Satellite Rainfall Estimates for Debris Flow Prediction: An Evaluation Based on Rainfall Accumulation–Duration Thresholds

E. I. NIKOLOPOULOS
Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut

E. DESTRO
Department of Land, Environment, Agriculture and Forestry, University of Padova, Legnaro, Italy

V. MAGGIONI
Department of Civil, Environmental, and Infrastructure Engineering, George Mason University, Fairfax, Virginia

F. MARRA
Institute of Earth Sciences, Hebrew University of Jerusalem, Jerusalem, Israel

M. BORGA
Department of Land, Environment, Agriculture and Forestry, University of Padova, Legnaro, Italy

(Manuscript received 29 March 2017, in final form 1 June 2017)

ABSTRACT

Rainfall thresholds are often used in early warning systems to identify rainfall conditions that, when reached or exceeded, are likely to result in debris flows. Rain gauges are typically used for the definition of these thresholds. However, in mountainous areas in situ observations are often sparse or nonexistent. Satellite-based rainfall estimates offer a solution to overcome the coverage problem at the global scale but are associated with significant estimation uncertainty. Evaluating satellite-based rainfall thresholds is thus necessary to understand their potential and limitations. In this work, an intercomparison among satellite-based precipitation products is presented in the context of estimating rainfall thresholds for debris flow prediction. The study is performed for the upper Adige River basin in the eastern Italian Alps during 2000–10. Large differences are observed between event-based characteristics (event duration and magnitude) derived from rain gauge and satellite-based estimates, revealing considerable interproduct variability in the debris flow–triggering rainfall characteristics. The parameters of the satellite-based thresholds differ less than 30% from the corresponding rain gauge–based parameters. Results further suggest that the adjustment of satellite-based estimates (either gauge based or by applying an error model) together with spatial resolution has an important impact on the estimation of the accumulation–duration thresholds.

1. Introduction

Rainfall-induced landslides and debris flows constitute natural hazards with severe impacts on human lives and property at global scale (Petley 2012; Dowling and Santi 2014). Alleviation of the risk associated with this hazard has therefore significant societal and economic benefits. A crucial element for successful mitigation strategies is the development of accurate early warning systems to forecast the occurrence of landslides and debris flows in space and time. A majority of operational landslide/debris flow warning systems at regional and national scales are based on the concept of rainfall thresholds (Segoni et al. 2015; Baum and Godt 2010; Alemonti 2004). Rainfall thresholds define the rainfall conditions above which landslides and debris flows are likely to occur and are expressed as a function of rainfall characteristics, commonly duration and accumulation or intensity (Guzzetti et al. 2007).

Corresponding author: E. I. Nikolopoulos, efthymios.nikolopoulos@uconn.edu

DOI: 10.1175/JHM-D-17-0052.1
© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
Rain gauge data are the typical source of information for the definition of these rainfall thresholds. However, in situ observations over mountainous areas, where these hazards mainly occur, are sparse or nonexistent, making the use of gauge-based rainfall thresholds impossible in many landslide/debris flow-prone areas over the globe. Conversely, space-based platforms are able to provide global rainfall observations over complex terrain. The potential of satellite-based rainfall estimates for landslide warning applications was first investigated almost a decade ago by Hong et al. (2006, 2007), who prototyped a global landslide prediction framework based on a global rainfall threshold estimated using the National Aeronautics and Space Administration (NASA) Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) rainfall product (Huffman et al. 2007). Later on, Kirschbaum et al. (2012) demonstrated that the derivation of satellite rainfall thresholds at regional (rather than global) scale appeared to be a more effective approach for landslide warning procedures, which relates to the strong regional dependence of rainfall thresholds reported by many studies (Guzzetti et al. 2007, and references therein). More recently, Mathew et al. (2014) derived intensity–duration thresholds from TMPA estimates for the Garhwal Himalaya area in India, and Posner and Georgakakos (2015) used rainfall estimates from the global Hydro-Estimator of the National Oceanic and Atmospheric Administration (NOAA; Scofield and Kuligowski 2003) for real-time landslide prediction in El Salvador. Robbins (2016) used TMPA estimates for deriving probabilistic rainfall thresholds for landslide triggering in Papua New Guinea, and Rossi et al. (2017) compared TMPA and rain gauge-based rainfall thresholds for the Umbria region in central Italy.

The aforementioned studies greatly highlight the potential of satellite-derived rainfall thresholds. However, all these applications are based on the use of a single satellite rainfall product and thus do not allow deriving conclusions on the relative performance of other available products. Interestingly, despite the large body of literature available on the evaluation of multiple satellite rainfall products at a pixel, catchment, regional, and global scale (see, e.g., Villarini and Krajewski 2007; Mei et al. 2014; Yong et al. 2015), nothing exists, so far, on the intercomparison of satellite rainfall products in the context of landslide and debris flow prediction. In this work, we address this issue by presenting a comparative analysis between rain gauge data and several widely used satellite rainfall estimates—described in the next section—for a large number (~300) of debris flow events in the upper Adige region in the eastern Italian Alps. Results are discussed with respect to different characteristics of satellite rainfall products considered with the aim to highlight current advantages and limitations in the use of satellite rainfall estimates for landslide/debris flow warning applications.

2. Study area and data

The upper Adige River basin is located in the eastern Italian Alps and covers an area of ~7400 km² (Fig. 1). This region is characterized by steep topographic gradients (elevation ranges from 200 to 2900 m MSL) and is affected by significant societal risk due to rainfall-induced hazards such as landslides and debris flows (Salvati et al. 2010). Debris flows are frequent and occur predominantly in the summer season, usually as a result of mesoscale convective systems (Nikolopoulos et al. 2015b). A detailed catalog of debris flows in the study area has been compiled by local authorities (South Tyrol province) based on field surveys that provide information on debris flow initiation points, georeferenced with an accuracy of 50 m, and time occurrence with a daily accuracy. In this study we analyzed the same record as in Nikolopoulos et al. (2014), which includes data for 442 debris flows during 2000–10 along with hourly rain gauge observations from approximately 100 stations covering the area (Fig. 1).

The satellite rainfall products involved in the analysis were selected on the basis of their differences in 1) space–time resolution, 2) retrieval algorithm, and 3) adjustment procedure. These include the version 7 TMPA products in the near-real-time (3B42RT) and past real-time gauge-adjusted (3B42) versions, which are based on rainfall retrievals from both microwave (MW) and infrared (IR) satellite observations. Both 3B42RT and 3B42 are available at 0.25°/3-hourly resolution. We also considered the gauge-adjusted version of the NOAA Climate Prediction Center morphing technique (CMORPH) at both coarse (CMORPH-Coarse at 0.25°/3-hourly resolution) and high (CMORPH-High at 0.07°/0.5-hourly resolution) space–time resolutions. CMORPH rainfall estimates are derived from MW-based retrievals and are propagated in space using IR observations (Joyce et al. 2004). Last, we considered the ensemble mean (EnsMean) of 50 error-corrected realizations of 3B42RT stochastically generated using the two-dimensional stochastic rainfall error model (SREM2D) developed by Hossain and Anagnostou (2006), which was set up and calibrated for the study area [see Maggioni et al. (2017) for details]. Selection of the total number of realizations was based on the analysis presented by Maggioni et al. (2017), who showed that ensemble statistics stabilize for a number of realizations above 40. SREM2D is able to model the spatial structure of satellite rainfall retrieval error as well as the detection of rainy and nonrainy areas of a
given rainfall product [for details, refer to Hossain and Anagnostou (2006)], aspects that are both very important in the context of accurate detection of debris flow–triggering rainfall events.

3. Analysis of debris flow–triggering rainfall

The preprocessing of rainfall datasets for the identification of debris flow–triggering rainfall events involved two main steps. First, debris flow locations available from the catalog were spatially matched to satellite pixels and rain gauges. For the satellite-based products, the pixel covering the debris flow initiation point was considered while for the rain gauges, the single nearest neighbor to the debris flow location was considered, following what is commonly done in gauge-based debris flow–triggering rainfall estimation (Nikolopoulos et al. 2014, 2015a). Second, the time series extracted for each gauge/pixel were processed to identify and separate the rainfall events that triggered the debris flows. Events covering the date of occurrence of the debris flows were considered and a 24-h no-rain threshold was used as minimum interevent duration. Among the 442 debris flows in the database, a total of 298 debris flows were considered, for which all the analyzed rainfall products did not report missing data.

The analysis focuses on the accumulation and duration of event rainfall, as these characteristics are primarily considered in the derivation of rainfall thresholds for landslide and debris flow warnings (Guzzetti et al. 2007). Results are shown and discussed in terms of differences with respect to the rain gauge–based estimates, which are considered the reference for the comparison with the satellite precipitation products.

First, bulk statistics considering all events are calculated to provide information on the differences between satellite and rain gauge rainfall estimates. Table 1 reports bias, centered root-mean-square difference (CRMSD), and Pearson’s correlation coefficient calculated according to

\[
\text{bias} = E(S) - E(G),
\]

\[
\text{CRMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [(S_i - E(S)) - (G_i - E(G))]^2},
\]

and

\[
\text{correlation} = \frac{\text{cov}(S, G)}{\sigma(S)\sigma(G)},
\]

where \(S\) and \(G\) are vectors, including the satellite \(S_i\) and rain gauge \(G_i\) rainfall estimates for each event \(i\); \(N\) is the total number of events (298); and operators \(E(\cdot)\), \(\sigma(\cdot)\), and \(\text{cov}(\cdot)\) correspond to the expectation, standard deviation,
and covariance, respectively. Overall, results suggest that the majority of satellite rainfall estimates are associated with an underestimation (i.e., negative bias) of the event rainfall accumulation and an overestimation of the event duration. Correlation is lower for rainfall accumulation (0.2–0.3) than duration (0.4–0.6) and the random error (represented by CRMSD) is high for both rainfall properties, but particularly for rainfall accumulation, for which it appears to be 2–3 times higher in magnitude than the systematic error. It is worth noting that the ensemble mean of error-corrected fields is associated with the lowest bias (in both accumulation and duration) among all satellite estimates. This demonstrates the improvement gained from the application of the stochastic error model (Maggioni et al. 2017).

Apart from the bias that provides information on the absolute magnitude of the differences, it is useful to examine the magnitude of the relative differences, which permit comparison for different rainfall magnitudes. The bias ratio, defined as

\[ \text{bias ratio} = \frac{S}{G}, \]

is analyzed for different classes of duration and rainfall accumulation (Fig. 2). Classes of rainfall accumulation (or duration) are created according to the quartiles of the distribution of rain gauge estimates. According to the rain gauge estimates, the quartiles (i.e., 25th, 50th, and 75th) of rainfall accumulation (duration) correspond to values of 20, 40, and 88 mm (12, 27, and 52 h), respectively. Figure 2 reports the distribution of bias ratios (presented as box plots) per quartile class. The direction of bias ratio varies with rainfall accumulation: lower (higher) rainfall accumulation events are generally overestimated (underestimated) (Fig. 2a). Estimates for both lowest and highest rainfall accumulation classes exhibit significant deviation from the reference gauge data, while variability among products is considerable and also depends on different classes (Fig. 2a). The highest-duration classes are associated with the largest underestimation in rainfall accumulation (Fig. 2b). On the other hand, satellite estimates overestimate event duration and the extent of overestimation reduces for larger rainfall accumulations and for longer durations (Figs. 2c,d). The same results (Figs. 2c,d) suggest that satellite estimates of duration are in close agreement for all products for long-duration events (bias ratio close to 1 and reduced variability). Overestimation of duration can be related to two possible causes. The first is false detection, which is a known issue in satellite rainfall estimates. The second is the fact that rainfall can occur within part of the satellite pixel but not necessarily covering the point where rain gauge is located. In this case discrepancy is caused by differences in sampling area, which is expected to be more important for the coarser-resolution products.

Overall, the main outcome of the results portrayed in Fig. 2 is that there are large differences in the representation of triggering rainfall characteristics among products in terms of both rainfall accumulation and duration. Investigation of the characteristics of the events in terms of rainfall intensity (Table 2) suggests that different rainfall types may dominate different classes. For example, a short (long)-duration class that corresponds to higher (lower) rainfall intensity events can be dominated by convective (stratiform) type. However, a relationship between the differences reported in Fig. 2 and rainfall intensity characteristics (Table 2) is not conclusive and more research is required to clarify the extent of this connection. Again, the spatial-scale mismatch in sampling area between rain gauge and satellite can partially explain some of the observed differences. Furthermore, the considerable variability among the satellite-based products suggests that other factors contribute to these discrepancies. Differences between 3B42RT and 3B42 are associated with gauge adjustment applied to the latter product. Differences between CMORPH-Coarse and CMORPH-High provide an appreciation of the effect of resolution, which appears to be substantial for both total rainfall and duration for the higher classes. Last, the comparison of 3B42RT and EnsMean highlights the differences in rainfall representation produced after applying a complex stochastic error correction method. Note that the stochastic error model accounts for missed and false detections, thus allowing for altering rainfall occurrence and therefore...
duration relative to the original product (in our case 3B42RT). It is important to note that the main purpose of the analysis presented here is not to identify the “best” satellite rainfall dataset, but rather to highlight the relative differences among products in estimating the amount and the duration of the debris flow–triggering rainfall events. Propagation of those differences on the identification of rainfall thresholds used for debris flow prediction is the primary objective of this study and is examined in the next section.

4. Comparison of rainfall thresholds and prediction accuracy

A commonly used functional form of rainfall thresholds for landslide/debris flows prediction is the power-law relationship (Guzzetti et al. 2007; Peruccacci et al. 2012), defined as

\[ E = \alpha D^\beta, \]

where \( E \) is the rainfall accumulation; \( D \) is the duration of the triggering event; and \( \alpha \) and \( \beta \) are the scaling and shape parameters, respectively. According to this formulation we used the frequentist estimation method proposed by Brunetti et al. (2010) to estimate the event rainfall accumulation–duration (hereinafter ED) threshold for all rainfall datasets. Thresholds, in this work, are estimated at 5% exceedance level, which means that the probability of a

<table>
<thead>
<tr>
<th>Rainfall accumulation quantiles (mm h(^{-1}))</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;25th</td>
<td>0.5</td>
<td>1.3</td>
<td>4.2</td>
</tr>
<tr>
<td>25–50th</td>
<td>0.7</td>
<td>1.9</td>
<td>5.6</td>
</tr>
<tr>
<td>50–75th</td>
<td>0.5</td>
<td>1.3</td>
<td>4.8</td>
</tr>
<tr>
<td>&gt;75th</td>
<td>0.3</td>
<td>1.3</td>
<td>4.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rainfall duration quantiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;25th</td>
</tr>
<tr>
<td>25–50th</td>
</tr>
<tr>
<td>50–75th</td>
</tr>
<tr>
<td>&gt;75th</td>
</tr>
</tbody>
</table>
FIG. 3. ED thresholds, estimated at 5% exceedance level for all rainfall products analyzed.
debris flow–triggering rainfall event (E–D pair) to not exceed the ED threshold is less than 5% [see Brunetti et al. (2010) for more details]. The visual manifestation of the differences in E–D values among the different products corresponds to the differences in the scatter of points shown in Fig. 3. Table 3 reports the parameters of the thresholds derived from rain gauge and from the different satellite products. Despite the relative differences in estimation of E and D, deviation of the satellite-based threshold parameters from the rain gauge–based threshold parameters is within 30%, with some products exhibiting differences lower than 10%. All products underestimate β, while, for the scale parameter, the 3B42- and CMORPH-based products over- and underestimate, respectively. The “perfect match” of EnsMean and gauge-based thresholds is a good example of a case where we have the “right result for the wrong reason” considering that their corresponding E–D spectra (Figs. 3a,f) have many differences. This highlights the complexity and nonlinearity of the propagation of differences in rainfall characteristics (accumulation and duration) to the corresponding differences in exceedance thresholds. Last, it is worth noting that factors like gauge adjustment (3B42RT vs 3B42), higher resolution (CMORPH-Coarse vs CMORPH-High), and error model correction (3B42RT vs EnsMean) resulted in consistent improvement (i.e., better agreement with the reference) of both satellite-based threshold parameters. The decrease in threshold parameters with coarser spatial resolution is in agreement with the recent finding of Marra et al. (2017), who investigated the impact of rainfall spatial aggregation on the estimation of threshold parameters.

5. Conclusions

The vast advancements in satellite-based rainfall estimation over the last couple of decades have led to a number of global rainfall products with various spatio-temporal resolutions. Although several evaluation studies have shown that these datasets can be associated with significant systematic and random errors (Maggioni et al. 2016a,b), they still provide the only source of rainfall information over many areas around the globe and thus offer a unique opportunity for hydrological and geomorphological studies in those areas. In this work, we present an intercomparison of several satellite-based products in the context of rainfall thresholds for debris flow prediction. Comparison between debris flow–triggering rainfall characteristics derived from rain gauge and satellite-based estimates showed large differences in both duration and magnitude of triggering rainfall events. It is important to acknowledge that debris flow–triggering rainfall estimates from rain gauges can be associated with several sources of uncertainty, including, among others, the distance and elevation difference between gauge and actual debris flow initiation point. In fact, comparisons between ED thresholds revealed differences on the order of 30%, which are within the range of uncertainty of rain gauge–based thresholds impacted by errors from rainfall sampling and spatial estimation (Marra et al. 2014; Nikolopoulos et al. 2015a; Destro et al. 2017). Therefore, while results presented in this study clearly demonstrate the relative differences among satellite rainfall products, the absolute accuracy of threshold estimation should be further examined by using high-resolution radar rainfall estimates as reference (Marra et al. 2014, 2016). Irrespective of the relative differences among the products, it is important to highlight that the power-law structure of the E–D spectrum was manifested in all products, suggesting that the traditional power-law model could still be considered in a satellite-based threshold warning system. Future work should focus on different regions and involve satellite products associated with the recently launched Global Precipitation Measurement mission.

Acknowledgments. We sincerely acknowledge Ripartizione Opere Idrauliche, Autonomous Province of Bolzano (Italy), for providing access to the archive of debris flows and Ufficio Idrografico, Autonomous Province of Bolzano (Italy), for providing the rain gauge data. The authors thank the two anonymous reviewers for providing constructive comments that helped us improve this work. Marra was supported by the Lady Davis Fellowship Trust (Project: RainFreq), Grant 1007/15.

REFERENCES


