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

Elsa Culler

Andrew Badger

Toby Minear

Kristy Tiampo

Ben Livneh

Extreme precipitation can have profound consequences for communities, resulting in flooding and rainfall-triggered landslides, causing 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. First, we compare the average intensity, peak intensity at the smallest interval available, duration and NOAA Atlas return periods of the landslide-triggering storms, at 257 landslide locations across the continental US and Canada. Precipitation data are taken from five 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), gauge (North American Land Data Assimilation System v. 2 Forcing precipitation), and numerical weather prediction (High-Resolution Rapid Refresh real-time precipitation). These products also cover a range of spatial and temporal resolutions as well as spatial extent and real-time or near real-time 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 is also potentially 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. Next, we compare the intensity and duration of storms at landslide sites to existing published Intensity-Duration Thresholds to determine which products achieve the highest Equitable Threat Score for landslide predictions using these existing models. Finally, we discuss the implications of precipitation uncertainty in the context of real-time landslide predictions, to provide guidance for practitioners and researchers on strengths and weaknesses of different products and approaches.

# Introduction

In spite of the destructive nature of landslides, these events remain challenging to forecast (Kirschbaum and Stanley 2018). There are many sources of uncertainty that contribute to poor landslide predictions such as soil property, vegetation, and anthropogenic modifications to surface and subsurface soil structure. Perhaps the largest source of uncertainty in landslide probability estimates, is hydrologic uncertainty, here defined 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 range precipitation datasets ranging from in situ observations, ground-based radar and remotely sensed retrievals. The goal of this analysis is to investigate the role of precipitation uncertainty and subsequently the uncertainty in landslide risks. A greater understanding areas of relative agreement and divergence across products may provide guidance to practitioners and researchers choosing precipitation products for studying landslides.

## Precipitation sensors and estimates

The precipitation products chosen for this inter-comparison represent three broad categories of precipitation measurements as their primary component: precipitation gauges, ground-based radar, and microwave satellite. Precipitation gauges operate by periodically measuring how much precipitation has landed in a bucket. Their main strength is they directly measure the amount of water that lands in the bucket, but nonetheless they suffer from spatial and temporal inconsistencies as a result of wind (Pollock et al. 2018), instrument design (Duchon, Fiebrich, and Grimsley 2014; Duchon and Biddle 2010), poor instrument placement (Vose et al. 2014), lack representativeness of the surrounding area, and global sensor density (Kidd et al. 2017). Ground-based radar can detect precipitation based on propagation and backscatter of radar, and therefore can detect variation in precipitation potentially hundreds of kilometers away. Like any indirect measurement of precipitation, radar must convert the radar signal to precipitation volume and to remove the influence of other objects from buildings or insects in the radar’s path (Fornasiero et al., n.d.; 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 is an advantage of ground-based radar over many satellite sensors (Chandrasekar et al. 2008). Satellites can use any of a number of sensors to detect precipitation including active and passive microwave, infrared, radar, or any combination, and they can also be deployed in geostationary or low Earth orbits that cover particular regions at particular intervals. The key advantage of satellite-based precipitation measurements is that unlike ground-based sensors they can provide frequent, spatially homogenous, global precipitation measurements, although typically multiple satellites are required to develop such a product (Tapiador et al. 2012). However, since each type of satellite-based sensor has its own set of strengths and weaknesses that are beyond the scope of this study, many of the challenges of satellite-based precipitation measurement are related to sensor calibration, bias-correction relative to ground-based measurements (Ebert 2007), and the development of algorithms for merging measurements from diverse sources (Huffman et al. 2007).

## Precipitation datasets for forecasting or nowcasting landslides

Though precipitation measurements have been compared on the basis of any number of metrics in prior studies ranging from annual totals to the largest number of consecutive dry days, in this study the choice of metrics was guided by what might be useful to researchers and practitioners interested in forecasting and responding to rainfall-triggered landslides.

While some landslides are triggered by short, intense precipitation events, others are triggered by the complete saturation of the soil column that occurs over a longer period of time (Cannon and Gartner 2005). However, in both of these 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, we selected precipitation products with hourly or finer temporal resolution (see sec. 2.2) and evaluated them over individual storm events. Satellite products tend to capture the higher-intensity precipitation (Sun et al. 2018) that can be key in triggering landslides. This may be due to the measurement method or the generally higher temporal resolution of satellite products.

When precipitation is used to provide warning systems or guide recovery efforts from landslides, it is important to be able to provide that information in a timely manner (Kirschbaum et al. 2012). Low latency is therefore vital in a precipitation product used to forecast or nowcast landslides. This study also will assess whether the low latency comes at a cost relative to landslide forecasting skill for the selected products.

Finally, many precipitation products struggle in mountainous regions (Sun et al. 2018), precisely where landslides are most likely to occur due to higher slopes. This study includes an analysis in the spatial variation in performance to assess the role of topography in the results.

## Precipitation product comparisons

Sun et al. (2018) reviewed 30 gauge-based, satellite-based, and reanalysis global precipitation products, comparing annual precipitation estimates, 90th percentile of daily precipitation, systematic and random error for daily precipitation, and regional differences in performance. They found a great deal of variability even within the same class of product (e.g. a deviation of 300 mm in annual precipitation for some). They conclude that cross validating across multiple datasets is crucial to account for errors, and that the placement and density of gauges accounts for many of the errors in gauge-based or gauge-corrected products.

Adler et al. (2001) similarly analyzed 31 gauge-based, satellite-based, model-based, and climatological datasets, comparing monthly precipitation, precipitation by latitude, and inter-annual change. They found 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 found that products incorporating both in situ and satellite information (e.g. the Global Precipitation Climatology Project [GPCP] (Adler et al. 2003)) perform better than products based on a single data source.

## Inter-comparison of extreme precipitation

Fewer studies comparing extreme precipitation exist (Amitai et al. 2012; Manzanas et al. 2014; Hashmi, Shamseldin, and Melville 2011; Tryhorn and DeGaetano 2011), primarily looking at extreme precipitation indicators like 90th percentile precipitation, extreme one-day precipitation and maximum number of consecutive wet days. 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). Because this study is focusing on rainfall-triggered landslides, it will focus instead on the total storm depth, duration, average intensity, and peak intensity of precipitation events known to precede landslides in North America.

## Intensity-Duration Thresholds for landslide prediction

Intensity-Duration Thresholds are a type of single-parameter statistical model used for landslide early warning systems, where rainstorms above the curve are predicted to cause landslides (Scheevel et al. 2017). The curves are typically valid in a particular region or climate and for a range of durations based on the training data (Guzzetti et al. 2008). This study will use a selection of power-law Intensity-Duration Thresholds from those included in a review by Guzzetti et al. (2008) to preliminarily compare the suitability of precipitation measurements from different sources for providing early warning or near-real time support to disaster response organizations.

Given the wide-ranging issues associated with precipitation observations cited above, as well as the importance of understanding and anticipating landslide events, this study presents a multi-product, multi-site analysis to understand landslide-triggering storms. We thereby address a gap in the literature when it comes to evaluating extreme precipitation through the lens of natural hazards. This work furthers the analysis by Rossi et al. (2017) who compared gauge and satellite precipitation for the purposes of landslide modeling by additionally including a ground-based radar product and by singling out observations preceding specific landslide events.

In sec. 2, we will discuss landslide site and precipitation product selection, followed by procedures for splitting precipitation into storms and the metrics used in the comparison. Sec. 3 begins with the cumulative precipitation over the 30-days preceding the landslide for 5 characteristic example sites. Next, we compare each product using storm characteristics of total depth, duration, total intensity, peak intensity, and return period. To test whether peak intensity might be accounting for low return period storms causing landslides, we compare the two. Finally, we use established intensity-duration thresholds to test which products have the best separation between landslides and other rainfall, comparing the hit ratio and the false alarm ratio for each product and threshold.

# Methods

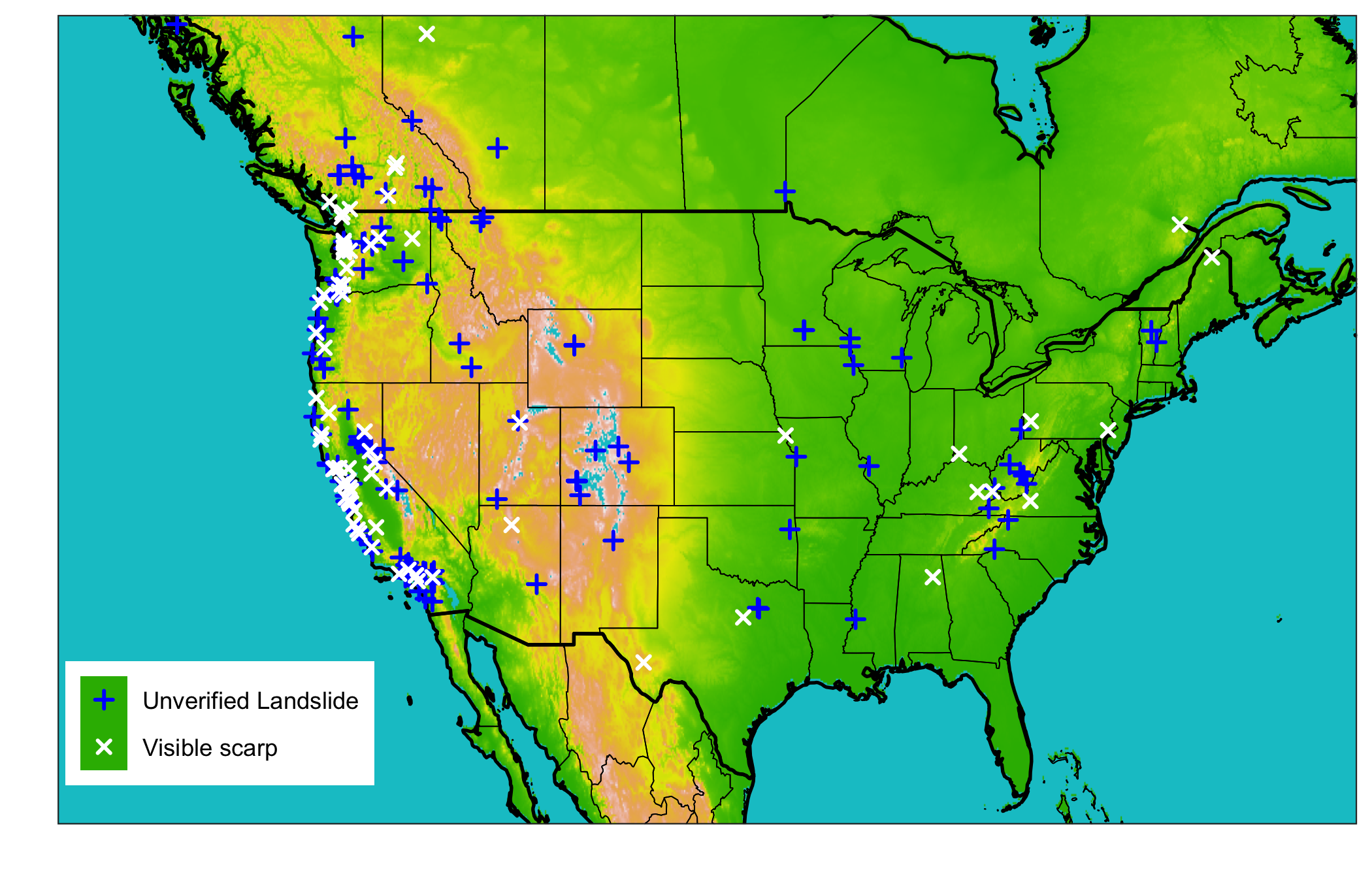
The overall goals of the methods are to evaluate precipitation characteristics at known landslide sites by first examining the features of triggering storms and then subsequently comparing precipitation estimates in the context of published intensity-duration models of landslides occurence. Rainfall-triggered landslide sites were chosen from the NASA Global Landslide Catalog 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 then the precipitation time series were split into individual storms events, and key characteristics of total depth, duration, intensity, peak intensity, and return period were 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 (GLC) (Kirschbaum et al. 2010). [Need to justify WHY you used the GLC and explain WHAT it is—here’s an example from which you can edit: The GLC was chosen for this study, since it provides a large sample of landslide locations useful for evaluating heavy rainfall events, etc etc. It’s key limitations are reporting from NEWS?, with only approximate location accuracies, ranging from A to B km etc etc. It shares strengths and weaknesses with other regional databases (include references here). Despite these limitations, it was deemed fit for purpose for this study since the primary focus here is to compare precipitation products in the vicinity of hydrologically-triggered landslides, where heavy rainfall events are likely to be present]. A compromise was reached to maximize the number of landslide locations for this study, while ensuring quality control and data availability. The following selection criteria were used to qualify landslide sites:

* Landslide events were reported as hydrologically-driven, either triggered by rain, downpour, continuous rain, or flooding according to the GLC;
* Landslide events took place in the CONUS or Canada below N after May 2015 so as to ensuring data availability across the selected precipitation products; and
* The landslide location accuracy was estimated to be km or less according to the 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 landslides, 8 locations were also verified by manually locating a visible scarp in satellite images; the location specified by the GLC was used for the remaining landslides. 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 GLC. The verified landslides seem to be spatially distriubuted fairly evenly relative to the distribution of the full selection of landslides.



**Figure 1 | Study locations**: Map of all landslide sites considered in this analysis (n=228), colored by whether the location was approximate (n=ABC) or verified using aerial satellite imagery to identify a visible scarp (n=DEF); Source of landslide locations was the GLC (Kirschbaum ref), source of the DEM data used for the basemap (ref).

## 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 continental US. 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 Table 1 and summarized in the following paragraphs.

### 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 Climate Prediction Center (NCEP-CPC) and hourly radar-based National Weather Service WSR-88D precipitation (Cosgrove et al. 2003). The gauge-based estimates are disaggregated to hourly using the WSR-88D radar-based estimates, resulting in a real-time hourly gridded product at a (~ km) resolution across North America going back to 1999 with a latency of approximately 4 days. Though it has low horizontal resolution relative to the other precipitation products used here, NLDAS-2 forcing is a widely used gauge-based product that has been extensively validated over more the past 20 years.

### 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.

### 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 “IMERG-Early” (~4 hour latency) and the “IMERG-Final” (~3.5 month latency) in this study. Since IMERG products use the GPM active and passive microwave data as a standard with only monthly gauge-based bias-correction, 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: Four precipitation products used to characterize the degree of hydrologic uncertainty present immediately before and during landslide events

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |
| IMERG final run, “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) forcing (Xia et al. 2012) | Disaggregation of Climate Prediction Center daily precipitation using bias-corrected radar | (~ 12 km) | 1 hour | 4 days |
| NOAA High-Resolution Rapid Refresh (HRRR) model (Alexander et al. 2016) | Numerical Weather Prediction with radar assimilation. | km | 1 hour | 1-36 hour forecasts updated hourly |

## 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 depth, duration, intensity, and peak intensity were computed and compared. Depth, duration, and frequency were chosen since they reflect the most common metrics used in extreme precipitation analysis (e.g. England et al., 2019), whereas peak intensity was important because… In addition, the distribution of storm depth, rank, and z-score were computed for the day of the landslide, the full preciptitation record. Storm depth was computed to evaluate whether any particular product had and unusual distribution of precipitation. 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.

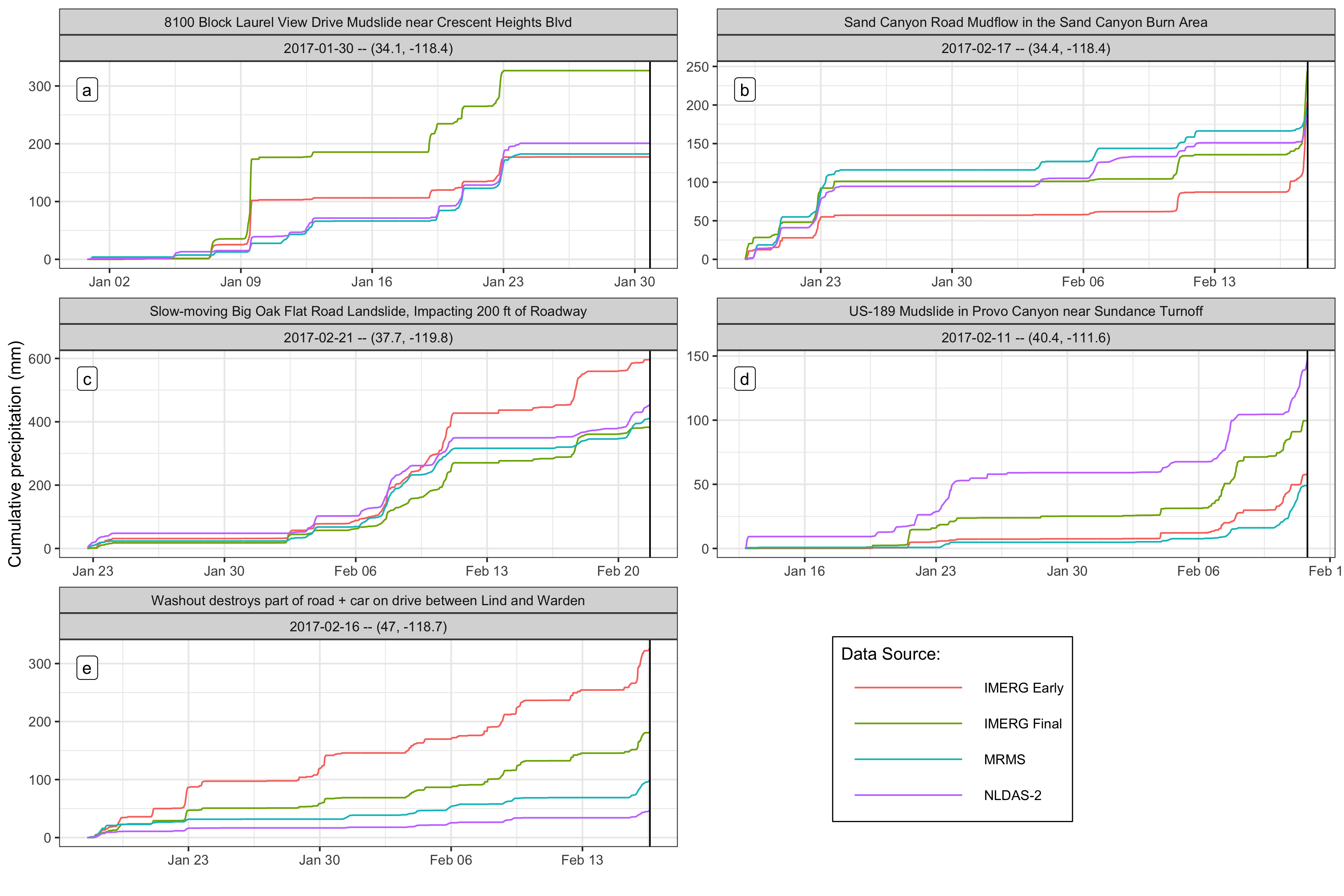
To facilitate comparison of storm characteristics within a single over-arching framework, the return period of the landslide-triggering storms was also computed using the NOAA precipitation atlas frequency estimations (US Department of Commerce 2013). In order to define a consistent return period for each storm, the maximum precipitation value for each applicable NOAA atlas duration (1, 2, 3, 6, 12, 24, 48, 72, 96, and 168 hours) was extracted. For example, for the 3-hour duration, the cumulative 3-hour precipitation total was calculated for the entire storm, and the maximum of the total chosen to look up in the NOAA atlas. We then selected the maximum return period for each of the 10 durations noted above for each landslide.

## 2.4 Comparison of storm events relative to Intensity-Duration Thresholds

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 … a is … etc etc. 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 (Scheevel et al. 2017; Caine 1980; Kirschbaum et al. 2012). Three thresholds for this study (Caine 1980; Cannon and Gartner 2005; Guzzetti et al. 2007) were obtained from a review by Guzzetti et al. (2008). Thresholds were used on applicable subsets of the data based on climate or other conditions. For example, the coastal threshold etc was used on the coastal data. 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)

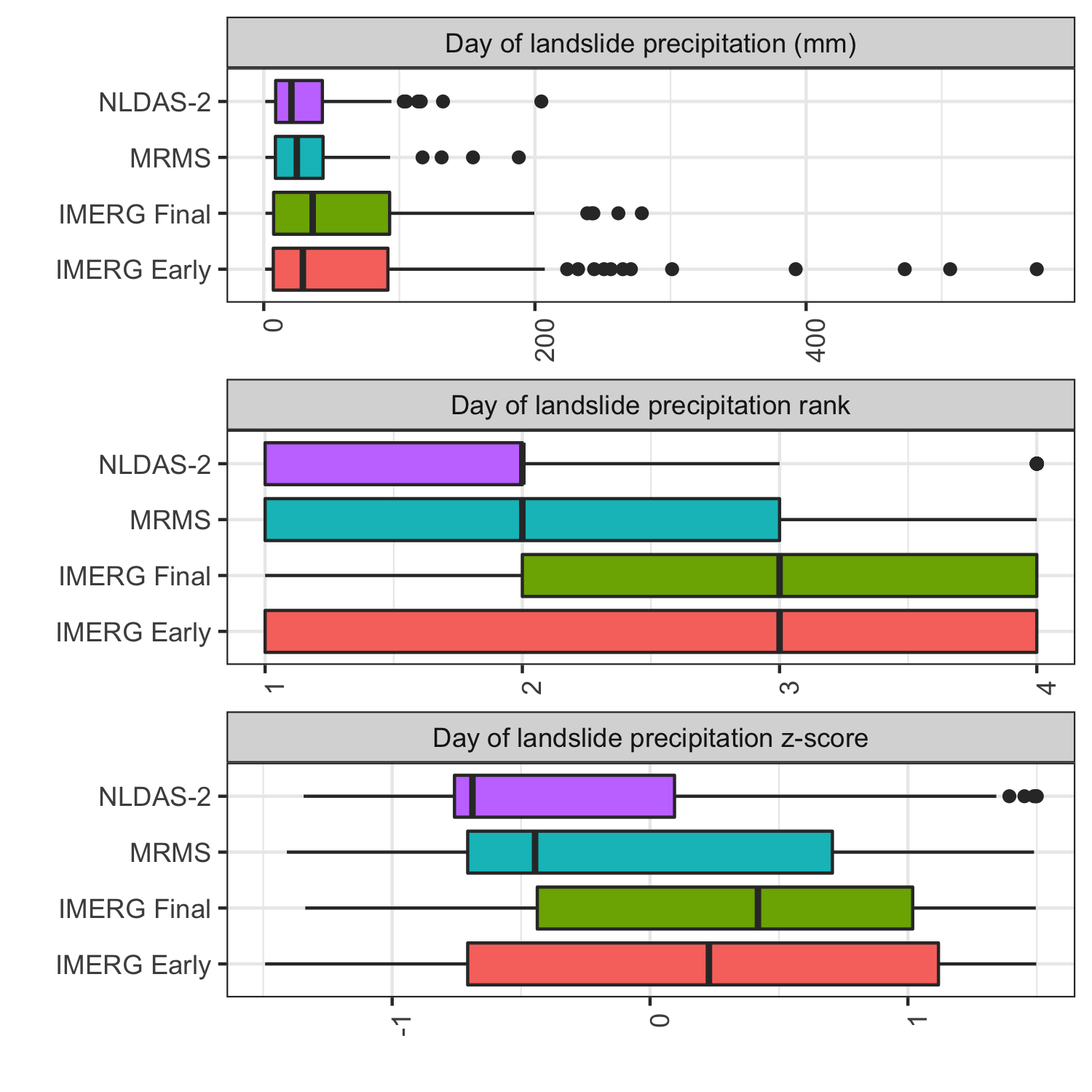
# Results

An exposition into the variety of ways in which the precipitation from multiple products can differ is presented in Fig. 2, showing the cumulative precipitation in the 30-days before a landslide for 5 example sites that showcase different categories of differences. For example, in panel (b) the products clearly diverge, but ultimately the 30-day depth is nearly identical across products. The precipitation in panel (c) shows the case of a close match for all products notwithstanding a single outlier, whereas in panel (d) there is a wide spread across products of approximately two-thirds the maximum total amount of precipitation. The precipitation in panel (e) also demonstrates a factor of 6 spread of precipitation values but storm events still appear to be highly correlated. Panel (a) 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 no landslide-triggering precipitation was detected by any product, suggesting a location error in the landslide record. In panel (b), 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. Panel (c) shows similar measurements among all products while in panel (d) there is a wide spread of approximately two-thirds the maximum total amount of precipitation. Finally, in panel (e) all products are well correlated, but the accumulated depths greatly differ.

The relative magnitude of the different precipitation products on the day of the landslide is shown in Fig. 3 in terms of the depth, rank, and z-score for day-of-landslide precipitation. The IMERG products tend to have higher rank than MRMS, which typically exceeds NLDAS-2 measurements. IMERG Early has the highest precipitation measurements by nearly 300mm, suggesting that the further interpolation in the IMERG-Final product reduces these outliers relative to IMERG-Early, although the median value for IMERG-Final is the largest overall. The z-scores reflect the same order as the rank, further underscoring the relative range of variability across all products and landslide sites. The IMERG products tend to be above the mean while NLDAS-2 and MRMS tend to be below, but the highest and lowest z-score values are similar for all products.



**Figure 3 | Relative magnitude of precipitation products on the day of the landslide**: Depth, rank, and z-score of precipitation as measured by each product for the day of the landslide for 228 events.

Fig. 4 shows the characteristics for the last storm leading up to the landslide event plotted against the ensemble mean of all the products. 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 GLC location accuracy for sites with only ‘approximate locations’.

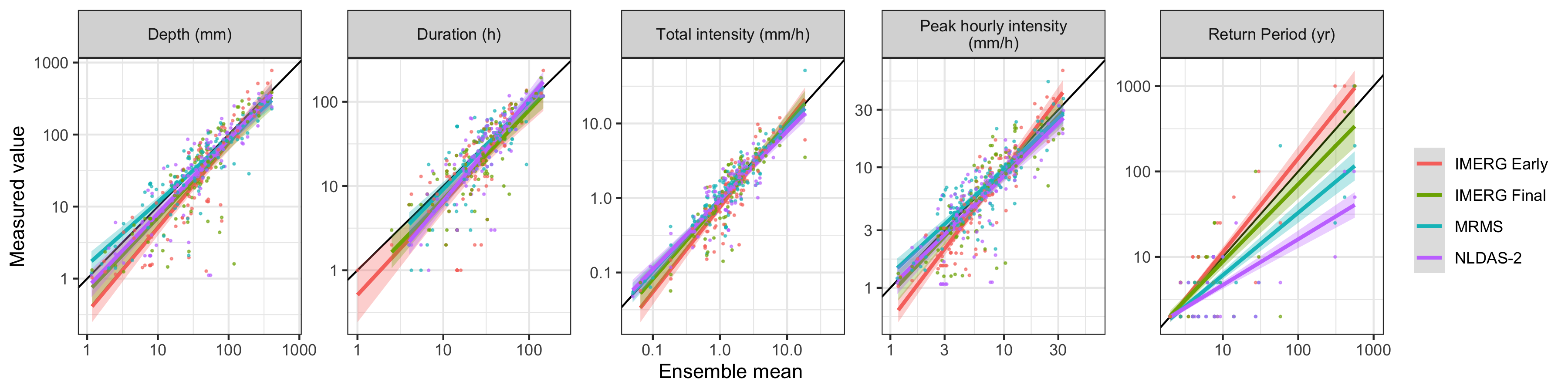
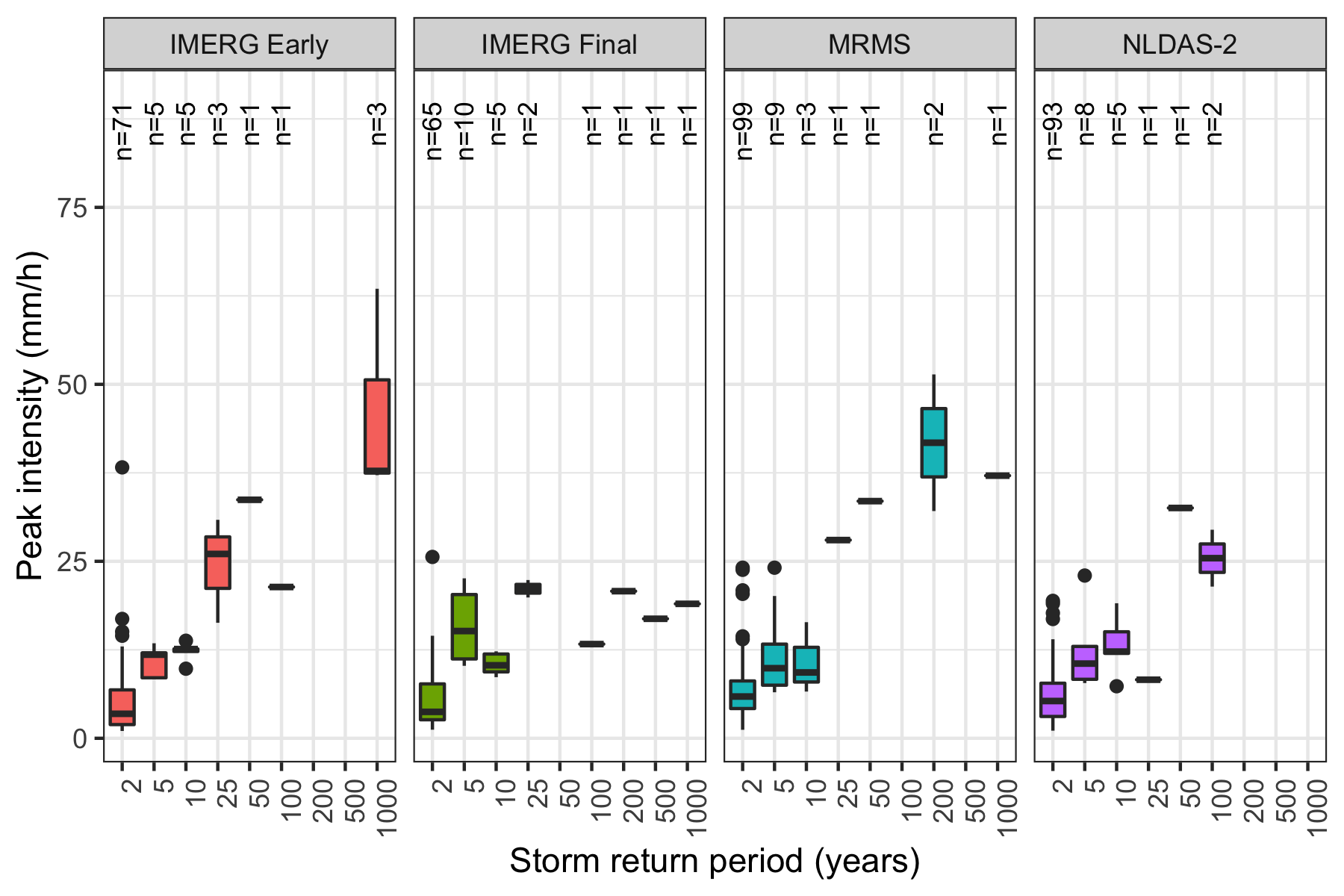


Figure 4: Storm characteristics vs. the ensemble mean as measured by each product along with trend lines.



Surprisingly, many of the landslide-triggering storms show very low return periods of less than 2 years according the NOAA atlas. One possible explanation is that large peak intensity values for these storms are causing landslides even when part of a relatively low return period longer storm event. Fig. 5 further investigates this hypothesis, showing a clear trend of increasing peak intensity with return period. Finer temporal resolutions would be necessary to further test this hypothesis, since for MRMS and NLDAS-2 the shortest storm duration is also the same as the product resolution. Peak hourly intensities high enough to cause a landslide amidst otherwise ordinary precipitation should show up as a high return period ordinary storm.



**Figure 5 | The relationship between peak storm intensity and storm return period**: [Add a description of the figure here] There appears to be direct proportionality between return period and peak intensity, but this relationship drops off for most products among the higher return periods.

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 models 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 for all products that is likely an artifact of the storm identification algorithm. It appears that many of the threshold misses (landslides not identified by the threshold) fall into this section of the data, suggesting that improvements to storm delineation might boost threshold performance for all products but particularly for the IMERG data.

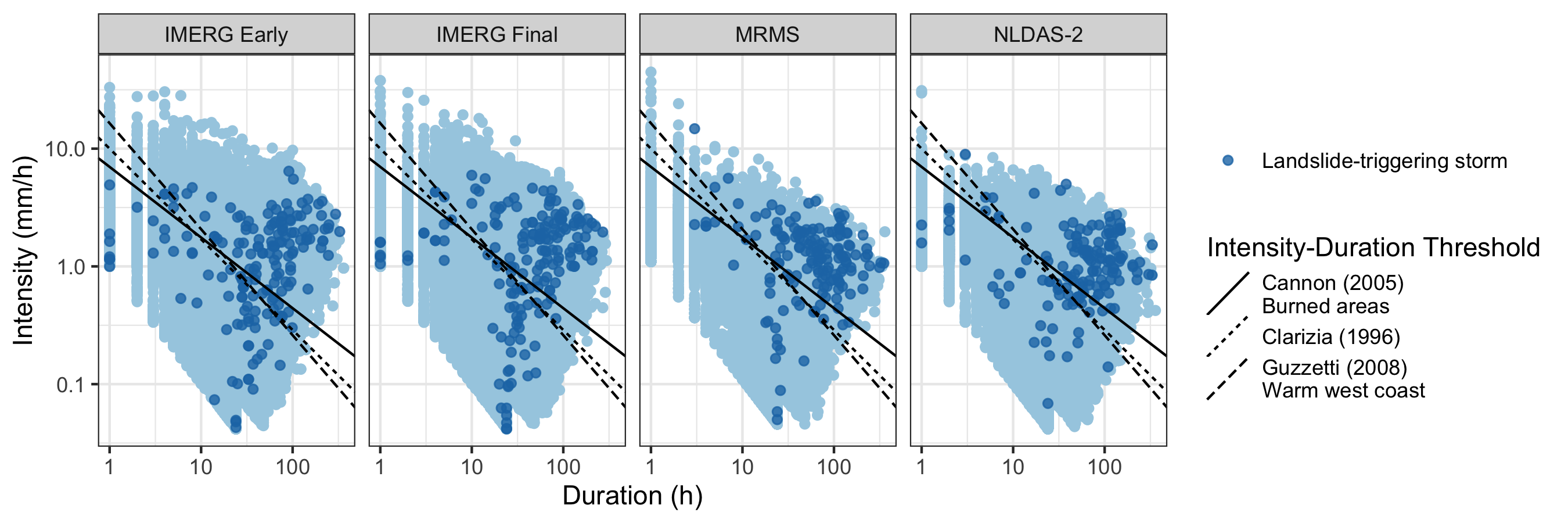


Figure 6: Each storm in the precipitation record and established global or climactic Intensity-Duration Thresholds. Landslide-triggering storms are marked. It appears that these models generally perform better when using MRMS or NLDAS-2 data, since the IMERG products detect a larger number of low intensity values for landslide-triggering storms.

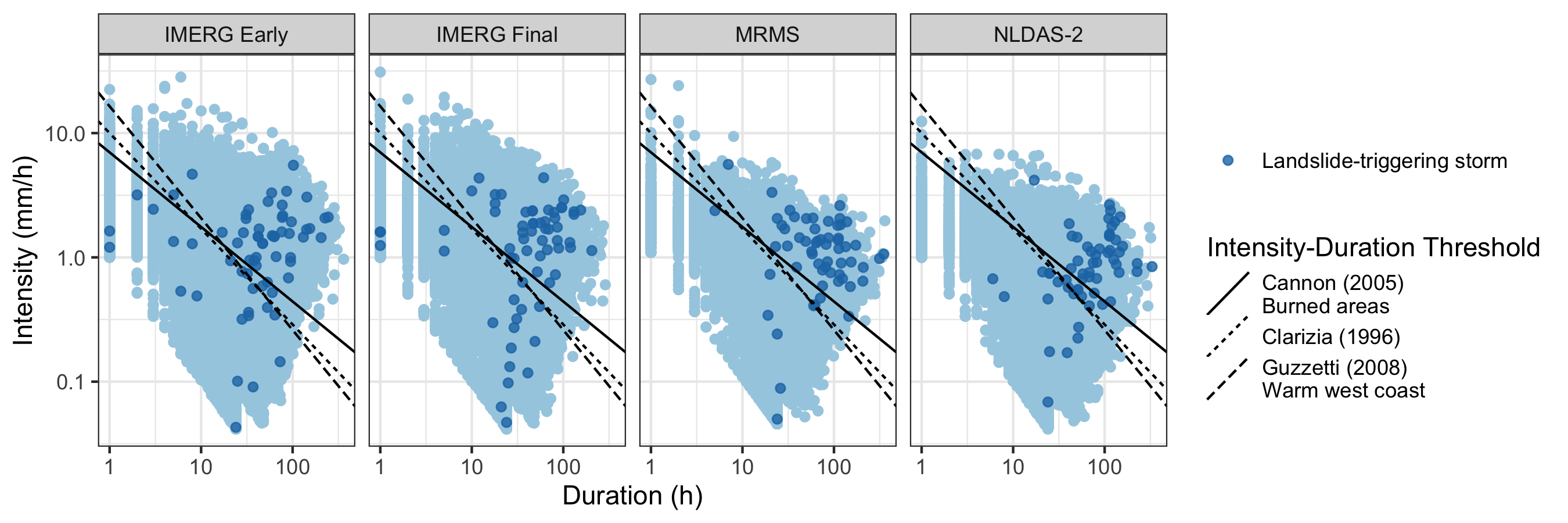


Table 2: Threat score, hit ratio, and false alarm ratio for each product and the Guzzetti et al. (2007) Intensity-Duration Threshold

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Product | Include | **Hits** | **Misses** | **Hit ratio** | **False alarm ratio** |
| GPM IMERG Early | All | 114 | 62 | 0.6477273 | 0.2694975 |
|  | Verified | 44 | 21 | 0.6769231 | 0.2980977 |
| GPM IMERG Final | All | 117 | 60 | 0.6610169 | 0.3074026 |
|  | Verified | 45 | 19 | 0.7031250 | 0.3389533 |
| NLDAS-2 | All | 114 | 40 | 0.7402597 | 0.2213864 |
|  | Verified | 45 | 14 | 0.7627119 | 0.2228354 |
| MRMS | All | 130 | 26 | 0.8333333 | 0.2433511 |
|  | Verified | 52 | 7 | 0.8813559 | 0.2635528 |

## Resolution

### Do products produce comparable results when compared at equal temporal and spatial resolution, or are there other underlying differences?

* FIGURE 8: Scatter volume, intensity, frequency, and peak intensity for each product with matched temporal resolution

# Discussion

Among the precipitation products chosen for this study, both IMERG products identify both higher peak intensities and longer? 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. 6), 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 many other factors beyond intensity and duration that can influence landslide occurrence such as topography, soil type, recent wildfire or disturbance or land development, or due to location errors of the landslide events such that some of the high-intensity precipitation may not be reflective of the landslide events but of adjacent areas. A landslide’s location in a highly susceptible location (e.g. high slope, poor land cover, soil, etc) could also affect the ability of a threshold to detect the landslide, since high susceptibility would mean less intense rain would be required to trigger the landslide. Even the 1.1 km resolution of the MRMS data could contain substantial variation in landslide susceptibility within an individual grid cell. Nonetheless, the reliance on ground-based sensors seems to have been as much of a factor as resolution to the performance of Intensity-Duration Thresholds in identifying landslides; the second-best model performace was using NLDAS-2 data. This could be due to the storm identification challenges mentioned above that put the IMERG data at disadvantage, or because IMERG data are fundamentally less well-suited to landslide identification due producing a precipitation distribution that is erroneously heavy on the low end.

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.

# 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 intead that a lack of noise on the low end may be more important for accurate landslide identification.

A key limitation to studies like this is the lack of exact and verified landslides locations, as reflected in the results presented here, where the exact landslide locations had higher precipitation and greater landslide prediction accuracy as compared to inexact locations. This can be addressed by a manual search as in this study or perhaps in the future by machine learning.

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](https://doi.org/10.1175/1525-7541(2003)004%3C1147:TVGPCP%3E2.0.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.

Alexander, Curtis, David Dowell, Stan Benjamin, Stephen Weygandt, Joseph Olson, Jaymes Kenyon, Georg Grell, et al. 2016. “The Hourly Updated US High-Resolution Rapid Refresh (HRRR) Storm-Scale Forecast Model.” In *EGU General Assembly Conference Abstracts*, 1.

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>.

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, 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 MEASUREMENTSRationale 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>.

Cosgrove, Brian A., Dag Lohmann, Kenneth E. Mitchell, Paul R. Houser, Eric F. Wood, John C. Schaake, Alan Robock, et al. 2003. “Real-Time and Retrospective Forcing in the North American Land Data Assimilation System (NLDAS) Project.” *Journal of Geophysical Research: Atmospheres* 108 (D22). <https://doi.org/10.1029/2002JD003118>.

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): 1330–6. <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>.

Fornasiero, A, R Amorati, P P Alberoni, L Ferraris, and A C Taramasso. n.d. “Impact of Combined Beam Blocking and Anomalous Propagation Correction Algorithms on Radar Data Quality,” 8.

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>.

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>.

Hashmi, Muhammad Zia, Asaad Y. Shamseldin, and Bruce W. Melville. 2011. “Comparison of SDSM and LARS-WG for Simulation and Downscaling of Extreme Precipitation Events in a Watershed.” *Stochastic Environmental Research and Risk Assessment* 25 (4): 475–84. <https://doi.org/10.1007/s00477-010-0416-x>.

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>.

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>.

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>.

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>.

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>.

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>.

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>.

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>.

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>.

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>.