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 [@kirschbaumSatelliteBasedAssessmentRainfallTriggered2018]. 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 [@chowdhuryUncertaintiesRainfallinducedLandslide2002]. 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.

# Background

## Precipitation product comparisons

@sunReviewGlobalPrecipitation2018 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.

@adlerIntercomparisonGlobalPrecipitation2001 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) [@houGlobalPrecipitationMeasurement2014] 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] [@adlerVersion2GlobalPrecipitation2003a]) perform better than products based on a single data source.

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 [@sunReviewGlobalPrecipitation2018; @manzanasPrecipitationVariabilityTrends2014]. 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 Threshold 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 [@scheevelPrecipitationThresholdsLandslide2017].
* [@guzzettiRainfallIntensityDuration2008] 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

FIGURE: map of sites with overlaid graphs showing precipitation product comparison for select sites for ‘typical' storms

## Precipitation product selection

TABLE 1: precipitation products

## Identify storm events

* Discrete storm events: a storm is continuous separated by no more than 24 hours of below-threshold precipitation

## 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

* Compute the precipitation values matching the methods used to create the Intensity-Duration Threshold (e.g. normalizing by the mean annual precipitation or not)

# Results

## Precipitation comparison for typical storms

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

FIGURE 1: Cumulative precipitation comparison for select sites

## Precipitation comparison for landslide-triggering storms

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

* FIGURE 2 : Scatter volume, intensity, frequency, and peak intensity for each product against the ensemble mean
* TABLE 1: Bias and variability for volume and intensity for each product; for exact locations only; and for matched spatial/temporal/both resolutions

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

* FIGURE 3 : Precipitation frequency vs. peak intensity for landslide-triggering storm by site by product

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

* FIGURE(s) 4: Example Intensity-Duration Threshold with multiple precipitation products plotted on top, landslide-triggering events in bold
* TABLE 2: True positives, true negatives, false positives, false negatives, and threat score for each product and threshold (For exact locations only and matched resolutions in parentheses)

## Location accuracy

### Are comparison results affected by location accuracy?

* FIGURE 5: Scatter volume, intensity, frequency, and peak intensity for each product for exact locations only
* TABLE 1
* FIGURE 6: Intensity-Duration Threshold example for exact locations only
* TABLE 2

## Resolution

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

* FIGURE 7: Scatter volume, intensity, frequency, and peak intensity for each product with matched spatial resolution, temporal resolution, and both
* TABLE 1
* FIGURE 7: Intensity-Duration Threshold example for each product with matched spatial resolution, temporal resolution, and both
* TABLE 2

# Discussion

* [Differences in center and spread for different precipitation products]
* [Differences in center and spread for different location accuracy]
* [Differences in center and spread for different resolutions]
* [Degree to which Intensity-Duration Thresholds were accurate in general across broad regions and using different precipitation products]
* [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