Received: 23 November 2020 | Revised: 8 May 2021 | Accepted: 28 May 2021

DOI: 10.1002/hyp.14260

WILEY

IMPACTS OF OBSERVATIONAL UNCERTAINTY ON ANALYSIS AND MODELLING OF HYDROLOGICAL PROCESSES

A multi-sensor evaluation of precipitation uncertainty for landslide-triggering storm events

Elsa S. Culler[1,2] | Andrew M. Badger[3,4] | Justin Toby Minear[2] | Kristy F. Tiampo[2,5] | Spencer D. Zeigler[5] | Ben Livneh[1,2]

[1]Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, Colorado, USA

[2]Cooperative Institute for Research in Environmental Science (CIRES), University of Colorado Boulder, Boulder, Colorado, USA

[3]Universities Space Research Association, Columbia, Maryland, USA

[4]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

[5]Department of Geological Sciences, University of Colorado Boulder, Boulder, Colorado, USA

Correspondence
Elsa S. Culler, Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, 216 UCB, Boulder, CO 80309, USA.
Email: elsa.culler@colorado.edu

Funding information
National Aeronautics and Space Administration, Grant/Award Number: 16-IDS16-0075

Abstract

Extreme precipitation can have profound consequences for communities, resulting in natural hazards such as rainfall-triggered landslides that cause casualties and extensive property damage. A key challenge to understanding and predicting rainfall-triggered landslides comes from observational uncertainties in the depth and intensity of precipitation preceding the event. Practitioners and researchers must select from a wide range of precipitation products, often with little guidance. Here we evaluate the degree of precipitation uncertainty across multiple precipitation products for a large set of landslide-triggering storm events and investigate the impact of these uncertainties on predicted landslide probability using published intensity–duration thresholds. The average intensity, peak intensity, duration, and NOAA-Atlas return periods are compared ahead of 177 reported landslides across the continental United States and Canada. Precipitation data are taken from four products that cover disparate measurement methods: near real-time and post-processed satellite (IMERG), radar (MRMS), and gauge-based (NLDAS-2). Landslide-triggering precipitation was found to vary widely across precipitation products with the depth of individual storm events diverging by as much as 296 mm with an average range of 51 mm. Peak intensity measurements, which are typically influential in triggering landslides, were also highly variable with an average range of 7.8 mm/h and as much as 57 mm/h. The two products more reliant upon ground-based observations (MRMS and NLDAS-2) performed better at identifying landslides according to published intensity–duration storm thresholds, but all products exhibited hit ratios of greater than 0.56. A greater proportion of landslides were predicted when including only manually verified landslide locations. We recommend practitioners consider low-latency products like MRMS for investigating landslides, given their near-real time data availability and good performance in detecting landslides. Practitioners would be well-served considering more than one product as a way to confirm intense storm signals and minimize the influence of noise and false alarms.

KEYWORDS
extreme precipitation, intensity–duration thresholds, natural hazards, precipitation intercomparison, rainfall-triggered landslides

1 | INTRODUCTION

Precipitation measurements and their uncertainties play a key role in understanding and mitigating rainfall-triggered landslides because they drive excess runoff and soil saturation that initiate these natural disasters (Highland & Bobrowsky, 2008). In spite of the destructive nature of landslides, which cause tens of thousands of deaths each year (Froude & Petley, 2018) these events remain challenging to diagnose, in part due to uncertainty in antecedent precipitation amounts (Kirschbaum & Stanley, 2018). There are many other sources of uncertainty that contribute to poor landslide diagnosis and prediction, such as unknown soil properties, vegetation, and anthropogenic modifications to surface and subsurface soil structure. However, perhaps the largest source of uncertainty in estimating landslide probability is hydrologic uncertainty, defined here as uncertainty in the depth and intensity of liquid precipitation leading up to the event (Chowdhury & Flentje, 2002). A confounding factor is the wide array of precipitation datasets ranging from in situ observations, ground-based radar and satellite retrievals. This study compares precipitation from various sources immediately preceding landslide events to assess the implications of precipitation uncertainty for evaluating landslide hazards.

The precipitation products chosen for this inter-comparison represent three broad categories of primary measurement techniques: precipitation gauges, ground-based radar, and microwave satellite. Precipitation gauges operate by periodically measuring the volume of precipitation collected at a gauge. Their main strength is that they directly measure the amount of collected water, but nonetheless they can suffer from issues of persistent bias driven by under-catch from wind (Pollock et al., 2018) instrument malfunctions (Duchon et al., 2014; Duchon & Biddle, 2010), gauges placed too close to other structures (Vose et al., 2014), and limited spatial representativeness due to sparse sensor density (Kidd et al., 2017). Despite these limitations, multiple studies (Maggioni et al., 2016; Sapiano & Arkin, 2009) have found gauge-based data products to have lower biases than satellite-based products over North America, especially in areas with complex topography including those where many of the landslides included in this study took place.

In contrast, ground-based radar detects precipitation indirectly using the backscatter of radar and can measure subtle variations in precipitation over regions of several hundreds of square kilometres (Zhang et al., 2015). Since ground-based radar is an indirect measurement of precipitation, its performance is dependent on skilful conversion of the radar signal to precipitation volume. Beam blockage and interference from buildings or even insects in the radar's path are another limitation (Bousquet & Smull, 2003; Fornasiero et al., 2004; Nikahd et al., 2016). Most ground-based radars use multiple bands of radar and multiple polarities to compute the raindrop shape and size distributions used in the processing and limit the impact of known sources of error, which offers an advantage over other indirect techniques such as some of those incorporated into satellite-based measurements (Chandrasekar et al., 2008).

Satellite techniques vary in terms of which sensors they use to detect precipitation, including active and passive microwave, infrared, radar, or any combination. Depending on the sensor type these satellites are deployed in either geostationary or low Earth orbits that cover particular spatial regions at particular intervals (Huffman et al., 2020). The key advantage of satellite-based precipitation measurements over ground-based in situ or radar measurements is that they can deliver frequent and spatially continuous measurements, although multiple satellites (Tapiador et al., 2012) with a variety of sensors and orbits (Ashouri et al., 2015) are required to provide global coverage. For example, the satellite products used in this analysis incorporate a fleet of geostationary satellites in addition to a single low Earth orbit reference satellite (Kidd et al., 2020). Many of the challenges associated with satellite-based precipitation measurement are related to sensor calibration and bias-correction relative to ground-based measurements (Ebert, 2007), as well as the development of algorithms for merging measurements from diverse sources (Huffman et al., 2007; Skofronick-Jackson et al., 2017). Estimating rain drop size distributions also presents a challenge when measuring precipitation with satellite-based microwave instruments, though it can be at least partially addressed through the use of either ground- or satellite-based radar.

Existing precipitation products have been compared and evaluated using a number of metrics in prior studies, for example annual and monthly totals (Adler et al., 2001) or the frequency of wet or dry days (Manzanas et al., 2014). Less attention has been paid to metrics most directly useful for analysing rainfall-triggered landslides. While many mudslides or debris flows are triggered by short, intense precipitation events, other landslides such as shallow slope failures or soil slips are triggered by saturation of the soil column that can develop over a longer period of time (Cannon & Gartner, 2005). In both cases the triggering event occurs over the course of hours or days rather than months or years, and for some landslides the critical time period may be less than an hour.

Published inter-comparisons of precipitation products based on different measurement techniques typically focus on specific applications such as evaluating grid-based products over complex terrain, portraying hydrologic phenomena (Ahmadalipour & Moradkhani, 2017), climate model downscaling efforts (Gutmann et al., 2014; Wang et al., 2020), or for merging data from multiple sensors together (Beck et al., 2017). A general review of 30 gauge-based, satellite-based, and reanalysis global precipitation products by Sun et al. (2018) compared systematic and random errors for daily and annual precipitation, reporting large disagreements even within the same class of product, that is, a deviation of 300 mm in annual precipitation over global land among satellite products. They conclude that the placement and spatial density of gauges accounts for many of the errors in gauge-based or gauge-corrected products, further suggesting that cross validation across multiple datasets is crucial to account for errors. Adler et al. (2003) similarly analysed 31 gauge-based, satellite-based, model-based, and climatological datasets in terms of monthly precipitation, finding that 'quasi-standard' products, for example, those like the global precipitation measurement (GPM) mission (Hou et al., 2014) that have undergone substantial testing, perform better. Additionally, they report that products incorporating both in situ and satellite information (e.g., the Global Precipitation Climatology Project) perform better than products based on a single data source.

Fewer studies comparing extreme precipitation events (e.g., events above the 90th percentile) exist. Many focus on climate model simulations (Sunyer et al., 2015; Tryhorn & DeGaetano, 2011) and trends (Bao et al., 2017; Janssen et al., 2014) while others evaluate observations and satellite data (AghaKouchak et al., 2011; Lockhoff et al., 2014; Pendergrass & Knutti, 2018). AghaKouchak et al. (2011) compared extreme precipitation across four satellite platforms and found trade-offs across products in terms of correct identification of precipitation above a threshold and measurements of the volume of extreme storms. While they showed that some datasets performed better than others in certain contexts, they ultimately concluded that no single precipitation product was ideal for detecting extremes because all of them failed to detect certain storms in certain regions. Lockhoff et al. (2014) found that satellite retrieved extreme precipitation values generally matched station-based precipitation when using fuzzy metrics to evaluate agreement at larger spatiotemporal scales of ~330 km and 5 days. Pendergrass and Knutti (2018) showed that precipitation was less temporally variable in coarser versus finer-resolution satellite precipitation datasets, suggesting that coarser precipitation products may be unable to capture extreme precipitation to the same extent as higher resolution datasets. Other studies primarily evaluated extreme precipitation indicators like 90th percentile daily precipitation, extreme one-day precipitation and maximum number of consecutive wet days (Amitai et al., 2012; Manzanas et al., 2014). These measures are meant to capture large storms that happen on at least an annual basis rather than storms that trigger natural disasters (Manzanas et al., 2014; Sun et al., 2018).

This works builds on the handful of studies that have specifically evaluated multiple precipitation products in the context of landslide triggering (Brunetti et al., 2018; Chikalamo et al., 2020; Rossi et al., 2017; Tajudin et al., 2020). For example, Rossi et al. (2017) compared satellite and gauge precipitation data preceding landslide events in Italy, using intensity–duration thresholds as a part of the comparison. They found that data from tropical rainfall measuring mission (TRMM) satellite products (Kummerow et al., 1998) tend to underestimate gauge data, particularly in mountainous areas where landslides are most likely to occur. Brunetti et al. (2018) similarly found that satellite precipitation from four products tended to underestimate rainfall relative to ground based observations. Both studies ultimately concluded that the satellite data could still be useful for forecasting landslides as long as issues of local bias could be accounted for.

The intensity–duration threshold is a type of two-parameter statistical model used for landslide early warning systems, where rainstorms plotting above the threshold curve are predicted to cause landslides (Scheevel et al., 2017). The curves are typically based on a power law (e.g., $I = aD^{-b}$) of average precipitation intensity during a window of time (I) as a function of duration (D) with fitted parameters a and b. These power laws are valid in a particular region or climate and for a range of durations depending on the training data (Guzzetti et al., 2008). Other statistical rainfall thresholds have been proposed, but generally rely upon either intensity or duration or both (Galanti et al., 2018; Leonarduzzi et al., 2017.

The focus of this analysis is to quantify precipitation uncertainty associated with known historical landslides, and to examine the role of this uncertainty in modelling landslide hazards. Given the wide-ranging issues associated with precipitation estimation cited above, this study presents a multi-product, multi-site analysis focused on landslide-triggering storms. We address an existing gap in evaluating extreme precipitation through the lens of rainfall-triggered landslide hazards, while conducting inter-product analyses into storm characteristics of potential relevance for the hydrological community. The inter-comparison includes four precipitation products including those based primarily in gauge, radar, and satellite measurements. We thereby advance the analyses by Rossi et al. (2017) and Brunetti et al. (2018) comparing satellite and gauge products in the evaluation of landslide hazards, by additionally including ground-based radar and by rigorously analysing each precipitation estimate preceding specific landslide events. Finally, when precipitation is within warning systems or guide recovery efforts from landslides, the timeliness, that is, low latency, of the information matters (Kirschbaum et al., 2012), such that the issue of latency will also be considered in the investigation of intensity–duration thresholds. Greater understanding of the areas of relative agreement and any divergence across products may provide guidance to practitioners and researchers choosing among precipitation products for studying landslides.

2 | METHODS

We compared precipitation characteristics at known landslide sites across the features of triggering storms, as well as relative to intensity–duration thresholds of landslide occurrence. Rainfall-triggered landslide sites were chosen from the NASA Global Landslide Catalogue (GLC; Kirschbaum et al., 2010) with a subset of landslide locations verified with ancillary satellite imagery (see Section 2.1). For each landslide location, precipitation was obtained from four different products (see Section 2.2) and the precipitation time series were split into individual storm events. For each storm, key characteristics of total depth, duration, intensity, peak intensity, and return period were calculated (Section 2.3). Finally, the storm events were plotted relative to landslide intensity–duration curves, with hit ratios (true positive rate) and false alarm ratios (false positive rate) compared for each model-product combination (Section 2.4).

2.1 | Study domain and landslide site selection

The NASA Global Landslide Catalog (GLC; Kirschbaum et al., 2010) was chosen as the source of landslide locations for this study, since it provides a large sample of rainfall-triggered landslide locations useful for evaluating heavy rainfall events. The GLC shares many strengths and weaknesses with other regional and global databases available (Kirschbaum et al., 2010; Mirus et al., 2020). Though the GLC covers a broad spatial and temporal domain, it suffers from problems of precision and completeness. The catalogue is comprised of a collection of second-hand landslide reports made by organizations like the news media, governmental organizations like departments of transportation,

along with available scientific reports. This means that landslides that nearby infrastructure and people are reported more frequently, resulting in a substantial spatial bias towards populated areas. Landslide location accuracies range from 'exact' locations, to location uncertainties between 1 km up to 50 km, depending on how specific the source article was about the location (Kirschbaum et al., 2010). Despite these limitations, the GLC was deemed fit for the purposes of this study, which is not to study landslide mechanisms and spatial distribution, but rather to compare precipitation data from different products in the vicinity of hydrologically triggered landslides where heavy rainfall events are likely to be present. Overall, the GLC provided a substantial number of landslide locations ($n = 177$) for this study that met the following selection criteria:

- Only landslide events reported as rainfall-driven, with a GLC trigger category of 'rain', 'downpour', 'continuous_rain', or 'flooding' were included. Though all the events in the GLC are rainfall-triggered, some of the designated triggering mechanisms are more complex, involving other factors such as earthquakes or dam embankment collapse. In particular, landslides with snow- and freeze–thaw-related triggers were not included even though these are hydrologically driven, because their precipitation is not contemporaneously linked with landslide triggers.
- The GLC contains many types of mass movements. The term 'landslide' in particular is somewhat ambiguous, and is used by the catalogue when no additional information was available to classify the mass movement. We have eliminated categories that are suggestive of deep-seated landslides such as 'translational slide', 'rotational slide', or 'complex', in an effort to ensure that the analysed landslide categories are suited to this analysis.
- Landslide events took place in the continental United States (CONUS) or Canada below 60°N and after May 2015 ensuring data availability across each of the selected precipitation products.
- The landslide location accuracy was reported to be 10 km or less. The value of 10 km was chosen since it is approximately equal to the spatial resolution of two of the precipitation products.
- The landslide size was reported as 'medium' or larger so as to reduce the likelihood that precipitation datasets at a resolution of 1.1 km or coarser might not be able to detect the triggering storms
- Precipitation above the threshold of 1 mm/h was measured by at least one of the four precipitation sources selected for the intercomparison at some point on the day of the landslide.

In total, 177 landslides were selected (see Figure 1). The GLC also contains information about the underlying mass movement mechanism, or landslide type. According to the GLC classification, the mass

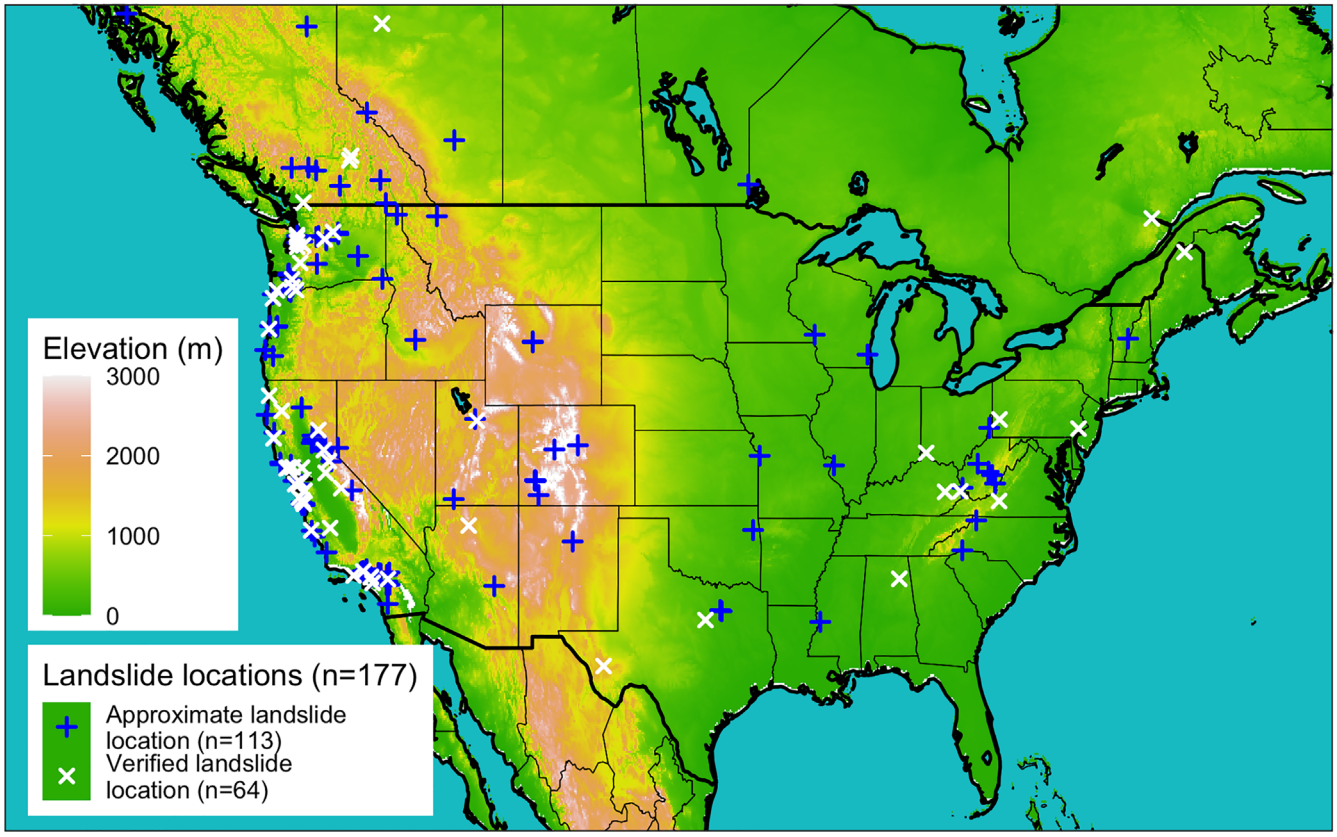

FIGURE 1 Map of all landslide sites considered in this analysis: 177 landslide sites coloured by whether the location was approximate ($n = 113$) or verified using aerial satellite imagery to identify a visible scarp ($n = 64$); source of landslide locations was the GLC (Kirschbaum et al., 2010), source of the DEM data used for the base map (North America Elevation 1-Kilometer Resolution, 2007). Elevations above 3000 m are indicated as the highest value on the colour scale

movements included in this study were dominated by 'landslide' and 'mudslide or debris flow' types (36% and 50%, respectively), with the remaining 14% being 'rock fall' and other types. A prior study (Guzzetti et al., 2008) of intensity–duration thresholds (the primary method of comparison used in this study) concluded that there was no significant difference between thresholds fit to shallow slope failures as compared to debris flows. Since there is no mass movement type corresponding to shallow slope failures in the GLC, a visual comparison of the distributions of intensity and duration of the storms is included in this study in Figure 2. We note that there is little difference in the storm intensity among different types of landslides, but that the mean duration of the storms that trigger mudslides or debris flows seems to be less than other types of mass movements by 6–11 h depending on the precipitation product. As illustrated by this example, the inclusion of multiple mass movement types is a potential source of error in the intensity–duration curve analysis for events that lie close to the threshold. Nonetheless, we conclude that the larger sample size available by using all landslide types was preferred to separating by type for the purposes of this analysis.

Of the included events, the exact locations for 64 sites were verified by a trained technician searching for a landslide scarp in visible satellite images of the terrain near the specified landslide location. The location specified by the GLC was used for the remaining landslides where 25 were marked in the GLC as 'exact' locations, 39 as 1 km, 41 as 5 km, and 8 as 10 km accuracy. Figure 1 shows that many of the sites are located near the Pacific coast, likely due to the presence of complex topography associated with landslides, as well as the population reporting bias of the GLC. The verified landslides are generally distributed evenly relative to the locations of the full selection of landslides.

2.2 | Precipitation data sources

The gridded precipitation datasets in this study were chosen to be reflective of three common measurement methods: gauges, ground-based radar, and satellite. We were interested in products that are freely available, have undergone extensive verification, and extend over at least the CONUS. An important additional criterion was that products be available at an hourly temporal resolution or finer in order to compute the characteristics of individual storm events. We further sought to include products with multiple latencies where available. The above criteria resulted in the four precipitation products described in Table 1 and summarized below. For each product, data were included from 2 June 2015 to 30 June 2018 in order to provide coverage for all the included landslide events, which span from 4 September 2015 to 11 July 2017.

2.2.1 | North American land data assimilation system version 2 meteorological dataset

The North American Land Data Assimilation System version 2 (NLDAS-2) meteorological dataset (Xia et al., 2012) is a combination of daily gauge-based National Center for Environmental Prediction (NCEP) Climate Prediction Center (CPC) precipitation with orographic corrections and hourly NCEP Doppler radar-based precipitation. The gauge-based estimates are disaggregated to hourly data using the radar-based estimates, resulting in a near real-time hourly gridded product at 0.125° (~12 km) resolution across North America going back to 1979 with a latency of approximately 4 days. Though it has coarser horizontal resolution relative to the other precipitation

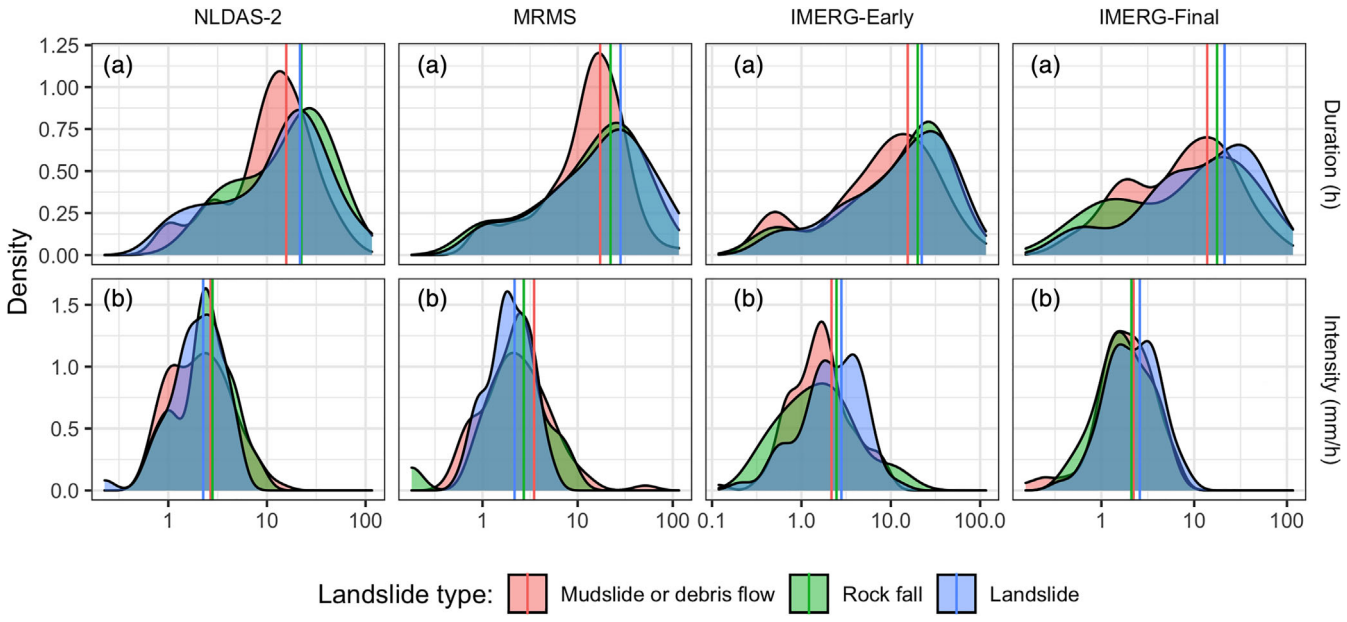

FIGURE 2 Storm characteristics by landslide type: Kernel density estimates for each precipitation product are shown for the duration (h, in panel (a)) and intensity (mm/h, in panel (b)) of landslide-triggering storms. Distributions are separated by the landslide type, where 'landslide' describes unknown types

TABLE 1 The four precipitation products included in the comparison, representing gauge-, radar-, and satellite-based measurements

Precipitation product	Description	Spatial resolution	Temporal resolution	Typical latency
North American Land Data Assimilation System version 2 (NLDAS-2) meteorological dataset (Xia et al., 2012)	Disaggregation of climate prediction centre daily precipitation using bias-corrected radar	0.125° (~12 km)	1 h	4 days
Multi-radar multi-sensor (MRMS; Zhang et al., 2015)	Integrates data from radars, satellites, precipitation gauges, and other sensors to provide near real-time decision support	0.01° (~1.1 km)	1 h	<5 min
Integrated Multi-satellitE Retrievals for Global precipitation measurement early run (IMERG-Early; Hou et al., 2014)	Global network of satellites unified by measurements from a single reference radar/ radiometer satellite	0.1° (~10 km)	30 min	4 h
Integrated Multi-satellitE Retrievals for Global precipitation measurement (IMERG-Final; Hou et al., 2014)	In addition to the satellite data included in IMERG-Early, IMERG-Final includes late-arriving microwave overpasses, monthly gauge-based adjustments, and an algorithm that interpolates forward as well as backward in time	0.1° (~10 km)	30 min	3.5 months

products used here, NLDAS-2 meteorological is a widely used gauge-based product that has been extensively validated over a recent period overlapping with this study (Livneh et al., 2015; Long et al., 2014; Xia et al., 2016).

2.2.2 | Multi-radar multi-sensor quantitative precipitation estimate

Multi-radar multi-sensor (MRMS) precipitation estimates are primarily based on a centralized radar mosaic with 2-min resolution over the United States and Canada. This study uses an hourly version that also integrates data from numerical weather prediction, satellites, gauges, lightning sensors, and precipitation models (Zhang et al., 2015). While both NLDAS-2 and MRMS estimates contain common information from gauges and radar, the NLDAS-2 product is primarily a gauge-based estimate while MRMS focuses on radar inputs. MRMS is the precipitation product with the shortest period of record among the products selected for this study, and so there are relatively few years of data for validation. However, it has by far the highest resolution at 0.01° (~1.1 km) and represents the state of the art in terms of leveraging computing resources to take advantage of a multitude of overlapping radars augmented by other types of sensors.

2.2.3 | Integrated Multi-satellitE Retrievals for GPM

Integrated Multi-satellitE Retrievals for GPM (IMERG) precipitation estimates are a combination of multiple satellite measurements, including the GPM Core Observatory Microwave Imager which is considered the standard for the other included satellites. In addition to active and passive microwave sensors, IMERG estimates include infrared sensors, satellite-based radar, and precipitation gauge adjustments. The gauge data are used for monthly bias correction (Huffman et al., 2020). There are three IMERG products, Early, Late, and Final, of which we use the Early (~4-h latency) and the Final (~3.5-month latency) in this study. The IMERG-Early product is available much more promptly than the IMERG-Final, but as a result some of the satellite retrievals do not arrive soon enough to be incorporated into the product. In addition, IMERG-Early cannot take advantage of some processing steps such as monthly gauge correction because they require data that have yet to be collected at the time of the release (O et al., 2017). IMERG-Final is recommended for research applications as being the most accurate but would not be useful for predicting landslides in a timely fashion (Huffman et al., 2020). Since IMERG products use the GPM active and passive microwave data as a standard with little-to-no information from gauges, they are fundamentally different from many other precipitation products available.

2.3 | Precipitation inter-comparison and computation of storm characteristics

For each of the precipitation products, data were extracted for the precipitation grid enclosing the landslide location for the period between May 2015 (the earliest date MRMS data are available) and May 2020 (the latest release of IMERG-Final data at the time of this analysis). Following Dinku et al. (2008), a minimum threshold of 1 mm/h was applied to the precipitation data to reduce noise. The data were then split into storm events, where a minimum inter-event time (MIT) criterion of 24 h was considered to mark the end of one storm and the beginning of the next as described in Dunkerley (2008). Though Dunkerley (2008) notes that a longer MIT is associated with detecting lower mean event intensity values, we have chosen this length because it is consistent with the temporal accuracy of the vast

majority of the GLC landslides which are reported with a date alone. At a minimum, the 24-h MIT ensures that the landslide-triggering precipitation will be captured.

For each storm, the characteristics of total depth (mm), duration (h), mean intensity over the total event duration (mm/h), peak intensity (mm/h), and return period were computed and compared. The peak intensity for a storm was the intensity of the single maximum precipitation measurement of the storm at the resolution of the product in question. Depth and return period were chosen since they reflect the most common metrics used in extreme hydrologic events (England Jr. et al., 2019). Mean intensity and duration were included because they are parameters commonly used to study rainfall-triggered landslides (Kirschbaum et al., 2012). Previous studies have suggested that in certain cases, high peak intensity can contribute significantly to triggering a landslide independent of the overall storm depth, duration or intensity (Corominas et al., 2002; Yu et al., 2006). This idea is supported for example by observations that landslides are commonly initialized within hours of the peak intensity (Premchitt et al., 1986). For each metric, the range of values among each precipitation product was also computed by subtracting the minimum from the maximum measured values for each event. The average and maximum ranges are reported in order to illustrate the level of variability among products.

To facilitate comparison of storm characteristics within a single over-arching framework, the return period of the landslide-triggering storms was computed using the NOAA precipitation atlas frequency estimations (US Department of Commerce, 2013). The NOAA atlas provides return periods for discrete precipitation durations, namely 1, 2, 3, 6, 12, 24, 48, 72, 96, and 168 h. To define a consistent return period for each storm, we used the maximum precipitation value for each applicable NOAA atlas duration rather than attempting to expand the storm duration to one of the NOAA atlas durations which might have artificially lowered the return periods. For example, for the 3-h duration, cumulative 3-h precipitation totals were calculated for each time step of the storm, and the maximum value chosen. The return period for this maximum value was then retrieved from the NOAA atlas. We then selected the maximum return period from among the 10 possible durations noted above for each landslide. For example, if the maximum 3-h precipitation during the MIT-defined storm event had a 25-year return period while the maximum 48-h precipitation during the storm event only had a 2-year return period, the return period of the 3-h interval would be used in preference over the return period of the 48-h interval or any other duration where the return period was less than 25 years. This procedure ensured that we used the maximum applicable return period available from the NOAA atlas that occurred during each landslide-triggering storm. Values less than a 2-year return period are not included in the NOAA atlas, such that return period values were only assigned for a subset of landslide-triggering storms that exceeded that threshold. Return period data were also unavailable for Canadian sites.

The measurement uncertainty in the total magnitude of precipitation occurring on the day of the landslide as measured by each product was also compared using two metrics: rank and *z*-score. Rank was calculated for each event among the four products, and so it is reported as a number ranging from one to four for each event that shows the tendency of each product to measure higher or lower values than other products. We expect that products with consistently higher ranks will tend to have more false positive landslide predictions, while those with consistently lower ranks may tend to miss true positive events. By using the rank, we quantify uncertainty in the form of biases relative to the other products in the study. The *z*-score conversely was calculated for each event and product among all the other days of precipitation as measured by the same product in the May 2015–May 2020 historical record. The distribution of *z*-scores highlights the variability of each product relative to the others. The differences in *z*-scores also quantify uncertainty in the degree to which landslide-triggering precipitation can be statistically separated from other precipitation for each product.

2.4 | Application of intensity–duration thresholds using different precipitation products

Intensity–duration thresholds are a category of simple models of landslide occurrence whereby a threshold is defined as a power law of the storm duration

$$I = aD^{-b} \tag{1}$$

where I is intensity, D is duration, and a and b are fitted parameters to a particular dataset. Intensities above the threshold are used to predict the occurrence of a landslide (Segoni et al., 2014). A range of thresholds have been calculated under different climates and over multiple scales (Caine, 1980; Kirschbaum et al., 2012; Scheevel et al., 2017). Four thresholds for this study (see Table 2) were obtained from a review by Guzzetti et al. (2008) as a way to test the sensitivity of our results to the choice threshold. From the myriad thresholds presented in the Guzzetti et al. (2008) review, we chose those that met the following requirements:

- The threshold type was an intensity–duration threshold and not, for example, normalized by mean annual precipitation, since the relatively short precipitation record of some of the data sources in this study (<5 years) was deemed too short to compute a reliable climatology.
- A spatial domain covering a large portion of the study area.
- A sample size of $n > 25$ events once climatic, mass movement type, and duration restrictions were applied to the data.
- Only thresholds based directly on data, as opposed to those based on probability estimates of precipitation, were considered.
- One threshold (Caine, 1980) was eliminated because it was based on data from sites that had been verified to be undisturbed prior to the landslide, which resulted in a much higher threshold than those based on inventories that, like the GLC, did not screen for disturbance.

TABLE 2 Four intensity–duration (ID) thresholds used in the analysis

Threshold source	Equation	Restrictions	GLC landslide types
Clarizia et al. (1996)	$I = 10 \times D^{-0.77}$	Soil slips	Landslide
Crosta and Frattini (2001)	$I = 0.48 + 7.2 \times D^{-1.00}$	Soil slips and debris flows	Landslide, mudslide, debris flow
Guzzetti et al. (2008)	$I = 4.81 \times D^{-0.49}$	Csa climate zone	All
Guzzetti et al. (2008)	$I = 3.57 \times D^{-0.41}$	Csb climate zone	All

Note: The two Guzzetti et al. thresholds were derived by the same method for data from different climates. Landslide types are translated from the threshold source to GLC categories. All thresholds are valid in the duration range of 0.1 h < D < 1000 h.

Further, thresholds were only applied to applicable subsets of the data based on the restrictions of each threshold, which included climate zone, mass movement type, and triggering storm duration. Specifically, since the Guzzetti et al. (2008) thresholds were defined for Koppen climate zones Csa and Csb (temperate, dry summer, hot and warm summer respectively) only data falling in the corresponding climate zones (Beck et al., 2018) were considered for those thresholds. The other two thresholds were limited on the basis of landslide types. Table 2 includes the GLC landslide types that correspond to the threshold definition in addition to applied restrictions.

For each threshold-product combination, we computed a hit ratio (correctly predicted landslides over the total number of landslides), a false alarm ratio (incorrectly predicted landslides over the total number of non-landslides), and a frequency bias (predicted landslides over observed landslides). The hit ratio and false alarm ratio tend to be inversely linked measures of whether the model tends to over- or under-predict landslides. Frequency bias adds an additional measure of whether or not an ID threshold is identifying landslide events at the same overall rate that they occur. In computing these metrics for multiple ID thresholds, we examine the degree to which landslide hazard assessments are sensitive to both uncertainty in precipitation measurements and uncertainty in the threshold itself.

3 | RESULTS

The four precipitation products examined in this study exhibit a great deal of variability in the time period leading up to a landslide event. As an example of the magnitudes and qualitative characteristics of that variability, Figure 3 shows the cumulative precipitation in the 30 days before a landslide at five sites. The selected sites showcase multiple ways in which precipitation can differ among the products. For example, while the precipitation in panel (a) matches fairly closely for all products, in panel (b) precipitation still appears to be correlated but also demonstrates a factor of two spread of precipitation values. In panel (c) the IMERG products diverge substantially from the ground-based products early on in cumulative volume, but the landslide-triggering storm is recorded as nearly twice as large by the satellite-based products, demonstrating that the differences in precipitation measurement can partially cancel out in the right situation. In panel (d) the IMERG-Early product reports nearly doubled precipitation values throughout while all three of the remaining products are very similar. Among the events where IMERG-Final recorded a much lower than average value, it was common for the high average to be driven by the IMERG-Early measurements almost exclusively, as shown here in panel (d). Panel (e) shows a likely landslide location error since none of the products register any precipitation close to the time of the event. Such events were not included in this analysis of landslide-triggering storms because no such storm could be identified.

The variability among products is also evident in the distribution of daily precipitation rank among products and *z*-score within products. The relative magnitude of the different precipitation products on the day of the landslide is shown in Figure 4 in terms of the rank among the four products for each day, and *z*-score among all non-zero data for a particular product. Both day-of-landslide precipitation and all other non-zero days in the study period are shown for comparison. The ranks of each product do not reveal substantial biases across the entire precipitation record, with the exception of MRMS which has a larger proportion of above-median ranks than the other products (see Figure 4(a)). On landslide-days only, the IMERG products have lower ranks overall, revealing greater uncertainty in measurements of these extreme events as well as a systematic bias in the satellite products relative to ground-based products. This idea is reinforced by the *z*-scores of the precipitation among measurements from the product and landslide site. Though the median *z*-score is similar for all products across the entire record (see Figure 4(c)), it is lower for both IMERG products on the day of the landslide. Conversely, some outliers in the IMERG-Early have the highest *z*-scores among day-of-landslide precipitation (see Figure 4(d)), even though the median and third quartile values are higher for IMERG-Final. This suggests that the further processing of the IMERG-Final product reduces unusually high precipitation measurements while also increasing low values. For all products, each quartile of the day-of-landslide precipitation is larger than that of the non-landslide-triggering precipitation, though none of the maximum precipitation *z*-scores appears to have occurred on the day of the landslide.

Variability among the precipitation products is also revealed in a comparison of day-of-landslide precipitation with mean values among each of the products. Figure 5 shows the characteristics of the landslide-triggering storms plotted against the ensemble mean of all the products for all the landslide sites and separately for the verified locations. It also includes depth, duration, mean intensity, and peak intensity plotted against the ensemble mean for all storms in the study

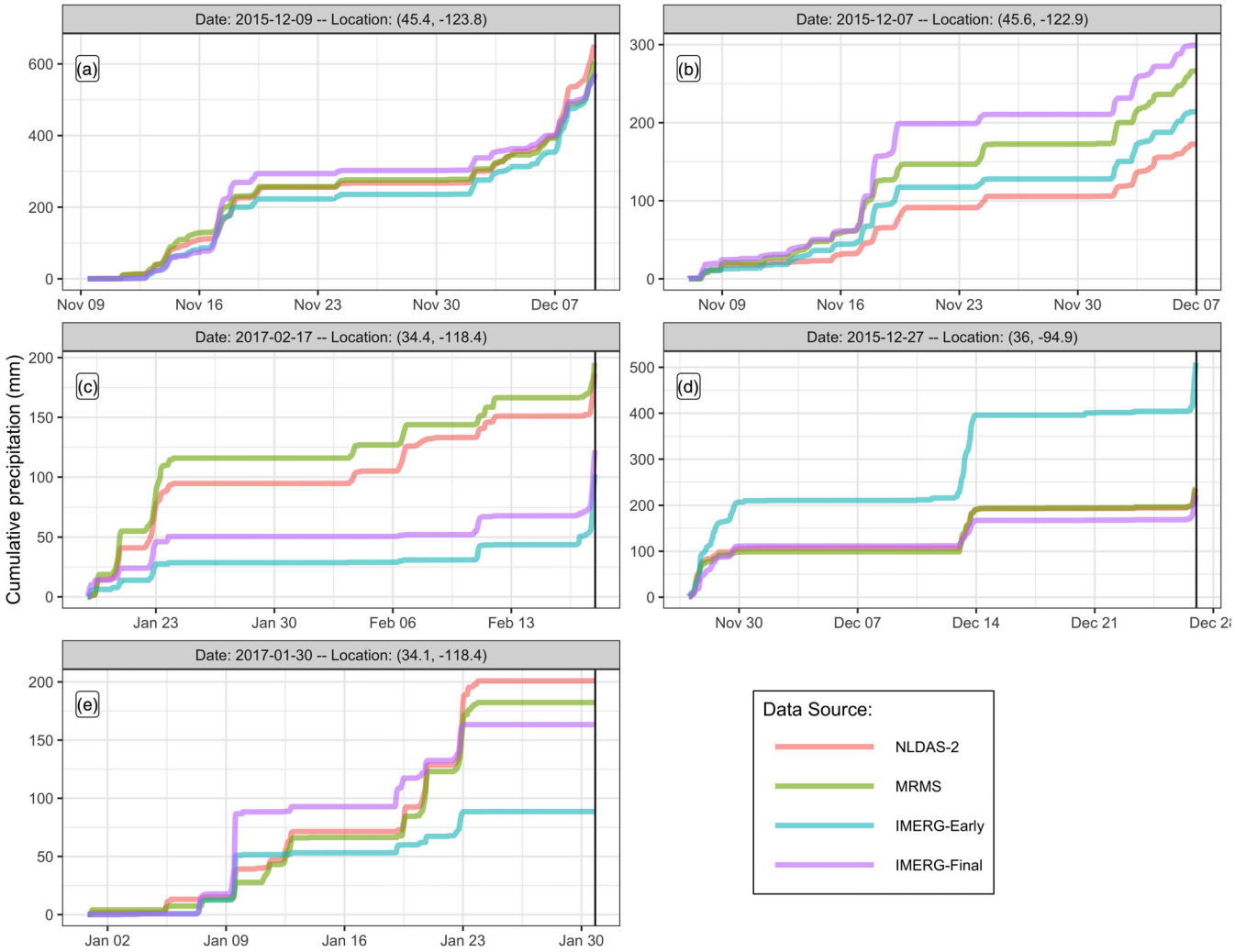

FIGURE 3 Exposition into the types of precipitation differences leading up to landslide events: Cumulative precipitation measurements at select landslide sites for the 30 days before the event. The date of the landslide event is indicated by a vertical black line in each panel. Precipitation is variable across the different products, and the selected sites each demonstrate diverse types of variability. Panel (a) shows similar measurements among all products throughout the 30 days. In panel (b), all products are well correlated, but the accumulated depths greatly differ. In panel (c) both IMERG products report less precipitation until the landslide-triggering storm when they reverse and report more precipitation than MRMS and NLDAS-2. In panel (d) IMERG-Early reports much more precipitation than the other products. Finally, in panel (e) no landslide-triggering precipitation was detected by any product, suggesting a location error in the landslide record

period record. Landslide-triggering precipitation varies widely across precipitation products with the depth of individual storm events diverging by as much as 296 mm with an average range of 51 mm. Peak intensity measurements which are typically influential in triggering landslides, were also highly variable with an average range of 7.8 mm/h and as much as 57 mm/h. Variation in storm depth is particularly visible among a set of outliers below a value of 10 mm —corresponding with a fairly modest storm depth.

Among the verified locations, there is less variability than among approximate locations. For depth, 53% of the approximate locations have a greater percent difference than the median, while only 46% are greater than the median for approximate locations. Likewise with duration, 54% of the approximate locations have a percent difference greater than the median while only 44% of the exact locations do. This suggests that some of the uncertainty in precipitation for these events is related to uncertainty in the landslide locations. Interestingly, this result also implies that an accurate landslide location can decrease precipitation uncertainty, and further suggests that there is greater agreement among the precipitation products used in this study about extreme events than about nearby, presumably less extreme precipitation. Notably, many fewer points appear on the return period plot than the other plots. This is partially due to NOAA Atlas return periods being unavailable for 69 of the 177 events, for example because they took place outside the CONUS. Another factor is the categorical nature of the NOAA Atlas return period, which results in overplotted data points when return periods are identical across all precipitation data sources. In the cases of an additional 90 of the 108 records, the return period was calculated to be the

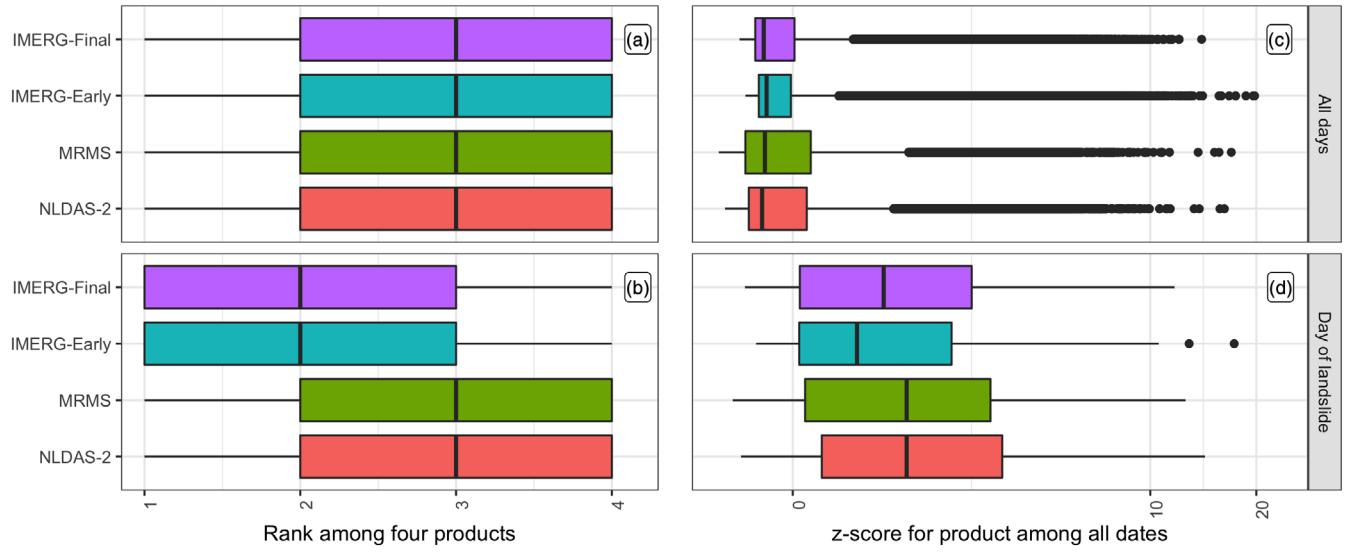

FIGURE 4 Relative magnitude of precipitation products on the day of the landslide: Rank among all products for each day, and *z*-score of daily precipitation as measured by each product for each of 177 events. Panels (a) and (c) show the entire precipitation record while panels (b) and (d) show only the day-of-landslide precipitation for comparison. *Z*-scores are plotted on a pseudo-log scale, a combination of a linear scale near zero and a log scale for higher values

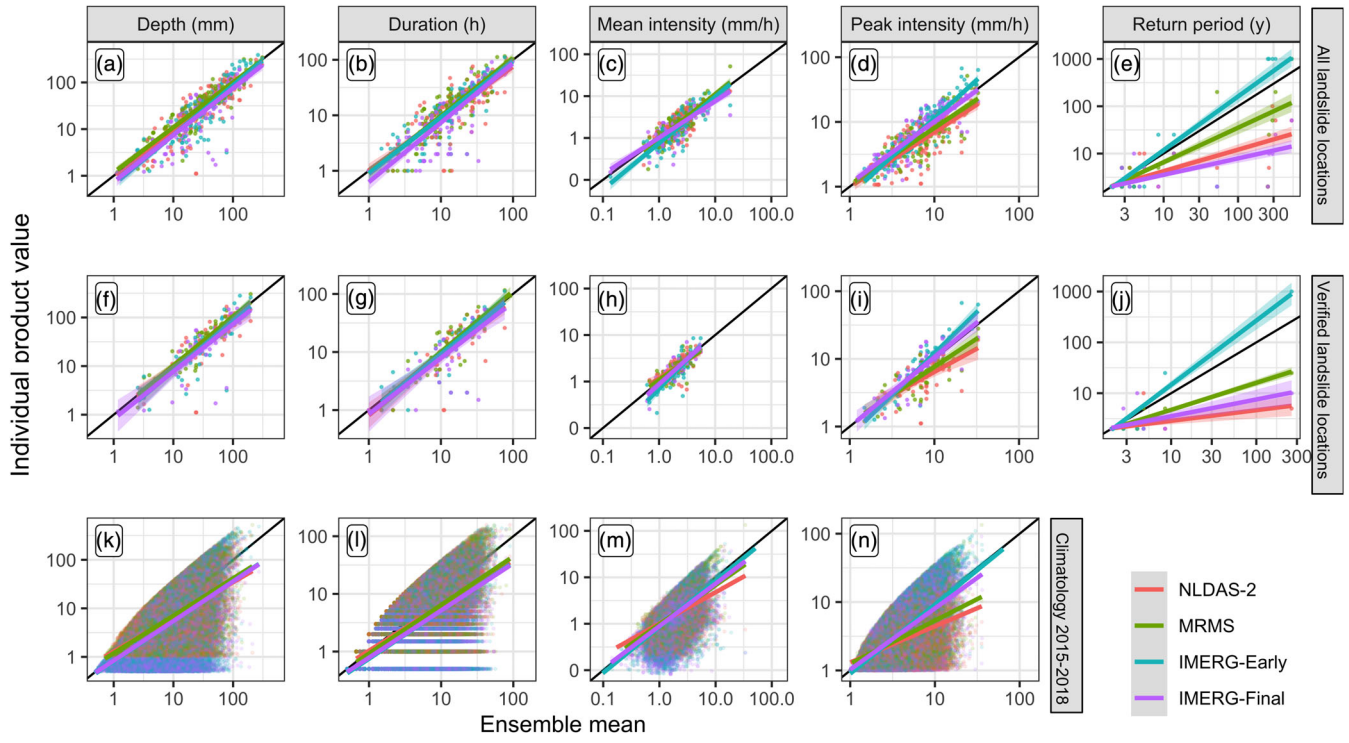

FIGURE 5 Storm characteristics versus the ensemble mean: Depth (mm), duration (h), mean intensity (mm/h), peak intensity (mm/h), and return period (year) for each of the landslide-triggering storms as measured by four precipitation products. Least-squares regression lines with 95% confidence intervals are also shown. Panels (a)–(e) show all 177 sites while panels (f)–(j) show the 64 verified locations. Panels (k)–(n) show the depth (mm), duration (h), mean intensity (mm/h) and peak intensity (mm/h) for all storms in the climatology at all locations as a comparison

minimum value of 2 years for every precipitation data source. As a result, trends in return periods are highly sensitive to the handful of events with meaningful return period information.

The two satellite products have a pattern of distinctive readings relative to ground-based products in that they contain both the highest and lowest values of several storm characteristics. The IMERG

products generally report higher peak hourly intensities for the storms (see Figure 5(d),(i)) with the highest mean peak intensity, which is likely at least partially due to the shorter 30-min time step. This phenomenon is visible across all the storms in the study period, not just those that triggered landslides. However, the trendline for the peak intensity for both IMERG products crosses over to above average for the landslide-triggering storms, while it does not for all storms, suggesting that a larger proportion of landslide-triggering storms have a larger-than-ensemble-average peak intensity than is generally the case. The higher peak intensities are also reflected to some degree in longer return periods for the IMERG-Early data, but not for the IMERG-Final data. The return periods are based on hourly durations or longer for comparison with the NOAA Atlas and therefore may be less sensitive to half-hourly peak intensities. An examination of the relationship between return period and peak intensity in Figure 6 shows a clear correlation between the $\log_{10}$ of the return period and the peak intensity. The relationship between peak intensity and return period is not surprising given that the return period values were calculated by searching for the most intense period of each NOAA atlas duration. However, IMERG-Final has lower return periods overall when all landslide locations are included despite reporting high peak intensities. This anomaly disappears in an examination of verified locations alone. An examination of the 30-day precipitation record prior to the landslide for sites where the IMERG-Final return period was much lower than the average revealed that in most of those cases the higher mean was driven primarily by anomalously high IMERG-Early values not reflected in any of the other datasets. The IMERG-Early product has higher return periods when considering the verified locations only, suggesting that the other products may not detect the highest return period precipitation events as consistently.

The precipitation products are examined in the context of landslide triggering thresholds in Figure 7, with the performance summarized in Table 3. Interestingly, the choice of intensity–duration threshold does not appear to make a large difference in performance because the threshold curves are more similar than the variation in the precipitation data across sites and among products. The MRMS or NLDAS-2 products tend to perform better than either IMERG product, with hit ratios between 0.77–0.81 and 0.73–0.81 rather than 0.61–0.65 and 0.64–0.67 among the verified landslide locations, respectively. Though NLDAS-2 performs similarly to MRMS on the hit rate, its frequency bias is higher (62.2–85.5 among verified locations as opposed to 49.1–66.5 for MRMS), indicating that it over-predicts to a larger degree. All products perform comparably or better when using only the verified landslide locations than they do when including the approximate locations as well. All products also have very large frequency biases, indicating excessive over-prediction.

Figure 7 shows a concentration of long-duration, low-intensity storms that are in the vicinity of a 12-h duration for all products. This pattern is illustrated in Figure 8, where it can be clearly seen that distributions of mean intensity values are comparable across precipitation data source for low and high duration values, but that between 6 and 18 h duration the values for the IMERG products are somewhat lower. Many of the storms that did trigger landslides but were not correctly identified by the intensity–duration threshold fall into this group of approximately 12-h low-intensity storms. We further note that there are few landslide-triggering storms at the short-duration, high-intensity end of the ID thresholds.

4 | DISCUSSION

The results of this analysis revealed substantial variation in the measurements of the four precipitation products, perhaps to an even greater extent than the already notable uncertainty in precipitation measurements reported by other inter-comparisons (Adler et al., 2003; AghaKouchak et al., 2011; Rossi et al., 2017; Sun et al., 2018). We note that the uncertainty due to measurement source in precipitation depths accumulated over these 30-day periods are of the same order of magnitude as the *annual* error in depth reported for products of the same category by Sun et al. (2018). This discrepancy may be due to increased variability across products of different categories, for example, satellite versus radar, in contrast with the figure from Sun et al. (2018) which includes only satellite products.

Among the precipitation products chosen for this study, the two IMERG products identify higher peak intensities relative to the other products. These peak intensities are lower than the highest landside-triggering intensities for data in the Guzzetti et al., 2007 review of ID thresholds, which lie above 100 mm/h and typically have hourly or half-hourly durations. Nonetheless they are within typical values for landslide-triggering storms from a variety of areas compared in Guzzetti et al., 2007. This discrepancy could be due to the 24-h MIT chosen for this study, which is on the longer side, chosen to correspond with the temporal accuracy of the landslide data. These storms

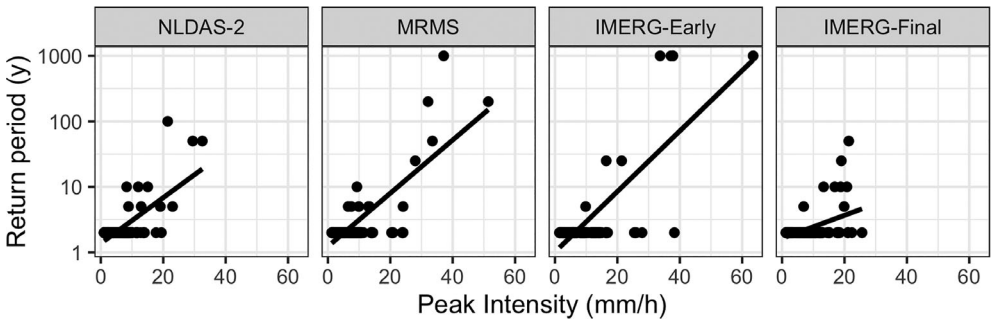

FIGURE 6 Relationship between peak intensity and return period: Scatter plots of the peak intensity (mm/h) and return period (year) for each of four precipitation products. A least-squares regression line is also shown

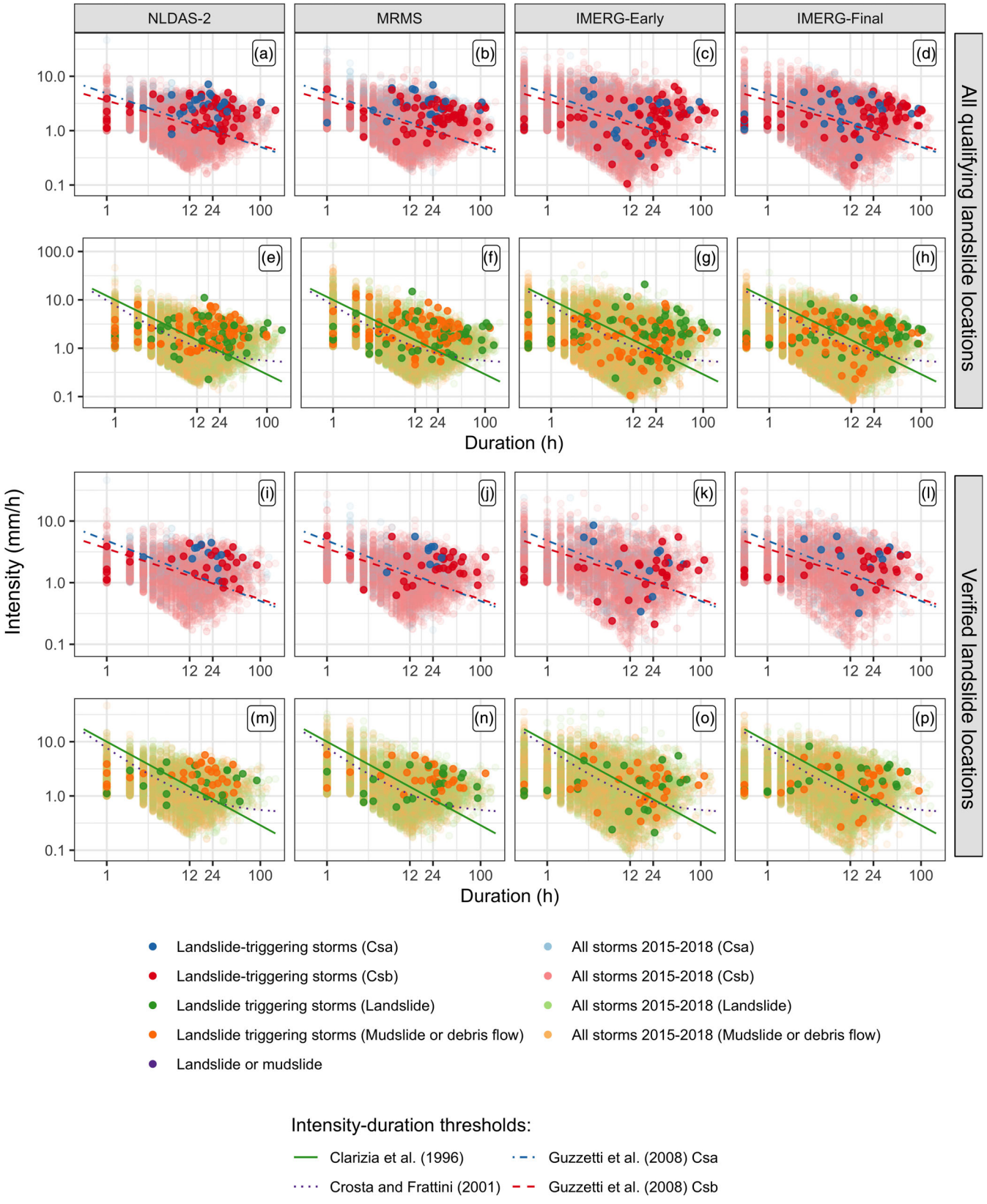

FIGURE 7 Comparison of landslide-triggering precipitation relative to intensity–duration thresholds: Each storm in the precipitation record and established global or climactic intensity–duration thresholds. Landslides are coloured according to their climate or landslide-type category and ID threshold curves are coloured by their restrictions. Landslide-triggering storms are shaded in darker colours. The panels (a)–(h) contains precipitation data for all sites while in panels (i)–(p) only verified sites are included. Points above each threshold are predicted by the threshold to be landslides, and so a larger proportion of landslides plotting above the threshold indicates better performance

TABLE 3 Hit ratio and false alarm ratio for each product and the Clarizia et al. (1996) [1], Crosta and Frattini (2001) [2], and Guzzetti et al. (2008) Csa [3] and Csb [4] intensity–duration thresholds

ID threshold:		Hit ratio				False alarm ratio				Frequency bias			
		1	2	3	4	1	2	3	4	1	2	3	4
Product	**Include sites**												
NLDAS-2	All ($n = 131$)	0.67	0.72	0.66	0.71	0.22	0.31	0.26	0.32	57.7	80.9	67.6	83.0
	Verified ($n = 55$)	0.75	0.78	0.73	0.81	0.25	0.34	0.28	0.34	62.2	85.0	70.9	85.0
MRMS	All ($n = 134$)	0.71	0.73	0.71	0.74	0.24	0.34	0.29	0.35	40.8	56.3	47.9	58.1
	Verified ($n = 56$)	0.77	0.77	0.78	0.81	0.27	0.37	0.31	0.37	49.1	66.2	55.3	66.5
IMERG-Early	All ($n = 146$)	0.57	0.63	0.61	0.63	0.2	0.27	0.25	0.29	65.1	86.1	78.1	93.4
	Verified ($n = 59$)	0.63	0.65	0.61	0.61	0.23	0.3	0.27	0.32	68.4	89.0	80.3	95.0
IMERG-Final	All ($n = 145$)	0.6	0.64	0.6	0.62	0.22	0.3	0.27	0.32	69.9	95.0	84.3	100.9
	Verified ($n = 56$)	0.65	0.67	0.64	0.64	0.25	0.34	0.3	0.36	75.9	100.6	90.4	106.1

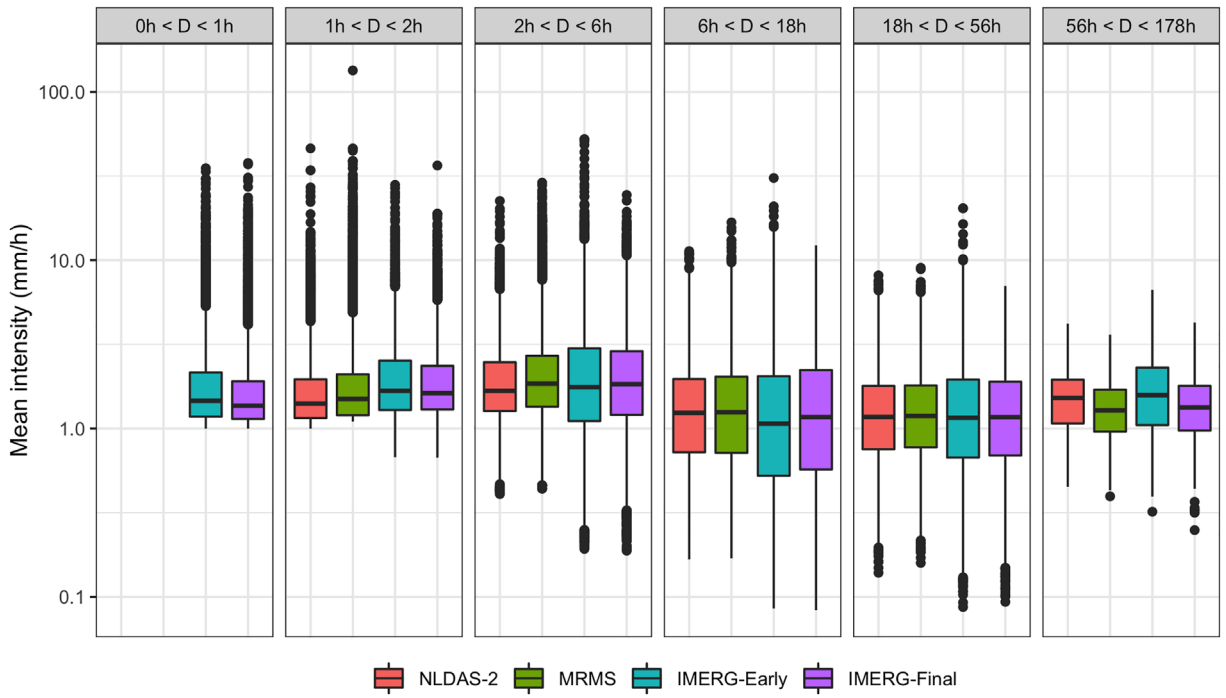

FIGURE 8 Comparison of mean intensity values for each precipitation source: The distribution of mean intensity values for each precipitation product shown as boxplots. The data are split into seven duration bands of equal width on a logarithmic scale

may be an artefact of the MIT storm identification algorithm. Since the landslides did not have times specified, the entire day of the landslide was always included unless there was no rain until the end of the day, and this may have extended some storms past when the landslide occurred. This would have the effect of computing lower total intensity values for storms that lasted only through the time of the landslide but persisted at a much lower intensity thereafter. In addition, Dunkerley (2008) notes that MIT and the mean event precipitation intensity are inversely correlated. An alternative landslide database with more accurate temporal information could have helped to address this issue, although at the time of publishing we are not aware of any landslide inventories in the study area that provide larger numbers of landslide sites with sub-daily temporal resolution. Another possibility is that the precipitation data used in Guzzetti et al., 2007 were available in sub-half-hourly increments, and so were able to detect higher intensity events than even the IMERG products used here. Finally, a substantial portion of the events occurred in regions that are arid and prone to wildfire. It is possible that some of the landslides took place in areas affected by some kind of disturbance that increased landslide hazard, or simply places where 10–40 mm/h of precipitation was sufficient to trigger a landslide.

Interestingly, the IMERG products also detect more anomalously low precipitation values, leading to greater uncertainty in precipitation detection and artificially extended storm durations. Low-intensity

precipitation in all products was associated with long duration storm events (see Figure 7), which may occur because of low-intensity precipitation slightly above the 1 mm/h threshold that extended the MIT-computed duration of the storm and thereby reduced its overall intensity. The IMERG products were particularly vulnerable to the identification of long-duration low-intensity storms, resulting in a high miss rate that is unsuitable for landslide prediction or hazard detection. However, it is possible that the low-intensity storms are an artefact of the 24-h MIT method used in this study to separate storms. This hypothesis is supported by the concentration of low-intensity 24-h events identified by the MIT algorithm. Those long-duration low-intensity storms had the effect of producing misses (false negatives), thereby lowering the hit ratio to such an extent as to practically disqualify the use of the IMERG products for prediction purposes. Because the IMERG products were able to identify higher intensity precipitation than the other products, it is possible that they would in fact perform better for identifying landslides if the low-intensity storms were eliminated through use of a different method of precipitation event separation. Further investigation is needed to determine a more suitable method, such as the method used in Rossi et al. (2017) in which events were separated not only if there was no rainfall for a specified period of time but also if precipitation fell below a minimum value.

A caveat to this study is that many of the events used in the analysis were labelled 'landslide' in the GLC, meaning that the specific mechanism was unknown. Though there is some evidence in a prior study of ID thresholds (Guzzetti et al., 2008) that the same ID threshold can be used for shallow slope failures and debris flows despite differing mass movement mechanisms, it is possible that some of the events labelled as landslides were in fact neither of these two but instead a rock fall or deep-seated landslide. We observed a small difference between the precipitation distributions for debris flows/mudslides and other types (see Figure 2), with debris flows/mudslides having a lower duration than landslides or rock falls. However, despite this difference the different types of mass movements had relatively similar distributions of both intensity and duration. This result suggests that even if some deep-seated landslides were unclassified and subsequently included in the study, they are not unduly impacting the results.

All precipitation products performed reasonably well at identifying landslides using the published intensity–duration thresholds particularly considering that these thresholds were developed on training data from different datasets spanning large regions. However, they did not perform as well at excluding false alarms, most likely because of factors beyond intensity and duration that can influence landslide occurrence such as topography, soil type, recent wildfire or disturbance or land development. The prevalence of false alarms is an inherent weakness of ID threshold models, which do not include many of the variables which affect landslide occurrence. In addition, as noted above, ID thresholds are sensitive to uncertainty in precipitation measurements. As a result, such threshold models cannot be used to predict specific landslide events and are better suited for representing regional changes in landslide hazards as a result of climate change or other disturbances.

An additional weakness of ID threshold models limiting their usefulness for specific predictions is the false alarm rate. This indicator of landslide hazards has the potential to help quantify regional changes in landslide hazards as a result of projected changes in precipitation patterns. Furthermore, as the boundaries of climate regions migrate over time, applying alternative intensity-duration thresholds developed in locations that are more representative of the future climate in a particular region may be more informative than relying exclusively on historical intensity-duration thresholds developed for the area.

Additionally, while the issue of different mass movement mechanisms does not appear to affect ID threshold performance on a large scale with current levels of model performance (Guzzetti et al., 2007), it may explain the incidence of, for example, high-intensity storms that do not trigger landslides because they are located in an area that is more prone to a landslide type commonly triggered by soil saturation. For all of these reasons, some of the high-intensity precipitation that did not trigger any recorded landslides could be more reflective of adjacent areas that are not as susceptible to landslides. Conversely a landslide at a highly susceptible location, such as an area with high slopes that had recently been burned by wildfire, could be triggered by less intense rain, potentially resulting as a miss on an intensity–duration curve. Even the 1.1 km resolution of the MRMS data could contain substantial variation in landslide susceptibility within an individual grid cell, leading to location-based uncertainty in the precipitation values, especially when combined with uncertain landslide locations. The poorest performing products were the IMERG products because despite their detecting more high-intensity precipitation events they also detected many low-intensity long-duration events that nonetheless triggered landslides.

Both Rossi et al. (2017) and Brunetti et al. (2018) also found satellite products did not perform as well as gauge data relative to intensity–duration thresholds, as a result of underestimating precipitation. Therefore, it was not surprising that the intensity–duration thresholds performed more poorly at identifying landslides when applied to satellite data in this study as well. However, Rossi et al. (2017) found that adjusting the threshold accounted for precipitation bias, suggesting that the pattern of lower intensities was more or less uniform across different durations. Our study supports this finding. The lower false alarm ratios and comparable frequency bias suggest that the satellite products might perform better with a bias-adjusted threshold. By contrast, as illustrated in Figure 8 we found that the low-intensity values were more often clustered around a relatively narrow duration band, which would be more challenging to bias-correct. Though the intensity–duration thresholds still show promise for diagnosing landslides using satellite-based data, the adjustments to improve performance may prove more complex for the IMERG products across the broader spatial domain of the continental United States and Canada.

MRMS and NLDAS-2 are relatively low latency products. In the case of IMERG-Early the short latency seemed to come at a cost of an exaggeration of the weaknesses and strengths of IMERG in identifying landslides. In particular, IMERG-Early had the greatest prevalence of low storm intensities, and so it ultimately performed the worst at

landslide identification (see Table 3). Without changes to the precipitation processing, the low latency does indeed appear to be a liability in this case.

Precipitation measurements at verified landslide sites tended to be of higher magnitude than those at sites with approximate locations for all products. The intensity–duration thresholds subsequently performed better at verified locations across all precipitation products. This finding corresponds with the conclusion of Leonarduzzi et al. (2017), who found that a high-quality landslide database improved ID threshold performance far more than other interventions such as regionalization. Though this difference remains unexplained, one possibility is that some of the approximate landslide locations were too far away from the true landslide location for the precipitation measurements to be representative. The issue of poor location accuracy is especially likely to cause difficulties for evaluating landslide hazards, since landslides typically occur in areas with steeper slopes, and the precipitation in these mountainous environments has a high degree of spatial variability (Buytaert et al., 2006; Diodato, 2005). Alternatively, there may have been other factors such as sparse vegetation cover that made it more difficult to locate landslides on satellite imagery and also lowered the precipitation threshold that would trigger a landslide. Since work on this study began, a compilation of U.S. landslides has been released by the USGS (Mirus et al., 2020) which would also be a suitable source of landslide locations with perhaps greater location precision that could help resolve this question in future work along the same lines.

5 | CONCLUSION

The precipitation products chosen for this study represent diverse measurement techniques that often recorded large differences in precipitation leading up to the landslide events evaluated here. As a result of this uncertainty, each precipitation product differed in overall performance in predicting landslides using ID thresholds. Overall, the choice of ID threshold curves was not as consequential as the choice of precipitation product in identifying landslides. Products that rely on ground-based sensors showed a more consistent landslide signal despite generally recording lower peak intensities.

Though it was hypothesized that half-hour or 1-h peak intensity would be an important factor in identifying landslides, the results suggest instead that removal of noise on the low end may be more important. A particular challenge was the presence of low-intensity, long-duration storms preceding landslide events, most prevalent in the IMERG products. A more expansive evaluation of processing techniques for separating storms may potentially mitigate these issue, although each technique will produce artefacts in the comparisons. Another potential avenue for addressing this problem is to combine multiple datasets, since the low-intensity long-duration storms did not appear in all datasets to the same degree.

Another limitation to the study of landslide-triggering storms is the general lack of both exact landslide locations and specific time of day of the landslide events. The location limitation was reflected in better performance for verified landslide locations as compared to approximate locations, which implies that some of the approximate locations were incorrect to such an extent that the precipitation measurements were misaligned. This problem could be addressed by more extensive manual searches such as the one used in this study that identified the 64 verified landslide locations, or perhaps in the future by machine learning methods. An additional option for potentially more accurate landslide location and timing data is the recently released USGS database mentioned above.

Using the methods tested in this study, those practitioners attempting to use intensity–duration thresholds as operational landslide models would do well to select a product like MRMS that has extremely low latency and performs well at identifying landslides. The other ground-based product with substantial gauge- and radar-based inputs, NLDAS-2, did not perform as well as MRMS, suggesting that the high spatial resolution of MRMS (~1.1 km) is a key feature when it comes to landslide identification. None of the products were particularly good at filtering out false alarms of landslides. Therefore, practitioners are advised to use intensity–duration thresholds as a hazard and risk assessment tool over larger regions rather than as direct predictions for specific locations. An additional recommendation would be for practitioners to consider more than one precipitation product, that is, multiple precipitation estimates simultaneously, as a way to confirm stronger precipitation signals and to minimize the influence of noise.

ACKNOWLEDGEMENTS

This research was supported by funding from the National Aeronautics and Space Administration's Interdisciplinary Research in Earth Science (IDS) program grant 16-IDS16-0075, The Interaction of Mass Movements with Natural Hazards Under Changing Hydrologic Conditions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

These data were derived from the following resources available in the public domain: The NASA Global Landslide Catalog was downloaded from NASA's Open Data Portal (https://data.nasa.gov/Earth-Science/Global-Landslide-Catalog/h9d8-neg4); Both IMERG products were retrieved from the Global Precipitation Measurement data portal (https://gpm.nasa.gov/data/directory); NLDAS-2 data is available on NASA's EarthData site (https://disc.gsfc.nasa.gov/datasets/NLDAS_FORA0125_H_002/summary?keywords=NLDAS/); and MRMS data was retrieved from a public archive at Iowa State University (https://mtarchive.geol.iastate.edu/). The verified locations of GLC landslides used in this study are available from the corresponding author upon reasonable request.

ORCID

Elsa S. Culler https://orcid.org/0000-0003-1151-3111
Andrew M. Badger https://orcid.org/0000-0003-4537-9993
Justin Toby Minear https://orcid.org/0000-0001-9496-2056
Kristy F. Tiampo https://orcid.org/0000-0002-5500-7600
Spencer D. Zeigler https://orcid.org/0000-0001-9338-010X
Ben Livneh https://orcid.org/0000-0001-5445-2473

REFERENCES

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., & Nelkin, E. (2003). The Version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979Present). *Journal of Hydrometeorology*, *4*(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2

Adler, R. F., Kidd, C., Petty, G., Morissey, M., & Goodman, H. M. (2001). Intercomparison of global precipitation products: The third precipitation intercomparison project (PIP-3). *Bulletin of the American Meteorological Society*, *82*(7), 1377–1396.

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the Central United States. *Journal of Geophysical Research: Atmospheres*, *116*(D02115), 1–11. https://doi.org/10.1029/2010JD014741

Ahmadalipour, A., & Moradkhani, H. (2017). Analyzing the uncertainty of ensemble-based gridded observations in land surface simulations and drought assessment. *Journal of Hydrology*, *555*, 557–568. https://doi.org/10.1016/j.jhydrol.2017.10.059

Amitai, E., Petersen, W., Llort, X., & Vasiloff, S. (2012). Multiplatform comparisons of rain intensity for extreme precipitation events. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(3), 675–686. https://doi.org/10.1109/TGRS.2011.2162737

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R., & Prat, O. P. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, *96*(1), 69–83. https://doi.org/10.1175/BAMS-D-13-00068.1

Bao, J., Sherwood, S. C., Alexander, L. V., & Evans, J. P. (2017). Future increases in extreme precipitation exceed observed scaling rates. *Nature Climate Change*, *7*(2), 128–132. https://doi.org/10.1038/nclimate3201

Beck, H. E., van Dijk, A. I., Levizzani, V., Schellekens, J., Miralles, D., Martens, B., & de Roo, A. (2017). MSWEP: 3-Hourly 0.25 global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, *21*(1), 589–615. https://doi.org/10.5194/hess-21-589-2017

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Bousquet, O., & Smull, B. F. (2003). Observations and impacts of upstream blocking during a widespread orographic precipitation event. *Quarterly Journal of the Royal Meteorological Society*, *129*(588), 391–409. https://doi.org/10.1256/qj.02.49

Brunetti, M. T., Melillo, M., Peruccacci, S., Ciabatta, L., & Brocca, L. (2018). How far are we from the use of satellite rainfall products in landslide forecasting? *Remote Sensing of Environment*, *210*, 65–75. https://doi.org/10.1016/j.rse.2018.03.016

Buytaert, W., Celleri, R., Willems, P., Bièvre, B. D., & Wyseure, G. (2006). Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes. *Journal of Hydrology*, *329*(3), 413–421. https://doi.org/10.1016/j.jhydrol.2006.02.031

Caine, N. (1980). The rainfall intensity—Duration control of shallow landslides and debris flows. *Geografiska Annaler: Series A, Physical Geography*, *62*(1–2), 23–27. https://doi.org/10.1080/04353676.1980.11879996

Cannon, S. H., & Gartner, J. E. (2005). Wildfire-related debris flow from a hazards perspective. In *Debris-flow hazards and related phenomena* (p. 363–385). Chichester, UK: Praxis Publishing Ltd.

Chandrasekar, V., Hou, A., Smith, E., Bringi, V. N., Rutledge, S. A., Gorgucci, E., Petersen, W. A., & Jackson, G. S. (2008). Potential role of dual-polarization radar in the validation of satellite precipitation measurements: Rationale and opportunities. *Bulletin of the American Meteorological Society*, *89*(8), 1127–1146. https://doi.org/10.1175/2008BAMS2177.1

Chikalamo, E. E., Mavrouli, O. C., Ettema, J., van Westen, C. J., Muntohar, A. S., & Mustofa, A. (2020). Satellite-derived rainfall thresholds for landslide early warning in Bogowonto Catchment, Central Java, Indonesia. *International Journal of Applied Earth Observation and Geoinformation*, *89*, 102093. https://doi.org/10.1016/j.jag.2020.102093

Chowdhury, R., & Flentje, P. (2002). Uncertainties in rainfall-induced landslide hazard. *Quarterly Journal of Engineering Geology and Hydrogeology*, *35*(1), 61–69. https://doi.org/10.1144/qjegh.35.1.61

Clarizia, M., Gullà, G., & Sorbino, G. (1996). Sui meccanismi di innesco dei soil slip. *International Conference Prevention of Hydrogeological Hazards: The Role of Scientific Research*, *1*, 585–597.

Corominas, J., Moya, J., & Hürlimann, M. (2002). Landslide rainfall triggers in the Spanish eastern Pyrenees. *Mediterranean Storms, Proceedings of the 4th Plinius Conference*.

Crosta, G. B., & Frattini, P. (2001). Rainfall threshold for triggering soil slips and debris flow. Proc. of the 2nd EGS Plinius Conference on Mediterranean Storms: Publication CNR GNDCI, 3547, 24.

Dinku, T., Chidzambwa, S., Ceccato, P., Connor, S. J., & Ropelewski, C. F. (2008). Validation of high-resolution satellite rainfall products over complex terrain. *International Journal of Remote Sensing*, *29*(14), 4097–4110. https://doi.org/10.1080/01431160701772526

Diodato, N. (2005). The influence of topographic co-variables on the spatial variability of precipitation over small regions of complex terrain. *International Journal of Climatology*, *25*(3), 351–363. https://doi.org/10.1002/joc.1131

Duchon, C. E., & Biddle, C. J. (2010). Undercatch of tipping-bucket gauges in high rain rate events. *Advances in Geosciences*, *25*, 11–15. https://doi.org/10.5194/adgeo-25-11-2010

Duchon, C., Fiebrich, C., & Grimsley, D. (2014). Using high-speed photography to study undercatch in tipping-bucket rain gauges. *Journal of Atmospheric and Oceanic Technology*, *31*(6), 1330–1336. https://doi.org/10.1175/JTECH-D-13-00169.1

Dunkerley, D. (2008). Identifying individual rain events from pluviograph records: A review with analysis of data from an Australian dryland site. *Hydrological Processes*, *22*(26), 5024–5036. https://doi.org/10.1002/hyp.7122

Ebert, E. E. (2007). Methods for verifying satellite precipitation estimates. In V. Levizzani, P. Bauer, & F. J. Turk (Eds.), *Measuring precipitation from space* (pp. 345–356) Netherlands: Dordrecht: Springer.

England, J. F., Jr., Cohn, T. A., Faber, B. A., Stedinger, J. R., Thomas Jr., W. O., Veilleux, A. G., Kiang, J. E., & Mason, R. R., Jr. (2019). Guidelines for determining flood flow frequency Bulletin 17C. In Guidelines for determining flood flow frequency Bulletin 17C (USGS Numbered Series No. 4-B5; Techniques and Methods, p. 168). U.S. Geological Survey.

Fornasiero, A., Amorati, R., Alberoni, P. P., Ferraris, L., & Taramasso, A. C. (2004). Impact of combined beam blocking and anomalous propagation correction algorithms on radar data quality. Proceedings of ERAD (pp. 216–222).

Froude, M. J., & Petley, D. N. (2018). Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, *18*(8), 2161–2181. https://doi.org/10.5194/nhess-18-2161-2018

Galanti, Y., Barsanti, M., Cevasco, A., D'Amato Avanzi, G., & Giannecchini, R. (2018). Comparison of statistical methods and multitime validation for the determination of the shallow landslide rainfall thresholds. *Landslides*, *15*(5), 937–952. https://doi.org/10.1007/s10346-017-0919-3

Gutmann, E., Pruitt, T., Clark, M. P., Brekke, L., Arnold, J. R., Raff, D. A., & Rasmussen, R. M. (2014). An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resources Research*, *50*(9), 7167–7186. https://doi.org/10.1002/2014WR015559

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2008). The rainfall intensity duration control of shallow landslides and debris flows: An

update. *Landslides*, *5*(1), 3–17. https://doi.org/10.1007/s10346-007-0112-1

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorology and Atmospheric Physics*, *98*(3–4), 239–267. https://doi.org/10.1007/s00703-007-0262-7

Highland, L., & Bobrowsky, P. (2008). *The landslide handbook*: *A guide to understanding landslides* (Circular No. 1325; Circular, p. 129). United States Geological Survey.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., & Iguchi, T. (2014). The global precipitation measurement Mission. *Bulletin of the American Meteorological Society*, *95*(5), 701–722. https://doi.org/10.1175/BAMS-D-13-00164.1

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K.-L., Joyce, R. J., Kidd, C., Nelkin, E. J., Sorooshian, S., Stocker, E. F., Tan, J., Wolff, D. B., & Xie, P. (2020). Integrated multi-satellite retrievals for the global precipitation measurement (GPM) mission (IMERG). In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds.), *Satellite precipitation measurement: Volume 1* (pp. 343–353). Springer International Publishing.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., & 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. https://doi.org/10.1175/JHM560.1

Janssen, E., Wuebbles, D. J., Kunkel, K. E., Olsen, S. C., & Goodman, A. (2014). Observational- and model-based trends and projections of extreme precipitation over the contiguous United States. *Earth's Future*, *2*(2), 99–113. https://doi.org/10.1002/2013EF000185

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. https://doi.org/10.1175/BAMS-D-14-00283.1

Kidd, C., Takayabu, Y. N., Skofronick-Jackson, G. M., Huffman, G. J., Braun, S. A., Kubota, T., & Turk, F. J. (2020). The global precipitation measurement (GPM) mission. In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds.), *Satellite precipitation measurement* (Vol. 67, pp. 3–23). Springer International Publishing.

Kirschbaum, D. B., Adler, R., Hong, Y., Hill, S., & Lerner-Lam, A. (2010). A global landslide catalog for hazard applications: Method, results, and limitations. *Natural Hazards*, *52*(3), 561–575. https://doi.org/10.1007/s11069-009-9401-4

Kirschbaum, D. B., Adler, R., Hong, Y., Kumar, S., Peters-Lidard, C., & Lerner-Lam, A. (2012). Advances in landslide nowcasting: Evaluation of a global and regional modeling approach. *Environmental Earth Sciences*, *66*(6), 1683–1696. https://doi.org/10.1007/s12665-011-0990-3

Kirschbaum, D., & Stanley, T. (2018). Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness. *Earth's Future*, *6*(3), 505–523. https://doi.org/10.1002/2017EF000715

Kummerow, C., Barnes, W., Kozu, T., Shiue, J., & Simpson, J. (1998). The tropical rainfall measuring mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, *15*(3), 809–817. https://doi.org/10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2

Leonarduzzi, E., Molnar, P., & McArdell, B. W. (2017). Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data. *Water Resources Research*, *53*(8), 6612–6625. https://doi.org/10.1002/2017WR021044

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D. R., & Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950. *Scientific Data*, *2*(1), 150042. https://doi.org/10.1038/sdata.2015.42

Lockhoff, M., Zolina, O., Simmer, C., & Schulz, J. (2014). Evaluation of satellite-retrieved extreme precipitation over Europe using gauge observations. *Journal of Climate*, *27*(2), 607–623. https://doi.org/10.1175/JCLI-D-13-00194.1

Long, D., Longuevergne, L., & Scanlon, B. R. (2014). Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resources Research*, *50*(2), 1131–1151. https://doi.org/10.1002/2013WR014581

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. https://doi.org/10.1175/JHM-D-15-0190.1

Manzanas, R., Amekudzi, L. K., Preko, K., Herrera, S., & Gutiérrez, J. M. (2014). Precipitation variability and trends in Ghana: An intercomparison of observational and reanalysis products. *Climatic Change*, *124*(4), 805–819. https://doi.org/10.1007/s10584-014-1100-9

Mirus, B. B., Jones, E. S., Baum, R. L., Godt, J. W., Slaughter, S., Crawford, M. M., Lancaster, J., Stanley, T., Kirschbaum, D. B., Burns, W. J., Schmitt, R. G., Lindsey, K. O., & McCoy, K. M. (2020). Landslides across the USA: Occurrence, susceptibility, and data limitations. *Landslides*, *17*(10), 2271–2285. https://doi.org/10.1007/s10346-020-01424-4

Nikahd, A., Hashim, M., & Nazemosadat, M. J. (2016). A review of uncertainty sources on weather ground-based radar for rainfall estimation. In *Applied mechanics and materials (Vol. 818*, 254–271). Trans Tech Publications Ltd.

North America Elevation 1-Kilometer Resolution. (2007). [Map]. Commission for Environmental Cooperation.

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 and Earth System Sciences*, *21*(12), 6559–6572. https://doi.org/10.5194/hess-21-6559-2017

Pendergrass, A. G., & Knutti, R. (2018). The uneven nature of daily precipitation and its change. *Geophysical Research Letters*, *45*(21), 980–988. https://doi.org/10.1029/2018GL080298

Pollock, M. D., O'Donnell, G., Quinn, P., Dutton, M., Black, A., Wilkinson, M. E., Colli, M., Stagnaro, M., Lanza, L. G., Lewis, E., Kilsby, C. G., & O'Connell, P. E. (2018). Quantifying and mitigating wind-induced undercatch in rainfall measurements. *Water Resources Research*, *54*(6), 3863–3875. https://doi.org/10.1029/2017WR022421

Premchitt, J., Brand, E. W., & Phillipson, H. B. (1986). Landslides caused by rapid groundwater changes. *Geological Society*, *London*, *Engineering Geology Special Publications*, *3*(1), 87–94. https://doi.org/10.1144/GSL.ENG.1986.002.01.09

Rossi, M., Kirschbaum, D., Valigi, D., Mondini, A. C., & Guzzetti, F. (2017). Comparison of satellite rainfall estimates and rain gauge measurements in Italy, and impact on landslide modeling. *Climate*, *5*(4), 90. https://doi.org/10.3390/cli5040090

Sapiano, M. R. P., & Arkin, P. A. (2009). An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *Journal of Hydrometeorology*, *10*(1), 149–166. https://doi.org/10.1175/2008JHM1052.1

Scheevel, C. R., Baum, R. L., Mirus, B. B., & Smith, J. B. (2017). Precipitation thresholds for landslide occurrence near Seattle, Mukilteo, and Everett, Washington (Open-File Report No. 2017-1039; Open-File Report). U.S. Department of the Interior; U.S. Geological Survey.

Segoni, S., Rossi, G., Rosi, A., & Catani, F. (2014). Landslides triggered by rainfall: A semi-automated procedure to define consistent intensity duration thresholds. *Computers & Geosciences*, *63*, 123–131. https://doi.org/10.1016/j.cageo.2013.10.009

Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., Kakar, R., Braun, S. A., Huffman, G. J., Iguchi, T., Kirstetter, P. E., Kummerow, C., Meneghini, R., Oki, R., Olson, W. S., Takayabu, Y. N., Furukawa, K., & Wilheit, T. (2017). The global precipitation measurement (GPM) mission for science and society. *Bulletin of the American Meteorological Society*, *98*(8), 1679–1695. https://doi.org/10.1175/BAMS-D-15-00306.1

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. https://doi.org/10.1002/2017RG000574

Sunyer, M. A., Hundecha, Y., Lawrence, D., Madsen, H., Willems, P., Martinkova, M., Vormoor, K., Bürger, G., Hanel, M., Kriaučiūnienė, J., Loukas, A., Osuch, M., & Yücel, I. (2015). Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe. *Hydrology and Earth System Sciences*, *19*(4), 1827–1847. https://doi.org/10.5194/hess-19-1827-2015

Tajudin, N., Ya'acob, N., Mohd Ali, D., & Adnan, N. A. (2020). Estimation of TRMM rainfall for landslide occurrences based on rainfall threshold analysis. *International Journal of Electrical and Computer Engineering (IJECE)*, *10*(3), 3208. https://doi.org/10.11591/ijece.v10i3.pp3208-3215

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T., Angelis, C. F., Salio, P., Kidd, C., Huffman, G. J., & de Castro, M. (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, *104–105*, 70–97. https://doi.org/10.1016/j.atmosres.2011.10.021

Tryhorn, L., & DeGaetano, A. (2011). A comparison of techniques for downscaling extreme precipitation over the northeastern United States. *International Journal of Climatology*, *31*(13), 1975–1989. https://doi.org/10.1002/joc.2208

US Department of Commerce. (2013). NOAA Atlas 2 precipitation frequency estimates in GIS compatible formats. In NOAA Atlas 2 Precipitation Frequency Estimates in GIS Compatible Formats. US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service. Retrieved from https://www.nws.noaa.gov/ohd/hdsc/noaaatlas2.htm

Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, M. J., Williams, C. N., Fenimore, C., Gleason, K., & Arndt, D. (2014). Improved historical temperature and precipitation time series for U.S. climate divisions. *Journal of Applied Meteorology and Climatology*, *53*(5), 1232–1251. https://doi.org/10.1175/JAMC-D-13-0248.1

Wang, G., Kirchhoff, C. J., Seth, A., Abatzoglou, J. T., Livneh, B., Pierce, D. W., Fomenko, L., & Ding, T. (2020). Projected changes of precipitation characteristics depend on downscaling method and training data: MACA versus LOCA using the U.S. northeast as an example. *Journal of Hydrometeorology*, *21*, 20.

Xia, Y., Cosgrove, B. A., Mitchell, K. E., Peters-Lidard, C. D., Ek, M. B., Brewer, M., Mocko, D., Kumar, S. V., Wei, H., Meng, J., & Luo, L. (2016). Basin-scale assessment of the land surface water budget in the National Centers for Environmental Prediction operational and research NLDAS-2 systems. *Journal of Geophysical Research: Atmospheres*, *121*(6), 2750–2779. https://doi.org/10.1002/2015JD023733

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., & Mocko, D. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, *117*(D03109), 1–27. https://doi.org/10.1029/2011JD016048

Yu, F.-C., Chen, T.-C., Lin, M.-L., Chen, C.-Y., & Yu, W.-H. (2006). Landslides and rainfall characteristics analysis in Taipei City during the typhoon Nari event. *Natural Hazards*, *37*(1–2), 153–167. https://doi.org/10.1007/s11069-005-4661-0

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., Grams, H., Wang, Y., Cocks, S., Martinaitis, S., Arthur, A., Cooper, K., Brogden, J., & Kitzmiller, D. (2015). Multi-radar multi-sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, *97*(4), 621–638. https://doi.org/10.1175/BAMS-D-14-00174.1

How to cite this article: Culler, E. S., Badger, A. M., Minear, J. T., Tiampo, K. F., Zeigler, S. D., & Livneh, B. (2021). A multi-sensor evaluation of precipitation uncertainty for landslide-triggering storm events. *Hydrological Processes*, *35*(7), e14260. https://doi.org/10.1002/hyp.14260