Validation of GPM IMERG extreme precipitation in the Maritime Continent by station and radar data

Nicolas A. Da Silva[1], Benjamin G. M. Webber[1], Adrian J. Matthews[2], Matthew M. Feist[3], Thorwald H.M. Stein[3], Christopher E. Holloway [3], Muhammad Firdaus Ammar Bin Abdullah[4]

[1]Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich, Norfolk, NR4 7TJ, United Kingdom
[2]Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences and School of Mathematics, University of East Anglia, Norwich, Norfolk, NR4 7TJ, United Kingdom
[3]Department of Meteorology, University of Reading, Reading, Berkshire, RG6 6ET, United Kingdom
[4]Malaysian Meteorological Department, Petaling Jaya, Malaysia

Key Points:

- Spatial sampling error severely impacts the comparison of IMERG data with point-wise precipitation
- The 95$^{\text{th}}$ percentile is the optimum choice for comparison of NWP precipitation extremes against IMERG
- Above the 95$^{\text{th}}$ percentile IMERG overestimates daily precipitation rates compared with rain gauges

Corresponding author: Benjamin Webber, `b.webber@uea.ac.uk`

Abstract

The Maritime Continent (MC) is a region subject to high impact weather (HIW) events, which are still poorly predicted by numerical weather prediction (NWP) models. To improve predictability of such events, NWP need to be evaluated against accurate measures of extreme precipitation across the whole MC. With its global spatial coverage at high spatio-temporal resolution, the Global Precipitation Measurement (GPM) dataset is a suitable candidate. Here we evaluate extreme precipitation in the Integrated Multi-Satellite Retrieval for GPM (IMERG) V06B product against station data from the Global Historical Climatology Network (GHCN) in Malaysia and the Philippines. We find that the high intra-grid spatial variability of precipitation extremes results in large spatial sampling errors when each IMERG gridbox is compared with individual co-located precipitation measurements, a result that may explain discrepancies found in earlier studies in the MC. Overall, IMERG daily precipitation is similar to station precipitation between the 85$^{\mathrm{th}}$ and 95$^{\mathrm{th}}$ percentile, but tends to overestimate above the 95$^{\mathrm{th}}$. IMERG data were also compared with radar data in western Peninsular Malaysia for sub-daily timescales. Allowing for uncertainties in radar data, the analysis suggests that the 95$^{\mathrm{th}}$ percentile is still suitable for NWP evaluation of extreme sub-daily precipitation, but that the rainfall rates diverge at higher percentiles. Hence, our overall recommendation is that the 95$^{\mathrm{th}}$ percentile be used to evaluate NWP forecasts of HIW on daily and sub-daily time scales against IMERG data, but that higher percentiles (i.e., more extreme precipitation) be treated with caution.

Plain Language Summary

Extreme rainfall is a major hazard in many parts of the tropics, leading to flooding and social and economic impacts. Accurate weather forecasting of extreme rainfall events is needed by national and regional government planners and disaster relief organisations, as well as by agriculture and industry. The skill of weather forecast computer models needs to be tested against a reliable data set of observed rainfall, so that scientists can improve the models to give better forecasts of extreme rainfall. Observed rainfall data sets need to be evaluated prior to their use for testing models. Here, we evaluate the reliability of the IMERG rainfall data set for this purpose. IMERG is based on satellite and rain gauge measurements of rainfall from across the planet. We focus on the area known as the western Maritime Continent. After comparing IMERG rainfall against local measurements of rainfall from weather radar in Malaysia, and weather station data across the region, the recommendation is that IMERG can be used as a reliable measure of fairly extreme rainfall (the top 5% of daily rainfall totals), but tends to overestimate and therefore should be used with caution for very extreme rainfall (the top 1% of daily rainfall totals).

1 Introduction

Precipitation has a considerable impact on human society. In excess, precipitation produces devastating floods that have a high destructive capacity for both infrastructure and human lives. Conversely, a lack of precipitation can lead to drought, lack of drinking water and crop failure. Being one of the wettest places on Earth, the Maritime Continent (MC) separates the Indian Ocean from the Pacific and encompasses the countries of Indonesia, Malaysia and the Philippines, among others. This region experiences significant extreme precipitation (Hai et al., 2017; Warlina & Guinensa, 2019), which, combined with the high vulnerability of the local population (Takama et al., 2017; Karki, 2019; Abd Majid et al., 2019; Cabrera & Lee, 2020), can lead to severe consequences. Accurate prediction of extreme precipitation in the MC is therefore of crucial importance for society. Numerical weather prediction (NWP) models still struggle to correctly predict such extreme events in the MC. Progress in the prediction of extreme precipitation needs accurate evaluations of NWP. This requires the use of an accurate observation system of actual precipitation.

Current observations of precipitation are made through the use of station gauge networks, ground-based radars, and satellite measurements. While prone to errors due to evaporation and wind effects (Lorenz & Kunstmann, 2012; Maggioni et al., 2016; Du et al., 2018), gauge measurements are expected to be more accurate as they provide a direct measure of precipitation (Sun et al., 2018). However, gauge measurements are limited by their localised (point) spatial nature (Kidd et al., 2017), which result in sampling errors when interpolated onto larger areas (Lorenz & Kunstmann, 2012; Rana et al., 2015). Ground-based radars can significantly increase the extent of precipitation observations, and still retain a high spatial resolution. However, because of the indirect way in which they measure precipitation, ground-based radar are affected by errors from contamination, attenuation of signal, and the uncertainty associated with the reflectivity–rain-rate (Z–R) relationship (Iguchi et al., 2009; Berne & Krajewski, 2013; Maggioni et al., 2016). Furthermore, the MC is poorly covered by ground-based measurements of precipitation (Kidd et al., 2017). Hence, NWP evaluation in the MC particularly relies on satellite precipitation measurements, with their potentially global spatial coverage. Although errors in estimation methods still remain (Derin et al., 2016; Camici et al., 2018), the use of precipitation data from satellites has increased and has enabled new applications (Kucera et al., 2013; Kirschbaum et al., 2017).

To benefit from the advantages of both satellite (higher spatial coverage) and gauge measurements (higher accuracy), considerable effort has been invested in the development of mixed gauge–satellite precipitation datasets (Huffman et al., 1995; Xie & Arkin, 1997; Huffman et al., 2007; Adler et al., 2018; Huffman et al., 2019). The Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) is one such dataset. The IMERG precipitation dataset was built with the use of over ten satellites, including the GPM Core Observatory satellite launched in 2014. It carries the Ku- and Ka-band Dual-frequency Precipitation Radar (DPR) and the GPM Microwave Imager (GMI) sensors, two of the most sophisticated satellite precipitation sensors currently in space (Skofronick-Jackson et al., 2018). These instruments are complemented by both Passive Micro-Wave (PMW) and Infra-Red (IR) sensors on board the IMERG satellite constellation.

The IMERG product has been evaluated in many locations globally (Sharifi et al., 2016; Prakash et al., 2016; Omranian & Sharif, 2018; Fang et al., 2019; Kim et al., 2017; Dezfuli et al., 2017; Mayor et al., 2017; Navarro et al., 2019), and is generally an improvement with respect to its predecessors. Thus, IMERG is a suitable candidate for the systematic evaluation of NWP extreme precipitation in the MC. However, IMERG is not exempt from errors, some of which are already well documented (J. Tan et al., 2016; Oliveira et al., 2016; O et al., 2017; O & Kirstetter, 2018; J. Tan et al., 2019). The IMERG precipitation estimates were shown to better match gauge data at the monthly timescale than at the daily/sub-daily timescales (M. L. Tan & Duan, 2017; Yuda et al., 2020).

Although accurate at measuring mean precipitation rates, such global satellite precipitation products often show deficiencies in their representation of extreme precipitation, and their accuracy may be regionally and climatically dependent (Rajulapati et al., 2020). The IMERG product does not seem to be an exception; it underestimates extreme precipitation over Mexico (Mayor et al., 2017), the eastern coast of the United States (J. Tan et al., 2016), Singapore (M. L. Tan & Duan, 2017), and Austria (O et al., 2017), and overestimates extreme precipitation in the central Amazon (Oliveira et al., 2016), the Tibetan plateau (Zhang et al., 2018), and the Netherlands (Gaona et al., 2016). Previous analysis of IMERG performance over the MC (M. L. Tan & Duan, 2017; M. L. Tan & Santo, 2018; J. Tan et al., 2019; Yuda et al., 2020; Liu et al., 2020) found that IMERG underestimates extreme precipitation and performs better during the wettest season. However, these studies were subject to potentially large spatial sampling errors, i.e., errors incurred when interpolating gauge precipitation data onto the IMERG grid. By degrading the same precipitation product onto different spatio-temporal resolutions, Behrangi and Wen (2017) showed that these errors can be large, especially over land areas. Similarly, Tian et al. (2018) and Tang et al. (2018) found that rain gauge density has a large impact on IMERG skill metrics over China.

Previous IMERG evaluation studies in the MC were done over relatively short periods of 1–2 years. By definition, extreme precipitation is very infrequent, hence small sample sizes may have a detrimental effect here. Consequently, these studies do not provide a practical range of precipitation from which IMERG can be used with the aim of evaluating extreme precipitation events simulated by NWP in the MC.

Therefore, the objective of the present study is to reassess the performance of IMERG in the detection of extreme precipitation over the MC, with an estimation of spatial sampling error, and to provide practical information for use in NWP evaluation. For this purpose, the IMERG V06B dataset is evaluated against the Global Historical Climatology Network (GHCN) gauge dataset over Malaysia and the Philippines, and against a ground-based weather radar dataset from western Peninsular Malaysia. Section 2 describes the precipitation datasets used in this study. Section 3 presents an evaluation of IMERG in the MC. Finally, Section 4 describes key findings and practical guidance for the use of IMERG in NWP evaluation.

2 Data

2.1 IMERG data

The main analysis in this study is based on the Integrated Multi-Satellite Retrievals (IMERG) product, version V06B, from the Global Precipitation Measurement (GPM) project (Huffman et al., 2019). This product is based on measurements from a constellation of satellites, equipped with Passive Micro-Wave (PMW) and geo-infrared (IR) sensors. The PMW measurements give more accurate direct estimations of precipitation rate but have limited spatial and temporal coverage. Meanwhile, the IR measurements only measure precipitation indirectly, but have almost complete spatial and temporal coverage.

The PMW precipitation estimates are first converted from brightness temperature to precipitation rate following the Goddard profiling algorithm (GPROF) (Kummerow et al., 2015) or the Precipitation Retrieval and Profiling Scheme (Kidd et al., 2018). Among PMW satellites, the GPM core observatory is considered to carry the most advanced instruments for precipitation detection (Skofronick-Jackson et al., 2018). It was launched in February 2014 and is the successor to the Tropical Rainfall Measuring Mission (TRMM, Huffman et al. (2007)) satellite, which was launched in 1997. As well as providing accurate precipitation measurements for the IMERG product, the TRMM satellite and the GPM core observatory serve for the inter-calibration of the whole IMERG PMW satellite constellation, in their respective eras. Several studies have identified improvements of precipitation estimates by IMERG relative to its predecessors in South East Asia (Prakash et al., 2016; Kim et al., 2017; M. L. Tan & Duan, 2017; F. Xu et al., 2019).

Prior to inter-calibration, the TRMM and GPM core observatory estimates are seasonally corrected over land areas by the climatological values from the Global Precipitation Climatology Project (GPCP) satellite-gauge product (Adler et al., 2018). The PMW inter-calibration is achieved through quantile matching, using a method similar to Miller (1972); Krajewski and Smith (1991). The IR data, which essentially measure cloud top features rather than precipitation directly, are trained and calibrated against the PMW estimates using an artificial neural network cloud classification system (PERSIANN-CSS; Nguyen et al. (2018)).

All precipitation estimates are gridded on to a $0.1^\circ \times 0.1^\circ$ longitude–latitude spatial grid. A Kalman smoother is then used to combine all precipitation estimates into a single half-hourly estimate (Joyce & Xie, 2011). In this step, the closest PMW estimates forward and backward in time from the analysis time of the half-hourly window are propagated to the analysis time using precipitable water vapor motion vectors from the Goddard Earth Observing System Forward Processing (IMERG early and late runs; GEOS FP; Keller et al. (2021)) or the Modern-Era Retrospective Analysis for Research and Applications, version 2 (IMERG final run; MERRA-2; Gelaro et al. (2017)). A weighted average of the two resultant estimates is then performed. The IR data are used only if the nearest PMW measurement is more than 30 minutes from the target time. In this, the IR estimates are incorporated into a Kalman filter in the form of an observation correcting the PMW "forecast". The resulting half-hourly estimates over land are then multiplied by the ratio between the Global Precipitation Climatology Centre (GPCC) (Schneider et al., 2008) monthly gauge estimate with the monthly sum of half-hourly estimates derived in the early steps of the IMERG algorithm. This step is only performed in the final version of the product, which is used in the present study. The IMERG product is thus a multi satellite-gauge precipitation dataset for which data are provided with a 30-minute time interval on a global $0.1^\circ \times 0.1^\circ$ grid.

The diurnal cycle of precipitation is reasonably well captured by IMERG, when compared to rain gauge (J. Tan et al., 2019; Li et al., 2018; Mayor et al., 2017; O & Kirstetter, 2018; Tang et al., 2016; Zhang et al., 2018) or ground-based radar precipitation estimates (Oliveira et al., 2016), although a phase delay of about 40 minutes was found in the presence of frozen hydrometeors aloft (O et al., 2017; O & Kirstetter, 2018; J. Tan et al., 2019; You et al., 2019). Potential sources of IMERG errors were attributed to the precision of the instruments on board the satellite constellation (J. Tan et al., 2016; Li et al., 2018). IMERG retrievals that only used IR measurements were found to be the least accurate, because precipitation is measured indirectly from cloud-top brightness temperatures. However, PMW sensors tend to underestimate warm cloud precipitation (Dinku et al., 2007; Shige et al., 2013), which can affect the performance of IMERG (O & Kirstetter, 2018). The IMERG algorithm itself was sometimes identified as a source of error, notably through its morphing and GPROF precipitation retrieval schemes (J. Tan et al., 2016; Oliveira et al., 2016).

In this study, 19 years of the IMERG precipitation dataset from 1 January 2001 to 31 December 2019 over Malaysia and the Philippines (Fig. 1) were used. When IMERG data were compared to radar data, IMERG accumulations were calculated only using data from times at which radar data were also available.

2.2 GHCN station data

The Global Historical Climatology Network (GHCN) dataset comprises several meteorological variables measured by surface weather stations across the Earth (Menne, Durre, Korzeniewski, et al., 2012; Menne, Durre, Vose, et al., 2012). Data are available at daily (UTC) time resolution, and have undergone a common suite of quality assurance reviews (Durre et al., 2010). In the present study, only the daily mean precipitation data from Malaysia and the Philippines were used to evaluate the IMERG data. First, the gauge time series were truncated to the IMERG period examined (2001–2019) to ensure time coherence between both datasets. Then, only GHCN stations having at least 1000 days of data within this period were selected for analysis. The GHCN dataset also included weather station time series from Indonesia but the lengths of these time series did not satisfy the latter criteria. The exact locations of the gauges used are shown in Fig. 1. The gauges are spread over large areas with different climate characteristics. Previous studies found that IMERG may have variable skill, depending on regional characteristics within the Maritime Continent (M. L. Tan & Santo, 2018). Hence, six groups of weather stations were defined in the following regions (red markers in Fig. 1): Western Peninsular Malaysia (5 stations); Eastern Peninsular Malaysia (3 stations); Northwest Borneo (6 stations); Western Philippines (except mountain regions, 6 stations); Eastern Philippines (11 stations); Philippines mountain region (1 station).

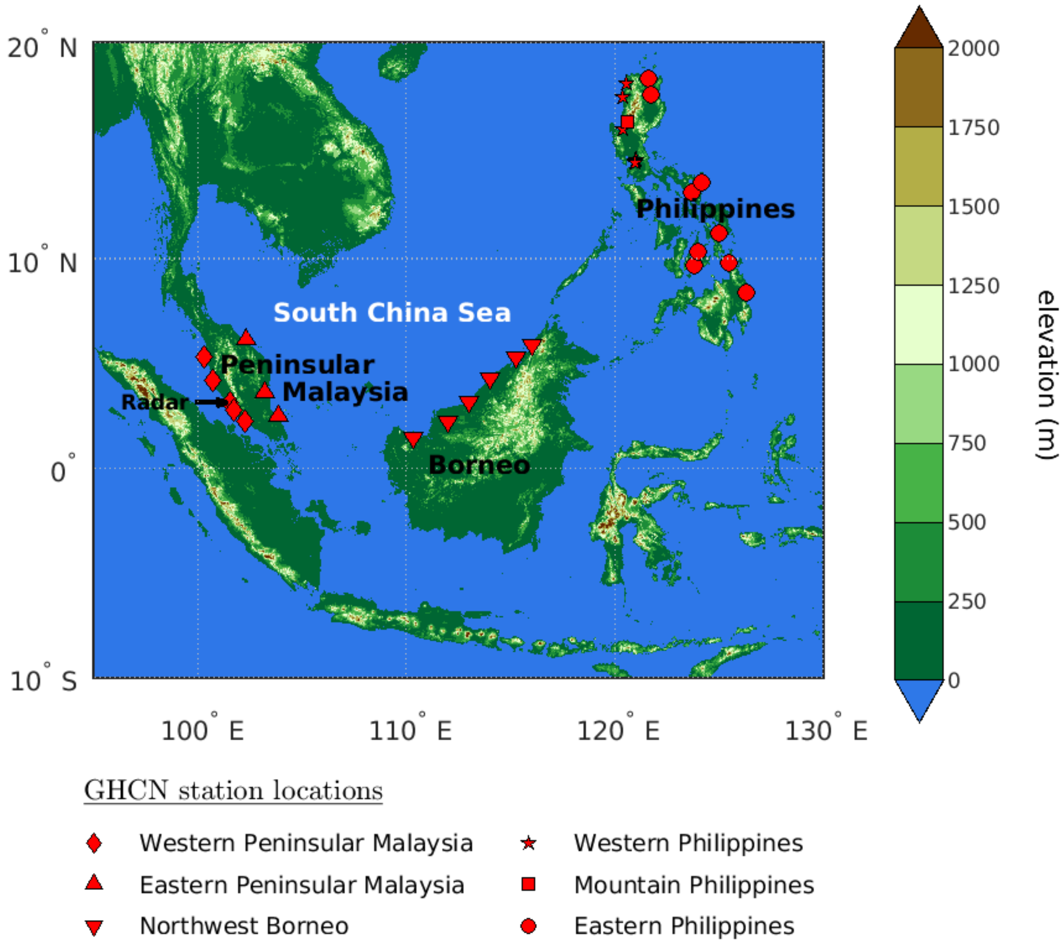

Figure 1. Topography of the Maritime Continent (shaded). Locations of GHCN stations are shown by red markers: diamonds for western Peninsular Malaysia; upward triangles for eastern Peninsular Malaysia; downward triangles for northwest Borneo; stars for western Philippines; circles for eastern Philippines; a square for the mountain Philippines station.

2.3 Radar data

Data from an S-band Doppler weather radar at Subang, Kuala Lumpur (101.559°E, 3.145°N), operated by MetMalaysia, were also used to evaluate the IMERG data. There are 89 days of radar data in a period spanning 94 days, from 11 January to 15 April 2019. The radar measurements were calibrated first using a relative calibration against clutter points and second using the DPR aboard GPM, following Warren et al (2018) and Louf et al. (2019). Following calibration, the radar data were interpolated on to a Cartesian grid at 2-km height above the radar location, from which precipitation values were retrieved using the Weather Surveillance Radar (WSR) Z–R relationship (Fulton et al., 1998). The WSR Z-R relationship is known to give correct estimations for convective precipitation. The Marshall–Palmer (Marshall and Palmer, 1947) and the Rosenfeld (Rosenfeld et al.,1993) Z–R relationships, which perform well for stratiform and tropical precipitation (respectively), were also tested and taken into account in the study in the form of uncertainties.

Instantaneous precipitation values are provided every 10 minutes, at 0000, 0010, 0020, . . . 2350, each day. The spatial resolution of the radar data is 0.0045°, or approximately 400 m. A spatial bilinear interpolation was performed on the radar data, to map

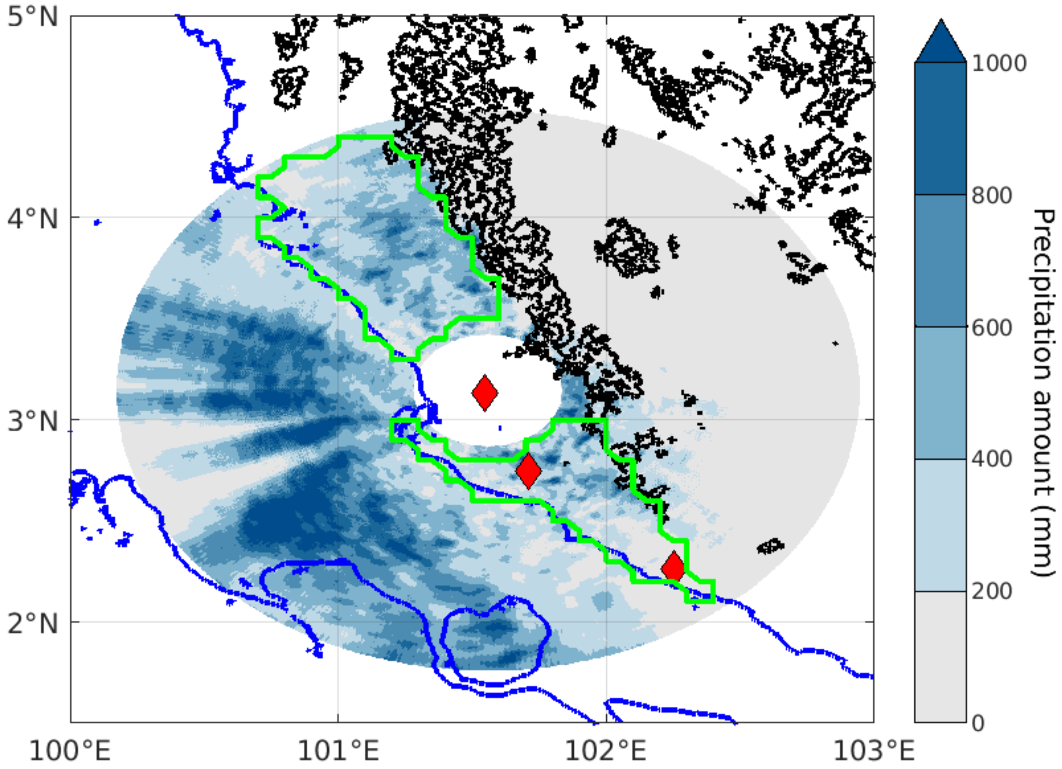

Figure 2. Total accumulated precipitation from the Subang radar, from 11 January to 15 April 2019. The locations of the GHCN stations are shown by red diamonds. Topography is line contoured, with an interval of 500 m (blue for the 0 meter level and black for the other levels). The green line delimits the low-land grid points used in this study.

it from its original grid to the 0.1° IMERG grid, for comparison. Both the 0.0045° and the 0.1° radar data were used in this study, the 0.0045° radar data being used as an estimate of pointwise precipitation in order to quantify the spatial sampling error.

The Subang radar is located on the coastal plain of western peninsular Malaysia, with the prominent Titiwangsa mountain range to the east (Fig. 2). The mountains clearly block the radar signal to the east, as evidenced by the near zero accumulations in this region. Hence, all radar grid points over and to the east of the Titiwangsa mountains were removed from the analysis.

The IMERG data are available every 30 minutes, at 0015, 0045, 0115, . . . , 2345, each day. When there is no passive-microwave measurement in the corresponding 30 minutes windows, the IMERG values are calculated as an average of the closest previous passive-microwave measurement advected forward in time by MERRA-2 motion vectors, and the closest following passive-microwave measurement advected backward in time by MERRA-2 motion vectors. Infra-red precipitation data are also incorporated in the calculation when no passive microwave measurements are available within ± 30 minutes of the time window. This effectively gives an approximately 25-minute mean precipitation value (O et al., 2017). Hence, for direct comparison of "instantaneous" radar and IMERG data, the two closest instantaneous radar values (backward and forward in time) from the IMERG output time were averaged. For example, the IMERG precipitation value at 1415 was compared with the average of the instantaneous 1410 and 1420 radar precipitation values. For the sake of simplicity, this average is still referred to as "instantaneous" in this study. While such an averaging procedure is the best estimate of precipitation intensity between two radar output time steps, it tends to underestimate extremes of instantaneous precipitation (and conversely, overestimate low precipitation). This averaging procedure was only carried out for the comparison of "instantaneous" precipitation values.

Rainfall accumulations were also calculated from the 10-minute instantaneous radar data, for periods of 30 minutes, and 1, 3, 6, 12, and 24 hours. A weighted average was calculated from all instantaneous precipitation measures within the period. Each 10-minute instantaneous radar scan was interpreted as the representation of averaged precipitation over a 10-minute window centered on the nominal time and the weightings were chosen accordingly.

There was a significant fraction of missing radar data (13%). Gaps in the radar time series were filled using linear time interpolation before the accumulations were calculated. To reduce potential errors from this interpolation, all accumulation periods for which more than half of the data were missing were discarded from the analysis. This restriction does not completely avoid errors, especially for the longest accumulation periods. A discussion of these errors is provided in Section 3.

2.4 Topography data

The General Bathymetric Chart of the Oceans (GEBCO) topography data set was used to distinguish between sea, lowland and mountain regions. It was regridded from its native 30 arc-second resolution to the coarser $0.1^\circ \times 0.1^\circ$ longitude–latitude IMERG grid (Fig. 1).

3 Validation of IMERG precipitation data over the Maritime Continent

3.1 Comparison of IMERG with GHCN station data

First, IMERG precipitation is evaluated against the GHCN dataset over the six regions of interest: Western Peninsular Malaysia, Eastern Peninsular Malaysia, Northwestern Borneo, Northwestern Philippines, Eastern Philippines, and a high elevation (mountain) station located in the Western Philippines. The correlation coefficient, root mean square error (RMSE), and relative bias were calculated for daily, weekly and monthly precipitation accumulations (Table 1). For the relative bias, we first calculated the bias and then we divided it by total accumulated precipitation over the time period (thus this metric does not vary with time scale). All of these statistics were initially calculated for each station (using the time series of IMERG precipitation from the nearest grid point, on the $0.1 \times 0.1^\circ$ IMERG grid) and then averaged over the region.

Correlation coefficients of daily precipitation values range from 0.5 in Western Peninsular Malaysia to 0.74 in Eastern Peninsular Malaysia, while correlation coefficients of monthly precipitation values are typically above 0.8. In each region, the correlation coefficient increases with increasing accumulation time, and RMSE decreases with increasing accumulation time. This increase in performance of IMERG at the seasonal time scale compared with the daily time scale was also observed in Singapore (M. L. Tan & Duan, 2017), Bali (Yuda et al., 2020), and the USA (J. Tan et al., 2017). Our analysis of daily correlation coefficients and RMSEs in Malaysia confirms and extends the results of M. L. Tan and Santo (2018) who used an earlier version of IMERG and a shorter time period.

Although the daily correlation coefficient values reflect a moderate–to–good representation of IMERG in capturing the day–to–day variability of precipitation, the high daily RMSE values in every location emphasise the magnitude of errors in IMERG precipitation intensity, ranging from 13.6 mm day^{-1} in Western Peninsular Malaysia up to 33.2 mm day^{-1} at the sole mountain station in the Western Phillipines. The relative bias tends to be positive for low-level land locations, but IMERG displays a substantial negative bias at the sole mountain station of −28%. With only one mountain station we cannot conclude that this bias is a consistent feature, but this result is consistent with previous findings that passive microwave sensors may underestimate warm orographic rain because they use ice loads for their detection of precipitation (Dinku et al., 2007; Derin et al., 2016; Kim et al., 2017; R. Xu et al., 2017; O & Kirstetter, 2018; Navarro et al., 2019). It is also worth noting that IMERG does not explicitely account for orographic enhancement, unlike Global Satellite Mapping of Precipitation (GSMaP) which should have an improved representation of precipitation over mountainous regions (Yamamoto & Shige, 2015).

These statistics were calculated from the comparison of time series of local GHCN gauge measurements with time series of 0.1° gridded IMERG precipitation (Section 2). We expect that the pointwise precipitation measurements will not be representative of the average precipitation over the relatively large $0.1 \times 0.1°$ (approximately 120 km^2) area covered by the IMERG nearest grid point. This discrepancy is referred to as the spatial sampling error, and is examined quantitatively below.

Table 1. Correlation coefficients, root mean square error (RMSE), and relative bias, of IMERG precipitation versus GHCN precipitation, and (in *italics*) the Subang radar data on the $0.1 \times 0.1°$ IMERG grid vs the radar data on its native grid, for daily, weekly, and monthly accumulation times. The relative bias does not vary with timescale.

Location	Duration	Correlation coefficient	RMSE (mm day^{-1})	Relative bias (%)
Western Peninsular Malaysia	1 day	0.50	13.6	+15.9
	7 days	0.63	5.3	
	30 days	0.74	2.7	
Radar (vs itself)	*1 day*	*0.72*	*9.16*	*+11.4*
Eastern Peninsular Malaysia	1 day	0.74	14.4	+2.2
	7 days	0.88	5.59	
	30 days	0.94	2.73	
Northwestern Borneo	1 day	0.57	18.1	+12.7
	7 days	0.69	7.23	
	30 days	0.82	3.48	
Northwestern Philippines	1 day	0.63	22.6	+16.5
	7 days	0.78	9.24	
	30 days	0.85	5.04	
Eastern Philippines	1 day	0.62	19.4	+1.3
	7 days	0.73	7.85	
	30 days	0.83	3.92	
Mountain Western Philippines	1 day	0.56	33.2	−28.0
	7 days	0.73	15.1	
	30 days	0.83	8.69	

3.2 Spatial sampling error between IMERG and GHCN precipitation

Several studies evaluated the uncertainties related to the sampling of precipitation measurements when estimating areal precipitation (Villarini et al., 2008; Behrangi & Wen, 2017; Tian et al., 2018; Tang et al., 2018). Here, the spatial sampling error is estimated by comparing the native resolution Subang radar precipitation (on a 0.0045° grid) against itself, but regridded onto the coarser 0.1° IMERG grid. The "Radar" row in Table 1 shows the daily correlation coefficient, RMSE, and relative bias from these calculations. These statistics were initially calculated for each radar grid point at native resolution and the nearest 0.1° neighbour, and subsequently averaged over all low-land grid points (delimited by the green lines in Fig. 2).

As the same product is being compared at two different spatial resolutions, the calculated values of correlation coefficient, RMSE and relative bias can be interpreted as the optimum values attainable, given the spatial sampling error between a 0.1° area-averaged precipitation dataset and a (nearly) point-wise precipitation dataset in Western Peninsular Malaysia. The daily radar–radar correlation coefficient is only 0.72, i.e., significantly less than the maximum theoretical value of 1. This is a similar value to that of Tang et al. (2018), who used a high density gauge network in the Ganjiang River basin (South China) to assess the expected sampling error. It shows that the sampling error contributes substantially to reducing the correlation coefficient for the IMERG–GHCN comparison, which has a value of 0.5.

A similar conclusion can be drawn for the RMSE which is 9.2 mm day^{-1} for the radar–radar comparison. Contributions to mean square error (MSE) can be added linearly, whereas those to RMSE cannot. With this in mind, the radar–radar MSE has a value that is 45% of the value of the IMERG–GHCN MSE. Hence, approximately 45% of the IMERG–GHCN MSE can be attributed to the spatial sampling error, with the remainder being a "genuine" physical error between the two systems. Finally, the radar–radar relative bias is +11.4%, compared with +15.9% for the IMERG–GHCN comparison. Hence, approximately two thirds of the IMERG–GHCN relative bias can be accounted for by spatial sampling error, the remainder being again a "genuine" bias between the two different data sets.

It is likely that precipitation extremes contribute disproportionately to the high RMSE values observed in all the regions. We define extreme precipitation days as those on which the precipitation rate exceeds 20 mm day^{-1}, in either IMERG or GHCN (or both). Retaining only extreme precipitation days, we were able to retrieve 86% of the MSE in Western Peninsular Malaysia, confirming that high RMSE values are almost entirely due to discrepancies between IMERG and GHCN measurements on extreme precipitation days.

To investigate the distribution of errors for such events, the probability density function (PDF) of daily precipitation differences between IMERG and the three nearest GHCN stations in the Subang area was calculated, following the method of Holloway et al. (2012). Precipitation bins were defined following a regular logarithmic increase in magnitude from 0.5 mm day^{-1} to 100 mm day^{-1} for both positive and negative differences. The PDF at bin i was calculated using the following formula:

$$P(i) = \frac{n(\Delta pr_i < \Delta Pr < \Delta pr_{i+1})}{N \times (\Delta pr_{i+1} - \Delta pr_i)}\,, \qquad (1)$$

where $n(\Delta pr_i < \Delta Pr < \Delta pr_{i+1})$ designates the number of extreme precipitation days (as defined above) for which the precipitation difference (ΔPr) is within the bin limits set by Δpr_i and Δpr_{i+1}, and N is the total number of extreme precipitation days.

The resulting distribution of IMERG versus GHCN daily extreme precipitation differences is bi-modal with one local maximum near -20 mm day^{-1} and another one near $+20$ mm day^{-1} (solid line in Fig. 3). The maximum near $+20$ mm day^{-1} mostly reflects precipitation events that occurred in the IMERG data but did not occur in the GHCN stations, and vice-versa for the maximum near -20 mm day^{-1}. Notably, such discrepancies are more frequent (note the logarithmic vertical axis in Fig. 3) than events where the difference in precipitation intensity is less than 20 mm day^{-1}. There is also a non-negligible frequency of events for which the differences between IMERG and GHCN daily precipitation are much higher, above 50 mm day^{-1}. These events contribute the most to the RMSE.

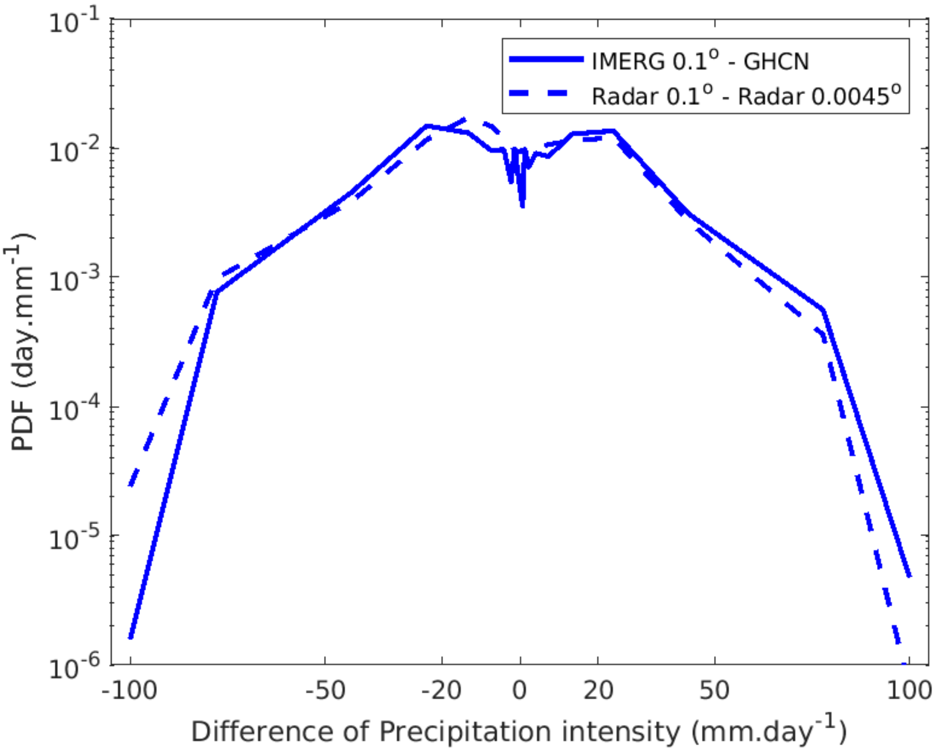

Figure 3. Probability density function (PDF) of the difference between IMERG and GHCN daily precipitation, for the three GHCN weather stations closest to the Subang radar (solid line). The PDF of the difference between daily land precipitation from the Subang radar on its native grid and the radar precipitation averaged over the nearest IMERG grid box (dashed line) is also shown for ease of comparison. Both PDFs are conditioned on extreme daily precipitation, defined as days for which at least one of the products exhibits daily precipitation above 20 mm day^{-1}.

These observations are not reassuring for the use of IMERG in evaluating NWP of extreme precipitation, unless they are the consequence of the spatial sampling error.

To ascertain whether the very large IMERG–GHCN precipitation differences can be attributed to the spatial sampling error, we examine the equivalent PDF for differences between the two different spatial resolutions of the Subang radar data. Each 0.0045° radar daily precipitation data point was subtracted from the daily precipitation estimate of its nearest 0.1° grid point equivalent. The PDF of the radar data (dashed line in Fig. 3) was constructed, retaining only the low land radar grid points for a better comparison with the IMERG–GHCN distribution. The two distributions are very similar. The radar–radar distribution also displays a bimodal shape with local maxima at ±20 mm day^{-1} and a local minimum at 0 mm day^{-1} of the same amplitude as the IMERG–GHCN distribution. This again highlights the large contribution of the spatial sampling error in explaining the large RMSE values, especially for extreme precipitation. This error cannot be ignored for a correct validation of IMERG extreme precipitation in the Maritime Continent, which in turn will serve for NWP evaluation.

3.3 Evaluation of IMERG reliability for extreme precipitation thresholds

Extreme precipitation is often defined in relative terms by using the local statistical distribution of precipitation to calculate a threshold such as the 95$^{\text{th}}$ percentile of precipitation over a given accumulation period. In this context, it is useful to know for which percentiles IMERG gives reliable estimates and those that should be avoided when using IMERG for NWP evaluation.

3.3.1 Subang region of Western Peninsular Malaysia

To evaluate the reliability of IMERG at various percentile thresholds we examine a quantile–quantile plot of IMERG versus GHCN precipitation for the three Malaysian stations closest to the Subang radar for northern winter (October–March; blue line in Fig. 4). The uncertainty of the percentile values is shown by error bars that cover the 95% confidence interval. If there was a perfect correspondence, the blue line would follow the black 1:1 control line.

However, in practice there will be errors due to spatial sampling (Section 3.2) and other sources. The spatial sampling error can be accounted for by the use of radar data at both the 0.1° and native (0.0045°) resolution, giving an expected theoretical quantile–quantile relationship due to spatial sampling alone (solid green control R–R line in Fig. 4). The solid green spatial sampling line does not follow the black 1:1 line. In particular, for extreme precipitation (95$^{\text{th}}$ and higher), the green line is below the 1:1 line, indicating that the (e.g.,) 95$^{\text{th}}$ percentile of radar precipitation on the native high resolution grid is larger than the 95$^{\text{th}}$ percentile of radar precipitation on the coarser IMERG grid. This neatly illustrates that the effect of spatial averaging is to reduce extremes. This effect works in the opposite sense at the lower percentiles. Here, the green line is above the 1:1 line. Hence, a very low rainfall rate (of a given value, e.g., 0.5 mm day^{-1}) is more likely to be observed in low spatial resolution data than in high resolution data, due to spatial aggregation. In summary, we would not expect the IMERG–GHCN quantile–quantile line to follow the black 1:1 line, because of the spatial sampling effect. We might expect it to follow the green R–R control line, however.

The control R–R quantile–quantile (solid green) line was calculated using the radar data with time interpolation to fill the missing values. For a rough estimation of the interpolation uncertainty, the R–R quantile–quantile line was recalculated by substituting missing values with zero (green dashed line in Fig. 4). This lies below the original control R–R line for the whole range of precipitation percentiles with a difference of about 25%.

The radar precipitation product itself presents multiple uncertainties that need to be taken into account in the analysis. In particular, the reflectivity–rainfall (Z–R) relationship is a substantial source of uncertainty. These uncertainties were taken into account in our study by the use of three different Z–R relationships: Marshall–Palmer (Marshall et al., 1947), Rosenfeld (Rosenfeld et al., 1993), and WSR (Fulton et al., 1998). The Marshall–Palmer relationship resulted generally in the weakest rainfall rates, with the Rosenfeld relationship produced the highest rainfall rates, and the WSR relationship led to rainfall rates in between. Solid particles such as hail can also alter the radar signal by amplifying it. The uncertainty related to that was estimated by capping extreme reflectivities at 53 dB. The uncertainty linked to potential hail contamination is non-negligible, although weaker than that linked to the Z-R relationship (not shown). In the following, we use the WSR Z-R relationship without capping as default. The total radar uncertainties were calculated using the minimum and maximum values of the 6 radar estimates emanating from the 3 different radar Z–R relationships with and without cap. The union of the 95% confidence intervals of these minimum and maximum values was taken to account for the percentile uncertainty. The resulted intervals are represented by a shaded grey area and the IMERG 95% confidence intervals are represented by errors bars in Fig. 4).

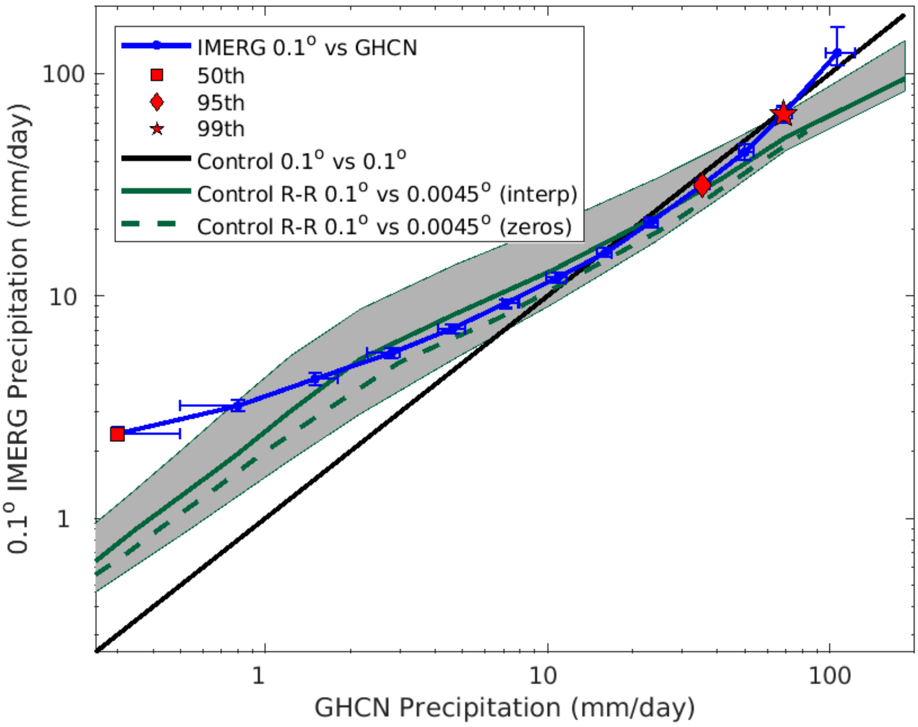

Figure 4. Quantile–quantile diagram of GHCN daily precipitation of the three weather stations at Subang in Fig. 2 versus their nearest neighbour IMERG daily precipitation (blue line). Quantiles are calculated at 5% intervals from the 50$^{\text{th}}$ to the 95$^{\text{th}}$ percentile, then at the 97.5$^{\text{th}}$, 99$^{\text{th}}$, and 99.9$^{\text{th}}$ percentiles. The red markers highlight the 50$^{\text{th}}$ (square), 95$^{\text{th}}$ (diamond) and 99$^{\text{th}}$ (asterisk) percentiles. Error bars show the 95% confidence interval. The black line shows the 1:1 control line. To account for spatial sampling error, the green lines represent the quantile–quantile diagram of Subang radar daily precipitation in low-land areas versus the corresponding (nearest neighbour) daily precipitation of the Subang radar averaged on the IMERG grid, with temporal interpolation over missing values (solid green line; control R–R), and by substituting each instantaneous missing value by zero (green dashed line). The grey shading corresponds to the merged 95% confidence intervals of the green lines.

The blue IMERG–GHCN quantile-quantile line remains within the two green control R–R lines from the 60$^{\text{th}}$ (approximately 1.5 mm day^{-1}) to the 95$^{\text{th}}$ percentile (35 mm day^{-1}), thus displaying a high fidelity in estimating this range of precipitation values. In particular, the 95$^{\text{th}}$ quantile is consistent with the control R–R line (solid green line, using interpolation for missing values) with a relatively low uncertainty of about 20%. The 95$^{\text{th}}$ percentile thus appears to be a reliable choice for evaluation of extreme precipitation in NWP against IMERG.

For percentiles above the 95$^{\text{th}}$, IMERG remains close to GHCN (i.e., close to the black 1:1 control line), but increasingly deviates above the solid green R–R control lines for higher percentiles. Indeed, the 99$^{\text{th}}$ percentile of IMERG is approximately 70 mm day^{-1} against an expected value of about 50 mm day^{-1} (from the green R–R lines). The 99$^{\text{th}}$ percentile of IMERG lies beyond the R–R uncertainty envelope, which means that the overestimation is significant. This reflects a tendency for IMERG to overestimate very extreme precipitation and reach values that tend to be higher than expected for its resolution. It should be noted that IMERG values are corrected by GPCC monthly accumulations (Section 2.1). Given that only one GPCC station was used to make this correction in Malaysia (M. L. Tan & Santo, 2018), it may not be surprising that IMERG precipitation extremes have the same magnitude as station precipitation extremes, and thus overestimate area averaged precipitation extremes. The fact that IMERG remains close to GHCN for these extreme percentiles can be useful for estimating the potential values that extreme precipitation could reach in local areas. However, these high percentiles are not recommended for NWP evaluations against IMERG since NWP are gridded products that usually do not output such local point measures of precipitation.

IMERG tends to overestimate the number of low precipitation rate days (< 1.5 mm day^{-1}, or the 60$^{\text{th}}$ percentile), compared to the solid green R–R line. The overestimation is significant for precipitation below < 0.9 mm day^{-1} where the IMERG line lies above the R–R uncertainty envelope. It should be noted that percentiles below the 50$^{\text{th}}$ were not represented in Fig. 4 because they are all equal to 0 mm day^{-1} for GHCN, and thus do not fit a log–log representation. The number of dry days is lower for IMERG than for GHCN (not shown). Non-meteorological targets such as insects affect the radar retrievals, making it impossible to detect dry days and thus evaluate more accurately if IMERG detects less dry days than it should at its resolution.

3.3.2 Other regions in the Maritime Continent

We now investigate whether these conclusions hold for areas outside of the Subang area (Western Peninsular Malaysia) and for seasons other than northern winter, using six selected areas in Malaysia and in the Philippines (Fig. 5). The absence of a high resolution dataset equivalent to the radar in Subang makes it difficult to precisely determine IMERG performance against the location-specific spatial sampling error in these regions. However, in most regions, the percentile relationships between IMERG and GHCN are very similar to the one observed in Subang: IMERG displays higher precipitation rates than GHCN for percentiles below the 90$^{\text{th}}$ percentile and is similar to GHCN for percentiles above the 90$^{\text{th}}$ percentile. This is the case in Western Peninsular Malaysia, Eastern Peninsular Malaysia, Northwest Borneo, and Western Philippines during northern summer, and Eastern Philippines during northern winter. While the optimal percentile cannot be precisely determined for these regions, the similarity with Subang suggests that the IMERG 95$^{\text{th}}$ percentile is also likely to be a suitable percentile to evaluate NWP extreme precipitation against in these regions. Conversely, higher percentiles are not recommended for NWP evaluation as they will tend to overestimate area averaged precipitation.

The performance of IMERG also shows seasonal dependence (Oliveira et al., 2016; M. L. Tan & Santo, 2018). This is particularly true in both the Western and Eastern Philippines (Fig. 5d,e). Indeed, IMERG displays higher precipitation rates than GHCN for every precipitation percentile during northern winter in the Western Philippines, whereas this is only the case for the lowest precipitation during northern summer (Fig. 5d). Thus, the positive bias for IMERG extreme precipitation is stronger during northern winter in the Western Philippines. This stronger overestimation might be explained by enhanced errors in the IMERG morphing scheme in this region, which is subjected to easterlies during the northern winter, such that most of the precipitating systems (including tropical cyclones) come from the east and cross the Cordillera Central mountain range. The propagation of precipitation in IMERG is based on the motion of total precipitable water vapor fields of the MERRA-2 reanalysis that may underestimate the mountain blocking effect on precipitation due to its relatively coarse spatial resolution. The use of IMERG for NWP evaluation of extreme precipitation in this region during northern winter should therefore be approached with caution.

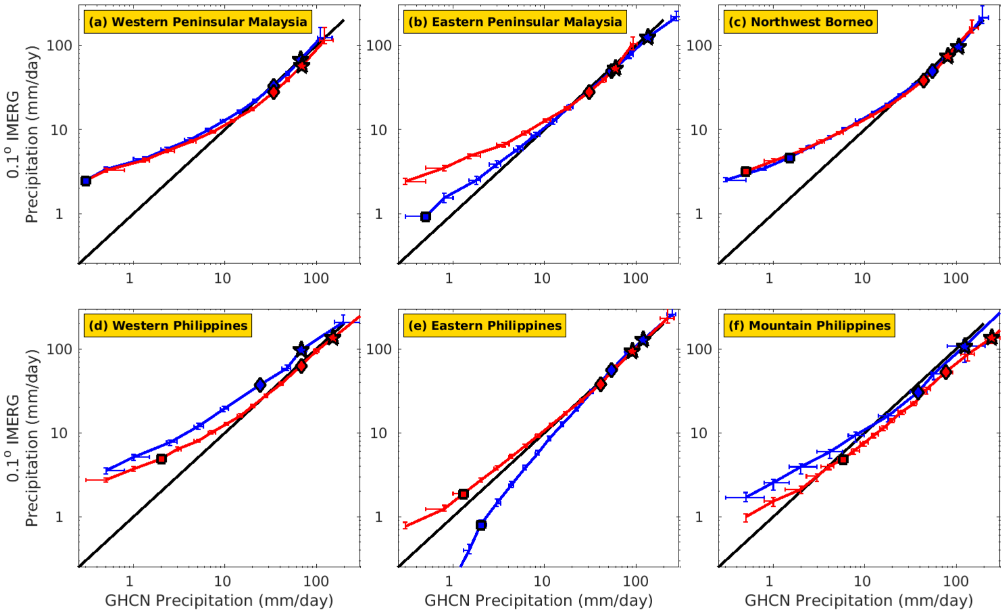

Figure 5. Quantile–quantile diagrams of GHCN daily precipitation versus nearest grid point IMERG daily precipitation during northern winter (October–March, blue) and northern summer (April–September, red) for: (a) Western Peninsular Malaysia, (b) Eastern Peninsular Malaysia, (c) North Western Borneo, (d) Western Philippines, (e) Eastern Philippines, (f) Mountain Philippines. The red markers highlight the 50^{th} (square), 95^{th} (diamond) and 99^{th} (asterisk) percentiles. The black line shows the 1:1 control line.

In the Eastern Philippines, the weak precipitation is underestimated by IMERG during northern winter but overestimated in northern summer (Fig. 5e); the rainfall matches GHCN station data above the 90^{th} percentile for both seasons, suggesting that the 95th percentile choice for evaluating extreme precipitation also holds during the northern winter in this region.

The case of the mountain Philippines station (Fig. 5f) remains undetermined because of the use of only one GHCN station, on the western side of the Cordillera Central mountain range. In mountain regions, the statistical distribution of precipitation extrema will vary spatially within a single IMERG grid box (approximately 11 km) due to topographic effects largely absent in coastal land areas. Indeed, precipitation will tend to be systematically heavier at high altitude than low altitude or on the windward side compared to the leeward side of individual mountains. These patterns of precipitation will persist between events, in contrast to the more random spatial distribution of rainfall over flat topography. These topographic controls will lead to spatial biases even in perfect observations.

Overall, the 95^{th} percentile appears to be a suitable choice for evaluating NWP daily precipitation in most of the regions evaluated here. However, this choice of percentile may not necessarily be appropriate for sub-daily precipitation extremes, which are examined in Section 3.4.

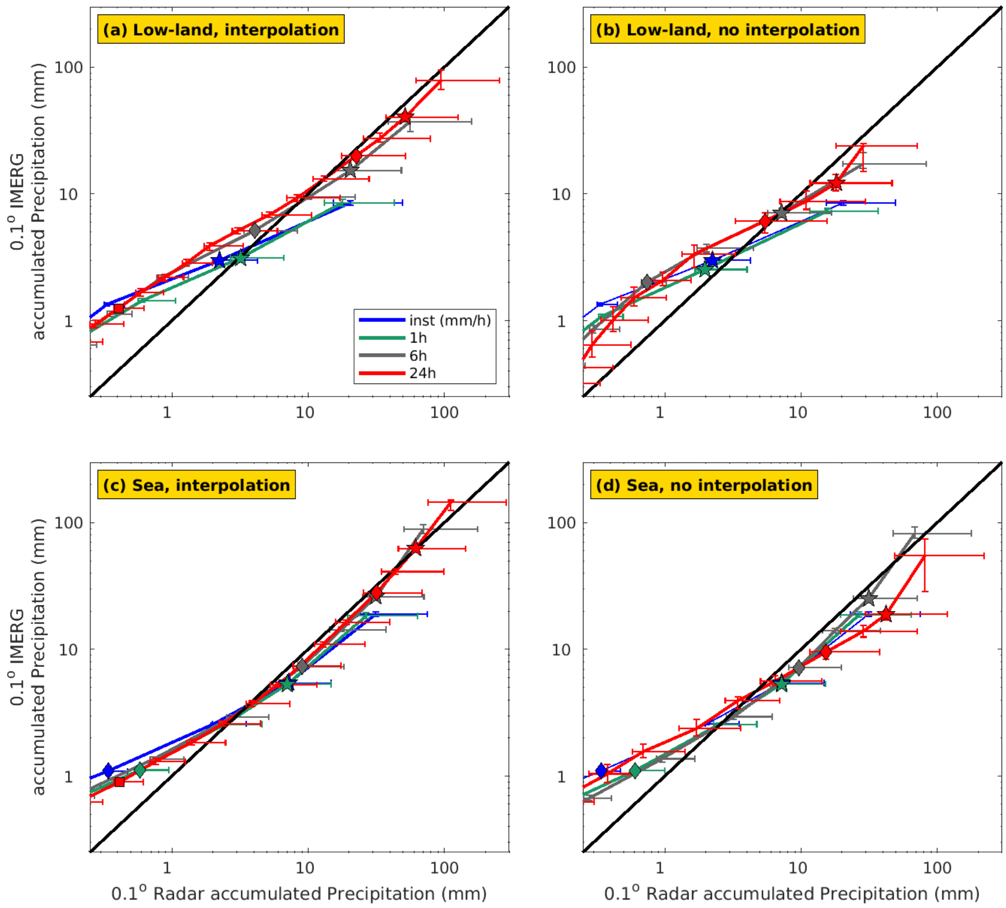

Figure 6. Quantile–quantile diagrams of precipitation accumulation from the Subang radar averaged onto the IMERG grid, versus precipitation accumulation from IMERG. In each panel, accumulations are shown for instantaneous precipitation (blue line), 1 hr (green), 6 hr (grey), and 24 hr (red). (a) Low-land grid points only, using the whole time period with interpolation over missing radar data values. (b) As (a), but only using data for periods where radar data exists. (c), (d) As (a) and (b) but for sea grid points. The black line shows the 1:1 control line. The markers highlight the 50th (square), 95th (diamond) and 99th (asterisk) percentiles.

3.4 Evaluation of sub-daily IMERG precipitation accumulation against radar

The Subang radar makes it possible to evaluate IMERG precipitation on sub-daily time scales. By comparing the IMERG data to the radar data gridded onto the same 0.1° IMERG grid, the spatial sampling error disappears. The uncertainties related to the Z–R relationship and potential hail contamination are evaluated in a similar way as in the previous section. The resultant intervals, as well as the IMERG 95% confidence intervals are represented by errors bars in Fig. 6. The uncertainties are far larger for the radar data than the IMERG data (Fig. 6), mainly associated with the choice of the Z–R relationship.

Sub-daily rainfall accumulations in IMERG were evaluated against radar data by constructing quantile–quantile diagrams of IMERG accumulated precipitation against 0.1° gridded radar accumulated precipitation, for various accumulation times (from instantaneous to daily), for low-land and sea grid points separately (Fig. 6). Despite the uncertainties, the comparison over land (Fig. 6a) shows that IMERG overestimates the lowest precipitation amounts compared to the radar, for all accumulation time scales from instantaneous to daily. This overestimation is consistent with the previous daily comparison with GHCN station data. For higher percentiles, IMERG tends to underestimate extreme precipitation for sub-hourly timescales compared with radar. Note that this underestimation only holds for the highest percentile used here, i.e. the 99.9^{th} percentile, thus corresponding to a very small number of cases.

Overall, the results for sea grid points are qualitatively similar to those for the land grid points (Fig. 6c, d). The overestimation of IMERG at low precipitation intensities is similar to the land case. The underestimation of IMERG sub-hourly extreme precipitation is less pronounced and no more robust than over land. Similarly to the land regions, the temporal interpolation error does not significantly affect the quantile-quantile relationship between IMERG and radar in the sea areas around Subang (Fig. 6d).

In contrast to the IMERG-GHCN comparison, we do not find any overestimation of daily IMERG precipitation at percentiles above the 95th percentile and there are no robust differences between IMERG and radar percentiles for longer accumulation times. In addition to the aforementioned radar uncertainties, there are several possible explanations for this. Temporal interpolation was necessary to fill gaps in the radar data, which may have induced errors; we estimate the potential impact of these by drawing a similar quantile-quantile diagram retaining only periods without any missing values (Fig. 6b). While this subsetting induces a significant decrease in the number of events (from 89 days to 10 days), the qualitative findings remain the same and they are also replicated over the sea (Fig. 6c, d). We therefore conclude that our findings are not dependent on the temporal interpolation method. Another potential reason for the apparent discrepancy between the radar and GHCN comparisons is the difference of period considered in each comparison. The IMERG versus GHCN comparison was done using nearly 20 years of data between 2001 and 2019 (without removing missing values) whereas the IMERG versus radar comparison is done with spatially aggregated data from 11 January to 15 April 2019. The 95% confidence interval error bars drawn in the IMERG–GHCN comparison account for the uncertainty linked to the representativeness of chosen period for the distribution of precipitation. However, these same errors bars in the IMERG–radar comparison mostly account for the spatial representativeness rather than the temporal representativeness, since time series from many grid points (86) were aggregated in this case compared to 3 for the GHCN-GPM comparison. Consequently, qualitative differences between the comparisons can be observed without contradiction. This suggests that although IMERG tends to overestimate the very high percentiles of daily precipitation, this overestimation is not necessarily present for all heavy precipitation events.

3.5 Representation of the diurnal cycle by IMERG

One of the major issues of NWP is its ability to correctly represent the diurnal cycle of precipitation. This is especially important for precipitation extremes, which often result from a complex interaction between the diurnal cycle and large-scale, slowly-evolving forcings. With its 30-minute output frequency, the IMERG product appears to be a good candidate to evaluate the diurnal cycle in NWP models. In this section, we use the Subang radar to assess the fidelity of IMERG in capturing the diurnal cycle of precipitation. Fig. 7 shows the 90^{th}, 95^{th}, 99^{th} percentile and mean instantaneous precipitation as a function of the time of day, for both the Subang radar and IMERG in both low-land and sea grid points. Despite the large uncertainties, IMERG agrees with the radar data with regards to the mean precipitation peak time in both low-land and sea areas. Mean precipitation peaks at about 6 UTC+8 over the sea and at 17 UTC+8 over the low-land areas for both IMERG and radar (Fig. 7a). For most times, the mean precipitation intensities are not significantly

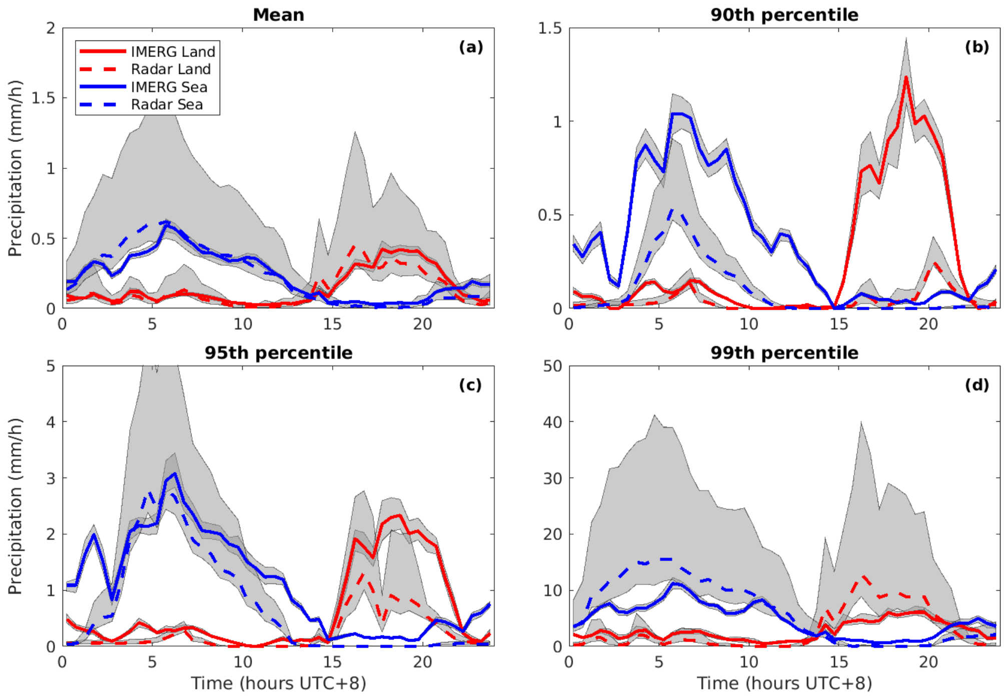

Figure 7. Mean (a), 90th (b), 95th (c), and 99th percentile (d) of instantaneous precipitation as of function of the time of the day for the IMERG product and for the Subang radar averaged on the IMERG grid. Diurnal cycles are represented for both land (red) and sea (blue) grid points. The grey shading areas display the 95% confidence intervals.

different between IMERG and radar, although the uncertainty in the radar data is very large.

This good agreement of mean precipitation hides some disparities in the statistical distribution of instantaneous precipitation, as seen previously in the quantile-quantile diagrams (Fig. 6). At the 90th percentile, IMERG consistently overestimates precipitation compared with the radar, especially for the peaks. The 95th percentile of IMERG precipitation remains quite close to the radar 95th percentile of precipitation especially over the sea. In the low-land areas, the IMERG 95th percentile precipitation peak is still stronger than the radar one but the differences are generally not significant with respect to the Z-R relationship uncertainty. However, the 99th percentile of precipitation tends to be underestimated by IMERG compared with the radar at the precipitation peak times in both land and sea regions. Despite these deficiencies in the amplitude of the diurnal cycle of extreme precipitation, the diurnal phase of extreme precipitation (the 90th 95th, and 99th percentiles) is reasonably well captured by IMERG.

4 Conclusion

Precipitation extremes have dramatic impacts on the population of the Maritime Continent. Improved predictions of such events can help to mitigate their negative effects. The evaluation of NWP models against reliable observation datasets is essential in order to understand model deficiencies. In this study, we evaluated the ability of the IMERG satellite product to detect extreme precipitation with the purpose of assessing its suitability for use in NWP model evaluations in the Maritime Continent.

We evaluated the global skill of IMERG with respect to the GHCN weather station dataset in Malaysia and in the Philippines. Our findings are similar to previous comparisons of IMERG with station data, with the best performance for longer accumulation times. However, we showed that the comparison of 0.1° grid versus pointwise precipitation is subjected to a spatial sampling error. Using the high resolution radar at Subang, we were able to estimate this spatial sampling error in western Peninsular Malaysia. We found that the sampling error may represent around 45% of the mean square error of daily precipitation between the GHCN weather station data and IMERG. This suggests that the skill of IMERG in detecting daily precipitation may have been underestimated in previous studies in this area and likely in the whole Maritime Continent.

When the spatial sampling error described above is taken into account, IMERG was found to overestimate low intensity daily precipitation. The overestimation of low precipitation may be due to erroneous detection of precipitation by IR sensors, as suggested by previous studies. Meanwhile, for very extreme precipitation over the 95th percentile, the IMERG precipitation coincides with the GHCN measurements in most regions. Given the identified spatial sampling error, this implies that IMERG is overestimating very extreme daily precipitation compared to the true area-averaged daily precipitation. This coincidence of both IMERG and GHCN extreme daily precipitation percentiles may be related to the use of only one gauge per grid point in the GPCC gauge–analysis product (which serves for the calibration of IMERG), as individual gauges unavoidably have higher extreme values than a grid average.

The use of radar data in western Peninsular Malaysia makes it possible to estimate more precisely the ideal choice of percentile to evaluate NWP extreme daily precipitation against IMERG. Our analysis shows that it is preferable to use the 95th percentile rather than the 99th percentile of daily precipitation to evaluate NWP against IMERG in western Peninsular Malaysia. We estimated that the IMERG 95th percentile is accurate with less than 20% potential error. Therefore, a 20% difference between NWP and IMERG is the minimum threshold for identification of model deficiencies, at least for the case of daily extreme precipitation at 0.1° horizontal resolution.

The lack of other very high resolution observational datasets in the Maritime Continent prevented us from performing the analysis with the same degree of confidence in the other selected areas. However, it was found that IMERG daily extreme percentiles match with those of GHCN in (the whole of) western Peninsular Malaysia, Eastern Peninsular Malaysia, Northwest Borneo, western Philippines during northern summer, and in eastern Philippines. Assuming that the 0.1° spatial variability of daily extreme precipitation does not vary much between regions, this implies that the findings for western Peninsula Malaysia are applicable across all these regions and likely across the whole Maritime Continent. Therefore it is not recommended to use very extreme percentiles for NWP evaluation against IMERG in these regions.

We found robust overestimation of low-level sub-daily IMERG precipitation when compared against Subang radar data. This overestimation was found for percentiles up to the 99th percentile for sub-hourly precipitation. However, very extreme (above the 99th percentile) sub-hourly precipitation was found to be robustly underestimated by IMERG compared to the radar in low-land areas. The differences of extreme precipitation at longer accumulation times were not significant at the 95% confidence interval when considering the uncertainties linked to the radar Z-R relationship and potential hail contamination on radar reflectivities. Further work aimed at reducing these uncertainties could help in diagnosing more precisely the behavior of IMERG, which would in turn improve the evaluation of NWP forecasts of extreme precipitation across the Maritime Continent.

The mean diurnal cycle of precipitation is fairly well reproduced by IMERG both in timing and intensity when compared with radar data. However, the peaks of precipitation remain either overestimated for percentiles below the 95th percentile or underestimated for percentiles above the 95th. This suggests that the 95th percentile of sub-hourly precipitation would also be preferable to higher percentiles for evaluation of NWP diurnal peak precipitation against IMERG. Finally, there was no obvious decrease of IMERG performances over the sea despite the absence of gauges.

In conclusion, we find that the spatial sampling error of precipitation can not be neglected when comparing IMERG against point-wise observations, particularly for extreme precipitation. Taking this into account, the combined evaluation of station and radar data supports the key finding that IMERG data is reliable for use in evaluating NWP simulations of extreme precipitation at the 95th percentile, with lower reliability at both higher and lower percentiles.

Acknowledgments

We thank Andy Hartley for liaising with Malaysian in country partners, which made it possible to obtain the Subang radar data. The GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06B (Huffman et al., 2019) was used in this study. Data were downloaded for the Maritime Continent through the National Aeronautics and Space Administration website at https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGHH_06/summary?keywords=%22IMERG%20final%22. The gauge precipitation data are from the Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.12 (Menne, Durre, Korzeniewski, et al., 2012) and can be downloaded at https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00861 for both Malaysia and the Philippines. The topography data were supplied by the General Bathymetric Chart of the Oceans: GEBCO Compilation Group (2019) GEBCO 2019 Grid (doi:10.5285/836f016a-33be-6ddc-e053-6c86abc0788e). Finally, the radar precipitation data in Subang (Malaysia) is freely available at https://doi.org/10.6084/m9.figshare.14330345.v2 (Da Silva, 2021). 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. NADS, BGMW, AJM, MMF, THMS, and CEH were supported through the Forecasting in Southeast Asia (FORSEA) project, funded by the Newton Fund through the Weather and Climate Science for Service Partnership (WCSSP) of the UK Met Office (award DN373682).

References

Abd Majid, N., Muhamad Nazi, N., Idris, N. D. M., & Taha, M. R. (2019). GIS-Based Livelihood Vulnerability Index Mapping of the Socioeconomy of the Pekan Community. *Sustainability*, *11*(24), 6935. doi: 10.3390/su11246935

Adler, R. F., Sapiano, M. R., Huffman, G. J., Wang, J.-J., Gu, G., Bolvin, D., ... Shin, D. B. (2018). The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, *9*(4), 138. doi: 10.3390/atmos9040138

Behrangi, A., & Wen, Y. (2017). On the spatial and temporal sampling errors of remotely sensed precipitation products. *Remote Sensing*, *9*(11), 1127. doi: 10.3390/rs9111127

Berne, A., & Krajewski, W. F. (2013). Radar for hydrology: Unfulfilled promise or unrecognized potential? *Advances in Water Resources*, *51*, 357–366. doi: 10.1016/j.advwatres.2012.05.005

Cabrera, J. S., & Lee, H. S. (2020). Flood risk assessment for Davao Oriental in the Philippines using geographic information system-based multi-criteria analysis and the maximum entropy model. *Journal of Flood Risk Management*, *13*(2). doi: 10.1111/jfr3.12607

Camici, S., Ciabatta, L., Massari, C., & Brocca, L. (2018). How reliable are satellite precipitation estimates for driving hydrological models: A verification study over the Mediterranean area. *Journal of hydrology*, *563*, 950–961. doi: 10.1016/j.jhydrol.2018.06.067

Da Silva, N. (2021). *Validation of GPM IMERG extreme precipitation in the Maritime Continent by station and radar data - Dataset.* doi: 10.6084/m9.figshare.14330345.v2

Derin, Y., Anagnostou, E., Berne, A., Borga, M., Boudevillain, B., Buytaert, W., … Yilmaz, K. K. (2016). Multiregional Satellite Precipitation Products Evaluation over Complex Terrain. *Journal of Hydrometeorology*, *17*(6), 1817–1836. doi: 10.1175/JHM-D-15-0197.1

Dezfuli, A. K., Ichoku, C. M., Huffman, G. J., Mohr, K. I., Selker, J. S., van de Giesen, N., … Annor, F. O. (2017). Validation of IMERG Precipitation in Africa. *Journal of Hydrometeorology*, *18*(10), 2817–2825. doi: 10.1175/JHM-D-17-0139.1

Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S. J., & Ropelewski, C. F. (2007). Validation of satellite rainfall products over East Africa's complex topography. *International Journal of Remote Sensing*, *28*(7), 1503–1526. doi: 10.1080/01431160600954688

Du, M., Zhang, M., Wang, S., Che, Y., Wang, J., Ma, R., & Yang, S. (2018). Precipitation measurement biases in an arid setting of central Asia: using different methods to divide precipitation types. *Climate Research*, *76*(1), 73–86. doi: 10.3354/cr01527

Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., & Vose, R. S. (2010). Comprehensive Automated Quality Assurance of Daily Surface Observations. *Journal of Applied Meteorology and Climatology*, *49*(8), 1615–1633. doi: 10.1175/2010JAMC2375.1

Fang, J., Yang, W., Luan, Y., Du, J., Lin, A., & Zhao, L. (2019). Evaluation of the TRMM 3B42 and GPM IMERG products for extreme precipitation analysis over China. *Atmospheric research*, *223*, 24–38. doi: 10.1016/j.atmosres.2019.03.001

Fulton, R. A., Breidenbach, J. P., Seo, D.-J., Miller, D. A., & O'Bannon, T. (1998). The wsr-88d rainfall algorithm. *Weather and forecasting*, *13*(2), 377–395. doi: 10.1175/1520-0434(1998)013⟨0377:TWRA⟩2.0.CO;2

Gaona, M. F. R., Overeem, A., Leijnse, H., & Uijlenhoet, R. (2016, 11). First-Year Evaluation of GPM Rainfall over the Netherlands: IMERG Day 1 Final Run (V03D). *Journal of Hydrometeorology*, *17*(11), 2799–2814. doi: 10.1175/JHM-D-16-0087.1

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., … Zhao, B. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. doi: 10.1175/JCLI-D-16-0758.1

Hai, O. S., Samah, A. A., Chenoli, S. N., Subramaniam, K., & Ahmad Mazuki, M. Y. (2017). Extreme rainstorms that caused devastating flooding across the east coast of Peninsular Malaysia during November and December 2014. *Weather and Forecasting*, *32*(3), 849–872. doi: 10.1175/WAF-D-16-0160.1

Holloway, C. E., Woolnough, S. J., & Lister, G. M. S. (2012). Precipitation distributions for explicit versus parametrized convection in a large-domain high-resolution tropical case study. *Quarterly Journal of the Royal Meteorological Society*, *138*(668), 1692–1708. doi: 10.1002/qj.1903

Huffman, G. J., Adler, R. F., Rudolf, B., Schneider, U., & Keehn, P. R. (1995). Global Precipitation Estimates Based on a Technique for Combining Satellite-Based Estimates, Rain Gauge Analysis, and NWP Model Precipitation Information. *Journal of Climate*, *8*(5), 1284–1295. doi: 10.1175/1520-0442(1995)008⟨1284:GPEBOA⟩2.0.CO;2

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., … Stocker, E. F. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *Journal of Hydrometeorology*, *8*(1), 38–55. doi: 10.1175/JHM560.1

Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2019). *GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC).* (Accessed: [18 Dec 2019]) doi: 10.5067/GPM/IMERG/3B-HH/06

Iguchi, T., Kozu, T., Kwiatkowski, J., Meneghini, R., Awaka, J., & Okamoto, K. (2009). Uncertainties in the rain profiling algorithm for the TRMM precipitation radar. *Journal of the Meteorological Society of Japan. Ser. II*, *87*, 1–30. doi: 10.2151/jmsj.87A.1

Joyce, R. J., & Xie, P. (2011). Kalman Filter–Based CMORPH. *Journal of Hydrometeorology*, *12*(6), 1547–1563. doi: 10.1175/JHM-D-11-022.1

Karki, T. K. (2019). Flood resilience in Malaysian cities: a case study of two towns in Johor state. *International Journal of Disaster Resilience in the Built Environment*, *11*(3), 329–342. doi: 10.1108/IJDRBE-06-2019-0037

Keller, C. A., Knowland, K. E., Duncan, B. N., Liu, J., Anderson, D. C., Das, S., ... Pawson, S. (2021). Description of the nasa geos composition forecast modeling system geos-cf v1.0. *Journal of Advances in Modeling Earth Systems*, *13*(4), e2020MS002413. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002413` (e2020MS002413 2020MS002413) doi: https://doi.org/10.1029/2020MS002413

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, How Much of the Earth's Surface Is Covered by Rain Gauges? *Bulletin of the American Meteorological Society*, *98*(1), 69–78. doi: 10.1175/BAMS-D-14-00283.1

Kidd, C., Tan, J., Kirstetter, P.-E., & Petersen, W. A. (2018). Validation of the Version 05 Level 2 precipitation products from the GPM Core Observatory and constellation satellite sensors. *Quarterly Journal of the Royal Meteorological Society*, *144*(S1), 313–328. doi: 10.1002/qj.3175

Kim, K., Park, J., Baik, J., & Choi, M. (2017). Evaluation of topographical and seasonal feature using GPM IMERG and TRMM 3B42 over Far-East Asia. *Atmospheric Research*, *187*, 95–105. doi: 10.1016/j.atmosres.2016.12.007

Kirschbaum, D. B., Huffman, G. J., Adler, R. F., Braun, S., Garrett, K., Jones, E., ... Zaitchik, B. F. (2017). NASA's Remotely Sensed Precipitation: A Reservoir for Applications Users. *Bulletin of the American Meteorological Society*, *98*(6), 1169–1184. doi: 10.1175/BAMS-D-15-00296.1

Krajewski, W. F., & Smith, J. A. (1991). On the Estimation of Climatological Z–R Relationships. *Journal of Applied Meteorology*, *30*(10), 1436–1445. doi: 10.1175/1520-0450(1991)030⟨1436:OTEOCR⟩2.0.CO;2

Kucera, P. A., Ebert, E. E., Turk, F. J., Levizzani, V., Kirschbaum, D., Tapiador, F. J., ... Borsche, M. (2013). Precipitation from Space: Advancing Earth System Science. *Bulletin of the American Meteorological Society*, *94*(3), 365–375. doi: 10.1175/BAMS-D-11-00171.1

Kummerow, C. D., Randel, D. L., Kulie, M., Wang, N.-Y., Ferraro, R., Munchak, S. J., & Petkovic, V. (2015). The evolution of the goddard profiling algorithm to a fully parametric scheme. *Journal of Atmospheric and Oceanic Technology*, *32*(12), 2265 - 2280. doi: 10.1175/JTECH-D-15-0039.1

Li, R., Wang, K., & Qi, D. (2018). Validating the Integrated Multisatellite Retrievals for Global Precipitation Measurement in Terms of Diurnal Variability With Hourly Gauge Observations Collected at 50,000 Stations in China. *Journal of Geophysical Research: Atmospheres*, *123*(18), 10,423–10,442. doi: 10.1029/2018JD028991

Liu, C.-Y., Aryastana, P., Liu, G.-R., & Huang, W.-R. (2020). Assessment of satellite precipitation product estimates over Bali Island. *Atmospheric Research*, 105032. doi: 10.1016/j.atmosres.2020.105032

Lorenz, C., & Kunstmann, H. (2012). The Hydrological Cycle in Three State-of-the-Art Reanalyses: Intercomparison and Performance Analysis. *Journal of Hydrometeorology*, *13*(5), 1397–1420. doi: 10.1175/JHM-D-11-088.1

Maggioni, V., Meyers, P. C., & Robinson, M. D. (2016). A Review of Merged High-Resolution Satellite Precipitation Product Accuracy during the Tropical Rainfall Measuring Mission (TRMM) Era. *Journal of Hydrometeorology*, *17*(4), 1101–1117. doi: 10.1175/JHM-D-15-0190.1

Marshall, J. S., Langille, R. C., & Palmer, W. M. K. (1947). Measurement of rainfall by radar. *Journal of Atmospheric Sciences*, *4*(6), 186 – 192. doi: 10.1175/1520-0469(1947)004⟨0186:MORBR⟩2.0.CO;2

Mayor, Y. G., Tereshchenko, I., Fonseca-Hernández, M., Pantoja, D. A., & Montes, J. M.

(2017). Evaluation of error in IMERG precipitation estimates under different topographic conditions and temporal scales over Mexico. *Remote Sensing*, *9*(5), 503. doi: 10.3390/rs9050503

Menne, M. J., Durre, I., Korzeniewski, B., McNeal, S., Thomas, K., Yin, X., ... Houston, T. G. (2012). *Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.12.* NOAA National Climatic Data Center. (Accessed: Sep. 2019) doi: 10.7289/V5D21VHZ

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology*, *29*(7), 897–910. doi: 10.1175/JTECH-D-11-00103.1

Miller, J. R. (1972). A Climatological Z-R Relationship for Convective Storms in the Northern Great Plains. In 15th Conf. on Radar Meteor. (Ed.), (p. 153-154).

Navarro, A., García-Ortega, E., Merino, A., Sánchez, J. L., Kummerow, C., & Tapiador, F. J. (2019). Assessment of IMERG Precipitation Estimates over Europe. *Remote Sensing*, *11*(21), 2470. doi: 10.3390/rs11212470

Nguyen, P., Ombadi, M., Sorooshian, S., Hsu, K., AghaKouchak, A., Braithwaite, D., ... Thorstensen, A. R. (2018). The PERSIANN family of global satellite precipitation data: a review and evaluation of products. *Hydrology and Earth System Sciences*, *22*(11), 5801–5816. doi: 10.5194/hess-22-5801-2018

O, S., Foelsche, U., Kirchengast, G., Fuchsberger, J., Tan, J., & Petersen, W. A. (2017). Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria. *Hydrology & Earth System Sciences*, *21*(12), 6559—6572. doi: 10.5194/hess-21-6559-2017

O, S., & Kirstetter, P.-E. (2018). Evaluation of diurnal variation of GPM IMERG-derived summer precipitation over the contiguous US using MRMS data. *Quarterly Journal of the Royal Meteorological Society*, *144*(S1), 270–281. doi: 10.1002/qj.3218

Oliveira, R., Maggioni, V., Vila, D., & Morales, C. (2016). Characteristics and diurnal cycle of GPM rainfall estimates over the central Amazon region. *Remote Sensing*, *8*(7), 544. doi: 10.3390/rs8070544

Omranian, E., & Sharif, H. O. (2018). Evaluation of the Global Precipitation Measurement (GPM) Satellite Rainfall Products over the Lower Colorado River Basin, Texas. *JAWRA Journal of the American Water Resources Association*, *54*(4), 882–898. doi: 10.1111/1752-1688.12610

Prakash, S., Mitra, A. K., Pai, D., & AghaKouchak, A. (2016). From TRMM to GPM: How well can heavy rainfall be detected from space? *Advances in Water Resources*, *88*, 1–7. doi: 10.1016/j.advwatres.2015.11.008

Rajulapati, C. R., Papalexiou, S. M., Clark, M. P., Razavi, S., Tang, G., & Pomeroy, J. W. (2020). Assessment of extremes in global precipitation products: How reliable are they? *Journal of Hydrometeorology*, 1–48. doi: 10.1175/JHM-D-20-0040.1

Rana, S., McGregor, J., & Renwick, J. (2015). Precipitation Seasonality over the Indian Subcontinent: An Evaluation of Gauge, Reanalyses, and Satellite Retrievals. *Journal of Hydrometeorology*, *16*(2), 631–651. doi: 10.1175/JHM-D-14-0106.1

Rosenfeld, D., Wolff, D. B., & Atlas, D. (1993). General probability-matched relations between radar reflectivity and rain rate. *Journal of Applied Meteorology and Climatology*, *32*(1), 50 – 72. doi: 10.1175/1520-0450(1993)032⟨0050:GPMRBR⟩2.0.CO;2

Schneider, U., Fuchs, T., Meyer-Christoffer, A., & Rudolf, B. (2008). Global precipitation analysis products of the gpcc. *Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation*, *112*.

Sharifi, E., Steinacker, R., & Saghafian, B. (2016). Assessment of GPM-IMERG and other precipitation products against gauge data under different topographic and climatic conditions in Iran: Preliminary results. *Remote Sensing*, *8*(2), 135. doi: 10.3390/rs8020135

Shige, S., Kida, S., Ashiwake, H., Kubota, T., & Aonashi, K. (2013). Improvement of TMI Rain Retrievals in Mountainous Areas. *Journal of Applied Meteorology and Climatology*, *52*(1), 242–254. doi: 10.1175/JAMC-D-12-074.1

Skofronick-Jackson, G., Kirschbaum, D., Petersen, W., Huffman, G., Kidd, C., Stocker, E., & Kakar, R. (2018). The Global Precipitation Measurement (GPM) mission's scientific achievements and societal contributions: reviewing four years of advanced rain and snow observations. *Quarterly Journal of the Royal Meteorological Society*, *144* (S1), 27–48. doi: 10.1002/qj.3313

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K.-L. (2018). A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, *56* (1), 79–107. doi: 10.1002/2017RG000574

Takama, T., Aldrian, E., Kusumaningtyas, S. D., & Sulistya, W. (2017). Identified vulnerability contexts for a paddy production assessment with climate change in Bali, Indonesia. *Climate and Development*, *9* (2), 110–123. doi: 10.1080/17565529.2016.1167658

Tan, J., Huffman, G. J., Bolvin, D. T., & Nelkin, E. J. (2019). Diurnal Cycle of IMERG V06 Precipitation. *Geophysical Research Letters*, *46* (22), 13584–13592. doi: 10.1029/2019GL085395

Tan, J., Petersen, W. A., Kirstetter, P.-E., & Tian, Y. (2017). Performance of IMERG as a Function of Spatiotemporal Scale. *Journal of Hydrometeorology*, *18* (2), 307–319. doi: 10.1175/JHM-D-16-0174.1

Tan, J., Petersen, W. A., & Tokay, A. (2016). A Novel Approach to Identify Sources of Errors in IMERG for GPM Ground Validation. *Journal of Hydrometeorology*, *17* (9), 2477–2491. doi: 10.1175/JHM-D-16-0079.1

Tan, M. L., & Duan, Z. (2017). Assessment of GPM and TRMM precipitation products over Singapore. *Remote Sensing*, *9* (7), 720. doi: 10.3390/rs9070720

Tan, M. L., & Santo, H. (2018). Comparison of GPM IMERG, TMPA 3B42 and PERSIANN-CDR satellite precipitation products over Malaysia. *Atmospheric Research*, *202*, 63–76. doi: 10.1016/j.atmosres.2017.11.006

Tang, G., Behrangi, A., Long, D., Li, C., & Hong, Y. (2018). Accounting for spatiotemporal errors of gauges: A critical step to evaluate gridded precipitation products. *Journal of hydrology*, *559*, 294–306. doi: 10.1016/j.jhydrol.2018.02.057

Tang, G., Ma, Y., Long, D., Zhong, L., & Hong, Y. (2016). Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. *Journal of hydrology*, *533*, 152–167. doi: 10.1016/j.jhydrol.2015.12.008

Tian, F., Hou, S., Yang, L., Hu, H., & Hou, A. (2018, 02). How Does the Evaluation of the GPM IMERG Rainfall Product Depend on Gauge Density and Rainfall Intensity? *Journal of Hydrometeorology*, *19* (2), 339–349. doi: 10.1175/JHM-D-17-0161.1

Villarini, G., Mandapaka, P. V., Krajewski, W. F., & Moore, R. J. (2008). Rainfall and sampling uncertainties: A rain gauge perspective. *Journal of Geophysical Research: Atmospheres*, *113* (D11). doi: /10.1029/2007JD009214

Warlina, L., & Guinensa, F. (2019). Flood Susceptibility and Spatial analysis of Pangkalpinang City, Bangka Belitung, Indonesia. *Journal of Engineering Science and Technology*, *14* (6), 3481–3495.

Xie, P., & Arkin, P. A. (1997). Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs. *Bulletin of the American Meteorological Society*, *78* (11), 2539–2558. doi: 10.1175/1520-0477(1997)078⟨2539:GPAYMA⟩2.0.CO;2

Xu, F., Guo, B., Ye, B., Ye, Q., Chen, H., Ju, X., … Wang, Z. (2019). Systematical Evaluation of GPM IMERG and TRMM 3B42V7 Precipitation Products in the Huang-Huai-Hai Plain, China. *Remote Sensing*, *11* (6), 697. doi: 10.3390/rs11060697

Xu, R., Tian, F., Yang, L., Hu, H., Lu, H., & Hou, A. (2017). Ground validation of GPM IMERG and TRMM 3B42V7 rainfall products over southern Tibetan Plateau based on a high-density rain gauge network. *Journal of Geophysical Research: Atmospheres*, *122* (2), 910–924. doi: 10.1002/2016JD025418

Yamamoto, M. K., & Shige, S. (2015). Implementation of an orographic/nonorographic rainfall classification scheme in the gsmap algorithm for microwave radiometers. *Atmospheric Research*, *163*, 36-47. Retrieved from `https://www.sciencedirect.com/`

science/article/pii/S0169809514002993 (6th Workshop of the International Precipitation Working Group) doi: https://doi.org/10.1016/j.atmosres.2014.07.024

You, Y., Meng, H., Dong, J., & Rudlosky, S. (2019). Time-Lag Correlation Between Passive Microwave Measurements and Surface Precipitation and Its Impact on Precipitation Retrieval Evaluation. *Geophysical Research Letters*, *46*(14), 8415–8423. doi: 10.1029/2019GL083426

Yuda, I. W. A., Prasetia, R., As-syakur, A. R., Osawa, T., & Nagai, M. (2020). An assessment of IMERG rainfall products over Bali at multiple time scale. In *E3s web of conferences* (Vol. 153, p. 02001). doi: 10.1051/e3sconf/202015302001

Zhang, S., Wang, D., Qin, Z., Zheng, Y., & Guo, J. (2018). Assessment of the GPM and TRMM precipitation products using the rain gauge network over the Tibetan Plateau. *Journal of Meteorological Research*, *32*(2), 324–336. doi: 10.1007/s13351-018-7067-0