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

Elsa Culler

Andrew Badger

Toby Minear

Kristy Tiampo

Ben Livneh

Rainfall-triggered landslides result in numerous casualties and extensive damage each year in populated mountainous regions. There are many sources of uncertainty that present challenges to skillful predictions and therefore effective mitigation of these destructive natural disasters. One of the largest source of uncertainty in landslide probability comes from the volume and spatial distribution of precipitation volume and intensity during and immediately preceding the landslide event. A key challenge for practitioners is selecting among the wide range of precipitation measurements across different available datasets. Here we investigate the degree of precipitation uncertainty in these landslide-triggering storm events and the impact of that uncertainty on predicted landslide probability using established operational models. 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. In order to evaluate the effects of resolution on the results, we also included a comparison of each dataset re-gridded to match the coarsest spatial and temporal resolution (NLDAS2). 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 acheive 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, indentifying strengths and weaknesses of different products and approaches.

# Introduction

In spite of the destructive nature of landslides, these events are challenging to forecast (Kirschbaum and Stanley 2018). There are many sources of uncertainty that contribute to poor landslide predictions such as anthropogenic modifications to the area and the subsurface structure of the slope. However, perhaps the largest source of uncertainty in landslide probability estimates, is hydrologic uncertainty, here defined as uncertainty in the volume and spatial distribution of both antecedent soil moisture and precipitation during and immediately preceding the event (Chowdhury and Flentje 2002). The key challenge is the wide range of values represented in different available precipitation datasets ranging across in situ observations, remotely sensed retrievals, and numerical weather prediction models. The goal of this chapter is to investigate the role of precipitation uncertainty, and subsequently the uncertainty in modeled soil saturation, in the uncertainty of a contributing to landslide probability.

## Precipitation sensors and estimates

### Ground-based precipitation gauges

* Measure precipitation directly
* Rain gauges are typically considered reference measurements because they the most direct measurement of precipitation (Tapiador et al. 2012).
* Gauges cover a small proportion of land area and are not uniformly distributed, and so only 6.5% of the Earth’s land area between N and S is within 5 km of a gauge (Chris Kidd et al. 2017). As a result, many precipitation products used gauge measurements to correct biases in more homogeneous but indirect measurements.

### Ground-based radar

* Indirect estimate of precipitation based on the return echo power from radar.
* Unlike gauges, radar can detect variability in precipitation over an area rather than a single measurement that may not be representative. Radar estimates are commonly combined with gauge measurements to fill sampling voids in the gauge network (Tapiador et al. 2012).
* Though ground-based radar can cover an area instead of a single point like a gauge, a high-density sensor network is still required for continuous spatial coverage.

### Satellite

* Satellite-based sensors included in the products used in this study are active and passive microwave, infrared and radar.(Christopher Kidd et al. 2020)
* Satellites can provide global, homogeneous coverage. However, as indirect measurements of precipitation where the relationship between the measurement and precipitation can vary by season and location, satellite products require extensive validation and calibration using ground-based methods (Tapiador et al. 2012).

### Models

* Numerical weather prediction is used to forecast weather conditions based on current measurements and physically-based atmospheric models.
* Can incorporate many different types of observations and provide homogeneous and high-resolution estimates over continental scales, but nonetheless vulnerable to biases especially in conditions where the model has less skill since it is not based on current direct observations (Tapiador et al. 2012).

## Precipitation datasets for near-real time landslide predictions

* Low latency: landslide predictions are most useful in advance or hours after rather than weeks or months
* Satellite products tend to capture higher-intensity precipitation (Sun et al. 2018) that can be key in triggering landslides, particularly mudslides and debris flows (Cannon and Gartner 2005). This may be due to the measurement method or the generally higher temporal resolution of satellite products.
* Many precipitation products struggle in mountainous regions (Sun et al. 2018), precisely where landslides are most likely to occur due to higher slopes.

## 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.

Robert F. 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] (Robert F. Adler et al. 2003)) perform better than products based on a single data source.

## Inter-comparison of extreme precipitation

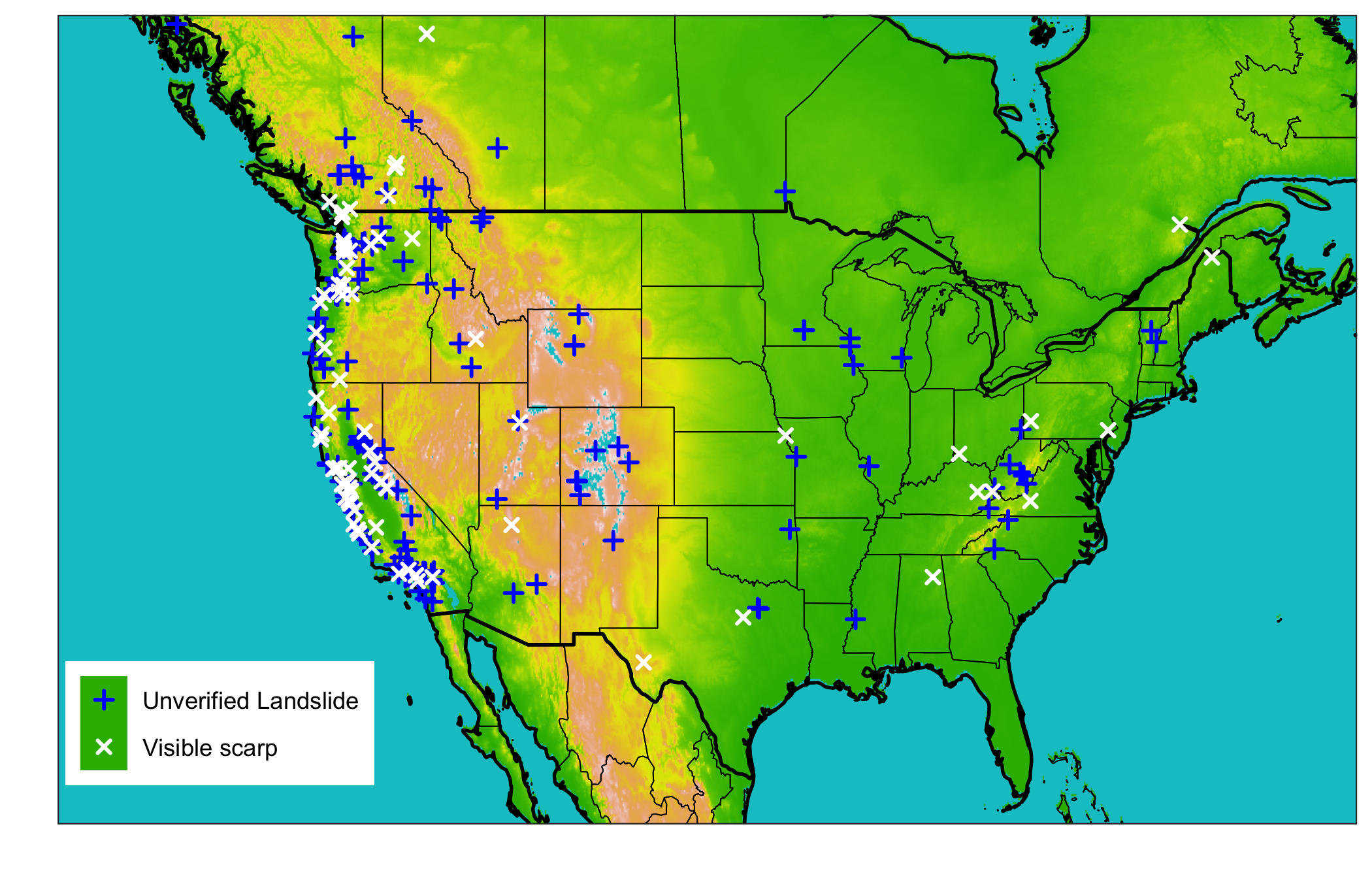
Of the past studies I was able to find, all focused on metrics like total annual or monthly volume and variability. A few looked at extreme precipitation indicators such as 90th percentile precipitation, extreme 1-day precipitation and maximum number of consecutive wet days, but these measures are meant 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 some of the most extreme precipitation 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).
* (Guzzetti et al. 2008) reviewed published Intensity-Duraion Thresholds worldwide.
* This study will use a selection of power-law Intensity-Duration Thresholds to compare the suitability of precipitation measurements from different sources for providing early warning or near-realtime support to disaster response organizations.

# Methods

## Study domain and site selection



Map of all landslide sites colored by whether or not the location was verified using aerial satellite imagery.

## Precipitation product selection

Table 1: 3 Precipitation products that will be 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 (IMERG) early run (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 run (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 km | 2 minutes | < 5 minutes |
| National 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 |

## Identify storm events

* A storm is continuous separated by no more than 24 hours of below-threshold precipitation
* For the purposes of calculating frequency, the maximum precipitation period for each applicable NOAA atlas duration was identified. For example, to find the 5-hour storm return period, the 5-hour precipitation total within the landslide-triggering storm was used. Where a single frequency value was required

## Precipitation comparison for all storms

* Measure of the bias of each product relative to the rest of the measurements
* Measure of the variability of the measurements for each storm

## Precipitation comparison for landslide-triggering storms

* Measure of the bias of each product relative to the rest of the measurements
* Measure of the variability of the measurements for each storm

## Precipitation comparison for use in Intensity-Duration Thresholds

* Intensity-Duration thresholds are simple landslide models whereby a threshold is defined as a power law of the storm duration, and raw or normalized intensities above the threshold predict a landslide. Thresholds were obtained from (Guzzetti et al. 2008), who compiled them from the literature. Thresholds were used on applicable subsets of the data based on climate or other conditions.
* Computed the hit ratio, false alarm ratio, and threat score for each product and threshold.

# Results

## Precipitation comparison for typical storms

### What are some of the ways in which precipitation measurements differ among different products at select sites?

* The precipitation measurements differ substantially in correlation, cumulative volume, and landslide-triggering storm volume.
* Example sites in fig. 1 demonstrate some different degrees and types of variation that occurred at various sites.
* As shown in fig. 1, products that share data sources such as the IMERG products are sometimes but not always more similar to each other than to other products.

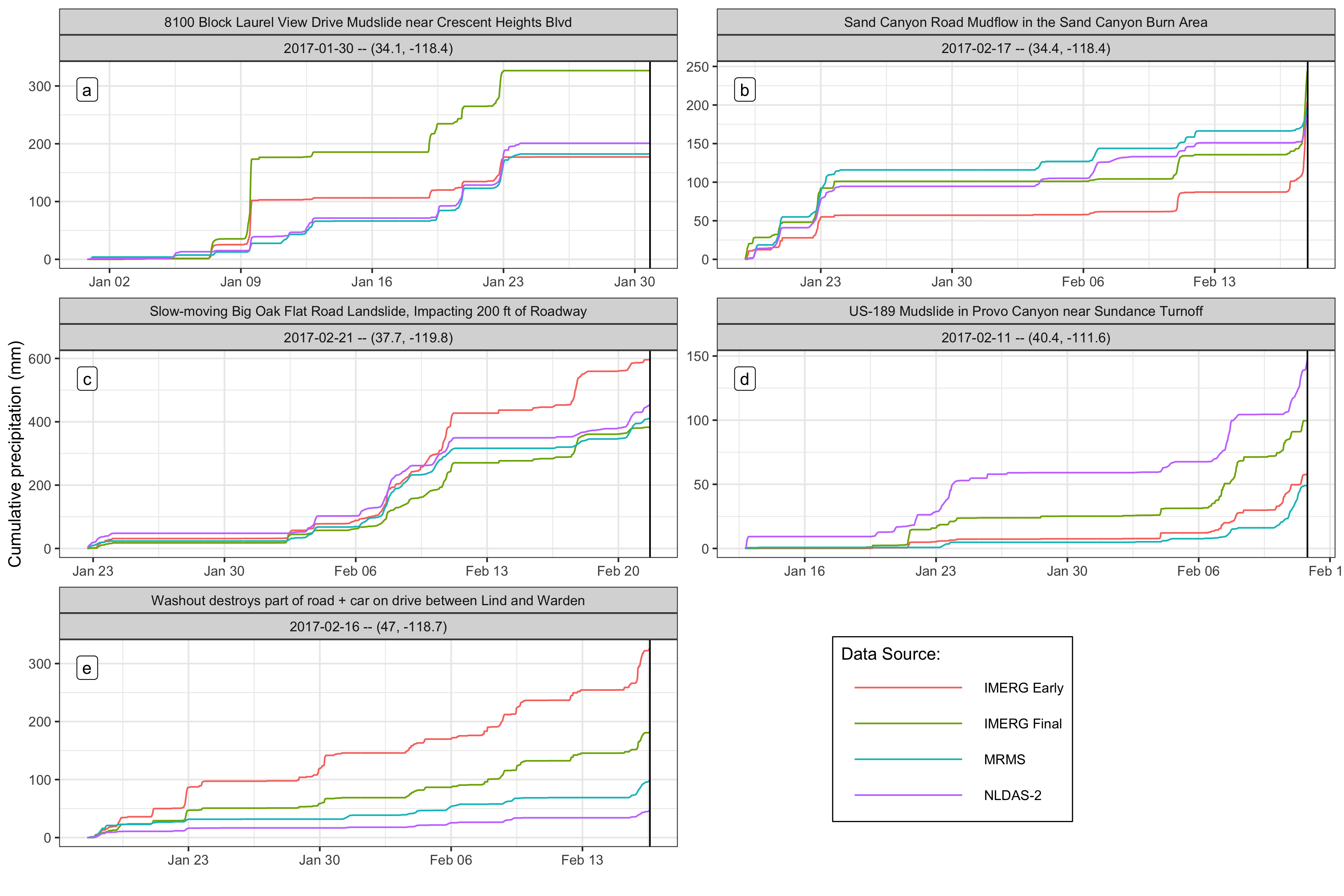
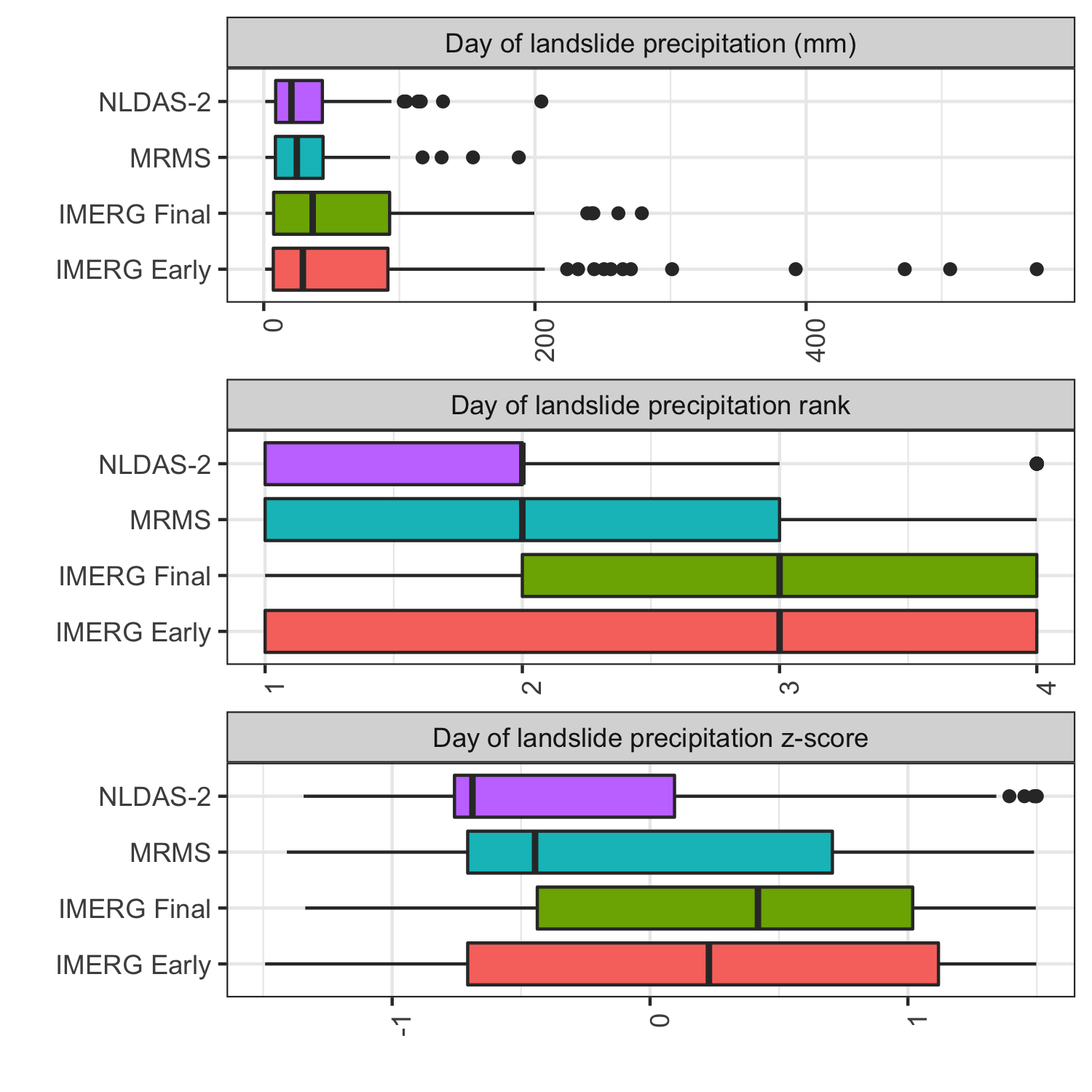


Figure 1: Cumulative precipitation measurements at selected landslide sites for the 30 days before the event. The precipitation is variable across the different products, and the selected sites each demonstrate diverse types of variability. Panel (a) shows a site where no landslide-triggering precipitation 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 volumes do not match.

## Precipitation comparison for landslide-triggering storms

### Is there consistent bias among precipitation products?

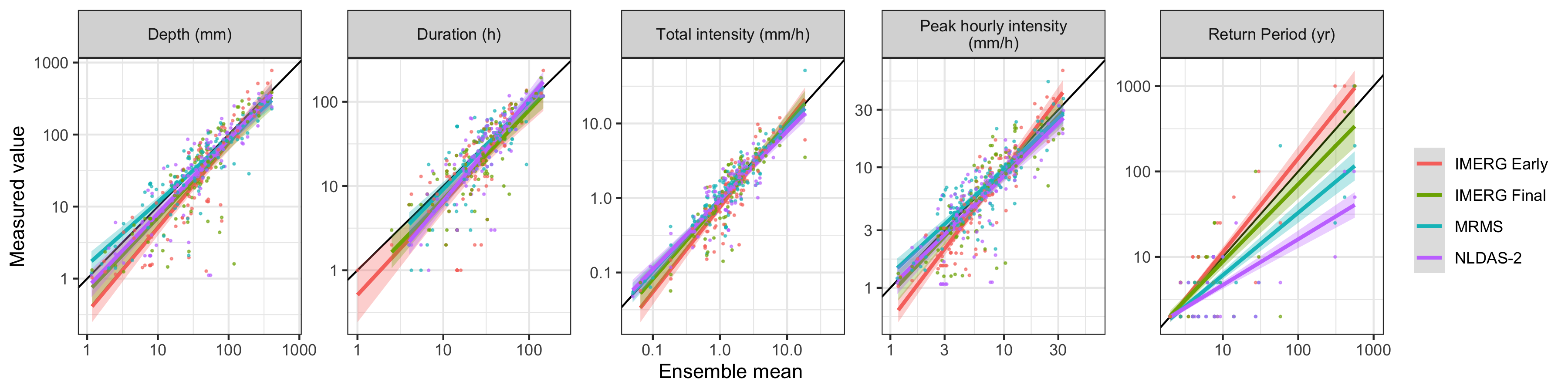
* Generally, the IMERG products have higher day-of-landslide precipitation rank than MRMS which has higher rank than NLDAS-2
* IMERG Early has by some 300mm the highest precipitation measurements in millimeters.
* The range of z-scores for each product is comparable, suggesting that each product is an outlier at some sites



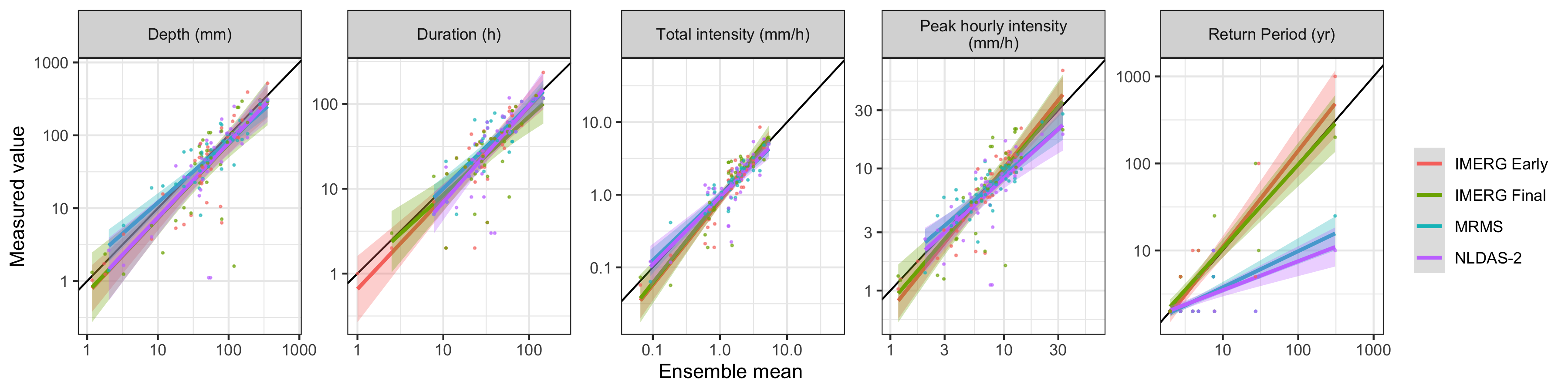
Rank and z-score for day-of-landslide precipitation as measured by each product. The IMERG products tend to have higher rank than MRMS, which typically exceeds NLDAS-2 measurements. The z-scores reflect the same order, but also a similar range of variability across all products.

### How does each precipitation product capture key elements of landslide-triggering storms?

* The IMERG products measure higher peak hourly intensities, which is likely at least partially due to the shorter 30-minute time step.
* The higher peak intensities are also reflected in longer return periods.
* There are many outliers on the low end of the depth and duration measurements, while the total intensity measurements remain close to the ensemble mean.
* Among the verified locations, there are not as many low values either close to the mean or outliers.
* There are also fewer high return periods among the ground-based products MRMS and NLDAS-2 among the verified locations

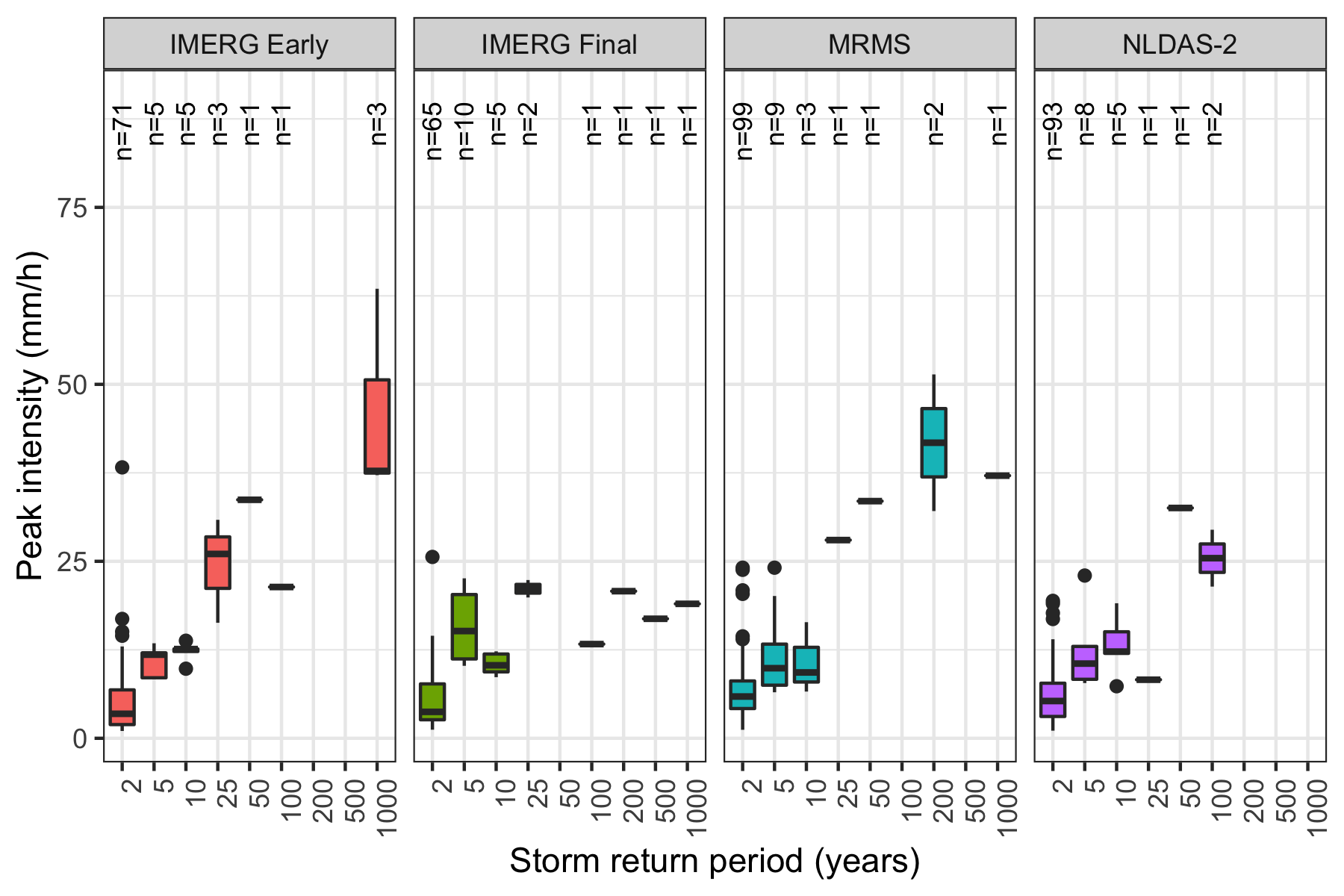


Storm characteristics as measured by each product along with trend lines. The IMERG products measure higher peak hourly intensities, which is likely at least partially due to the shorter 30-minute time step. The higher peak intensities are also reflected in longer return periods. In general there appears to be good agreement among products on the depth and duration of storms, with the exception of outlying low measurements.



### Can peak intensity account for relatively high return period storms causing landslides?

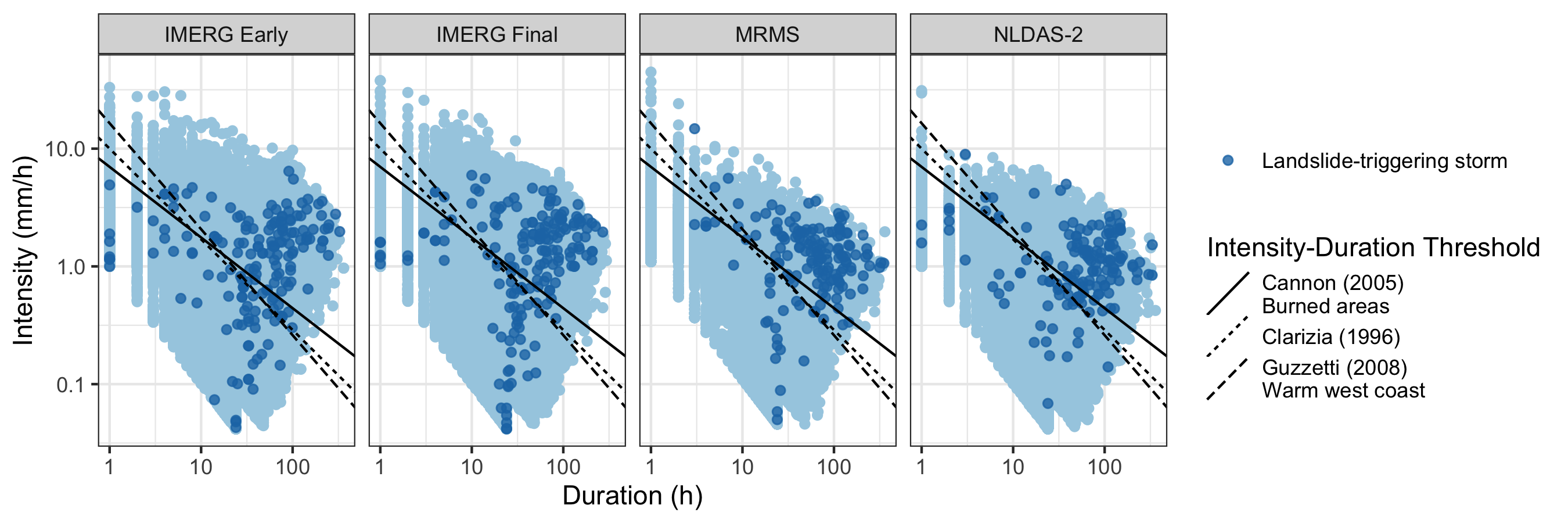
* There appears to be a positive correlation between return period and peak intensity, but this relationship drops off for most products for the the higher return periods.



Peak intensity vs. storm return period. There appears to be a positive correlation between return period and peak intensity, but this relationship drops off for most products among the higher return periods.

### How does each product compare if it were used to predict landslides using an industry standard method of intensity-duration curves?

* These models tend to perform better using MRMS or NLDAS-2 data than using either IMERG product.
  + The IMERG products seem to be more sensitive to both high intensity precipitation and low intensity precipitation
  + The low intensity precipitation may be erroneous noise slightly above the 1mm threshold that causes the storm detection algorithm to select too long of a storm in some cases.
  + Artificially lengthened storms would be expected to have lower intensity values for the whole storm.
* The choice of model does not appear to make as much difference in performance as the choice of precipitation measurement source.
* All models have a better hit ratio when using only verified landslide sites.



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.

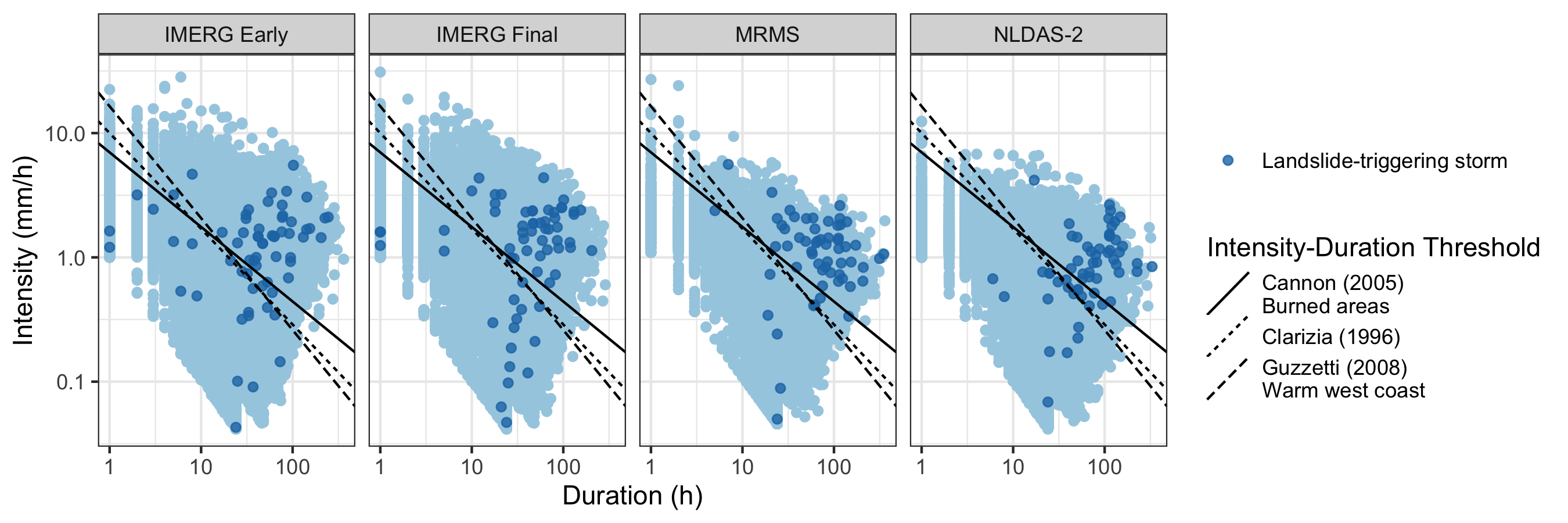


Table 2: Threat score, hit ratio, and false alarm ratio for each product and the Guzzetti (2008) Intensity Duration Threshold

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Product | Include | **Hits** | **Misses** | **Threat score** | **Hit ratio** | **False alarm ratio** |
| GPM IMERG Early | All | 114 | 62 | 0.00676598 | 0.6477273 | 0.2694975 |
|  | Verified | 44 | 21 | 0.006588799 | 0.6769231 | 0.2980977 |
| GPM IMERG Final | All | 117 | 60 | 0.00631068 | 0.6610169 | 0.3074026 |
|  | Verified | 45 | 19 | 0.006249132 | 0.7031250 | 0.3389533 |
| NLDAS-2 | All | 114 | 40 | 0.01433421 | 0.7402597 | 0.2213864 |
|  | Verified | 45 | 14 | 0.014768625 | 0.7627119 | 0.2228354 |
| MRMS | All | 130 | 26 | 0.02492331 | 0.8333333 | 0.2433511 |
|  | Verified | 52 | 7 | 0.023245418 | 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 spatial resolution, temporal resolution, and both
* FIGURE 9: Intensity-Duration Threshold example for each product with matched spatial resolution, temporal resolution, and both

# Discussion

* The satellite products identify to have higher peak intensities and return periods. They also were more sensitive at detecting anomalously low precipitation values, in particular the IMERG Early product.
* Precipitation measurements at verified landslide sites tended to be higher than those at other sites, suggesting that the actual landslide location was too far away from the recorded location for the precipitation measurements to be representative.
* Intensity-Duration Thresholds performed reasonably well at identifying landslides considering that they were trained on different types of data and designed to cover large regions. However, they fared more poorly at excluding false alarms.
* [Degree to which resolution and location accuracy affected performance of Intensity-Duration Thresholds]
* Other factors impacting precipitation measurements could include climate and topography of landslide locations, the density of ground-based sensors,
* Landslide susceptibility caused by slope, soil type, recent wildfire or disturbance, and infrastructure placement could also affect the precipitation intensity or duration needed to trigger a landslide

# Conclusion

* A major limitation to studies like this is the lack of exact and verified landslides, shown in the results for exact landslide locations 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.
* Precipitation products differ greatly in measurement values for the same time and location
* As a result, precipitation products differ in their ability to predict landslides
* Implications for developing early warning systems for landslides across broad regions using remotely sensed precipitation

# 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.

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.

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

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>.

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>.

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>.

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>.

Kidd, Christopher, Yukari N. Takayabu, Gail M. Skofronick-Jackson, George J. Huffman, Scott A. Braun, Takuji Kubota, and F. Joseph Turk. 2020. “The Global Precipitation Measurement (GPM) Mission.” In *Satellite Precipitation Measurement*, edited by Vincenzo Levizzani, Christopher Kidd, Dalia B. Kirschbaum, Christian D. Kummerow, Kenji Nakamura, and F. Joseph Turk, 67:3–23. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-24568-9_1>.

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>.

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.

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>.

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>.