Generated using the official AMS LaTeX template v5.0

A local-to-large scale view of Maritime Continent rainfall: control by ENSO, MJO and equatorial waves

Simon C. Peatman*, Juliane Schwendike, Cathryn E. Birch, John H. Marsham

Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, UK

Adrian J. Matthews

Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences and School of Mathematics, University of East Anglia, Norwich, UK

Gui-Ying Yang

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

**Corresponding author*: Simon C. Peatman, Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK, earspe@leeds.ac.uk

ABSTRACT

The canonical view of the Maritime Continent (MC) diurnal cycle is deep convection occurring over land during the afternoon and evening, tending to propagate offshore overnight. However, there is considerable day-to-day variability in the convection, and the mechanism of the offshore propagation is not well understood. We test the hypothesis that large-scale drivers such as ENSO, the MJO and equatorial waves, through their modification of the local circulation, can modify the direction or strength of the propagation, or prevent the deep convection from triggering in the first place. Taking a local-to-large scale approach we use *in situ* observations, satellite data and reanalyses for five MC coastal regions, and show that the occurrence of the diurnal convection and its offshore propagation is closely tied to coastal wind regimes we define using the ~~using the~~ k-means cluster algorithm. Strong prevailing onshore winds are associated with a suppressed diurnal cycle of precipitation; while prevailing offshore winds are associated with an active diurnal cycle, offshore propagation of convection and a greater risk of extreme rainfall. ENSO, the MJO, equatorial Rossby waves and westward mixed Rossby-gravity waves have varying levels of control over which coastal wind regime occurs, and therefore on precipitation, depending on the MC coastline in question. The large-scale drivers associated with dry and wet regimes are summarised for each location as a reference for forecasters.

Significance statement. Extreme precipitation can be life-threatening in the Maritime Continent region, for example due to flash floods and landslides. The main form of variability of convective storms is the diurnal cycle, but this can be modulated by large-scale weather drivers. By quantifying the effect of these drivers on local-scale weather regimes for a range of Maritime Continent locations, we identify which drivers are most important (and in which phase) to consider when understanding the local risk of extreme rainfall. Given that these large-scale drivers may be forecast with greater skill than is possible for quantitative precipitation forecasts, this study provides crucial extra information for forecasters to aid prediction of life-threatening weather conditions.

1. Introduction

The Maritime Continent (MC; figure 1), the archipelago situated on the equator between 90°E and 160°E, experiences some of the most intense rainfall on Earth, due to its location in the Indo-Pacific warm pool (Ramage 1968). Convection exhibits a strong diurnal cycle driven by the land-sea temperature contrast, with the heaviest rainfall over land generally occurring in the late afternoon and evening, and over ocean in the early morning. In some regions, such as south-west of Sumatra, north-west of Borneo, north and south of Java, and north of New Guinea, convection can propagate offshore overnight (e.g., Qian 2008; Love et al. 2011). Several physical mechanisms have been proposed for the nocturnal offshore propagation, including roles for gravity waves (Warner et al. 2003; Mapes et al. 2003; Love et al. 2011; Hassim et al. 2016), cold pool outflow (Mori et al. 2004; Wu et al. 2009) and the land-sea breeze circulation (Houze et al. 1981).

The diurnal convection and its propagation vary in strength day-to-day, and on some days at any given location in the MC it may not occur at all. We use a novel local-to-large scale framework, as explained below, to investigate the hypothesis that these forms of local variability are influenced by large-scale drivers through their control on the local circulation, and to quantify the relative contributions of such drivers for a range of locations in the MC.

Previous work has shown an interaction between the diurnal cycle and the intraseasonal Madden-Julian Oscillation (MJO; e.g., Oh et al. 2012; Peatman et al. 2014; Birch et al. 2016; Vincent and Lane 2016; Sakaeda et al. 2020; Qian 2020; Muhammad et al. 2021). The MJO consists of alternating large-scale envelopes of active and suppressed convection, propagating slowly eastwards at the equator, from the Indian Ocean across the MC into the Pacific (Madden and Julian 1971, 1972). Changes in the diurnal cycle account for 81% of the variability in land-based precipitation during an MJO cycle, with the largest diurnal amplitude occurring just ahead of the active MJO envelope (Oh et al. 2012; Peatman et al. 2014).

The El Niño-Southern Oscillation (ENSO) also modulates the MC diurnal cycle, enhancing it during El Niño events and suppressing it during La Niña events. As a result, even though El Niño suppresses MC convection and La Niña enhances MC convection on the large scale, precipitation anomalies over the islands have the opposite sign (Rauniyar and Walsh 2013). Several types of convectively-coupled equatorial wave also modulate the diurnal cycle as they propagate through the MC, affecting the probability of extreme rainfall (Ferrett et al. 2020; Lubis and Respati 2021). According to Sakaeda et al. (2020), equatorial Kelvin waves predominantly modulate the diurnal amplitude over ocean and over land to a lesser extent, with the strongest convection leading the active convection phase of the wave. Baranowski et al. (2016) showed that Kelvin waves may enhance the diurnal cycle over Sumatra and Borneo depending on the time of day of their arrival, due to phase locking. Sakaeda et al. (2020) also showed equatorial Rossby waves modulate the diurnal cycle more strongly over ocean than land; and over land the strongest diurnal cycle leads the convective phase of the wave on the east side of islands, but lags on the west side. Nocturnal offshore propagation of convection from south-west Sumatra, north-west Borneo and south Java is enhanced by the active phase of a Kelvin wave~~;~~. Ahead of the convective phase of a Rossby wave,~~while~~ the propagation is enhanced for south-west Sumatra and suppressed for north-west Borneo ~~ahead of a Rossby wave~~.

In diagnosing such scale interactions, most studies use a large-to-local scale approach, compositing local conditions as a function of the large-scale state. However, the diurnal cycle is still variable even within an MJO phase. An example is the offshore propagation south-west of Sumatra, which is strongest in a composite sense during MJO phase 2, but on any individual day during that phase the propagation may be weak or non-existent. This means the large-to-local scale approach is of limited use for forecasters. Given the potential for convection in this region to produce dangerous and possibly life-threatening conditions (e.g., Xavier et al. 2014; Ferrett et al. 2020; Mohd Nor et al. 2020; Lubis and Respati 2021), we require further metrics.

Therefore, we adopt a local-to-large scale approach, considering first the local conditions associated with propagating convection, before understanding how these local regimes are set up by phenomena at larger scales. This novel approach, combined with the use of *in situ* data from one of the few intensive atmospheric field campaigns to have been carried out in the region, allows us to quantify the large-scale drivers' influence on the coastal winds over several MC locations; and thus document which large-scale drivers are key to determining when and where intense convection associated with extreme rainfall is likely to occur.

2. Data and methods

a. Field campaign observations

We use observations from two Japanese field campaigns associated with the Years of the Maritime Continent (YMC; Yoneyama and Zhang 2020) International programme, located around Bengkulu in Sumatra (figure 1). These are referred to as "pre-YMC" (2015 campaign, November–December 2015) and "YMC" (2017 campaign, November 2017 – January 2018). We use radiosonde observations of wind from both campaigns from Bengkulu, from 2015/11/09 to 2015/12/25 and 2017/11/16 to 2018/01/15 (108 days in total). Radiosondes were released every 3 hours at 00, 03, . . . , 21 UTC; or 07, 10, 13, 16, 19, 22, 01, 04 local time (LT), taking LT to be UTC+7. Data are linearly interpolated to the same pressure levels as used in the European Centre for Medium-range Weather Forecasting (ECMWF) Fifth Reanalysis (ERA5; Hersbach et al. 2020; see below).

b. Other data sets

To diagnose precipitation, including the offshore propagation of rainfall, Global Precipitation Measurement (GPM; Heale et al. 2019) data sets are used. GPM data are provided every 30 minutes on a $0.1^\circ \times 0.1^\circ$ grid. The "high-quality" data set (GPM-HQ) uses intercalibrated observations from passive microwave (PMW) instruments on a number of satellites, which are gridded then further calibrated using monthly gauge accumulations. The PMW satellites are largely "satellites of opportunity" (that is, their orbits, operations and so on are out of the control of the GPM mission) and there are missing regions between swaths. The Integrated Multi-satellitE Retrievals for GPM (IMERG) algorithm fills these gaps to provide a complete gridded product, at the same temporal and spatial resolutions. We use IMERG version 06, which fills gaps by morphing the GPM-HQ data according to motion vectors derived from total column water vapour in the Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al. 2017).

Where possible, we use GPM-HQ as this includes only the direct measurements of precipitation from PMW instruments (with calibration). When producing composites over a long time period ($O(10^2)$ or more days), the effects of missing data are negligible. However, when the number of days is small ($O(10^1)$) we use IMERG, to benefit from the improved data coverage. We use GPM during December, January and February (DJF) for the period of its availability, from DJF 2000/01 to 2019/20. For simplicity, we always exclude 29 February from DJF.

When extending our analysis beyond the field campaign periods, where possible we use hourly ERA5 for all 41 available DJFs, from 1979/80 to 2019/20. ERA5 is on a 0.25° grid. We also use the ECMWF interim reanalysis (ERA-Interim; Dee et al. 2011) when comparing wind values against those in equatorial waves (see below). MJO phases are defined using the Realtime Multivariate MJO (RMM) indices (Wheeler and Hendon 2004). To investigate the effect of ENSO we use the Oceanic Niño Index (ONI; Climate Prediction Center 2020) version 5, which is the 3-month running mean of the monthly Niño3.4 anomaly (with the subtracted climatology being a 30-year mean updated every 5 years). El Niño and La Niña events are defined as a period of at least 5 months with ONI ≥ 0.5°C or ≤ -0.5°C, respectively. All other times are defined as neutral ENSO phase. We also analyse tropical cyclone (TC) tracks, using the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al. 2010).

We use a data set of equatorial waves identified using the methodology of Yang et al. (2003) and described in Ferrett et al. (2020). Wind and geopotential height data are regridded to a 1° grid and filtered to retain variability with period 2–30 days and zonal wavenumbers 2–40. Eastward- and westward-components are separated out in wavenumber-frequency space and projected onto the theoretical horizontal structures of equatorial Kelvin, $n = 1$ Rossby (R1), $n = 2$ Rossby (R2) and westward mixed Rossby-gravity (WMRG) waves, using a meridional trapping scale of 6° (Yang et al. 2003). This is performed using ERA-Interim from 1997 to 2018, separately at each pressure level. The resulting data set consists of the wind and geopotential height anomaly contributions from each wave type.

c. k-means clustering

To differentiate between coastal wind regimes, the k-means clustering algorithm is used (MacQueen 1967). k-means is an iterative algorithm which sorts data points into clusters by minimizing the total Euclidean distance between a cluster's data points and their mean. For example, in section 3a we cluster zonal wind u from 108 days of 3-hourly radiosondes, concatenating the 8 sondes each day to produce a field of shape 108 days × 8 times of day × 16 pressure levels. We use k-means to sort each day into a cluster, hence we are clustering 108 data points in $8 \times 16 = 128$-dimensional space.

The number of clusters k is an *a priori* choice which is made subjectively, albeit with physical justifications as detailed in the text, having run the algorithm for a range of k values. The initialization of the algorithm is random, with no guarantee that different initializations will converge to the same result. Therefore, a number of initializations are performed and the best solution is selected (i.e., that which produces the minimal total Euclidean distance). For field campaign data, 500 initializations are performed; when extending to 41 DJFs, this is increased to 20,000. These values were chosen by experiment as sufficient to make the results of the clustering robust. To avoid confusion, we label the clusters from observations (section 3a) as 0, 1, . . . , $(k-1)$; and the clusters from reanalysis (sections 3b,c) as A, B, ... Coastal wind clusters are always ordered by the mean value of the cluster centre field, from the most strongly onshore to the most strongly offshore.

d. Coastal wind associated with large-scale drivers

To investigate possible large-scale causes of the coastal wind regimes derived from k-means clustering, we quantify the contribution of several large-scale drivers to the wind at a given coastal location using the following methods. ERA-Interim is used throughout for consistency since it is the reanalysis used in the equatorial waves data set. Here we assume the required coastal wind is u_{850}, although the same technique is also applied to meridional wind v_{850} for certain coastlines. The 850 hPa level is chosen as it is one of the levels used in the RMM indices to monitor the MJO and it is representative of all levels of the wind clusters.

For Kelvin, R1, R2 and WMRG waves we take u'_{850}, where the prime denotes an anomaly, directly from the equatorial waves data set (see section 2b) and average along the nominal coastline in the respective coloured box in figure 1.

To derive a u'_{850} value associated with ENSO, we take the monthly ONI values and, to avoid having a sharp jump at the start of each month, apply a 31-day running mean. We take the daily mean ERA-Interim u_{850} averaged along the nominal coastline and subtract the climatology and leading 3 harmonics of the seasonal cycle to produce a time series of u'_{850}. We take these u'_{850} values and the daily smoothed ONI values, for all 39 available DJFs (1979/80 to 2017/18), and perform least-squares linear regression:

$$u'_{850}(t) = m_{\text{ONI}} \times \text{ONI}(t) + c_{\text{ONI}}, \tag{1}$$

where t is time, and m_{ONI} and c_{ONI} are the parameters to be fitted.

Similarly, we perform linear regression with the RMM indices to find u'_{850} associated with the MJO. However, because there are two RMMs we need to perform multiple linear regression and because they are not independent we need to include a cross-term, which represents the fact that the prevailing wind at the coast does not necessarily blow in opposite directions in opposite MJO phases:

$$u'_{850}(t) = m_1 \times \text{RMM1}(t) + m_2 \times \text{RMM2}(t) + m_3 \times \text{RMM1}(t) \times \text{RMM2}(t) + c_{\text{RMM}}, \tag{2}$$

where m_1, m_2, m_3 and c_{RMM} are the parameters to be fitted, and the RMM time series are taken from Wheeler and Hendon (2004). Note that, although u'_{850} is not modelled as a linear function of the RMMs, equation 2 is linear in the fitted parameters, so least-squares multiple linear regression can be used to find the parameter values.

3. Results

a. Pre-YMC and YMC field campaign observation periods

This section of the analysis uses observations from the field campaigns on the south-west coast of Sumatra. Several of the possible mechanisms for offshore propagation of convection mentioned in section 1 depend either directly or indirectly on land-sea breezes. Therefore, we explore the variability in land-sea breeze and its relation to diurnal convection, including offshore propagation. We also wish to investigate the development of each land-sea breeze regime in terms of large-scale phenomena. Since phenomena such as the MJO and most equatorial waves are (close to the equator) predominantly associated with variations in zonal not meridional wind, we analyse the zonal component u only at this stage. The zonal and meridional components were also rotated to onshore and alongshore components, and the the analysis repeated using the onshore component only (not shown); this did not substantially change the results.

We take u from the 108 days of 3-hourly radiosondes at Bengkulu, performing k-means clustering as described in section 2c. By experiment, it was found that $k = 3$ best represents the land-sea breeze regimes (figures 2a–c; other k values not shown). This achieves the best balance between separating out physically distinct regimes (with $k = 2$ being too few) while preserving a reasonable sample size for each cluster (for $k = 4$, one of the clusters has only 4 days). Moreover, with $k = 4$ two of the clusters are very similar to cluster 0 in figure 2a so are not physically distinct.

The clusters consist of strong onshore winds all day (cluster 0), moderate onshore winds all day (cluster 1) and offshore winds all day except near to the ground in the afternoon (cluster 2). The diurnal anomaly profiles of u (relative to the cluster's own daily mean profile; figures 2d–f) are broadly similar for each cluster, but the daily mean is sufficiently different that only the offshore cluster (2) has an absolute change in wind direction at any level (figure 2c). The diurnal anomalies show all three clusters have the strongest onshore wind in the afternoon, but slightly earlier (13 LT) in the offshore cluster (2) than in the onshore clusters (0 and 1; 16 LT). The strong land-sea breeze is shallow, reaching up to around 925 hPa, close to a typical boundary layer depth over ocean. However, in the onshore clusters (0 and 1) the wind is onshore at all levels shown at the time of the strongest sea breeze, whereas the offshore cluster (2) has a return flow between 925 and 775 hPa at this time.

The clustering of the land-sea breeze is a valuable tool as it neatly divides the days into distinct diurnal convection regimes. Figure 3 shows the composite diurnal cycle of precipitation as Hovmöller diagrams using IMERG data in the red dashed box of figure 1, averaged along the coastline direction. The diagrams are composited over each cluster, with time of day running down the page and extended 12 hours into the next day, to capture the full offshore propagation. Distances on the horizontal axis are negative offshore (to the south-west) and positive onshore (to the north-east).

Although all three clusters have a discernible diurnal cycle over both land and sea, the offshore cluster (2) alone exhibits the canonical view of heavy precipitation (around 2 mm hr^{-1}) over land near the coast at 13–19 LT, propagating offshore during the evening and night-time. All the precipitation over the sea is associated with this propagation, which extends into the following afternoon. There appear to be two modes of propagation, causing the propagation region to widen from about 19 LT onwards, with the two modes becoming distinct around 04–07 LT. Preceding the convection is a corresponding propagating region of suppressed conditions.

In the strong onshore cluster (0) the propagation is mostly in the same direction as the wind, progressing inland before continuing to propagate over the sea to the north-east of Sumatra. Precipitation over the sea reduces in the afternoon and evening but is on a larger scale than the organised propagation in the offshore cluster (2), suggesting it is associated with a large-scale phenomenon such as the MJO. Precipitation over land does not feature a strong burst near the coast in the afternoon and evening as in the offshore cluster (2), indicating a suppression of the canonical diurnal cycle. Cluster 1 has features which are elements of each of the other clusters, with coherent propagation inland and weak precipitation propagating offshore. While the offshore cluster (2) has strong convection on the sea-facing flanks of the mountains (see orography at the bottom of the plot), the strongest precipitation in the moderate onshore cluster (1) is inland, the other side of the mountain range. Hence, the stronger convection always favours the leeward side of the mountains.

To investigate possible large-scale causes of the coastal wind regimes and therefore understand which large-scale conditions can lead to each convection regime, we consider the zonal wind at Bengkulu associated with a number of large-scale drivers and compare them to the total wind. Daily mean u'_{850} values associated with several drivers are computed as described in section 2d and shown in figure 4. The linear regression of u'_{850} onto ONI (equation 1) to derive the ENSO contribution is shown in figure 5a. The plot shows considerable variability in u'_{850} which is not explained by variability in ONI, due to other factors (which, according to our hypothesis, are mainly the other large-scale drivers in figure 4). By modelling the response to ENSO as linear and regressing onto ONI, we aim to estimate the contribution to u'_{850} which arises from ENSO forcing alone. Similarly, multiple linear regression of u'_{850} onto the RMM indices was performed (equation 2) to find the MJO contribution and the resultant surface in RMM space is shown as a contour plot in figure 5b.

Time series are plotted for ENSO (dark green curve), the MJO (dark blue), Kelvin (orange), R1 (dark pink), R2 (light pink) and WMRG (light blue). If the large-scale drivers described here account for the total wind field, then the sum of the six coloured curves and the mean seasonal cycle (climatology plus leading 3 harmonics; thin black ~~dotted~~ curve in figure 4), which is shown by the thick black ~~solid~~ curve, should match the total field shown in grey. During the 2015 campaign (figure 4a) there is a close match between the two. For approximately the first month of the campaign the wind was offshore and all days were in the cluster 2 except 26 and 27 November, when the wind was moderately onshore in cluster 1. The time series show this was caused by an R1 wave propagating through the region. From 12 December onwards the winds again shifted to being onshore, with all days being in cluster 1 or 0. On these days the sum of the large-scale contributions ~~in~~(thick black) does not match the total field so well, but both are positive and the strongest contribution is from the MJO. As discussed in section 2d, the MJO u'_{850} may be less accurate than for the other drivers considered, which may explain the discrepancy between the thick black and grey curves on these days. However, the MJO u'_{850} calculation is still sufficiently accurate to indicate days when the MJO is chiefly responsible for the coastal wind regime, such as the example discussed here.

In the 2017 campaign (figure 4b) there were no offshore cluster (2) days. The main difference between the two campaigns was an El Niño event in 2015/16 and a La Niña event in 2017/18. Hence, the u'_{850} contribution from ENSO is approximately -3.5 m s^{-1} in the 2015 campaign and approximately $+1.3$ m s^{-1} in the 2017 campaign. Individual peaks and troughs in the grey and ~~solid~~thick black curves for the 2017 campaign tend to match each other, with the Kelvin, R1 and R2 waves generally dominating variability on time scales of under a week. There are some periods when the magnitude of the grey and ~~solid~~thick black curves differ considerably, such as 22–30 November and 4–15 January. Such periods tend to have either a strong MJO signal or a nearby TC (any day with a TC centre within 10° geodesic of Bengkulu according to IBTrACS is shown with a red star).

Thus, we have examples from the field campaign periods of ENSO, the MJO and equatorial waves (although there is no considerable contribution from WMRG waves) controlling the coastal wind regimes at Bengkulu (figure 4); and we find that these coastal wind regimes themselves determine the regimes of the diurnal cycle of convection (figure 3). In the following subsection we extend the analysis beyond the field campaign periods and perform a more quantitative analysis of the contributions of the large-scale drivers.

b. Extension to 41 DJFs along Sumatran coastline

Although the analysis thus far has the advantage of using *in situ* observations to determine the wind regimes, it is limited to a few weeks in just two boreal winters, and is based on wind at a point location. Using ERA5 we now extend the analysis spatially and temporally. To confirm that ERA5 is an appropriate research tool for this study, we take ERA5 u at Bengkulu on the 108 field campaign days and perform k-means clustering again (not shown), to compare against the radiosonde clusters in figure 2. The results are very similar, although the ERA5 land-sea breeze circulation is slightly weaker and deeper than in the radiosondes. Therefore, we accept that the land-sea breeze in ERA5 is verified by the observations and we are confident that ERA5 is realistic enough to use it for our analysis.

We restrict ourselves to DJF, matching approximately the season already considered in section 3a, but extend the analysis to the 41 DJFs available in ERA5, going back to December 1979. Furthermore, noting that convection tends to be spatially heterogeneous and wanting to capture as much of the convection and its propagation as possible, we extend the analysis region to the entire solid red box in figure 1. Thus we cover as much of the coastline as possible where it is fairly straight whilst avoiding Siberut, the largest of the Mentawai Islands just offshore from Sumatra.

We take hourly ERA5 u averaged along the nominal west coastline of Sumatra indicated by the thick red line in figure 1 and cluster the 3,690 days from DJFs 1979/80 to 2019/20 (excluding 29 February for simplicity). Whereas for point data at Bengkulu we chose $k = 3$ to give us distinct physical regimes, for this case it was found by experiment that $k = 4$ was preferable (figures 6a–d), and we label these A, B, C and D. Clusters A (strong onshore) and D (offshore) correspond to clusters 0 and 2 respectively in figures 2a,c but are spatially smoother due to averaging over more days. Clusters B (moderate onshore) and C (weak onshore) both roughly correspond to cluster 1 in figure 2b. These two clusters, and in particular their corresponding offshore Hovmöller diagrams of precipitation (figures 7b,c – see below), are distinct enough to warrant both being included in the analysis without either suffering from a small sample size (indeed, they are the two most frequent clusters).

The diurnal anomalies for each cluster (figures 6e–h) are very similar to each other, with daytime onshore winds beginning at low levels and rapidly spreading to higher levels in all cases. This is unlike the observed diurnal anomalies in figures 2d–f. In observations, the differences in diurnal anomalies between clusters may be due to errors associated with the convective parametrization scheme in the ERA5 model (e.g., Birch et al. 2015; Love et al. 2011); land-air feedbacks which the ERA5 model is unable to capture; or the small sample size if these anomalies vary considerably between days.

To create the Hovmöller diagrams of precipitation (figure 7) we use GPM-HQ, since we are compositing over enough days for the lesser data coverage to be unimportant. Both the strong and moderate onshore clusters (A and B) exhibit large-scale precipitation over the sea with almost all propagation being inland. The weak onshore cluster (C), although similar to cluster B in the u

field, is not associated with large-scale heavy rain over the sea. The weak onshore (C) and offshore (D) clusters, while very different in u, have similar precipitation Hovmöller diagrams, differing only in that inland propagation is stronger than offshore for cluster C, and *vice versa* for cluster D.

The probability of extreme precipitation, defined here as exceeding the 99$^{\text{th}}$ percentile for the daily mean DJF, varies by cluster (figure 8). In the weak onshore (C) and offshore (D) clusters, over the south-west coast and just offshore, the probability of extreme rainfall can be around 50% more than average (over 70% more in some areas). In the strong and moderate clusters (A and B) the probability is reduced by a similar amount. Where there is a greater chance of extreme rainfall over and around the south-west coast region in clusters C and D, the Hovmöller diagrams in figure 7 show that these are due to precipitation systems forming over land and propagating offshore, not propagating in from the sea.

The results of section 3a suggest ENSO and the MJO each play a role in determining the coastal wind regime at Bengkulu. Now, with a much larger data set, we examine the relationship between these phenomena and the wind clusters for the full south-west Sumatra coast. Figure 9a shows the number of days in each ENSO phase and figure 9b shows the same separated out by cluster. Consistent with the findings from the field campaigns (figure 4), La Niña events are more often associated with more onshore wind (accounting for 52% of the strong onshore cluster (A) days) and El Niño events with more offshore wind (58% of the offshore cluster (D) days). The large-scale circulation component of ENSO consists of a strengthened Walker circulation during La Niña and weakened during El Niño, so La Niña events are associated with large-scale ascent over the MC and El Niño with large-scale suppression. Therefore, we expect large-scale convergence into the MC region in the lower troposphere during La Niña events, consistent with winds blowing onshore over the south-west coast of Sumatra; and large-scale divergence out of the MC during El Niño, consistent with winds blowing offshore.

Figures 9c,d show the equivalent statistics for the MJO. Days with RMM amplitude < 1 are shown in grey as "no MJO". See the figure caption for a list of regions experiencing active large-scale convection in each MJO phase. Cluster A, with the strongest onshore wind and weakest diurnal cycle of precipitation, favours phases 4–6 (each accounting for 15–25% of cluster A days); while cluster D, with offshore wind and the strongest diurnal cycle of precipitation, favours phases 8 and 1–3 (11–15% of cluster D days). This is consistent with existing theories of MJO propagation, with surface easterlies which blow offshore from Sumatra occurring ahead of the MJO when it is over the Indian Ocean (e.g., Matthews 2000). It is also consistent with Peatman et al. (2014), who showed the strongest diurnal cycle occurs just to the east of the active MJO; and the diurnal cycle is most greatly suppressed just ahead of the suppressed MJO. Here we see the offshore wind regime, associated with a strong diurnal cycle, favouring MJO phases when the active envelope is propagating through the Indian Ocean and approaching the MC, with this regime becoming far less common once the envelope reaches the MC in phases 4 and 5. Similarly the onshore wind regime, with a suppressed diurnal cycle, favours MJO phases when the active envelope has already propagated through the western MC where Sumatra is located, and is propagating into the west Pacific Ocean, with the suppressed envelope now approaching Sumatra.

Although figure 9 shows some correspondence between the clusters and both ENSO and the MJO, there are examples of every ENSO phase and every MJO phase coinciding with every wind cluster (although some are so rare that they are difficult to discern in the plot). Hence, while previous studies such as Rauniyar and Walsh (2013), Rauniyar and Walsh (2011) and Peatman et al. (2014) demonstrate the variation of the diurnal cycle by large-scale environment in a composite sense, we show here that ENSO or MJO phase do not uniquely determine the diurnal cycle regime and, from an operational forecasting perspective, is not sufficient information to issue a forecast of the likelihood or otherwise of severe thunderstorms in our region of interest. This very concern is why in section 1 we chose a local-to-large scale approach by examining the local diurnal cycle and understanding what large-scale conditions are associated with different regimes, as opposed to the large-to-local scale approach of other studies.

Section 3a also showed equatorial waves partially determine the wind cluster. Figures 10b,g show composites of wind and geopotential height anomalies at 850 hPa summed over all four of the identified equatorial wave types (clusters A and D only) for the period covered by the waves data set (1997/98 to 2017/18), which is a subset of the ERA5 period. For comparison, the composite wind anomaly from ERA-Interim (the same reanalysis as was used for the wave identification) is shown in figures 10a,f. If the wind field in these clusters were determined entirely by the identified equatorial waves then panels (a) and (b) would match, as would panels (f) and (g). In fact, we have already seen that ENSO and the MJO also contribute to determining the wind regime, but there are broad similarities between the ERA-Interim composites and the wave composites which confirm the waves also have a substantial contribution. Note that the ERA-Interim composites have a considerably larger amplitude than the wave composites due to the filtering applied before the identification of the waves, as explained in section 2b. Applying the same filtering before compositing the winds as in figures 10a,f (not shown) gives values of similar amplitude to the wave composites, as well as removing some features not associated with the waves.

In the strong onshore cluster (A; figure 10b) there are cyclonic circulation patterns either side of the equator, slightly stronger and with a larger zonal extent in the southern hemisphere, which are also seen in figure 10a. In the offshore cluster (D; figure 10g) there is an anti-cyclone in the southern hemisphere and a small region of anti-cyclonic vorticity around 5°N, 90°E in the northern hemisphere. Again, the southern hemisphere vorticity in figure 10f is in approximately the same location, and the vorticity in the northern hemisphere is weak but also visible, albeit slightly further north (around 10°N). The asymmetry about the equator, with stronger anomalies in the southern hemisphere, is because the clustering is performed over the southern half of Sumatra only. Southern hemisphere structures are more coherent in the composites whereas any variability between hemispheres causes northern hemisphere structures to be partially averaged over.

Figures 10c,h, 10d,i and 10e,j show, respectively, the Kelvin, R1 and R2 contributions to figures 10b,g. The WMRG contribution (not shown) is negligible. Figures 10d,i indicate the onshore wind regime is associated with the low-pressure (cyclonic) phase of an R1 wave and the offshore regime with the high-pressure (anti-cyclonic) phase. R1 is the largest contributor to the total wave composites in figures 10b,f. The theoretical R1 structure is symmetrical about the equator in geopotential height, while R2 is anti-symmetric. Therefore, if the real circulation has some asymmetry it can project onto both R1 and R2. Here, R2 reinforces the R1 signal in the southern hemisphere, where the analysis region is located (see figure 1) and opposes it in the opposite hemisphere, due to R1 being more coherent in the hemisphere of the analysis region. Given how precisely the pressure centres of R1 and R2 line up in longitude, we conclude that the projection onto both R1 and R2 arises from the same circulation pattern, which mostly resembles the theoretical R1 structure but with some asymmetry.

The winds in the Kelvin wave composites (figures 10c,h) ~~contribute very little to the composite wind~~are small in magnitude but their composite structures are coherent, with the high-pressure phase associated with the onshore regime and the low-pressure phase with the offshore regime. Thus, ~~Kelvin waves may be an indicator of the likely coastal wind regime even though their contribution is too small to be the cause of the coastal wind direction~~although Kelvin waves are correlated with the coastal wind, their contribution is small.

The analysis of figure 4 in section 3a suggested that, during the field campaign periods, ENSO, the MJO and equatorial waves all played a role in determining the coastal wind regime. In this section we have demonstrated the role of these large-scale drivers ~~during 41 DJFs~~over a longer time period. We now demonstrate that these drivers are sufficient to explain nearly all of the variability in the coastal wind. As in figure 4, the wind from the identified equatorial waves and the wind regressed onto ONI and the RMMs were summed, along with the mean seasonal cycle, and compared to the total field from ERA-Interim. This was carried out for all ~~4~~21 DJFs (1997/98 to 2017/18) covered by all the contributing data sets; the seasons 1997/98, 1998/99 and 1999/2000 are shown as examples in figure S1. The residual between the grey and ~~solid~~thick black curves was calculated for each day and a histogram of the absolute value is presented in figure 11. Blue bars show the number of days in each bin and the orange curve is the cumulative distribution, displayed as a percentage. The distribution peaks in the lowest bin; a similar histogram (not shown) of the signed residual (i.e., not the absolute value) is symmetrical about 0, so there is no overall tendency for the theoretical value to be more onshore or more offshore than the true value.

The majority of days have a residual < 1.4 m s^{-1}. The vertical red line is the season-mean standard deviation in coastal u_{850}. On 83% of days the residual is less than one standard deviation, indicating that our theoretical coastal wind reconstructed from the large-scale drivers has a high degree of accuracy. Inaccuracies may result from inaccuracies in computing the associated wind for each driver or due to the influence of other drivers such as TCs.

c. Extension to other MC coastlines

Sections 3a and 3b investigated the wind regimes on the south-west coast of Sumatra, how they relate to the diurnal cycle of convection and what large-scale conditions give rise to each regime. This region was chosen because of the *in situ* data from the pre-YMC and YMC field campaigns, and ERA5 was used to extend the analysis to gain more robust results. We now use ERA5 to repeat the analysis over the other MC coastlines in figure 1: north-west Borneo, north Java, south Java and north New Guinea.

Again, ERA5 wind between 1000 hPa and 500 hPa was averaged along the nominal coastline and the k-means algorithm was used with $k = 4$. For north-west Borneo, as for south-west Sumatra, zonal wind u was used; but v was used for the other three coastlines as they are oriented approximately east-west. Clusters were again sorted with cluster A being the most onshore and D the most offshore. Clusters A and D only are shown in figures 12a–h. Values of v tend to be weaker than u so the colour bar shown applies to three of the coastlines only; the north-west Borneo clusters use the same colour bar as south-west Sumatra (figure 6). Because onshore wind is always plotted as positive, the north Java and north New Guinea plots show $-v$. GPM-HQ composite Hovmöller diagrams are shown for clusters A and D only in figures 12i–p, with the mean orography plotted beneath panels (m)–(p).

The north-west Borneo clusters are similar to south-west Sumatra, except the mean u is less westerly so the onshore cluster is weaker and the offshore cluster is stronger. The associated precipitation is also similar to Sumatra, with large-scale precipitation dominating in the onshore regime and organised offshore propagation, preceded by a propagating region of suppression, in the offshore regime.

Since Java is a long, thin island oriented east-west, the winds over the north and south coasts are very similar (albeit plotted with the opposite sign in figures 12b,f and 12c,g). Thus, north Java's strong onshore cluster (A) approximately corresponds to south Java's offshore cluster (D), and *vice versa*. The exception is the land-sea breeze within the boundary layer, which is always anomalously onshore during the day, peaking around 13–14 LT. The same correspondence is seen in the Hovmöller diagrams in figures 12j,n and 12k,o, remembering that the positive onshore direction is southward for north Java and northward for south Java. However, the relationship between wind regime and propagation of convection is different for Java from the Sumatra and Borneo cases. Java has strong northward propagation (i.e., offshore for the north coast and inland for the south coast) regardless of wind direction (i.e., in both clusters A and D). However, the strong southward propagation (i.e., inland for the north coast and offshore for the south coast) occurs only when the wind is from the north (i.e., cluster A for north Java and cluster D for south Java).

North New Guinea has a very different coastal wind structure from the other coastlines investigated here, with all four clusters having onshore wind almost all day in at least the lowest part of the troposphere (up to around 825 hPa in cluster D; up to around 650 hPa in cluster B; and over the entire range 1000–500 hPa in clusters A and C, albeit weakly in cluster C). Convection over New Guinea is strongest on the north and south flanks of the New Guinea Highlands, which run east-west across the middle of the island and can be seen in the orography cross-section below figure 12p. The mean altitude along the section is around 2.2 km and the maximum is around 4.8 km, which is considerably higher than the orography on the other islands studied; and the mountains are considerably further from the coast (hundreds rather than tens of kilometres). Convection forms on both flanks of the mountain range regardless of coastal wind cluster, as also found by Hassim et al. (2016) in convection-permitting simulations. The convection propagates away from the mountains in both directions, but more strongly on the leeward side.

As for south-west Sumatra, we now investigate the large-scale drivers associated with the wind regimes at each coastline and quantify the extent to which they account for the variability in clusters (figure 13). The impact of ENSO is weaker over north-west Borneo (m_{ONI} from equation 1 is -0.74 m s^{-1} °C^{-1} with a correlation of $\rho = -0.28$; figure 13a) than south-west Sumatra ($m_{\mathrm{ONI}} = -1.38$ m s^{-1} °C^{-1} with $\rho = -0.41$; figure 5a). The impact of ENSO over the other three coastlines is negligible ($|m_{\mathrm{ONI}}|$ is never larger than 0.15 m s^{-1} °C^{-1}; figures 13d,g,j). The impact of the MJO is also less for north-west Borneo than south-west Sumatra, with the values in figure 13b around half the magnitude of figure 5b. For the other coastlines (figures 13e,h,k), the values are small (up to around 0.7 m s^{-1}) compared to the magnitude of the clusters (around ± 3 m s^{-1}; see figures 12b–d,f–h), suggesting the MJO also has a limited role in determining the wind regime for these coastlines. It is notable that the cross-term in equation 2 is very small for south-west Sumatra and north-west Borneo (the surfaces in figures 5b and 13b and nearly flat) but is much larger for the other coastlines.

Composites of 850 hPa wind anomaly and equatorial waves for clusters A and D are shown for selected coastlines south Java in figure 14 and north New Guinea in figure 15. Each figure shows the contributions from R1 (panels (c) and (h)), R2 (panels (d) and (i)) and WMRG (panels (e) and (j)) waves. Equivalent plots for the other coastlines are in the supplementary material – north-west Borneo (figure S2; Kelvin, R1 and R2 contributions are shown) is very similar to south-west Sumatra; and north Java (figure S3) is very similar to south Java as discussed below, but for a change of sign due to being on the opposite coast.

For south Java, R1 and WMRG waves both contribute strongly to v_{850} over the coast, in each cluster. The strong onshore cluster (A) has the R1 high pressure centres just west of Java so the eastern edge of the southern hemisphere anti-cyclone contributes positive v over the coastline (figure 14c). For the offshore cluster (D), the low pressure centres are just west of Java (figure 14h). WMRG waves consist of a quadrupole in pressure centred on the equator, with one phase having high pressure to the south-west and north-east, and low pressure to the north-west and south-east. This results in a dipole of vorticity, with clockwise rotation to the east and anti-clockwise to the west. This is the WMRG phase which exists in cluster A (figure 14e), with the vorticity centres located either side of Java and the two regions of vorticity together contributing positive v over the south coast. The opposite phase, with the opposite signs of pressure anomaly and vorticity, is seen in cluster D (figure 14j).

For north New Guinea, by far the greatest contributor to coastal wind is WMRG waves (figures 15e,j). R1 waves are very weak in the composites for both clusters shown (figures 15c,h), suggesting they do not have a consistent phase in these clusters, so they cancel each other out during the averaging. R2 waves have much more coherent structures than R1 so, unlike for south-west Sumatra (see section 3b), there is strong R2 propagation. Like WMRG waves, R2 has a quadrupole pressure structure but further away from the equator. Centred on the equator are regions of clockwise and anti-clockwise vorticity, with the western edges of these regions contributing to v over the coast in clusters A and D (figures 15d,i). In Cluster A, the western edge of a high pressure region is to the north of New Guinea and a low pressure region to the south; and *vice versa* for cluster D.

In section 3b it was shown that the residual u_{850} for south-west Sumatra, when calculating the total wind as the sum of the mean seasonal cycle and the six drivers investigated (ENSO, MJO and four equatorial waves), is less than one standard deviation on 83% of days (figure 11). The equivalent histograms for the other coastlines are figures 13c,f,i,l. The percentage of days with residual less than one standard deviation is 78% for north-west Borneo, 82% for north Java, 83% for south Java and 79% for north New Guinea. Again, this indicates the large-scale phenomena investigated can, to a high degree of accuracy, explain the total wind seen over the coastlines studied, thus exerting a control on the diurnal cycle of convection.

4. Summary and discussion

The MC has a pronounced diurnal cycle of precipitation due to warm oceans and a large number of islands, with the land-sea temperature contrast creating the conditions for deep convection in the afternoon and evening as moist air converges over land. On some islands the convection may be enhanced by significant orography. A weaker diurnal cycle exists over ocean and in some locations is modified by convection propagating offshore overnight. However, there is a lack of consensus regarding the mechanism of offshore propagation and its forcing by the large scale. Sources of variability in the diurnal cycle have been investigated in the literature but further understanding is required, including quantifying the contributions of large-scale controls. This is expected to aid forecasters in predicting the occurrence of extreme rainfall. Previous work has uncovered scale interactions with large-scale phenomena such as ENSO, the MJO and equatorial waves, but the variability of the diurnal cycle within any given phase of these phenomena is still considerable.

This study takes a fresh approach to diagnosing the interaction between the diurnal cycle and the large-scale environment. We take a local-to-large scale approach, rather than the large-to-local scale technique of producing local composites conditional on the large scale. We use a combination of *in situ* observations, satellite measurements and reanalysis data to investigate these scale interactions. We test the hypothesis that a range of large-scale drivers exert a control on MC precipitation through their modulation of the coastal wind regimes, and quantify the relative contributions of each driver.

The k-means clustering algorithm is used to define the coastal wind regimes, first using 108 days of field campaign radiosonde data on the coast of Sumatra before extending the analysis using ERA5 to 41 DJFs and the other coastlines labelled in figure 1. For most coastlines, the regime with strong onshore winds all day has suppressed convection over land and an active region of convection over the sea, with a lower than average probability of extreme rainfall (above the 99^{th} percentile) occurring. The regime with predominantly offshore winds produces the canonical diurnal cycle with strong convection over the land, a higher than average probability of extreme rainfall, and propagation offshore overnight. This is consistent with Yokoi et al. (2019) who showed, also using the pre-YMC and YMC observations, that around 800 hPa the wind is more offshore from south-west Sumatra on days with offshore propagation than on days without.

The 850 hPa wind anomaly associated with ENSO, the MJO and convectively-coupled equatorial waves was computed at each coastline, and used to quantify the contribution of each of these drivers to the coastal wind regime. The sum of these contributions explains the total wind to a high degree of accuracy, with the residual value having an amplitude smaller than one standard deviation on 78% and 83% of days. Remaining discrepancies are likely due to deficiencies in the technique for deriving the wind associated with each driver, or the influence of other drivers such as TCs, tropical depressions, cold surges, Borneo vortices or the Indian Ocean dipole. Composites of large-scale conditions with large residuals (not shown) suggest that the largest inaccuracies tend to be associated with either active phases of the MJO or R1 waves.

It is important to note that studies such as Peatman et al. (2014), which show the mean diurnal cycle of precipitation for each MJO phase, do not necessarily imply that the diurnal cycle plotted will always occur on days in that phase. Rather, the results of such papers are true only in a composite sense. MJO phase alone does not determine the small-scale weather regime and it cannot be used to predict the local distribution of thunderstorms on an individual day. Instead, a combination of large-scale phenomena, which may occur in any combination, should be considered.

Figure 16 summarises the drivers contributing to the onshore and offshore wind regimes for each coastline, with details of which phase of the large-scale phenomenon contributes to each cluster. Drivers are listed if the 1997/98 to 2017/18 DJF variance in their associated coastal 850 hPa wind (see figure S1 for the first three years) exceeds 10% of the variance of the total value, for each coastline (all variances are listed in table S4). The schematic diagrams in figure 17 indicate how these drivers affect coastal wind and show the associated variability in precipitation. Cumulonimbus clouds indicate deep convection and cumulus clouds indicate less intense precipitation. Grey arrows indicate propagation of convection, with the strength of propagation shown by the arrow size.

For south-west Sumatra there is a strong contribution from Rossby waves. Deep convection over the mountains and nocturnal offshore propagation are more likely during the high pressure phase of the R1 wave. The stronger high pressure signal is in the southern hemisphere because the coastline is in that hemisphere. When there is some asymmetry in the Rossby wave signal there is also some projection onto the R2 structure. Therefore, the two wave types are considered together in the top row of figure 16 and the asymmetry is illustrated in figure 17e. The opposite phase, with low pressure centres, is associated with more moderate, large-scale precipitation which propagates onshore (figure 17a). The same result is found for north-west Borneo (figures 17b,f), but with the stronger signals in the northern hemisphere because the coastline investigated is north of the equator.

The MJO and ENSO also contribute to the coastal wind for south-west Sumatra. The enhanced diurnal cycle is most often found in phases prior to the arrival of the active MJO envelope (8–3), consistent with Oh et al. (2012), Peatman et al. (2014) and other studies, when the wind is more offshore. The diurnal cycle is more suppressed during phases 4 to 6. El Niño is associated with offshore wind, consistent with large-scale suppression and low-level divergence over the region; and is therefore associated with an enhanced diurnal cycle, consistent with Rauniyar and Walsh (2013). The effect of ENSO is weaker over north-west Borneo than south-west Sumatra, possibly because the ENSO wind signal is due to large-scale convergence or divergence at the edge of the MC region. However, north-west Borneo does have a contribution from the MJO, with onshore and offshore wind regimes tending to occur one phase later than over south-west Sumatra due to the longitudinal difference.

Over Java and north New Guinea, where the coastal wind we consider is meridional, there is no substantial contribution from ENSO or the MJO, which is consistent with the fact that both these phenomena are associated with anomalies in a zonal overturning circulation. Variability in coastal wind regimes instead arises from the propagation of Rossby and WMRG waves. Since Java is a thin island in the north-south direction, onshore wind on the north coast often coincides with offshore wind on the south coast, and *vice versa*. The control of large-scale drivers on the coastal wind regimes is, therefore, roughly the same at each coastline but for a change of sign in wind direction (see yellow and blue rows of figure 16) so only south Java is shown in figure 17 (panels (c) and (g)). Rossby waves projecting onto R1 and R2 again have a strong contribution over Java, as do WMRG waves. Deep convection occurs over the Java mountains in both the onshore and offshore wind regimes, with northward propagation occurring in both. Southward propagation also occurs in both regimes but is considerably stronger when winds are offshore from the south coast.

Over north New Guinea, WMRG waves have the greatest contribution to coastal wind. Although the variance in coastal wind due to R1 waves is 14.9% of the variance in the total coastal wind, composites do not show a coherent signal (figures 15c,h) so there is no consistent phase of the wave associated with a given wind regime. Therefore, we do not provide information about the phase of these waves in figure 16 and we omit them from figures 17d,h. However, there is a coherent signal for R2 waves. The R2 structure has high and low pressure centres located away from the equator, and vorticity centred on the equator, as indicated in figures 17d,h. It is these equatorial vorticity centres which contribute to wind on the north New Guinea coast.

The New Guinea highlands are higher and broader than most of the orography in other MC regions studied in this paper (see figure 1). Unlike on other islands, there are two distinct regions of convective initiation, on the north and south flanks of the mountain range. As shown by the schematic diagram and as was seen in the Hovmöller diagrams in figures 12l,p, the most intense precipitation forms on the leeward side of the mountains. In both regimes the convection on the north side propagates northwards and on the south side propagates southwards, with the strongest propagation arising from the more intense convection.

The stark differences in the results for Java and north New Guinea, compared with south-west Sumatra and north-west Borneo, demonstrate the diversity in the behaviour of the diurnal cycle across the MC. More work is required to understand the differences in the underlying physical mechanisms occurring in each location.

In summary, we have shown that the strength and absolute wind direction within the coastal land-sea breeze circulation exerts a strong control on the diurnal convection and offshore propagation over MC coastlines. Our local-to-large scale method, using a clustering algorithm, is a powerful tool which has allowed us to identify what large-scale conditions set up each local regime. The summary information in figure 16 and figure 17 have applications for forecasting in the region. Global NWP models struggle to forecast the local-scale spatial and temporal distribution of rainfall accurately in the tropics, at least partly due to errors in convection parametrization schemes (e.g., Birch et al. 2015; Argüeso et al. 2020). However, large-scale drivers can be forecast skilfully several days or weeks in advance. For example, operational forecasting systems may skilfully predict the MJO on time scales of 3–4 weeks (e.g., Klingaman and Woolnough 2014; Kim et al. 2014); and Rossby and WMRG waves on time scales of around 1 week (e.g., Yang et al. 2021). Our results allow forecasters to harness such skill through understanding the impact of the large-scale environment and suggest there may be opportunities to infer the risk of high-impact weather from predicted large-scale weather regimes (cf. Neal et al. 2016, 2020). This information can be used alongside NWP forecasts to improve prediction of extreme rainfall.

Acknowledgments. This work is part of the Vertical Structure of Weather project, funded by the Met Office Weather and Climate Science for Service Partnership (WCSSP) Southeast Asia, as part of the Newton Fund; and the TerraMaris project, funded by National Environment Research Council (NERC) grant NE/R016739/1. The “pre-YMC” and “YMC” field campaign data were collected as part of YMC, led by investigator Dr Kunio Yoneyama. The equatorial wave data set was produced as part of the Newton Fund project under the auspices of the WCSSP Southeast

Asia project by Dr Gui-Ying Yang of the National Centre for Atmospheric Science, University of Reading. Data analysis and plots used the Iris Python package for Earth science data, version 2.4.0 (Met Office 2020). k-means clustering and multiple linear regression were performed using the Scikit-learn Python package for machine learning, version 0.22.1 (Pedregosa et al. 2011). The authors are grateful to Dr Steve Woolnough for providing Python code for filtering data in wavenumber-frequency space. The authors would like to thank three anonymous reviewers for their comments on this paper.

Data availability statement. The "pre-YMC" and "YMC" field campaign data are archived by the Japan Agency for Marine-Earth Science and Technology (JASMTEC) at `http://www.jamstec.go.jp/ymc/ymc_data.html`.

References

Argüeso, D., R. Romero, and V. Homar, 2020: Precipitation features of the Maritime Continent in parameterized and explicit convection models. *Journal of Climate*, **33 (6)**, 2449–2466, doi:10.1175/JCLI-D-19-0416.1.

Baranowski, D. B., M. K. Flatau, P. J. Flatau, and A. J. Matthews, 2016: Impact of atmospheric convectively coupled equatorial Kelvin waves on upper ocean variability: AIR-SEA INTERACTIONS IN CCKWS. *Journal of Geophysical Research: Atmospheres*, **121 (5)**, 2045–2059, doi:10.1002/2015JD024150.

Birch, C. E., M. J. Roberts, L. Garcia-Carreras, D. Ackerley, M. J. Reeder, A. P. Lock, and R. Schiemann, 2015: Sea-breeze dynamics and convection initiation: The influence of convective parameterization in weather and climate model biases. *Journal of Climate*, **28 (20)**, 8093–8108, doi:10.1175/JCLI-D-14-00850.1.

Birch, C. E., S. Webster, S. C. Peatman, D. J. Parker, A. J. Matthews, Y. Li, and M. E. E. Hassim, 2016: Scale interactions between the MJO and the western Maritime Continent. *Journal of Climate*, **29 (7)**, 2471–2492, doi:10.1175/JCLI-D-15-0557.1.

Climate Prediction Center, 2020: Oceanic Niño Index. URL https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**, 553–597, doi:10.1002/qj.828.

Ferrett, S., G.-Y. Yang, S. J. Woolnough, J. Methven, K. Hodges, and C. E. Holloway, 2020: Linking extreme precipitation in Southeast Asia to equatorial waves. *Quarterly Journal of the Royal Meteorological Society*, **146 (727)**, 665–684, doi:10.1002/qj.3699.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2). *Journal of Climate*, **30 (14)**, 5419–5454, doi:10.1175/JCLI-D-16-0758.1.

Hassim, M. E. E., T. P. Lane, and W. W. Grabowski, 2016: The diurnal cycle of rainfall over New Guinea in convection-permitting WRF simulations. *Atmospheric Chemistry and Physics*, **16 (1)**, 161–175, doi:10.5194/acp-16-161-2016.

Heale, C. J., J. B. Snively, A. N. Bhatt, L. Hoffmann, C. C. Stephan, and E. A. Kendall, 2019: Multilayer Observations and Modeling of Thunderstorm-Generated Gravity Waves Over the Midwestern United States. *Geophysical Research Letters*, doi:10.1029/2019GL085934.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146 (730)**, 1999–2049, doi:10.1002/qj.3803.

Houze, R. A., S. G. Geotis, F. D. Marks, and A. K. West, 1981: Winter monsoon convection in the vicinity of north Borneo. Part I: Structure and time variation of the clouds and precipitation. *Monthly Weather Review*, **109 (8)**, 1595–1614, doi:10.1175/1520-0493(1981)109<1595:WMCITV>2.0.CO;2.

Kim, H.-M., P. J. Webster, V. E. Toma, and D. Kim, 2014: Predictability and prediction skill of the MJO in two operational forecasting systems. *Journal of Climate*, **27 (14)**, 5364–5378, doi:10.1175/JCLI-D-13-00480.1.

Klingaman, N. P., and S. J. Woolnough, 2014: Using a case-study approach to improve the Madden-Julian oscillation in the Hadley Centre model. *Quarterly Journal of the Royal Meteorological Society*, **140 (685)**, 2491–2505, doi:10.1002/qj.2314.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS) project: Unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, **91**, 363–376, doi:10.1175/2009BAMS2755.1.

Love, B. S., A. J. Matthews, and G. M. S. Lister, 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, doi:10.1002/qj.809.

Lubis, S. W., and M. R. Respati, 2021: Impacts of convectively coupled equatorial waves on rainfall extremes in Java, Indonesia. *International Journal of Climatology*, doi:10.1002/joc.6967.

MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Volume I: Theory of Statistics*, 281–297.

Madden, R. A., and P. R. Julian, 1971: Detection of a 40-50 day oscillation in the zonal wind in the tropical Pacific. *Journal of the Atmospheric Sciences*, **28 (5)**, 702–708, doi:10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2.

Madden, R. A., and P. R. Julian, 1972: Description of global-scale circulation cells in the tropics with a 40-50 day period. *Journal of the Atmospheric Sciences*, **29 (6)**, 1109–1123, doi:10.1175/1520-0469(1972)029<1109:DOGSCC>2.0.CO;2.

Mapes, B. E., T. T. Warner, and M. Xu, 2003: Diurnal patterns of rainfall in northwestern South America. Part III: Diurnal gravity waves and nocturnal convection offshore. *Monthly Weather Review*, **131 (5)**, 830–844, doi:10.1175/1520-0493(2003)131<0830:DPORIN>2.0.CO;2.

Matthews, A. J., 2000: Propagation mechanisms for the Madden-Julian Oscillation. *Quarterly Journal of the Royal Meteorological Society*, **126 (569)**, 2637–2651, doi:10.1002/qj.49712656902.

Met Office, 2020: Iris: A Python library for analysing and visualising meteorological and oceanographic data sets. *http://scitools.org.uk*.

Mohd Nor, M. F. F., C. E. Holloway, and P. M. Inness, 2020: The role of local orography on the development of a severe rainfall event over western peninsular Malaysia: A case study. *Monthly Weather Review*, **148 (5)**, 2191–2209, doi:10.1175/MWR-D-18-0413.1.

Mori, S., H. Jun-Ichi, Y. I. Tauhid, and M. D. Yamanaka, 2004: Diurnal land-sea rainfall peak migration over Sumatera Island, Indonesian Maritime Continent, observed by TRMM satellite and intensive rawinsonde soundings. *Monthly Weather Review*, **132 (8)**, 2021–2039, doi:10.1175/1520-0493(2004)132<2021:DLRPMO>2.0.CO;2.

Muhammad, F. R., S. W. Lubis, and S. Setiawan, 2021: Impacts of the Madden-Julian Oscillation on precipitation extremes in Indonesia. *International Journal of Climatology*, doi:10.1002/joc.6941.

Neal, R., D. Fereday, R. Crocker, and R. E. Comer, 2016: A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteorological Applications*, **23 (3)**, 389–400, doi:/10.1002/met.1563.

Neal, R., J. Robbins, R. Dankers, A. Mitra, A. Jayakumar, E. N. Rajagopal, and G. Adamson, 2020: Deriving optimal weather pattern definitions for the representation of precipitation variability over India. *International Journal of Climatology*, **40 (1)**, 342–360, doi:/10.1002/joc.6215.

Oh, J. H., K. Y. Kim, and G. H. Lim, 2012: Impact of MJO on the diurnal cycle of rainfall over the western Maritime Continent in the austral summer. *Climate Dynamics*, **38 (5–6)**, 1167–1180, doi:10.1007/s00382-011-1237-4.

Peatman, S. C., A. J. Matthews, and D. P. Stevens, 2014: Propagation of the Madden-Julian Oscillation through the Maritime Continent and scale interaction with the diurnal cycle of precipitation. *Quarterly Journal of the Royal Meteorological Society*, **140 (680)**, 814–825, doi:10.1002/qj.2161.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12 (85)**, 2825–2830.

Qian, J.-H., 2008: Why precipitation is mostly concentrated over islands in the Maritime Continent. *Journal of the Atmospheric Sciences*, **65 (4)**, 1428–1441, doi:10.1175/2007JAS2422.1.

Qian, J.-H., 2020: Mechanisms for the dipolar patterns of rainfall variability over large islands in the Maritime Continent associated with the Madden-Julian Oscillation. *Journal of the Atmospheric Sciences*, **77 (6)**, 2257–2278, doi:10.1175/JAS-D-19-0091.1.

Ramage, C. S., 1968: Role of a Tropical "Maritime Continent" in the atmospheric circulation. *Monthly Weather Review*, **96 (6)**, 365–370, doi:10.1175/1520-0493(1968)096<0365:ROATMC>2.0.CO;2.

Rauniyar, S. P., and K. J. E. Walsh, 2011: Scale interaction of the diurnal cycle of rainfall over the Maritime Continent and Australia: Influence of the MJO. *Journal of Climate*, **24 (2)**, 325–348, doi:10.1175/2010JCLI3673.1.

Rauniyar, S. P., and K. J. E. Walsh, 2013: Influence of ENSO on the diurnal cycle of rainfall over the Maritime Continent and Australia. *Journal of Climate*, **26 (4)**, 1304–1321, doi:10.1175/JCLI-D-12-00124.1.

Sakaeda, N., G. Kiladis, and J. Dias, 2020: The diurnal cycle of rainfall and the convectively coupled equatorial waves over the Maritime Continent. *Journal of Climate*, **33 (8)**, 3307–3331, doi:10.1175/JCLI-D-19-0043.1.

Vincent, C. L., and T. P. Lane, 2016: Evolution of the diurnal precipitation cycle with the passage of a Madden-Julian Oscillation event through the Maritime Continent. *Monthly Weather Review*, **144 (5)**, 1983–2005, doi:10.1175/MWR-D-15-0326.1.

Warner, T. T., B. E. Mapes, and M. Xu, 2003: Diurnal patterns of rainfall in northwestern South America. Part II: Model simulations. *Monthly Weather Review*, **131 (5)**, 813–829, doi:10.1175/1520-0493(2003)131<0813:DPORIN>2.0.CO;2.

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

Wu, P., M. Hara, J.-i. Hamada, M. D. Yamanaka, and F. Kimura, 2009: Why a large amount of rain falls over the sea in the vicinity of western Sumatra Island during nighttime. *Journal of Applied Meteorology and Climatology*, **48 (7)**, 1345–1361, doi:10.1175/2009JAMC2052.1.

Xavier, P., R. Rahmat, W. K. Cheong, and E. Wallace, 2014: Influence of Madden-Julian Oscillation on Southeast Asia rainfall extremes: Observations and predictability. *Geophysical Research Letters*, **41 (12)**, 4406–4412, doi:10.1002/2014GL060241.

Yang, G.-Y., S. Ferrett, S. Woolnough, J. Methven, and C. Holloway, 2021: Real-time identification of equatorial waves and evaluation of waves in global forecasts. *Weather and Forecasting*, **36 (1)**, 171–193, doi:10.1175/WAF-D-20-0144.1.

Yang, G.-Y., B. J. Hoskins, and J. M. Slingo, 2003: Convectively coupled equatorial waves: A new methodology for identifying wave structures in observational data. *Journal of the Atmospheric Sciences*, **60 (14)**, 1637–1654, doi:10.1175/1520-0469(2003)060<1637:CCEWAN>2.0.CO;2.

Yokoi, S., S. Mori, F. Syamsudin, U. Haryoko, and B. Geng, 2019: Environmental conditions for nighttime offshore migration of precipitation area as revealed by in situ observation off Sumatra Island. *Monthly Weather Review*, **147 (9)**, 3391–3407, doi:10.1175/MWR-D-18-0412.1.

Yoneyama, K., and C. Zhang, 2020: Years of the Maritime Continent. *Geophysical Research Letters*, **47 (12)**, e2020GL087 182, doi:10.1029/2020GL087182.

LIST OF FIGURES

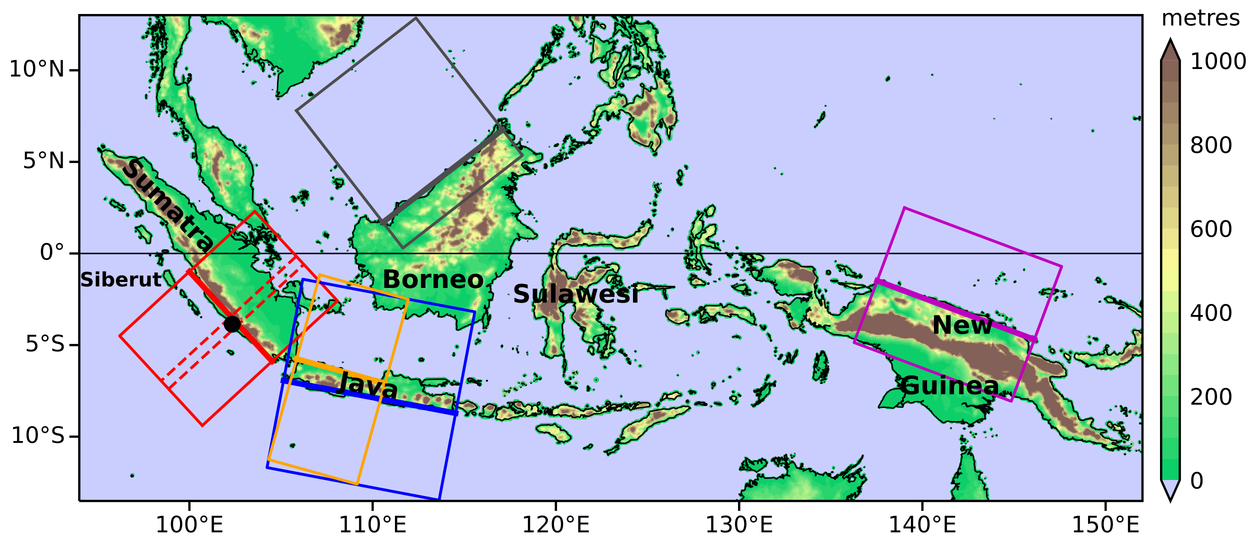

FIG. 1. Orography of the MC with selected islands labelled. Solid boxes indicate rotated grids used in the analysis, with the internal thick line indicating the nominal coastline, for south-west Sumatra (red), north-west Borneo (grey), north Java (orange), south Java (blue) and north New Guinea (magenta). The red dashed box is used instead of the solid box when analysing data from the field campaigns at Bengkulu. Bengkulu is shown by the black dot inside the dashed box.

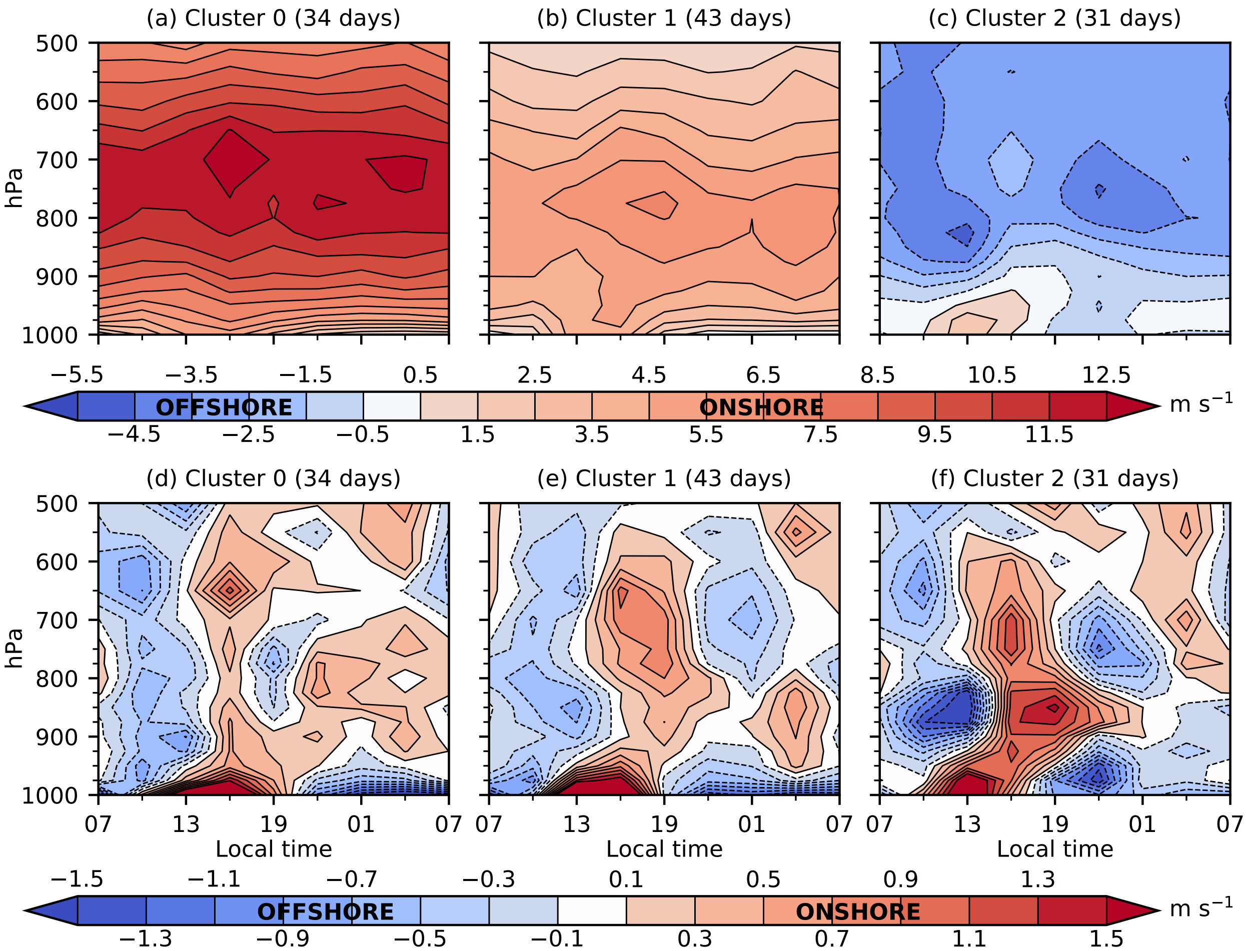

FIG. 2. (a–c) Diurnal cycle of vertical profile of zonal wind from radiosondes at Bengkulu during the 108 days 2015/09/11–2015/12/25 and 2017/11/16–2018/01/15. Days are sorted into 3 clusters using the *k*-means algorithm (see main text for details), with the number of days in each cluster shown above each panel. Clusters are ordered according to the mean value of zonal wind averaged over time and pressure. Radiosonde data are every 3 hours and interpolated to the same pressure levels used by the ERA5 reanalysis, from 1000 to 500 hPa, as indicated by the tick marks. Local time is taken to be UTC+7. (d–f) As (a–c) but plotted as a diurnal anomaly; that is, with the respective daily mean profile for each cluster subtracted.

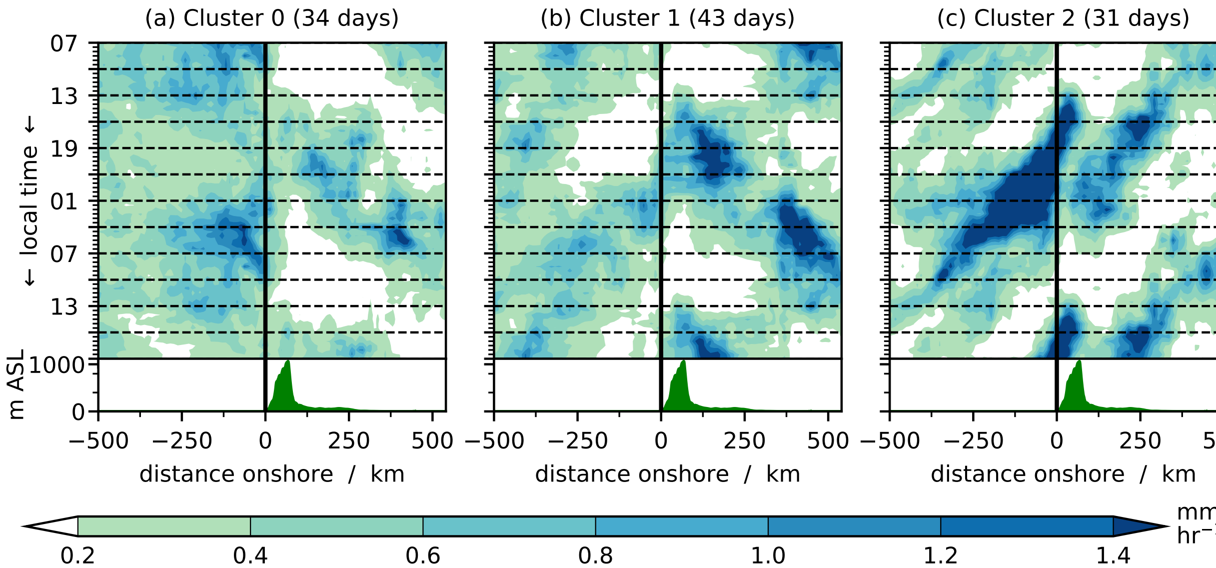

Fig. 3. Hovmöller diagrams of the composite diurnal cycle of precipitation from IMERG for each of the clusters in figure 2, extended by 12 hours to 19 LT (12 UTC) the following day. Precipitation is composited over the red dashed box in figure 1 and averaged along the coastline, every 30 minutes. The mean orography is shown below each panel in metres above sea level (ASL).

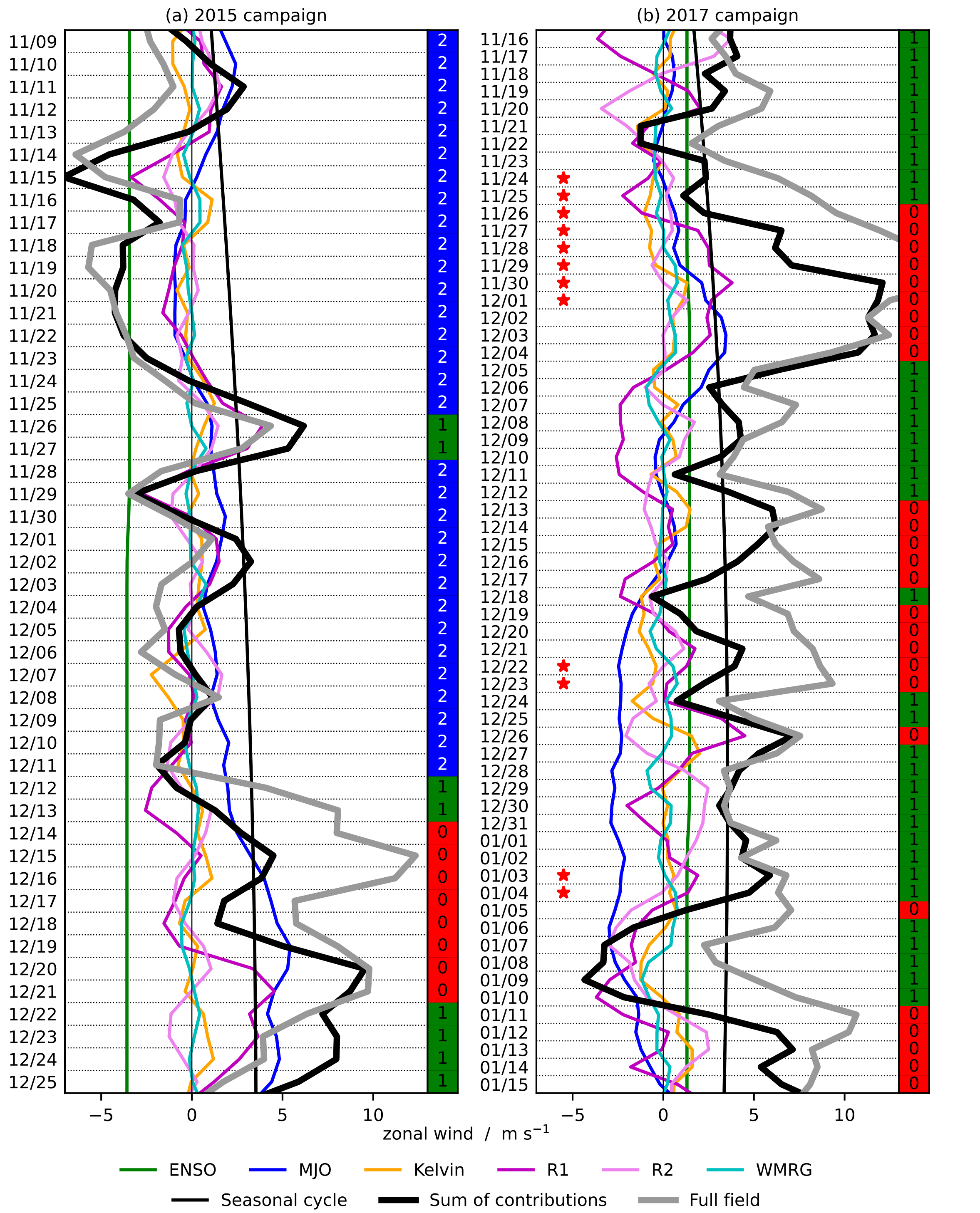

FIG. 4. Time series of u_{850} (grey) from ERA-Interim at Bengkulu during the two field campaigns. Coloured curves show u'_{850} due to ENSO (dark green; regression onto ONI), the MJO (dark blue; multiple regression onto RMMs), equatorial Kelvin (orange), R1 (dark pink), R2 (light pink) and WMRG (light blue) waves (identified in ERA-Interim). Thin b~~B~~lack ~~dotted~~ curve shows the mean seasonal cycle. Thick b~~B~~lack ~~solid~~ curve is the sum of the six coloured curves and the mean seasonal cycle. Red stars indicate a TC centre within 10° geodesic of Bengkulu. (All TCs shown are in the southern hemisphere.) Red, green and blue boxes denote clusters from figures 2a–c.

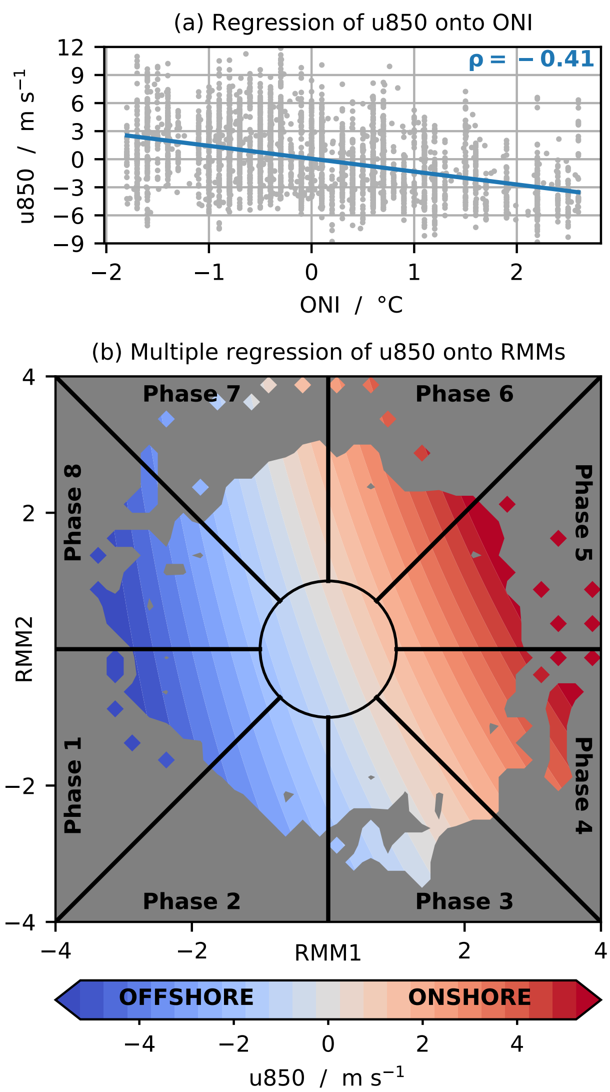

FIG. 5. (a) Linear regression of south-west Sumatra coast u'_{850} onto ONI index of ENSO, according to equation 1. The data points used for the regression are shown in grey. ρ is the correlation coefficient. (b) Multiple linear regression of south-west Sumatra coast u'_{850} onto RMM indices of the MJO, according to equation 2. Areas with no MJO days in the 39 DJFs used are greyed out.

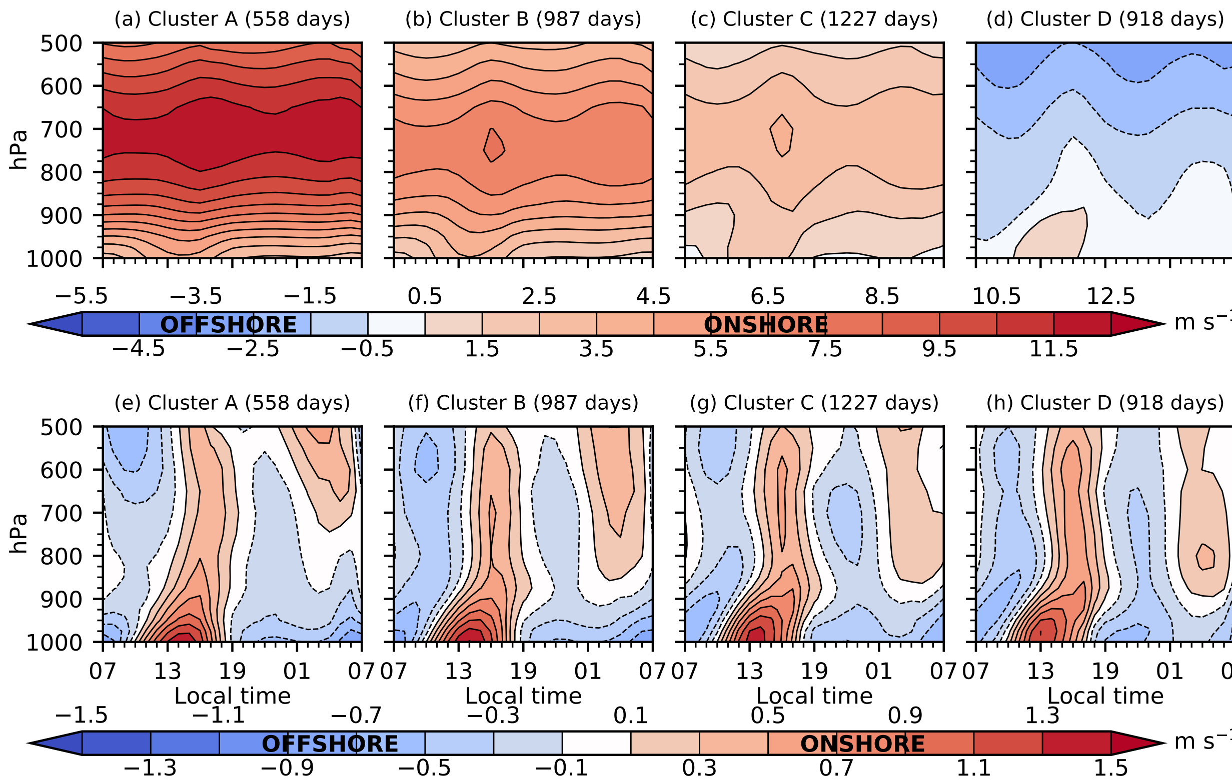

FIG. 6. (a–d) Hourly ERA5 zonal wind averaged along the nominal coastline of the solid red box (south-west Sumatra) in figure 1, clustered using the k-means algorithm. 41 DJFs (1979/80–2019/20, excluding 29 February) are used with $k = 4$ (see main text for details of the choice of k) and clusters labelled A, B, C and D to avoid confusion with the $k = 3$ clustering in figure 2. (e–h) As (a–d) but plotted as a diurnal anomaly; that is, with the respective daily mean profile for each cluster subtracted.

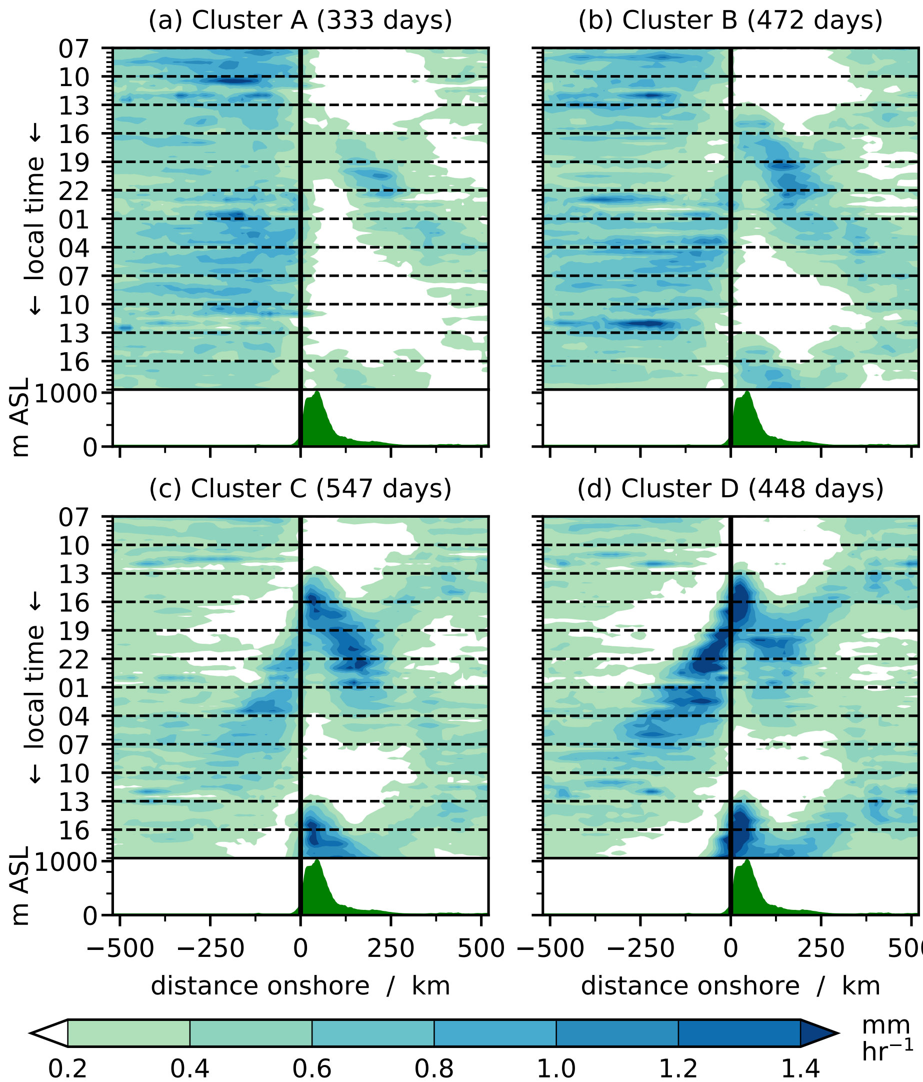

FIG. 7. Hovmöller diagrams of the composite diurnal cycle of precipitation from GPM-HQ for each of the clusters in figure 6. These composites use a subset (20 DJFs) of the days in figure 6 as GPM is not available before the year 2000.

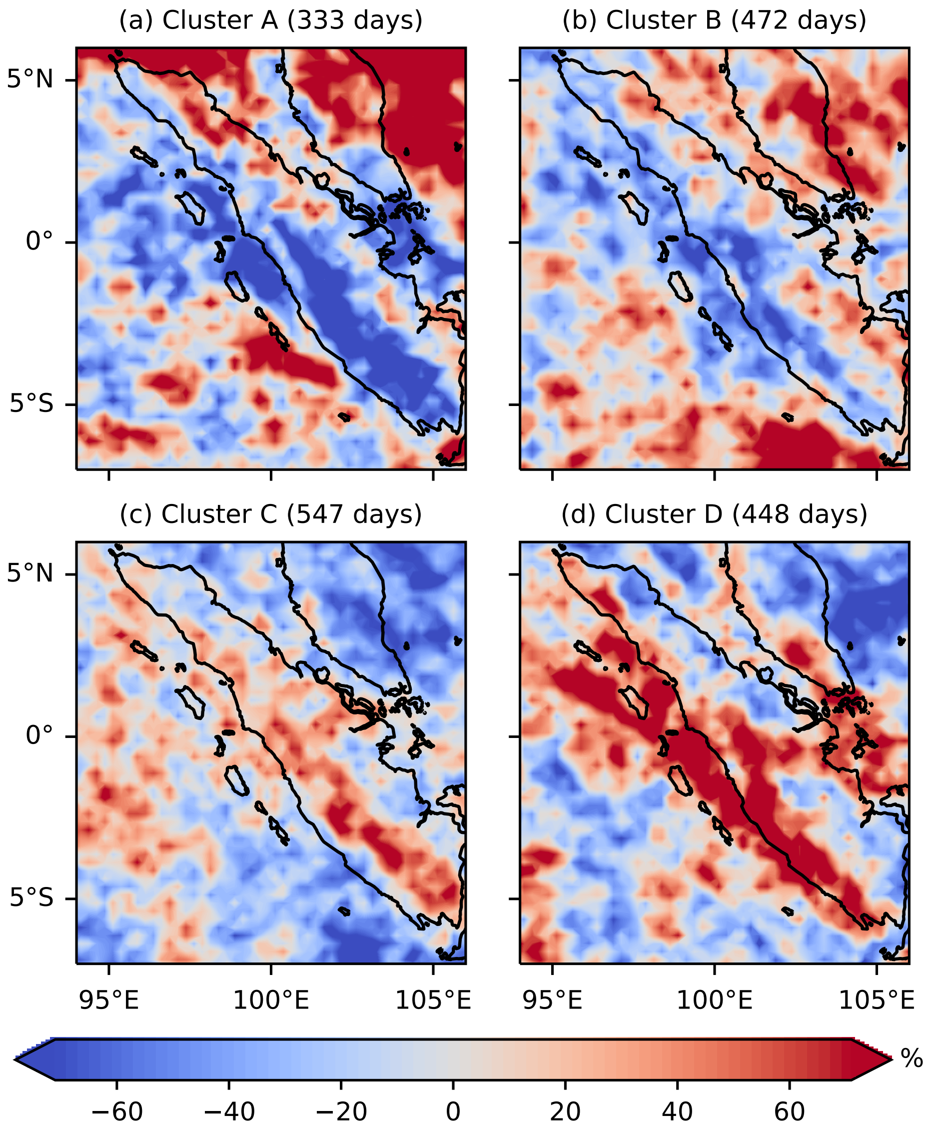

FIG. 8. Fractional change (expressed as a percentage), relative to climatology, of the probability of extreme precipitation occurring in each cluster, using daily mean ~~IMERG~~GPM-HQ rainfall for the 20 DJFs from 2000/01 onwards. Extreme precipitation is defined as exceeding the 99th percentile, so the climatological probability is 0.01 everywhere. Data are regridded to 0.25° for plotting purposes.

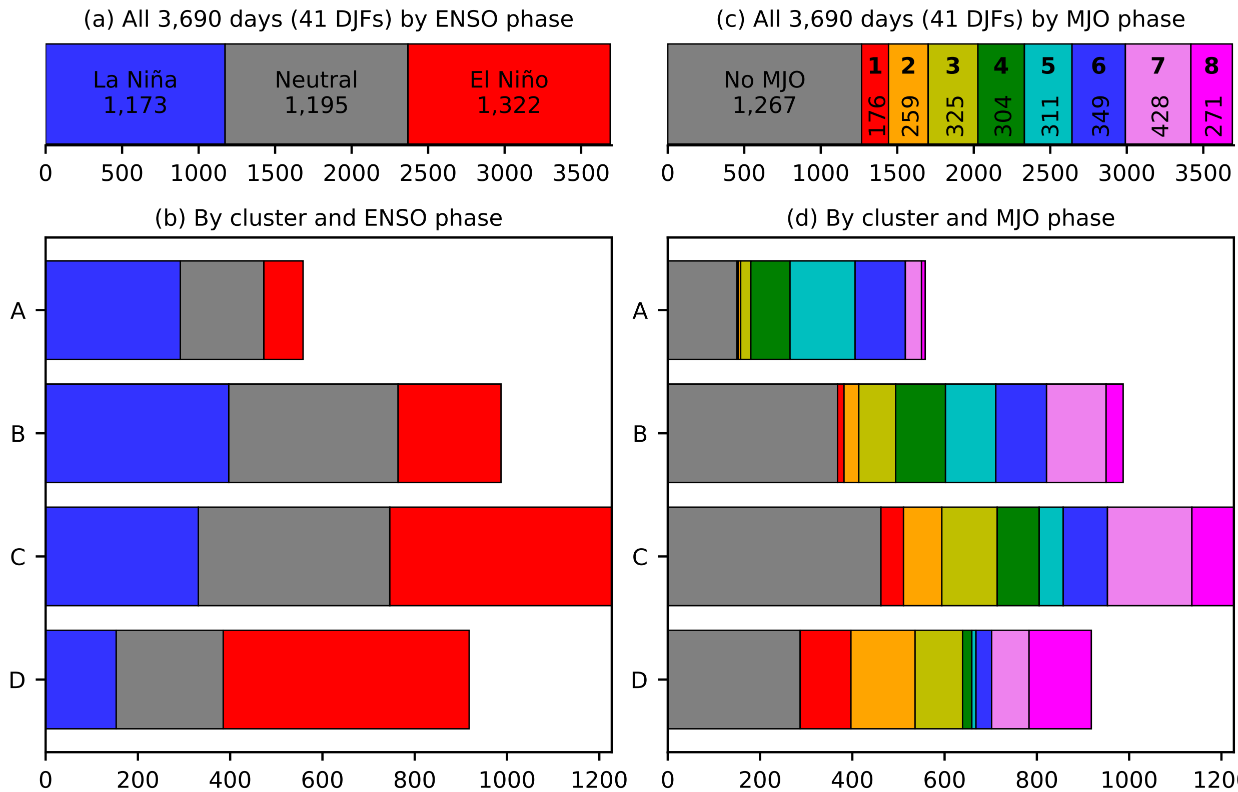

FIG. 9. (a) Number of days in DJFs 1979/80–2019/20 in each phase of ENSO. (b) As (a) but separately for each of the clusters in figure 6. (c,d) As (a,b) but for the MJO, with grey indicating RMM amplitude < 1. MJO large-scale convection is active over the Indian Ocean in phases 2–3, the MC in 4–5, the western Pacific Ocean in 6–7, and the western hemisphere and Africa in 8–1; and is suppressed in the corresponding opposite phases.

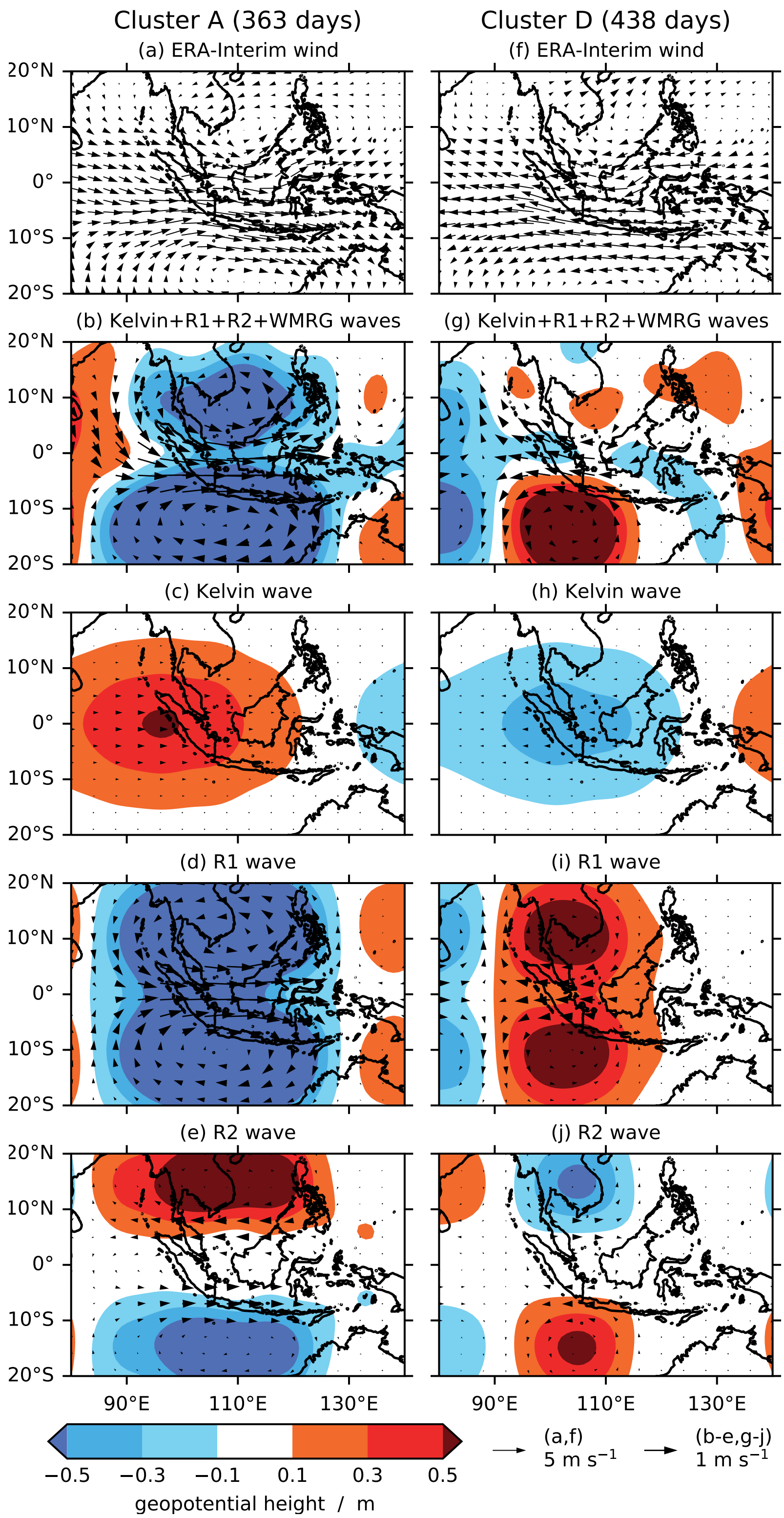

FIG. 10. Composites at 850 hPa for clusters A and D from figures 6a,d. (a,f) Daily mean horizontal wind anomaly using ERA-Interim reanalysis data. (b,g) Daily mean horizontal wind and geopotential height anomalies from structures of equatorial waves, summed over Kelvin, R1, R2 and WMRG waves. (c,h) As (b,g) but for Kelvin wave contribution only. (d,i) As (b,g) but for R1 wave contribution only. (e,j) As (b,g) but for R2 wave contribution only. All composites use a subset (21 DJFs) of the days in figure 6, for 1997/98 to 2017/18.

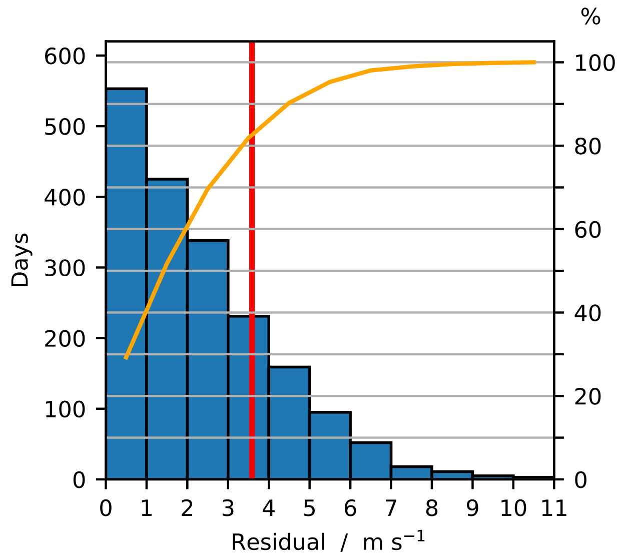

FIG. 11. Histogram of the absolute value of the residual south-west Sumatra coast u_{850}, with the theoretical value computed as the sum of the contributions from the MJO, ENSO, equatorial Kelvin, R1, R2 and WMRG waves, and the mean seasonal cycle. For the example days shown in figures 4 and S1, this is the absolute difference between the grey and ~~solid~~thick black curves. Blue bars show the number of days in each bin (left vertical axis); orange curve shows the cumulative distribution as a percentage (right vertical axis); and red line shows the inter-annual standard deviation of u_{850}, averaged over each day of DJF.

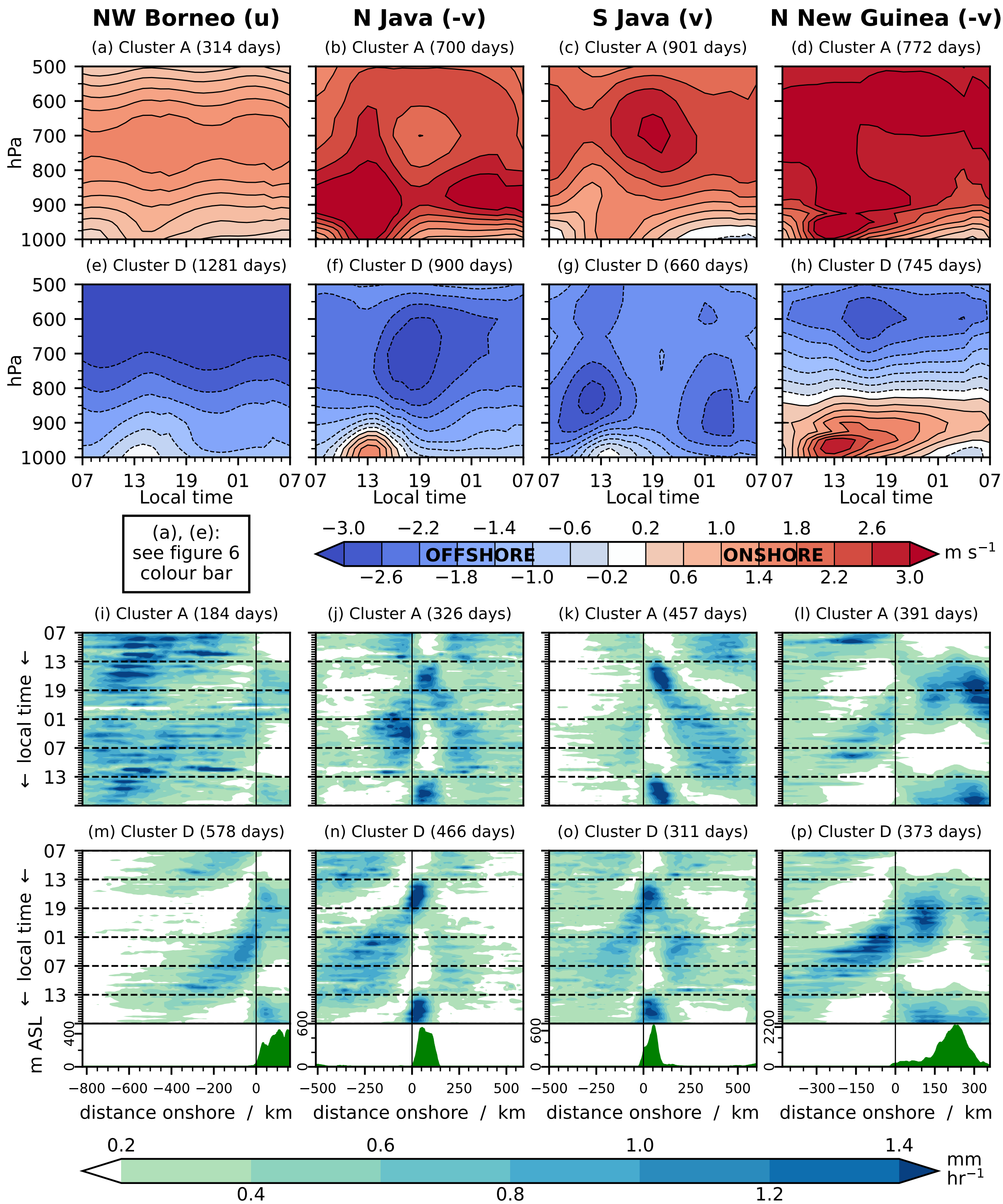

FIG. 12. (a–h) As figure 6 but extended to the other coastlines shown in figure 1, for $k = 4$ (clusters B and C not shown). u and v after each coastline name indicate whether zonal or meridional wind was clustered. Wind in the onshore direction is plotted in red and the offshore direction in blue, with cluster A as the most onshore cluster. For north Java and north New Guinea this requires the sign to be reversed so $-v$ is plotted. Note that panels (a) and (e) use the same colour bar (not shown here) as figure 6. (i–p) As figure 7 but for the clusters shown here in panels a–h.

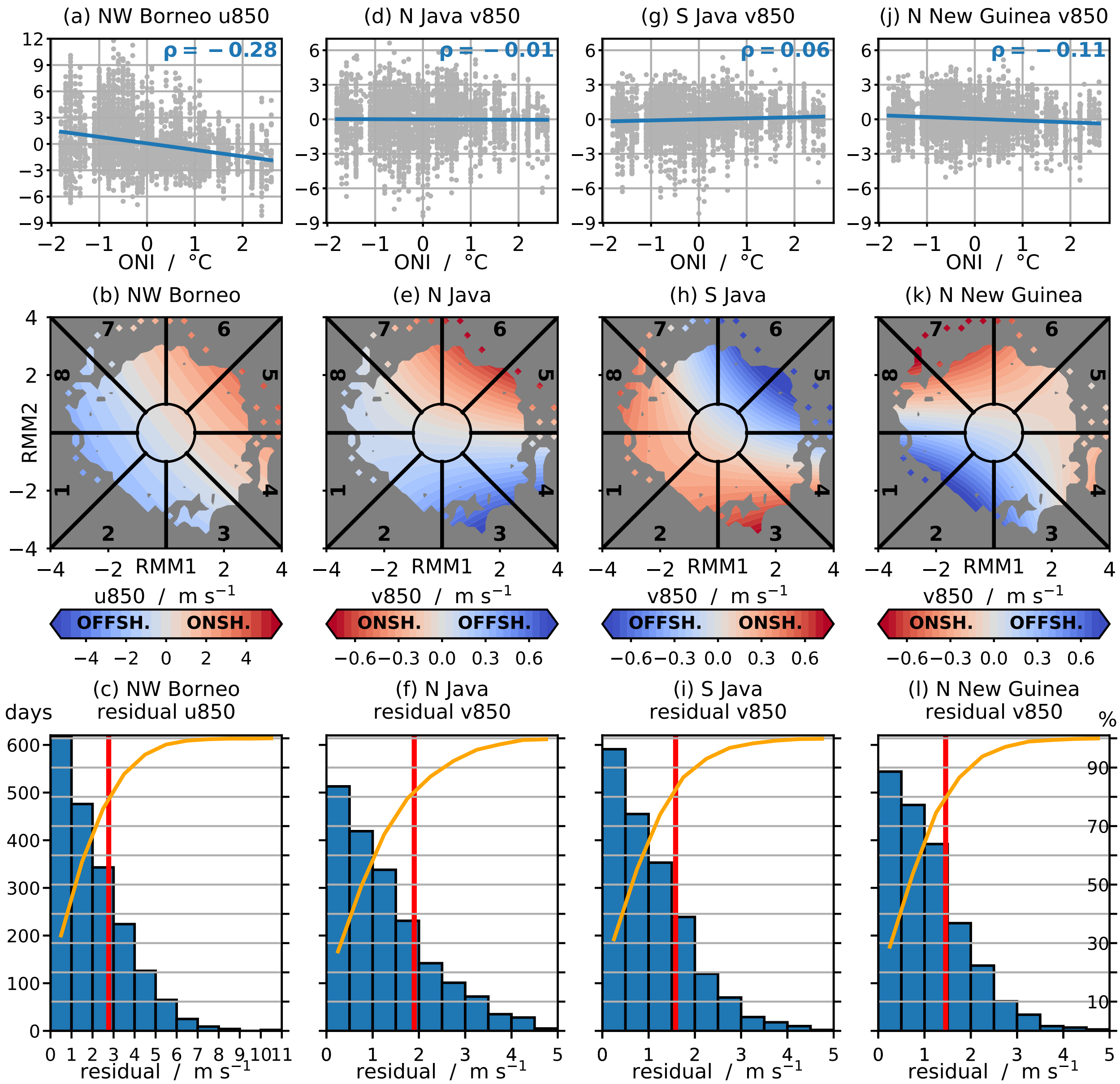

FIG. 13. Top two rows as figure 5 and bottom row as figure 11, but for (a–c) north-west Borneo, (d–f) north Java, (g–i) south Java and (j–l) north New Guinea. For middle row, colours are chosen so that red is always onshore wind. For bottom row, tick marks up to 600 (number of days in each bin; left axis) relate to blue bars; and horizontal grid lines and tick marks up to 100 (cumulative distribution as a percentage; right axis) relate to orange curves.

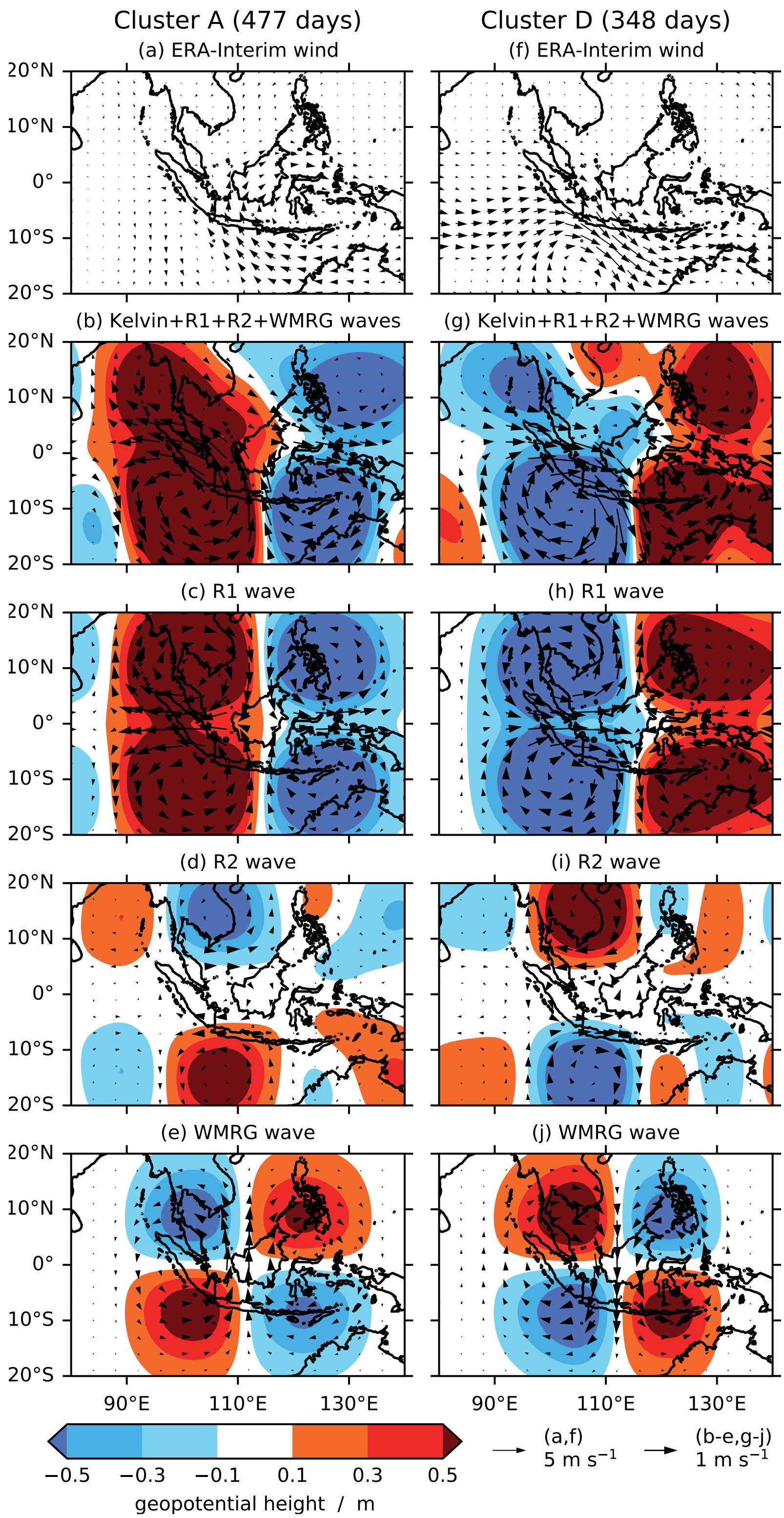

FIG. 14. (a,b,f,g) As figure 10 but for south Java clusters A and D (see figures 12c,g). (c,h) R1, (d,i) R2 and (e,j) WMRG contributions only.

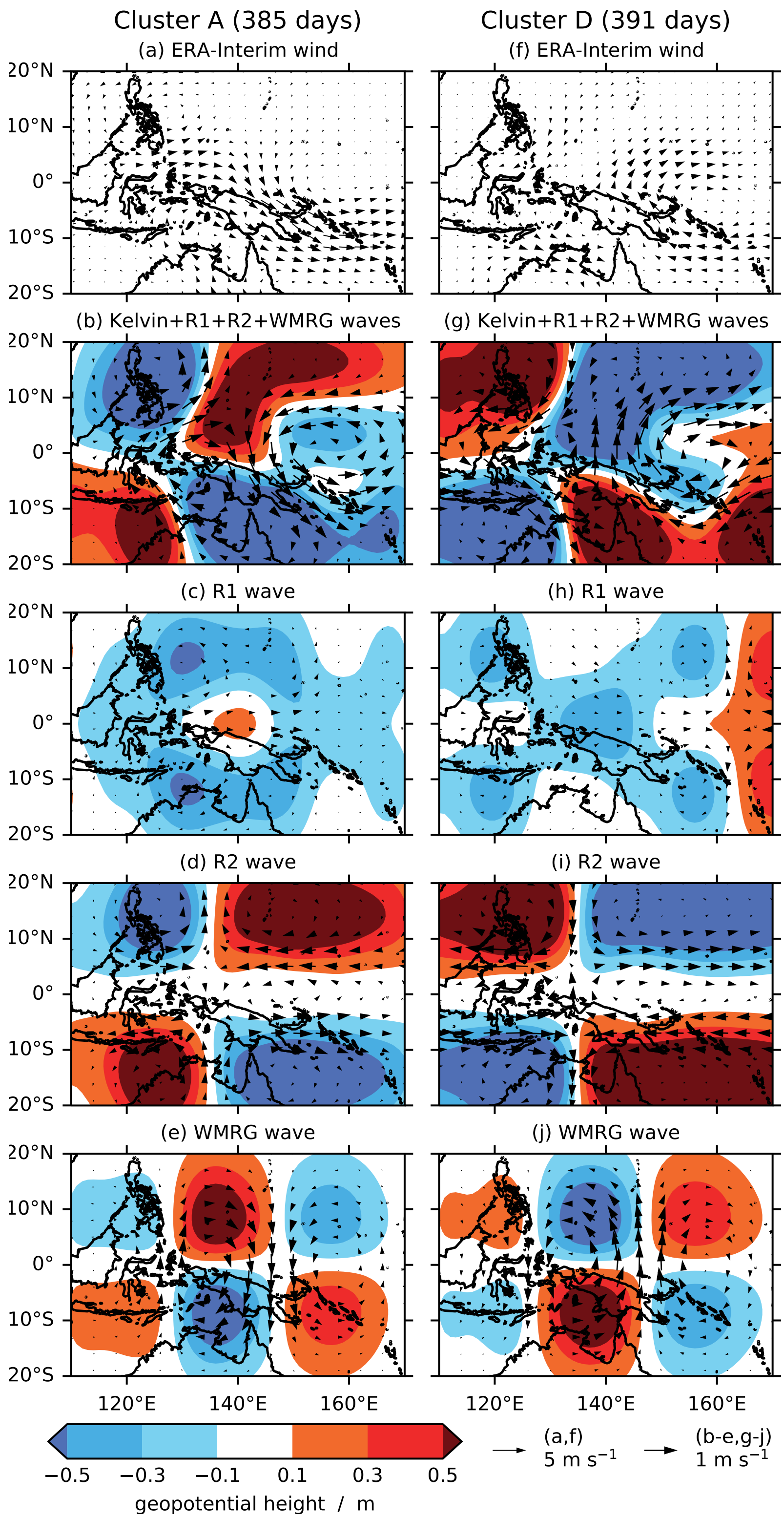

FIG. 15. As figure 14 but for north New Guinea clusters A and D (see figures 12d,h).

Coastline	Driver and % variance	Onshore regime (suppressed DC)	Offshore regime (enhanced DC)
South-west Sumatra	R1+R2* (43.3%)	Cyclonic phase (strong in SH)	Anti-cyclonic phase (strong in SH)
	MJO (24.8%)	Phases 4–6	Phases 8–3
	ENSO (17.2%)	La Niña	El Niño
North-west Borneo	R1+R2* (40.4%)	Cyclonic phase (strong in NH)	Anti-cyclonic phase (strong in NH)
	MJO (10.0%)	Phases 5–7	Phases 1–4
North Java	WMRG (32.4%)	High pressure NW & SE Low pressure SW & NE	High pressure SW & NE Low pressure NW & SE
	R1+R2* (29.2%)	Cyclone centre slightly to W (strong in SH)	Anti-cyclone centre slightly to W (strong in SH)
South Java	R1+R2* (48.3%)	Anti-cyclone centre slightly to W (strong in SH)	Cyclone centre slightly to W (strong in SH)
	WMRG (25.8%)	High pressure SW & NE Low pressure NW & SE	High pressure NW & SE Low pressure SW & NE
North New Guinea	WMRG (44.0%)	High pressure N & SE Low pressure S & NE	High pressure S & NE Low pressure N & SE
	R1** (14.9%)	—	—
	R2 (10.3%)	High pressure SW & N/NE Low pressure NW & S/SE	High pressure NW & S/SE Low pressure SW & N/NE

FIG. 16. Summary of the large-scale drivers with the strongest control on coastal wind regime, and the phases associated with the strongest onshore and offshore regimes, for each of the coastlines analysed. Forcings are listed if the variance in their associated u'_{850} or v'_{850} (depending on the coastline) is at least 10% of the variance in the full u_{850} or v_{850} field during DJFs 1997/98 to 2017/18; for a full list, see table S4. Colours correspond to the boxes in figure 1, with more intense shades highlighting larger contributions to the variability. *For coastlines other than north New Guinea, identified R2 waves appear to be an artefact of the same circulation pattern which projects more strongly onto the theoretical R1 structure, so the two are shown in combination. **For north New Guinea, although the variance associated with R1 waves is high enough to be listed here, their

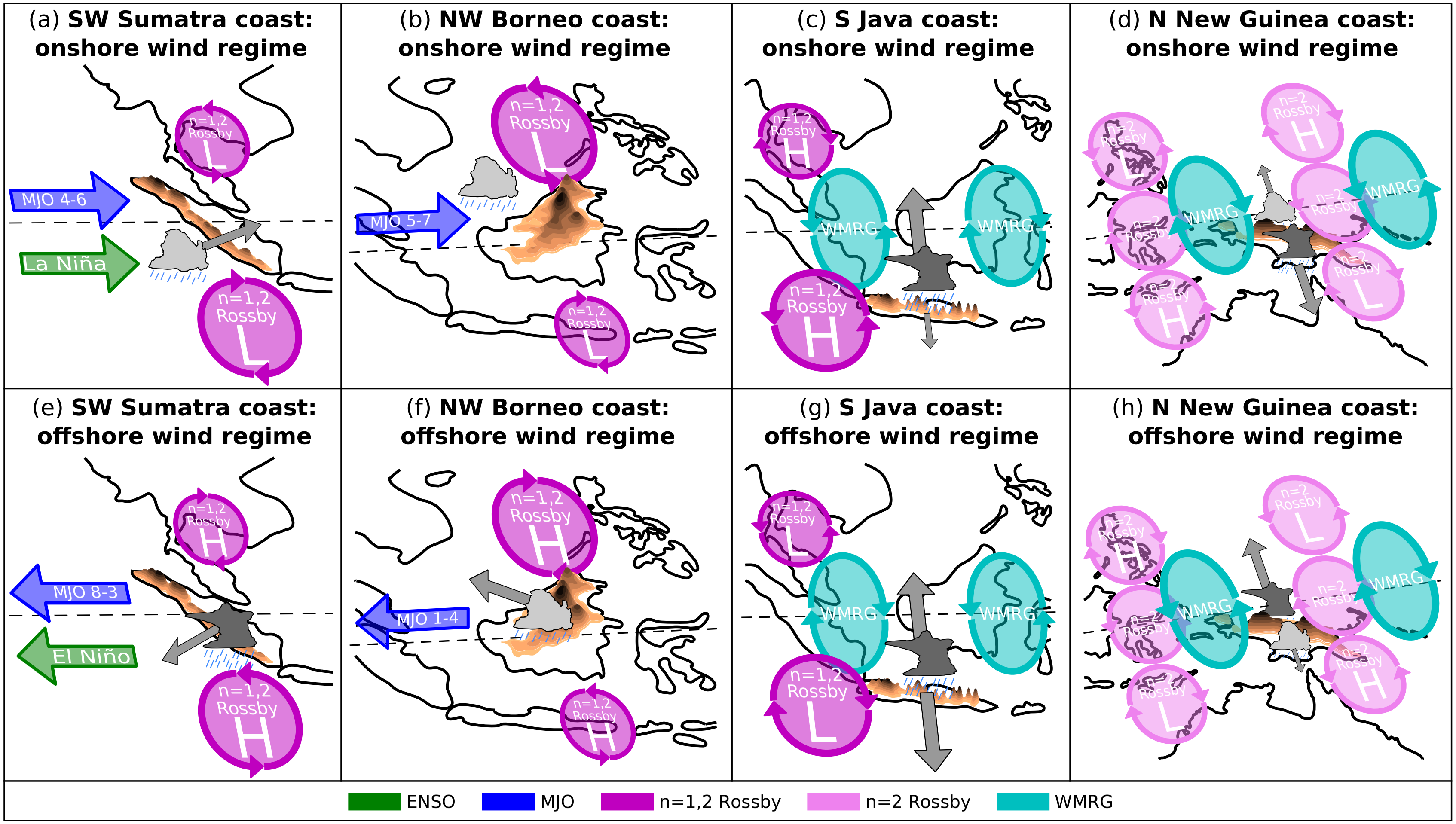

FIG. 17. Schematic diagram of important large-scale drivers, and their effect on convection and its propagation, for strong onshore and offshore coastal wind regimes for the Maritime Continent coastlines detailed in figure 16. North Java is not shown due to its similarity to the results for south Java (with onshore and offshore reversed). Dark grey cumulonimbus clouds indicate intense convection and rainfall; light grey cumulus clouds indicate more moderate rainfall. Grey arrows indicate propagation, with the largest arrow in an image indicating the strongest propagation (not to scale between images). These are drawn with reference to the Hovmöller diagrams in figures 7 and 12i–p. Coloured arrows indicate the wind direction associated with large-scale drivers. H and L refer to high and low pressure centres. For $n = 1, 2$ Rossby waves these differ in size between hemispheres (see main text for details). MJO numbers refer to phases from Wheeler and Hendon (2004). The dashed line marks the equator. Major mountains are shown for each island (relative size between images not to scale).