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

Elsa Culler - Cooperative Institute for Research in Environmental Science (CIRES), University of Colorado Boulder, Boulder, CO 80309, USA; Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, CO 80309, USA

Andrew Badger - Universities Space Research Association, Columbia, MD, 21046 USA; Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, 20771 USA

Toby Minear - Cooperative Institute for Research in Environmental Science (CIRES), University of Colorado Boulder, Boulder, CO 80309, USA

Kristy Tiampo - Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, Boulder, CO 80309, USA; Department of Geological Sciences, University of Colorado Boulder, Boulder, CO 80309, USA

Ben Livneh - Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, CO 80309, USA; Cooperative Institute for Research in Environmental Science (CIRES), University of Colorado Boulder, Boulder, CO 80309, USA

Spencer Ziegler - Department of Geological Sciences, University of Colorado Boulder, Boulder, CO 80309, USA

Extreme precipitation can have profound consequences for communities, resulting in flooding and rainfall-triggered landslides that cause casualties and extensive damage each year. A key challenge to understanding and predicting these natural hazards comes from uncertainties in the depth and intensity of precipitation preceding the landslide event. Practitioners and researchers must select among a wide range of precipitation products, often with little guidance. Here we investigate the degree of precipitation uncertainty across multiple precipitation products for a large set of landslide-triggering storm events and assess the impact of 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 228 reported landslide e across the continental US and Canada. Precipitation data are taken from four products that cover disparate measurement methods: near real-time and post-processed satellite (Global Precipitation Mission IMERG Early and Final calibrated precipitation), radar (Multi-Radar Multi-Sensor gauge bias-corrected precipitation), and gauge-based (North American Land Data Assimilation System v. 2 Forcing precipitation). These products also cover a range of spatial and temporal resolutions as well as spatial extent and near real-time or longer latency of data availability. Landslide-triggering precipitation was found to vary extensively on the basis of the measurement source with the depth of individual storm events diverging by as much as 247 mm with an average range of 38 mm. Peak intensity measurements, which are typically influential in triggering landslides, were also highly variable with an average range of 8.8 mm/hr and at times as much as 72 mm/hr. The two products using more ground-based observations (MRMS and NLDAS2) tended to perform better at identifying landslides using the published intensity and duration storm thresholds, but all products exhibited hit-ratios of greater than 0.68. While not all storms predicted landslides successfully, a greater proportion of predicted landslides were seen when sub-setting the data to include only verified landslide locations. Overall, we recommend practitioners consider low-latency products like MRMS for investigating landslides, given their near-realtime data availability and good performance in detecting landslides, although practitioners would be well-served considering more than one product as a way to confirm intense storm signals and minimize the influence of noise.

# 1 Introduction

Precipitation measurements are intrinsically linked with the study and mitigation of rainfall-triggered landslides because these natural disasters occur in places that are otherwise susceptible to mass movements as a result of some combination of excess runoff and saturation of the soil column (Highland and Bobrowsky 2008). In spite of the destructive nature of landslides, which cause tens of thousands of deaths each year (Froude and Petley 2018)these events remain challenging to forecast in part due to uncertainty in precipitation leading up to the event (D. Kirschbaum and Stanley 2018). There are many other sources of uncertainty that contribute to poor landslide predictions such as unknown soil properties, vegetation, and anthropogenic modifications to surface and subsurface soil structure. However, perhaps the largest source of uncertainty in landslide probability estimates is hydrologic uncertainty, defined here as uncertainty in the depth and intensity of liquid precipitation leading up to the event (Chowdhury and Flentje 2002). A confounding factor is the wide array of precipitation datasets ranging from in situ observations, ground-based radar and satellite retrievals. The goal of this analysis is to investigate the role of precipitation uncertainty preceding known historical landslide events, and to assess the implications of that uncertainty for evaluating landslide hazards. Greater understanding areas of relative agreement and divergence across products may provide guidance to practitioners and researchers choosing precipitation products for studying landslides.

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 how much precipitation has landed in the gauge. Their main strength is that they directly measure the amount of collected water, but nonetheless they suffer from issues of persistant bias driven by by under-catch from wind (Pollock et al. 2018) instrument malfunctions C. E. Duchon and Biddle (2010), placement of gauges too close to other structures (Vose et al. 2014), ), and sparse sensor density relative to surrounding areas limiting representativeness (Kidd et al. 2017). Ground-based radar can detect precipitation based on propagation and backscatter of radar, and therefore can detect subtle variations in precipitation potentially hundreds of kilometers away. However, radar is an indirect measurement of precipitation that requires conversion of the radar signal to precipitation volume and is further limited by beam blockage and interference from buildings or even insects in the radar’s path (Fornasiero et al. 2004; Bousquet and Smull 2003; Nikahd, Hashim, and Nazemosadat 2016). Most ground-based radars use multiple bands of radar and multiple polarities in order to compute the raindrop shape and size distributions used in the processing, which offers an advantage over other indirect techniques such as 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, and depending on the sensor type these can be deployed in either geostationary or low Earth orbits that cover particular regions at particular intervals (Huffman et al. 2020). The key advantage of satellite-based precipitation measurements is that unlike ground-based in situ or radar sensors they can deliver frequent, spatially continuous, precipitation measurements, although typically multiple satellites (Tapiador et al. 2012) with a variety of sensors and orbits (Ashouri et al. 2015) are required to provide global coverage. 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 drop size distributions is also a challenge, though it can be addressed through the use of either ground- or satellite-based radar.

Precipitation measurements have been compared on the basis of a number of metrics in prior studies ranging from annual and monthly totals (Robert F. Adler et al. 2001) to the largest number of consecutive dry days (Manzanas et al. 2014). Less attention has been paid to metrics most directly useful for predicting and understanding rainfall-triggered landslides. While some landslides are triggered by short, intense precipitation events, others are triggered by saturation of the soil column that can develop over a longer period of time (Susan H. Cannon and 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 of intense rainfall. As a result, this study focuses on precipitation products with hourly or finer temporal resolution to facilitate an evaluation of individual storm events.

Existing precipitation intercomparisons often focus on specific applications, for example for evaluating grid-based products over complex terrain, or relevant for portraying hydrologic phenomena Ahmadalipour and Moradkhani (2017) for climate model downscaling efforts (Gutmann et al. 2014; Wang et al. 2020) or for merging 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, i.e. a deviation of 300 mm in annual precipitation among satellite products. They conclude that the placement and 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. Robert F. Adler et al. (2003) similarly analyzed 31 gauge-based, satellite-based, model-based, and climatological datasets in terms of monthly precipitation, finding that ‘quasi-standard’ products, e.g. those like the Global Precipitation Measurement mission (GPM) (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 [GPCP] ) perform better than products based on a single data source.

Fewer studies comparing extreme precipitation exist, with many focusing on climate model simulations (Sunyer et al. 2015; Tryhorn and DeGaetano 2011) and trends (Janssen et al. 2014; Bao et al. 2017) while others focusing on observations and satellites (Pendergrass and Knutti 2018; AghaKouchak et al. 2011; Lockhoff et al. 2014). AghaKouchak et al. (2011) compared extreme precipitation across four satellite platforms finding tradeoffs across products in terms of correct identification of precipitation above a threshold and measurements of the volume of identified extreme storms. Though they found 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 a significant amount of precipitation. Lockhoff et al. (2014) found that satellite retrieved extreme precipitation values 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 uneven 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 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 rise to the level of a natural disaster (Sun et al. 2018; Manzanas et al. 2014).

The intensity-duration threshold is a type of two-parameter statistical model used for landslide early warning systems, where rainstorms 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 = a D^{-b}) and are valid in a particular region or climate and for a range of durations based on the training data (Fausto 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, Molnar, and McArdell 2017). Here, we will investigate several power-law intensity-duration thresholds reviewed by Fausto Guzzetti et al. (2008) as a basic way to compare precipitation measurements from different sources in the context of concerns specific to landslide hazard estimation. Furthermore, when precipitation is used to provide warning systems or guide recovery efforts from landslides, the timeliness, i.e. latency, of the information matters (D. B. Kirschbaum et al. 2012), such that the issue of latency will also be considered in the investigation of intensity-duration thresholds.

In a comparison of satellite and gauge precipitation data by specific to landslide sites in Italy, Rossi et al. (2017) used intensity-duration thresholds as an aspect of the precipitation 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. They concluded that the satellite data are still useful for forecasting landslides using intensity-duration thresholds as long as they are scaled appropriately to correct for local bias.

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. This work addresses an existing gap in evaluating extreme precipitation through the lens of natural hazards, while conducting inter-product analyses into storm characteristics of broader relevance for the hydrological community. This work furthers the analysis by Rossi et al. (2017) by including ground-based radar and by rigorously analyzing each precipitation estimate preceding specific landslide events.

In sec. 2, we present the criteria used to select landslide sites and precipitation products, followed by procedures for splitting precipitation into storms and the metrics used in the comparison. Sec. 3 begins with an exposition into cumulative observed precipitation over the 30-days preceding the landslide for five characteristic example sites. Next, we compare each product using storm characteristics of total depth, duration, total intensity, peak intensity, and return period. We further test whether peak intensity might be accounting for low return period storms causing landslides by comparing the two quantities. Finally, we use p intensity-duration thresholds to test which products best detect landslides, comparing the hit ratio and the false alarm ratio for each product and threshold.

# 2 Methods

The overall goals of the methods are to evaluate precipitation characteristics at known landslide sites by first examining the features of triggering storms and then subsequently comparing precipitation estimates in the context of intensity-duration thresholds of landslide occurrence. Rainfall-triggered landslide sites were chosen from the NASA Global Landslide Catalog (NASA GLC) with a subset of landslide locations verified with ancillary satellite imagery (see sec. 2.1). For each landslide location, precipitation was obtained from four different products (see sec. 2.2) and the precipitation time series were split into individual storm events, with key characteristics of total depth, duration, intensity, peak intensity, and return period calculated relative to a reference dataset (sec. 2.3). Finally, the storm events were plotted relative to landslide intensity-duration curves, with hit-ratios and false-alarm-ratios computed for each model-product combination (sec. 2.4).

## 2.1 Study domain and landslide site selection

Landslides were selected from the NASA Global Landslide Catalog (NASA GLC) (D. B. Kirschbaum et al. 2010). The NASA GLC was chosen for this study, since it provides a large sample of landslide locations useful for evaluating heavy rainfall events. The NASA GLC shares many strengths and weaknesses with other regional and global databases available (D. B. Kirschbaum et al. 2010; Mirus et al. 2020). Though the NASA GLC covers a broad spatial and temporal domain, it suffers from problems of precision and completeness. The source of these data are a collection of second-hand observations made by organizations like the news media, governmental organizations like departments of transportation, along with some available scientific reports. This means that landslides that are causing problems for people are reported more frequently, resulting in a substantial spatial bias towards populated areas. Some of the landslide reports suffer from relatively high location error (as much as 50 km) depending on how specific the source article is about the location (D. B. Kirschbaum et al. 2010), and this was evident in that it was only possible to definitively locate one third of the landslides in this study on satellite imagery. Despite these limitations, it was deemed fit for purpose for this study which is not to study a selection of landslides with a representative spatial distribution but rather to compare precipitation products in the vicinity of hydrologically triggered landslides where heavy rainfall events are likely to be present. Overall, the NASA GLC provide a substantial number of landslide locations for this study that met the following selection criteria:

* Only landslide events reported as rainfall-driven, with a NASA GLC trigger category of rain, downpour, continuous rain, or flooding were included. Snow-related triggers were not included even though these are hydrologically related, but precipitation is not contemporaneously linked with landslide triggers;
* Landslide events took place in the continental United States (CONUS) or Canada below N and after May 2015 ensuring data availability across each of the selected precipitation products; and
* The landslide location accuracy was estimated to be km or less according to the NASA GLC. The value of 10km was chosen since it is approximately equal to the spatial resolution of two of the precipitation products.

In total, 228 landslides were selected. Of those, 80 were also 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 NASA GLC was used for the remaining landslides where 31 were marked in the NASA GLC as exact locations, 51 as 1 km, 52 as 5 km, and 14 as 10 km accuracy. Fig. 1 shows that many of the sites are located near the Pacific coast, likely due to the complex topography associated with landslides, as well as the population reporting bias of the NASA GLC. The verified landslides are generally distributed fairly evenly relative to the locations of the full selection of landslides.

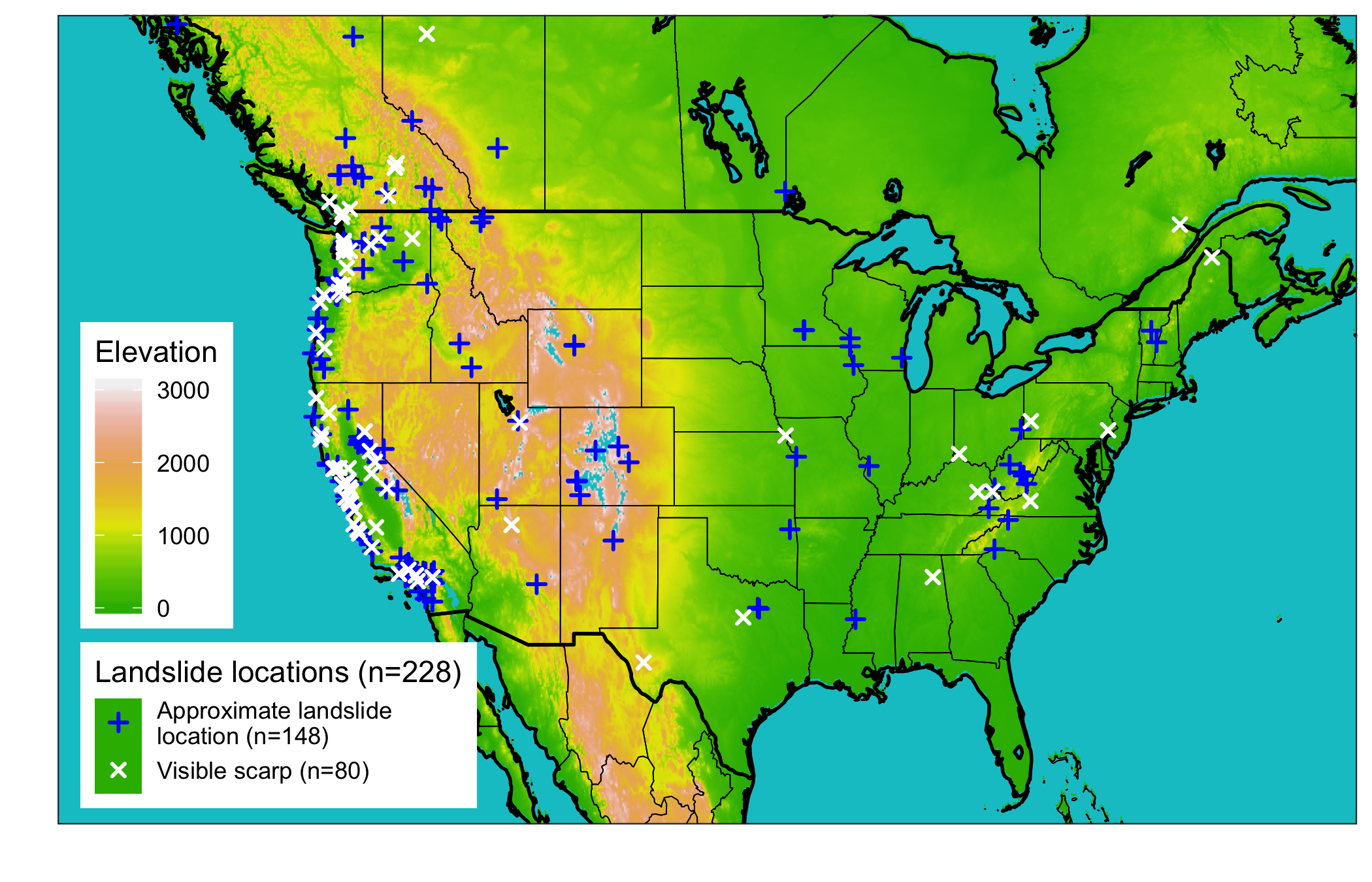


Figure 1: Map of all landslide sites considered in this analysis (n=228), colored by whether the location was approximate (n=148) or verified using aerial satellite imagery to identify a visible scarp (n=80); Source of landslide locations was the NASA GLC (D. B. Kirschbaum et al. 2010), source of the DEM data used for the basemap (“North America Elevation 1-Kilometer Resolution” 2007).

## 2.2 Precipitation data sources

The gridded precipitation datasets selected for this study were chosen to be reflective of three common measurement methods: gauges, ground-based radar, and satellite. In addition, we focused on products that were both freely available, having undergone extensive verification, and with coverage over at least the CONUS. An important additional criteria 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 precipitation products and features described in tbl. 1 and summarized in the following paragraphs.

### 2.2.1 North American Land Data Assimilation System version 2 (NLDAS-2) meteorological dataset

The NLDAS-2 meteorological dataset 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 (Xia et al. 2012). The gauge-based estimates are are disaggregated to hourly using the radar-based estimates, resulting in a near real-time hourly gridded product at (~ km) resolution across North America going back to 1979 with a latency of approximately 4 days. Though it has low horizontal resolution relative to the other precipitation products used here, NLDAS-2 meteorological is a widely used gauge-based product that has been extensively validated over the 20 years.

### 2.2.2 Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimate

MRMS precipitation estimates are primarily based on a centralized radar mosaic with 2 minute resolution over the US 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 (~1.1 km) and represents the state of the art in terms of leveraging computing resources to take advantage of a multitude of overlapping radar and other types of sensors.

### 2.2.3 Global Precipitation Mission (GPM) Integrated Multi-satellitE Retrievals for Global (IMERG) precipitation measurement

GPM IMERG precipitation estimates are a combination of multiple satellite measurements, including the GPM Core Observatory Microwave Imager which is considered the standard for other included satellites. In addition to active and passive microwave sensors, IMERG estimates include Infrared sensors, satellite-based radar, and precipitation gauge adjustments. The gauges are used for monthly bias correction (Huffman et al. 2020). There are 3 IMERG products, Early, Late, and Final, of which we use the Early (~4 hour latency) and the Final (~3.5 month latency) in this study. The IMERG-Early product is available much more quickly than the IMERG-Final, but as a result some of the satellite retrievals are not included because they have not yet arrived, and it cannot take advantance of some processing steps or monthly gauge correction (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. IMERG-Early also has extremely small latency, making it the most suitable among the products explored here for operational landslide modeling in the context of near real-time data availability.

Table 1: 3 The four precipitation products included in the comparison, representing gauge-, radar-, and satellite-based measurements

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Precipitation product | Description | Spatial Resolution | Temporal resolution | Typical Latency |
| 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. | (~10 km) | 30 minutes | 4 hours |
| Integrated Multi-satellitE Retrievals for Global precipitation measurement (IMERG-Final)(Hou et al. 2014) | In addition to the satellite data included in the IMERG early run, the final run includes late-arriving microwave overpasses, monthly gauge-based adjustments, and an algorithm that interpolates forward as well as backward in time. | (~10 km) | 30 minutes | 3.5 months |
| Multi-Radar Multi-Sensor (MRMS) (Zhang et al. 2015) | Integrates data from radars, satellites, precipitation gages, and other sensors to provide real-time decision support | (~1.1 km) | 2 minutes | < 5 minutes |
| North American Land Data Assimilation System version 2 (NLDAS-2) meteorological (Xia et al. 2012) | Disaggregation of Climate Prediction Center daily precipitation using bias-corrected radar | (~ 12 km) | 1 hour | 4 days |

## 2.3 Precipitation inter-comparison and computation of storm characteristics

For each of the above precipitation products, data were extracted for the nearest grid location for the period between May 2015 (the earliest date MRMS data are available) and May 2020 (the latest release of IMERG-Final data). A minimum threshold of 1 mm was applied to the precipitation data to reduce noise. The data were then split into storm events, where a gap of at least 24 hours was considered to mark the end of one storm and the beginning of the next.

For each storm, the characteristics of total depth, duration, intensity, and peak intensity were computed and compared. The peak intensity for a storm was the intensity of the single maximum precipitation measurement of the storm. Depth and frequency were chosen since they reflect the most common metrics used in extreme hydrologic events in general (England Jr. et al. 2019). Intensity and duration were included because they are parameters commonly used to forecast rainfall-triggered landslides (D. B. Kirschbaum et al. 2012). Some studies have suggested that for certain landslides high peak intensity can contribute significantly to triggering a landslide independent of the overall storm depth, duration or intensity (Yu et al. 2006; Corominas, Moya, and Hürlimann 2002). This idea is supported for example by observations that landslides are commonly initialized within hours of the peak intensity (Premchitt, Brand, and Phillipson 1986). The precipitation rank and z-score among the four products for each landslide event were also computed for the day of the landslide as well as for the full precipitation record. Rank was chosen as in indicator of the relative magnitude of each product relative to the others, and the z-score is an indicator of the variability of each product relative to the others.

To facilitate comparison of storm characteristics within a single over-arching framework, the return period of the landslide-triggering storms was also 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 hours. In order 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 round the storm duration to one of the NOAA atlas durations which might have artificially lowered the return periods. For example, for the 3-hour duration, cumulative 3-hour 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 looked up in the NOAA atlas. We then selected the maximum return period from among the 10 possible durations noted above for each landslide. This procedure ensured that we used the maximum applicable return period available from the NOAA atlas that occurred during each landslide-triggering storm.

## 2.4 Performance of intensity-duration thresholds using different precipitation products

Intensity-Duration thresholds represent simple models of landslide occurence whereby a threshold is defined as a power law of the storm duration (, where I is intensity and D is duration), where either raw or normalized intensities above the threshold predict the occurance of a landslide (Segoni et al. 2014). Thresholds have been calculated under different climates and over multiple scales, including globally D. B. Kirschbaum et al. (2012). Three thresholds for this study F. Guzzetti et al. (2007) were obtained from a review by Fausto Guzzetti et al. (2008). Thresholds were used on applicable subsets of the data based on climate or other conditions. For example, a threshold for “mild, marine west coast climates” was only used on data west of longitude 155W. For each threshold-product combination, we computed a hit ratio (correctly predicted landslides over the total number of landslides) and a false alarm ratio (incorrectly predicted landslides over the total number of non-landslides)

# 3 Results

## 3.1 Precipitation pre-landslide time series

Fig. 2 shows the cumulative precipitation in the 30-days before a landslide at 5 example sites. The selected sites showcase a variety of ways in which the precipitation from multiple products can differ. For example, while the preciptiation in panel (a) matches closely for all products, in panel (d) there is a wide spread of over 100 mm of precipitation that happens suddenly during the landslide-triggering event. The preciptiation in panel (b) also demonstrates a factor of 6 spread of precipitation values but appears to be more strongly correlated than other sites. In panel (c) the IMERG-Early product separates substatially but ultimately the 30-day depth is nearly identical for all products. Panel (e) shows a likely landslide location error since none of the products register any precipitation at all. We note that the differences 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 could be because using products from different categories introduces much more variability, or that the large landslide-triggering storms have a greater potential for error by virtue of containing more depth overall than other storms).

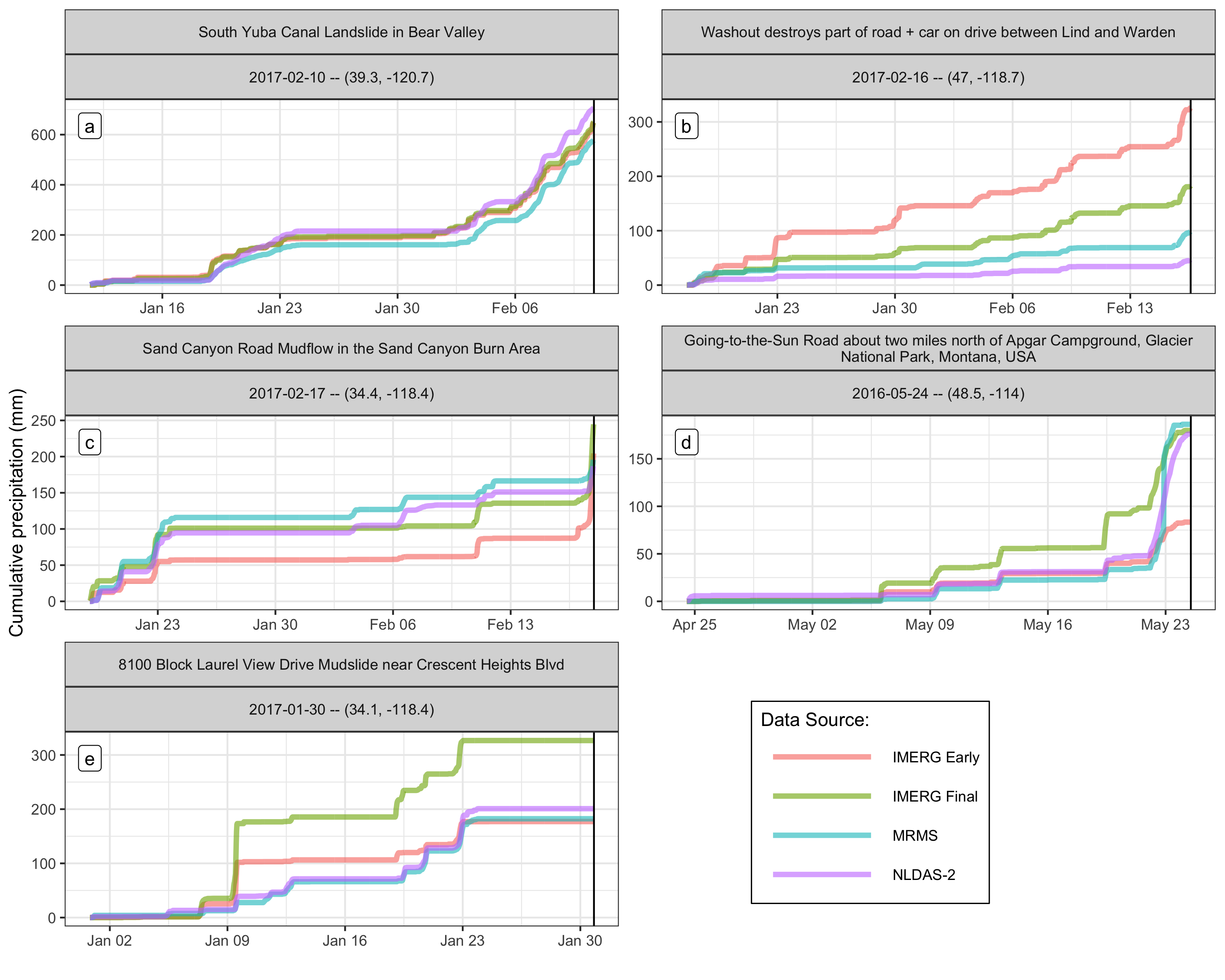


Figure 2: **Exposition into the types of precipitation differences leading up to landslide events:** Cumulative precipitation measurements at select landslide sites for the 30 days before the event. The precipitation is variable across the different products, and the selected sites each demonstrate diverse types of variability. Panel (a) shows a site where there were 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) shows the IMERG-Early product reports nearly 50mm less cumulative precipitation leading into the landslide-triggering storm, but then makes up the difference by detecting much more precipitation immediately before the landslide. In panel (d) there is a wide spread of approximately half the maximum total amount of precipitation. Finally, in panel (e) no landslide-triggering precipitation was detected by any product, suggesting a location error in the landslide record.

The relative magnitude of the different precipitation products on the day of the landslide is shown in Fig. 3 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 IMERG products appear to have identical ranks, which exceeds MRMS and is less than NLDAS-2 measurements. IMERG-Early has the highest z-scores among day-of-landslide preciptiation, suggesting that the further interpolation in the IMERG-Final product reduces these outliers relative to IMERG-Early, although the median and third quartile values for IMERG-Final are the larger overall. There are larger z-scores in the entire record than the most extreme day-of-landslide precipitation, but for all products the bulk of the day-of-landslide precipitation is larger than most of the non-landslide-triggering precipitation.

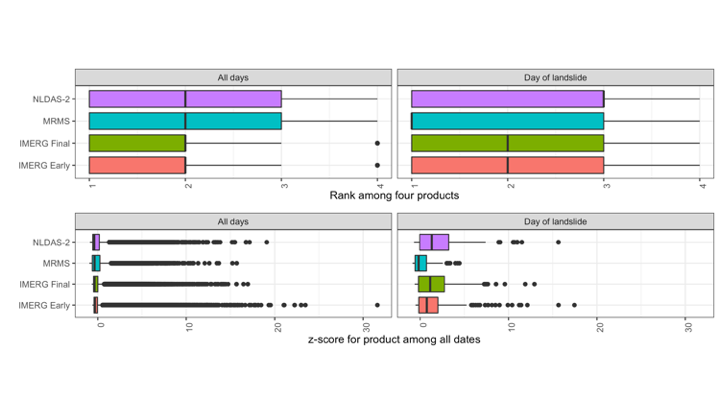


Figure 3: **Relative magnitude of precipitation products on the day of the landslide**: Rank among all products for the day, and z-score of precipitation as measured by each product for each of 228 events. Metrics are divided by the day of the landslide and the rest of the days.

Fig. 4 shows the characteristics of the landslide-triggering storms plotted against the ensemble mean of all the products by landslide event. Included are values for all the landslide sites and for the verified locations alone. The IMERG products generally report higher peak hourly intensities, which is likely at least partially due to the shorter 30-minute time step. However, the higher peak intensities are more clearly reflected in longer return periods, which are based on hourly durations or longer for comparison with the NOAA Atlas. MRMS and NLDAS-2 seem to have even lower return periods among the verified locations, suggesting that these products have difficulty detecting high return period precipitation consistently.

In general there appears to be good agreement among products on the depth and duration of storms, with the exception of outliers below 10mm of total depth—which is a fairly modest storm depth. Among the verified locations, there are fewer low depth or duration values that are either outliers or near to the mean, suggesting that low measurements may reflect limitations in the NASA GLC location accuracy for sites with only approximate locations.

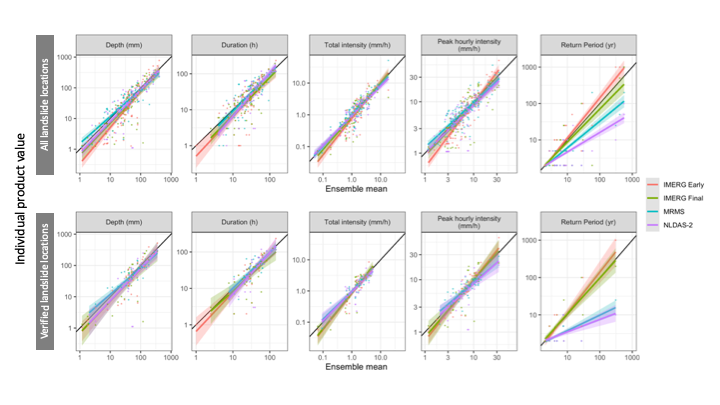


Figure 4: **Storm characteristics vs. the ensemble mean:** Depth, duration, intensity, peak intensity, and return period for each of the landslide-triggering storms as measured by four precipitation products. Least-squares regression lines with confidence intervals are also shown. Top panels show all 228 sites while the bottom panel only shows the 80 verified locations.

Fig. 5 shows the precipitation on intensity/duration axes with three intensity-duration thresholds plotted over them, and tbl. 2 summarizes the performance of each threshold. The choice of threshold does not appear to make a large difference in this context, since the models are very similar when compared to the variation in precipitation data across sites and among products. These models tend to perform better using MRMS or NLDAS-2 data than using either IMERG product, with hit ratios of 0.88 and 0.76 rather than 0.70 and 0.68 among the verified landslide locations. All products perform better when using only the verified landslide sites.

The precipitation products are examined in the context of landslide triggering threholds in Fig. 6, which shows the precipitation on intensity vs. duration axes with three intensity-duration thresholds plotted over them, with the performance summarized in Table 2. Interestingly, the choice of intensity-duration threshold does not appear to make a large difference in this context, since the threshold curves are very similar when compared to the variation in precipitation data across sites and among products. The MRMS or NLDAS-2 products tend to perform better than either IMERG product, with hit ratios of 0.88 and 0.76 rather than 0.70 and 0.68 among the verified landslide locations, respectively. All products perform better when using only the verified landslide locations relative to the approximate locations.

There is a concentration of long-duration, low-intensity storms around 24 hours for all products that may an artifact of the storm identification algorithm. Since the landslides did not have times specified, the whole day of the landslide was always included unless there was no rain all the way until the end of the day. This could have resulted in lower total intensity values for storms that lasted only through the day of the landslide but tapered off towards the end of the day. Many of the storms that did trigger landslides but were not correctly identified by the intensity-duration threshold fall into this section of approximately 24-hour low-intensity storms. Improvements to storm delineation through a different algorithm or a higher minimum threshold might boost performance. The IMERG data in particular might benefit from such improvements.

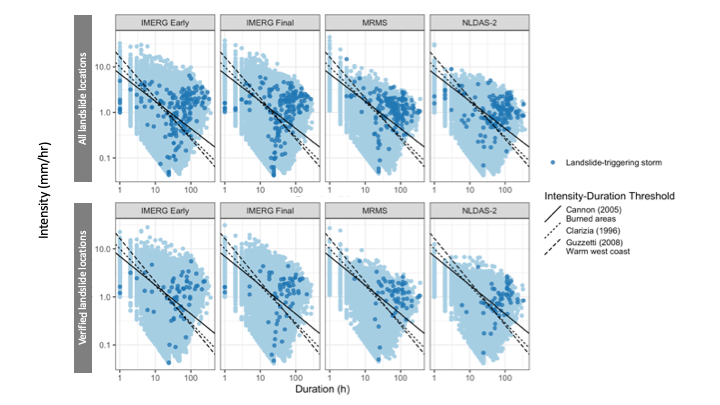


Figure 5: Each storm in the precipitation record and established global or climactic intensity-duration thresholds. Landslide-triggering storms are marked. It appears that these models generally perform better when using MRMS or NLDAS-2 data, since the IMERG products detect a larger number of low intensity values for landslide-triggering storms.

Table 2: Threat score, hit ratio, and false alarm ratio for each product and the F. Guzzetti et al. (2007) intensity-duration threshold

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Product | Include | **Hits** | **Misses** | **Hit ratio** | **False alarm ratio** |
| GPM IMERG Early | All | 114 | 62 | 0.648 | 0.269 |
|  | Verified | 44 | 21 | 0.677 | 0.298 |
| GPM IMERG-Final | All | 117 | 60 | 0.661 | 0.307 |
|  | Verified | 45 | 19 | 0.703 | 0.339 |
| NLDAS-2 | All | 114 | 40 | 0.740 | 0.221 |
|  | Verified | 45 | 14 | 0.763 | 0.223 |
| MRMS | All | 130 | 26 | 0.833 | 0.243 |
|  | Verified | 52 | 7 | 0.881 | 0.264 |

Table 3: Threat score, hit ratio, and false alarm ratio for each product and the Susan H. Cannon et al. (2008) intensity-duration threshold for burned areas. Since most of these sites were not burned, this threshold tends to capture more false alarms

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Product | Include | **Hits** | **Misses** | **Hit ratio** | **False alarm ratio** |
| GPM IMERG Early | All | 69 | 77 | 0.473 | 0.107 |
|  | Verified | 31 | 28 | 0.525 | 0.126 |
| GPM IMERG-Final | All | 76 | 69 | 0.524 | 0.135 |
|  | Verified | 33 | 23 | 0.589 | 0.160 |
| NLDAS-2 | All | 89 | 42 | 0.679 | 0.184 |
|  | Verified | 36 | 19 | 0.655 | 0.177 |
| MRMS | All | 52 | 41 | 0.559 | 0.208 |
|  | Verified | 26 | 15 | 0.634 | 0.210 |

Table 4: Threat score, hit ratio, and false alarm ratio for each product and the Clarizia, Gullà, and Sorbino (1996) intensity-duration threshold.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Product | Include | **Hits** | **Misses** | **Hit ratio** | **False alarm ratio** |
| GPM IMERG Early | All | 80 | 66 | 0.548 | 0.161 |
|  | Verified | 35 | 24 | 0.593 | 0.187 |
| GPM IMERG-Final | All | 89 | 56 | 0.614 | 0.200 |
|  | Verified | 54 | 39 | 0.581 | 0.251 |
| NLDAS-2 | All | 97 | 34 | 0.740 | 0.236 |
|  | Verified | 45 | 14 | 0.763 | 0.223 |
| MRMS | All | 54 | 39 | 0.581 | 0.251 |
|  | Verified | 42 | 13 | 0.764 | 0.232 |

# 4 Discussion

Among the precipitation products chosen for this study, both IMERG products identify both higher peak intensities and larger return periods relative to the other products. Interestingly, they also detect more anomalously low precipitation values. Low-intensity precipitation in all products was associated with long duration storm events (see fig. 5), which may occur because of noisy low-precipitation slightly above the 1 mm threshold extending the computed duration of the storm and reducing its overall intensity. As a result, it appears that while every product could benefit from an enhanced storm delineation process that prevents the intensity from being diluted, the IMERG products were particularly vulnerable to the identification of long-duration low-intensity storms as a result of the method used in this study to separate storms. Those long-duration low-intensity storms tended to bring the hit ratio down for the intensity-duration thresholds. It is possible that many of the long-duration low-intensity precipitation events could be effectively filtered out by using a different storm delineation algorithm. Since the IMERG products were both 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 storm problem were mitigated.

The precipitation products performed reasonably well at identifying landslides using the published intensity-duration thresholds particularly considering that these thresholds were developed on training data spanning large regions and different sources of precipitation data than those used in this study. However, they fared more poorly at excluding false alarms, most likely because there are factors beyond intensity and duration that can influence landslide occurrence such as topography, soil type, recent wildfire or disturbance or land development. Some of the high-intensity precipitation that did not trigger any recorded landslides may be more reflective 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 endgin up 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. The poorest performing products were the IMERG products despite their detecting more high-intensity events but because they also detected many low intensity events, causing the intensity-duration threshold to miss landslides.

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. 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 other sites with approximate locations. This suggests that some of the approximate landslide locations were too far away from the true landslide location for the precipitation measurements to be representative. The intensity-duration thresholds similarly performed better at verified locations across all precipitation products. 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 for future work along the same lines.

# 5 Conclusion

The precipitation products chosen for this study represent diverse measurement techniques that often reported large differences in precipitation leading up to the landslide events evaluated here. As a result, the precipitation products differed in their overall performance in predicting landslides on the basis of published intensity-duration thresholds. A particular challenge was the presence of low-intensity, long-duration storms preceding landslide events. This challenge could potentially be addressed by better filtering and aggregating the data into more pronounced storm events. Overall, the choice of intensity-duration threshold was not as consequential as the choice of precipitation product in identifying landslides. Performance from products that rely on ground-based sensors showed a more easily identifiable landslide signal despite generally recording lower peak intensities and return periods. Though it was hypothesized that peak intensity would be an important predictive factor, the results suggest instead that a lack of noise on the low end may be more important for accurate landslide identification.

A major limitation to studies of landslide-triggering precipitation is the lack of verifiable exact landslide locations and times. This limitation was reflected in the results for verified landslide locations as compared to approximate locations, which strongly imply that some of the approximate locations were incorrect to such an extent that the precipitation measurements were inaccurate. This problem can be addressed by more extensive manual searches such as the one used in this study that identified the 80 verified landslide locations, or perhaps in the future by machine learning methods.

Using the methods of this study, those practitioners attempting to use intensity-duration thresholds as operation landslide models would do well to select a product like MRMS that has extremely low latency and performs well at identifying landslides. None of the products was particularly good at filtering out false alarms of landslides. A model that takes into account landslide susceptibility has the potential to reduce false alarms. Therefore, an additional recommendation would be for practitioners to consider more than one precipitation product, i.e. multiple precipitation estimates simultaneously, as a way to quantify stronger precipitation signals and to minimize the influence of noise.

# Bibliography

Adler, Robert F., George J. Huffman, Alfred Chang, Ralph Ferraro, Ping-Ping Xie, John Janowiak, Bruno Rudolf, et al. 2003. “The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979Present).” *Journal of Hydrometeorology* 4 (6): 1147–67. <https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2>.

Adler, Robert F, Christopher Kidd, Grant Petty, Mark Morissey, and H Michael Goodman. 2001. “Intercomparison of Global Precipitation Products: The Third Precipitation Intercomparison Project (PIP-3).” *Bulletin of the American Meteorological Society*, 20.

AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai. 2011. “Evaluation of Satellite-Retrieved Extreme Precipitation Rates Across the Central United States.” *Journal of Geophysical Research: Atmospheres* 116 (D2). <https://doi.org/10.1029/2010JD014741>.

Ahmadalipour, Ali, and Hamid Moradkhani. 2017. “Analyzing the Uncertainty of Ensemble-Based Gridded Observations in Land Surface Simulations and Drought Assessment.” *Journal of Hydrology* 555 (December): 557–68. <https://doi.org/10.1016/j.jhydrol.2017.10.059>.

Amitai, E., W. Petersen, X. Llort, and S. Vasiloff. 2012. “Multiplatform Comparisons of Rain Intensity for Extreme Precipitation Events.” *IEEE Transactions on Geoscience and Remote Sensing* 50 (3): 675–86. <https://doi.org/10.1109/TGRS.2011.2162737>.

Ashouri, Hamed, Kuo-Lin Hsu, Soroosh Sorooshian, Dan K. Braithwaite, Kenneth R. Knapp, L. Dewayne Cecil, Brian R. Nelson, and Olivier P. Prat. 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, Jiawei, Steven C. Sherwood, Lisa V. Alexander, and Jason P. Evans. 2017. “Future Increases in Extreme Precipitation Exceed Observed Scaling Rates.” *Nature Climate Change* 7 (2): 128–32. <https://doi.org/10.1038/nclimate3201>.

Beck, Hylke E., Albert IJM van Dijk, Vincenzo Levizzani, Jaap Schellekens, Diego Miralles, Brecht Martens, and Ad de Roo. 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>.

Bousquet, Olivier, and Bradley F. Smull. 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>.

Caine, Nel. 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, Susan H., Joseph E. Gartner, Raymond C. Wilson, James C. Bowers, and Jayme L. Laber. 2008. “Storm Rainfall Conditions for Floods and Debris Flows from Recently Burned Areas in Southwestern Colorado and Southern California.” *Geomorphology*, Debris flows initiated by runoff, erosion, and sediment entrainment in western North America, 96 (3): 250–69. <https://doi.org/10.1016/j.geomorph.2007.03.019>.

Cannon, Susan H, and Joseph E Gartner. 2005. “Wildfire-Related Debris Flow from a Hazards Perspective.” In *Debris-Flow Hazards and Related Phenomena*, 23.

Chandrasekar, V., Arthur Hou, Eric Smith, V. N. Bringi, S. A. Rutledge, E. Gorgucci, W. A. Petersen, and Gail Skofronick Jackson. 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): 1127–46. <https://doi.org/10.1175/2008BAMS2177.1>.

Chowdhury, Robin, and Phil Flentje. 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, G Gullà, and G Sorbino. 1996. “Sui Meccanismi Di Innesco Dei Soil Slip.” In *International Conference Prevention of Hydrogeological Hazards: The Role of Scientific Research*, 1:585–97.

Corominas, J, J Moya, and M Hürlimann. 2002. “Landslide Rainfall Triggers in the Spanish Eastern Pyrenees.” In *Mediterranean Storms*, 4. Mallorca, Spain: Editrice.

Duchon, C. E., and C. J. Biddle. 2010. “Undercatch of Tipping-Bucket Gauges in High Rain Rate Events.” *Advances in Geosciences* 25 (March): 11–15. <https://doi.org/10.5194/adgeo-25-11-2010>.

Duchon, Claude, Christopher Fiebrich, and David Grimsley. 2014. “Using High-Speed Photography to Study Undercatch in Tipping-Bucket Rain Gauges.” *Journal of Atmospheric and Oceanic Technology* 31 (6): 1330–36. <https://doi.org/10.1175/JTECH-D-13-00169.1>.

Ebert, Elizabeth E. 2007. “Methods for Verifying Satellite Precipitation Estimates.” In *Measuring Precipitation From Space*, edited by Vincenzo Levizzani, Peter Bauer, and F. Joseph Turk, 345–56. Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-1-4020-5835-6_27>.

England Jr., John F., Timothy A. Cohn, Beth A. Faber, Jery R. Stedinger, Wilbert O. Thomas Jr., Andrea G. Veilleux, Julie E. Kiang, and Robert R. Mason Jr. 2019. “Guidelines for Determining Flood Flow FrequencyBulletin 17c.” USGS Numbered Series 4-B5. *Guidelines for Determining Flood Flow FrequencyBulletin 17c*. Techniques and Methods. Reston, VA: U.S. Geological Survey. <https://doi.org/10.3133/tm4B5>.

Fornasiero, A, R Amorati, P P Alberoni, L Ferraris, and A C Taramasso. 2004. “Impact of Combined Beam Blocking and Anomalous Propagation Correction Algorithms on Radar Data Quality.” *Proceedings of ERAD*, 216–22.

Froude, Melanie J., and David N. Petley. 2018. “Global Fatal Landslide Occurrence from 2004 to 2016.” *Natural Hazards and Earth System Sciences* 18 (8): 2161–81. <https://doi.org/10.5194/nhess-18-2161-2018>.

Galanti, Yuri, Michele Barsanti, Andrea Cevasco, Giacomo D’Amato Avanzi, and Roberto Giannecchini. 2018. “Comparison of Statistical Methods and Multi-Time Validation for the Determination of the Shallow Landslide Rainfall Thresholds.” *Landslides* 15 (5): 937–52. <https://doi.org/10.1007/s10346-017-0919-3>.

Gutmann, Ethan, Tom Pruitt, Martyn P. Clark, Levi Brekke, Jeffrey R. Arnold, David A. Raff, and Roy M. Rasmussen. 2014. “An Intercomparison of Statistical Downscaling Methods Used for Water Resource Assessments in the United States.” *Water Resources Research* 50 (9): 7167–86. <https://doi.org/10.1002/2014WR015559>.

Guzzetti, F., S. Peruccacci, M. Rossi, and C. P. Stark. 2007. “Rainfall Thresholds for the Initiation of Landslides in Central and Southern Europe.” *Meteorology and Atmospheric Physics* 98 (3-4): 239–67. <https://doi.org/10.1007/s00703-007-0262-7>.

Guzzetti, Fausto, Silvia Peruccacci, Mauro Rossi, and Colin P. Stark. 2008. “The Rainfall Intensityduration Control of Shallow Landslides and Debris Flows: An Update.” *Landslides* 5 (1): 3–17. <https://doi.org/10.1007/s10346-007-0112-1>.

Henn, Brian, Andrew J. Newman, Ben Livneh, Christopher Daly, and Jessica D. Lundquist. 2018. “An Assessment of Differences in Gridded Precipitation Datasets in Complex Terrain.” *Journal of Hydrology* 556 (January): 1205–19. <https://doi.org/10.1016/j.jhydrol.2017.03.008>.

Highland, Lynn, and Peter Bobrowsky. 2008. “The Landslide Handbook: A Guide to Understanding Landslides.” Circular 1325. Circular. Reston, VA: United States Geological Survey.

Hou, Arthur Y., Ramesh K. Kakar, Steven Neeck, Ardeshir A. Azarbarzin, Christian D. Kummerow, Masahiro Kojima, Riko Oki, Kenji Nakamura, and Toshio Iguchi. 2014. “The Global Precipitation Measurement Mission.” *Bulletin of the American Meteorological Society* 95 (5): 701–22. <https://doi.org/10.1175/BAMS-D-13-00164.1>.

Huffman, George J., David T. Bolvin, Dan Braithwaite, Kuo-Lin Hsu, Robert J. Joyce, Christopher Kidd, Eric J. Nelkin, et al. 2020. “Integrated Multi-Satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG).” In *Satellite Precipitation Measurement: Volume 1*, edited by Vincenzo Levizzani, Christopher Kidd, Dalia B. Kirschbaum, Christian D. Kummerow, Kenji Nakamura, and F. Joseph Turk, 343–53. Advances in Global Change Research. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-24568-9_19>.

Huffman, George J., David T. Bolvin, Eric J. Nelkin, David B. Wolff, Robert F. Adler, Guojun Gu, Yang Hong, Kenneth P. Bowman, and Erich F. Stocker. 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, Emily, Donald J. Wuebbles, Kenneth E. Kunkel, Seth C. Olsen, and Alex Goodman. 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, Chris, Andreas Becker, George J. Huffman, Catherine L. Muller, Paul Joe, Gail Skofronick-Jackson, and Dalia B. Kirschbaum. 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>.

Kirschbaum, Dalia Bach, Robert Adler, Yang Hong, Stephanie Hill, and Arthur Lerner-Lam. 2010. “A Global Landslide Catalog for Hazard Applications: Method, Results, and Limitations.” *Natural Hazards* 52 (3): 561–75. <https://doi.org/10.1007/s11069-009-9401-4>.

Kirschbaum, Dalia Bach, Robert Adler, Yang Hong, Sujay Kumar, Christa Peters-Lidard, and Arthur Lerner-Lam. 2012. “Advances in Landslide Nowcasting: Evaluation of a Global and Regional Modeling Approach.” *Environmental Earth Sciences* 66 (6): 1683–96. <https://doi.org/10.1007/s12665-011-0990-3>.

Kirschbaum, Dalia, and Thomas Stanley. 2018. “Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness.” *Earth’s Future* 6 (3): 505–23. <https://doi.org/10.1002/2017EF000715>.

Kummerow, Christian, William Barnes, Toshiaki Kozu, James Shiue, and Joanne Simpson. 1998. “The Tropical Rainfall Measuring Mission (TRMM) Sensor Package.” *Journal of Atmospheric and Oceanic Technology* 15 (3): 809–17. <https://doi.org/10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2>.

Leonarduzzi, Elena, Peter Molnar, and Brian W. McArdell. 2017. “Predictive Performance of Rainfall Thresholds for Shallow Landslides in Switzerland from Gridded Daily Data.” *Water Resources Research* 53 (8): 6612–25. <https://doi.org/10.1002/2017WR021044>.

Lockhoff, M., O. Zolina, C. Simmer, and J. Schulz. 2014. “Evaluation of Satellite-Retrieved Extreme Precipitation over Europe Using Gauge Observations.” *Journal of Climate* 27 (2): 607–23. <https://doi.org/10.1175/JCLI-D-13-00194.1>.

Lundquist, Jessica D., Mimi Hughes, Brian Henn, Ethan D. Gutmann, Ben Livneh, Jeff Dozier, and Paul Neiman. 2015. “High-Elevation Precipitation Patterns: Using Snow Measurements to Assess Daily Gridded Datasets Across the Sierra Nevada, California.” *Journal of Hydrometeorology* 16 (4): 1773–92. <https://doi.org/10.1175/JHM-D-15-0019.1>.

Manzanas, R., L. K. Amekudzi, K. Preko, S. Herrera, and J. M. Gutiérrez. 2014. “Precipitation Variability and Trends in Ghana: An Intercomparison of Observational and Reanalysis Products.” *Climatic Change* 124 (4): 805–19. <https://doi.org/10.1007/s10584-014-1100-9>.

Mirus, Benjamin B., Eric S. Jones, Rex L. Baum, Jonathan W. Godt, Stephen Slaughter, Matthew M. Crawford, Jeremy Lancaster, et al. 2020. “Landslides Across the USA: Occurrence, Susceptibility, and Data Limitations.” *Landslides* 17 (10): 2271–85. <https://doi.org/10.1007/s10346-020-01424-4>.

Nikahd, Ali, Mazlan Hashim, and Mohammad Jafar Nazemosadat. 2016. “A Review of Uncertainty Sources on Weather Ground-Based Radar for Rainfall Estimation.” *Applied Mechanics and Materials*. https://www.scientific.net/AMM.818.254; Trans Tech Publications Ltd. <https://doi.org/10.4028/www.scientific.net/AMM.818.254>.

“North America Elevation 1-Kilometer Resolution.” 2007. Map. Commission for Environmental Cooperation.

O, Sungmin, Ulrich Foelsche, Gottfried Kirchengast, Juergen Fuchsberger, Jackson Tan, and Walter A. Petersen. 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): 6559–72. <https://doi.org/10.5194/hess-21-6559-2017>.

Pendergrass, Angeline G., and Reto Knutti. 2018. “The Uneven Nature of Daily Precipitation and Its Change.” *Geophysical Research Letters* 45 (21): 11, 980–11, 988. <https://doi.org/10.1029/2018GL080298>.

Pollock, M. D., G. O’Donnell, P. Quinn, M. Dutton, A. Black, M. E. Wilkinson, M. Colli, et al. 2018. “Quantifying and Mitigating Wind-Induced Undercatch in Rainfall Measurements.” *Water Resources Research* 54 (6): 3863–75. <https://doi.org/10.1029/2017WR022421>.

Premchitt, J., E. W. Brand, and H. B. Phillipson. 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, Mauro, Dalia Kirschbaum, Daniela Valigi, Alessandro Cesare Mondini, and Fausto Guzzetti. 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>.

Scheevel, Caroline R., Rex L. Baum, Benjamin B. Mirus, and Joel B. Smith. 2017. “Precipitation Thresholds for Landslide Occurrence Near Seattle, Mukilteo, and Everett, Washington.” Open-File Report 2017-1039. Open-File Report. U.S. Department of the Interior; U.S. Geological Survey.

Segoni, Samuele, Guglielmo Rossi, Ascanio Rosi, and Filippo Catani. 2014. “Landslides Triggered by Rainfall: A Semi-Automated Procedure to Define Consistent Intensityduration Thresholds.” *Computers & Geosciences* 63 (February): 123–31. <https://doi.org/10.1016/j.cageo.2013.10.009>.

Skofronick-Jackson, Gail, Walter A. Petersen, Wesley Berg, Chris Kidd, Erich F. Stocker, Dalia B. Kirschbaum, Ramesh Kakar, et al. 2017. “The Global Precipitation Measurement (GPM) Mission for Science and Society.” *Bulletin of the American Meteorological Society* 98 (8): 1679–95. <https://doi.org/10.1175/BAMS-D-15-00306.1>.

Sun, Qiaohong, Chiyuan Miao, Qingyun Duan, Hamed Ashouri, Soroosh Sorooshian, and Kuo-Lin Hsu. 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., Y. Hundecha, D. Lawrence, H. Madsen, P. Willems, M. Martinkova, K. Vormoor, et al. 2015. “Inter-Comparison of Statistical Downscaling Methods for Projection of Extreme Precipitation in Europe.” *Hydrology and Earth System Sciences* 19 (4): 1827–47. <https://doi.org/10.5194/hess-19-1827-2015>.

Tapiador, Francisco J., F. J. Turk, Walt Petersen, Arthur Y. Hou, Eduardo García-Ortega, Luiz A. T. Machado, Carlos F. Angelis, et al. 2012. “Global Precipitation Measurement: Methods, Datasets and Applications.” *Atmospheric Research* 104-105 (February): 70–97. <https://doi.org/10.1016/j.atmosres.2011.10.021>.

Tryhorn, Lee, and Art DeGaetano. 2011. “A Comparison of Techniques for Downscaling Extreme Precipitation over the Northeastern United States.” *International Journal of Climatology* 31 (13): 1975–89. <https://doi.org/10.1002/joc.2208>.

US Department of Commerce, NOAA. 2013. “NOAA Atlas 2 Precipitation Frequency Estimates in GIS Compatible Formats.” *NOAA Atlas 2 Precipitation Frequency Estimates in GIS Compatible Formats*. https://www.nws.noaa.gov/ohd/hdsc/noaaatlas2.htm; US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service.

Vose, Russell S., Scott Applequist, Mike Squires, Imke Durre, Matthew J. Menne, Claude N. Williams, Chris Fenimore, Karin Gleason, and Derek Arndt. 2014. “Improved Historical Temperature and Precipitation Time Series for U.S. Climate Divisions.” *Journal of Applied Meteorology and Climatology* 53 (5): 1232–51. <https://doi.org/10.1175/JAMC-D-13-0248.1>.

Wang, Guiling, Christine J Kirchhoff, Anji Seth, John T Abatzoglou, Ben Livneh, David W Pierce, Lori Fomenko, and Tengyu Ding. 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, Youlong, Kenneth Mitchell, Michael Ek, Justin Sheffield, Brian Cosgrove, Eric Wood, Lifeng Luo, et al. 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 (D3). <https://doi.org/10.1029/2011JD016048>.

Yu, Fan-Chieh, Tien-Chien Chen, Meei-Ling Lin, Chien-Yuan Chen, and Wen-Hai Yu. 2006. “Landslides and Rainfall Characteristics Analysis in Taipei City During the Typhoon Nari Event.” *Natural Hazards* 37 (1-2): 153–67. <https://doi.org/10.1007/s11069-005-4661-0>.

Zhang, Jian, Kenneth Howard, Carrie Langston, Brian Kaney, Youcun Qi, Lin Tang, Heather Grams, et al. 2015. “Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities.” *Bulletin of the American Meteorological Society* 97 (4): 621–38. <https://doi.org/10.1175/BAMS-D-14-00174.1>.