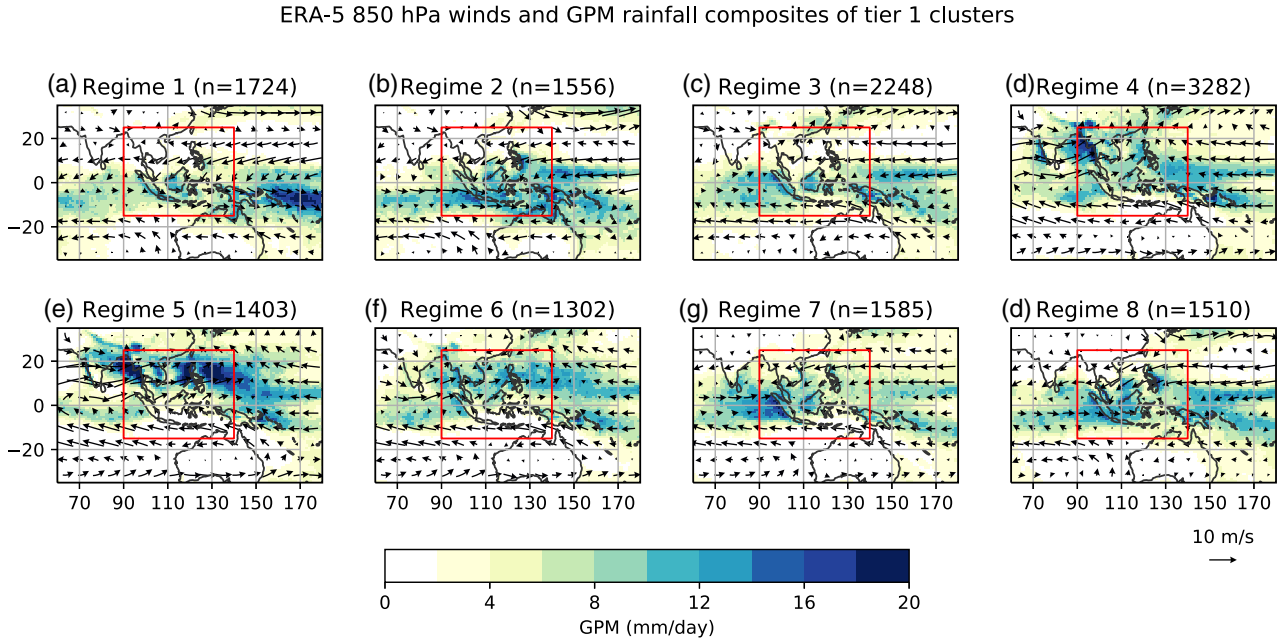

FIGURE 3 Tier-1 regime centroids. Vector wind at 850 hPa and GPM precipitation (shading) [Colour figure can be viewed at wileyonlinelibrary.com]

between itself and Regime 5, which represents an eastward extension of the monsoon westerlies over the South China Sea and the Philippines (Figure 3e). From mid-September to mid-December, three regimes which describe the transitional season have been identified. These are Regime 6 being the retreat of the boreal monsoon, Regime 7 containing the southward passage of the ITCZ across the Equator, and Regime 8 representing the onset of the austral monsoon.

During the transitional seasons, further dependence on the seasonal cycle is present in the tier-2 patterns. This is evident from Figure 4a, which shows the seasonal frequency of tier-2 pattern occurrence by day of year. A distinct diagonal distribution, which is indicative of strong seasonality, is present in the weather patterns associated with Regimes 3, 6, 7 and 8. However within the two monsoon seasons (tier-1 Regimes 1, 2, 4 and 5), the seasonal influence is reduced and tier-2 pattern frequencies are more evenly spaced within the timing envelopes of the tier-1 regimes.

The seasonal cycle of flat pattern frequencies is given in Figure 4b. Patterns 1–9 and 51 occur during the austral summer, while the northward transition season is divided into patterns 10–17 and Patterns 20 and 21. The boreal

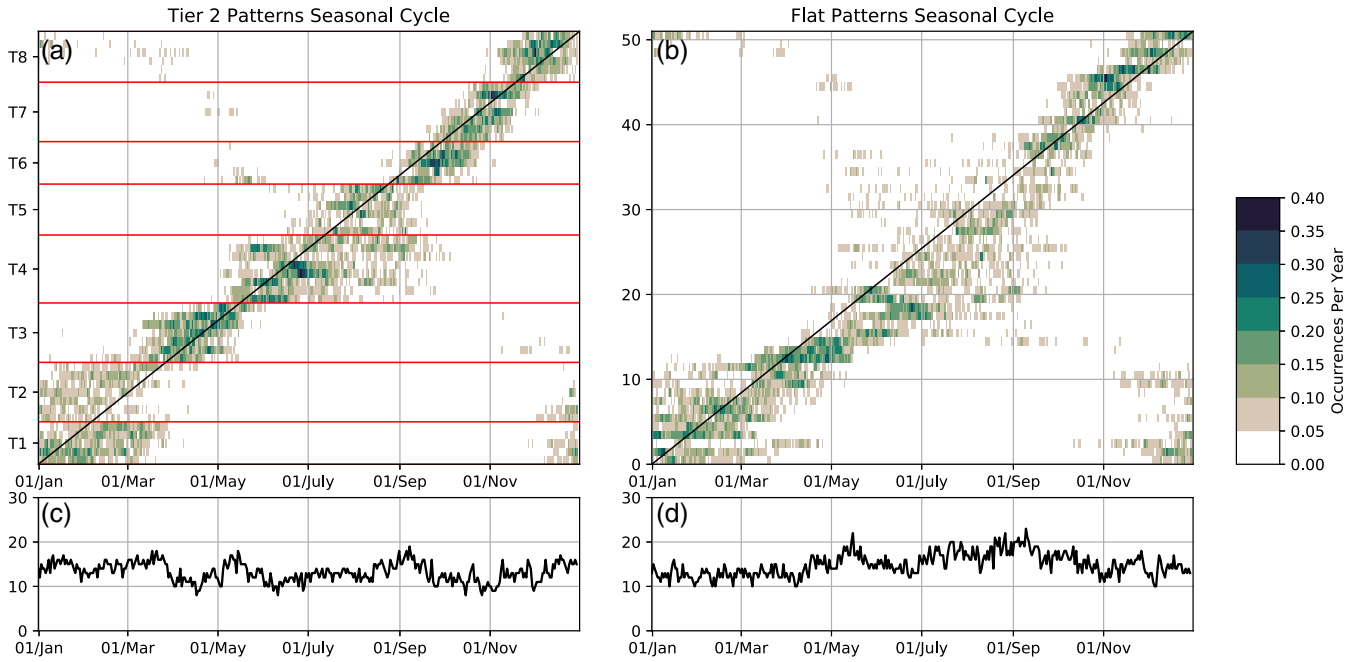

FIGURE 4 (a, b) seasonal frequency of each pattern (vertical axis), by day of the year (horizontal axis). Units are occurrences per year. (c, d) number of patterns that are likely to occur during rolling 7-day climatological window (using a 98% cumulative probability). (a, c) show tiered patterns, and (b, d) flat patterns [Colour figure can be viewed at wileyonlinelibrary.com]

monsoon occurs during patterns 18–36, with Patterns 18, 19, 22 and 23 occurring at the time of monsoon onset. The southward transition season then occurs with the progression from Pattern 37 through to Pattern 50. The diagonal structure of the figure indicates that, as in the tier-2 patterns, seasonal coherence is more present during the transition seasons than during the monsoon seasons. The density of flat patterns is greatest during the boreal monsoon, indicated by the location of the main band below the diagonal line. This is not the case for the tiered patterns.

3.2 | Pattern transitions and centroids

Statistics relating to the persistence of regime and pattern events are presented in Table 1. An event is defined as a spell of consecutive days in the same cluster. The distribution of event persistence is heavily negatively skewed. We find that the tier-1 regimes persist for 7 days on average, while the persistence of the tier-2 and flat patterns is very similar, both having a mean of 2.5–2.6 days. The median pattern persistence for the tier-2 and flat patterns is 2 days. With this median in mind, the empirical 2-day transition probabilities for the tiered and flat weather patterns are described by Figure 5a, b respectively. Consistent with the increased persistence of the tier-1 regimes, the majority of tier-2 patterns transition within the same tier-1 regime. Due to the ordering of the regimes and patterns by their seasonal cycles, a transition from a classification with an earlier seasonal timing to one with a later timing is denoted by a grid-cell below the diagonal. Patterns that occur during the transition seasons, as described in Section 3.1, generally have greater persistence (demonstrated by more probable transitions into the same pattern, along the diagonal), while the least persistence is present in the weather patterns associated with Regime 5 and in flat patterns 32–37.

TABLE 1 Statistics describing the number of consecutive days spent in each regime/pattern. P90 and P99 denote the 90th and 99th percentiles

	Mean	Median	P90	P99
Tier 1	7.0	3	18	45
Tier 2	2.6	2	5	11
Flat	2.5	2	5	11

Figure 6a, b indicates which of the tiered (a) and flat (b) patterns are closest to each other in the high-dimensional vector space containing the wind vectors at each grid-cell in the regional-scale domain[1]. Euclidean distances are normalised by the average Euclidean distance between the wind field at each point in time and the climatological mean wind field. Lines are drawn between markers indicating pairs of patterns for which this normalised

[1]In this instance, the vector space spanned by the principal components has not been used because the tier-2 patterns belonging to different tier-1 regimes were defined using different principal components.

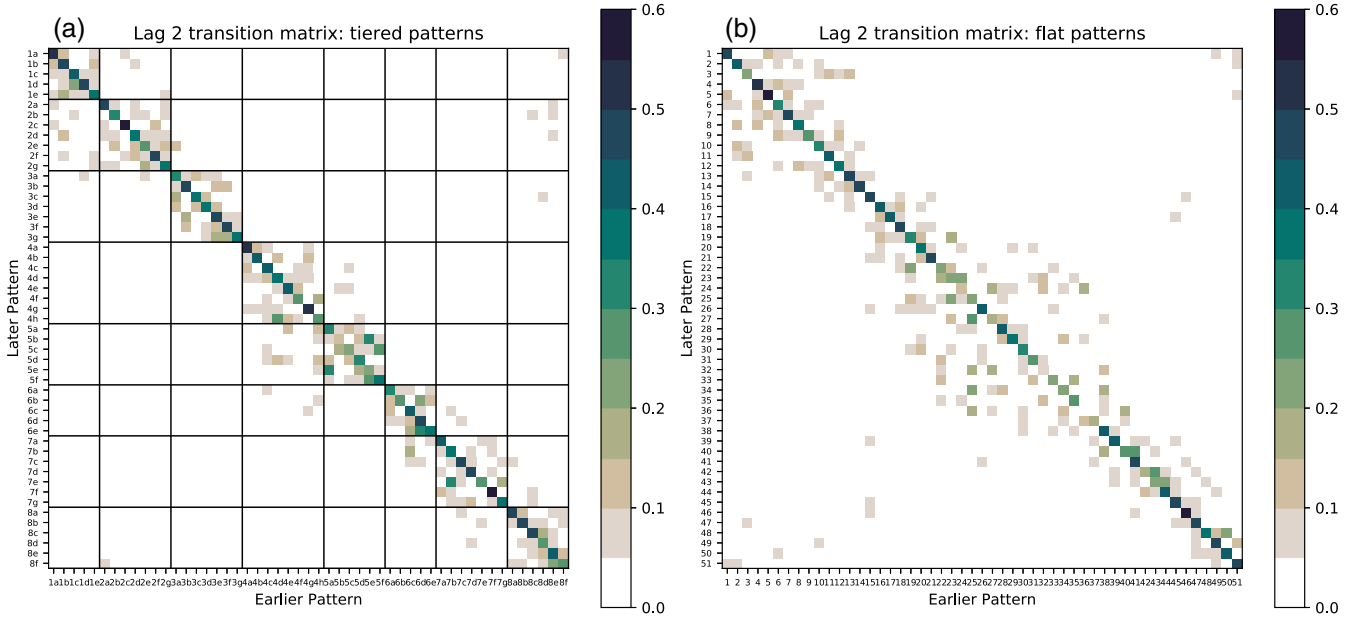

FIGURE 5 (a) Tiered and (b) flat pattern 2-day lagged transition matrix depicting the proportion of days that fall into each pattern (on the vertical axis) two days after the occurrence of the pattern indicated on the horizontal axis [Colour figure can be viewed at wileyonlinelibrary.com]

Euclidean distance between the centroids is less than 40%. In both sets of patterns, the boreal summer monsoon (tiered Regimes 4, 5 and 6; flat Regimes 18–36) possess centroids that are pair-wise close to each other. The tiered patterns show similar overlap for the austral summer monsoon (tiered Regimes 8, 1 and 2), however the flat patterns do not. This overlap is indicative of the redundancy of the tiered patterns described in Section 2.2, which was a consequence of tier-2 patterns in different tier-1 regimes representing the same weather type. The tiered regime set also contains pairs of close regimes which occur during different transition seasons (3b with 7f, 3c with 8d, 3f with 7b), whereas the flat regimes do not.

Figure 6c, d show the pairs of tiered patterns and flat patterns for which at least 30% of the days classified into the flat pattern were also classified into the tiered pattern. In the discussion below, we refer these pairs of patterns as 'linked'. Shading indicates the proportion of the flat pattern classified into the tiered pattern. 37 of the flat patterns are linked to a single tier-1 pattern, and the remaining 14 patterns are linked to two. Tier-1 Regime 4 is linked to the largest number of flat patterns, with 11 flat patterns linked only to this tier-1 regime and 4 linked to this tier-1 regime as well as one other tier-1 regime. This may be related to the construction of tier-1 Regime 4, which was the result of merging two clusters which featured a subtropical wave in quadrature in the planetary-scale domain. Across the tier-2 and flat patterns, 26 pattern pairs show a direct 1–1 correspondence, whereby each pattern is linked only to its partner. Many of these pairs will be found in Section 4.2 to feature similar modes of synoptic variability.

4 | REPRESENTATION OF MODES OF TROPICAL VARIABILITY

This section quantifies the degree to which each set of regimes and patterns distinguishes planetary-scale and synoptic modes of atmospheric variability. First we consider the representation of large-scale variability, including planetary-scale variability outside the regional domain. Secondly, synoptic variability, including weather events such as cold surges, tropical cyclones and equatorial waves is discussed.

4.1 | Large-scale variability

The following analysis compares the degree to which each set of regimes and patterns describes the variance of known modes of variability: ENSO, the IOD, the MJO and the seasonal cycle. We consider the variance of a set of climate index time-series explained by the respective sets of patterns. In this instance, variance explained is defined as:

$$\frac{\sum_t \left\{\widetilde{X}(r_t) - \overline{X}\right\}^2}{\sum_t \left(X - \overline{X}\right)^2}, \tag{2a}$$

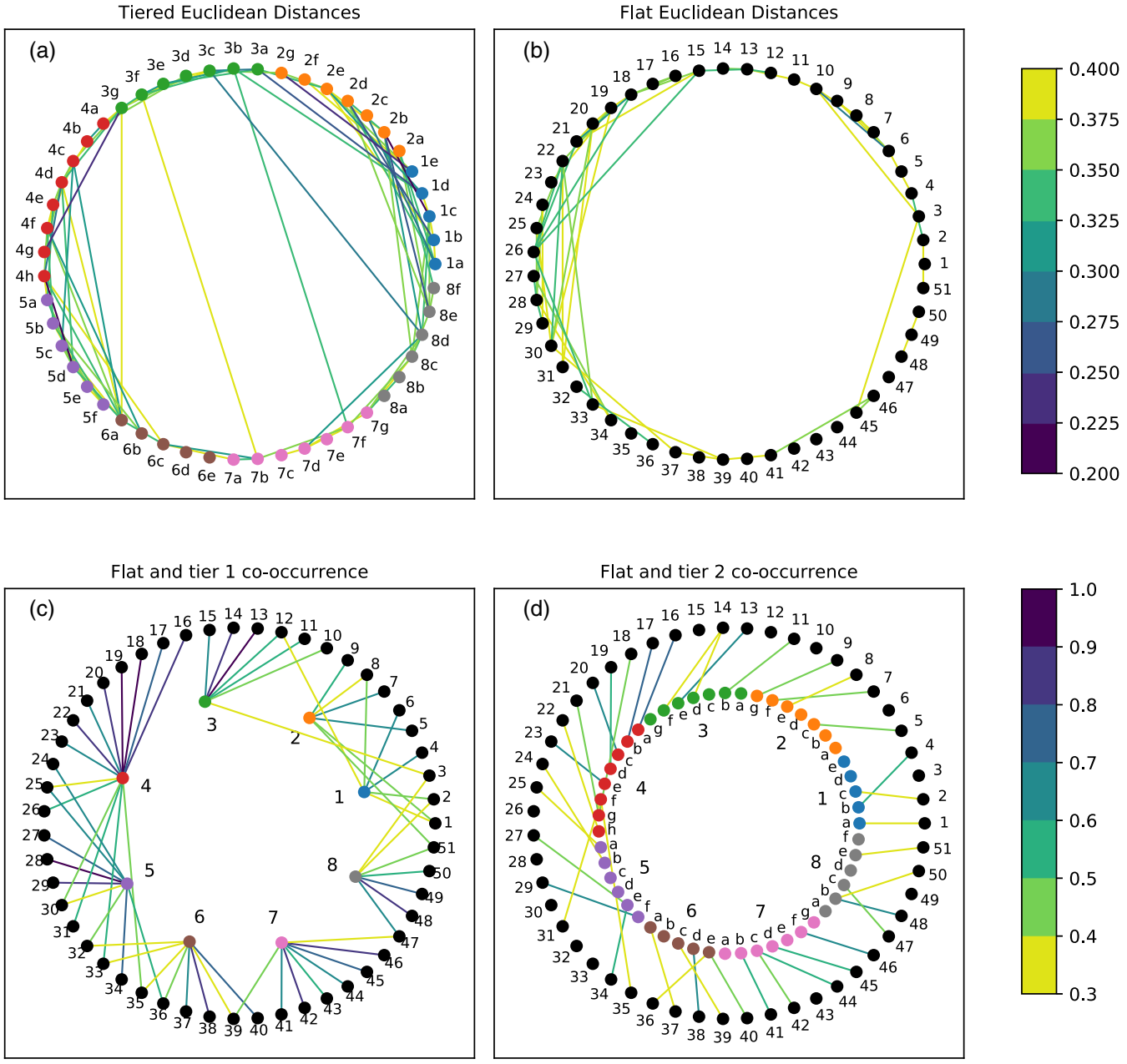

FIGURE 6 (a, b) show loom plots of pattern pairs with short pairwise Euclidean distances (a tiered, b flat). Line shading indicate Euclidean distances normalised by the mean Euclidean distance to the climatological mean. Patterns pairs with normalised distances greater than 0.4 are not shown. (c, d) show loom plots of the cross-over of (c) tier-1 regimes and (d) tier-2 patterns with the flat patterns. Line shading indicates the proportion of flat patterns that are also classified as the linked tiered pattern. Pattern pairs with overlap proportions below 0.3 are not shown [Colour figure can be viewed at wileyonlinelibrary.com]

$$\text{or} \quad \frac{\sum_t \left[\left\{ \widetilde{X_1}(r_t) - \overline{X_1} \right\}^2 + \left\{ \widetilde{X_2}(r_t) - \overline{X_2} \right\}^2 \right]}{\sum_t \left[\left(X_1 - \overline{X_1} \right)^2 + \left(X_2 - \overline{X_2} \right)^2 \right]} \tag{2a}$$

where X is the climate index, $\overline{X}$ is the climate index mean and $\widetilde{X}(r_t)$ is the mean of the climate index calculated over all days classified into classification r_t. If the climate index is two-dimensional, such as the real-time multivariate MJO (RMM), the expression (2b) is used, with $X = [X_1, X_2]$ representing the components of the index. The variance explained of each climate index is presented in Table 2 for the tier-1 regimes, for the tier-2 patterns and for the flat patterns in each column. The indices representing large-scale variability, presented in the first section of the table, are the seasonal cycle, the Niño 3.4 SST Index, the Dipole Mode Index (DMI), and the RMM. As well as the MJO, seasonal cycle is taken to be two-dimensional,

TABLE 2 Summary statistics describing the relationship between the tiered and flat patterns and modes of variability

		Tier-1	Tier-2	Flat
1	Day of Year	0.721	**0.748**	0.668
2	Niño 34 SST	0.066	**0.152**	0.128
3	DMI	0.031	**0.100**	0.088
4	RMM MJO	0.077	**0.247**	0.237
5	BSISO-1	0.099	0.339	**0.367**
6	BSISO-2	0.032	0.167	**0.187**
7	Cross-equatorial Surge	0.064	0.283	0.251
8	Easterly Surge	0.256	0.403	0.430
9	Meridional Surge	0.187	0.303	0.33
10	Merid.+East. Surge	0.066	0.144	0.167
11	Borneo Vortex	0.345	0.408	0.396
12	Tropical Cyclone	0.137	0.220	0.237
13	SCS Vortex	0.440	0.494	0.479
14	Kelvin Wave	0.027	0.049	0.049
15	Rossby-1 Wave	0.063	0.101	**0.110**
16	Rossby-2 Wave	0.050	0.090	0.099
17	WMRG Wave	0.050	0.086	**0.093**

Note: Rows 1-4: large-scale variability, rows 5-17 synoptic variability.Each entry indicates the fractional variance of a daily time-series that is explained using the pattern-conditioned means of each index using the tiered or flat pattern decomposition. The variance for the MJO, BSISO and the day of year are calculated in two-dimensional space, with the MJO and BSISO considered in EOF space and the day of year first projected onto a circle. The tier-2 or flat pattern set explaining the most variance for each index, where statistically significant for at least one component at a $p < 0.01$ level using both Bartlett and Levene tests, is shown in **bold**.

defined as:

$$[X1,\ X2] = [\cos(2\pi D/365.25),\ \sin(2\pi D/365.25)],$$

where D is the day of the year. This ensures that the index may be considered modulo the number of days in a year.

Though the first tier does distinguish El Niño and the MJO, inclusion of the second tier markedly improves the variance explained, likely due to the finer discretisation of tier-2. Across these large-scale indices, the tier-2 patterns have a significantly greater variance explained than the flat patterns do. This improvement can be directly linked to the conditioning on the tier-1 regimes: when days are assigned directly to tier-2 centroids (not shown), no significant improvement compared to the flat regimes was found. Owing to the extra information provided by the planetary-scale domain, tier-1 also has a higher explained variance for the seasonal cycle than the flat regimes. This implies that the tiered patterns are more skilful at representing large-scale variability. Although this does not translate into an improved discrimination of the precipitation variability, it may lead to the tiered patterns being more predictable than the flat patterns.

The relationship between the first tier regimes and ENSO is demonstrated in Figure 7, which shows the timing of each regime in each year from 1979–2018 by a coloured legend. Years have been reordered using annual March–March Niño 3.4 SST index, so that La Niña years are shown towards the bottom of the figure and El Niño years are shown towards the top. This means that the footprints in Figure 7 of regimes whose annual frequency is dependent on ENSO will be tapered, while the footprint of those whose timing is dependent on ENSO will be tilted.

The interannual frequency and timing of each regime has also been related to annual Niño 3.4 SST indices using regression analysis. For each regime, a Niño 3.4 SST was generated by weighting the monthly Niño 3.4 SST index by the climatological monthly regime frequency (according to Figure 2). Statistics for Regimes 1, 2 and 8 were calculated by splitting years in June, rather than January. The following text describes the direction and the correlation coefficient for the significant tier-1 regime dependencies on ENSO. All results were significant at the $p < .01$ level, unless otherwise stated. It is evident from Figure 7 that the strongest influence of ENSO on tier-1 regime assignment is present during the austral summer, with Regime 1 occurring more frequently during El Niño ($r = 0.76$), and Regime 2 occurring more frequently during La Niña ($r = -0.72$). Transition Regime 3 occurs earlier ($r = 0.35, p = 0.03$) and more frequently ($r = 0.48$) during El Niño. During the boreal monsoon, Regimes 5 and 6 are more frequent during El Niño ($r = 0.48$, $r = 0.68$), while Regime 4 is more frequent during La Niña ($r = -0.57$). During the southward monsoon transition, Regime 7 occurs earlier in the season and Regime 8 is more frequent during La Niña ($r = 0.50$, $r = 0.36$, respectively).

The distribution of MJO phases observed during each tier-1 regime is shown in Figure 8. In this figure, statistically significant increased (decreased) co-occurrences compared to seasonal climatologies are indicated by blue (red) dots. These were calculated by generating a 1,000-member bootstrapped sample of 1,000 days from the set of days assigned to each pattern and determining whether the climatological event occurrence lay within the bootstrapped P99 confidence interval of synoptic event frequencies. For each weather regime, the climatological synoptic event occurrences were calculated separately using a sample with replacement of 10,000 days whose days and years were assigned randomly and whose months were distributed identically to that of the weather pattern. The strongest relationship with the MJO is apparent in

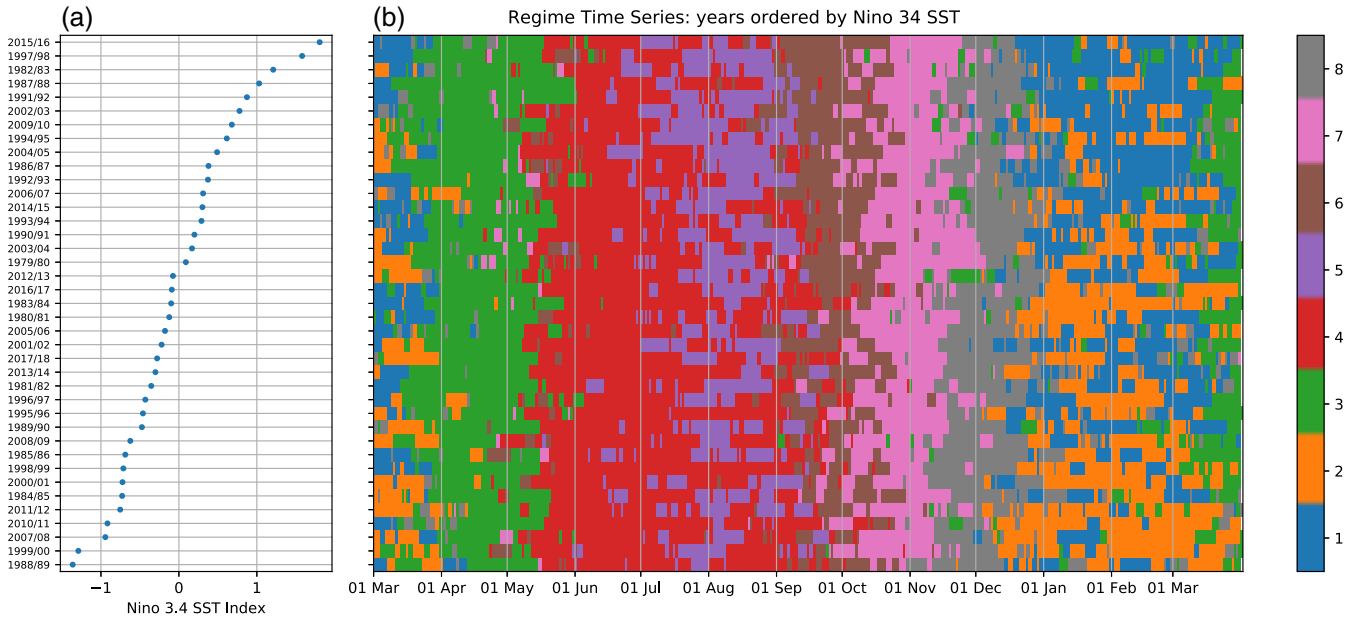

FIGURE 7 Tier-1 regime timing, reordered by ENSO. (a) shows the annual Niño 3.4 SST index (x−axis), averaged from March to March in each year from 1979 to 2018. Years (y−axis) have been reordered such that this index is increasing. (b) shows the regime assigned on each year between 1979 and 2018, ordered by the annual Niño 3.4 SST index. The x−axis shows the day of the year, starting in March and ending in April the following year. The y−axis shows the year, reordered by the Niño 3.4 SST index as in (a) [Colour figure can be viewed at wileyonlinelibrary.com]

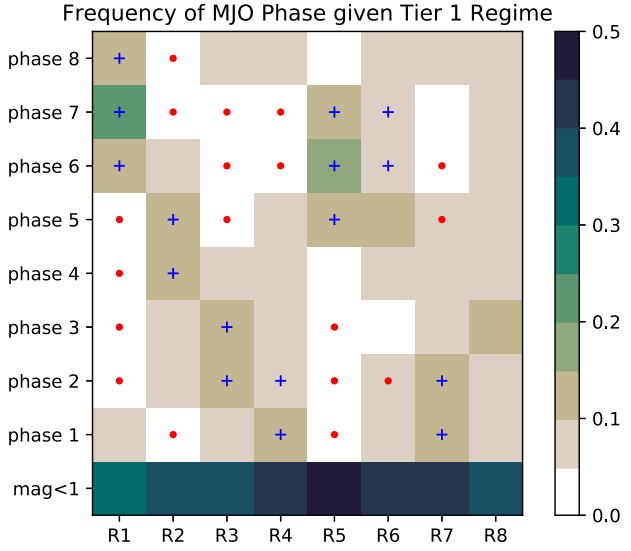

FIGURE 8 Frequency of MJO Phase conditioned on tier-1 regime. The x−axis is the Tier-1 regime, and the y−axis the MJO phase, calculated using RMM indices (Wheeler and Hendon, 2004). Plusses (dots) indicate that the event is significantly more (less) likely to co-occur with the assigned pattern than otherwise during the season in which the pattern occurs, based on a bootstrapped 99% confidence interval. [Colour figure can be viewed at wileyonlinelibrary.com]

Regimes 1, 2 and 5. Combined with our findings from Figures 2 and 7, we may conclude that the core large-scale variability of the austral summer monsoon has been divided into two categories: Regime 2, where the monsoon is enhanced by La Niña and locally active phases of the MJO (phases 3–5), and Regime 1, where the monsoon is subdued by El Niño and locally suppressed phases of the MJO (Phases 7–1). Meanwhile, during the boreal monsoon, the extension of the monsoon westerlies into the South China Sea, represented by Regime 5, occurs more frequently during MJO phases 5, 6 and 7, when the MJO is active over the western Pacific.

4.2 | Synoptic-scale variability

In order to quantify the degree to which the tier-2 and flat patterns distinguish synoptic-scale variability, we study the pattern-conditioned probability of the occurrence of synoptic weather events known to impact weather in Southeast Asia. We consider cold surges (Chang *et al.*, 2005; Hattori *et al.*, 2011), tropical cyclones (Hodges *et al.*, 2017), the MJO (Wheeler and Hendon, 2004), the Borneo vortex (Chang *et al.*, 2005), the BSISO (Lee *et al.*, 2013), vortices in the South China Sea (Dang-Quang *et al.*, 2016) and equatorial waves (Yang *et al.*, 2003), and use objective event definitions from previous studies. The following list describes the event definitions used to calculate the co-occurrence between synoptic events and the weather patterns. A weather index is also defined for each event in the list below and the lower sections of Table 2 indicate the

variance of these weather indices explained by the tier-1, tier-2 and flat classification sets.

- Cold surges and the Borneo vortex (Chang *et al.*, 2005; Hattori *et al.*, 2011). In Table 2, rows 7–11 indicate the variance of the binary indices explained by the patterns.
  - Meridional: ERA-Interim 925 hPa northerly wind averaged between 110° and 117.5°E along 15°N exceeds 8 $m \cdot s^{-1}$;
  - Easterly: ERA-Interim 925 hPa easterly wind averaged between 7.5° and 15°N along 120°E exceeds 8 $m \cdot s^{-1}$;
  - Both Meridional and Easterly: both of the above conditions are met;
  - Cross-equatorial: ERA-Interim 925 hPa northerly wind averaged over 105–115°E at 5.0°S exceeds 5 $m \cdot s^{-1}$; and
  - Borneo vortex: ERA-Interim 925 hPa stream function calculated using Dawson (2016) has a local minimum within the area of 2.5°S–7.5°N, 107.5°–117.5°E, and at least one grid-cell adjacent to the local minimum has a wind speed exceeding 2.5 $m \cdot s^{-1}$.
- Tropical cyclones (Hodges *et al.*, 2017): A cyclonic vortex with filtered relative vorticity exceeding a range of thresholds ($5–24 \times 10^{-5} s^{-1}$) is present in the Southeast Asia domain plus a 5° buffer. The cyclonic vortices were identified by objectively tracking vertically averaged (850–600 hPa) ERA-Interim relative vorticity at T63 resolution at 6-hourly time resolution in the Northern Hemisphere only. The objective tracking methodology described by Hodges *et al.* (2017) was applied. Here, the relative vorticity threshold was chosen so that it corresponds to the 40-year P90 of the daily maximum filtered vorticity across the domain. In Table 2, row 11 indicates the variance of the binary index of this vorticity exceeding its P90 ($1.3 \times 10^{-4} s^{-1}$) explained by the patterns.
- The MJO (Wheeler and Hendon, 2004): The RMM-based MJO is in each of its eight phases, with magnitude greater than 1. Weak MJO, with magnitude less than 1, is also considered as a separate event (denoted by W for Weak). In Table 2, row 4 indicates the variance of the 2D RMM indices explained by the patterns.
- South China Sea vortex: A cyclonic vortex with filtered vorticity exceeding $2 \times 10^{-5} s^{-1}$ in the region (100°–120°E, 0°–12°N). The cyclonic vortices were identified by tracking low-level (850 hPa) ERA-Interim vorticity at T63 resolution at 6-hourly time resolution. In Figure 9d, these vortices are binned by their central latitude. In Table 2, row 13 indicates the variance of the binary index of the tracked vorticity exceeding its P90 ($4.1 \times 10^{-4} s^{-1}$) explained by the regimes. South China Sea vortices towards the southern extent of this range may also be classified as Borneo vortices. However, this tracking methodology differs from the Borneo vortex stream-function approach above in that it is designed to capture propagating systems, whereas the stream-function approach also identifies stationary systems.
- Equatorial Waves (Yang *et al.*, 2003): the amplitude of an equatorial wave mode at a chosen latitude, averaged from 90° to 140 °E, exceeds a chosen threshold. Here, the amplitude for each wave mode is calculated as the Hilbert envelope of the 850 hPa winds projected onto the parabolic cylindrical basis functions associated with the horizontal structure of the equatorial wave. The threshold, wind component and latitude are chosen separately for each wave mode, as follows. The formulation of the projection methodology ensures that the results are not dependent on choice of latitude. In Table 2, rows 14–17 indicate the variance of the respective filtered wave amplitudes explained by the patterns.
  - (K) Kelvin wave: Filtered eastward wind amplitude at the Equator exceeds its P90 (2.0 $m \cdot s^{-1}$).
  - (R1) $n = 1$ Rossby wave: Filtered northward wind amplitude at 10° S exceeds its P90 (1.8 $m \cdot s^{-1}$).
  - (R2) $n = 2$ Rossby wave: Filtered northward wind amplitude at the Equator exceeds its P90 (1.1 $m \cdot s^{-1}$).
  - (WM) Westward mixed Rossby-gravity wave: Filtered northward wind amplitude at the Equator exceeds its P90 (2.0 $m \cdot s^{-1}$).
  - BSISO-1 (Lee *et al.*, 2013): the first mode of BSISO is in each of its eight phases, with magnitude greater than 1. Weak BSISO, with magnitude less than 1, is also considered as a separate event (denoted by W for Weak).
  - BSISO-2 (Lee *et al.*, 2013): the second mode of BSISO is in each of its eight phases, with magnitude greater than 1. Weak BSISO, with magnitude less than 1, is also considered as a separate event (denoted by W for Weak). In Table 2, rows 5 and 6 indicate the variance of the two sets of 2D BSISO EOF indices explained by the patterns.

For each event, the conditional probability of event occurrence during each tiered and flat weather pattern is displayed in Figures 9 and 11. Statistical significance was calculated according to Figure 8, with bootstrapped sample sizes reduced to 200 days to reflect the approximate number of days in each regime. Event seasonality is apparent; tropical cyclones (Figure 9b) only occur with

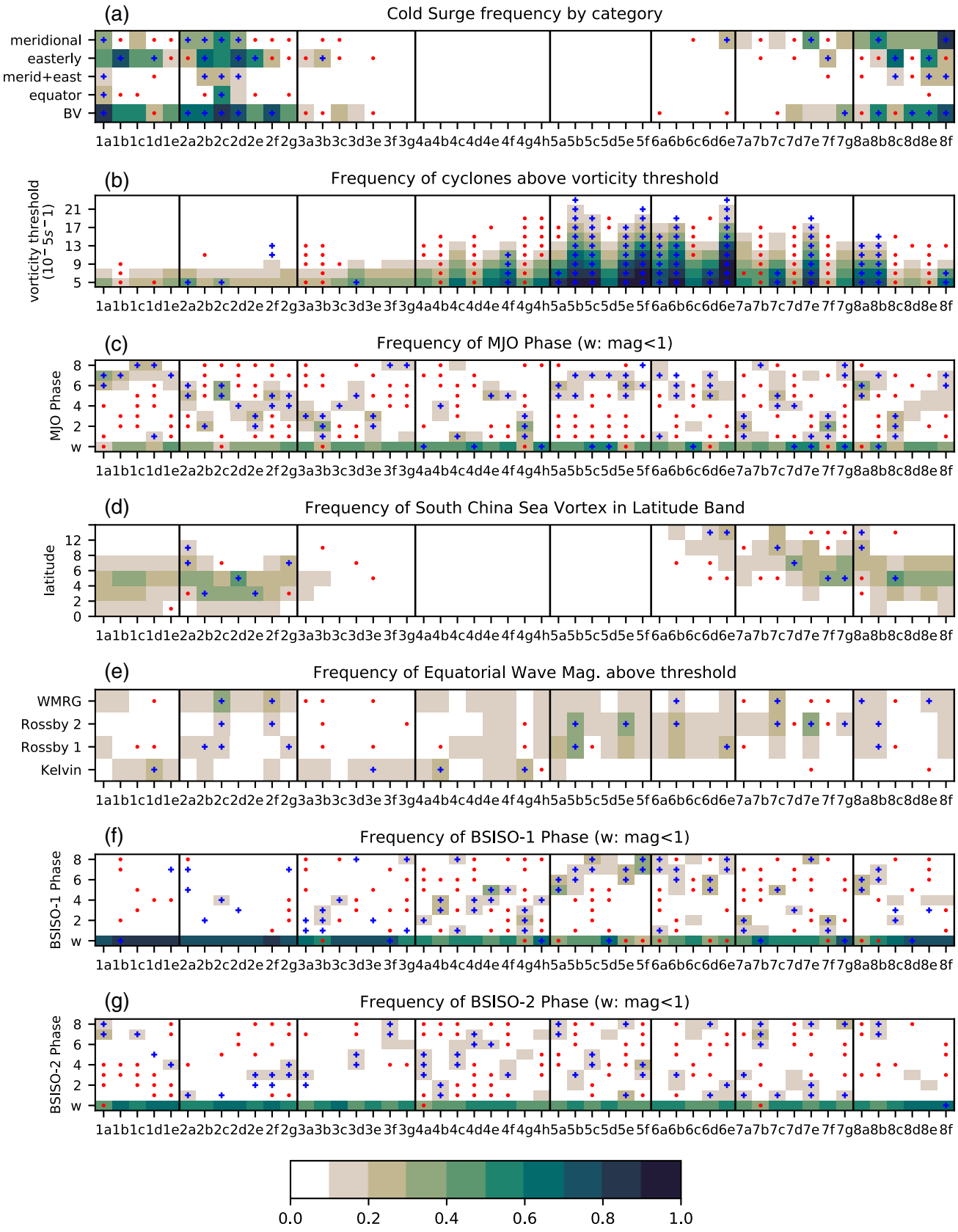

FIGURE 9 Frequency at which events defined in Section 4 occur in each tier-2 pattern. Panels indicate different event types, as labelled. Thresholds are as defined in the text. Plusses (dots) indicate a that the event is significantly more (less) likely to co-occur with the assigned pattern than otherwise during the season in which the pattern occurs, based on a bootstrapped 99% confidence interval [Colour figure can be viewed at wileyonlinelibrary.com]

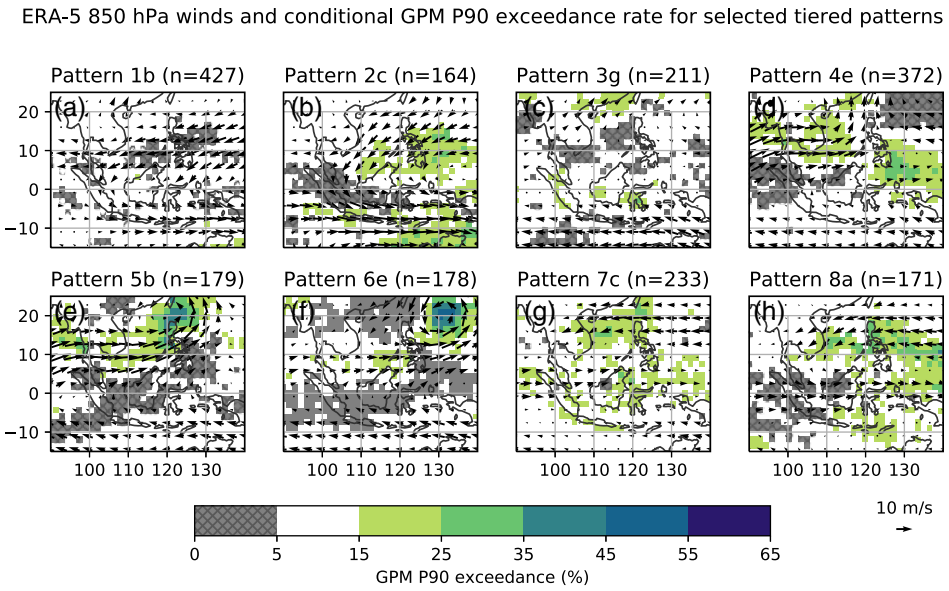

FIGURE 10 850 hPa wind vectors (arrows) and probability (colour shading) of exceeding the 20 year annual P90 precipitation on a 1.5 × 1.5° grid, for selected patterns from the tiered methodology. Exceedances in regions where the mean seasonal P90 during each pattern is less than 10 mm·day^{-1} are masked. Key phenomena identified in each pattern are: easterly surges (1b), cross-equatorial surges (2c), BSISO-1 phase 8 (3g), BSISO-1 phase 5 (4e), tropical cyclones (5b, 6e), and South China Sea vortices (7c, 8a) [Colour figure can be viewed at wileyonlinelibrary.com]

increased frequencies during the boreal monsoon patterns (tier-1 Regimes 4, 5, 6 and 7), while cold surges (Figure 9a) and Borneo vortices (Figure 9d) occur predominantly during the austral summer patterns (tier-1 Regimes 1, 2 and 8). Despite this, however, there is a considerable degree of variability displayed within each of tier-1 regimes, demonstrated by the increase of variance explained in Table 2 comparing tier-2 to tier-1. Since the flat patterns are ordered by their average day of year, the corresponding seasonality in the flat patterns show up as clusters of increased co-occurrence frequencies across adjacent columns of Figure 11.

To illustrate the links between the patterns and synoptic weather, we present several case-studies for a selected subset of the weather patterns. The subset was chosen to contain one tier-2 pattern from each tier-1 regime, and to highlight synoptic weather features with enhanced pattern co-occurrence rates, and a range of the seasonal cycle of flat regimes. The composite-mean 850 hPa winds and pattern-conditioned exceedance of the 20-year annual P90 GPM precipitation of the subset of tiered and flat patterns are presented in Figures 10 and 12). The full set of pattern centroids are given in Figures S1 and S2).

The classification of cold surges is clearly partitioned between the different austral summer patterns (Figure 9a and Figure 11a). A cross-equatorial cold surge in pattern 2c can be seen in the centroid of Figure 10b, together with elevated precipitation across Java and the southern islands of Indonesia. Meanwhile, the cyclonic circulation associated with a South China Sea vortex near 8°N is evident in the centroid of pattern 8a visible in the wind field shown in Figure 10h. Representation of cold surges in the flat pattern is similar, with pattern 48 featuring the strongest co-occurrence with meridional surges and patterns 2 and 5 showing the the strongest likelihood of cross-equatorial surges.

A total of 44 (47) of the tier-2 (flat) patterns co-occur and are associated with MJO phases (Figures 9c and 11c). During the boreal monsoon and transition season patterns, the frequencies of pattern-conditioned MJO phases typically match those of pattern-conditioned BSISO-1 phases. This is not surprising, since the first mode of BSISO often occurs in conjunction with the MJO during the boreal monsoon (Lee *et al.*, 2013). The EOF spatial patterns in the BSISO phase composites (figures 2 and 3 of Lee *et al.*, 2013) match up with the patterns of enhanced extreme precipitation likelihood in patterns with high increased BSISO phases. This is most apparent in tiered pattern 4e and flat pattern 23 (Figure 10d and Figure 12c) which feature BSISO-1 Phase 5 and show a band of high probability of heavy precipitation stretching out from Vietnam over the South China Sea with a northwest to southeast orientation. These regimes were identified as 'linked' patterns in Figure 5. Flat pattern 28 (Figure 12d) is associated with BSISO-1 phase 8 and shows high rainfall risk in the north of the domain and low rainfall risk in the centre near 10°N.

The patterns derived from tier-1 Regimes 5 and 6 are dominated by tropical cyclones (Figure 9b). Track frequency maps (not shown) indicate that the various patterns feature tropical cyclones in locations that match up with cyclonic vortices and regions of high probability of heavy precipitation in Figures 10e, 10f and S1. In pattern 5b, this region is located immediately over Luzon, while in pattern 6b it is located to the east of the Philippines. Meanwhile, nine boreal summer flat weather patterns feature the highest co-occurrence rates with tropical cyclones (Figure 11b). Preferential tropical cyclone locations for many of these patterns can once

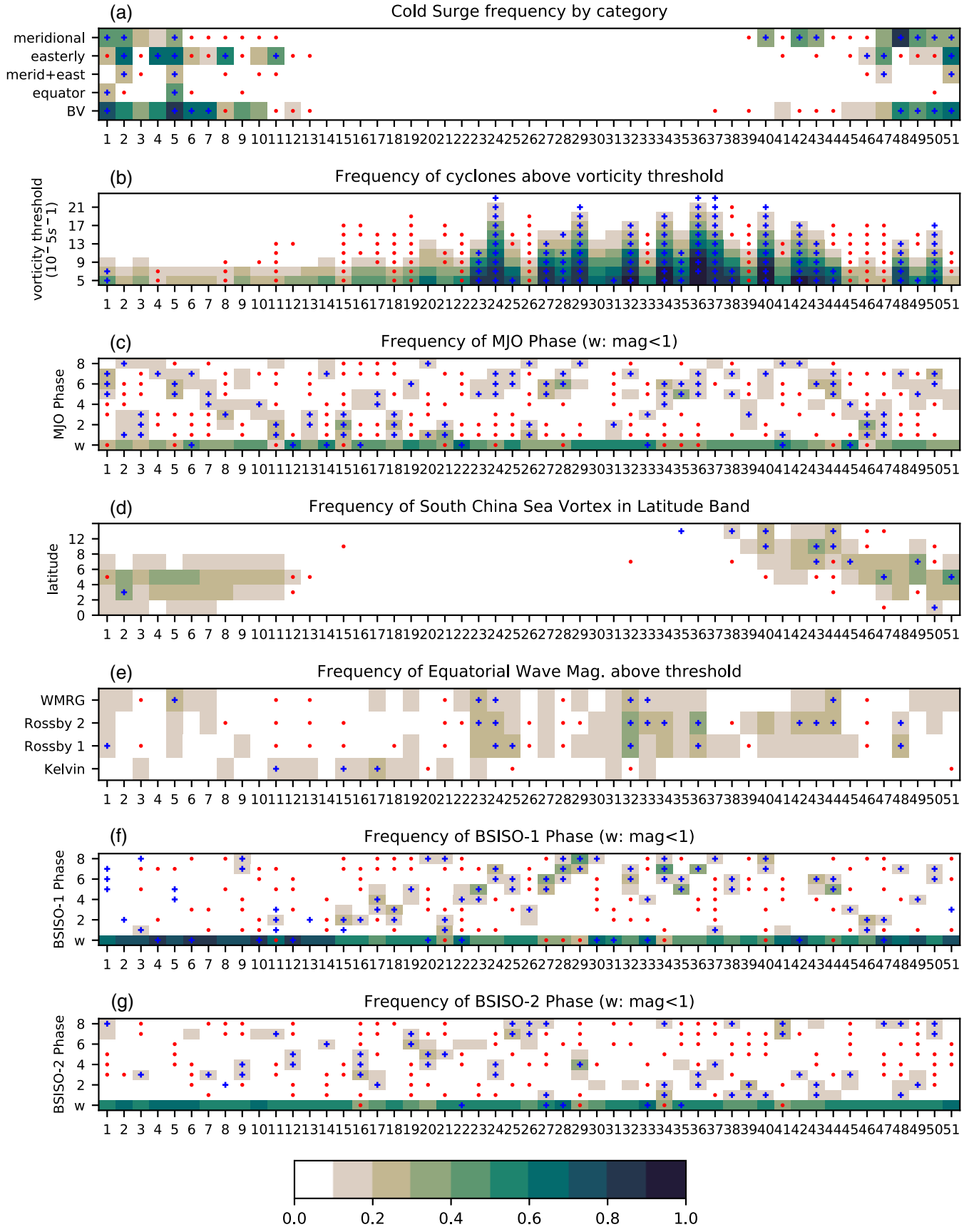

FIGURE 11 Frequency at which events defined in Section 4 occur in each flat pattern. Panels indicate different event types, as labelled. Plusses (dots) dots indicate a that the event is significantly more (less) likely to co-occur with the assigned pattern than otherwise during the season in which the pattern occurs, based on a bootstrapped 99% confidence interval. [Colour figure can be viewed at wileyonlinelibrary.com]

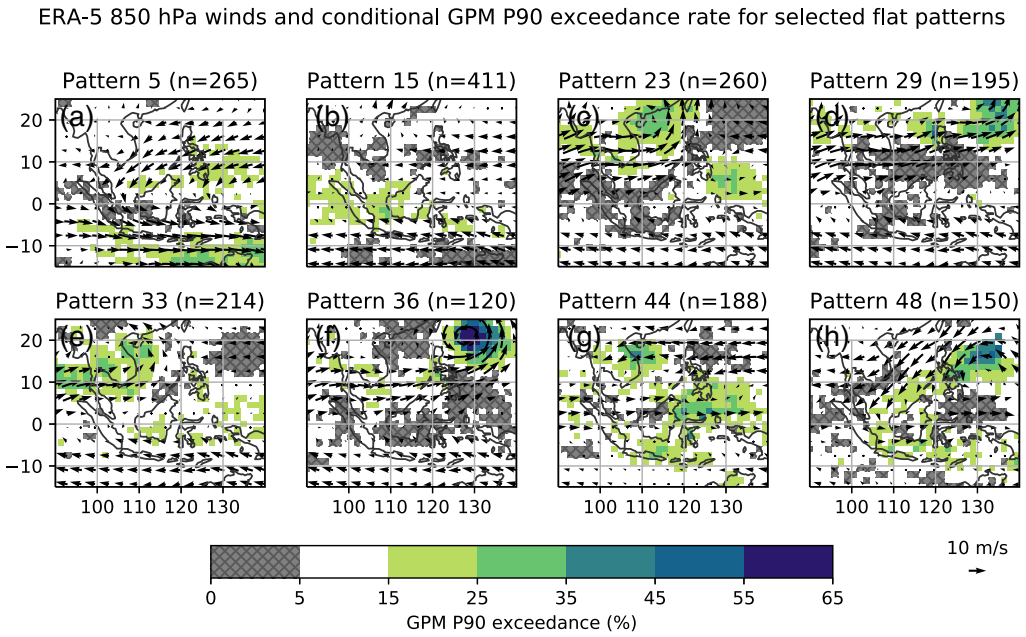

FIGURE 12 850 hPa wind vectors (arrows) and probability (colour shading) of exceeding the 20 year annual P90 precipitation on a 1.5 × 1.5° grid, for selected patterns from the flat methodology. Exceedances in regions where the mean seasonal P90 during each pattern is less than 10 mm·day^{-1} are masked. Key phenomena identified in each pattern are: cross-equatorial surges (5), BSISO-1 and MJO Phase 2 (15), BSISO-1 phase 5 (23), BSISO-2 phase 8 (29), equatorial Rossby waves (33), tropical cyclones (36), South China Sea vortices (44) and meridional surges (48) [Colour figure can be viewed at wileyonlinelibrary.com]

again be derived from Figure 12. The strongest tropical cyclone pattern was pattern 36, which is shown in Figure 12f. Some of these tropical cyclones appear to project onto the westward-propagating equatorial wave modes. Meanwhile, pattern 33 shows a low frequency of tropical cyclones but a high frequency of Rossby-2 waves (Figure 12e). During this pattern, a cyclonic vortex and elevated extreme precipitation likelihood in Vietnam are both evident from Figure 12e.

The variance in the equatorial wave-filtered wind amplitudes explained by the patterns is lower than that of the other phenomena considered. This is to be expected, since the waves are fast-propagating disturbances, while the patterns distinguish stationary features and last typically for 2–3 days. Some patterns that co-occur with tropical cyclones also feature increased occurrence frequency of a combination of Rossby-1, Rossby-2 and WM waves, which may be due to a projection of the tropical cyclones onto the equatorial wave modes.

5 | REPRESENTATION OF HEAVY PRECIPITATION

The two sets of patterns have been defined using 850 hPa winds since it is expected that this will be more accurately forecast by sub-seasonal forecast models, however the ultimate aim is to obtain a set of circulation patterns which are useful for forecasting extreme precipitation. In this section, we investigate the degree to which the weather patterns discriminate precipitation variability in Southeast Asia.

Figure 13 shows the precipitation variance explained over land GPM grid-cells. The total variance explained ranges from 0.1–0.4 over most land-based regions in the tier-2 and flat pattern sets. This value is consistent with Neal *et al.* (2020), who obtained values of 0.25–0.35 at grid-cells adjacent to Nilgiris and Darjeeling in their similar study of Indian weather patterns. A large degree of this variance explained results from the seasonal cycle, which is well represented by the weather patterns.

When the seasonal cycle is removed and the variance of seasonally anomalous precipitation is considered (Figure 13d–f), the variance explained by the tier-1 is close to zero and that explained by the tier-2 and flat patterns is reduced to 0.04–0.12 over the key regions of interest. The largest proportion of anomalous variance explained is located over the north of the Philippines. This signal has been attributed to the patterns associated with tropical cyclones. In this region, the flat patterns explain up to 5% more of the anomalous variance than the tiered regions. The flat patterns also perform better over Vietnam. In other regions, however, the variance explained by the tiered and the flat patterns is broadly comparable.

Analysis of Figure 13 is useful for understanding how the patterns discriminate precipitation variability

generally, but does not provide explicit information regarding heavy precipitation. For this, we instead consider the Brier Skill Score (BSS) of a hypothetical perfect patterns forecast. This answers the question: how much can be deduced about the likelihood of extreme precipitation on a given date from prior knowledge of the pattern this date belongs to? The BSS of a perfect pattern forecast may be considered to be an upper bound on the BSS of a pattern forecast, as presented in G21. This analysis calculates BSS using threshold exceedance of the seasonal P90. The BSS is calculated from the Brier Skill (BS) as follows:

$$\mathrm{BS}_R = \sum_{x,y,t} \left(O(x,y,t) - \underbrace{\frac{1}{n_r - 1}\left[n_r E\{x,y,r(t)\} - O(x,y,t)\right]}_{\text{Pattern conditioned exceedance excluding } t} \right)^2, \tag{3a}$$

$$\mathrm{BSS}_R = 1 - \frac{\mathrm{BS}_R}{\mathrm{BS}_{\mathrm{clim}}}, \tag{3a}$$

where R is the pattern set, O is the binary observed threshold exceedance, E is the probability of threshold exceedance conditioned on pattern $r(t) \in R$, and n_r is the number of days classified into pattern r that have precipitation data. The term denoted by an underbrace is the empirical probability of threshold exceedance conditioned on pattern $r(t) \in R$ calculated using all days excluding day t. Use of this term prevents spurious correlations between O and E due to potential under-sampling. The fraction in the right term of Equation (3a) acts to ensure that the observation being considered is excluded from the conditional exceedance calculation. The climatology is calculated using a 60-day rolling window and the 18 years of GPM data.

To aid with the interpretation of the BSS, a fourth set of weather patterns has been constructed using the MJO index. Each day was assigned a MJO pattern based on the season (separated into December–February, March–April, May–July and August–November) and the phase of the MJO, with weak MJO considered as a separate phase. The MJO-conditioned exceedance of the seasonal P90 was then calculated using 36 MJO patterns. The resultant exceedance maps are consistent with the well-known signal of the MJO influence on precipitation (Wheeler and Hendon, 2004). Since the MJO is frequently used as a forecasting tool in Southeast Asia, this provides a useful benchmark for the performance of the weather pattern-based forecasts.

Spatial maps of the annual BSS were found to feature patterns that mirrored the explained variance of the anomalous precipitation (Figure 13d–f), and so have not

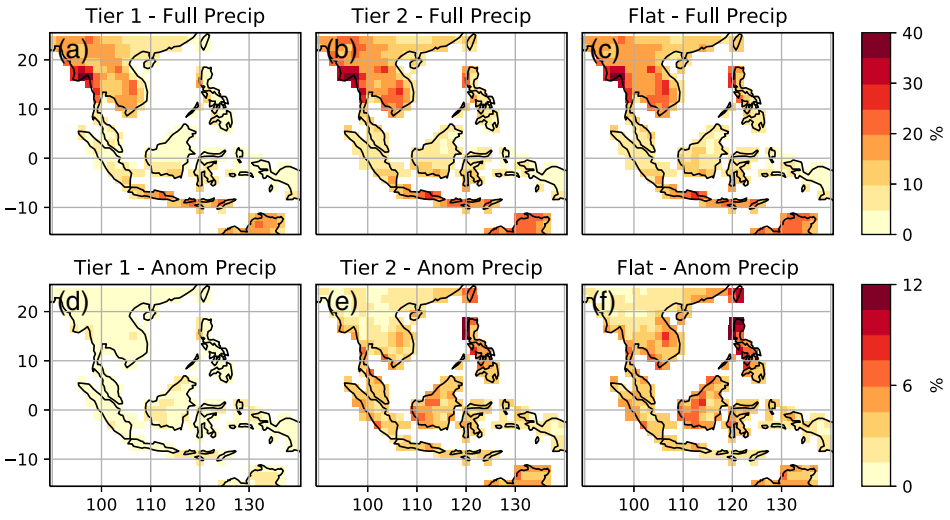

FIGURE 13 GPM Precipitation variance over land explained by the indicated patterns, calculated using (a–c) the full precipitation field and (d–f) anomalies from the 60-day running mean climatology [Colour figure can be viewed at wileyonlinelibrary.com]

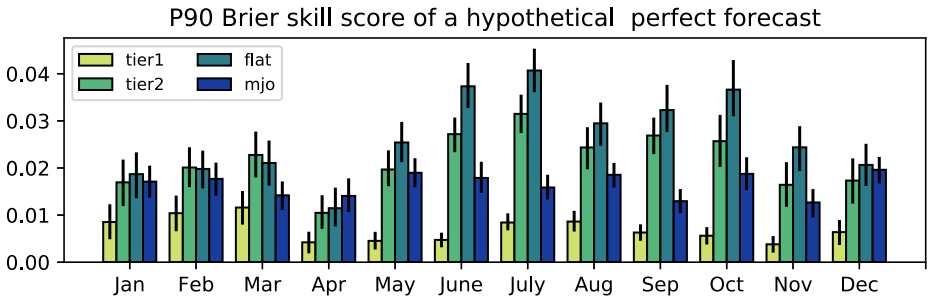

FIGURE 14 GPM precipitation Brier Skill Scores for exceedance of seasonal p90 precipitation in each month. Columns per block are (left to right) tier-1, tier-2, flat, MJO. Error bars show the bootstrapped 95% confidence interval [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Case-study precipitation and threshold exceedance likelihoods during selected extreme weather events

	Date	Location (°E, °N)	Precipitation (mm·day^{-1})	Clim. prob.(%)	Tier	Tier prob.	Flat	Flat prob.
1	03 January 2019	Southern Thailand (100.5, 7.5)	276.5	6.7 (0.7)	1b	×0.2 (×0.0)	51	**×2.9 (×6.4)**
2	28 December 2018	Central Philippines (123.0, 13.5)	263.5	12.0 (1.5)	2a	**×2.4 (×2.2)**	1	**×1.2 (×2.7)**
3	22 January 2019	South Sulawesi, Indonesia (120.0, −4.5)	25.5	6.2	2a	**×1.5**	1	**×2.0**
4	18 March 2019	Papua, Indonesia (139.5, −3.0)	66.4	8.5	1e	**×1.0**	6	**×1.1**
5	15 March 2019	Java, Indonesia (108.0, −6.0)	38.0	7.7	8e	**×1.1**	2	**×2.4**
6	26 April 2019	South Sumatra, Indonesia (103.5, −4.5)	42.4	6.5	3e	**×1.7**	13	**×1.8**
7	03 May 2019	Sumatra, Indonesia (99.0, 1.5)	25.8	6.4	6a	**×1.6**	14	**×1.2**
8	17 May 2019	Maluku and Timor Leste (129.0, −3.0)	33.7	9.5	6a	×0.5	20	×0.9
9	28 May 2019	North Vietnam (105.0, 22.5)	52.0	9.1	4g	×0.7	21	**×1.4**
10	02 June 2019	Sarawak, Malaysia (112.5, 0.0)	51.5	9.2	3g	**×2.0**	15	**×2.1**
11	08 June 2019	Sulawesi, Indonesia (121.5, −3.0)	77.8	7.7	4g	**×1.2**	18	**×2.0**
12	12 June 2019	East Kalimantan, Indonesia (114.0, 0.0)	34.0	8.0	4c	×0.7	20	**×1.1**
13	08 June 2019	Mindanao, Philippines (124.5, 7.5)	28.6	3.7	4g	×0.7	18	**×2.0**
14	24 June 2019	Northwest Vietnam (102.0, 21.0)	30.4	6.6	4h	×0.7	30	**×1.1**
15	03 July 2019	North central Vietnam (106.5, 18.0)	211.4	4.5 (0.5)	4e	**×1.6 (×4.3)**	23	**×3.2 (×6.5)**
16	16 July 2019	Luzon, Philippines (121.5, 18.0)	92.6	8.4 (1.0)	5e	**×1.9 (×2.9)**	32	**×6.3 (×19.7)**
17	03 August 2019	Laos and Vietnam (103.5, 19.5)	152.8	10.9 (0.3)	5e	**×1.3** (×0.0)	28	**×1.4 (×2.9)**
18	23 August 2019	Luzon, Philippines (121.5, 18.0)	162.6	10.2 (1.7)	5b	**×3.7 (×6.2)**	32	**×5.1 (×11.2)**
19	29 August 2019	Widespread Vietnam (108.0, 16.5)	206.5	9.0 (1.5)	4f	×0.7 (×0.8)	31	×0.5 (×0.0)
20	15 October 2019	Central Vietnam (106.5, 18.0)	56.5	12.2	7f	×0.3	46	×0.3
21	22 October 2019	Malaysia and Sumatra (103.5, 1.5)	40.6	4.7	6c	×0.9	15	**×1.3**
22	30 October 2019	Southern central Vietnam (109.5, 13.5)	177.7	13.9 (1.2)	7b	×0.7 (×0.6)	41	×0.3 (×0.0)
23	09 November 2019	Luzon, Philippines (121.5, 16.5)	48.5	10.4	6d	×1.0	38	×0.7
24	10 November 2019	Daklak, Vietnam (108.0, 12.0)	176.8	4.5 (0.2)	6d	**×1.5 (×3.9)**	38	**×2.1 (×4.7)**
25	20 November 2019	Southern Luzon, Philippines (121.5, 18.0)	48.0	9.5	8c	×0.5	46	**×1.0**
26	01 December 2019	South Thailand/Malaysia (102.0, 6.0)	88.8	18.4	7e	×0.2	48	×0.8
27	02 December 2019	Central Philippines (123.0, 13.5)	163.1	14.8 (1.9)	7e	**×1.2 (×1.1)**	48	**×1.5 (×2.1)**
28	05 December 2019	Luzon, Philippines (121.5, 18.0)	65.8	7.0	7g	**×1.3**	42	**×6.6**
29	09 December 2019	Sabah, Malaysia (118.5, 6.0)	67.1	10.2	7f	×0.2	46	×0.3
30	13 December 2019	Johor, Malaysia (103.5, 1.5)	86.2	8.1	8c	**×1.5**	11	×0.9
31	18 December 2019	Malaysia and Thailand (103.5, 4.5)	62.0	18.1	8c	**×1.7**	47	**×1.8**
32	18 December 2019	Riau, Indonesia (103.5, 0.0)	31.0	8.0	8c	**×1.8**	47	**×1.7**
33	30 December 2019	North Sumatra (97.5, 3.0)	26.5	8.8	1c	×1.0	2	**×1.3**
34	31 December 2019	Jakarta, Indonesia (106.5, −6.0)	81.4	10.1	1c	**×1.5**	2	**×1.6**

Note: Precipitation is based on GPM value at nearest grid-cell. Climatological probabilities denote the likelihood of exceeding 25 mm·day^{-1} at the event location in a 60-day climatological window surrounding the date of the event. Pattern probabilities indicate the enhanced or suppressed probability of exceedance given the occurrence of the assigned pattern, as a ratio over the climatological probability. When the observed precipitation exceeds 90 mm·day^{-1}, numbers in brackets denote the probability or enhanced/suppressed probability of exceeding 100 mm·day^{-1}. Bold indicates a factor greater than 1.

been included for brevity. Figure 14 indicates the seasonal cycle of the P90 BSS for each set of patterns and for the MJO, averaged over each month of the year. The BSS is calculated at a 1.5° resolution using GPM precipitation, and then averaged across all land grid-cells. The first tier of patterns shows very little skill beyond the seasonal climatology, as may be expected. Error-bars represent the 95% bootstrapped confidence interval, obtained by re-sampling the spatially averaged BSS for each month 10,000 times.

The annual average values of the values shown in Figure 14 are:

tier-1: 0.007 ± 0.0008, tier-2: 0.022 ± 0.0013,
flat: 0.027 ± 0.0014, and MJO: 0.017 ± 0.0009.

Thus, both the tiered and the flat pattern outperform the MJO overall. The BSS values are much less than 1 in all cases, indicating that a pattern-conditioned precipitation forecast will always have low sharpness. The flat patterns perform significantly better than the tier-2 patterns from May through to November, with the difference peaking in mid-July during the tropical cyclone season, consistent with the conclusions drawn from Figure 13. During February and March, the tier-2 patterns show a slightly improved representation of the P90 exceedance compared to the flat patterns, though this difference is not statistically significant. The flat patterns significantly out-perform the MJO from May to November and during March, and are not statistically distinguished from the MJO in the remaining months. The tiered patterns significantly outperform the MJO from June to October and during March.

In order to provide a heuristic sense of the information that a pattern-based forecast may provide, pattern membership of a limited set of heavy precipitation case-studies has been considered. This analysis is provided for illustrative purposes only. Case-studies have been selected by considering all HIW linked to heavy precipitation reported by floodlist.com during 2019 that affected Indonesia, Malaysia, Vietnam, Thailand, the Philippines, and Timor Leste (Davies *et al.*, 2021). For each event, the coarse-grained GPM grid-cell with the highest daily precipitation in the vicinity of the reported event and on a day occurring in the event duration was selected. If this precipitation was less than 25 mm·day^{-1} the event was excluded, in order to remove localised convection that the pattern methodology is not designed to detect.

Table 3 presents a list of the heavy precipitation event dates, locations, grid-cell maximum precipitation and assigned patterns. The climatological and pattern-based probabilities of precipitation exceeding 25 mm·day^{-1} are given. When the grid-cell mean precipitation exceeds 90 mm·day^{-1}, the climatological- and pattern-based probabilities of exceeding 100 mm·day^{-1} are also provided in parentheses. Entries in columns 7 and 9 reflect the ratio of the values of E used in the calculation of the pattern BS to the values of E used to calculate the climatology BS (Equation 3). Some events are well described by the pattern-based forecasts with pattern-conditioned likelihood of high precipitation far exceeding the climatological likelihoods. Other events are completely missed by both the tiered and the flat patterns. More of the events have higher than climatology probabilities based on the flat patterns than based on the tiered patterns, consistent with the increased BSS of the flat patterns. The tropical cyclone events in the Philippines (rows 16, 18 and 28) have particularly high flat-pattern-based exceedance likelihoods. In the median, however, when the pattern-conditioned probability of exceeding 25 mm is larger than the climatology, the enhanced probability of exceedance is only 1.5 (1.75) times greater than climatology when using the tiered (flat) patterns. This is reflective of the characteristic lack of sharpness that was observed when studying the BSS.

6 | DISCUSSION AND CONCLUSIONS

This paper has defined two sets of weather patterns over Southeast Asia suitable for the description of high-impact weather in the form of heavy precipitation. The general methodology for the definition of the weather patterns was based on k-means cluster analysis on the leading principal components of 850 hPa wind components.

The first set of weather patterns was defined following a two-tiered approach in which tier-1 weather regimes in a planetary-scale domain conditioned a second tier of patterns defined in a Southeast Asia regional domain. The planetary domain covers approximately one fifth of the Earth's surface, while the regional domain is approximately a quarter of the planetary domain. This approach was motivated by two factors: firstly, previous work strongly suggested that large sets of regional-scale weather patterns at midlatitudes (Neal *et al.*, 2016) and the Tropics (Neal *et al.*, 2020) can be further classified by their relationship with the circulation at larger scales, and secondly, the influence that large-scale modes of variability exert on the tropical circulation. The second set of weather patterns was defined following a single-tier or flat cluster analysis within the Southeast Asia regional domain.

The planetary-scale tier-1 weather regimes were described in terms of their relationship with large-scale modes of variability, while the tier-2 and flat patterns were described in term of their relationship with synoptic-scale phenomena. It was shown that tier-1 patterns are mainly an expression of the seasonal cycle, discerning between the monsoon circulation in and around Southeast Asia. However, they are also strongly influenced by large-scale

variability modes such as ENSO and the MJO, exhibiting distinct behaviour dependent on whether the year correspond to a La Niña or an El Niño year. The influence of the IOD was also investigated, but was found to be not significant.

Characteristics of the tier-1 weather regimes were comparable to previous studies of weather regimes conducted over the Maritime Continent. The seasonal timing and relationships with ENSO and the MJO of Regimes 1 and 2 are very similar to weather types 5 and 3 respectively identified by Moron *et al.* (2015). These regimes and weather types all occur at the peak of the austral monsoon, and while Regime 1 and weather type 5 favour El Niño and an inactive MJO over the Maritime Continent, Regime 2 and weather type 3 favour La Niña and active MJO phases. Meanwhile, the sequential timing of their weather types 1, 6 and 2 from September to December maps directly onto tier-1 Regimes 6, 7 and 8. This suggests that the tier-1 Regimes are not very sensitive to the differences in methodology between these two studies: namely restriction to a particular season and choice in the number of clusters. We do not find an equivalent of weather type 4 of Moron *et al.* (2015), which has similar timing to weather types 3 and 5, perhaps because our larger domain allows the large-scale influences of ENSO and the MJO to be more fully reflected. Hassim and Timbal (2019) also identified the same number of regimes as the first tier, despite applying a very different methodology. In this case, particular correspondence is present in the transition seasons: between tier-1 Regime 3 and their Regime 1, which both dominate the northward passage of the ITCZ before the onset of the boreal monsoon, and between tier-1 Regime 8 and their Regime 6, which both occur in October and November at the beginning of the austral monsoon.

Both the tiered and the flat patterns were shown to be able to highlight the occurrence of synoptic weather events: cold surges, the MJO, tropical cyclones, Borneo vortices and the BSISO. They were able to explain up to 10% of the variance in seasonally anomalous precipitation anomalies over land regions. A hypothetical perfect forecast of the assigned patterns was found to have positive skill at predicting the exceedance of heavy precipitation thresholds compared to the seasonal climatology. The tier-2 and flat pattern sets were also shown to have enhanced skill potential compared to an MJO-based forecast. The pattern sets are able to classify circulation patterns over Southeast Asia in a way that may be used to distinguish the likelihood of seasonally anomalous heavy precipitation. This suggests that a patterns-based forecast of the probability of heavy precipitation has the potential to improve the S2S forecastability of heavy precipitation. This hypothesis will be explored further in G21.

Both the proportion of precipitation explained by the weather patterns and the BSS of heavy precipitation take quite modest values. This is in part due to the weakness of the relationship between rainfall and large-scale circulation, particularly at relatively small spatial scales around $150 \times 150\,\mathrm{km}^2$. Examination of the grid-point correlation between ERA-5 wind convergence at 850 hPa and GPM precipitation on the small domain 1.5° grid reveals correlation values of $R = 0.23$. This corresponds to a R^2 value of 0.053, which is of a similar order of magnitude to the variance explained by our regimes. Based on this, it is unlikely that improvements in the choice of regimes could result in an improvement in the rainfall variance explained at this length-scale. On the other hand, aggregation of precipitation to larger scales has been shown to extend forecastability in the Southeast Asian region (e.g., Ferrett *et al.*, 2021). When considering the actual skill of these regimes in predicting precipitation, G21 will demonstrate that useful levels of skill are present when precipitation is aggregated to coarser resolutions.

While a traditional approach of clustering studies is to restrict focus to specific seasons, the tiered methodology applied here allows for seasonal and monsoonal patterns to be objectively identified in the first tier and considered separately in each set of tier-2 patterns. This approach has particular value in the equatorial Maritime Continent region, where the annual cycle of both monsoonal winds and rainfall is complex. This complexity derives from the bi-annual passage of the ITCZ, which is associated with three separate named monsoon winds patterns: the southwest and northeast monsoons in the Northern Hemisphere, and the austral monsoon in the Southern Hemisphere. The onset of the latter progresses from east to west, while some regions of the domain, including New Guinea, exhibit very little seasonal variation. All these subtleties of the local annual cycle are expressed though the large-scale tier-1 weather regimes.

A consequence of this tiered approach is that the total number of weather patterns identified, at 51, is considerable. Many other studies that consider weather patterns across the annual cycle including Lamb (1972), Hess and Brezowsky (1952) and Neal *et al.* (2016,2020) use a smaller number of patterns, typically around 30. We found that our tier-2 patterns had roughly the same wind inertia metric as a flat set of 31 patterns. This implies that redundancy within the tier-2 regimes accounts for the presence of 20 regimes. However, the strong conditioning of regime assignment on the seasonal cycle means that on average only 13 of the tiered regimes are likely to occur at any given day within the seasonal cycle.

The synoptic climatology of Southeast Asia is well-established, and it is well-known that synoptic-scale features such as cold surges, the MJO, equatorial waves

and tropical cyclones are crucial for the local rainfall climate (e.g., Yang *et al.*, 2003; Wheeler and Hendon, 2004; Chang *et al.*, 2005; Ferrett *et al.*, 2020). Rather than adding to this literature, this paper has sought to synthesise these discrete understandings of synoptic events into stand-alone classification systems. As such, the weather patterns described in these paper should not be thought of as synoptic phenomena in their own rights, but as a means to easily describe and assign synoptic conditions to more physically based synoptic weather systems and their associated risk of heavy precipitation. This will enable forecasters to easily and objectively classify large numbers of forecast ensembles by weather regime.

This paper set out to determine whether the use of a tiered methodology could improve the ability of the patterns to describe (a) planetary-scale variability, (b) synoptic circulation variability, and/or (c) precipitation variability, compared to a flat methodology. We found that the tiered methodology did indeed improve the representation of (a), while there was very little difference in the representation of (b) between the tiered and flat regimes (Table 2). Precipitation variability (c) was better described by the flat patterns (Section 5). The improved characterisation of the large-scale climate by the tiered regimes may have the potential to improve the forecastability of these patterns. One major unanswered question, which is the focus of G21, is whether constraining the choice of pattern membership based on the slowly varying large-scale flow may improve the prediction of synoptic weather patterns. A second question is whether probabilistic forecasts of the risk of heavy precipitation based on the projection of sub-seasonal to seasonal forecasts onto the tier-1 patterns, and the conditional dependence of precipitation on those patterns, has skill that pushes into the extended range. In other words, can either the tiered or the flat pattern identification approach yield skill in the prediction of high-impact weather risk beyond the medium-range?

ACKNOWLEDGEMENTS

This work and its contributors were supported by the Met Office Weather and Climate Science for Service Partnership (WCSSP) Southeast Asia as part of the Newton Fund. SJW and OM-A were also supported by the National Centre for Atmospheric Science ODA national capability program ACREW (NE/R000034/1), which is supported by NERC and the GCRF. The GPM DPR and GMI (Combined Precipitation) L3 daily 0.1° × 0.1° version 06 data were provided by the NASA's Goddard Earth Sciences Data and Information Services Center. ERA-5 850 hPa wind vectors were accessed from the Copernicus Climate Change Service (C3S), available at https://cds.climate.copernicus.eu/cdsapp#/home. The Borneo vortex, tropical cyclone and equatorial waves datasets were produced as part of the Newton Fund project under the auspices of the WCSSP Southeast Asia Equatorial Waves Project by Dr Kevin Hodges and Dr Gui-Ying Yang of the National Centre for Atmospheric Science and Department of Meteorology, University of Reading.

AUTHOR CONTRIBUTIONS

Emma Howard: data curation; formal analysis; investigation; methodology; writing – original draft; writing – review and editing. **Simon Thomas:** data curation; formal analysis; investigation; methodology; writing – review and editing. **Thomas H.A. Frame:** conceptualization; funding acquisition; methodology; supervision; writing – review and editing. **Paula L.M. Gonzalez:** data curation; formal analysis; investigation; methodology; writing – review and editing. **John Methven:** conceptualization; funding acquisition; methodology; supervision; writing – review and editing. **Oscar Martínez-Alvarado:** conceptualization; funding acquisition; methodology; supervision. **Steven J. Woolnough:** conceptualization; funding acquisition; methodology; supervision; writing – review and editing.

ORCID

Emma Howard https://orcid.org/0000-0003-0108-1220
Simon Thomas https://orcid.org/0000-0001-7119-9545
Paula L.M. Gonzalez https://orcid.org/0000-0003-0154-0087
Oscar Martínez-Alvarado https://orcid.org/0000-0002-5285-0379

REFERENCES

Baur, F., Hess, P. and Nagel, H. (1944) *Kalendar der Grosswetterlagen Europas 1881–1939*. Bad Homburg: DWD.

Birch, C., Webster, S., Peatman, S., Parker, D., Matthews, A., Li, Y. and Hassim, M. (2016) Scale interactions between the MJO and the western Maritime Continent. *Journal of Climate*, 29, 2471–2492.

Boer, G. (2003) Predictability as a function of scale. *Atmosphere–Ocean*, 41, 203–215.

Cassou, C. (2008) Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. *Nature*, 455, 523–527.

Chang, C.P., Harr, P.A. and Chen, H.J. (2005) Synoptic disturbances over the equatorial South China Sea and western Maritime Continent during boreal winter. *Monthly Weather Review*, 133, 489–503.

Dang-Quang, N., Renwick, J. and McGregor, J. (2016) On the presence of tropical vortices over the southeast Asian Sea–Maritime Continent region. *Journal of Climate*, 29, 4793–4800. https://doi.org/10.1175/JCLI-D-14-00468.1

Davies, R., Behrend, J. and Hill, E. (2021) Floodlist Asia. *Floodlist*. http://floodlist.com/asia; accessed 13 December 2021.

Dawson, A. (2016) Windspharm: a high-level library for global wind field computations using spherical harmonics. *Journal of Open Research Software*, 4. https://doi.org/10.5334/jors.129.

Eckstein, D., Künzel, V., Schäfer, L. and Winges, M. (2019). Global Climate Risk Index 2020, Berlin: GermanWatch. www.germanwatch.org/en/17307; accessed 13 December 2021.

Ferrett, S., Frame, T.H.A., Methven, J., Holloway, C.E., Webster, S., Stein, T.H.M. and Cafaro, C. (2021) Evaluating convection-permitting ensemble forecasts of precipitation over Southeast Asia. *Weather and Forecasting*, 36, 1217. https://journals.ametsoc.org/view/journals/wefo/36/4/WAF-D-20-0216.1.xml.

Ferrett, S., Yang, G.-Y., Woolnough, S.J., Methven, J., Hodges, K. and Holloway, C.E. (2020) Linking extreme precipitation in Southeast Asia to equatorial waves. *Quarterly Journal of the Royal Meteorological Society*, 146, 665–684.

Ghil, M., Groth, A., Kondrashov, D. and Robertson, A.W. (2018). Extratropical sub-seasonal to seasonal oscillations and multiple regimes: The dynamical systems view, pp. 119–142 in Sub-Seasonal to Seasonal Prediction, Robertson, A.W., Vitart, F. (eds). Amsterdam: Elsevier.

Gonzalez, P.L.M., Howard, E., Frame, T.H.A., Martínez-Alvarado, O., Methven, J. and Woolnough, S.J. (2021) Weather regimes in Southeast Asia: Enhancing high-impact weather sub-seasonal forecast skill. *Quarterly Journal of the Royal Meteorological Society*.

Grams, C.M., Beerli, R., Pfenninger, S., Staffell, I. and Wernli, H. (2017) Balancing Europe's wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, 7, 557–562.

Hassim, M.E. and Timbal, B. (2019) Observed rainfall trends over Singapore and the Maritime Continent from the perspective of regional-scale weather regimes. *Journal of Applied Meteorology and Climatology*, 58, 365–384.

Hattori, M., Mori, S. and Matsumoto, J. (2011) The cross-equatorial northerly surge over the Maritime Continent and its relationship to precipitation patterns. *Journal of the Meteorological Society of Japan*, 89, 27–47.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D.P., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hogan, R.J., Hólm, E.V., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049.

Hess, P. and Brezowsky, H. (1952). Katalog der Grosswetterlagen Europas [Catalogue of large-scale weather conditions over Europe], Bad Kissingen, Germany. Berichte des Deutschen Wetterdienstes in der US-Zone, No. 33.

Hodges, K., Cobb, A. and Vidale, P.L. (2017) How well are tropical cyclones represented in reanalysis datasets?. *Journal of Climate*, 30, 5243–5264. https://doi.org/10.1175/JCLI-D-16-0557.1.

Hohenegger, C. and Schäar, C. (2007) Atmospheric predictability at synoptic versus cloud-resolving scales. *Bulletin of the American Meteorological Society*, 88, 1783–1794.

Holsclaw, T., Greene, A.M., Robertson, A.W. and Smyth, P. (2016) A Bayesian hidden Markov model of daily precipitation over South and East Asia. *Journal of Hydrometeorology*, 17, 3–25. https://doi.org/10.1175/JHM-D-14-0142.1.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J. and Xie, P. (2015). Integrated multi-satellite retrievals for GPM (IMERG); technical documentation. NASA/GSFC Code 612, Greenbelt, MD.

Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Kyselỳ, J. and Tveito, O.E. (2008) Classifications of atmospheric circulation patterns: recent advances and applications. *Annals of the New York Academy of Sciences*, 1146, 105–152.

James, P. (2007) An objective classification method for Hess and Brezowsky Grosswetterlagen over Europe. *Theoretical and Applied Climatology*, 88, 17–42.

Lamb, H.H. (1972) *British Isles weather types and a register of the daily sequence of circulation patterns 1861–1971*. London: H.M. Stationery Office. Met Office Geophysical Memoirs No. 116.

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

Love, B.S., Matthews, A.J. and Lister, G.M. (2011) The diurnal cycle of precipitation over the Maritime Continent in a high-resolution atmospheric model. *Quarterly Journal of the Royal Meteorological Society*, 137, 934–947.

Michelangeli, P.-A., Vautard, R. and Legras, B. (1995) Weather regimes: recurrence and quasi-stationarity. *Journal of the Atmospheric Sciences*, 52, 1237–1256.

Moron, V., Robertson, A.W. and Qian, J.-H. (2010) Local versus regional-scale characteristics of monsoon onset and post-onset rainfall over Indonesia. *Climate Dynamics*, 34, 281–299.

Moron, V., Robertson, A.W., Qian, J.-H. and Ghil, M. (2015) Weather types across the Maritime Continent: from the diurnal cycle to interannual variations. *Frontiers in Environmental Science*, 2. https://doi.org/10.3389/fenvs.2014.00065.

Moron, V., Barbero, R. and Robertson, A. (2016) Subseasonal-to-interannual variability of rainfall over New Caledonia (SW Pacific). *Climate Dynamics*, 46, 2449–2468.

Moron, V., Barbero, R., Evans, J.P., Westra, S. and Fowler, H.J. (2019) Weather types and hourly to multiday rainfall characteristics in tropical Australia. *Journal of Climate*, 32, 3983–4011. https://doi.org/10.1175/JCLI-D-18-0384.1.

Neal, R., Fereday, D., Crocker, R. and Comer, R.E. (2016) A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteorological Applications*, 23, 389–400.

Neal, R., Robbins, J., Dankers, R., Mitra, A., Jayakumar, A., Rajagopal, E. and Adamson, G. (2020) Deriving optimal weather pattern definitions for the representation of precipitation variability over India. *International Journal of Climatology*, 40, 342–360.

North, G.R., Bell, T.L., Cahalan, R.F. and Moeng, F.J. (1982) Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review*, 110, 699–706.

Pal, I., Robertson, A., Lall, U. and Cane, M. (2015) Modeling winter rainfall in northwest India using a hidden Markov model: understanding occurrence of different states and their dynamical connections. *Climate Dynamics*, 44, 1003–1015.

Philipp, A., Bartholy, J., Beck, C., Erpicum, M., Esteban, P., Fettweis, X., Huth, R., James, P., Jourdain, S., Kreienkamp, F., Krennert,

T., Lykoudis, S., Michalides, S.C., Pianko-Kluczynska, K., Post, P., Alvarez, D.R., Schiemann, R., Spekat, A. and Tymvios, F.S. (2010) Cost733cat – a database of weather and circulation type classifications. *Physics and Chemistry of the Earth*, 35, 360–373. https://doi.org/10.1016/j.pce.2009.12.010.

Pope, M., Jakob, C. and Reeder, M. (2009) Regimes of the north Australian wet season. *Journal of Climate*, 22, 6699–6715.

Sáenz, F. and Durán-Quesada, A.M. (2015) A climatology of low-level wind regimes over Central America using a weather type classification approach. *Frontiers in Earth Science*, 3. https://doi.org/10.3389/feart.2015.00015.

Selz, T. and Craig, G.C. (2015) Upscale error growth in a high-resolution simulation of a summertime weather event over Europe. *Monthly Weather Review*, 143, 813–827.

Thomas, S.R., Martínez-Alvarado, O., Drew, D. and Bloomfield, H. (2020) Drivers of extreme wind events in Mexico for wind power applications. *International Journal of Climatology*, 41, E2321–E2340. https://doi.org/10.1002/joc.6848.

Tibshirani, R., Walther, G. and Hastie, T. (2001) Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B*, 63, 411–423.

Wheeler, M.C. and Hendon, H.H. (2004) An all-season real-time multivariate MJO index: development of an index for monitoring and prediction. *Monthly Weather Review*, 132, 1917–1932. https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Yang, G.-Y., Hoskins, B.J. and Slingo, J.M. (2003) Convectively coupled equatorial waves: A new methodology for identifying wave structures in observational data. *Journal of Atmospheric Sciences*, 60, 1637–1654. https://doi.org/10.1175/1520-0469(2003)060<1637:CCEWAN>2.0.CO;2.

Ying, Y. and Zhang, F. (2017) Practical and intrinsic predictability of multi-scale weather and convectively coupled equatorial waves during the active phase of the MJO. *Journal of the Atmospheric Sciences*, 74, 3771–3785.

Zhang, F., Bei, N., Rotunno, R., Snyder, C. and Epifanio, C.C. (2007) Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *Journal of the Atmospheric Sciences*, 64, 3579–3594.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Howard, E., Thomas, S., Frame, T.H., Gonzalez, P.L., Methven, J., Martínez-Alvarado, O. *et al.* (2022) Weather patterns in Southeast Asia: Relationship with tropical variability and heavy precipitation. *Quarterly Journal of the Royal Meteorological Society*, 148(743), 747–769. Available from: https://doi.org/10.1002/qj.4227