
ORIGINAL ARTICLE

Journal Section

Characterisation of the observed diurnal cycle of precipitation over the Maritime Continent

Jack M. Mustafa¹ | Adrian J. Matthews^{1,2} | Rob
A. Hall¹ | Karen J. Heywood¹ | Marina V. C. Azaneu¹

This study investigates the temporal and spatial complexities of the mean diurnal cycle (DC) of precipitation over the Maritime Continent during the wet season using the IMERG data product, and highlights systematic inaccuracies of amplitude and phase representation using the first diurnal harmonic (FDH). The first-order nature of the DC of precipitation is already well documented, typically featuring heavy precipitation over near-coastal land areas in the late afternoon and evening followed by maximum precipitation overnight over the surrounding seas, with offshore propagation evident in places. The DC is often concisely described in terms of an amplitude and phase based on the FDH parameters, however the omission of higher-order components of variability results in the FDH parameters being poor indicators of the magnitude and peak time of diurnal variability in many locations. This study improves the accuracy of the amplitude and phase parameters by characterising the DC using two novel waveforms – a skew-permitting waveform and a spike-permitting waveform – which are constructed to more accurately characterise single-peak cycles with rapid transitions. Key characterisation improvements include correction of a phase lag (averaging about one hour) over near-coastal land areas and capture of the short-lasting but extreme peak in precipitation rate over Java which increases the amplitude by the order of 20%. The new skew parameter shows that locations close to coastlines experience rapid intensification and gradual weakening of diurnal precipitation while there is a tendency toward gradual intensification and rapid weakening far inland and offshore. The new spike parameter shows that near-coastal land experiences a brief and precisely timed peak in precipitation, whereas diurnal activity over inland locations is less precisely timed, and waters surrounding Java experience a precisely timed suppression of precipitation. Other potential applications of the novel waveforms used in this study are discussed.

¹Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom

²School of Mathematics, University of East Anglia, Norwich, United Kingdom

Correspondence

J. Mustafa, School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom
Email: j.mustafa@uea.ac.uk

Funding information

Natural Environment Research Council through TerraMaris project, grant NE/R016704/1.

KEYWORDS

diurnal cycle, precipitation, Maritime Continent, characterisation, amplitude, phase, GPM-IMERG

1 | INTRODUCTION

The Maritime Continent (MC) is the name given to the archipelago situated in the Indo-Pacific warm pool between Southeast Asia and Australia. It features islands of a range of sizes, many with complex topography, interspersed among shallow shelf seas which are some of the Earth's warmest (e.g. Zhang et al., 2016; Roxy et al., 2019; Yin et al., 2022). Mean humidity and precipitation rates across the MC are also among the highest observed globally (e.g. Asyaktur et al., 2016). The mean diurnal cycle (DC) features extreme contrasts between maximum and minimum rates of precipitation, particularly over land and near-coastal marine regions, with near-coastal land typically experiencing maximum mean precipitation between the late afternoon and evening, near-coastal waters experiencing maximum mean precipitation shortly after midnight, and waters further offshore experiencing an increasingly late mean precipitation maximum up to late morning (e.g. Love et al., 2011; Sakaeda et al., 2017; Worku et al., 2019). The DC of convection, to which the DC of precipitation is coupled, is dominantly driven by land–sea breeze dynamics which are generated by the differential in surface warming and cooling rates between land and water surfaces under the intense equatorial DC of solar radiative forcing (e.g. Mori et al., 2004; Kikuchi and Wang, 2008).

The DC is a high-frequency mode of tropical weather variability that is landlocked over the MC. Lower-frequency modes of tropical variability that propagate along the equator, including the eastward-propagating Madden-Julian Oscillation (MJO; Madden and Julian, 1971 & 1972; Zhang, 2005) and convectively-coupled equatorial Kelvin waves (CCKWs; e.g. Kiladis et al., 2009), modulate and are modulated by the DC as they propagate into the MC area (e.g. Oh et al., 2012; Peatman et al., 2014; Birch et al., 2016; Lu et al., 2019; Worku et al., 2020). The probability of a CCKW successfully propagating across the MC has been shown to be highly dependent on whether its arrival time at the islands of the MC is in phase with active diurnal convection (Baranowski et al., 2016). Furthermore, there is increasing support for the hypothesis that a strong DC over the MC is an integral component of the so-called “barrier effect” experienced by the MJO over the MC (e.g. Zhang and Ling, 2017; Ling et al., 2019; Zhou et al., 2021); evidence to the contrary is also still emerging (e.g. Seiki et al., 2021), however all studies agree that the strength of the DC influences the MJO in some manner. Since the convectively active envelope of an MJO event frequently contains CCKW components (Kikuchi et al., 2018), it follows that the fate of an MJO active envelope as it interacts with the DC is highly sensitive to both the amplitude and phase of the DC. This sensitivity has implications for global medium-range weather forecasting, as the inability of a model to generate an accurate DC will limit the degree to which the widespread medium-range predictability offered by MJO dynamics can be harnessed (Cassou, 2008; Skinner et al., 2022).

Although coarse-gridded general circulation models still show mixed skill concerning the generation of an accurate DC (Baranowski et al., 2019), high-resolution regional models have become increasingly capable of generating a DC that reflects the observed DC quite well (Bhatt et al., 2016). There is, therefore, an increased need to be able to characterise the complexities of the DC beyond the first-order pattern (i.e. an amplitude and a phase) in order to provide an accurate benchmark against which to assess model performance to facilitate further improvements. With the availability of over two decades of high-resolution gridded satellite precipitation data at the time of this study, there is potential to refine the characterisation of the DC in order to capture more of the spatio-temporal complexities that cannot be fully encapsulated with a two-parameter (amplitude and phase) characterisation.

The first diurnal harmonic (FDH) is a reasonable simplification of the DC in most locations across the MC as it represents the single 24-hour oscillation between a diurnal precipitation maximum and minimum that is generally observed, and has been used widely in the past (e.g. Yang and Slingo, 2001; Peatman et al., 2014; Birch et al., 2016; Seiki et al., 2021). Using a simple waveform to approximate the DC at a location has the two main benefits of noise suppression and characterisation of the full 24-hour composite cycle. Studies that use only the diurnal range (difference between maximum and minimum precipitation) and the time of maximum precipitation are not able to constrain the details of the 24-hour cycle very well except for the maximum, which can complicate dynamical interpretation of the local DC. Bai and Schumacher (2022) note that precipitation events tend to approximate more closely to brief instantaneously-initiating pulses than to sinusoidally-varying oscillations, particularly where the precipitation is associated with the triggering of deep convection. As such, a waveform which is more adaptable to sharp temporal gradients in precipitation rate may offer a more appropriate approximation. The primary purpose of this study is to demonstrate the application of two novel waveforms developed to be able to capture and intuitively characterise rapid transitions in precipitation rate.

The remainder of this article is structured as follows: Section 2 details the data used in this study. Section 3 describes the observed DC with the use of the diurnal range and time of maximum, and notes some complexities of the DC which are difficult to resolve with these characteristics alone. Section 4 describes the DC with the use of the FDH parameters and notes how and where this description is especially misleading. Section 5 introduces two novel waveforms trialled for a more accurate characterisation and gives an overview of the output of these new characterisation methods. Section 6 discusses the benefits and complications of the new characterisation methods, and describes how the new results might be interpreted and applied to other fields. Conclusions are presented in Section 7.

2 | DATA

The current state-of-the-art high-resolution satellite-derived precipitation data product, the Integrated Multi-satellite Retrievals for GPM (IMERG) V06, is used (Huffman et al., 2015). Having stitched together data from the Tropical Rainfall Measuring Mission (TRMM) satellite (active 1997 – 2015) and the Global Precipitation Measurement (GPM) satellite (active 2014 – present), and having reprocessed the TRMM data to match the higher spatial and temporal resolution of GPM, IMERG V06 offers over two decades of global precipitation data with a 30-minute temporal resolution and a 0.1° by 0.1° spatial resolution. Tan et al. (2019) demonstrated the improved suitability of IMERG V06 for analysis of the DC of precipitation relative to the previous TRMM-derived precipitation data product by calculating a lower median phase lag (0.5 hr) relative to ground-based observations over the southeast United States.

To focus on diurnal dynamics in the wet season, a composite DC is calculated using the 651 January days from 2001 to 2021. Time is stored in Universally Coordinated Time (UTC), and converted to local solar time (LST) as a linear function of longitude after all data extraction and best-fitting has been performed. The study area, spanning 8°N to 10°S and 94°E to 120°E , encapsulates the western side of the MC including the major islands of Sumatra, Borneo and Java, part of the southeast Asian peninsula, and a number of smaller islands (Figure 1a). The complementary land–sea mask at IMERG resolution is used to isolate land and marine regions, with marine regions arbitrarily defined as having over 80% water surface coverage.

3 | BASIC CHARACTERISTICS OF THE COMPOSITE DIURNAL CYCLE

In this study, the term 'basic characteristic' is applied to simple quantities describing the DC that are not associated with a waveform. Mean precipitation (represented as c ; Figure 1b) is the primary basic characteristic around which diurnal variability occurs. Mean precipitation has a mean value of 0.38 mm hr^{-1} across the entire study area, or 0.41 mm hr^{-1} over land, with the wettest locations experiencing a mean precipitation rate of around 1 mm hr^{-1} . Here, and in many other studies, two further basic characteristics are extracted to inform on the strength and timing of the diurnal precipitation maximum. For later ease of comparison with waveform amplitudes, the 'half range' (half of the difference between the diurnal maximum and minimum of precipitation) is reported, along with the 'peak time' (the time of day of maximum precipitation at the half-hourly resolution of IMERG).

The half range (Figure 2a) is greatest over land, particularly Java and the western sides of the other major land masses, with an overall mean half range over land (around 0.38 mm hr^{-1}) comparable to the magnitude of mean precipitation (Figure 1b), and slightly exceeding 1 mm hr^{-1} in the most highly diurnally active locations. Near-coastal waters also show moderately high diurnal variability, with the half range decaying offshore toward background values of around 0.15 mm hr^{-1} over the eastern Indian Ocean and less than 0.10 mm hr^{-1} over the South China Sea, some hundred kilometres from their closest highly diurnally active coastlines. The unusually high marine diurnal variability observed near the northwest coast of Borneo and over the far western Java Sea is associated with very high mean precipitation as opposed to proportionally stronger diurnal dynamics. The pattern of peak time (Figure 2b) shows that diurnal precipitation over near-coastal land peaks in the mid-late afternoon with a distinct inland gradient within about 100-200 km of the coastlines, inferring inland propagation of precipitation through the evening. Inland Borneo and parts of inland Sumatra have a less spatially coherent peak time, and in some regions a sharp discontinuity is observed between locations with a mid-afternoon peak time and locations with a post-midnight peak time. In many near-coastal waters, offshore propagation of peak precipitation is also evident from the progression of peak time from around midnight near the coasts toward late morning further offshore. The less spatially coherent peak time signal over the offshore Indian Ocean and South China Sea is indicative of the dominance of outliers, associated with strong individual non-diurnal precipitation events, over the weak open ocean DC.

These basic characteristics are consistent with expectations of a DC regime driven by the land-sea contrast. The discontinuities of peak time in inland areas point to one of the shortcomings of these basic characteristics, which is their inability to characterise more than one major diurnal peak in precipitation if present. Lowland regions to the east of mountain ranges over Sumatra and Borneo are known to experience a mean DC with both an afternoon peak and a night peak (e.g. Bai and Schumacher, 2022), hence a subtle change in the relative strength of these local maxima can affect which one counts as the peak time. This is a dramatic example of the difficulties associated with trying to understand the full 24-hour cycle of precipitation at a location using only basic characteristics, and these difficulties also apply where there is only one major maximum. It is not necessarily accurate to assume that the diurnal minimum occurs 12 hours before/after the maximum, or that the intensification and weakening occur steadily or in an oscillatory manner. By characterising the data with a suitable waveform, the complete 24-hour cycle can be approximately represented which removes the need to make assumptions about the details of the DC away from the maximum. The following section considers the FDH characterisation of the DC and to what extent the FDH is a suitable waveform for the characterisation of observed DC variability.

4 | FORM-FITTING WITH DIURNAL HARMONICS

4.1 | Form-fitting with the First Diurnal Harmonic

The FDH is the least-squares best-fit sinusoidal wave with 24-hour period fitted to the observed diurnal variability, where 'diurnal variability' is defined as the composite DC minus the mean precipitation (c ; Figure 1b) at any location. Two parameters describe the FDH; an amplitude (A) and a phase (ϕ). The FDH characterisation of the composite DC (diurnal variability plus the mean) therefore follows the equation:

$$f(t) = A \cos\left(\frac{2\pi}{T} [t - \phi]\right) + c, \quad (1)$$

where f is the rate of precipitation indicated by the waveform at time t in hours, and T is the diurnal period (24 hours).

In meteorological time series, it has typically been found that the amplitude of the FDH dominates over that of higher-order harmonics. For instance Yang and Slingo (2001) found that the amplitudes of the second and third diurnal harmonics of mean brightness temperature are generally of the order 40% and 15% respectively of the FDH amplitude. When higher-order harmonics are negligible, the FDH amplitude and phase are accurate indicators of the magnitude and timing of the diurnal maximum and the FDH waveform can be used to accurately approximate the full 24-hour DC. In such cases, the FDH amplitude should closely resemble the half range (Figure 2a) while the FDH phase should closely resemble the peak time (Figure 2b).

To quantify the general accuracy of the FDH waveform approximation (and later waveform approximations), the proportion of diurnal variability accounted for by the waveform is calculated. The standard deviation of the residual, where the residual is defined as the composite DC (d) minus the best-fit waveform (f) evaluated at each of the N (48) half-hourly time steps, quantifies the error in the waveform characterisation. This error is inherently proportional to the magnitude of diurnal variability, therefore a 'scaled error' (ϵ ; Figure 3a) is calculated in order to quantify the appropriateness of the waveform characterisation at each grid point in the IMERG dataset by normalising against the standard deviation of the composite DC:

$$\epsilon = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N (d_n - f_n)^2}}{\sqrt{\frac{1}{N} \sum_{n=1}^N (d_n - c)^2}}. \quad (2)$$

This scaling enforces a $0 \leq \epsilon \leq 1$ range, where $\epsilon = 0$ indicates perfect waveform fit and $\epsilon = 1$ indicates no waveform component (i.e. $A = 0$). In this study, waveforms are considered to provide a sufficiently accurate characterisation of the DC only where $\epsilon \leq 0.4$; this arbitrary threshold balances the need for an accurate waveform with the reality that the waveforms considered here are highly idealised. In order to minimise the risk of misinterpreting parameters, visualisation and interpretation of waveform parameters are restricted to locations where the waveform provides a sufficiently accurate characterisation.

Over land, scaled error varies from low over central Borneo (Figure 3c) to very high over eastern Sumatra (Figure 3e). Broad areas of land including Java and much of coastal Sumatra and Borneo have intermediate scaled error, suggesting a characterisation which is fair to first order but with a large residual at certain times of day, which could be due to relatively strong noise or due to a temporally coherent difference between the shape of the DC and the shape of the idealised waveform (e.g. Figure 3d). Over water, there is a clear tendency from low scaled error in coastal

areas toward very high scaled error far offshore as the noise becomes increasingly dominant (e.g. Figure 3f). The half-range-weighted mean scaled error – that is, the mean scaled error with each grid point weighted proportionally to its half range – across the entire study area is 0.43.

The FDH amplitude and phase are visualised in Figure 2c&d, with those regions which are not sufficiently accurately characterised by the waveform (i.e. $\epsilon > 0.4$) hatched over. Where not hatched, the FDH amplitude compares strongly to the half range (Figure 2a), as anticipated, while the phase resembles a smoothed-out peak time (Figure 2b). The spatial smoothness of the FDH phase emphasises the divide between the diurnal regimes of land and water and is less complicated by outliers than the peak time, however it also affects the representation of propagation speeds. Over Sumatra and Borneo, the FDH phase features a shorter difference in timing between coastal and inland precipitation maxima than the peak time (removal of coastal hatching would further accentuate this), such that inland propagation of precipitation based on the FDH phase appears quicker. A similar comparison is observed over the Java Sea; the near-midnight peak in precipitation around the northern coast of Java slowly propagates northward until midday in the peak time map, while the FDH phase changes less rapidly with distance northward, indicating more rapid northward propagation of the precipitation peak. The FDH phase and peak time are more clearly contrasted in Figure 4b, which visualises the lag of the FDH phase relative to the peak time (this lag is described as the 'phase error' to reflect waveform misrepresentation of the time of maximum precipitation). There is a transition from robustly positive phase error near the coasts (where diurnal precipitation initiates) toward mostly negative phase error further inland and offshore (where the peak time is later than the respective initiation area).

The FDH phase errors observed are of the order of multiple hours in many areas, even where the FDH provides a sufficiently accurate characterisation of the DC. The FDH phase is within one hour of the peak time over only about 35% of the total study area or about 37% of land area, increasing slightly to 43% of the total area where the FDH characterisation is sufficiently accurate according to $\epsilon \leq 0.4$. Over land, phase lag (red) exceeding one hour is nearly three times more prevalent than phase lead (blue) exceeding one hour; this systematic deviation in land-based precipitation phase risks offsetting the perceived timescale over which deep diurnal convection develops over land. Bai and Schumacher (2022) performed a similar analysis demonstrating large and spatially variable differences between FDH phase and peak time.

As quantified by the analogous amplitude error (Figure 4a), the FDH amplitude tends to under-represent the magnitude of diurnal activity across the study area, particularly over certain land areas such as Java, parts of near-coastal Sumatra and Borneo, and many of the smaller islands. Over Java, the mean amplitude error is -0.16 mm hr^{-1} while the mean half range is 0.64 mm hr^{-1} , meaning that the FDH amplitude captures only 75% of the magnitude of diurnal variability in this key region. Regions with high amplitude error correlate strongly with regions of high scaled error (Figure 3a), particularly over land; since the correlation between the magnitude of the phase error and the scaled error is less strong, it is suggested that an accurate amplitude parameter is intrinsically more critical for accurate waveform characterisation than an accurate phase parameter.

As used here, the FDH does not offer accurate insight into diurnal processes across most coastal land areas, far offshore nor over almost all of Java and peninsular Malaysia (Figure 2c&d). Far offshore, the half range is small, so it is expected that any true diurnal signal has a magnitude indiscernible against the background noise – this may be addressed by spatial averaging of the DC, however spatial fidelity is prioritised in the visualisations of this study. Conversely, the half range over Java (and, to a lesser extent, other coastal regions) is large, indicative of a strong diurnal signal which is worth characterising more accurately than the FDH can achieve.

4.2 | Consideration of Diurnal Cycles with Multiple Maxima

In order to characterise the DC variability in locations that the FDH fails to sufficiently accurately characterise, a more complex waveform with more than two degrees of freedom is needed. One might consider adding to the FDH the second diurnal harmonic, a sinusoidal wave with 12 h period, which would introduce two additional best-fit parameters: the amplitude and the phase of the second diurnal harmonic. This approach reduces the half-range-weighted scaled error of the study area by around one third, from 0.43 to 0.29. Of course, yet higher-order diurnal harmonics may also be added to ultimately completely characterise the DC. While incorporation of the second diurnal harmonic is moderately effective at improving the accuracy of the waveform characterisation, it is difficult to make use of the characterisation intuitively because the shape of the combined waveform may be complex and the additive nature of harmonics makes it difficult to interpret any single parameter in isolation.

The intuitively ideal waveform would retain just one amplitude parameter and phase parameter, to facilitate direct comparisons with the half range and peak time, and incorporate one additional degree of freedom that can capture as much additional diurnal variability as possible by morphing the shape of the waveform in an intuitive manner. A waveform with a single amplitude-phase parameter pair (a 'single-peak waveform') is, however, inherently only suited to characterisation of DCs with one dominant maximum-minimum pair. This excludes parts of eastern Sumatra and Borneo that are known to experience dual diurnal precipitation maxima. However, the regions with the highest magnitude of diurnal variability are generally found to have one dominant maximum, therefore a single-peak waveform may have high characterisation potential in key areas. To demonstrate the spatial distribution of the degree of dominance of the primary maximum-minimum pair, this study utilises a simple metric referred to as the 'wave index' (W ; Figure 3b) which quantifies the relative size of perturbations attributed to additional maxima and minima in the DC. The wave index is defined as the ratio between the sum of the absolute differences between each pair of consecutive time steps (or, equivalently, each neighbouring maximum-minimum pair) and two times the DC range (or four times the half range):

$$W = \frac{\sum_{n=1}^N |d_{n+1} - d_n|}{2(d_{max} - d_{min})}. \quad (3)$$

W takes a minimum value of one when there is only one local maximum and minimum in the composite DC, and a higher value when additional maxima and minima are present. The wave index shows that most land features one dominant diurnal maximum-minimum pair, with the main exceptions being the eastern parts of Sumatra, Borneo and peninsular Malaysia, as anticipated. Near-coastal marine locations typically show a reasonably low wave index, tending toward a more complex multi-peak DC further offshore. Unlike over eastern land areas, the multi-peak behaviour offshore is mostly attributed to noise; this can be demonstrated by calculating the wave index after spatially averaging the DC which reduces the offshore wave index as the noise is damped, while the wave index over eastern land areas is not strongly affected (not shown). Locations with the highest half range feature one dominant maximum-minimum pair, and those locations with a high wave index tend to have a low half range. Across the study area, over 65% of grid points with half range over 0.2 mm hr^{-1} , and over 86% of grid points with half range over 0.3 mm hr^{-1} , have $W \leq 1.5$.

The half-range-weighted mean wave index across the entire study area is 1.54. This value shows that the combined magnitude of additional diurnal fluctuations is, on average, about 54% of the magnitude of the dominant diurnal oscillation. However, this weighted mean value decreases to 1.29 (or 1.25) when the contribution of offshore noise is reduced by performing the same calculation after spatially averaging the DC over 0.3° by 0.3° (or 0.5° by 0.5°) bins.

Based on the half-range-weighted mean wave index of 1.25 which better represents the relative strength of additional diurnal fluctuations, it follows that additional diurnal fluctuation are generally much weaker than the dominant diurnal oscillation ($\approx 25\%$). Despite being limited to characterisation of the dominant diurnal oscillation, it seems likely that a single-peak waveform may be able to characterise the key features of the DC well. The half-range-weighted mean scaled error across the study area that would be achieved by the optimal single-peak waveform (i.e. a waveform that characterises the dominant diurnal oscillation with absolute accuracy) can be estimated by considering an idealised cycle of diurnal variability with $W = 1.25$; two sinusoidal oscillations each with 12-hour period, where one oscillation has an amplitude four times greater than the other. The scaled error of this idealised DC can be calculated and is approximately 0.24. While this is a crude estimate, it is much lower than the half-range-weighted mean scaled error calculated above for the FDH (0.43), which suggests that there is significant potential to improve on the characterisation of the DC provided by the FDH without needing to consider the complication of characterising multiple diurnal maximum-minimum pairs with multiple pairs of amplitude and phase parameters. The challenge is to achieve this while retaining the independent intuitive value of each variable parameter, meaning the degrees of freedom added to the FDH should be minimised. The following section demonstrates DC characterisation using two single-peak waveforms that generalise the oscillatory nature of the FDH with one extra degree of freedom each, and explores the impact of each waveform on amplitude and phase representation and diurnal variability characterisation.

5 | FORM-FITTING WITH OTHER SINGLE-PEAK WAVEFORMS

In this section, two novel single-peak waveforms – a skew-permitting waveform and a spike-permitting waveform – are discussed, with description of the waveform properties and the best-fit characterisation procedure followed by a description of the resultant best-fit parameters and their comparisons with the FDH parameters. A reference to simplified versions of the programming code used to calculate these best-fit waveforms (in Python programming language) is supplied at the end of Section 7.

5.1 | Form-fitting with a Skew-permitting Waveform

5.1.1 | Form-fitting procedure

The skew-permitting waveform is designed to accommodate cases where the DC maximum and minimum are not half a cycle apart, such that the increase in diurnal precipitation occurs rapidly and the decrease is drawn out across more of the day, or vice versa. The skew-permitting waveform has the form:

$$f(t) = \frac{A}{\tan^{-1}\left(\frac{\alpha_0}{\sqrt{1-\alpha_0^2}}\right)} \tan^{-1}\left\{\frac{\alpha_0 \sin\left[\frac{2\pi}{T}(t-\phi) + \cos^{-1}(\alpha_0)\right]}{1 - \alpha_0 \cos\left[\frac{2\pi}{T}(t-\phi) + \cos^{-1}(\alpha_0)\right]}\right\} + c, \quad (4)$$

where α_0 is the skew parameter and can vary between -1 and 1 (except for 0 as discussed below). For all values of α_0 , as for the FDH, A represents half the difference between the waveform maximum and minimum and ϕ represents the time of the waveform maximum; these are treated as the fixed definitions of waveform amplitude and phase in this study. A linearisation calculation (explained in the following paragraph) is applied to convert from the raw skew parameter (α_0) to a rescaled skew parameter, $-1 \leq \alpha \leq 1$, which is equivalent to α_0 for the values of -1 , 0 and 1 . This cumbersome equation and process result in a rather intuitive waveform (Figure 5a). The waveform shape

converges on the shape of the FDH at low $|\alpha|$, hence the simplification to Equation (1) is made when $\alpha = 0$ to render it a solvable value in the best-fitting procedure. Equation (4) varies smoothly toward a saw-tooth function as $|\alpha|$ increases. Positive α corresponds to a quicker increase in precipitation, such that the extreme positive case of $\alpha = 1$ features an instantaneous transition from minimum to maximum precipitation followed by constant decrease throughout the cycle. Conversely, extreme negative α gives a constant increase throughout the day followed by an instantaneous drop from the maximum to the minimum. Note that the concept of skew described here is unrelated to the statistical definition of the skew of a distribution, however the direction of the tail away from the maximum does qualitatively mirror the statistical definition.

The skew parameter is linearised with respect to the mean time lag of precipitation following the waveform minimum. This is achieved by calculating the timing of the 'waveform centroid' between two consecutive waveform minima – the centre of the area bound between the waveform and the straight line connecting the consecutive waveform minima ($y = -1$ in Figure 5a) – for any given value of α_0 . When $\alpha_0 = 0$, the time lag of the waveform centroid after the minimum (t_{lag}) is exactly 12 hours, while t_{lag} may be as low as 8 hours for positive skew or as high as 16 hours for negative skew. By defining $\alpha = -(t_{lag} - 12)/4$, the magnitude of the rescaled skew parameter becomes a good indicator of how significantly the shape of the DC deviates from that of the FDH.

Unlike the FDH, the best-fit skew-permitting waveform cannot be extracted using a Fourier technique, therefore least-squares optimisation is applied. The mean (c) is subtracted to leave the diurnal variability, and the FDH amplitude and phase and $\alpha_0 \approx 0$ are applied as the initial guess. In almost all cases for the DCs in this study, the optimisation successfully converged on a best-fit solution, with only a handful of noisy offshore locations failing to converge (in which case the FDH was taken as the best fit).

5.1.2 | Results

The amplitude and phase of the best-fit skew-permitting waveform are shown in Figure 2e&f, while the associated rescaled skew parameter is shown in Figure 5b. As for the FDH, the scaled error was calculated for the best-fit skew-permitting waveform (Figure 6a), and only sufficiently accurately characterised locations with $\epsilon \leq 0.4$ are shown. The amplitude parameter is very similar to the FDH amplitude (Figure 2c), with a subtle increase in mean amplitude of 0.011 mm hr^{-1} relative to the FDH (approximately 5% increase) with the increase distributed reasonably evenly across the study area. The phase parameter distribution bears a closer resemblance to the peak time than the FDH phase parameter distribution does, with the breadth of timings across Borneo and the Java Sea well represented, though the waveform best-fitting still has a subtle smoothing effect. Systematic regional phase errors are not as evident as they are for the FDH phase (Figure 4b&d). Phase is within one hour of the peak time over 54% of the entire study area, or 58% of the region sufficiently accurately characterised by the best-fit skew-permitting waveform. While these are much greater proportions than for the FDH (35% and 43%, respectively), the best-fit skew-permitting waveform phase parameter does not precisely match the peak time in all cases. This suggests that this waveform is not over-fitting to the noise-associated outliers that influence the peak time measurement of the precipitation maximum, therefore the best-fit skew-permitting waveform phase parameter may offer a more robust quantification of the timing of maximum diurnal precipitation.

The rescaled skew parameter distribution (Figure 5b) shows that moderately positive skew values prevail over western Sumatra and near-coastal Borneo, with some negative values over eastern central Borneo. Over marine areas, positive skew also prevails within some hundred kilometres of coastlines (most strongly to the southwest of Sumatra and north of Java) and there is a tendency toward negative skew further offshore. Separately for both the land-based and marine diurnal regimes, skew tends to decrease as the peak time becomes later. Dynamical inferences

from the best-fit skew-permitting waveform parameter distributions are discussed in Section 6.2.

The half-range-weighted mean scaled error for the best-fit skew-permitting waveform is 0.37, which is an improvement of 0.07 on the FDH and about one-third closer to the estimated optimal single-peak waveform characterisation. The spatial distribution of characterisation improvement (Figure 6b) shows that the key regions that benefit most from a skew-permitting characterisation are parts of near-coastal Borneo and Sumatra and the seas around the Mentawi islands to the west of Sumatra (all of which have strongly positive skew). Some regions however, notably including Java, do not benefit substantially from a skew-permitting characterisation and remain insufficiently accurately characterised, meaning that the best-fit parameters over such regions are still liable to be misleading.

5.2 | Form-fitting with a Spike-permitting Waveform

5.2.1 | Form-fitting procedure

The spike-permitting waveform accommodates cases where the mean local precipitation is not the average value of the DC maximum and minimum. With this waveform, composite precipitation may be relatively constant for a long stretch of the DC, with a short duration where precipitation significantly increases or decreases (Figure 5c). The following equation is the basis for the spike-permitting waveform:

$$f(t) = A \cos \left(\frac{\pi \tan^{-1} \left\{ \left[\frac{2 \operatorname{mod}_T \left(t - \phi - \frac{T}{2} \right)}{T} - 1 \right] \tan(\beta_0) \right\}}{\beta_0} \right) + A g(\beta_0) + c, \quad (5)$$

where β_0 is the spike parameter and varies between 0 and 0.5π , mod_T is the remainder operator with base T , and $g(\beta_0)$ is a translation function, varying between -1 and 1 . When integrated across the cycle, the first term on the right side of the equation does not have a mean value of zero, so the second term translates the waveform to restore a time-integrated mean value of zero before the third term (c) is added. This ensures that the parameter c consistently represents the mean value. As with the FDH and the skew-permitting waveform, A denotes half the difference between the waveform maximum and minimum while ϕ corresponds to the time of the waveform maximum.

Equation (5) is undefined when the new parameter $\beta_0 = 0$. However, the function is designed to converge on the FDH as β_0 tends toward 0. As β_0 becomes more positive, the waveform peak becomes increasingly narrow, until at $\beta_0 = 0.5\pi$ the DC features a constant value except for an instantaneous intense peak (akin to a Dirac delta function). In order to achieve the opposite effect where the minimum becomes increasingly narrow, the waveform is inverted (i.e. A is replaced with $-A$ and ϕ with $\phi + 0.5T$) – the best-fitting procedure is therefore run twice, with the original initial guess and the inverted initial guess, and the converged solution with the lower least-squares residual is accepted. In cases where the inverted waveform provides the optimal fit, the true-to-definition values of A and ϕ are reported, and β_0 is reported as negative. This means that, as reported, β_0 may range between -0.5π and 0.5π .

In order to make the spike parameter a good indicator of how significantly the waveform shape deviates from that of the FDH, a rescaled spike parameter, β , is calculated by linearising with respect to the relative closeness of the waveform mean to the waveform maximum and minimum, and normalising onto a scale from -1 to 1 . Values of -1 , 0 and 1 for β correspond to values of -0.5π , 0 and 0.5π for β_0 , respectively. In practice, the definition of the rescaled spike parameter makes it equivalent to the translation function (i.e. $\beta = g(\beta_0)$). Due to the complexity of the function $g(\beta_0)$, it is treated as an independent variable in the form-fitting optimisation process alongside A , ϕ and β_0 (having subtracted c). This adds an extra degree of freedom in the least-squares optimisation procedure, and

therefore it is possible that the value of $g(\beta_0)$ that the optimisation produces will not precisely equal the value that it should equal based on its dependence on β_0 . The risk of high imprecision increases as the resolution of the cyclic data decreases, and therefore for the high-resolution temporal data used in this study, it is found that the output value of the independent variable $g(\beta_0)$ reliably matches the expected function value for the output β_0 value to very close approximation. The waveform that fits the diurnal anomaly therefore has a mean value very close to zero, as intended.

5.2.2 | Results

The amplitude and phase of the best-fit spike-permitting waveform are shown in Figure 2g&h, while the associated rescaled spike parameter is shown in Figure 5d. As for the other waveforms, the scaled error was calculated for the best-fit spike-permitting waveform (Figure 6c), and the best-fit parameters are only shown in regions with $\epsilon \leq 0.4$. Of the three waveforms discussed in this paper, the amplitude of the best-fit spike-permitting waveform offers the closest match to the half range. The domain-averaged amplitude is larger than that of either the FDH or best-fit skew-permitting waveform (a 0.019 mm hr^{-1} or 9% increase on the FDH). Regions with the largest FDH amplitude error (Figure 4a) are responsible for a high proportion of this increase, such that these regions do not stand out as having a highly under-representative amplitude (Figure 4e). Compared with the FDH, there are nearly two-thirds fewer grid points where the amplitude error is more negative than -0.1 mm hr^{-1} , resulting in a more uniform background amplitude error which accounts for about 14% of diurnal variability as quantified by the half range.

The best-fit spike-permitting waveform phase parameter (Figure 2h) is qualitatively similar to the FDH phase (Figure 2d). As such, use of the spike-permitting waveform results in a systematic phase lag over near-coastal land, much like the FDH (Figure 4b&f), though the magnitude of the lag is lower. 51% of the well-characterised area has a phase value within one hour of the peak time, compared with 43% and 58% for the FDH and best-fit skew-permitting waveform, respectively. Across both land and water, the best-fit spike-permitting waveform phase is more uniform than either the best-fit skew-permitting waveform or the peak time distribution (Figure 2b&f), therefore the apparent propagation speed of maximum precipitation away from coastlines is too fast, as discussed for the FDH.

The spike parameter distribution (Figure 5d) shows a prevalence of positive spike behaviour over most land areas, particularly over Java, western and central Sumatra, western peninsular Malaysia, coastal Borneo and some of the minor islands. The most positive values align with those areas that have the highest FDH amplitude error (Figure 4a). Eastern inland Sumatra and most of inland Borneo show slightly negative spike. Over most near-coastal marine areas, spike tends to be weakly to moderately positive, with no clear gradient with distance offshore. The notable exceptions to this are off the northern and southern coastlines of Java, where the spike is slightly negative.

As a result of significant characterisation improvements over positive spike land areas such as Java (Figure 6d), the half-range-weighted mean scaled error for the best-fit spike-permitting waveform is 0.33, compared with 0.43 and 0.37 for the FDH and best-fit skew-permitting waveform, respectively. Based on the scaled error metric and estimated value of 0.24 for the optimal single-peak waveform, the spike-permitting waveform is approximately 50% closer to the estimated optimal single-peak waveform characterisation than the FDH. Nearly all areas with high half range meet the $\epsilon \leq 0.4$ criterion (Figure 6c) and are therefore sufficiently accurately characterised by this new waveform.

6 | DISCUSSION AND APPLICATIONS

6.1 | Waveform Strengths and Deficiencies in Representation of the Diurnal Cycle

Best-fit waveforms have two key benefits; they are not highly sensitive to noise or outliers in the way that the basic characteristics of half range and peak time may be, and they offer a means to approximate the continuous 24-hour cycle. However, if the waveform cannot achieve a shape that closely follows the DC, the output parameters cannot be accurately interpreted in their usual manner (i.e. amplitude as an indicator of DC magnitude and phase as an indicator of the time of the maximum). As these results show, different waveforms have different susceptibilities regarding the intuitive reliability of their parameters.

The FDH is susceptible to misrepresenting both the DC magnitude and the time of maximum, particularly over land. The shape of the FDH cannot capture the brief intense peak in precipitation observed in areas like Java, nor can it capture the asymmetry of the rapid intensification and gradual weakening of precipitation over near-coastal areas (Bai and Schumacher, 2022). The suitability of the FDH characterisation is highly regionally variable; it characterises the DC very well over most of inland Borneo and many near-coastal marine areas, and is entirely unsuitable over eastern land areas.

The best-fit skew-permitting waveform is also susceptible to misrepresenting the DC magnitude, since the added degree of freedom does not allow for a much briefer peak. The phase parameter, however, is a reliable indicator of the time of day at which the most intense precipitation is focused, as shown by its regional accuracy in comparison with the peak time (Figure 4d). Over land, the mean phase error is very small (-0.11 hours), indicating an accurate representation of the peak time to within ten minutes, while the other two waveforms have a mean phase lag over land of more than 30 minutes (and greater if considering only near-coastal land). The standard deviation of best-fit skew permitting waveform phase error is lower than for the other waveforms (1.78 hours over the entire study area, in comparison with 2.05 and 2.11 hours for the FDH and best-fit spike-permitting waveform, respectively). The lower standard deviation is consistent with the best-fit skew-permitting waveform phase's closer comparison to the peak time, however it is still imprecise enough to suggest that the waveform does not over-fit to noise-related outliers.

The best-fit spike-permitting waveform produces a more accurate phase parameter than the FDH, however it still features a clear and significant phase lag over coastal land and a tendency toward phase lead or less phase lag away from coastlines, both inland and offshore. However, the waveform shape allows for accurate characterisation of the brief but extreme peak in precipitation over land areas including Java (Rais and Yunita, 2017), which means that the best-fit spike-permitting waveform amplitude is a much more reliable indicator of the magnitude of diurnal variability than the other waveform magnitudes. The reliability of amplitude representation leads to sufficiently accurate DC characterisation ($\epsilon \leq 0.4$) in the vast majority of cases with a single dominant peak (Figure 7). The best-fit spike-permitting waveform sufficiently accurately characterises 92.5% of observed DCs with $W < 1.5$ (Figure 7d), in comparison with 68.1% and 83.0% sufficiently accurate characterisation of low wave index areas by the FDH and best-fit skew-permitting waveform, respectively (Figure 7b-c). For the best-fit spike-permitting waveform, the insufficiently-characterised low wave index locations (magenta) are mostly well-dispersed across the region, which suggests that many of these 7.5% insufficiently-characterised low wave index locations are fringe or borderline cases. One exception is northwest Borneo (around 2.5°N , 111.5°E) where an otherwise roughly steady cycle features a brief negative diurnal perturbation for a few hours before local midday followed by a slightly less brief positive diurnal perturbation after midday (resulting in the negative perturbation not being characterised).

Both the FDH and best-fit skew-permitting waveform show several coherent areas over land where the wave index is low and characterisation is considered insufficiently accurate (magenta), suggesting that these are not borderline

failures sensitive to the choice of sufficient ϵ . Instead, these characterisation failures show that there is considerable systematic inability of these waveforms to appropriately characterise certain observed forms of single-peak DC. Of the three waveforms, the spike-permitting waveform is demonstrably the most successful waveform at characterising diurnal variability in the presence of a single dominant maximum-minimum pair. Such locations account for 41.8% of the western MC region (Figure 7a), including two-thirds of its land area. It also demonstrates slightly superior characterisation success at moderate wave index values (lavender in Figure 7). However none of the three waveforms considered are able to accurately characterise DCs with complex multi-peak variability (grey in Figure 7). This confirms the expected unifying weakness of these single-peak waveforms and shows that a multi-peak waveform will be necessary in order to achieve accurate 24-hour characterisation of regions with a complex DC.

The FDH has the benefit of being a guaranteed least-squares best fit as extracted by Fourier analysis, while the skew-permitting and spike-permitting waveforms both require an appropriate initial guess and optimisation methodology in order to converge on the best fit. The very low rates of convergence failure (only over weakly-varying regions dominated by noise) and the high spatial coherence of the best-fit skew and spike parameters give confidence in the general success of the best-fit optimisation methodology. The skew and spike parameters both offer clear qualitative insight into the temporal details of the observed diurnal precipitation variability, which may in turn be used to understand the regional dynamics. Common diurnal mechanisms may occur in different areas with comparable skew and/or spike. The following subsection considers the features and consistency of individual precipitation events to investigate possible physical mechanisms that lead to positive and negative values of skew and spike.

6.2 | Interpretation of Skew and Spike Parameters

Analysis and description of the DC thus far in this study has considered the composite DC, which isolates the mean diurnal signal from the transient non-diurnal weather systems that are observed on any given day. There is, of course, much day-to-day variability in practice because the convection and precipitation driven by the onset of solar radiative forcing every morning is highly dependent on the initial state of the atmosphere. Indeed, there may be so much day-to-day variability that there is no individual day during which precipitation closely follows the composite cycle. It is difficult to assess the degree of day-to-day variability from the composite alone. For example, if the degree of day-to-day variability is low, a DC with positive skew may be interpreted as having a systematic rapid intensification of precipitation followed by a systematic gradual weakening, whereas if the day-to-day variability is high, a positive skew DC may instead be interpreted as having a consistent, systematic timing of precipitation onset but an inconsistent precipitation event duration. In order to assess the degree of day-to-day variability and constrain dynamical hypotheses for positive and negative values of skew and spike, an event-based analysis was performed in select locations to determine the frequency of precipitation event initiations, event maxima and event terminations (as observed at the location) at each time of day. The same 21 January months were used, spatially bin-averaged into 0.5° by 0.5° boxes with similar parameter values throughout the box. A precipitation threshold of 0.02 mm hr^{-1} was chosen to identify precipitation initiations and terminations. Only 'major precipitation events' – those that registered a maximum precipitation rate of at least 5 mm hr^{-1} – were counted. The low proportion of major precipitation events lasting more than 24 hours were omitted in calculations involving duration.

6.2.1 | Positive spike

Positive spike is perhaps the case with the most well-constrained interpretation. At extreme positive values, the composite shows a rapid increase is shortly followed by a rapid decrease, meaning that the diurnal component of

precipitation must arrive at almost exactly the same time of day on any day that it manifests and last for a very short time. Less extreme positive values allow for longer precipitation events and/or more day-to-day variability in the exact timing of initiation and termination, but on any given day any diurnally-driven precipitation is still highly constrained. The most positive values observed are in coastal land areas which are close to great lengths of coastline, including many small islands, the thin island of Java and narrow northern Sumatra (Figure 5d). Such locations are likely to experience intense low-level wind convergence in the afternoon due to the convergence of moisture-rich sea-breezes from multiple directions, which would induce a strong updraught and trigger rapid initiation and intensification of precipitation. The timing of initiation is also likely to be more consistent than other land areas since there is less of a divide in diurnal regime between onshore and offshore background winds (Peatman et al., 2021) – background winds from most directions are likely to be onshore with respect to some part of the nearby coastline. The event-based analysis over the part of eastern Java marked in Figure 1a supports a very consistent timing of initiation (Figure 9a), with nearly 95% of major precipitation events initiating between 0800 and 1600 LST. A comparatively short median major precipitation event duration of 11 hours is also found.

6.2.2 | Positive skew

There is significant overlap between areas of positive spike and positive skew, and both of these idealised DC shapes have the common feature of a rapid increase (Figure 5). Three hypotheses are considered that may result in the gradual decrease observed after the rapid increase observed in positive skew areas: 1) day-to-day variability is insignificant and every individual day features rapid intensification and gradual weakening of precipitation at consistent times (Figure 8a); 2) precipitation events are alike in duration but variable in timing, with the distributions of initiation time, maximum time and termination time all featuring (statistical) positive skew (Figure 8b), and; 3) the time of initiation is consistent day-to-day, but the event duration (and therefore the termination time) is highly variable (Figure 8c). The event-based analysis finds that the timing of the precipitation maximum of major precipitation events is, on average, closer to the initiation time than the termination time over positive skew areas, particularly over a land box in southern Borneo (Figure 1a) where mean post-maximum precipitation endures nearly three hours longer than mean pre-maximum precipitation (Figure 9b). This frequent asymmetry in individual major precipitation events over positive skew areas is consistent with hypothesis 1 above. One dynamical possibility with which this hypothesis is consistent is a systematic transition from intense but brief convective precipitation to less intense but potentially longer-lasting stratiform precipitation (e.g. Worku et al., 2019).

The distribution of event durations has a standard deviation of over four hours, which infers a role for day-to-day variability in controlling the duration of precipitation events in positive skew areas, which is a key feature of **hypothesis 3** (Figure 8c). The contrast of consistent initiation timing and inconsistent termination timing over the southern Borneo area (Figure 9b) is also most consistent with hypothesis 3. A dynamical mechanism that may achieve this contrast in initiation and termination consistency is advection of diurnally-initiated weather systems by background winds. The development of diurnally-driven deep convective cells is relatively independent of the background weather conditions (e.g. Da Silva and Matthews, 2021), being driven by the strong, temporally consistent and land-locked surface heating contrast. However, once the diurnal convective system has developed and mid-troposphere stratiform cloud has spread broadly around the convection centre, the land–sea surface heating contrast is inhibited and stops anchoring the convection centre. This leaves the precipitating convection centre susceptible to horizontal advection by strong background winds. Variability in the speed and direction of these background winds may then be critical in determining the local precipitation termination time in positive skew areas. In summary, over positive skew areas there is clear evidence supporting a consistent precipitation initiation mechanism, and evidence that both a systematic transition

to weaker precipitation and day-to-day variability contribute to the gradual composite weakening associated with a positive value of the best-fit skew parameter.

6.2.3 | Negative skew

Negative skew of the DC points to the presence of a consistent rapid decrease in precipitation rate following a more gradual increase. Such cases are found predominantly far from coastlines, in inland areas and over relatively distal waters (Figure 5a). Mirroring the three hypotheses for positive skew, three hypotheses are considered that explain how day-to-day variability in precipitation events leads to negative skew DCs: 1) day-to-day variability is insignificant and every individual day features gradual intensification and rapid suppression of precipitation at consistent times (Figure 8d); 2) precipitation events are alike in duration but variable in timing, with the distributions of initiation time, maximum time and termination time all featuring (statistical) negative skew (Figure 8e), and; 3) precipitation may initiate at a **broad** range of times of day, but most events terminate at a consistent time (Figure 8f). Hypothesis 1 for negative skew is ruled out, since the event-based analysis finds that the event maxima at different negative skew locations tend to average midway through the event or closer to the initiation (Figure 9c). Over the negative skew area of the Java Sea, the distribution of termination times shows a well-constrained modal frequency in the hours around sunset (not shown, but comparable to Figure 9d), however the termination times are more broadly spread over negative skew Borneo (Figure 9c). Both negative skew areas experience an inconsistent initiation of precipitation, therefore the contrast between initiation and termination consistency over the Java Sea shows consistency with hypothesis 3. Over negative skew Borneo, the distribution of precipitation maxima is more consistent than either the distribution of initiation or termination time, and the modal frequency of the precipitation maxima distribution shows a subtle statistical negative skew (Figure 9c). This suggests that negative skew over land results from day-to-day variability more consistent with hypothesis 2.

The commonality amongst areas of negative skew is a broadly-spread distribution of initiation times. While the inconsistency of event terminations over positive skew areas was suggested above to be the result of advection by variable background winds, it follows that the inconsistency of event initiations over nearby negative skew areas may be the result of variable arrival time of diurnal convective centres. Over the Java Sea, there may be a consistent precipitation suppression mechanism leading to event terminations more consistent than either event initiations or maxima – over the southeastern Java Sea most event terminations occur within four hours of 1600 LST (Figure 9d). One possible precipitation suppression mechanism is a gravity wave that may nucleate over land as deep convection develops. The closest land areas, Java and southern Borneo, both have a DC with a consistent initiation time and maximum time (Figure 9a&b), and the strong diurnal signal over Java in particular suggests that strong and consistently-timed gravity waves may be nucleated there. The modal frequency of event terminations over the Java Sea is a few hours later than the modal frequency of event initiations over Java – those few hours of lag are plausibly related to the propagation time of a gravity wave, which would typically propagate at a speed of the order $15\text{-}20\text{ m s}^{-1}$ (Mapes et al., 2003), plus Doppler shift by upper-tropospheric background winds. The mechanism leading to the statistical negative skew of the distribution of precipitation maxima over negative skew Borneo (Figure 9c) is unclear.

6.2.4 | Negative spike

Negative spike tends to occur in inland areas, and is also observed in the waters surrounding Java (Figure 5d). Extreme negative values of spike are not observed; these would diagnose the peculiar case where the dominant feature of the DC is a strong suppression of precipitation followed shortly by a strong intensification, occurring at a very consistent

time day-to-day. Three hypotheses for how variability in day-to-day precipitation may lead to the moderate negative spike values observed are as follows: 1) day-to-day variability is insignificant and every individual day features precipitation intensification followed much later by suppression, both at a consistent time (Figure 8g); 2) precipitation events are alike (and comparatively short) in duration but highly variable in timing (Figure 8h), and; 3) there is a period of regular event termination, and a period of regular event initiation shortly after, but precipitation events need not last the entire day in between (Figure 8i). As discussed above, decreasing spike parameter allows for increased event duration and/or less consistent timing, and hypotheses 1 and 2 divide these possibilities; a higher typical event duration than positive spike areas would be expected with hypothesis 1, while the event duration may be similar to positive spike areas if hypothesis 2 is more accurate. Hypothesis 3 represents the case where a precipitation suppression mechanism is the dominant diurnal driver.

The event-based analysis shows different distribution patterns for the two key negative spike areas of inland Borneo and the southeastern Java Sea. Inland Borneo exhibits a wide spread of initiation and termination times, and a slightly more constrained modal frequency in the distribution of time of precipitation maxima (Figure 9c), and the median event duration is lower in negative spike areas of Borneo than in positive spike areas. This is consistent with hypothesis 2 (Figure 8h), and suggests that high day-to-day variability is the dominant cause of negative spike. This is consistent with the earlier inference that the timing of diurnal precipitation at locations far from coastlines is highly influenced by variable background winds. In contrast, the southeastern Java Sea experiences rather consistent event initiations and terminations, but inconsistent event maxima, with high frequency of event maxima between midnight and 1600 LST (Figure 9d). There is no evidence that event duration is systematically longer over the negative spike part of the Java Sea than the positive spike part of the Java Sea, however a higher standard deviation of event duration points to more day-to-day variability over the negative spike part of the Java Sea (not shown). This is most consistent with hypothesis 3 (Figure 8i), which suggests that a brief and consistent precipitation suppression mechanism is a dominant diurnal driver only over water. The concentration of negative spike seas around Java (Figure 5d) supports the suggestion above that the dominant gravity wave originates from this island, and a similar pattern is observed around the small islands between Sumatra and Borneo. A consistent precipitation termination is observed across much of the Java Sea, even toward the north where the spike is positive, therefore it may be that the diurnal consistency of the suppressive influence of a gravity wave originating from Java does not reduce substantially with distance from Java. Instead, it appears that the northward transition to positive spike results from the increasingly constrained distribution of precipitation maxima.

6.3 | Applications of New Waveforms

In this study, the skew-permitting waveform (Equation (4)) and the spike-permitting waveform (Equation (5)) have been applied to satellite-derived tropical precipitation to capture and characterise a range of diurnal behaviour that cannot be accurately described using the FDH. Future work on the DC generated by models will benefit from the skew-permitting and spike-permitting characterisations, which will provide more representative amplitude and phase parameters and the additional characterisation metrics of skew and spike. Having begun to establish the likely dynamical associations of positive and negative skew and spike, the skew and spike parameters may be used to compare and contrast the dynamics of models and observations.

The observed tropical DC of precipitation is one of many examples across the sciences of a complex but dominantly single-peak cycle. Other phenomena that these waveforms may be well suited to characterise include the annual cycle of sea surface temperature (e.g. Wang et al., 2022), the surface elevation signal associated with shallow water tidal constituents (e.g. Zhang et al., 2018) and the vertical isopycnal displacement signal associated with

internal solitary waves (e.g. Helfrich and Melville, 2006). For the latter examples, either time or distance may be used as the independent variable. Another example of a cyclic quantity for which these waveforms may be well suited is the phase of the MJO (e.g. Peatman et al., 2014; Bai and Schumacher, 2022). There are likely to be many further characterisation applications beyond the research experience of the authors.

The intuitive range of behaviours that these waveforms provide may also prove useful in idealised modelling studies if applied as parameter drivers. For instance, in an atmosphere-only climate model, the DC of sea surface temperature may be idealised to follow the shape of the spike-permitting waveform, and the implications of variable warm layer longevity may be investigated by varying the spike parameter of the driving waveform. Alternatively, a hydrological model might idealise the DC of precipitation to follow the shape of the skew-permitting waveform, and the sensitivity of land drainage to the rate of precipitation intensification may be investigated by varying the skew parameter of the driving waveform.

7 | CONCLUSIONS

In this study, high-resolution satellite-derived GPM-IMERG data were used to analyse the observed temporal and spatial complexities of the DC of precipitation over the MC, and two new waveform approximations were applied to enable intuitive characterisation of these complexities. In previous studies, the FDH has typically been the waveform of choice, but the DC cannot be accurately approximated by a sinusoidal wave in many places. Therefore, characterisation with this simple waveform misrepresents the magnitude of diurnal variability and the time of maximum precipitation and hinders accurate interpretation of the underlying dynamics. By restricting use of the best-fit waveform to areas where a sufficiently accurate characterisation is offered, accurate interpretation is more assured, but the FDH fails to offer an accurate characterisation over many areas with a strong DC such as Java. The two new waveforms developed and applied in this study, the skew-permitting and spike-permitting waveforms which each add one degree of freedom to the shape of the FDH, provide a more complete accurate characterisation of the DC across the MC.

By comparing the phase of the FDH with the time of the peak in the (non-best-fitted) composite DC, it was found that the phase tends to lag behind the peak time by an average of around an hour (and frequently up to three hours) over most land areas. The skew-permitting waveform was shown to be able to accurately capture the rapid intensification and gradual weakening of diurnal precipitation over land and produce a phase parameter which is more accurate with respect to the observed peak time. The amplitude parameter of the FDH was shown to significantly under-represent the true magnitude of diurnal variability observed over land areas including Java, parts of western Sumatra and many minor islands, and it was found that the spike-permitting waveform is able to capture the full magnitude of the brief but strong diurnal peak in precipitation in these areas. On average across the study area, the spike-permitting waveform was shown to provide the greatest improvement in diurnal variability characterisation, with waveform characterisation considered sufficiently accurate over more than 90% of the key subset of the study area with one dominant diurnal maximum-minimum pair. The DC over the eastern lowland areas of Sumatra, Borneo and peninsular Malaysia features two distinct diurnal maximum-minimum pairs, therefore these areas require a more complex multi-peak characterisation beyond the scope of this study, however these areas tend to experience a weaker DC than single-peak areas.

The more representative values of the phase parameter of the skew-permitting waveform and the amplitude parameter of the spike-permitting waveform will help to ensure unbiased assessment of model performance at DC generation. The newly developed skew parameter and spike parameter distributions can be used to constrain the

diurnal dynamical mechanisms at play locally. In the observational data considered here it was found that positive spike and positive skew are associated with precisely-timed rapid intensification of diurnal convection and precipitation, while negative spike and negative skew are associated with high day-to-day variability in the DC over land. Over the waters surrounding Java, negative spike may indicate a strong, precisely-timed precipitation suppression mechanism, which is argued to be the result of afternoon gravity waves nucleated over Java.

Future work will use this characterisation methodology to assess model performance, and investigate whether the same mechanisms are responsible for positive and negative skew and spike DCs. It is hoped that the new waveforms developed, investigated and made available here will be applied more broadly across the sciences in scenarios where qualitatively intuitive characterisation of a single-peak cycle is desired. Programming code (in Python programming language) to enable form-fitting of cyclic data with the skew-permitting waveform and the spike-permitting waveform is publicly available at https://github.com/jm-mustafa/Single_Peak_Waveforms.

ACKNOWLEDGEMENTS

The research presented in this paper was carried out on the High Performance Computing Cluster supported by the Research Computing Service at the University of East Anglia. JMM and MVCA were funded and AJM, RAH and KJH were partially funded by the Natural Environment Research Council through the TerraMaris project (grant NE/R016704/1).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

references

- As-syakur, A., Osawa, T., Miura, F., Nuarsa, I., Ekayanti, N., Dharma, I., Adnyana, I., Arthana, I. and Tanaka, T. (2016) Maritime continent rainfall variability during the trmm era: The role of monsoon, topography and el niño modoki. *Dyn. Atm. Oceans*, **75**, 58–77.
- Bai, H. and Schumacher, C. (2022) Topographic influences on diurnally driven mjo rainfall over the maritime continent. *J. Geophys. Res.: Atm.*, **127**.
- Baranowski, D., Flatau, M., Flatau, P. and Matthews, A. (2016) Phase locking between atmospheric convectively coupled equatorial kelvin waves and the diurnal cycle of precipitation over the maritime continent. *Geophys. Res. Lett.*, **43**, 8269–8276.
- Baranowski, D., Waliser, D., Jiang, X., Ridout, J. and Flatau, M. (2019) Contemporary gcm fidelity in representing the diurnal cycle of precipitation over the maritime continent. *J. Geophys. Res.*, **124**, 747–769.
- Bhatt, B., Sobolowski, S. and Higuchi, A. (2016) Simulation of diurnal rainfall variability over the maritime continent with a high-resolution regional climate model. *J. Met. Soc. Japan*, **94A**, 89–103.
- Birch, C., Webster, S., Peatman, S., Parker, D.J. Matthews, A., Li, Y. and Hassim, M. (2016) Scale interactions between the mjo and the western maritime continent. *J. Clim.*, **29**, 2471–2492.
- Cassou, C. (2008) Intraseasonal interaction between the madden-julian oscillation and the north atlantic oscillation. *Nature*, **455**, 523–527.

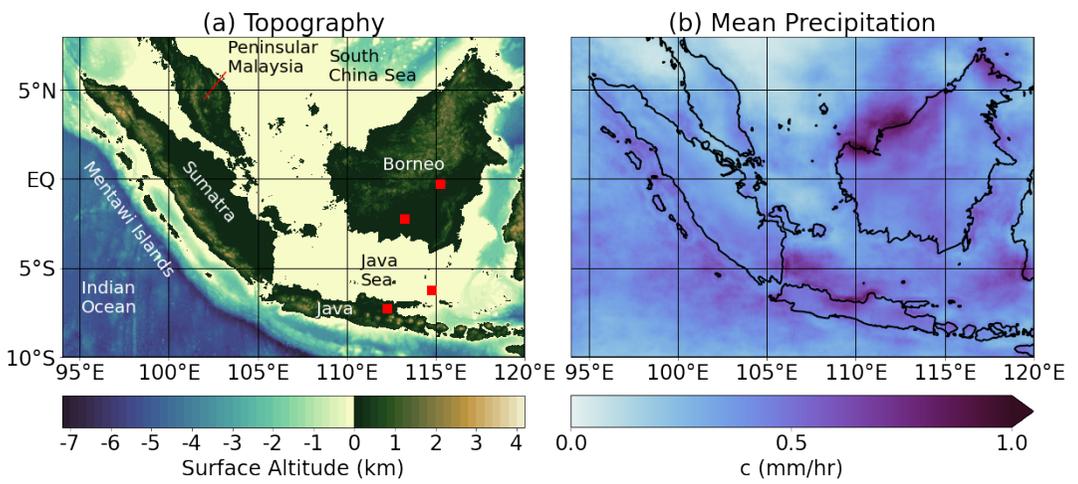


FIGURE 1 (a) Regional map of topography and bathymetry, with major land masses and water bodies labelled. The red squares mark four of the boxes across which the event-based analysis, described in Section 6.2, is performed. (b) Climatological distribution of January precipitation, based on GPM-IMERG data from 2001–2021.

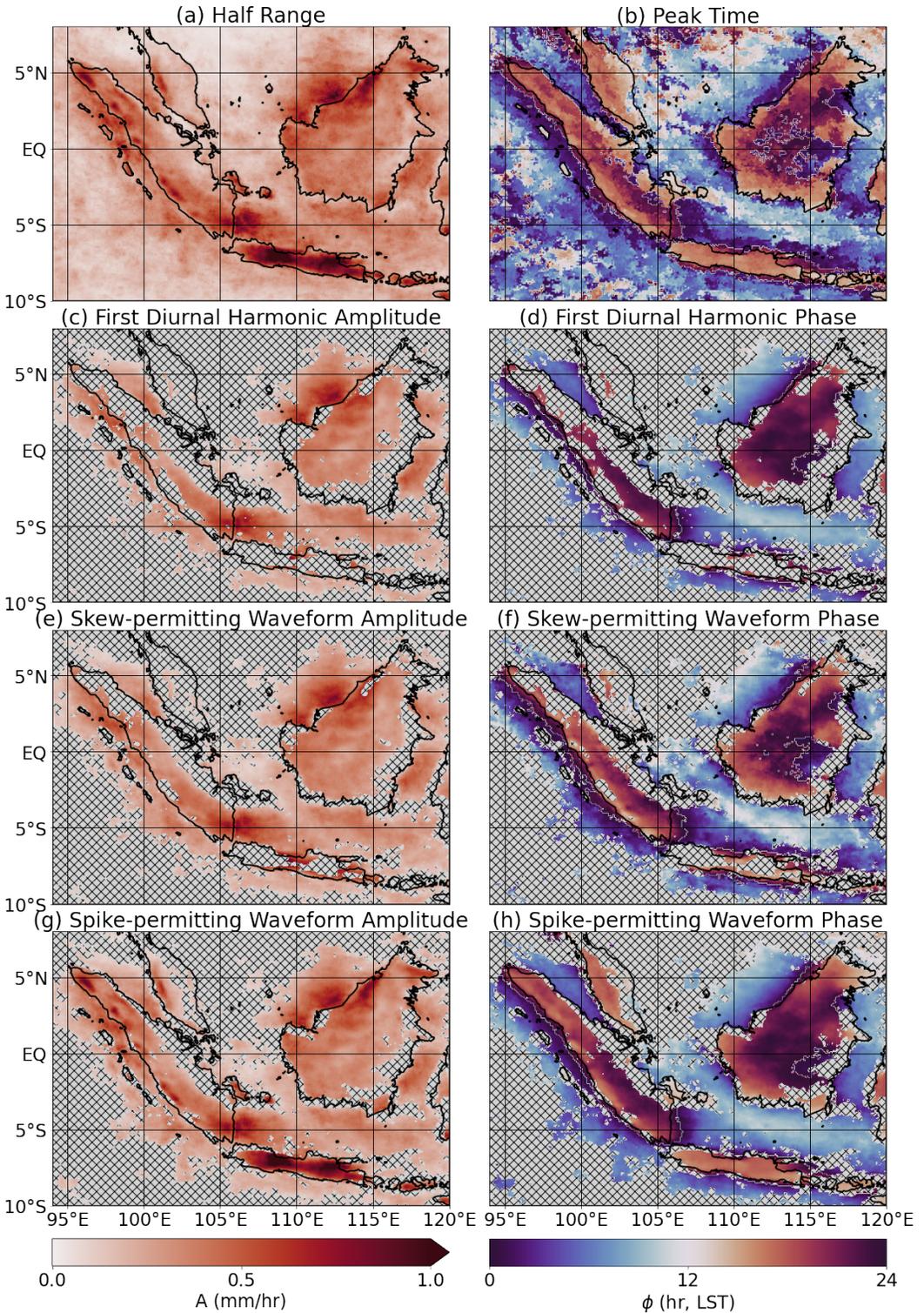


FIGURE 2 Indicators of the magnitude (left panels) and timing (right panels) of diurnal precipitation variability. (a) Half the difference between the maximum and minimum precipitation. (b) Time of maximum precipitation. (c–h) Amplitude and phase of the: (c&d) first diurnal harmonic; (e&f) best-fit skew-permitting waveform, and; (g&h) best-fit spike-permitting waveform. Hatching is applied where $\epsilon > 0.4$ for the relevant waveform (see Figures 3a and 6a&c).

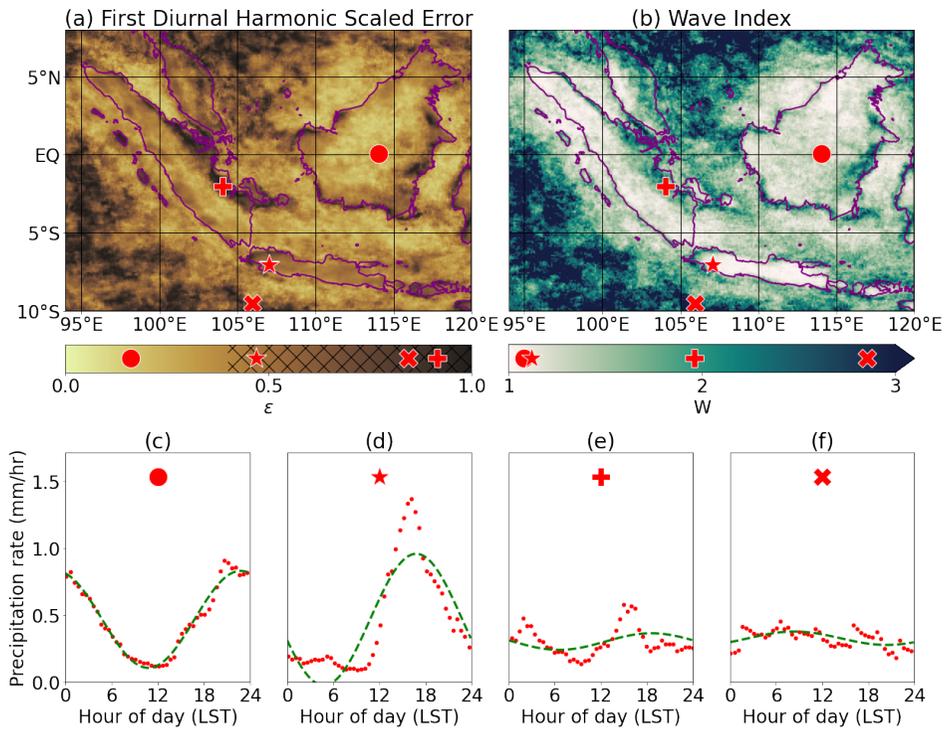


FIGURE 3 Indicators and examples of presence and character of diurnal activity. (a) Scaled error (ϵ) map for the first diurnal harmonic waveform. Colour bar hatching corresponds to the map hatching applied in Figure 2c&d. (b) Wave index map, quantifying the complexity of the diurnal cycle as defined in text. (c–f) Examples of the composite diurnal cycle of precipitation at point locations, with the first diurnal harmonic as the dashed green line.

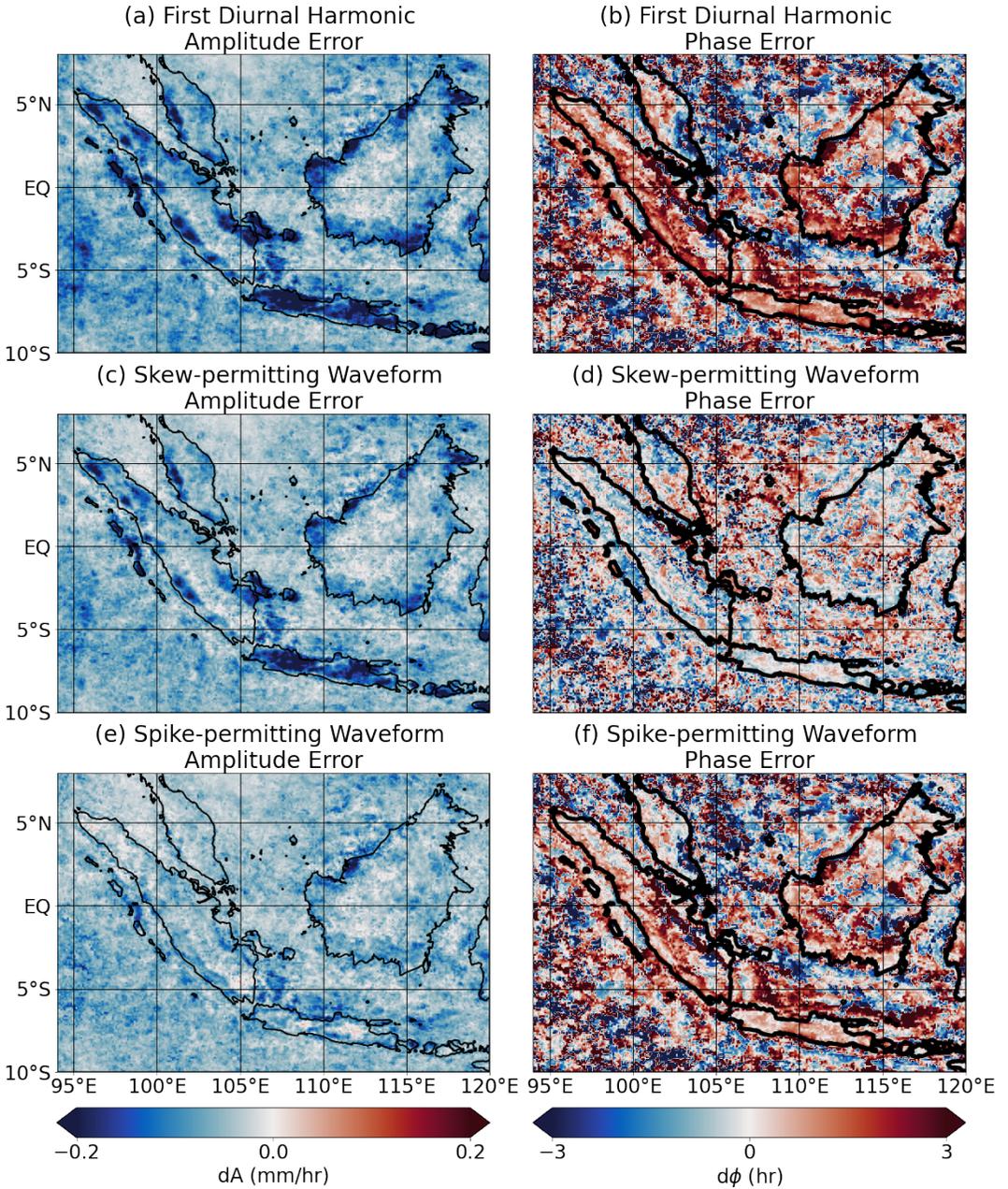


FIGURE 4 Amplitude error (left panels) and phase error (right panels) for: (a&b) the first diurnal harmonic; (c&d) the best-fit skew-permitting waveform, and; (e&f) the best-fit spike-permitting waveform. Errors defined as the waveform parameter (amplitude/phase) minus the comparative basic characteristic (half range/peak time).

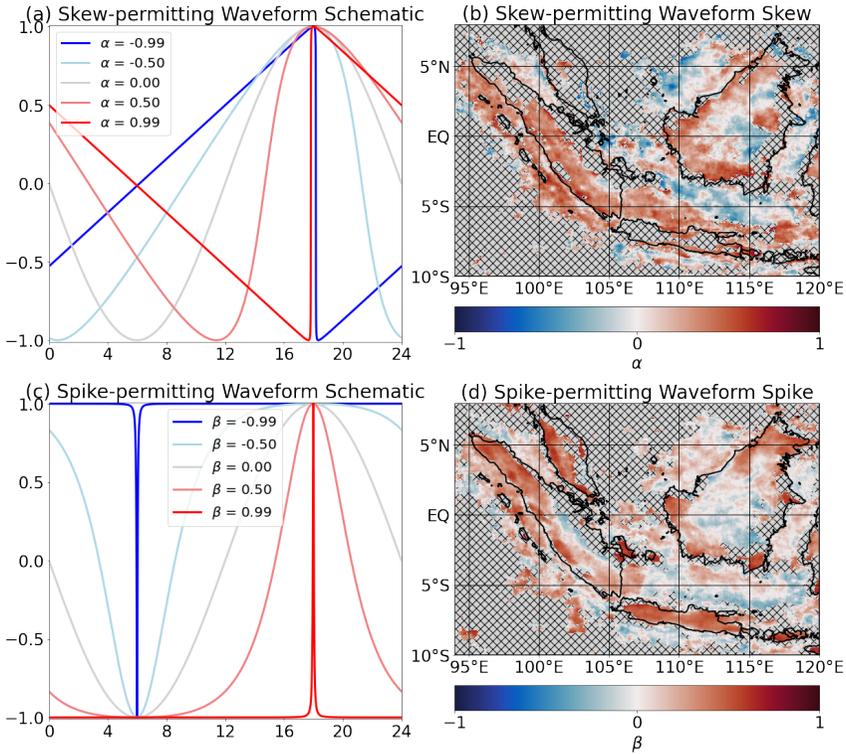


FIGURE 5 Skew and spike behaviour of the diurnal cycle (see Equations (4)&(5)). (a) Skew-permitting waveform structure according to variable rescaled skew parameter, α . $A = 1$, $\phi = 18$ and $c = 0$ for all lines. (b) Map of best-fit skew parameter, partnering the best-fit skew-permitting waveform parameters in Figure 2e&f. (c) Spike-permitting waveform structure according to variable rescaled spike parameter, β . $A = 1$, $\phi = 18$ and $A_g + c = 0$ for all lines. (d) Map of best-fit spike parameter, partnering the best-fit spike-permitting waveform parameters in Figure 2g&h. Hatching is applied where $\epsilon > 0.4$ for the relevant waveform (see Figure 6a&c).

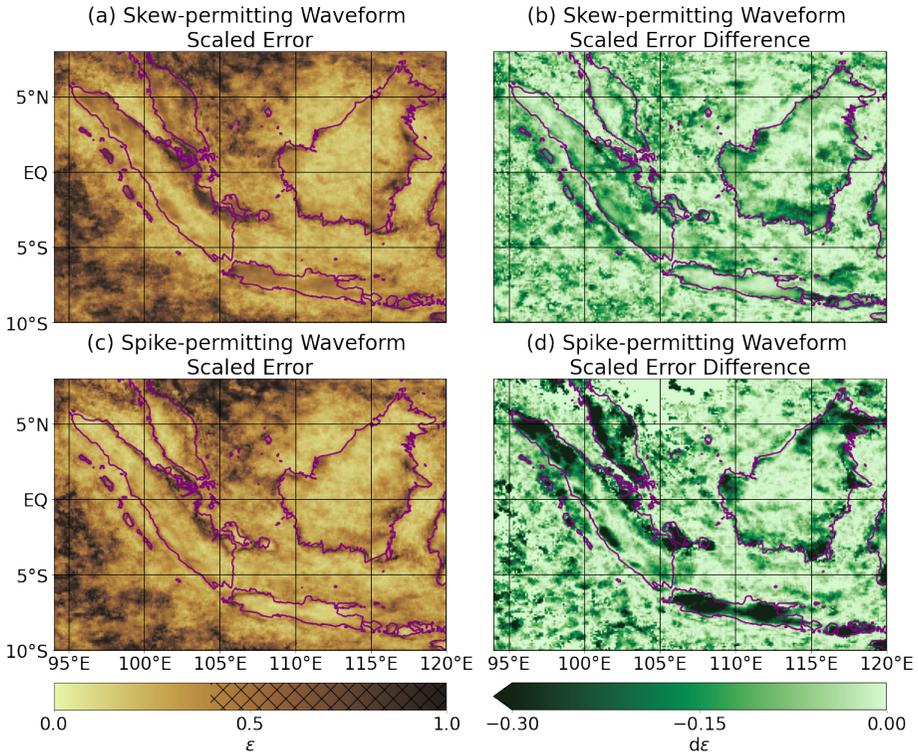


FIGURE 6 Maps of new waveform scaled error (left panels), and improvements relative to the first diurnal harmonic (right panels). Based on (a&b) the best-fit skew-permitting waveform and (c&d) the best-fit spike-permitting waveform. Colour bar hatching corresponds to the map hatching applied to the appropriate waveform in Figures 2e–h and 5b&d.

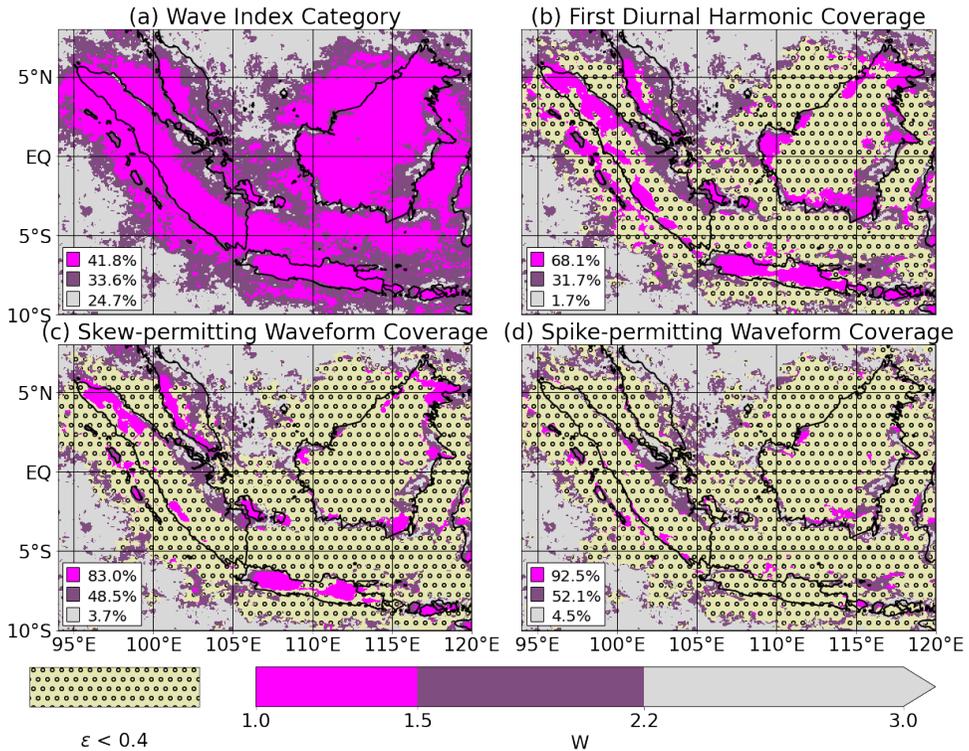


FIGURE 7 Demonstration of the relative skill of the three waveforms in characterising the composite diurnal cycle of precipitation where it is dominated by a single peak. (a) Simplified wave index from Figure 3b, where all magenta ($1.0 \leq W \leq 1.5$) and some lavender ($1.5 < W \leq 2.2$) regions might be expected to be well-characterised with a single-peak waveform. The legend percentages show the proportion of grid points within the study area in each wave index category. The same map features yellow circular hatching overlaid where $\epsilon \leq 0.4$ for: (b) the first diurnal harmonic; (c) the best-fit skew-permitting waveform, and; (d) the best-fit spike-permitting waveform. The legend percentages express the proportion of locations for which $\epsilon \leq 0.4$ for the relevant waveform for each wave index category.

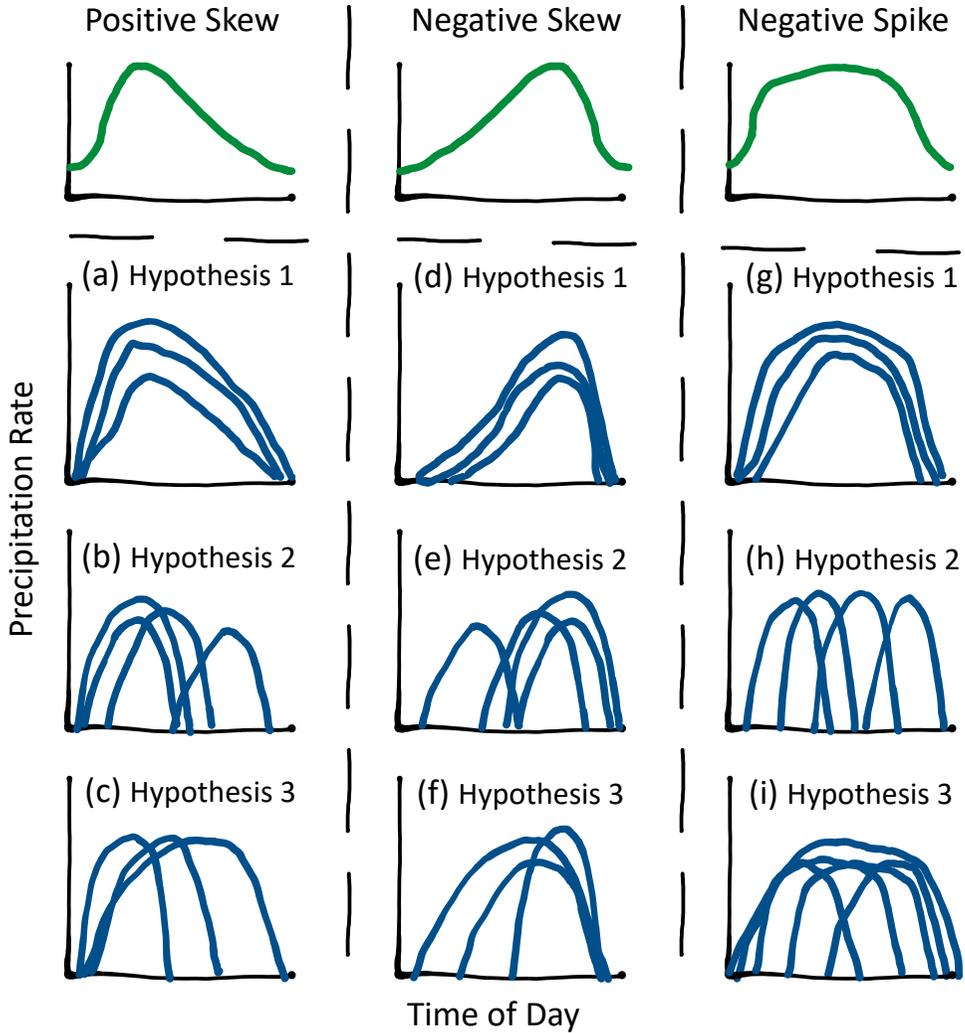


FIGURE 8 Schematic of hypotheses of the range of individual precipitation events (blue) that may combine to produce DCs with positive skew, negative skew and negative spike in the composite (green). (a-c) Hypotheses 1-3 for positive skew, (d-f) hypotheses 1-3 for negative skew, and (g-i) hypotheses 1-3 for negative spike. Note that hypotheses of the same number are independent of each other and need not have the same degree of accuracy.

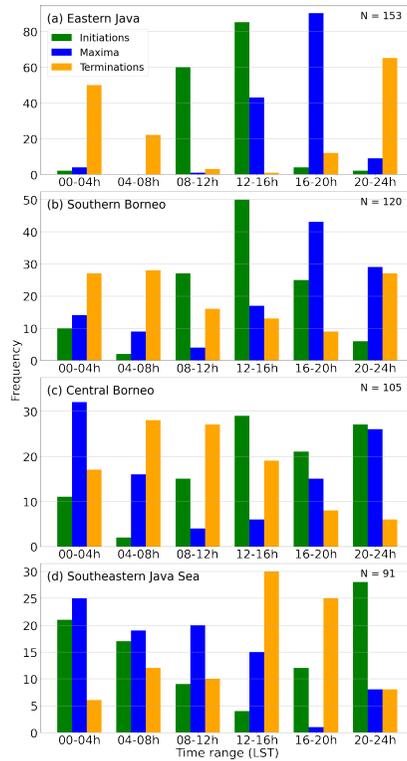


FIGURE 9 Frequency histograms for the timing of initiation (green), maximum (blue) and termination (orange) of individual major precipitation events at the case study locations shown in Figure 1a: a) eastern Java; b) southern Borneo; c) central Borneo, and; d) the southeastern Java Sea. The number of major precipitation events observed at each location is given in the top right of each subplot. Time ranges are in LST for each location.

- Da Silva, N. and Matthews, A. (2021) Impact of the madden–julian oscillation on extreme precipitation over the western maritime continent and southeast asia. *Q. J. Roy. Met. Soc.*, **147**, 3434–3453.
- Helfrich, K. and Melville, W. (2006) Long nonlinear internal waves. *Ann. Rev. Fluid Mech.*, **38**, 395–425.
- Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., Xie, P. and Yoo, S.-H. (2015) Nasa global precipitation measurement (gpm) integrated multi-satellite retrievals for gpm (IMERG). *Algorithm Theoretical Basis Document (ATBD) Version*, **4**, 26.
- Kikuchi, K., Kiladis, G., Dias, J. and Nasuno, T. (2018) Convectively coupled equatorial waves within the mjo during cindy/dynamo: slow kelvin waves as building blocks. *Clim. Dyn.*, **50**, 4211–4230.
- Kikuchi, K. and Wang, B. (2008) Diurnal precipitation regimes in the global tropics. *J. Clim.*, **21**, 2680–2696.
- Kiladis, G., Wheeler, M., Haertel, P., Straub, K. and Roundy, P. (2009) Convectively coupled equatorial waves. *Rev. Geophys.*, **47**.
- Ling, J., Zhang, C., Joyce, R., Xie, P. and Chen, G. (2019) Possible role of the diurnal cycle in land convection in the barrier effect on the mjo by the maritime continent. *Geophys. Res. Lett.*, **46**, 3001–3011.
- Love, B., Matthews, A. and Lister, G. (2011) The diurnal cycle of precipitation over the maritime continent in a high-resolution atmospheric model. *Q. J. Roy. Met. Soc.*, **137**, 934–947.
- Lu, J., Li, T. and Wang, L. (2019) Precipitation diurnal cycle over the maritime continent modulated by the mjo. *Clim. Dyn.*, **53**, 6489–6501.
- Madden, R. and Julian, P. (1971) Detection of a 40–50 day oscillation in the zonal wind in the tropical pacific. *J. Atm. Sci.*, **28**, 702–708.
- (1972) Description of global-scale circulation cells in the tropics with a 40–50 day period. *J. Atm. Sci.*, **29**, 1109–1123.
- Mapes, B., Warner, T. and Xu, M. (2003) Diurnal patterns of rainfall in northwestern south america. part iii: Diurnal gravity waves and nocturnal convection offshore. *Mon. Wea. Rev.*, **131**, 830–844.
- Mori, S., Jun-Ichi, H., Yamanaka, M., Okamoto, N., Murata, F., Sakurai, N. and Hashiguchi, H. (2004) Diurnal land-sea rainfall peak migration over sumatera island, indonesian maritime continent, observed by trmm satellite and intensive rawinsonde soundings. *Mon. Wea. Rev.*, **132**, 2021–2039.
- Oh, J., Kim, K. and Lim, G. (2012) Impact of mjo on the diurnal cycle of rainfall over the western maritime continent in the austral summer. *Clim. Dyn.*, **38**, 1167–1180.
- Peatman, S., Matthews, A. and Stevens, D. (2014) Propagation of the madden–julian oscillation through the maritime continent and scale interaction with the diurnal cycle of precipitation. *Q. J. Roy. Met. Soc.*, **140**, 814–825.
- Peatman, S., Schwendike, J., Birch, C., Marsham, J., Matthews, A. and Yang, G.-Y. (2021) A local-to-large scale view of maritime continent rainfall: Control by enso, mjo, and equatorial waves. *J. Clim.*, **34**, 8933–8953.
- Rais, A. and Yunita, R. (2017) Main diurnal cycle pattern of rainfall in east java. *AIP Conference Proceedings*, **1867**.
- Roxy, M., Dasgupta, P., McPhaden, M., Suematsu, T., Zhang, C. and Kim, D. (2019) Twofold expansion of the indo-pacific warm pool warps the mjo life cycle. *Nature*, **575**, 647–651.
- Sakaeda, N., Kiladis, G. and Dias, J. (2017) The diurnal cycle of tropical cloudiness and rainfall associated with the madden–julian oscillation. *J. Clim.*, **30**, 3999–4020.
- Seiki, A., Yokoi, S. and Katsumata, M. (2021) The impact of diurnal precipitation over sumatra island, indonesia, on synoptic disturbances and its relation to the madden–julian oscillation. *J. Met. Soc. Japan*, **99**, 113–137.

- Skinner, D., Matthews, A. and Stevens, D. (2022) North atlantic oscillation response to the madden–julian oscillation in a coupled climate model. *Weather*, **77**, 201–205.
- Tan, J., Huffman, G., Bolvin, D. and Nelkin, E. (2019) Diurnal cycle of imerg v06 precipitation. *Geophys. Res. Lett.*, **46**, 13584–13592.
- Wang, Y., Heywood, K., Stevens, D. and Damerell, G. (2022) Seasonal extrema of sea surface temperature in cmip6 models. *Ocean Sci.*, **18**, 839–855.
- Worku, L., Mekonnen, A. and Schreck, C. (2019) Diurnal cycle of rainfall and convection over the maritime continent using trmm and isccp. *I. J. Clim.*, **39**, 5191–5200.
- (2020) The impact of mjo, kelvin, and equatorial rossby waves on the diurnal cycle over the maritime continent. *Atmosphere*, **11**.
- Yang, G.-Y. and Slingo, J. (2001) The diurnal cycle in the tropics. *Mon. Wea. Rev.*, **129**, 784–801.
- Yin, Z., Dong, Q., Xiang, K. and Bian, M. (2022) Spatio-temporal characteristics of the indo-pacific warm pool and the corresponding rain pool. *Sustainability*, **14**, 10841.
- Zhang, C. (2005) Madden–julian oscillation. *Rev. Geophys.*, **43**.
- Zhang, C. and Ling, J. (2017) Barrier effect of the indo-pacific maritime continent on the mjo: Perspectives from tracking mjo precipitation. *J. Clim.*, **30**, 3439–3459.
- Zhang, H., Beggs, H., Wang, X., Kiss, A. and Griffin, C. (2016) Seasonal patterns of sst diurnal variation over the tropical warm pool region. *J. Geophys. Res.: Oceans*, **121**, 8077–8094.
- Zhang, W., Cao, Y., Zhu, Y., Zheng, J., Ji, X., Xu, Y., Wu, Y. and Hoitink, A. (2018) Unravelling the causes of tidal asymmetry in deltas. *J. Hydrol.*, **564**, 588–604.
- Zhou, Y., Fang, J. and Wang, S. (2021) Impact of islands on the mjo propagation across the maritime continent: a numerical modeling study of an mjo event. *Clim. Dyn.*, **57**, 2921–2935.