# Global Detection and Analysis of Coastline-Associated Rainfall Using an Objective Pattern Recognition Technique

MARTIN BERGEMANN AND CHRISTIAN JAKOB

*School of Earth, Atmosphere and Environment, Faculty of Science, and ARC Centre of Excellence for Climate System Science, Monash University, Melbourne, Victoria, Australia*

TODD P. LANE

*School of Earth Sciences, and ARC Centre of Excellence for Climate System Science, University of Melbourne, Melbourne, Victoria, Australia*

(Manuscript received 29 January 2015, in final form 5 June 2015)

ABSTRACT

Coastally associated rainfall is a common feature, especially in tropical and subtropical regions. However, it has been difficult to quantify the contribution of coastal rainfall features to the overall local rainfall. The authors develop a novel technique to objectively identify precipitation associated with land–sea interaction and apply it to satellite-based rainfall estimates. The Maritime Continent, the Bight of Panama, Madagascar, and the Mediterranean are found to be regions where land–sea interactions play a crucial role in the formation of precipitation. In these regions ~40%–60% of the total rainfall can be related to coastline effects. Because of its importance for the climate system, the Maritime Continent is a region of particular interest, with high overall amounts of rainfall and large fractions resulting from land–sea interactions throughout the year. To demonstrate the utility of this study's identification method, the authors investigate the influence of several modes of variability, such as the Madden–Julian oscillation and the El Niño–Southern Oscillation, on coastal rainfall behavior. The results suggest that during large-scale suppressed convective conditions, coastal effects tend to modulate the rainfall over the Maritime Continent leading to enhanced rainfall over land regions compared to the surrounding oceans. The authors propose that the novel objective dataset of coastally influenced precipitation can be used in a variety of ways, such as to inform cumulus parameterization or as an additional tool for evaluating the simulation of coastal precipitation within weather and climate models.

## 1. Introduction

Precipitation, one of the most important meteorological variables, is strongly affected by variations in solar forcing. As a result, tropical rainfall variability is strongly dominated by the seasonal and the diurnal cycle. Yang and Slingo (2001) showed the importance of rainfall variance within diurnal and subdiurnal frequencies for coastal tropical regions such as the Maritime Continent. In this area the diurnal rainfall variability is thought to be mostly generated by land–sea-breeze circulations (Mori et al. 2004). Land–sea-breeze systems are mainly forced by differential heating between land and the adjacent ocean but also affected by a variety of different factors, such as coastline curvature, latitude, topography, atmospheric stability, land use, and synoptic wind patterns (e.g., McPherson 1970; Haurwitz 1947; Pielke 2002; Estoque 1962; Mak and Walsh 1976; Mahrer and Pielke 1977). Crosman and Horel (2010) provide a comprehensive review of the studies that have been conducted about the nature of land–sea-breeze circulation systems. These circulation systems cause characteristic rainfall patterns in coastal regions. Mori et al. (2004) showed that the Maritime Continent rainfall between 2100 and 0900 LT is concentrated over the oceans, peaking in the early morning. The 0900–2100 LT precipitation is mainly located over land, with maxima occurring in the early evening. The rainfall patterns associated with land–sea interaction tend to propagate roughly 150 km onshore and offshore (Keenan and Carbone 2008). Further propagation can occur through the interaction with other phenomena, such as

*Corresponding author address:* Martin Bergemann, School of Earth, Atmosphere and Environment, Monash University, Melbourne VIC 3800, Australia.
E-mail: martin.bergemann@monash.edu

DOI: 10.1175/JCLI-D-15-0098.1

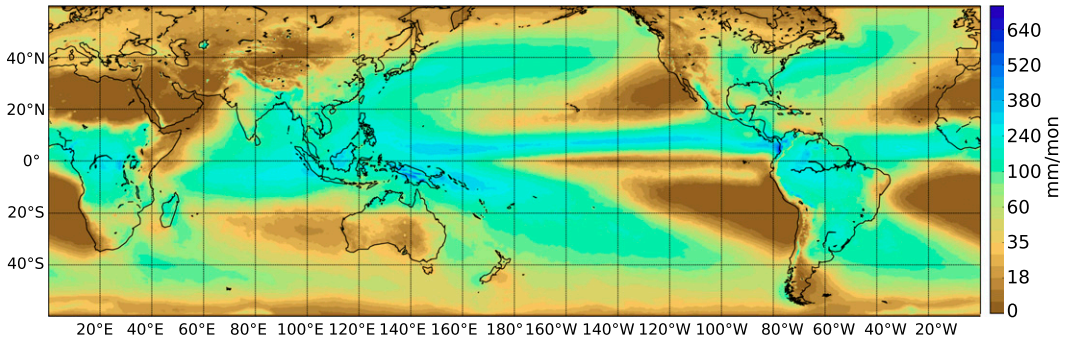


FIG. 1. Average monthly sum of total rainfall (1998–2013) for the region that is covered by the CMORPH satellite-based rainfall estimates.

mountain–valley breeze systems (Qian 2008) or gravity waves (Mapes et al. 2003).

Roughly 20% of the world's population lives within the area that is affected by coastal precipitation. Additionally, the mean population density is about 3 times higher near coasts than on global average (Small and Nicholls 2003). Coastal areas are also vulnerable to an increase of storm surges and heavy precipitation. Therefore, a more accurate simulation of coastal rainfall in global climate models can potentially contribute directly to a better assessment of climate impacts on coastal areas. Precipitation, on the other hand, remains a challenging meteorological variable in general circulation models. Several studies have documented the issues in representing precipitation in climate models. For example, Sun et al. (2006) and Dai (2006) compared rainfall simulations of 18 coupled models with observations and found that while most models are able to capture the broad pattern of precipitation amount and year-to-year variability, they fail to reproduce the diurnal cycle. Compared with observations, modeled rainfall is too weak and too frequent (Stephens et al. 2010). In coastal areas, the spatial pattern and timing of precipitation becomes worse (Collier and Bowman 2004). Especially over the Maritime Continent most climate models reveal a dry–wet rainfall bias, with too-wet conditions over the ocean and too-dry conditions over land, or vice versa. It is likely that this is related to the complex structure of islands in combination with steep terrain. This combination leads to complex coastal convective systems causing rainfall patterns that are not easily captured by the relatively coarse-resolution global climate models. Although advances in convection parameterization and model resolution have been made to tackle the problem of rainfall timing and intensity in climate models, the main issues in simulating the diurnal precipitation cycle remain unsolved (e.g., Warner et al. 2003; Slingo et al. 2004; Sato et al. 2009; Gianotti et al. 2012; Folkins et al. 2014).

One of the key problems for an accurate description of coastal convection and precipitation is a fundamental lack of a global dataset of coastally induced precipitation. Although many studies have been conducted to describe convection, precipitation, and land–sea-breeze circulation systems (e.g., Frizzola and Fisher 1963; Tijm et al. 1999; Zhuo et al. 2014; Wapler and Lane 2012), the vast majority of them are both local and phenomenological because precipitation that is induced by land–sea interaction has to be separated from the background state of overall rainfall. So far, to our knowledge, no method exists that attempts to objectively identify rainfall directly associated with land–sea interactions. The aim of this study is to describe and evaluate a method that objectively finds coastal precipitation patterns that are related to land–sea interaction. Once the method is developed and assessed, a global climatology of coastally induced rainfall and its diurnal cycle are presented. To give an example for the utility of the derived dataset, the roles of the Madden–Julian oscillation (MJO) and El Niño–Southern Oscillation (ENSO) in coastal precipitation in the Maritime Continent region are then investigated.

Section 2 describes the rainfall data used in the study. Section 3 discusses the broad features of the objective coastal rainfall detection technique, with a more technical description provided in the appendix. Sections 4 to 6 then describe its application to a global rainfall dataset to study the global climatology of coastal rainfall, its diurnal cycle, and the influence of larger-scale modes of variability on the coastal rainfall occurrence. This is followed by a summary and conclusions in section 7.

## 2. Rainfall observations

The rainfall data used in this study are based on satellite rainfall estimates of the Climate Prediction Center Morphing Method (CMORPH; Joyce et al. 2004). The dataset has a spatial resolution of 0.25° and covers the global area from 60°S to 60°N (Fig. 1). The

temporal resolution is 3 h, and the time period used is 1998–2013.

The morphing method uses motion vectors that are derived from half-hourly interval geostationary satellite infrared imagery to propagate the precipitation estimates derived from passive microwave scans. In addition, the shape and intensity of the precipitation features are modified during the time between microwave sensor scans by performing a time-weighted linear interpolation. This process yields a spatially and temporally complete microwave-derived precipitation analysis, independent of the infrared temperature field. The dataset shows substantial improvements both over simple averaging of the microwave estimates and over techniques that blend microwave and infrared information. Yet there are still several issues: One is the lack of accuracy of precipitation estimation over snowy regions. Another is related to precipitation that dissipates over regions that are not covered by the passive microwave satellite. This rainfall cannot be captured by the morphing algorithm. Although CMORPH ranks among the best available satellite-based rainfall estimates, the correlation with rain gauge products is far from being perfect (Joyce et al. 2004).

## 3. Pattern recognition

In the present work we wish to extract rainfall whose structure indicates that it is associated with coastal phenomena like land–sea-breeze convergence. Previous studies have shown that rainfall related to coastal land–sea interaction is mainly of higher intensity than the rainfall that is not affected by the presence of the coastline in that area. Furthermore, the precipitation patterns of interest are aligned with the coastline and should occur in the vicinity of the coastline (e.g., Mori et al. 2004; Keenan and Carbone 2008; Qian 2008; Hill et al. 2010). Based on this knowledge we define four heuristics that have to be met to identify precipitation as associated with coastal land–sea interaction:

(i) compared with the surrounding precipitation, the rainfall of interest has a higher intensity,
(ii) the recognized coastal precipitation is not large scale,
(iii) rainfall due to land–sea interaction occurs within ~250 km of the coast, and
(iv) the precipitation pattern is aligned with the coastline.

The above heuristics are applied in six steps, which we briefly summarize below and in Figs. 2a–f. A more technical description of our algorithm can be found in the appendix.

(i) Application of a rainfall threshold to the rain data.

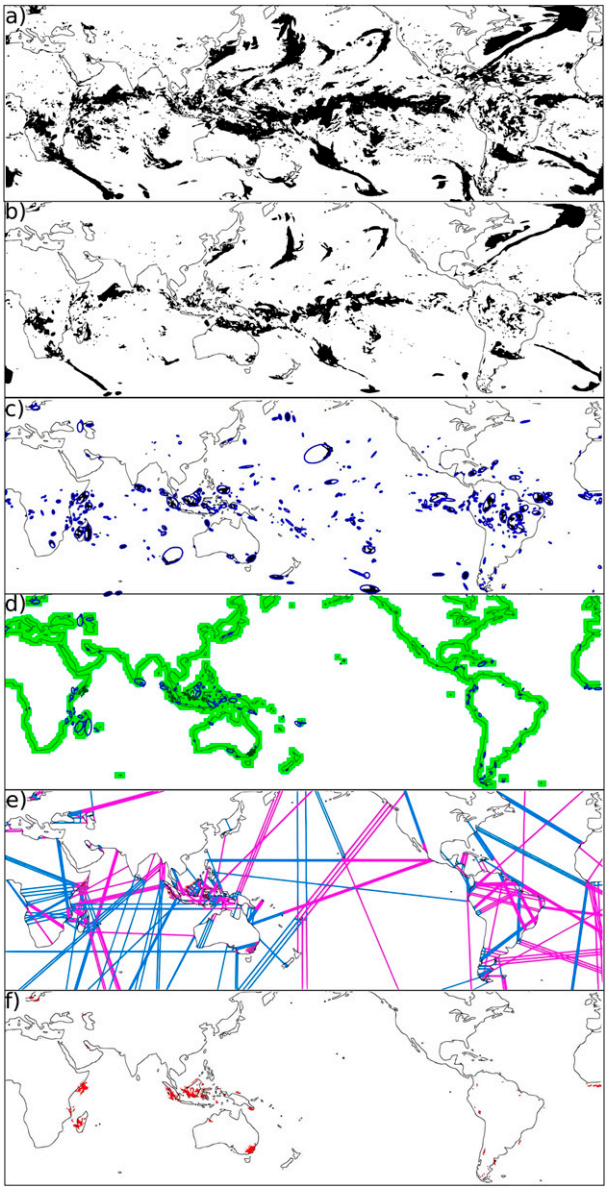


FIG. 2. The sequence of coastline-induced rainfall recognition algorithm. (a) Conversion of the rain data to a binary image. (b) Applying a rainfall threshold. (c) Deleting synoptic-scale rainfall patterns and fitting the remaining patterns to ellipses. (d) Defining a broad coastal area and deleting any rainfall domains not located in the green marked area. (e) Alignment testing with straight lines for the fit ellipse to the next coastline. (f) Result.

To apply the first-mentioned heuristic, we choose local monthly rainfall percentile thresholds instead of a hard threshold. At every grid point and for every month we determine the deciles of the 3-hourly rainfall distribution and apply thresholds of the 20th, 50th, and 80th percentile, respectively (see below). With monthly percentiles as a threshold,

regional and seasonal variations of the rainfall are taken into account.

(ii) Converting the rainfall data to binary data.

Here any rainfall greater than the rainfall threshold is set to 1, and lower values are assigned as 0. To test geometrical aspects like size and coastline alignment, each contiguous rainfall area has to be considered separately. The separation of the different rainfall areas uses a gradient method, which performs best on binary data (see appendix for further details).

(iii) Deletion of large-scale rainfall.

After separation, the total size of each contiguous rainfall area is measured. Using a size threshold of $10^6$ km$^2$, rainfall patches that can be accounted as synoptic-scale patterns, like large frontal systems, are then deleted from the data.

(iv) Fitting the rainfall domains with ellipses and applying an eccentricity threshold.

Since we are interested in rainfall features that stretch along a coastline, we assume that the detected features are elongated. Therefore, the rainfall domains are least squares fitted with ellipses. After fitting, each contiguous rain area is enclosed by an ellipse. We then apply a threshold for the eccentricity of the ellipses to delete rainfall patterns that are not elongated. The application of an eccentricity threshold is also necessary because the remaining rainfall patterns will be tested for alignment with the coastline. Alignment can only sensibly be determined if the rainfall area has a clear orientation. This is ensured by an eccentricity of the fit ellipse larger than a threshold.

(v) Deleting rainfall not occurring in the vicinity of the coast.

Here we define a broad area (~250 km) around the coastline and delete all contiguous rainfall domains not occurring in the defined coastal area. Specifically, if less than 20% of the contiguous rainfall area lies within the coastal vicinity it is deleted from the dataset. With this blurred threshold we do not cut off rainfall domains stretching over the borders of the coastal area. We are also able to detect patterns that are slightly farther onshore or offshore than the distance threshold defined in the heuristics. This has the advantage that the hard areal threshold becomes smoother.

(vi) Testing the alignment of the rainfall domains with the coastline.

For alignment testing we first define a narrow coastal area (~50 km). Any rainfall area with at least 90% of its area in this narrow coastal strip is assumed to be aligned with the coastline and already marked as a coastally influenced precipitation pattern. For each of the remaining objects, three straight lines, two from the tips and one from the center of the fit ellipse, are drawn orthogonally in both directions of the major axis. The distance from the origin to the next coastline intersection for each straight line is measured. If the standard deviation of the three distances is below a certain percentage of the mean distance, the contour is assumed to be aligned with a coastline. The object is then marked as aligned with the coast if the mean distance of all corresponding straight lines is not greater than 500 km.

TABLE 1. Threshold values used in the recognition algorithm yielding an ensemble of $3^3 = 27$ coastal rainfall estimates.

| Threshold | Values | | |
|---|---|---|---|
| Rainfall intensity (percentile) | 20 | 50 | 80 |
| Eccentricity | 0.2 | 0.5 | 0.8 |
| Variation of straight line length (%) | 5 | 25 | 50 |

All rainfall domains that have not been deleted during the six steps meet the previously mentioned heuristics and are assumed to be precipitation that is associated with coastal land–sea interaction. The algorithm is applied to all single 3-hourly time steps of the CMORPH dataset from 1998 to 2013. A more comprehensive technical description of the algorithm can be found in the appendix. The source code of the algorithm can be downloaded freely via GitHub (http://dx.doi.org/10.5281/zenodo.18173).

From the above discussion it is evident that thresholds have to be applied to run the pattern recognition algorithm. These thresholds are

(i) the area of the rainfall patches,
(ii) the intensity of the rainfall,
(iii) the eccentricity of the fitted ellipses, and
(iv) the threshold for the standard deviation of different straight lines from the main axis of the fit ellipse to the next coastline.

As the synoptic scale is well defined, the area threshold is set to a single value (~$10^6$ km$^2$). The choice of the other three thresholds is more difficult. We make reasonable choices for all three thresholds and generate an ensemble of 27 members by using all possible combinations of the threshold settings summarized in Table 1. The decision to use an ensemble was made to keep the method objective rather than subjectively choosing a ''perfect'' parameter set. For each parameter we chose the highest and lowest reasonable threshold and one value in between.

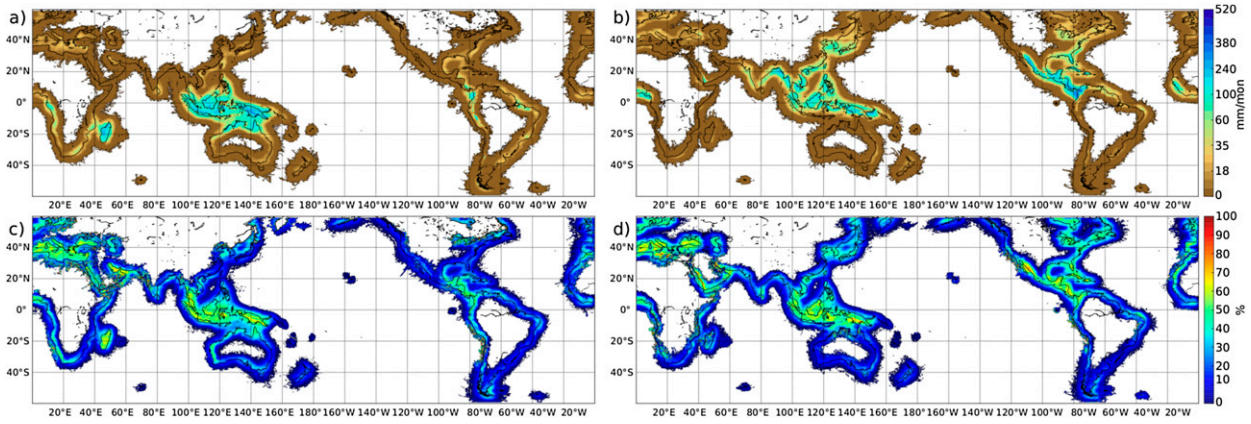


FIG. 3. Mean coastal rainfall for (a),(c) DJF and (b),(d) JJA. The ensemble mean of detected coastal rainfall is shown in (a) and (b), and the detected precipitation as a percentage of total rainfall is shown in (c) and (d).

The 20th percentile is chosen as a reasonable low rainfall intensity threshold. Values less than the 20th percentile would be contradictory to the a priori assumption that coastal rainfall is of relatively high intensity. Also, with low rainfall thresholds the rainfall data become noisier. This leads to a poor performance of the applied Canny edge detection.

Canny edge detection performed best by applying a rainfall threshold greater than the 80th percentile. With such high rainfall thresholds, however, only a small number of rainfall events are remaining. Therefore, we chose to set the highest reasonable threshold choice as the 80th percentile. The application of the eccentricity threshold as an orientation parameter is similar. Low values indicate rainfall patterns that are rather round. To assign a clear major axis that is needed for alignment testing, the objects should not be round but elliptical. We therefore chose 0.8 as a maximal threshold. Thresholds of less than 0.2 turned out to be effective for the alignment testing. On the other hand, too much precipitation was deleted when further decreasing the threshold. Therefore, we chose 0.2 as a reasonable lower threshold.

Rainfall patterns with an angle of less than 45° can be considered as aligned with the coastline. They would roughly correspond to the variation of the straight lines by 50%. Therefore, we choose the upper 50% as an upper threshold. Nevertheless, the testing with different straight-length variation thresholds indicates that this parameter has less impact on the results than the threshold choice for rainfall intensity and the eccentricity of the fit ellipse.

## 4. Overall climatology of detected coastal rainfall

All results presented here are based on the ensemble of objectively identified coastal rainfall described in the previous section. All ensemble members are taken into account and no weighting is applied.

The climatology of the detected coastal precipitation, shown in Figs. 3a,b, clearly reveals the expected seasonal variability in the tropics and subtropics. During December–February (DJF) coastal rainfall maxima occur in the Maritime Continent region and over Madagascar. June–August (JJA) shows peaks over the Bay of Bengal, the Central American Pacific coast, the North American Atlantic coast, and tropical West Africa. Also, relatively high amounts of precipitation are detected over the Gulf of Carpenteria (15°S, 140°E) during austral summer. In both seasons high amounts of coastal rainfall are detected in the Bight of Panama and around the Maritime Continent.

In the tropical coastal regions high values of coastal rainfall are accompanied by a high percentage of the contribution of coastal rainfall to the total (Figs. 3c,d). Perhaps the most prominent example all year round is the Maritime Continent, with the coastal rainfall contribution frequently exceeding 50% and approaching 66% in some regions. This highlights the importance of land–sea interaction for Maritime Continent rainfall. Relatively high amounts of coastally related rainfall are found in the highlands of New Guinea. This might originate from orographic effects caused by the mountain flanks being roughly parallel to the New Guinean coastline. The interaction of coastal and orographic effects is difficult to separate. Qian et al. (2012) showed the strong interconnection of orographic and coastal effects for the Maritime Continent rainfall. Furthermore, Warner et al. (2003) conjectured a complex interaction of stratified air above topography and moist convection triggered by land-breeze lifting. Therefore, we account these detected features as coastline associated.

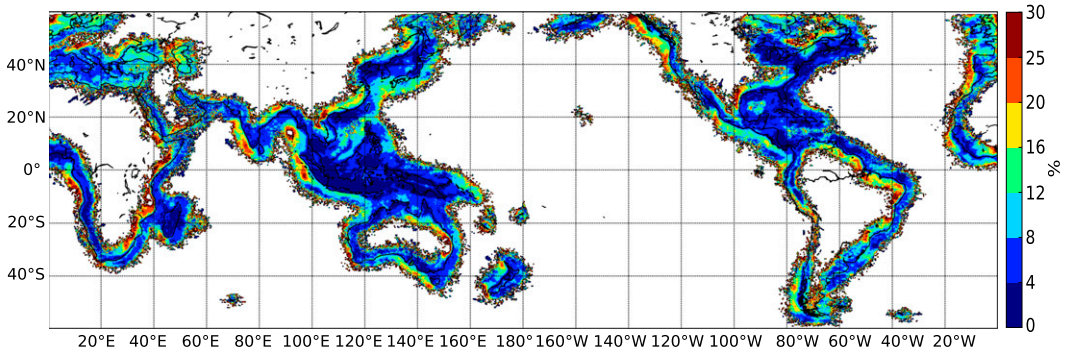


FIG. 4. The ensemble standard deviation as percentage of the ensemble mean.

The fraction of detected coastal rainfall also reveals the importance of land–sea interaction in arid and semiarid coastal regions. The Red Sea, the Persian Gulf, the Australian west coast, and the southwest African coast are examples where, despite relatively modest overall rainfall amounts, major portions of the precipitation are aligned with the coast. The Mediterranean is the second largest region (after the Maritime Continent) that is strongly influenced by land–sea interaction. Even in the Northern Hemisphere winter, the percentage of detected coastal precipitation remains high. This signal during DJF in the Mediterranean might have its origin in frontal systems that are aligned with the coastline and should therefore be treated with caution.

Figure 4 shows the ensemble standard deviation as a percentage of the ensemble mean. It is interesting to note that the ensemble members show better agreement over tropical land areas than over the adjacent ocean. The standard deviation tends to rise with an increasing distance from the coastline. It is known that with greater distance from the coast the precipitation patterns become more influenced by ambient flow patterns (Gilliam et al. 2004; Azorin-Molina and Chen 2009), and thus the alignment of the rainfall areas with the coast becomes more sensitive to the application of an alignment threshold. Relatively high standard deviations are found in the extratropics. A more detailed investigation reveals that in this area our algorithm is particularly sensitive to the application of the rainfall threshold, leading to larger variations across the different ensemble members.

## 5. The diurnal cycle of coastal rainfall

Figure 5 shows the local daytime (Figs. 5a–c) and nighttime (Figs. 5d–f) rainfall as a fraction of overall rainfall for all rainfall (Figs. 5a,d), the detected coastal rainfall events only (Figs. 5b,e), and the residual rainfall considered noncoastal in nature (Figs. 5c,f). Here, daytime is referred to as 0900–2100 LT and nighttime as 2100–0900 LT. For total rainfall (Figs. 5a,d) the diurnal variability is very large in the Maritime Continent, the Bight of Panama, the west coast of Central America, the eastern Bay of Bengal, the African east coast, and around the Horn of Africa. The fraction of detected daytime rainfall, shown in Figs. 5b,e, reveals a more pronounced diurnal variation than the total rainfall. Some regions have up to 95% more rainfall during the day over land than over the adjacent sea, and vice versa for nighttime rainfall. The most significant diurnal rainfall variations occur over the Maritime Continent, Madagascar, and the Bight of Panama. The fraction of local nighttime rainfall to overall rainfall is shown on the right-hand side of Fig. 5. Note that daytime and nighttime rainfall together add up to total rainfall. Strong nighttime signatures of detected coastal precipitation are evident over the Central American Pacific coast and the east coast of the United States as well as the Maritime Continent. Weaker but still discernible nighttime signals are detected in Southeast Asia, the Bay of Bengal, and around the Philippines. Comparing the daytime and nighttime fractions, it is evident that the diurnal rainfall cycle is stronger over land than over the adjacent ocean.

It is well known that the coastal diurnal rainfall cycle is strongly affected by land–sea interaction. Therefore we hypothesize that 1) diurnal rainfall variations of the detected coastal precipitation should be much larger than those of the overall rainfall near coasts, and 2) the residual rainfall that has not been detected as coastally influenced should reveal very little diurnal variation. Figure 5c shows this remaining residual daytime and nighttime rainfall as a fraction of overall residual rainfall. The residual fractions of daytime and nighttime precipitation are small, with values mostly between 35% and 65%. Note that no diurnal variation would be indicated by 50% (green areas). This supports our claim that the presented technique is able to capture the majority of rainfall that is due to land–sea interaction in

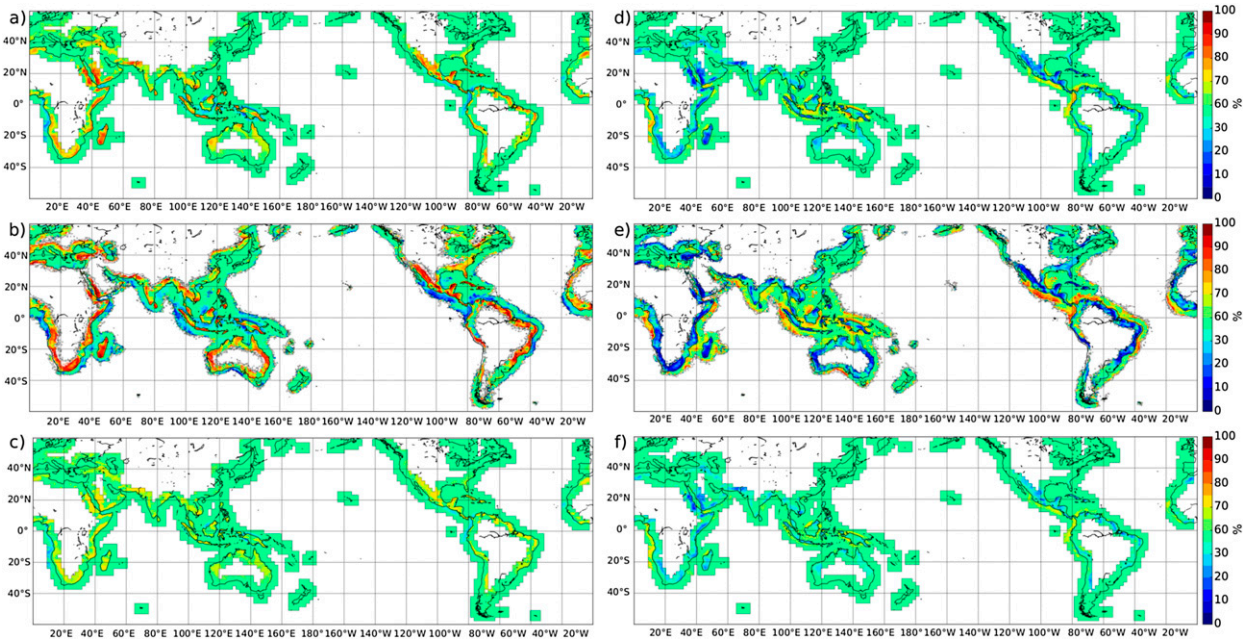


FIG. 5. Mean annual (left) daytime and (right) nighttime rainfall as a fraction of overall daily rainfall for (top) total precipitation, (middle) detected coastal precipitation, and (bottom) residual rainfall. Daytime is defined as 0900–2100 LT and nighttime as 2100–0900 LT. For (top) and (bottom) only rainfall within 250 km onshore and offshore is shown.

coastal areas. In some areas the residual rainfall variation remains significant though. In particular, the northwest coast of Borneo, the west coast of Sumatra, and the Bight of Panama show a relatively high amount of residual nighttime rainfall over the ocean.

## 6. Coastal rainfall over the Maritime Continent

The Maritime Continent, not surprisingly, has been identified as one of the regions where precipitation is particularly strongly influenced by land–sea interaction. Figure 6 shows the average timing of rainfall over land and over water for the Maritime Continent and on global average. Here, the dashed lines represent the detected coastal precipitation and the solid lines the total precipitation. For a better comparison, the rainfall is normalized by making the sum of a full cycle equal to one. The diurnal cycle over the Maritime Continent is clearly more pronounced than on global average. This is especially true for the land regions. The overall distribution of the maxima shows the expected behavior. Over land the rainfall peaks in the late afternoon, whereas the maximum over water occurs in the early morning. This is in agreement with the results of Mori et al. (2004), who were using TRMM-3B42 satellite rainfall estimates.

The detected precipitation shows a stronger diurnal variation, especially over water. The detected minimum of oceanic rainfall occurs roughly 3 h earlier than the total rainfall minimum. In general the detected coastal precipitation variation is most pronounced over land. The detected coastal rainfall over the Maritime Continent shows roughly the same diurnal variation as the recognized precipitation on global average.

To further assess the utility of the coastal rainfall dataset, we investigate the influence of large-scale modes of climate variability on coastal precipitation in the Maritime Continent. In particular, we continue on from earlier studies (Rauniyar and Walsh 2011, 2013) and investigate how the MJO and ENSO modulate the diurnal cycle. First, the difference of coastal rainfall in the Maritime Continent region during active and suppressed MJO phases is calculated. The MJO is considered to be active over the Maritime Continent in phases 4–6 of the Wheeler–Hendon real-time multivariate MJO (RMM) index (Wheeler and Hendon 2004). The inactive phase is defined with index values of 1, 2, 7, and 8. Only days with an RMM amplitude of equal to or greater than 1 were sampled. Figure 7 shows the difference of total DJF rainfall (Fig. 7a) and DJF detected coastal rainfall (Fig. 7b) between suppressed and active MJO phase.

In general there is more rainfall during the active phase over the ocean. However, as has been found in previous studies [such as Rauniyar and Walsh (2011), who were using nonobjective methods], over land there are large areas where rainfall during the suppressed

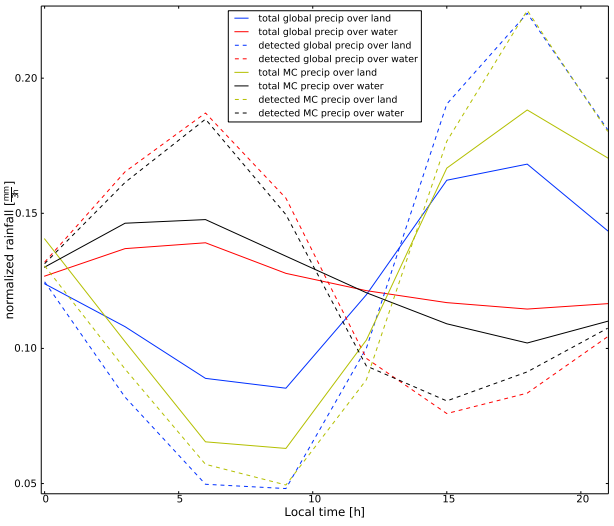


FIG. 6. The mean occurrence of global rainfall and rainfall over the Maritime Continent over land and ocean as a function of the time of the day. For a better comparison of the diurnal global and Maritime Continent rainfall, the sum of a full rainfall cycle is set to one.

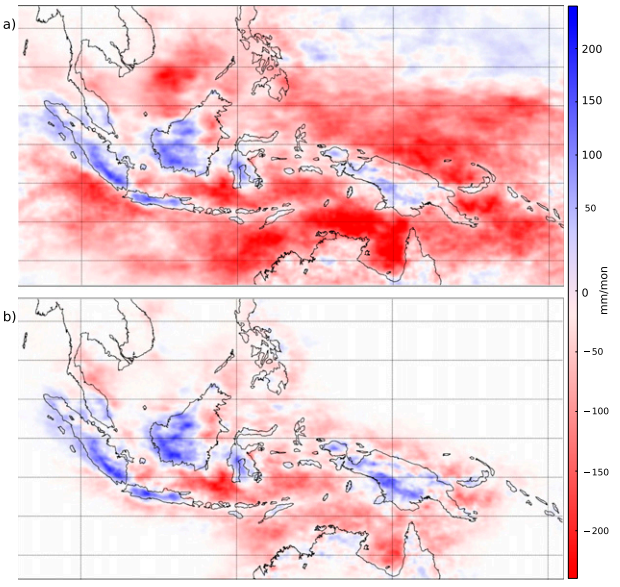


FIG. 7. Mean difference of DJF rainfall during suppressed and active MJO phase for (a) total rainfall and (b) detected coastal rainfall.

phases is larger than that during the active phase. Interestingly, both the magnitude and pattern of the overall signal are well reproduced when considering only coastally influenced precipitation. This leads us to conclude that it is mainly the precipitation due to land–sea interaction that leads to an increase of rainfall over land during the suppressed phase. This is an important finding, especially in the light that weather and climate models of coarse resolution are unlikely to be able to capture land–sea effects correctly.

The ENSO also influences the rainfall over the Maritime Continent. Global rainfall pattern differences tend toward dryer conditions in the western Pacific during the El Niño phase of the oscillation. Figure 8a shows the mean DJF difference of total rainfall between El Niño years and weak ENSO years. Our El Niño definition is based on the oceanic Niño index (ONI). The ONI is based on a 3-monthly running mean of the SST anomaly in the Niño-3.4 region (5°N–5°S, 120°E–170°W). ONI values of greater than 0.5 indicate El Niño events, and values within ±0.5 show weak ENSO episodes.

It can be clearly seen that there is more rainfall over the ocean during years in which ENSO is weak. On the islands, on the other hand, the situation looks different. There is more rainfall over land regions of the Maritime Continent during El Niño years. The rainfall pattern difference for the detected coastal rainfall looks very similar to the total precipitation (see Fig. 8). The magnitudes of the positive detected coastal rainfall anomalies over land are about the same order as the anomalies of the total rainfall.

Once again, as for the MJO events, the variations of rainfall with El Niño support the hypothesis that during suppressed convective large-scale conditions, coastal effects strongly modulate the rainfall over the Maritime Continent and contribute largely to an enhancement of rainfall (Qian et al. 2013; Rauniyar and Walsh 2013). The fact that the results using the objectively detected coastal rainfall over the islands are very similar to those

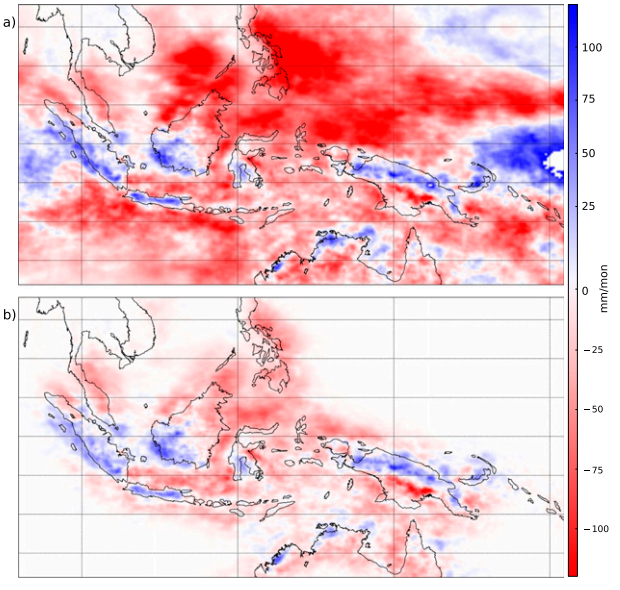


FIG. 8. Mean difference of DJF rainfall during weak ENSO and El Niño phase for (a) total rainfall and (b) detected coastal rainfall.

using total rainfall provides some evidence that the algorithm developed in this study is able to identify coastally driven precipitation events well.

## 7. Summary and conclusions

In the present study a pattern recognition technique is developed and applied to a global 3-hourly rainfall dataset to detect rainfall associated with land–sea interactions. The technique applies thresholds for four different characteristics to extract rainfall that is likely driven from coastal features. Since the algorithm employs several thresholds and since the implications of the threshold choice are not entirely knowable, an ensemble of 27 different threshold setups is created. The ensemble mean of the generated dataset reveals the expected seasonal and diurnal variability of coastal tropical precipitation. The standard deviation among the ensemble members, which is on average 4% of the ensemble mean, indicates the overall robustness of the algorithm. Over the Maritime Continent and the Bight of Panama major portions of total precipitation can be clearly related to coastal land–sea interaction. High fractions of rainfall associated with coastal rainfall over the Mediterranean, the Red Sea, the Persian Gulf, the South African coast, and the Australian west coast highlight the importance of coastal processes in arid and semiarid regions of the world. Most of the expected features of coastal convection are captured by the algorithm. This is indicated by the strong diurnal cycle of the detected rainfall and relatively weak diurnal rainfall variation of the residual rainfall.

The coastal rainfall recognition method is only based on a few geometrical aspects of precipitation; no further criteria have to be met. This differs from many statistical approaches, including cluster, spectral, or principal-component analysis, where assumptions like the occurrence of diurnal harmonics have to be applied. Therefore, already-known statistical properties, like the diurnal precipitation cycle, can be studied to evaluate the newly developed method. Moreover, further aspects of coastal precipitation can now be easily illustrated. For instance, coastline effects are known to be an important trigger for deep precipitating convection (Simpson et al. 1980, 1993; Qian 2008). The presented method reveals that up to two thirds of the total Maritime Continent rainfall can be related to coastal effects.

We are also able to objectively show that coastline-associated precipitation is influenced by modes of large-scale variability like the MJO and ENSO. The results suggested that during suppressed large-scale convective conditions the total precipitation over land areas is strongly modulated by coastline effects. This is in accordance with previous studies using nonobjective statistical methods (Rauniyar and Walsh 2011, 2013).

The strength of the method presented here is that it is objective and relatively straightforward to apply. This has allowed us, to our knowledge for the first time, to characterize coastal precipitation globally rather than just for a limited amount of regional cases. The dataset of coastal convection features will also be very useful for future studies. For instance, it can be applied to identify the background atmospheric state in which coastal precipitation occurs. This is crucial for the parameterization of such rainfall in models, as the MJO and ENSO analysis shows that coastal effects likely allow rainfall to exist in large scale conditions, in which rainfall is suppressed over the open ocean. Since it can be driven with any kind of gridded rainfall data, the algorithm can moreover be utilized to evaluate coastal precipitation projected within climate models.

*Acknowledgments.* We acknowledge the Australian Research Council's Centre of Excellence for Climate System Science (CE110001028) for funding this work. We would also like to thank Rit Carbone from NCAR and Chris Holloway from University of Reading for their useful suggestions and comments in the early stage of this work. Furthermore, we thank the three anonymous reviewers for their valuable comments and suggestions to improve the quality of the publication. The CMORPH satellite-based rainfall estimates were obtained from the Climate Prediction Center (CPC) of the National Oceanic and Atmospheric Administration (NOAA). The tools utilized for the pattern recognition are supplied by the open source image processing library OpenCV. The source code and documentation can be retrieved from GitHub (http://dx.doi.org/10.5281/zenodo.18173).

## APPENDIX

### The Pattern Recognition Algorithm

A key task in the mining of remote sensing imagery is the identification of static structures such as buildings, roads, and bridges. Nevatia and Price (1982) were one of the first to identify the airports in the San Francisco Bay area. They applied some simple characteristics about the airports in this area. Specifically, they assumed an airport has a straight runway of certain length and is located near the coast. Starting from these heuristics, they used simple image segmentation techniques to detect the target features. To detect the target precipitation features, we also use image segmentation. Image segmentation is a technique where information is extracted

from the background of an image until all a priori assumptions are met (see section 3 for the heuristics used here) and the remaining part of the image can be considered as the detected pattern.

We start the pattern detection by applying the rainfall intensity threshold and converting the rainfall data to a binary image. Areas less than the threshold are set to 0 and the ones above to 1. After converting the rainfall data to a binary image, small holes within the rainfall areas (1-domains) are closed. Closing of small holes within a connected domain can be done by dilation and erosion. Dilation and erosion are methods where a domain of an image (A) is probed with a structuring element (B), and the way in which the element fits inside an image object is quantified. The structuring element, in image processing also referred to as a kernel, is applied on the binary image. In the present case, a cross of $3 \times 3$ pixels is chosen as the kernel shape. Formally, dilation and erosion are defined as

$$\text{dilation: } A \oplus B = \bigcup_{b \in B} A_b$$

$$\text{erosion: } A \ominus B = \bigcap_{b \in B} A_{-b}.$$

For dilation, the kernel B is scanned over the image domain A, and the pixels overlapped by B are added to A. Thus, the black regions within A are growing. Erosion is very similar to dilation; the center of B is subtracted from A if A and B are only partly overlapping. Therefore, the black regions in A are shrinking. These two methods are applied consecutively to close small holes within the domain A.

To apply the heuristics mentioned in section 3, each rainfall domain has to be considered separately. The separation of the patterns is realized by Canny edge detection (Canny 1986). For Canny edge detection, an image is Gaussian filtered to remove noise from the image. After de-noising the image, local maxima of gradients within the domains that are to be separated and their surroundings are found. The Gaussian filter is defined as

$$G(i,j) = \left(\frac{1}{2\pi\sigma^2}\right) e^{-[(i-k)^2+(j-k)^2]/2\sigma^2},$$

where $\sigma$ is the standard deviation of the Gaussian filter and $k$ the size of the kernel. A standard deviation of 3 pixels combined with a kernel size of 3 pixels performed best for the application of the Gaussian filter. Eventually Canny edge detection is applied on the resulting image.

From now on every closed contour that has been detected is considered as an independent object. The first step after contour separation is to delete large domains. The remaining objects are fitted to ellipses. We chose ellipses as fit objects because geometric properties like orientation and aspect ratio of the fitted pattern are easily retrieved by the location of the main axis and the value of numerical eccentricity of the ellipse. The fit is based on least squares fitting and assumes that all pixels within an object belong to one ellipse. Further details are provided in Fitzgibbon et al. (1999) and in Mulchrone and Choudhury (2004).

The next step includes the a priori assumption that the target patterns occur no farther than roughly 250 km onshore or offshore. For this purpose we define a coastline by applying Canny edge detection to a land–sea mask of the same resolution as the data. The coastal area is defined by adding pixels to the neighborhood of the coastal edges. This technique can be seen as the reversal of box counting that is applied to define fractal geometries such as coastlines (Block et al. 1990). Finally, all domains that have less than 80% of their total area within the defined area are deleted from the image.

The remaining patterns are tested for alignment with the coastline. Alignment can be tested only if each fit ellipse has an assignable major and minor axis. This is guaranteed if the numerical eccentricity of the fit ellipse is considerably larger than 0. Therefore, all domains with an eccentricity of the fit ellipse lower than a certain threshold are deleted. The alignment with the coastlines of the remaining domains is tested in three ways. First, a narrow coastal area ($\sim 2 \times$ resolution) is defined by inverse box counting. Domains having more than 90% of their area overlapping with the narrow coastal area are considered as aligned with the coastline. The remaining domains are shifted into the narrow coastal area without changing their orientation. All domains that are again overlapping by at least 90% are marked as detected. For each of the remaining objects, three straight lines, two from the tips and one from the center of the fit ellipse, are drawn orthogonally in both directions of the major axis. The distance from the origin to the next coastline intersection for each straight line is measured. If the standard deviation of the three distances is below a certain percentage of the mean distance, the contour is assumed to be aligned with a coastline. The object is then marked as aligned with the coast if the mean distance of all corresponding straight lines is not greater than 500 km.

All domains that have not been deleted meet all of the heuristics mentioned in section 3 and are finally labeled as coastally influenced rainfall features.

## REFERENCES

Azorin-Molina, C., and D. Chen, 2009: A climatological study of the influence of synoptic-scale flows on sea breeze evolution in the Bay of Alicante (Spain). *Theor. Appl. Climatol.,* **96,** 249–260, doi:10.1007/s00704-008-0028-2.

Block, A., W. von Bloh, and H. J. Schellnhuber, 1990: Efficient box-counting determination of generalized fractal dimensions. *Phys. Rev.,* **42A,** 1869–1874, doi:10.1103/PhysRevA.42.1869.

Canny, J., 1986: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.,* **8,** 679–698, doi:10.1109/TPAMI.1986.4767851.

Collier, J. C., and K. P. Bowman, 2004: Diurnal cycle of tropical precipitation in a general circulation model. *J. Geophys. Res.,* **109,** D17105, doi:10.1029/2004JD004818.

Crosman, E., and J. Horel, 2010: Sea and lake breezes: A review of numerical studies. *Bound.-Layer Meteor.,* **137,** 1–29, doi:10.1007/s10546-010-9517-9.

Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate,* **19,** 4605–4630, doi:10.1175/JCLI3884.1.

Estoque, M. A., 1962: The sea breeze as a function of the prevailing synoptic situation. *J. Atmos. Sci.,* **19,** 244–250, doi:10.1175/1520-0469(1962)019<0244:TSBAAF>2.0.CO;2.

Fitzgibbon, A., M. Pilu, and R. B. Fisher, 1999: Direct least square fitting of ellipses. *IEEE Trans. Pattern Anal. Mach. Intell.,* **21,** 476–480, doi:10.1109/34.765658.

Folkins, I., T. Mitovski, and J. R. Pierce, 2014: A simple way to improve the diurnal cycle in convective rainfall over land in climate models. *J. Geophys. Res. Atmos.,* **119,** 2113–2130, doi:10.1002/2013JD020149.

Frizzola, J. A., and E. L. Fisher, 1963: A series of sea breeze observations in the New York City area. *J. Appl. Meteor.,* **2,** 722–739, doi:10.1175/1520-0450(1963)002<0722:ASOSBO>2.0.CO;2.

Gianotti, R. L., D. Zhang, and E. A. B. Eltahir, 2012: Assessment of the Regional Climate Model version 3 over the Maritime Continent using different cumulus parameterization and land surface schemes. *J. Climate,* **25,** 638–656, doi:10.1175/JCLI-D-11-00025.1.

Gilliam, R., S. Raman, and D. Niyogi, 2004: Observational and numerical study on the influence of large-scale flow direction and coastline shape on sea-breeze evolution. *Bound.-Layer Meteor.,* **111,** 275–300, doi:10.1023/B:BOUN.0000016494.99539.5a.

Haurwitz, B., 1947: Comments on the sea-breeze circulation. *J. Meteor.,* **4,** 1–8, doi:10.1175/1520-0469(1947)004<0001:COTSBC>2.0.CO;2.

Hill, C. M., P. J. Fitzpatrick, J. H. Corbin, Y. H. Lau, and S. K. Bhate, 2010: Summertime precipitation regimes associated with the sea breeze and land breeze in southern Mississippi and eastern Louisiana. *Wea. Forecasting,* **25,** 1755–1779, doi:10.1175/2010WAF2222340.1.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.,* **5,** 487–503, doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

Keenan, T. D., and R. E. Carbone, 2008: Propagation and diurnal evolution of warm season cloudiness in the Australian and Maritime Continent region. *Mon. Wea. Rev.,* **136,** 973–994, doi:10.1175/2007MWR2152.1.

Mahrer, Y., and R. A. Pielke, 1977: The effects of topography on sea and land breezes in a two-dimensional numerical model. *Mon. Wea. Rev.,* **105,** 1151–1162, doi:10.1175/1520-0493(1977)105<1151:TEOTOS>2.0.CO;2.

Mak, M. K., and J. E. Walsh, 1976: On the relative intensities of sea and land breezes. *J. Atmos. Sci.,* **33,** 242–251, doi:10.1175/1520-0469(1976)033<0242:OTRIOS>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. *Mon. Wea. Rev.,* **131,** 830–844, doi:10.1175/1520-0493(2003)131<0830:DPORIN>2.0.CO;2.

McPherson, R. D., 1970: A numerical study of the effect of a coastal irregularity on the sea breeze. *J. Appl. Meteor.,* **9,** 767–777, doi:10.1175/1520-0450(1970)009<0767:ANSOTE>2.0.CO;2.

Mori, S., and Coauthors, 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, doi:10.1175/1520-0493(2004)132<2021:DLRPMO>2.0.CO;2.

Mulchrone, K. F., and K. R. Choudhury, 2004: Fitting an ellipse to an arbitrary shape: Implications for strain analysis. *J. Struc. Geol.,* **26,** 143–153, doi:10.1016/S0191-8141(03)00093-2.

Nevatia, R., and K. E. Price, 1982: Locating structures in aerial images. *IEEE Trans. Pattern Anal. Mach. Intell.,* **4,** 476–484, doi:10.1109/TPAMI.1982.4767291.

Pielke, R. A., 2002: Examples of mesoscale models. *Mesoscale Meteorological Modeling,* R. A. Pielke, Ed., International Geophysics Series, Vol. 78, Academic Press, 472–530.

Qian, J.-H., 2008: Why precipitation is mostly concentrated over islands in the Maritime Continent. *J. Atmos. Sci.,* **65,** 1428–1441, doi:10.1175/2007JAS2422.1.

——, A. W. Robertson, and V. Moron, 2013: Diurnal cycle in different weather regimes and rainfall variability over Borneo associated with ENSO. *J. Climate,* **26,** 1772–1790, doi:10.1175/JCLI-D-12-00178.1.

Qian, T., C. C. Epifanio, and F. Zhang, 2012: Topographic effects on the tropical land and sea breeze. *J. Atmos. Sci.,* **69,** 130–149, doi:10.1175/JAS-D-11-011.1.

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. *J. Climate,* **24,** 325–348, doi:10.1175/2010JCLI3673.1.

——, and ——, 2013: Influence of ENSO on the diurnal cycle of rainfall over the Maritime Continent and Australia. *J. Climate,* **26,** 1304–1321, doi:10.1175/JCLI-D-12-00124.1.

Sato, T., H. Miura, M. Satoh, Y. N. Takayabu, and Y. Wang, 2009: Diurnal cycle of precipitation in the tropics simulated in a global cloud-resolving model. *J. Climate,* **22,** 4809–4826, doi:10.1175/2009JCLI2890.1.

Simpson, J., N. Westcott, R. Clerman, and R. Pielke, 1980: On cumulus mergers. *Arch. Meteor. Geophys. Bioklimatol.,* **29A,** 1–40, doi:10.1007/BF02247731.

——, T. Keenan, B. Ferrier, R. Simpson, and G. Holland, 1993: Cumulus mergers in the Maritime Continent region. *Meteor. Atmos. Phys.,* **51,** 73–99, doi:10.1007/BF01080881.

Slingo, A., K. I. Hodges, and G. J. Robinson, 2004: Simulation of the diurnal cycle in a climate model and its evaluation using data from *Meteosat* 7. *Quart. J. Roy. Meteor. Soc.,* **130,** 1449–1467, doi:10.1256/qj.03.165.

Small, C., and R. J. Nicholls, 2003: A global analysis of human settlement in coastal zones. *J. Coastal Res.,* **19,** 584–599.

Stephens, G. L., and Coauthors, 2010: Dreary state of precipitation in global models. *J. Geophys. Res.,* **115,** D24211, doi:10.1029/2010JD014532.

Sun, Y., S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate,* **19,** 916–934, doi:10.1175/JCLI3672.1.

Tijm, A. B. C., A. A. M. Holtslag, and A. J. van Delden, 1999: Observations and modeling of the sea breeze with the return current. *Mon. Wea. Rev.,* **127,** 625–640, doi:10.1175/1520-0493(1999)127<0625:OAMOTS>2.0.CO;2.

Wapler, K., and T. Lane, 2012: A case of offshore convective initiation by interacting land breezes near Darwin, Australia. *Meteor. Atmos. Phys.,* **115,** 123–137, doi:10.1007/s00703-011-0180-6.

Warner, T. T., B. E. Mapes, and M. Xu, 2003: Diurnal patterns of rainfall in northwestern South America. Part II: Model simulations. *Mon. Wea. Rev.,* **131,** 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. *Mon. Wea. Rev.,* **132,** 1917–1932, doi:10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Yang, G.-Y., and J. Slingo, 2001: The diurnal cycle in the tropics. *Mon. Wea. Rev.,* **129,** 784–801, doi:10.1175/1520-0493(2001)129<0784:TDCITT>2.0.CO;2.

Zhuo, H., P. Zhao, and T. Zhou, 2014: Diurnal cycle of summer rainfall in Shandong of eastern China. *Int. J. Climatol.,* **34,** 742–750, doi:10.1002/joc.3718.