ORIGINAL PAPER

Dynamic rainfall thresholds for landslide early warning in Progo Catchment, Java, Indonesia

Ratna Satyaningsih[1,2] · Victor Jetten[1] · Janneke Ettema[1] ·
Ardhasena Sopaheluwakan[2] · Luigi Lombardo[1] · Danang Eko Nuryanto[2]

Abstract

This study aims to derive and evaluate new empirical rainfall thresholds as the basis for landslide early warning in Progo Catchment, Indonesia, using high-resolution rainfall datasets. Although attempts have been made to determine such thresholds for regions in Indonesia, they used coarse-resolution data and fixed rainfall duration that might not reflect the characteristics of rainfall events that induced the landslides. Therefore, we evaluated gauge-adjusted global satellite mapping of precipitation (GSMaP-GNRT) and bias-corrected climate prediction center morphing method (CMORPH-CRT) hourly rainfall estimates against measurements at rainfall stations. Based on this evaluation, a minimum rainfall of 0.2 mm/h was used to identify rain events, in addition to a minimum of 24 h of consecutive no-rain to separate two rainfall events. Rainfall thresholds were determined at various levels of non-exceedance probability, using accumulated and duration of rainfall events corresponding to 213 landslide occurrences from 2012 to 2021 compiled in this study. Receiver operating characteristics (ROC) analysis showed that thresholds based on rainfall station data, GSMaP-GNRT, and CMORPH-CRT resulted in area under ROC curve values of 0.72, 0.73, and 0.64, respectively. This result indicates that the performance of high-resolution satellite-derived data is comparable to that of ground observations in the Progo Catchment. However, GSMaP-GNRT outperformed CMORPH-CRT in discriminating the occurrence/non-occurrence of landslide-triggering rainfall events. For early warning purposes, the rainfall threshold is selected based on the probability exlevel at which the threshold maximizes the true skill score, i.e., at 10% if based on station data, or at 20% if based on GSMaP-GNRT.

Keywords Rainfall thresholds · Landslides · Early warning · GSMaP-GNRT · CMORPH-CRT

✉ Ratna Satyaningsih
r.satyaningsih@utwente.nl

1 Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands

2 Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG), Jakarta, Indonesia

1 Introduction

The complexity of the Indonesian landscape and the unique climatic regime the country is exposed to make it prone to landslides (Cepeda et al. 2010; Kirschbaum et al. 2015). The number of landslide occurrences has locally seen a significant increase in the last decade (BNPB 2021). Specifically, 5822 landslides have been documented in Indonesia from 2010 to 2020, out of which approximately 79% took place solely within Java (BNPB 2021). As a result, at least 1044 lives were lost, to which an appalling count of 971 people injured and 119 missing still need to be added. These numbers attest to the high-risk locals exposed to landslides (Cepeda et al. 2010; Sartohadi et al. 2010; Hadmoko et al. 2017). These studies also stressed the need for a suitable early warning system (EWS) meant to inform the population of the threat they may face and the appropriate response to be taken in imminent situations of hydro-morphological hazard.

One EWS for landslides (LEWS) is already in place in Indonesia (Hidayat et al. 2019), employing the rainfall threshold valid for the entire Indonesian landscape. A single rainfall threshold might be problematic as it may not reflect the climatic regime responsible for landslide occurrences in such a diverse territory, nor the terrain characteristics. As a result, the warnings issued for specific areas may be unsuitable, thus leading to false alarms or landslide events that have not been forecasted (Guzzetti et al. 2008; Segoni et al. 2014). This issue is relatively common, with examples of national thresholds failing when their performance is assessed on a local basis (Guzzetti et al. 2020). For this reason, scientific efforts have been made to create a spatially distributed network of rainfall thresholds for specific countries. For instance, Wang et al. (2021) subdivided the Chinese landscape according to six geomorphological regions and assigned a different threshold to each one. Moreover, Peruccacci et al. (2017) defined thresholds over environmental sub-zones in Italy clustered based on topography, lithology, land use, land cover, climate, and meteorology. An example from another tropical country comes from Malaysia, whose territory is divided into six zones, each relying on an independent EWS (Althuwaynee et al. 2015). However, even assigning thresholds to sub-zones of a given country may not be sufficient to support LEWS, and an even higher spatiotemporal detail may be required.

For operational LEWS, rainfall thresholds are typically derived for territorial units or alert zones surrounding a reference rain gauge (e.g., Lagomarsino et al. 2013; Segoni et al. 2014; Piciullo et al. 2017). The representation of the rainfall over the zone depends on the spatial distribution of the rain gauge. This distribution and spatial variability of the rainfall led to significant uncertainties in the estimation of landslide-triggering rainfall (Nikolopoulos et al. 2014), especially in complex topography and in case of localized convective rainfall (Segoni et al. 2018). The temporal resolution of the rain gauge data also contributes to the uncertainty. Previous studies showed that the usage of daily rainfall may underestimate rainfall thresholds, leading to false alarm if the thresholds are implemented in a LEWS (Marra 2019; Gariano et al. 2020). Nevertheless, rain gauge data represent direct measurement of rain amount that reach the surface.

On the other hand, defining thresholds over large territories also has advantages because one can safely rely on satellite-based precipitation products (SPPs) to describe the spatiotemporal rainfall distribution. For instance, the Tropical Rainfall Measuring Mission/TRMM rainfall estimates (Huffman et al. 2007) were used to derive rainfall thresholds in Papua New Guinea (Robbins 2016) and Indonesia (Hidayat et al. 2019).

However, due to its coarse resolution, the TRMM product failed to capture the spatially heterogeneous rainfall patterns within a catchment in Indonesia (Chikalamo et al. 2020). Brunetti et al. (2018) explored various SPPs to assess their capabilities to forecast landslides using rainfall thresholds for Italy and found better accuracy when using higher spatiotemporal resolution. Recent developments are attempting to mitigate the coarse-resolution issue with new-generation satellite products such as the Japan Aerospace Exploration Agency's (JAXA) Global Satellite Mapping of Precipitation/GSMaP (Kubota et al. 2020) and the Climate Prediction Center/National Oceanic and Atmospheric Administration's (CPC/NOAA) morphing technique (CMORPH) (Xie et al. 2017) ensuring a higher spatial and temporal resolution.

Irrespective of the data one may choose to use, it is common that LEWS relies on a fixed time window of rainfall duration from which the thresholds are to be derived. It is also found in the existing LEWS in Indonesia that uses 1-day and 3-day accumulated rainfall to define rainfall thresholds (Hidayat et al. 2019; Yuniawan et al. 2022). This approach may overlook the effects of longer antecedent days of rainfall. For example, at least 15-day antecedent rainfall was indicated to mostly trigger landslides in the Bogowonto Catchment, Indonesia (Chikalamo et al. 2020). There may be differences in the time window length, but the threshold definition itself boils down to a fixed rainfall duration (e.g., Robbins 2016; Chikalamo et al. 2020). This element requires further consideration. Weather systems are dynamic phenomena and do not always manifest in the same rainfall characteristics. Therefore, a fixed event duration implies that a threshold may miss some important rainfall characteristics and meteorological variability characteristics. A more appropriate solution would be a dynamic time window capable of adapting to the rainfall event responsible for landslides by extending or shortening its duration depending on the persistence of the rainfall signal.

Thus, the main objective of this study is to derive and evaluate new empirical rainfall thresholds for landslides in the Progo Catchment using high-resolution satellite-based precipitation products (SPPs) and rain gauge observations. To achieve this, we first evaluate the performance of SPPs in detecting rainfall measured by rain gauge observations. Then, we define rainfall events responsible for landslides by implementing a dynamic time window. This approach implicitly takes the initial state of soil into account. Afterward, we derive rainfall thresholds using the frequentist method (Brunetti et al. 2010, 2018; Peruccacci et al. 2012) that allows multiple thresholds corresponding to different non-exceedance probability levels.

The work presented here fits into the recent Indonesian effort to develop LEWS demonstrated in the work of Hidayat et al. (2019) and revised by Yuniawan et al. (2022). In particular, we move away from a unique national LEWS and instead move toward a nested system, where specific areas of the country can rely on their local forecast. There are several important areas where this study makes an original contribution to advance the LEWS in Indonesia. First, the dynamic time window we introduced here differs from a fixed and predefined duration of antecedent rainfall used in the thresholds commonly used in rainfall threshold analysis for LEWS in the Indonesia region (e.g., Irawan et al. 2019; Hidayat et al. 2019; Chikalamo et al. 2020; Yuniawan et al. 2022). Second, we offer probabilistic analysis with robustness that was not produced before for the area. Third, we generate multiple thresholds depending on the non-exceedance probability levels, ensuring an adaptive tool for operational purposes in landslide early warning in the area.

2 Study area and data sets

2.1 Study area

The study area is the Progo Catchment, which is in central Java, Indonesia, and covers an area of approximately 3100 km² (Fig. 1a, b). The Progo River, the main river, flows through several mountain ranges, such as Sumbing, Sundoro, Merbabu, Merapi, and Menoreh, which are located in Central Java and Yogyakarta Provinces. Those mountain ranges also form boundaries at the catchment's west, north, and east sides. Like other regions in Java Island, Progo Catchment is predominated by monsoon climate and experiences the wet northwest monsoon from November to March and the dry southeast monsoon from May to September (Aldrian and Susanto 2003). From 1991 through 2000, the average annual rainfall in the catchment was 2358 mm.
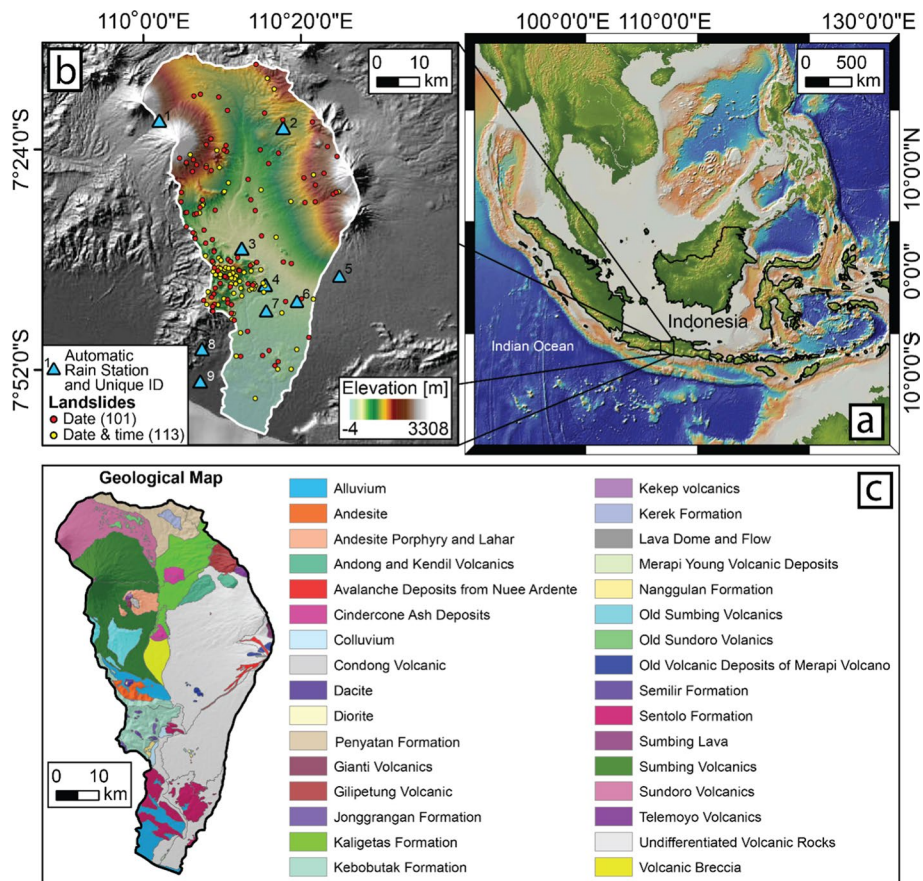

Fig. 1 **a** Location of Progo Catchment within Java Island, Indonesia. **b** Map of Progo Catchment, including distribution of landslide events and location of automatic rain stations. The ID number of automatic rain stations refers to stations presented in Table 1. **c** Geological map of the catchment. The distribution of different rock types, formations, and deposits was obtained from the Geological Agency of Indonesia (2022)

Geologically, the Progo Catchment is situated in two volcanic arcs: the Central Volcanic Zone and the Southern Java Mountain Zone (Amin et al. 1999). Figure 1c shows that the area reflects the volcanic origins in its geological outcrops (Geological Agency of Indonesia 2022). There are primarily Quaternary volcanic rocks or deposits in the form of ash, breccia, and lahar. Other geological types mainly consist of alluvium, colluvium, and local igneous rocks. In situ studies using geophysical methods revealed the existence of sliding planes in several sites within the Catchment (Katriani et al. 2019; Pratiwi et al. 2019; Widyadarsana and Hartantyo 2021; Dzakiya et al. 2023).

The complex landscape of the Progo Catchment makes the area prone to multiple natural hazards. For example, Mount Merapi is one of Indonesia's most active volcanoes; meanwhile, the flanks of Merapi, which are vulnerable to debris flow, are inhabited by more than 50,000 people (Hapsari et al. 2019). Furthermore, human settlements are situated on gentle slopes or very steep slopes, the instability of which is aggravated due to slope-cut and slope-fill practices to build houses (Hadmoko et al. 2010). In addition, Tropical Cyclone Cempaka, which hit the southern coast of Java on November 27, 2017 (BMKG 2017), caused floods and landslides in the catchment, among other affected areas.

2.2 Landslide inventory

We collected landslide information from authorized sources, including (1) online reports of landslide events compiled by the National Agency for Disaster Management (BNPB) and Regional Agency for Disaster Management (BPBD), (2) event reports by the territorial police department, (3) reports of disaster compiled by the Center for Health Crisis, Indonesian Ministry of Health, and (4) technical report by the Volcanology Survey Indonesia. We also gathered information on landside occurrences from online mass media. The landslide information collected and stored included (1) the exact landslide location, (2) the date and the time of the slope failures, (3) the type of landslide, and (4) the trigger. Since most of the reports were prepared by agencies responsible for disaster response, not all information is available in the inventory. The landslide locations were primarily reported by mentioning the name of the governmental administrative unit where the landslide occurred. The administrative unit includes Province, Regency/City, District, Village/Sub-district, and Hamlet, where the province is the highest level while Hamlet is the lowest. However, the lowest level of the administrative unit up to which the landslide location was reported varied. Only a few reports provided location coordinates. The coordinates are essential for determining objectively the representable rain station for determining rainfall event responsible for a landslide (Sect. 3.2). Therefore, we geocoded the locations through Google Geocoding API using the ggmap R package (Kahle and Wickham 2013) based on the names of the administrative units reported. The coordinates were used to select the representative automatic rain gauge station from which the rainfall event triggering the landslide was identified. Since the timing of landslide occurrences is crucial to derive a rainfall threshold, only the landslide events with at least a known date of occurrence were considered.

After reviewing the inventory, we obtained 214 landslide events in the study area from 2012 to 2021, out of which 213 events have a known date of occurrence. Note that the same rainfall event can trigger multiple landslides. The precise or approximate time of landslide occurrence was known for 110 (51.6%) events for these dated events.

2.3 Rainfall datasets

We used measurements by automatic rainfall stations and rainfall estimates obtained from SPPs to identify the dynamical rainfall thresholds triggering landslides in the study area. The SPPs include the Global Satellite Mapping of Precipitation (GSMaP) data by the Japan Aerospace Exploration Agency (JAXA) and the Climate Prediction Center morphing method (CMORPH) by the National Oceanic and Atmospheric Administration Climate Prediction Center (NOAA CPC). The products have been provided by their respective data owner for the World Meteorological Organization Space-based Weather and Climate Extremes Monitoring (SWCEM). Kuleshov et al. (2019) demonstrated that the two SPPs provide valuable information for monitoring heavy rainfall over the Maritime Continent.

2.3.1 GSMaP-GNRT data

JAXA's GSMaP Near-real-time Gauge-adjusted Rainfall Product version 6 (hereafter GSMaP-GNRT) is processed 4 h after observation and distributed with a spatial resolution of $0.1° \times 0.1°$ and hourly temporal resolution (Kubota et al. 2020). The dataset is available from April 2000 to the present. For the Southeast Asia region and Pacific Ocean, the data can be obtained from (ftp://swcem@hokusai.eorc.jaxa.jp/EAWP/GSMaP_GNRT/DATA/, last accessed 10 July 2022). The accuracy of the gauge-adjusted product is generally higher than the uncorrected ones (Kubota et al. 2020; Zhou et al. 2020). A few studies have evaluated the product on a sub-daily scale. Rauniyar et al. (2017) showed that the gauge-adjusted product outperformed other multi-satellite-based rainfall estimate products in characterizing rain–no rain over the Maritime Continent.

2.3.2 CMORPH-CRT data

Bias-corrected CMORPH precipitation estimates (hereafter, CMORPH-CRT) are originally generated with a spatial resolution of 8 km × 8 km over the global domain (60°S–60°N) in 30-min intervals (Joyce et al. 2004; Xie et al. 2017). For SWCEM, CMORPH-CRT is available from January 1998 to the present with a spatial resolution of $0.1° \times 0.1°$ and temporal resolution of 1 h. The data are updated at a latency of 2 h and can be obtained from (ftp://ftp.cpc.ncep.noaa.gov/precip/PORT/SEMDP/CMORPH_CRT/DATA/, last accessed 10 July 2022).

The reprocessing and bias correction of CMORPH original product make CMORPH-CRT better at representing the magnitude and spatial–temporal variations of precipitation over the quasi-global domain (Xie et al. 2017). This product is similar to the GSMaP-GNRT in the way that both are derived from PMW observations and adjusted using the CPC gauge-based analysis of global daily precipitation (Chen et al. 2008). The key difference lies in the usage of the infrared dataset. GSMaP derived rainfall estimates from infrared data between overpasses of passive microwave satellites (Ushio et al. 2009). Meanwhile, CMORPH retrieves rainfall estimates only from passive microwave data and uses the moving vector of cloud systems derived from infrared data to propagate precipitation estimated based on passive microwave data (Joyce et al. 2004; Xie et al. 2017). A study at 3-hourly temporal resolution showed that CMORPH-CRT could accurately capture the amplitude of the diurnal cycle over the Maritime Continent (Rauniyar et al. 2017).

2.3.3 Automatic rain stations data

The data measured by automatic rain stations were obtained from the Indonesian Agency for Meteorology, Climatology and Geophysics (BMKG). The automatic stations record cumulative rainfall every 10 min starting from past midnight and ending at midnight Coordinated Universal Time (UTC). The distribution of rain stations with such high temporal resolution in Indonesia is inhomogeneous and sparse in mountainous areas (please see BMKG 2022 for the map of the automatic weather station network). The network density is sparser than the density of a gauge network that typically measures at 3-hourly and daily periods at fixed synoptic hours (Kidd et al. 2017).

Before processing the data for analysis, we undertook a test to check for gross errors. Missing time intervals were added to create a complete 10-min time series. To detect implausible values, we checked if they fell within a range of 0–300 mm, the measure of the instruments. Values out of range were then manually inspected. If the implausible values were likely caused by mistakes in the algorithm for recording the rainfall value, the values were corrected using the appropriate formula. For instance, in case the values recorded in each 10-min interval during a day were 10-min intensity, instead of accumulation, we corrected the value by accumulating from the starting time of accumulation for the day. Otherwise, the values were considered suspicious and excluded from the analysis. Suppose there are dips or empty values, while the preceding and following values are the same and occur before the accumulated ending hours. In that case, the value is corrected as the same as the preceding value.

After the gross check, the 10-min data were aggregated into hourly rainfall to match the lowest temporal resolution of SPPs used in this study. The data were also aggregated into daily data for consistency checks with nearby sites. We inspected if the rainfall values > 20 mm/h (very heavy rainfall) had fallen over the site by comparing the daily data to the rainfall of the collocated observations. We accepted the value if the collocated observation exhibited a rainfall event with a comparable amount. Otherwise, we excluded the suspicious value. In the absence of collocated observation, we compared the data with rainfall observations from the neighboring site(s), typically within a distance of ~ 7 km.

Considering the length of rainfall time series, non-missing value > 70% from the available time series, and landslide area representativeness, we selected nine automatic rainfall stations within and near the study (Fig. 1). The data availability and information of each station are shown in Table 1.

3 Method

The rainfall estimates of SPPs are indirect measurements of rain rates. Hence, we first benchmark the performance of SPPs against the surface measurements by the automatic rainfall stations in the study area. Considering the data availability, a performance evaluation period of 1 October 2017–31 December 2021 was used. Afterward, we identified the rainfall events that triggered past landslides in the study area. Ultimately, we derived the rainfall thresholds based on the accumulated rainfall and the duration of the rainfall events.

Table 1 Meta-information on the automatic rain stations in the Progo catchment in Central Java, Indonesia

ID	Station name	Lon	Lat	Elevation (meter)	Available period
1	ARG Temanggung	110.0335°E	7.339445°S	234	2016-01-01–present
2	ARG Pakisdadu	110.2957°E	7.354779°S	620	2015-08-24–present
3	ARG SMPK Borobudur	110.2080°E	7.61007°S	248	2016-09-01–present
4	ARG Kalibawang	110.25779°E	7.687176°S	160	2017-09-06–present
5	AWS Pakem	110.4190°E	7.667140°S	415	2015-01-01–present
6	AWS Staklim Mlati	110.3354°E	7.716165°S	182	2015-04-01–present
7	ARG Moyudan	110.2620°E	7.740457°S	150	2017-09-06–present
8	ARG Waduk Sermo	110.123754°E	7.823245°S	169	2017-09-06–present
9	AWS Kulonprogo	110.1202°E	7.89177°S	8	2016-11-01–present

3.1 Statistical metrics

We evaluated the detection capacity of SPPs using four indicators based on a standard contingency table that includes (1) hit (H), where both SPP and rain station detect the rain occurrence; (2) miss (M), the rain station detects the rain occurrence, but the SPP does not; (3) false alarm (FA), SPP detects rain occurrence that does not occur; and (4) correct negative, neither the rain station nor the SPP detects the rain occurrence. The indicators are the probability of detection (POD), success ratio (SR), bias score (BIAS), and critical success index (CSI) (Wilks 1995). POD indicates the fraction of observed rains that SPP correctly detects. SR denotes the fraction of rain occurrences detected by SPP that actually occur. BIAS reflects whether the SPP tends to underestimate (BIAS < 1) or overestimate (BIAS > 1) rain events. CSI measures the overall fraction of rain occurrences correctly detected by SPP. Those four indicators are summarized in a performance diagram (Roebber 2009). To measure the discrepancy between SPP rain estimates and ground observations, we calculated the statistical metrics such as mean absolute error (MAE), relative bias (RB), and correlation coefficient (CC). The formula, range, and perfect score for each metric are listed in Table 2. The statistical metrics were calculated and visualized using the *verification* R package (NCAR 2015).

Following Haile et al. (2013), we calculated the statistical metrics for paired datasets of rainfall rates greater than a threshold value of rain detection for at least one of the two datasets. This approach was carried out to minimize the effects of rainless hours. Rain rates of 0.1 mm/h are considered suitable for defining no-rain conditions since such a low intensity is unlikely to generate surface ponding and runoff (Dunkerley 2015). To assess the performance of the SPPs in capturing higher rainfall intensities, we extended the rain detection threshold for other values: 0.2 mm/h, 0.5 mm/h, 1 mm/h, 5 mm/h, and 10 mm/h.

3.2 Determination of rainfall events

When the exact or estimated location and the date of landslide events were known, we selected the representable rain station based on the nearest distance. We also considered the station elevation by finding the minimum difference in elevation between the station and the landslide location. However, we prioritized the distance over elevation due to the sparseness of the

Table 2 Statistical metrics for evaluating the performance of SPPs

Indicator	Formula	Range	Perfect Score
Mean absolute error (MAE)	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert S_i - O_i \rvert$	$[0,\infty)$	0
Relative bias (RB)	$RB = \frac{\sum_i^n (S_i - O_i)}{\sum_i^n O_i}$	$(-\infty,\infty)$	0
Correlation coefficient (CC)	$CC = \frac{\sum_i^n (S_i-\bar{s})(O_i-\bar{O})}{\sqrt{\sum_i^n (S_i-\bar{s})^2}\sqrt{\sum_i^n (O_i-\bar{O})^2}}$	$[-1,1]$	1
Probability of detection (POD)	$POD = \frac{H}{H+M}$	$[0,1]$	1
Success ratio (SR)s	$SR = \frac{H}{H+FA}$	$[0,1]$	1
Bias score (BIAS)	$BIAS = \frac{H+FA}{H+M}$	$[0,\infty)$	1
Critical success index (CSI)	$CSI = \frac{H}{H+M+FA}$	$[0,1]$	1

S: SPP rain estimate, O: rainfall observed by automatic rain station, H: hit, M: miss, FA: false alarm

automatic rain station network that records in a sub-hourly time scale in the study area. If the station with the minimum difference in elevation was at a distance beyond twice SPP spatial resolution (20 km), we opted for the nearest station.

Once the appropriate rain station was identified and the precise or approximate time of landslide occurrence was known, we determined the possible rainfall event that triggered the landslide. We took the time of rainfall measurement coinciding with the time of landslide occurrence as the ending time of rainfall event. When the approximate time of landslide occurrence was unknown, we identified the ending time of rainfall as the time of peak rainfall intensity during the day of the landslide event. Considering the temporal variation of rainfall at the sub-daily scale, identifying the starting time of rainfall events might be complicated. Therefore, we considered a minimum of 24 h of consecutive no-rain to separate two rainfall events. The rainfall event defined in this manner is equivalent to the effective cumulative rainfall in Hong et al. (2018). The interval between the starting and ending times is the duration of the rainfall event. The procedure allows us to have any duration of rainfall events, hence a dynamic time window. We then accumulated the rainfall from the starting to the ending time. In the end, we had a pair of accumulated rainfall and duration for each landslide event to derive rainfall thresholds (Sect. 3.3). The same procedure was applied to the SPP rainfall database extracted from the grid corresponding to the location of the reference rain stations.

For threshold performance evaluation, the non-triggering rainfall events were taken into account. Using the similar definition above, we extracted the non-triggering rainfall events from the period between the oldest and the most current landslide event in our inventory.

3.3 Calculation of rainfall thresholds

To derive an empirical rainfall threshold, we adopted the frequentist method (Brunetti et al. 2010, 2018; Peruccacci et al. 2012) that assumes a threshold curve of power-law form, relating the accumulated rainfall E (mm) to the rainfall duration D (hour):

$$E = (\alpha \pm \Delta\alpha) \cdot D^{(\gamma \pm \Delta\gamma)} \qquad (1)$$

where α is a scaling constant (the intercept), and γ is the shape parameter that defines the slope of the threshold curve. We use accumulated rainfall instead of rainfall intensity for several reasons. First, using peak rainfall intensities instead of the actual triggering intensity may lead to overestimating rainfall thresholds (Staley et al. 2013). Moreover, when using SPPs, one should consider the SPPs' capacity performance in detecting high-intensity rainfall. As we show later in Sect. 4.1, the SPPs' skill in representing high-intensity rainfall is limited even at an hourly time scale. Secondly, the mean intensity that represents the average value of rainfall over a period of time has different physical meanings for short and long periods (Guzzetti et al. 2007). Since mean intensity has a functional dependency on duration through the accumulated rainfall, using accumulated rainfall to derive rainfall thresholds statistically is more appropriate (Peruccacci et al. 2012).

Multiple thresholds corresponding to different non-exceedance probabilities were derived, ranging from 1 to 50%. The non-exceedance probability level of 50% corresponds to the best-fit line of all (D, E) scatter points, while lower levels represent threshold lines parallel to the best-fit line. When the rainfall in the study area exceeds the threshold, a landslide is likely to occur.

In order to be statistically robust, following Brunetti et al. (2018), we randomly selected 70% of the landslide-triggering rainfall events and used the remaining to assess the threshold performance. This procedure was repeated 100 times, allowing us to obtain the mean and standard deviation of α and γ in Eq. 1 and to construct a contingency table. The contingency table lists four possible outcomes with the threshold as a binary classifier of the rainfall conditions ((D, E) points) that triggered or did not trigger landslides. A landslide-triggering (D, E) located above the threshold is a true positive (TP), and below the threshold is a false negative (FN). A non-triggering (D, E) above the threshold is a false positive (FP), and below the threshold is a true negative (TN). Using the contingency table, we calculated three skill scores, i.e., true positive rate (TPR) and false positive rate (FPR) (Fawcett 2006), as well as true skill statistic (TSS):

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{3}$$

$$\mathrm{TSS} = \mathrm{TPR} - \mathrm{FPR} \tag{4}$$

TPR represents the proportion of the landslide-triggering (D, E) above the threshold (correctly predicted), and FPR is the proportion of non-triggering (D, E) above the threshold (incorrectly predicted). TSS measures the performance of the rainfall threshold to separate the triggering- and non-triggering rainfall events, where value of 1 indicates the best performance and value of 0 or less suggests a performance no better than a random chance. TSS is also known as Hanssen and Kuipers discriminant (Hanssen and Kuipers 1965).

For each non-exceedance probability, we classified the rainfall events (E, D) into four categories of contingency (TP, FP, FN, and TN) and calculated TPR and FPR skill scores. We then constructed a receiver operating characteristic (ROC) graph (Fawcett 2006) based on TPR and FPR to evaluate the predicting capability of rainfall threshold for different non-exceedance probability levels. The area under the ROC curve (AUC) measures the goodness of each dataset. AUC values range from 0 (perfectly inaccurate) to 1 (perfectly accurate). In general, an AUC of 0.5 indicates no discrimination (i.e.,

landslide/no-landslide-triggering rainfall events), $0.5 < AUC < 0.7$ is considered poor, $0.7 \leq AUC < 0.8$ is considered acceptable, $0.8 \leq AUC < 0.9$ is considered excellent, and $AUC \geq 0.9$ is considered outstanding (Hosmer et al. 2013).

4 Results

4.1 Performance of GSMaP-GNRT and CMORPH-CRT

The capacity of GSMaP-GNRT and CMORPH-CRT in detecting the hourly rainfall observed by automatic rain stations in the study area is summarized in performance diagrams (Fig. 2). Given the perfect scores of statistical metrics shown in Table 2, a perfect detection capacity lies in the upper right of the performance diagram. In general, the detection capacity of both SPPs over each station site lies in approximately the same location, i.e., near the center for low rainfall intensities (Fig. 2a–d) and lower left for higher rainfall intensity (Fig. 2e, f). The performance diagrams suggest that the capability of SPPs in detecting rainfall at an hourly time scale tends to be poorer as the threshold of rain detection increases. In general, the performance of GSMaP-GNRT is comparable to that of CMORPH-CRT, and both SPPs do not exhibit a trend with regard to the station elevation.
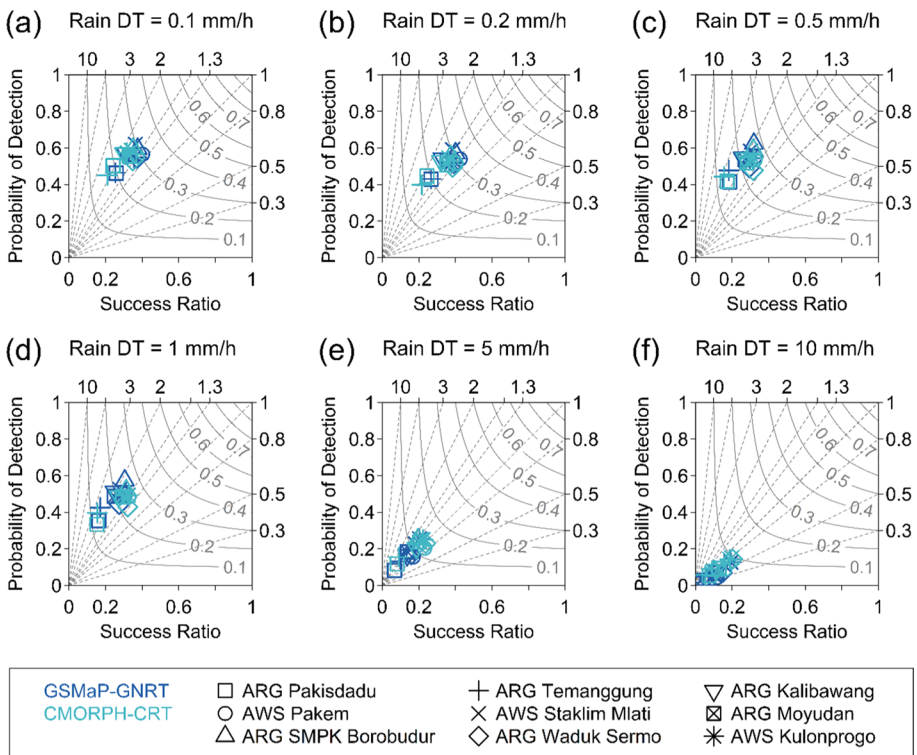

Fig. 2 Performance diagram summarizing POD, SR, CSI, and BIAS of SPPs. Light gray contours represent CSI, whereas the dashed lines denote BIAS. Rain DT above each graph indicates the rain detection threshold (in mm/hour)

However, there are two groups in the performance diagrams depicting the detection of low rain intensity (Fig. 2a–d). Specifically, if we inspect this figure further, the lower group exhibiting poorer performance represents the performances of the SPPs over the stations located leeward of the mountain and hill ranges, i.e., ARG Pakisdadu and ARG Temanggung. Except for BIAS, the performance indicator values of this group are persistently lower than those of the upper group.

POD values imply that < 63% of low rainfall intensity observed by automatic rain stations in the study area are correctly detected by both SPPs (Fig. 2a–d). Less than 26% of the rain rates ≥ 5 mm/h observed in the study area were correctly detected by GSMaP-GNRT and CMORPH-CRT (Fig. 2e). The fraction even has a lower value for rain rates ≥ 10 mm/h, below 16% (Fig. 2f). Along with low POD values, low SR values are observed as well. Since SR equals to 1—false alarm ratio, the low SR values imply that the SPP algorithms detected a significant number of rains that did not occur. Less than 43% of low rain intensity detected by either GSMaP-GNRT or CMORPH-CRT were observed at the surface in the study area, and the proportion is even less (< 24%) for higher rainfall intensity. The combination of low POD and SR values resulted in low CSI values, indicating that the overall fraction of rainfall correctly detected by both SPPs is relatively small. BIAS values suggest that both GSMaP-GNRT and CMORPH-CRT tend to overestimate the low-intensity rainfall (Fig. 2a–d) and underestimate the high-intensity rainfall in our study area (Fig. 2f). The SPPs' skill in detecting low-intensity rainfall varies from station to station. Overall, both SPPs have very low skill in detecting high rain intensity at an hourly time scale.

Based on the performance in detecting rain over various detection thresholds described above and the spread of the indicators in the performance diagram, we consider using 0.2 mm/h as the basis for rain/no-rain cut-off in the subsequent analysis. At this rain detection threshold, the spread of BIAS values is the narrowest among other detection thresholds for low-intensity rainfall.

Figure 3 shows a significant scatter between GSMaP-GNRT hourly rain estimates and the corresponding surface measurements at automatic rain stations in Progo Catchment. Few data points spread along the diagonal line, and false detection (missed rain) by the GSMaP-GNRT along the vertical (horizontal) axis is observed (Fig. 3). As shown in Fig. 3, GSMaP-GNRT tends to overestimate low rainfall intensity and underestimate high-intensity rainfall. However, most of the rain estimates and rain observed in the stations are low intensity. Therefore, GSMaP-GNRT exhibits very low CC at an hourly time scale (Table 3). This result suggests that GSMaP-GNRT could not reproduce hourly rainfall variation. Moreover, GSMaP-GNRT appears to provide rain estimates with a maximum value of ~ 30 mm/h.

A significant scatter also can be found between CMORPH-CRT rain estimates and the corresponding observed rain data from rain stations (Fig. 4). There are few data points spread along the diagonal line, while false detection and missed rain are visible. Like GSMaP-GNRT, CMORPHT-CRT could not capture hourly rainfall observed in the Progo Catchment.

In general, the performance of GSMaP-GNRT in detecting rain ≥ 0.2 mm/h is comparable to that of CMORPH-CRT in terms of MAE, RB, and CC (Table 3). MAE and CC values are relatively unvarying from station to station in the Progo Catchment. GSMaP-GNRT has MAE values ranging from 2.32 mm (in Temanggung) to 3.03 mm (in Pakem), while MAE of CMORPH-CRT ranges from 2.37 mm (in Temanggung) to 2.94 (in Pakem). Regarding the RB indicator, RB values vary from station to station but do not show a trend regarding elevation. The values support the results in Fig. 2b, where GSMaP-GNRT and
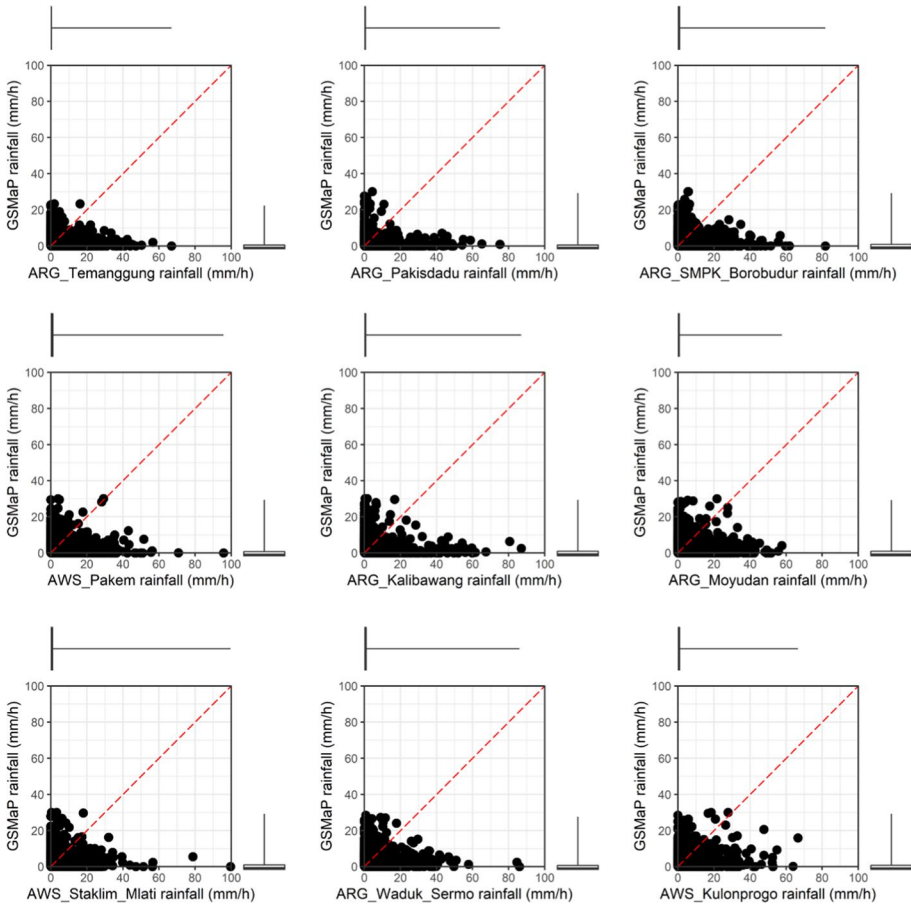
Fig. 3 Scatter plots of GSMaP-GNRT and hourly rain rates observed at automatic rain stations in the study area. Each black dot represents a pair of observations, with at least one of the pair having a value of ≥ 0.2 mm/h. Box and whisker plot at horizontal (vertical) axis shows the spread of observed rain at the station (GSMaP-GNRT rain estimates) at hourly time scale. The whiskers extend to the minimum and maximum values

CMORPH-CRT slightly overestimate the rain observed at almost all stations except at AWS Pakem and ARG Waduk Sermo. The largest RB is observed at ARG Temanggung. The large bias is likely related to the fact that the station is located between two mountains on the leeward side. Regarding correlation, CC values show a very weak relationship between rain estimates obtained from each SPP and observed rain in all stations in the Progo Catchment at an hourly time scale.

4.2 Empirical rainfall thresholds

The duration of rainfall events (D) measured by automatic rain stations that triggered landslides in the Progo Catchment ranged from 1 to 289 h (~12 days) with an average of ~87 h. GSMaP-GNRT provided the duration of rainfall events triggering landslides

Table 3 Values of statistical metrics for hourly GSMaP-GNRT and CMORPH-CRT rain estimates with a rain detection threshold of 0.2 mm/h over the Progo Catchment

	TMG (1)	PKD (2)	SBR (3)	KLB (4)	PKM (5)	SML (6)	MYD (7)	WDS (8)	KLP (9)
MAE									
GSMaP-GNRT	2.32	2.72	2.66	3.01	3.03	2.70	2.90	2.79	2.61
CMORPH-CRT	2.37	2.65	2.55	2.87	2.94	2.54	2.72	2.66	2.43
RB									
GSMaP-GNRT	0.40	0.12	0.16	0.16	−0.12	0.26	0.23	0.03	0.19
CMORPH-CRT	0.70	0.26	0.18	0.11	−0.09	0.21	0.12	−0.01	0.15
CC									
GSMaP-GNRT	0.03	−0.02	0.07	0.03	0.07	0.09	0.08	0.08	0.19
CMORPH-CRT	0.06	0	0.14	0.09	0.15	0.20	0.14	0.17	0.21

TMG: ARG Temanggung, PKD: ARG Pakisdadu, SBR: ARG SMPK Borobudur, KLB: ARG Kalibawang, PKM: AWS Pakem, SML: AWS Staklim Mlati, MYD: ARG Moyudan, WDS: ARG Waduk Sermo, KLP: AWS Kulonprogo. The numbers below the abbreviations of station names indicate the ID displayed in Fig. 1 and Table 1

ranging from 1 to 903 h (~38 days) with an average of 130.5 h. CMORPH-CRT produced a duration of rainfall events triggering landslides ranging from 1 to 387 h (~16 days) with an average of ~90 h. Note that the temporal resolution of the rainfall datasets to calculate the thresholds might contribute to the minimum duration of 1 h.

As for the accumulated rainfall (E) triggering landslides, measurements by automatic rain stations show the values ranged from 0.2 to 314.6 mm with an average of 102.2 mm. GSMaP-GNRT showed accumulated rainfall between 0.2 and 755.4 mm with an average of 121.9 mm. CMORPH-CRT produced accumulated rainfall as low as 0.2 mm and as high as 390.3 mm, with an average of 92.3 mm.

From the reconstructed rainfall events (D, E) for each dataset that triggered landslides in our landslide inventory, we calculated rainfall thresholds at several levels of non-exceedance probability. Table 4 presents equations of threshold curves defined from each dataset at several levels of non-exceedance probability, ranging from 1 to 50%. Close inspection of the equations shows that the station data generally produced the lowest thresholds at non-exceedance probability levels of 1–20%. At non-exceedance probability levels of 20–35%, station data exhibited only slightly lower thresholds than those derived from GSMaP-GNRT. However, station data produced slightly higher thresholds than GSMaP-GNRT for D > 17 h, > 1.3 h, and > 5.5 h at non-exceedance probability levels of 40%, 45%, and 50%, respectively. Compared to CMORPH-CRT, the station data exhibit higher thresholds for shorter rainfall duration but higher for longer duration. The rainfall duration cut-off of this transition rises as the non-exceedance probability level increases. For instance, at level of 5%, the station data produce a lower threshold for D < 15 h. In comparison, at level of 20%, the data exhibit a lower threshold for D < 5 days. Comparing the thresholds derived from GSMaP-GNRT and CMORPH-CRT, at non-exceedance probability levels of 1–35%, GSMaP-GNRT consistently produces higher thresholds higher than CMORPH does, except for rainfall duration longer than ~11 days. At non-exceedance probability levels of 40%, 45%, and 50%, CMORPH-CRT exhibits higher thresholds for rainfall duration longer than 7 days, 3 days, and ~6 days, respectively.
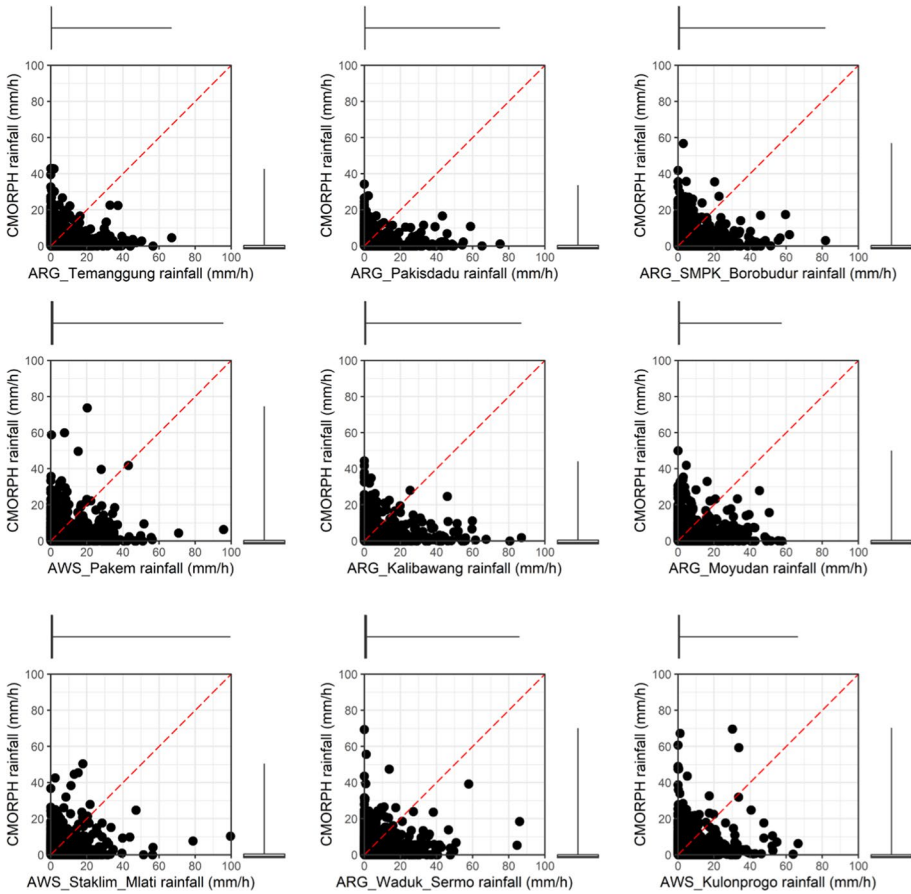
Fig. 4 Similar to Fig. 3, but the scatter plots compare CMORPH-CRT and hourly rain rates observed at automatic rain stations in the study area

We assessed the performance of the thresholds using the remaining 30% of the landslide-triggering rainfall events and the landslide-non-triggering rainfall events. Figure 5 represents one out of 100 samples of derived rainfall thresholds at the non-exceedance probability level of 5% for three different datasets. The total number of rainfall events in each category of a contingency table for performance evaluation is also provided in Fig. 5. Note that the number of landslide-triggering rainfall events captured by the three datasets might be different because rainfall data were sometimes unavailable. Figure 6 shows the constructed ROC describing the performance of the threshold at different levels of non-exceedance probability. It can be seen from Fig. 6 that the automatic rainfall station dataset provides slightly better performance compared to GSMaP-GNRT, particularly at a non-exceedance probability level below 15%. CMORPH-CRT performs the worst for all non-exceedance probabilities.

Further, a summary statistic of ROC in terms of AUC was calculated to be more objective. The results show that the AUC for automatic rainfall station, GSMaP-GNRT, and CMORPH-CRT is 0.72, 0.73, and 0.64, respectively. The AUC values also suggest that automatic rainfall station data and GSMaP-GNRT have acceptable capacity in

Table 4 Equations describing rainfall ED thresholds at different non-exceedance probability levels (labeled as T_x) derived for Progo Catchment	Dataset	Label	Equation
	Automatic rain stations	T_1	$E = (-0.30 \pm 0.19) \cdot D^{(0.82 \pm 0.09)}$
		T_5	$E = (-0.10 \pm 0.19) \cdot D^{(0.82 \pm 0.09)}$
		T_{10}	$E = (0.01 \pm 0.19) \cdot D^{(0.82 \pm 0.09)}$
		T_{15}	$E = (0.09 \pm 0.19) \cdot D^{(0.82 \pm 0.09)}$
		T_{20}	$E = (0.15 \pm 0.18) \cdot D^{(0.82 \pm 0.09)}$
		T_{25}	$E = (0.20 \pm 0.18) \cdot D^{(0.82 \pm 0.09)}$
		T_{30}	$E = (0.25 \pm 0.19) \cdot D^{(0.82 \pm 0.09)}$
		T_{35}	$E = (0.30 \pm 0.19) \cdot D^{(0.82 \pm 0.09)}$
		T_{40}	$E = (0.34 \pm 0.18) \cdot D^{(0.82 \pm 0.09)}$
		T_{45}	$E = (0.35 \pm 0.17) \cdot D^{(0.82 \pm 0.09)}$
		T_{50}	$E = (0.37 \pm 0.17) \cdot D^{(0.82 \pm 0.09)}$
	GSMaP-GNRT	T_1	$E = (-0.07 \pm 0.10) \cdot D^{(0.79 \pm 0.03)}$
		T_5	$E = (0.08 \pm 0.09) \cdot D^{(0.79 \pm 0.03)}$
		T_{10}	$E = (0.17 \pm 0.09) \cdot D^{(0.79 \pm 0.03)}$
		T_{15}	$E = (0.22 \pm 0.08) \cdot D^{(0.79 \pm 0.03)}$
		T_{20}	$E = (0.27 \pm 0.08) \cdot D^{(0.79 \pm 0.03)}$
		T_{25}	$E = (0.31 \pm 0.08) \cdot D^{(0.79 \pm 0.03)}$
		T_{30}	$E = (0.34 \pm 0.08) \cdot D^{(0.79 \pm 0.03)}$
		T_{35}	$E = (0.37 \pm 0.08) \cdot D^{(0.79 \pm 0.03)}$
		T_{40}	$E = (0.38 \pm 0.06) \cdot D^{(0.79 \pm 0.03)}$
		T_{45}	$E = (0.35 \pm 0.06) \cdot D^{(0.79 \pm 0.03)}$
		T_{50}	$E = (0.39 \pm 0.07) \cdot D^{(0.79 \pm 0.03)}$
	CMORPH-CRT	T_1	$E = (-0.34 \pm 0.11) \cdot D^{(0.90 \pm 0.05)}$
		T_5	$E = (-0.19 \pm 0.10) \cdot D^{(0.90 \pm 0.05)}$
		T_{10}	$E = (-0.11 \pm 0.10) \cdot D^{(0.90 \pm 0.05)}$
		T_{15}	$E = (-0.05 \pm 0.10) \cdot D^{(0.90 \pm 0.05)}$
		T_{20}	$E = (-0.01 \pm 0.10) \cdot D^{(0.90 \pm 0.05)}$
		T_{25}	$E = (0.03 \pm 0.10) \cdot D^{(0.90 \pm 0.05)}$
		T_{30}	$E = (0.06 \pm 0.10) \cdot D^{(0.90 \pm 0.05)}$
		T_{35}	$E = (0.10 \pm 0.10) \cdot D^{(0.90 \pm 0.05)}$
		T_{40}	$E = (0.13 \pm 0.10) \cdot D^{(0.90 \pm 0.05)}$
		T_{45}	$E = (0.14 \pm 0.09) \cdot D^{(0.90 \pm 0.05)}$
		T_{50}	$E = (0.15 \pm 0.09) \cdot D^{(0.90 \pm 0.05)}$

E is accumulated rainfall (in mm) and D is rainfall duration (in hour)

discriminating the occurrence/non-occurrence of rainfall events resulting in landslides, while CMORPH-CRT has poor capacity.

Since the aim of defining rainfall thresholds is for an early warning system, choosing at which non-exceedance probability level the threshold provides the minimum number of false alarms (FP) and missed occurrences (FN) is essential. In other words, we find the threshold that maximizes TPR and minimizes FPR and consequently maximizes TSS,
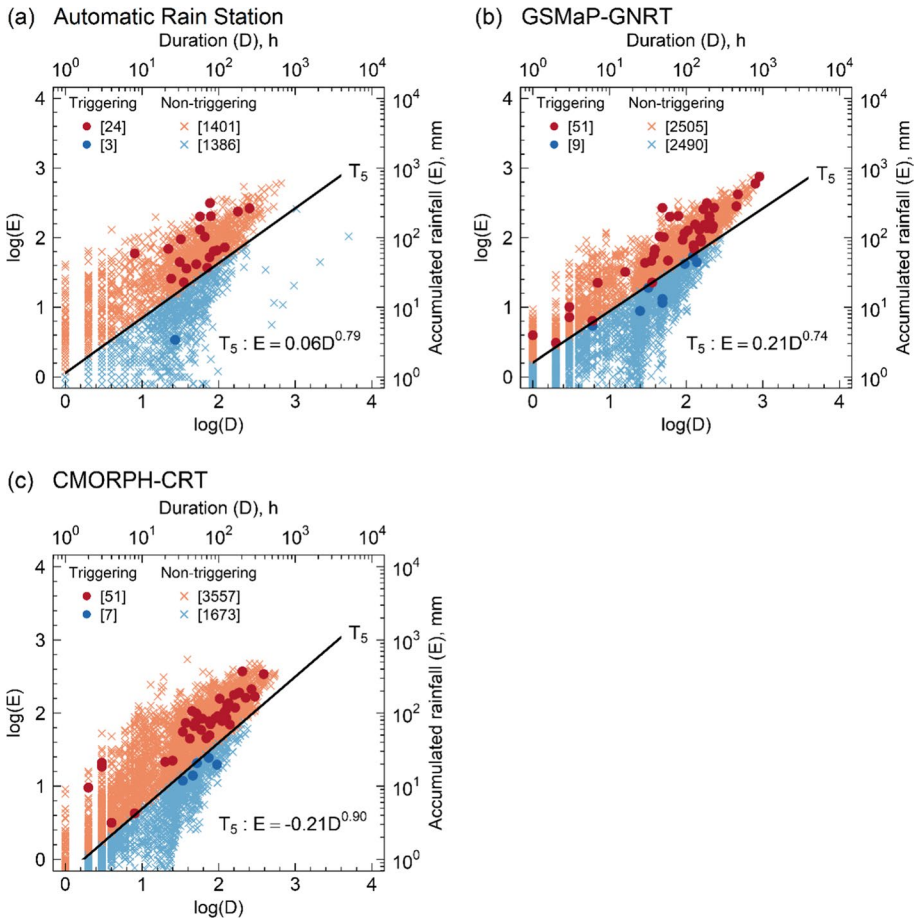
Fig. 5 A sample of rainfall thresholds at a non-exceedance probability level of 5% derived from automatic rain stations (**a**), GSMaP-GNRT (**b**), and CMORPH-CRT (**c**) datasets using the frequentist method. The number inside the square bracket denotes the number of rainfall events (E-D pairs) corresponding to the categories in a contingency table

which is commonly applied (e.g., Leonarduzzi et al. 2017; Brunetti et al. 2018). Figure 7 presents TSS scores at different non-exceedance probability levels of rainfall thresholds for the three datasets. The figure shows that the maximum TSS for automatic rain station data is achieved when the threshold is defined at a non-exceedance probability level of 10%. For GSMaP-GNRT (CMORPH-CRT) data, the highest TSS is obtained by defining the threshold at a non-exceedance probability level of 20% (15%), which also maximizes TPR.

5 Discussion

Comparison of GSMaP-GNRT and CMORPH-CRT rainfall estimates to rain gauge data showed that the performance of these products at hourly time scale is less than excellent despite the bias-corrections using gauge-based analysis data. Both SPPs tend to

Fig. 6 ROC curves built based on TPR and FPR skill scores for three datasets: automatic rain station (blue), GSMaP-GNRT (green), and CMORPH-CRT (orange). Each dot corresponds to the mean value of TPR and FPR skill scores of the repeated samplings. The labels beside the points represent the non-exceedance probability levels of rainfall thresholds
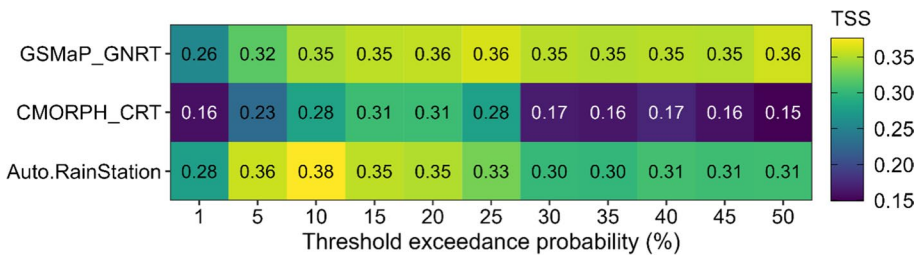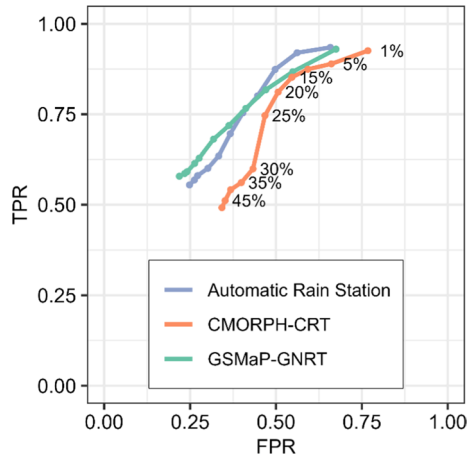

Fig. 7 Values of true skill statistics (TSS) obtained at different non-exceedance probability levels of rainfall thresholds for the three datasets

overestimate low-intensity rain and underestimate high-intensity rain. This result corroborates the findings of previous studies for other regions with different climatic characteristics and varied topography, e.g., in Mexico (Bruster-Flores et al. 2019), Myanmar (Yuan et al. 2019), Ethiopia (Haile et al. 2013), and Tibetian Plateau (Li et al. 2021). The same tendency persists at a daily time scale in different mountainous regions around the world (Derin et al. 2016). These results may be explained partially by the fact that both SPPs correct their rainfall estimates using the NOAA Climate Prediction Center (CPC) daily gauge analyses (Chen et al. 2008), whose accuracy depends on the rain gauge network density (Chen et al. 2008; Barros and Arulraj 2020) and the representativeness of the gauges (Derin et al. 2016). Moreover, both SPPs perform poorly over mountainous areas despite the high spatial resolution and incorporation of an orographic rainfall detection scheme. This result is likely related to the technical limitations of satellite sensors in detecting orographic rainfall regarding ground-clutter effects, heterogeneity in the vertical and horizontal structure of cloud systems, satellite viewing angle, and overpass frequency (Barros and Arulraj 2020).

The comparison results also show a weak correlation between SPP rainfall estimates and gauge observations. One explanation for this weak correlation is that the satellite algorithms failed to capture rainfall events with a duration shorter than or equal to satellite retrieval intervals (Haile et al. 2013). The point-area comparison also contributes to

significant differences since high-intensity rainfall commonly occurs over an area much smaller than the SPP pixel size (Haile et al. 2013).

Another finding is that GSMaP-GNRT failed to capture rain rate > 30 mm/h. This finding reflects that of Turk and Xian (2013), who also found that GSMaP products do not sufficiently capture the rainfall extremes at a sub-daily scale. Turk and Xian (2013) suggested that the failure could be attributed to the effect of Kalman filtering that averages the extreme values.

The above results may indicate that potential bias may arise if hourly rainfall estimates of those SPPs are used as direct input for a hydrological model to simulate hourly hydrological response and potential landslides in the study area. Current studies that evaluated GSMaP products demonstrated that the newer version of the product is not superior to the previous one (Huang et al. 2021; Aryastana et al. 2022). Those studies and the evaluation results presented here highlight the importance of improving the SPP capacity in detecting rainfall (both high intensity and lower intensity) and of using hourly ground observations.

Rainfall events determined from automatic rain stations, GSMaP-GNRT, and CMORPH-CRT that resulted in landslides have a duration of up to ~ 12 days, ~ 38 days, and ~ 16 days, respectively. These results indicate that landslides in Progo Catchment can also be triggered by less intense but prolonged rainfall. von Ruette et al. (2014) suggested that prolonged rainfall resulted in highest landslide activity. On the other hand, the existing LEWS in Indonesia limited the antecedent days up to 3 days to define the threshold (Hidayat et al. 2019; Yuniawan et al. 2022). Therefore, it is possible that the amount of rainfall triggering landslides was underestimated by the existing LEWS, leading to more false alarms.

The analysis of rainfall thresholds showed that the rainfall measured at the stations performed slightly better than the estimated rainfall from GSMaP-GNRT, particularly at a non-exceedance probability level below 15% (Fig. 6). Meanwhile, CMORPH-CRT performs the worst at all non-exceedance probabilities. AUC values indicated that the performance of the thresholds derived using SPPs is comparable to those derived using automatic rain station data in Progo Catchment. However, GSMaP-GNRT outperformed CMORPH-CRT in discriminating the occurrence/non-occurrence of landslide-triggering rainfall events. Brunetti et al. (2021) showed that satellite-derived rainfall products outperformed ground observations for landslide prediction in India. However, their ground observation dataset has a coarser spatiotemporal resolution compared to satellite-derived rainfall products. Therefore, these results support previous studies that suggested a high temporal dataset is more appropriate for determining rainfall threshold for landslide early warning (Brunetti et al. 2018; Gariano et al. 2020).

For early warning purposes, the rainfall threshold is selected based on the non-exceedance probability level at which the TSS value is maximum. For automatic rain station data, TSS is maximum when the threshold is derived at a non-exceedance probability level of 10% (the equation with label T_{10} for automatic rain station dataset presented in Table 4). The threshold at a non-exceedance probability level of 20% is suitable for GSMaP-GNRT (the equation with label T_{20} for the GSMaP-GNRT dataset presented in Table 4). The levels are higher than the level applied by other studies, i.e., 5% (e.g., Vennari et al. 2014; Roccati et al. 2018; Leonarduzzi and Molnar 2020) but within the threshold levels at 10–25% suggested by Brunetti et al. (2018). It is important to bear in mind that this direct comparison might be unfair because of the differences in data sources, resolution, climate, and the coverage area valid for the early warning. To issue a warning, accumulated rainfall at different time intervals is calculated and compared to the threshold to confirm if the value exceeds the threshold. The warning should be regarded as a first-level warning of LEWS since the

location of slope failures likely to occur is unknown. Based on this warning, a follow-up by hydrological experts can be taken by applying a physical-based hydrological model to predict the location (second-level warning).

This study is subject to uncertainties and has limitations. The first source of uncertainty is the landslide inventory we compiled for this study. Considering our sources of landslide information, the inventory is very likely spatial biased toward landslides occurring in populated areas and creating damage to infrastructure. Another source of uncertainties is the representation of area-average hourly rainfall in the grid-scale of each SPP as a point measurement of rain gauge (Apip et al. 2010; Haile et al. 2013). However, this is unavoidable due to the limited number of rain stations in our study area. Because the rainfall thresholds derived here are based on landslide events that occurred in Progo Catchment only, the thresholds might be untransferable to other areas due to differences in the local weather system and landslide preparatory factors of catchments. Moreover, different catchments might have different hydrological responses to various rainfall events (e.g., Marin et al. 2020; Jaramillo-Gonzalez et al. 2023). Nevertheless, given a sufficient inventory of landslides, the approach presented here can be adopted for other areas. A study by Johnston et al. (2021)showed that landslide hazard in urbanized areas is sensitive to rainfall variations. However, in our analysis, we did not distinguish landslide events in urbanized and non-urbanized areas due to the limited number of events and the spatial bias of our inventory. Finally, our datasets have no rainfall forecast value. Nevertheless, considering the relatively short latency, they can be applied for nowcasting (van Natijne et al. 2020; Hartke et al. 2020; Stanley et al. 2021).

6 Conclusion

This study set out to derive and evaluate new rainfall thresholds for landslides in the Progo Catchment using various high-resolution rainfall datasets. The determination of the thresholds involved a frequentist approach. Results show that the performance of GSMaP-GNRT and CMORPH-CRT at an hourly time scale is less than excellent. The poor capacity in correctly detecting hourly rainfall might limit the usage of SPP for hydrological modeling to simulate hydrological response and potential landslides. Nevertheless, when the rainfall estimates were aggregated according to the definition of rainfall events corresponding to landslides in our study area, the results showed that both SPPs could be used to determine rainfall thresholds as the basis for a landslide early warning. The performance of the thresholds derived using SPPs is comparable to those derived using automatic rain station data, particularly GSMaP-GNRT thresholds. For early warning purposes, the suitable threshold is the threshold derived using automatic rain station data at a non-exceedance probability level of 10% or the GSMaP-GNRT threshold at 20%.

Our study is the first attempt to be a statistically rigorous probabilistic approach that has not been carried out before for a region in Indonesia. The results may assist the existing LEWS, which applies a fixed threshold limited to up to 3 days of rainfall amount. Conversely, we offer a dynamic window of rainfall events to determine rainfall thresholds that show a range of possible rainfall conditions in terms of accumulation and duration that trigger the landslides in the Progo Catchment. The dynamic time window, applied to hourly rainfall data, implicitly accounts for rainfall intermittency and antecedent rainfall that influence predisposing conditions of landslides.

The major limitation of this study is that the datasets used to determine rainfall thresholds have no prediction value. Further study should be done using high-resolution numerical weather prediction (NWP) models that generate weather forecasts. Since both satellite products and NWP output have their uncertainties, it is unlikely to directly translate the thresholds derived here to the thresholds for NWP output. Therefore, a technique such as bias correction for adjusting NWP model output to ensure that its statistical properties are coherent with observations (e.g., Rogelis and Werner 2018; Jabbari and Bae 2020; Schauwecker et al. 2021) needs to be implemented before applying the model output into an operational early warning system. Since the NWP model is sensitive to the choice of physical parameter schemes (e.g., Umer et al. 2021), it is essential to assess the performance skills of the model in forecasting large/mesoscale weather systems producing the rainfall events associated with the landslides in the study area. Such studies eventually could help advance the landslide early warning system.

Appendix

Variables and acronyms used in text.

Variables	Description
AUC	Area under an ROC curve
BIAS	Bias score
CC	Correlation coefficient
CSI	Critical success index
D	Duration of rainfall event
E	Accumulated rainfall
FA	False alarm
FN	False negative
FP	False positive
FPR	False positive rate
H	Hit
M	Miss
MAE	Mean absolute error
O	Rainfall observed by automatic rain station
POD	Probability of detection
RB	Relative bias
S	Satellite-derived rainfall estimates
SR	Success ratio
TN	True negative
TP	True positive
TPR	True positive rate
TSS	True skill statistics
Acronym	Description
BMKG	Badan Meteorologi Klimatologi dan Geofisika (Indonesian Agency for Meteorology, Climatology and Geophysics)

Acronym	Description
BNPB	Badan Nasional Penanggulangan Bencana (National Agency for Disaster Management of Indonesia)
BPBD	Badan Penanggulangan Bencana Daerah (Regional Agency for Disaster Management)
CMORPH	CPC/NOAA morphing technique
CMORPH-CRT	Bias-corrected version of CMORPH
CPC	Climate Prediction Center
EWS	Early Warning System
GSMaP	Global Satellite Mapping of Precipitation
GSMaP-GNRT	gauge-adjusted version of the Global Satellite Mapping of Precipitation
JAXA	Japan Aerospace Exploration Agency
LEWS	Landslide Early Warning System
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
ROC	Receiver operating characteristic
SPP	Satellite-based precipitation products
SWCEM	Space-based Weather and Climate Extremes Monitoring
TRMM	Tropical Rainfall Measuring Mission
UTC	Coordinated Universal Time
WMO	World Meteorological Organization

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Aldrian E, Susanto RD (2003) Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. Int J Climatol 23:1435–1452. https://doi.org/10.1002/joc.950

Althuwaynee OF, Pradhan B, Ahmad N (2015) Estimation of rainfall threshold and its use in landslide hazard mapping of Kuala Lumpur metropolitan and surrounding areas. Landslides 12:861–875. https://doi.org/10.1007/s10346-014-0512-y

Amin TC, Ratman N, Gafoer S (1999) Geological Map of Middle Part of Jawa

Apip TK, Yamashiki Y et al (2010) A distributed hydrological—geotechnical model using satellite-derived rainfall estimates for shallow landslide prediction system at a catchment scale. Landslides 7:237–258. https://doi.org/10.1007/s10346-010-0214-z

Aryastana P, Liu C, Jong-Dao Jou B et al (2022) Assessment of satellite precipitation datasets for high variability and rapid evolution of typhoon precipitation events in the Philippines. Earth Sp Sci. https://doi.org/10.1029/2022ea002382

Barros AP, Arulraj M (2020) Remote sensing of orographic precipitation. Adv Glob Change Res 69:559–582. https://doi.org/10.1007/978-3-030-35798-6_6

BMKG (2017) Siklon Tropis "CEMPAKA" Lahir, Siaga Cuaca Ekstrem 3 Hari Ke Depan. https://www.bmkg.go.id/berita/?p=siklon-tropis-cempaka-waspadai-hujan-lebat-disertai-angin-kencang-dan-gelombang-tinggi-di-wilayah-selatan-indonesia&tag=berita-utama&lang=ID. Accessed 30 Oct 2020

BMKG (2022) AWS (Automatic Weather Station). https://awscenter.bmkg.go.id/. Accessed 1 Mar 2022

BNPB (2021) Data Informasi Bencana Indonesia. http://dibi.bnpb.go.id/. Accessed 3 Feb 2021

Brunetti MT, Peruccacci S, Rossi M et al (2010) Rainfall thresholds for the possible occurrence of landslides in Italy. Nat Hazards Earth Syst Sci 10:447–458. https://doi.org/10.5194/nhess-10-447-2010

Brunetti MT, Melillo M, Peruccacci S et al (2018) How far are we from the use of satellite rainfall products in landslide forecasting? Remote Sens Environ 210:65–75. https://doi.org/10.1016/j.rse.2018.03.016

Brunetti MT, Melillo M, Gariano SL et al (2021) Satellite rainfall products outperform ground observations for landslide forecasting in India. Hydrol Earth Syst Sci Discuss. https://doi.org/10.5194/hess-2021-42

Bruster-Flores JL, Ortiz-Gómez R, Ferriño-Fierro AL et al (2019) Evaluation of precipitation estimates CMORPH-CRT on regions of Mexico with different climates. Water (switzerland) 11:1–15. https://doi.org/10.3390/w11081722

Cepeda J, Smebye H, Vangelsten B, et al (2010) Landslide risk in Indonesia Global Assessment Report on Disaster Risk Reduction. ISDR. Available via PreventionWeb. https://www.preventionweb.net/english/hyogo/gar/2011/en/bgdocs/Cepeda_et_al._2010.pdf. Accessed 3 Feb 2021

Chen M, Shi W, Xie P et al (2008) Assessing objective techniques for gauge-based analyses of global daily precipitation. J Geophys Res 113:D04110. https://doi.org/10.1029/2007JD009132

Chikalamo EE, Mavrouli OC, Ettema J et al (2020) Satellite-derived rainfall thresholds for landslide early warning in Bogowonto Catchment, Central Java, Indonesia. Int J Appl Earth Obs Geoinf 89:102093. https://doi.org/10.1016/j.jag.2020.102093

Derin Y, Anagnostou E, Berne A et al (2016) Multiregional satellite precipitation products evaluation over complex terrain. J Hydrometeorol 17:1817–1836. https://doi.org/10.1175/JHM-D-15-0197.1

Dunkerley D (2015) Intra-event intermittency of rainfall: an analysis of the metrics of rain and no-rain periods. Hydrol Process 29:3294–3305. https://doi.org/10.1002/hyp.10454

Dzakiya N, Laksmana RB, Hidayah RA (2023) Identifying probable slip surface in Wanurejo, Borobudur District, by subsurface analysis utilizing the dipole–dipole configuration of resistivity method. J Appl Geospatial Inf 7:739–746. https://doi.org/10.30871/jagi.v7i1.5055

Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27:861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Gariano SL, Melillo M, Peruccacci S, Brunetti MT (2020) How much does the rainfall temporal resolution affect rainfall thresholds for landslide triggering? Nat Hazards 100:655–670. https://doi.org/10.1007/s11069-019-03830-x

Geological Agency of Indonesia (2022) GeoMap. https://geologi.esdm.go.id/geomap

Guzzetti F, Peruccacci S, Rossi M, Stark CP (2007) Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorol Atmos Phys 98:239–267. https://doi.org/10.1007/s00703-007-0262-7

Guzzetti F, Peruccacci S, Rossi M, Stark CP (2008) The rainfall intensity-duration control of shallow landslides and debris flows: an update. Landslides 5:3–17. https://doi.org/10.1007/s10346-007-0112-1

Guzzetti F, Gariano SL, Peruccacci S et al (2020) Geographical landslide early warning systems. Earth-Sci Rev 200:102973. https://doi.org/10.1016/j.earscirev.2019.102973

Hadmoko DS, Lavigne F, Sartohadi J et al (2010) Landslide hazard and risk assessment and their application in risk management and landuse planning in eastern flank of Menoreh Mountains, Yogyakarta Province, Indonesia. Nat Hazards 54:623–642. https://doi.org/10.1007/s11069-009-9490-0

Hadmoko DS, Lavigne F, Samodra G (2017) Application of a semiquantitative and GIS-based statistical model to landslide susceptibility zonation in Kayangan Catchment, Java, Indonesia. Nat Hazards 87:437–468. https://doi.org/10.1007/s11069-017-2772-z

Haile AT, Habib E, Rientjes T (2013) Evaluation of the climate prediction center (CPC) morphing technique (CMORPH) rainfall product on hourly time scales over the source of the Blue Nile River. Hydrol Process 27:1829–1839. https://doi.org/10.1002/hyp.9330

Hanssen AW, Kuipers WJA (1965) On the relationship between the frequency of rain and various meteorological parameters: with reference to the problem of objective forecasting, mededeling. Koninklijk Nederlands Meteorologisch Instituut, De Bilt

Hapsari RI, Oishi S, Syarifuddin M et al (2019) X-MP radar for developing a lahar rainfall threshold for the merapi volcano using a bayesian approach. J Disaster Res 14:811–828. https://doi.org/10.20965/jdr.2019.p0811

Hartke SH, Wright DB, Kirschbaum DB et al (2020) Incorporation of satellite precipitation uncertainty in a landslide hazard nowcasting system. J Hydrometeorol 21:1741–1759. https://doi.org/10.1175/JHM-D-19-0295.1

Hidayat R, Sutanto SJ, Hidayah A et al (2019) Development of a landslide early warning system in Indonesia. Geosciences 9:451. https://doi.org/10.3390/geosciences9100451

Hong M, Kim J, Jeong S (2018) Rainfall intensity-duration thresholds for landslide prediction in South Korea by considering the effects of antecedent rainfall. Landslides 15:523–534. https://doi.org/10.1007/s10346-017-0892-x

Hosmer DW, Lemeshow S, Sturdivant RX (2013) Applied logistic regression, 3rd edn. Wiley, New York

Huang WR, Liu PY, Hsu J et al (2021) Assessment of near-real-time satellite precipitation products from GSMaP in monitoring rainfall variations over Taiwan. Remote Sens 13:1–17. https://doi.org/10.3390/rs13020202

Huffman GJ, Adler RF, Bolvin DT et al (2007) The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J Hydrometeorol 8:38–55. https://doi.org/10.1175/JHM560.1

Irawan AM, Virgianto RH, Safril A et al (2019) Rainfall threshold and soil moisture indexes for the initiation of landslide in Banjarmangu sub-district, central Java, Indonesia. IOP Conf Ser Earth Environ Sci. https://doi.org/10.1088/1755-1315/243/1/012028

Jabbari A, Bae DH (2020) Improving ensemble forecasting using total least squares and lead-time dependent bias correction. Atmosphere (basel). https://doi.org/10.3390/atmos11030300

Jaramillo-Gonzalez R, Aristizábal-Giraldo E, García-Aristizábal E, Marin RJ (2023) Physically-based models applied to rainfall thresholds for shallow landslides: literature review. Rev La Asoc Geol Argent 80:164–178

Johnston EC, Davenport FV, Wang L et al (2021) Quantifying the effect of precipitation on landslide hazard in urbanized and non-urbanized areas. Geophys Res Lett 48:e2021GL094038

Joyce RJ, Janowiak JE, Arkin PA, Xie P (2004) CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. J Hydrometeorol 5:487–503. https://doi.org/10.1175/1525-7541(2004)005%3c0487:CAMTPG%3e2.0.CO;2

Kahle D, Wickham H (2013) ggmap: Spatial visualization with ggplot2. R J 5:144–161. https://doi.org/10.32614/rj-2013-014

Katriani L, Darmawan D, Ruwanto B et al (2019) Identification of subsurface lithology in Sendang Mulyo, Purwoharjo Village, Samigaluh Subdistrict, Kulon Progo Regency. J Phys Conf Ser 1397:012008. https://doi.org/10.1088/1742-6596/1397/1/012008

Kidd C, Becker A, Huffman GJ et al (2017) So, how much of the Earth's surface is covered by rain gauges? Bull Am Meteorol Soc 98:69–78. https://doi.org/10.1175/BAMS-D-14-00283.1

Kirschbaum D, Stanley T, Zhou Y (2015) Spatial and temporal analysis of a global landslide catalog. Geomorphology 249:4–15. https://doi.org/10.1016/j.geomorph.2015.03.016

Kubota T, Aonashi K, Ushio T et al (2020) Global Satellite Mapping of Precipitation (GSMaP) Products in the GPM Era. In: Levizzani V, Kidd C, Kirschbaum DB et al (eds) Satellite precipitation measurement. Advances in global change research. Springer, Cham, pp 355–373

Kuleshov Y, Kurino T, Kubota T et al (2019) WMO space-based weather and climate extremes monitoring demonstration project (semdp): first outcomes of regional cooperation on drought and heavy precipitation monitoring for Australia and Southeast Asia. In: Abbot J, Hammond A (eds) Rainfall—extremes, distribution and properties. IntechOpen, London

Lagomarsino D, Segoni S, Fanti R, Catani F (2013) Updating and tuning a regional-scale landslide early warning system. Landslides 10:91–97. https://doi.org/10.1007/s10346-012-0376-y

Leonarduzzi E, Molnar P (2020) Deriving rainfall thresholds for landsliding at the regional scale: daily and hourly resolutions, normalisation, and antecedent rainfall. Nat Hazards Earth Syst Sci 20:2905–2919. https://doi.org/10.5194/nhess-20-2905-2020

Leonarduzzi E, Molnar P, McArdell BW (2017) Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data. Water Resour Res 53:6612–6625. https://doi.org/10.1002/2017WR021044

Li Q, Wei J, Yin J et al (2021) Multiscale comparative evaluation of the GPM and TRMM precipitation products against ground precipitation observations over Chinese Tibetan plateau. IEEE J Sel Top Appl Earth Obs Remote Sens 14:2295–2313. https://doi.org/10.1109/JSTARS.2020.3047897

Marin RJ, García EF, Aristizábal E (2020) Effect of basin morphometric parameters on physically-based rainfall thresholds for shallow landslides. Eng Geol 278:105855. https://doi.org/10.1016/j.enggeo.2020.105855

Marra F (2019) Rainfall thresholds for landslide occurrence: systematic underestimation using coarse temporal resolution data. Nat Hazards 95:883–890. https://doi.org/10.1007/s11069-018-3508-4

NCAR (2015) verification: Weather forecast verification utilities. https://cran.r-project.org/package=verification

Nikolopoulos EI, Crema S, Marchi L et al (2014) Impact of uncertainty in rainfall estimation on the identification of rainfall thresholds for debris flow occurrence. Geomorphology 221:286–297. https://doi.org/10.1016/j.geomorph.2014.06.015

Peruccacci S, Brunetti MT, Luciani S et al (2012) Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central Italy. Geomorphology 139–140:79–90. https://doi.org/10.1016/J.GEOMORPH.2011.10.005

Peruccacci S, Brunetti MT, Gariano SL et al (2017) Rainfall thresholds for possible landslide occurrence in Italy. Geomorphology 290:39–57. https://doi.org/10.1016/j.geomorph.2017.03.031

Piciullo L, Gariano SL, Melillo M et al (2017) Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. Landslides 14:995–1008. https://doi.org/10.1007/s10346-016-0750-2

Pratiwi ES, Sartohadi J (2019) Geoelectrical prediction for sliding plane layers of rotational landslide at the volcanic transitional landscapes in Indonesia. IOP Conf Ser Earth Environ Sci 286:012028. https://doi.org/10.1088/1755-1315/286/1/012028

Rauniyar SP, Protat A, Kanamori H (2017) Uncertainties in TRMM-Era multisatellite-based tropical rainfall estimates over the Maritime Continent. Earth Sp Sci. https://doi.org/10.1002/2017EA000279

Robbins JC (2016) A probabilistic approach for assessing landslide-triggering event rainfall in Papua New Guinea, using TRMM satellite precipitation estimates. J Hydrol 541:296–309. https://doi.org/10.1016/j.jhydrol.2016.06.052

Roccati A, Faccini F, Luino F et al (2018) Rainfall events with shallow landslides in the Entella catchment, Liguria, northern Italy. Nat Hazards Earth Syst Sci 18:2367–2386. https://doi.org/10.5194/nhess-18-2367-2018

Roebber PJ (2009) Visualizing multiple measures of forecast quality. Weather Forecast 24:601–608. https://doi.org/10.1175/2008WAF2222159.1

Rogelis MC, Werner M (2018) Streamflow forecasts from WRF precipitation for flood early warning in mountain tropical areas. Hydrol Earth Syst Sci 22:853–870. https://doi.org/10.5194/hess-22-853-2018

Sartohadi J, Samodra G, Hadmoko DS (2010) Landslide susceptibility assessment using heuristic statistically Method in Kayangan Catchment Kulon Progo Yogyakarta-Indonesia. Int J Geoinformat 6:23–28

Schauwecker S, Schwarb M, Rohrer M, Stoffel M (2021) Heavy precipitation forecasts over Switzerland—an evaluation of bias-corrected ECMWF predictions. Weather Clim Extrem 34:100372. https://doi.org/10.1016/j.wace.2021.100372

Segoni S, Rosi A, Rossi G et al (2014) Analysing the relationship between rainfalls and landslides to define a mosaic of triggering thresholds for regional-scale warning systems. Nat Hazards Earth Syst Sci 14:2637–2648. https://doi.org/10.5194/nhess-14-2637-2014

Segoni S, Piciullo L, Gariano SL (2018) A review of the recent literature on rainfall thresholds for landslide occurrence. Landslides 15:1483–1501. https://doi.org/10.1007/s10346-018-0966-4

Staley DM, Kean JW, Cannon SH et al (2013) Objective definition of rainfall intensity-duration thresholds for the initiation of post-fire debris flows in southern California. Landslides 10:547–562. https://doi.org/10.1007/s10346-012-0341-9

Stanley TA, Kirschbaum DB, Benz G et al (2021) Data-driven landslide nowcasting at the global scale. Front Earth Sci 9:1–15. https://doi.org/10.3389/feart.2021.640043

Turk FJ, Xian P (2013) An assessment of satellite-based high resolution precipitation datasets for atmospheric composition studies in the maritime continent. Atmos Res 122:579–598. https://doi.org/10.1016/j.atmosres.2012.02.017

Umer Y, Ettema J, Jetten V et al (2021) Evaluation of the WRF model to simulate a high-intensity rainfall event over Kampala, Uganda. Water (switzerland) 13:1–25. https://doi.org/10.3390/w13060873

Ushio T, Sasashige K, Kubota T et al (2009) A kalman filter approach to the global satellite mapping of precipitation (GSMaP) from combined passive microwave and infrared radiometric data. J Meteorol Soc Jpn 87A:137–151. https://doi.org/10.2151/jmsj.87A.137

van Natijne AL, Lindenbergh RC, Bogaard TA (2020) Machine learning: new potential for local and regional deep-seated landslide nowcasting. Sensors (switzerland) 20:1–18. https://doi.org/10.3390/s20051425

Vennari C, Gariano SL, Antronico L et al (2014) Rainfall thresholds for shallow landslide occurrence in Calabria, southern Italy. Nat Hazards Earth Syst Sci 14:317–330. https://doi.org/10.5194/nhess-14-317-2014

von Ruette J, Lehmann P, Or D (2014) Effects of rainfall spatial variability and intermittency on shallow landslide triggering patterns at a catchment scale. Water Resour Res 50:7780–7799. https://doi.org/10.1002/2013WR015122

Wang N, Cheng W, Lombardo L et al (2021) Statistical spatiotemporal analysis of hydro-morphological processes in China during 1950–2015. Stoch Environ Res Risk Assess 2021:1–21. https://doi.org/10.1007/S00477-021-02007-Y

Widyadarsana SN, Hartantyo E (2021) Lithological modelling based on shear wave velocity using horizontal to vertical spectral ratio (HVSR) inversion of ellipticity curve method to mitigate landslide hazards at the main road Of Samigaluh District, Kulon Progo Regency, Yogyakarta. E3S Web Conf 325. https://doi.org/10.1051/e3sconf/202132501009

Wilks DS (1995) Statistical methods in the atmospheric sciences. Academic Press, Cambridge

Xie P, Joyce R, Wu S et al (2017) Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. J Hydrometeorol 18:1617–1641. https://doi.org/10.1175/JHM-D-16-0168.1

Yuan F, Zhang L, Soe KMW et al (2019) Applications of TRMM- and GPM-era multiple- satellite precipitation products for flood simulations at sub-daily scales in a sparsely gauged watershed in Myanmar. Remote Sens. https://doi.org/10.3390/rs11020140

Yuniawan RA, Rifa'i A, Faris F et al (2022) Revised rainfall threshold in the Indonesian landslide early warning system. Geosciences 12:129. https://doi.org/10.3390/geosciences12030129

Zhou Z, Guo B, Xing W et al (2020) Comprehensive evaluation of latest GPM era IMERG and GSMaP precipitation products over mainland China. Atmos Res 246:105132. https://doi.org/10.1016/j.atmosres.2020.105132

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.