Incorporation of Satellite Precipitation Uncertainty in a Landslide Hazard Nowcasting System

SAMANTHA H. HARTKE AND DANIEL B. WRIGHT

Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, Wisconsin

DALIA B. KIRSCHBAUM

Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland

THOMAS A. STANLEY

Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, and Universities Space Research Association/GESTAR, Columbia, Maryland

ZHE LI

Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, Wisconsin

(Manuscript received 24 December 2019, in final form 5 May 2020)

ABSTRACT

Many existing models that predict landslide hazards utilize ground-based sources of precipitation data. In locations where ground-based precipitation observations are limited (i.e., a vast majority of the globe), or for landslide hazard models that assess regional or global domains, satellite multisensor precipitation products offer a promising near-real-time alternative to ground-based data. NASA's global Landslide Hazard Assessment for Situational Awareness (LHASA) model uses the Integrated Multisatellite Retrievals for Global Precipitation Measurement (IMERG) product to issue hazard "nowcasts" in nearreal time for areas that are currently at risk for landsliding. Satellite-based precipitation estimates, however, can contain considerable systematic bias and random error, especially over mountainous terrain and during extreme rainfall events. This study combines a precipitation error modeling framework with a probabilistic adaptation of LHASA. Compared with the routine version of LHASA, this probabilistic version correctly predicts more of the observed landslides in the study region with fewer false alarms by high hazard nowcasts. This study demonstrates that improvements in landslide hazard prediction can be achieved regardless of whether the IMERG error model is trained using abundant ground-based precipitation observations or using far fewer and more scattered observations, suggesting that the approach is viable in data-limited regions. Results emphasize the importance of accounting for both random error and systematic satellite precipitation bias. The approach provides an example of how environmental prediction models can incorporate satellite precipitation uncertainty. Other applications such as flood and drought monitoring and forecasting could likely benefit from consideration of precipitation uncertainty.

1. Introduction

Landslides result in thousands of fatalities, property loss, and infrastructure damage around the world every year (Dilley et al. 2005; Froude and Petley 2018; Petley 2012). They occur across a broad range of geographic, climatic, and land use settings and can range from minor slope failures to kilometers-long debris flows. Factors

Corresponding author: Samantha H. Hartke, shartke@wisc.edu

that determine landslide hazard can be sorted into two categories: 1) *static factors* that determine an area's preexisting susceptibility to landsliding, such as slope, aspect, forest loss, road cut activity, lithology, and distance to fault zones, and 2) *dynamic factors* that trigger landslides (Dai et al. 2002; Sassa et al. 2014). Static factors can be conceptualized as determining where landslides are most likely to occur and dynamic factors as determining *when* they occur within susceptible areas. Though landslides can be initiated by seismic

DOI: 10.1175/JHM-D-19-0295.1

and human activity, rainfall is recognized as the most widespread and frequent trigger (Dai et al. 2002; Guzzetti et al. 2007; Petley et al. 2005).

Most existing landslide hazard monitoring systems use ground-based precipitation measurements. Japan and Norway, for example, operate nationwide early-warning systems that use radar rainfall measurements (Krøgli et al. 2018; Devoli et al. 2015; Osanai et al. 2010), while Italy's early-warning system for rainfall-induced landslides, SANF, and Rio de Janiero's "Alerta Rio" system rely on rain gauge networks (Calvello et al. 2015; Piciullo et al. 2017; Rossi et al. 2012). In many parts of the globe, however, including in steep terrain and developing countries, such measurements are often lacking (Gebregiorgis and Hossain 2014; Kidd et al. 2017), hampering real time monitoring and warning of potential landslide hazards.

Satellite multisensor precipitation products (SMPPs) provide near-real-time estimates of precipitation with near-global coverage, potentially enabling prediction of landslides and other environmental phenomena in locations and at scales not previously possible. SMPPs use algorithms that merge passive microwave and infrared sensing data from multiple satellites (e.g., Kidd and Levizzani 2011; Kidd and Huffman 2011; Tapiador et al. 2012; Wright 2018). Commonly used SMPPs include the TRMM Multisatellite Precipitation Analysis (TMPA; Huffman et al. 2007), the Climate Prediction Center (CPC) morphing technique (CMORPH; Joyce et al. 2004), and the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN; Sorooshian et al. 2000). In this study, we use the recent Integrated Multisatellite Retrievals for GPM (IMERG; Huffman et al. 2017), which is available at 0.1°, 30-min resolution available for 2000-present. IMERG retrieves passive microwave (PMW) precipitation estimates using the Goddard profiling algorithm (GPROF) and blends PMW estimates with the IR-based PERSIANN-Cloud Classification System algorithm using the CMORPH-Kalman filter Lagrangian time interpolation scheme (Huffman et al. 2017; Joyce et al. 2004; Sorooshian et al. 2000).

Hong et al. (2006) first demonstrated the potential for near-real-time global landslide hazard assessment by combining TMPA with a global susceptibility map (Hong et al. 2007). Farahmand and AghaKouchak (2013) developed a global landslide model based on Support Vector Machines that used PERSIANN precipitation data to globally classify landslide events. Brunetti et al. (2018) and Nikolopoulos et al. (2017) evaluated the suitability of SMPPs for use in landslide prediction in Italy, finding that rain gauge products and SMPPs may differ in their estimation of landslide-triggering precipitation thresholds. Kirschbaum et al. (2015b)

developed a model to assess landslide hazard in Central America using precipitation data from TMPA. This effort evolved into the global Landslide Hazard Assessment for Situational Awareness (LHASA) model framework, which uses precipitation data from IMERG to provide publicly available near-real-time "nowcasts" of landslide hazard around the world (Kirschbaum and Stanley 2018). Nowcasts identify areas with currently elevated landslide hazard by indicating either moderate hazard (yellow shading) or high hazard (red shading) in LHASA model output.

None of these existing global landslide hazard models explicitly address the systematic biases and random errors that are prevalent in SMPPs. These errors pose a key obstacle to the usage of satellite precipitation in landslide hazard prediction and environmental prediction more generally (e.g., AghaKouchak et al. 2011; Tian et al. 2009; Wright et al. 2017). SMPP performance declines at high latitudes and over ice-covered land surfaces (Ferraro et al. 2013; Tian and Peters-Lidard 2010). SMPPs have difficulty accurately depicting the extreme rainfall rates and orographic enhancement that typify landslide triggering conditions (Shige et al. 2013; AghaKouchak et al. 2011). Furthermore, retrospective studies that characterize bias and other error statistics of satellite precipitation based on comparison to ground reference data, which constitute the bulk of existing SMPP error studies (see Maggioni et al. 2016) are not directly applicable to models that ingest precipitation estimates. Instead, it is necessary to use an error model, which can generate a corrected estimate, range, or distribution of errors or values as soon as new satellite estimates are made available. Development of such models is nontrivial, since SMPP errors tend to be non-Gaussian, heteroscedastic, and can be both discrete and continuous (Maggioni et al. 2014; Tian et al. 2013).

To date, SMPP error modeling has relied on ground-based precipitation data (e.g., rain gauges; bias-corrected radar) to use as "ground truth" in order for model fitting (henceforth referred to as "training"). Unfortunately, many parts of the world lack spatially or temporally complete records of ground-based precipitation (Kidd et al. 2017; Sun et al. 2018). This paucity of ground-truth information (e.g., limited numbers of rain gauges) has led previous error modeling studies to note that some form of "regionalization" of error estimates or error model parameters would be necessary (Gebregiorgis and Hossain 2014, 2013; Tang and Hossain 2009, 2012). Modeling SMPP errors regionally may reduce finescale variability in error structure, but the alternative is no error models in data-limited regions.

In addition to producing error estimates in the first place, another challenge is enabling environmental models to ingest such estimates. One approach is to generate ensembles consisting of multiple realizations of precipitation time series or space-time fields and then use each ensemble member to drive a prediction model. This approach allows prediction models to be used without any particular modification, as demonstrated in a number of studies using stochastic rainfall input for soil moisture and landslide hazard modeling (Maggioni et al. 2011; Nikolopoulos et al. 2010; White and Singham 2012). Though ensemble prediction models have been developed for some applications, flood forecasting in particular (Cloke and Pappenberger 2009), assembling such ensembles can be nontrivial, while for complex physics-based models, the requisite multiple simulations may not be computationally feasible for real-time applications. An alternative approach is to directly ingest precipitation distributions generated by SMPP error models into environmental prediction models. This latter approach is challenging to implement and has received little attention.

In this study, we combine IMERG with a recent, relatively simple, SMPP error model (Wright et al. 2017) and a probabilistic adaptation of the existing LHASA model framework. The approach is evaluated in the mountainous Appalachian region of the southeastern United States. This region features high-quality groundbased precipitation observations, which are used to train the error model. We examine the sensitivity of both SMPP error estimates and landslide predictions to the quantity of ground reference data used to train the error model. The results highlight the potential value of incorporating both systematic and random precipitation errors into environmental models, as well as the benefits of modifying those models' structures to explicitly accommodate such error estimates. We consider the latter point to be critical, since probabilistic representations of precipitation are anticipated to become more common in the near future (Kirstetter et al. 2018; Wright 2018), and existing prediction models are not generally configured to directly ingest probabilistic estimates of precipitation.

Study region, precipitation data, and landslide susceptibility and inventory data are presented in section 2. The existing LHASA model, the SMPP error model, and a new probabilistic formulation of LHASA are introduced in section 3. Results are presented in section 4, and a discussion follows in section 5. A closing summary and conclusions are provided in section 6.

2. Study region and data

a. Study region

The study region encompasses the Appalachian Mountains in western North Carolina and eastern

Tennessee and extends north into Virginia and Kentucky and south into Georgia and South Carolina (Figs. 1a,b). Extreme precipitation is the primary natural hazard in this region, causing floods and landslides that result in deaths and economic losses (Moore et al. 2015). These can result from tropical cyclones, mesoscale convective systems, orographic uplift, and atmospheric rivers that interact with the region's complex terrain (Barros et al. 2014; Mahoney et al. 2016; Moore et al. 2015). Hurricane Frances was followed by Hurricane Ivan within a 2-week period in September 2004, for example, and caused approximately 400 landslides, 11 deaths, and widespread property damage (Boyle 2014). Landslides continue to pose a threat to the region, with three fatalities occurring in May 2018 (Carter 2018; Doom 2018).

b. Rainfall data

The study period is 2002–18. IMERG Version 6B Early and Late, both of which are used in the operational version of LHASA and are produced with latencies of 4 and 14 h, respectively (Huffman et al. 2015), are used. The Stage IV radar-gauge merged precipitation product (Lin 2011), available over the continental United States at hourly, roughly 4-km resolution, serves as the ground reference. Stage IV has been used previously to validate SMPPs (e.g., AghaKouchak et al. 2011). Though ground-reference data such as Stage IV can contain errors, we compared Stage IV against rain gauge observations (results not shown) and assume that such errors are negligible in comparison to SMPP errors for this study. Stage IV, IMERG Early, and IMERG Late are aggregated to a daily scale. Within the study area, IMERG observations of daily rainfall often differ from Stage IV and IMERG observations of high precipitation are especially error-prone (Figs. 1c,d).

Rather than use precipitation directly, LHASA uses an antecedent rainfall index (ARI) to define the dynamic component of a landslide hazard. ARI_t is a 7-day weighted accumulation of precipitation p_t for day t and the prior 6 days:

$$ARI_{t} = \frac{\sum_{i=0}^{6} p_{t-i} w_{i}}{\sum_{i=0}^{6} w_{i}},$$
 (1)

where $w_i = (i + 1)^{-2}$. ARI is formulated to account for the combined impacts of current and recent precipitation on slope stability since soil moisture, pore water pressure, and other physical processes are not explicitly modeled in LHASA. More details on ARI and assigned

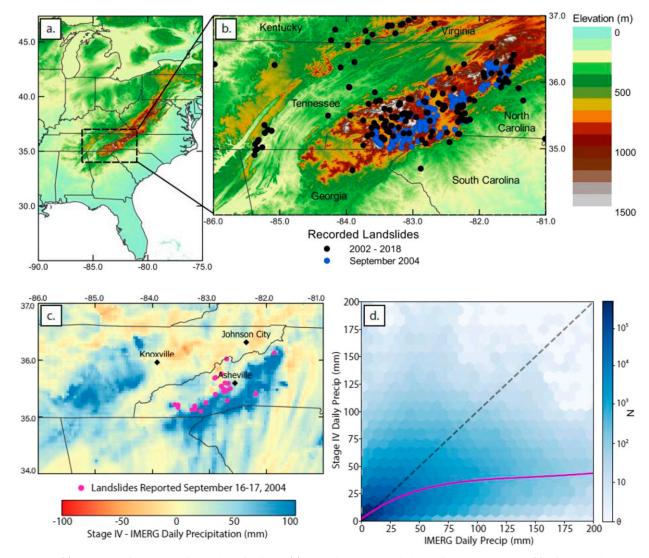


FIG. 1. (a) Eastern United States with study region inset. (b) Elevation and recorded landslides for study area. (c) Difference between Stage IV and IMERG daily precipitation observations on 17 Sep 2004 during Hurricane Ivan. (d) Comparison of IMERG with Stage IV throughout study area during 2002–18. Observations are fitted with a polynomial trend line. Observations less than 1 mm are excluded from this plot.

weights can be found in Kirschbaum et al. (2015b) and Kirschbaum and Stanley (2018).

c. Global landslide susceptibility map and landslide inventory

The LHASA model (section 3a) uses the 1-km global landslide susceptibility map developed by Stanley and Kirschbaum (2017). The map depicts the static susceptibility index (SI) determined using a fuzzy overlay model to combine gridded datasets of five static factors: slope, geology, distance to fault zones, presence of roads, and forest loss. SI consists of integer values from 1 to 5, corresponding to very low, low, moderate, high, and very high susceptibility (Fig. 2). The methodology used

to create the susceptibility map can be found in Stanley and Kirschbaum (2017).

A landslide inventory was obtained from the North Carolina Geological Survey (Wooten et al. 2016) and the Global Landslide Catalog (GLC; Kirschbaum et al. 2010, 2015a), which provide databases of rainfall-triggered landslides based on media reports and other disaster databases. Landslide reports in the GLC include information on the date, location, and impacts of the events and provide an estimate of the landslide's location accuracy.

Only landslides occurring between 2002 and 2018 were used due to the limited length of IMERG and Stage IV records. Additionally, landslide reports were

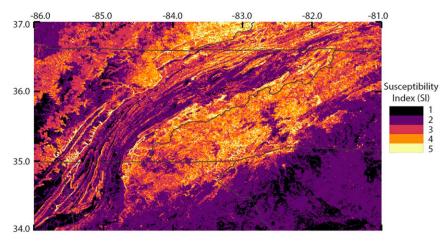


FIG. 2. 1-km static susceptibility map for the study region. Mountainous terrain has the highest susceptibility to landslides in the region.

only considered if the location accuracy in the inventory was within 1 km, if the range of the date of occurrence spanned less than 5 days, and if the trigger was listed as rainfall. The inventory was inspected to remove any duplicate landslide reports. This resulted in an inventory of 214 landslides, approximately half of which are associated with Hurricane Frances and Hurricane Ivan in September 2004 (Fig. 1b). Over 80% of landslides recorded during the 2002–18 study period occurred in high or very high susceptibility areas (SI \geq 4; Table 1).

3. Methods

a. CSGD-based error modeling framework

SMPPs exhibit three types of error: false alarms, in which the SMPP reports precipitation but none actually occurs; missed cases, in which precipitation occurs but the SMPP does not "see" it; and hits, in which the SMPP correctly reports nonzero precipitation, but of the wrong magnitude. The error modeling framework based on censored shifted gamma distributions (CSGDs), developed by Scheuerer and Hamill (2015) for post-processing ensemble numerical precipitation forecasts and adapted by Wright et al. (2017) to model errors in SMPPs, is capable of quantifying all three error types. The CSGD is an adaptation of the two-parameter

gamma distribution (here written in terms of its mean and standard deviation, but which can be reparametrized in terms of shape and scale parameters) with an additional "shift" parameter δ that shifts the probability density function (PDF) leftward. The distribution is left-censored at zero, replacing all negative values with zero. The probability density left of zero thus represents the probability of zero precipitation, while the density at any value greater than zero represents the likelihood of that amount of precipitation (Figs. 3a,b). The CSGD is thus able to describe both precipitation occurrence and magnitude. The cumulative distribution function (CDF; Fig. 3a) is defined by

$$F_{\mu,\sigma,\delta}(x) = \begin{cases} F_{\mu,\sigma}(x-\delta), & \text{for } x > 0, \\ 0, & \text{for } x \le 0. \end{cases}$$
 (2)

A "climatological CSGD" with parameters μ , σ , and δ is fitted to the record of ground-truth precipitation (Fig. 3a). A nonlinear regression system is then trained based on contemporaneous collocated SMPP and ground-truth observations to produce regression parameters α_1 , α_2 , α_3 , α_4 and, at any time t, unique "conditional" CSGD parameters $\mu(t)$, $\sigma(t)$, and $\delta(t)$:

$$\mu_{t} = \frac{\mu}{\alpha_{1}} \log \left[1 + (e^{\alpha_{1}} - 1) \left(\alpha_{2} + \alpha_{3} \frac{\hat{R}_{t}}{\overline{R}} \right) \right], \quad (3)$$

TABLE 1. Static susceptibility zone coverage and landslide occurrence in study region using 388 landslides reported during 2002–18 study period. A total of 214 landslides were used for LHASA model evaluation; the remainder either occurred outside of the 2002–18 study period or failed to meet other necessary reporting requirements.

Static susceptibility	1 (very low)	2 (low)	3 (moderate)	4 (high)	5 (very high)
Study area	8.0%	42.0%	29.0%	16.0%	5.0%
No. of landslides	0 (0%)	5 (1.3%)	68 (17.5%)	225 (58.0%)	90 (23.2%)

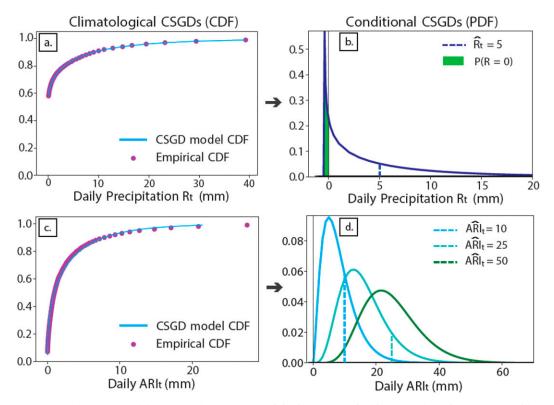


FIG. 3. The CSGD Model fits ground reference precipitation from a pixel in the study region near Asheville, North Carolina, to (a) a climatological CSGD for daily precipitation and optimizes a nonlinear regression to produce (b) conditional CSGDs based on IMERG observation \hat{R}_t . The same training process is used to fit (c) a climatological CSGD for Stage IV ARI and to generate (d) conditional CSGDs for any given IMERG A $\hat{R}I_t$. Panel (d) shows three conditional CSGDs for three possible values of IMERG A $\hat{R}I_t$. The range of error tends to increase with A $\hat{R}I_t$ magnitude. Both the climatological ARI distribution shown in (c) and the conditional CSGDs shown in (d) exhibit a very low probability of ARI $_t$ = 0, since it is a time-integrated precipitation metric.

$$\sigma_t = \alpha_4 \sigma \sqrt{\frac{\mu_t}{\mu}},\tag{4}$$

$$\delta_t = \delta,$$
 (5)

where \overline{R} is the mean of the satellite observations. The terms μ_t , σ_t , and δ_t define the distribution of possible "true" precipitation at time t based on the SMPP observation \hat{R}_t (Fig. 3b). The CSGD framework implicitly downscales estimates if the spatial resolution of the satellite and ground-reference precipitation records differ, as shown in Scheuerer and Hamill (2015). In this study, we use the error model to downscale from IMERG's 0.1° resolution to the $1/24^\circ$ (roughly 4km) resolution of Stage IV. The CSGD error model thus characterizes the relationship between 0.1° IMERG estimates of precipitation and $1/24^\circ$ Stage IV estimates. This relationship will differ from what would be obtained if Stage IV were used at 0.1° resolution, since some discrepancies between the two datasets would be smoothed out at that coarser scale.

The CSGD error modeling framework could be applied to daily precipitation to generate an ensemble of

daily precipitation time series for input to LHASA. Such an approach would be complicated, however, since it would be necessary to preserve temporal autocorrelation in daily precipitation. Instead, we train the CSGD model on ARI time series since ARI is a time-integrated precipitation value [Eq. (1)]. Climatological CSGDs fit to ARI time series show good agreement with empirical ARI CDFs (Fig. 3c), while the probability of zero precipitation both in the climatological CSGD of ARI and in the conditional CSGDs of ARI (Fig. 3d) are unsurprisingly much lower than that for daily precipitation. It has been previously shown that IMERG error depends on the amount and source of passive microwave and infrared data used (Tan et al. 2016). Since this data availability varies over relatively short time scales (generally subhourly), it is not feasible to consider it when modeling multiday ARI.

Since both Stage IV and IMERG observations of ARI are available over the entire study domain, we can train a CSGD error model for any collocated set of IMERG and Stage IV ARI time series. This training method produces unique error model parameters for

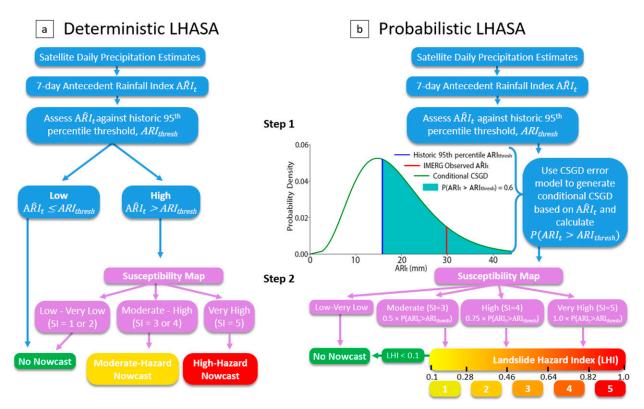


FIG. 4. Decision trees of the (a) LHASA model (adapted from Kirschbaum and Stanley 2018) and (b) probabilistic LHASA model.

each of the 8640 four-kilometer pixels in the study area. We henceforth referred to this as the "localized model." Since ground-based precipitation data are more limited in many parts of the world, we also adopt a regional parameter estimation scheme that may better reflect data-limited settings. The ARI time series of all pixels in the region are concatenated into a single satellite and ground-truth time series to use as input into the CSGD error model framework, generating one set of error model parameters. This regional training scheme neglects the heterogeneity in error characteristics across the region and provides a more generalized model of IMERG error in the study area. The result of this regional approach is henceforth referred to as the "regional model." We also explore the possibility of training a regional model on data from fewer pixels by using ten randomly selected locations (instead of all study area pixels) in the regional parameter estimation scheme. This random regional training, further detailed in section 4a, better reflects operational conditions in data-limited regions where extensive precipitation records are likely unavailable.

Regional and localized CSGD error models for daily ARI from IMERG are trained on the 5-yr period from 2002 to 2006. Both models are incorporated into the probabilistic LHASA model [section 3b(3)]. We assess the consequences of using these different training methods

by comparing the probabilistic LHASA model performance using each model (section 4b).

b. LHASA model

1) Existing "Deterministic LHASA"

The LHASA model characterizes rainfall-induced landslide potential worldwide in near-real time using a heuristic decision tree model (Fig. 4; Kirschbaum and Stanley 2018). Its outputs are 1-km resolution nowcasts, i.e., predictions that indicate either moderate hazard or high hazard if the combination of static and dynamic factors meet specific criteria.

The LHASA model decision tree assesses landslide hazard in two steps. In step 1, satellite-estimated ARI on day t, or $A\hat{R}I_t$, is calculated and compared against the 95th percentile of historical (e.g., climatological) satellite-estimated ARI (ARI_{thresh}) for that location. Thresholds other than ARI_{thresh} could be used; this is outside the scope of our study. IMERG-Early observations are used for the current day due to their lower latency, while IMERG-Late observations are used for the prior 6 days due to their improved accuracy. If $A\hat{R}I_t$ is below ARI_{thresh}, no nowcasts are issued to any of the 1-km pixels covered by the 0.1° IMERG grid cell. If $A\hat{R}I_t$ exceeds ARI_{thresh} , the 1-km static susceptibility map (section 2c) is consulted in step 2.

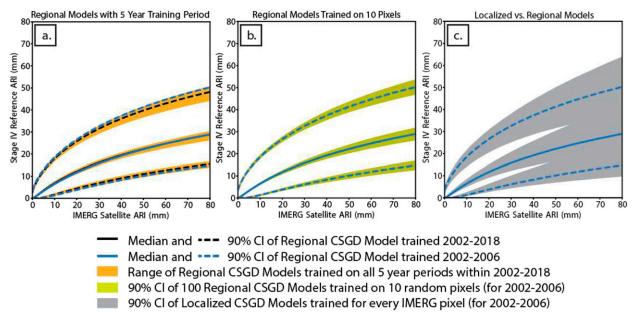


FIG. 5. (a) The 90% confidence interval (CI) and median of regional CSGD error models trained over 2002–18 and 2002–06 and the full range of the 90% CI and median of regional models trained on each 5-yr consecutive period in 2002–18. (b) The 2002–06 regional CSGD model and the 90% CI of the 5th percentile, median, and 95th percentile of 100 CSGD models each trained regionally on 10 random collocated pixels in study area. (c) The 90% CI and median of regional CSGD model trained 2002–06 and the 90% CI of the 5th percentile, median, and 95th percentile of all localized CSGD models, which are trained individually on each collocated IMERG and Stage IV pixel in the study region.

One-kilometer pixels with moderate or high static susceptibility (SI = 3 or 4) are issued a moderate hazard nowcast and very high susceptibility pixels (SI = 5) are issued a high hazard nowcast. All nowcasts are issued at a 1-km resolution of the static susceptibility map.

This existing version of LHASA is referred to as "deterministic LHASA" in recognition that only the IMERG observation is used, without any estimate of the potential uncertainty associated with it. In this study, deterministic LHASA is run on a daily scale using IMERG for the period 2002–18.

2) DETERMINISTIC LHASA USING STAGE IV AND CSGD-MEDIAN

Deterministic LHASA is additionally run using Stage IV precipitation as input for the period 2002–18. Accordingly, the ARI_{thresh} values used in this run of deterministic LHASA are based on Stage IV data. This simulation is conducted to compare satellite-based LHASA predictions to those produced by ground-reference rainfall. This helps to contextualize how much of LHASA's predictive power is associated with precipitation uncertainty, as opposed to other possible sources of prediction error.

We also evaluate a version of deterministic LHASA in which the median of the conditional CSGD, rather than the original IMERG observation, is used to calculate daily $A\hat{R}I_t$. Using the median or mean of the conditional CSGD corrects for systematic bias in SMPPs, but does not address random error. As will be shown in section 4, using this biascorrected value actually degrades LHASA's predictive capability, while considering the full conditional CSGD improves prediction.

3) PROBABILISTIC LHASA WITH IMERG ERROR MODELING

We adapted the LHASA framework so that it is able to incorporate probabilistic precipitation estimates produced by the CSGD error model (section 3a). This modified version is referred to as "probabilistic LHASA" and is shown schematically in Fig. 5. It should be emphasized that our purpose was not to develop a new model, but rather to make as few modifications to the existing framework as possible to allow it to ingest probabilistic precipitation estimates. Key elements, therefore, including the fixed 95th percentile threshold of ARI, remain in the probabilistic version. Concepts such as probabilistic ARI thresholds or a continuous scale of static susceptibility may be useful, but are beyond the scope of this study.

In step 1, a conditional CSGD for day t is generated for each pixel based on the satellite-observed $A\hat{R}I_t$ and the CSGD error model. The conditional CSGD is compared to ARI_{thresh} of the same pixel; $P(ARI_t > ARI_{thresh})$ is the area of the conditional CSGD PDF that lies above

ARI_{thresh} (highlighted as the light blue area in Fig. 4b), where ARI_t is the true ARI at time t, which is unknown. In the CSGD PDF in Fig. 4b, IMERG A $\hat{R}I_t$ is depicted in red to illustrate how deterministic LHASA evaluates rainfall hazard: whether or not A $\hat{R}I_t$ is above or below ARI_{thresh}; probabilistic LHASA, in contrast, utilizes the probability that the actual ARI value, ARI_t, is above this threshold based on the SMPP observation.

After calculating $P(ARI_t > ARI_{thresh})$, probabilistic LHASA factors in static susceptibility using the global susceptibility map (section 2c) in step 2. No nowcasts are issued to pixels with very low or low susceptibility (SI \leq 2). $P(ARI_t > ARI_{thresh})$ for pixels with moderate, high, and very high susceptibility (SI \geq 3) are multiplied by 0.5, 0.75, and 1.0, respectively, to produce a landslide hazard index (LHI). The 0.5 and 0.75 multipliers in probabilistic LHASA were chosen to allow a category 4 (3) hazard nowcast to be issued in high (moderate) susceptibility pixels provided that there is a near certain probability that $ARI_t > ARI_{thresh}$. Pixels with LHI less than 0.1 are assigned as no nowcast. The resulting 0.1-1.0 LHI scale is continuous and reflects a range of moderate to very high landslide hazard that considers precipitation uncertainty. We divide this continuous scale into five equally spaced categories: category 1 LHI = (0.1, 0.28], category 2 LHI = (0.28, 0.46], and so on up to category 5 (Fig. 4b). While these LHI category ranges as well as the aforementioned 0.5 and 0.75 multipliers are assigned somewhat arbitrarily, landsliding events are too sparse in the study area to justify calibration of these values. We show in section 4 that this uncalibrated formulation nonetheless improves prediction compared with deterministic LHASA.

It should be noted that these nowcast categories 1–5 in probabilistic LHASA are not the same as in deterministic LHASA, which instead issues moderate and high hazard nowcasts. The continuous nature of the LHI scale in probabilistic LHASA may lend itself to a variety of nowcast communication approaches that are not available in deterministic LHASA.

c. LHASA evaluation metrics

To assess the performance of deterministic and probabilistic LHASA, the true positive rate (TPR), false positive rate (FPR), and area-wide false alarm ratio (FAR) are calculated. A nowcast is considered correct if it is issued within 1 km of a recorded landslide the day of or the day before the date of occurrence. A correct nowcast is said to "capture" a landslide. A "pixel-day" refers to a 1-km pixel on any given day in the study period, to which a landslide may be reported or a nowcast may be issued. TPR is the percentage of reported landslides that are correctly detected by a landslide hazard nowcast. FPR is the percentage of pixel-days when a nowcast was issued but should not have been (i.e., no landslide was reported within 1 km). Area-wide FAR is the percentage of pixeldays in which a nowcast is issued but no landslide occurs anywhere in the study area on the day of or the day after the nowcast. Area-wide FAR allows for nowcasts to not be considered false alarms as long as a landslide is reported within the study area. This is relevant because the static and dynamic conditions at a particular location can indeed be hazardous even if no landslide actually occurs, and thus a nowcast in such conditions should not necessarily be considered erroneous. These three evaluation metrics are provided in the form of equations below:

$$TPR = \frac{\text{reported landslides captured by a nowcast}}{\text{reported landslides}},$$
 (6)

$$FPR = \frac{\text{pixel-days when nowcast was issued and no landslide occurred}}{\text{all pixel-days in which no landslide occurred}},$$
 (7)

$$FAR = \frac{\text{pixel-days when nowcast was issued and no landslide occurred in study area}}{\text{all pixel-days when nowcast was issued}}.$$
 (8)

4. Results

a. CSGD error model

To explore how SMPP error models may be applied to areas without extensive ground-truth precipitation estimates, it is critical to evaluate the robustness of error model parameter estimation. Localized and regional CSGD error models for IMERG observations of ARI were trained over the 2002–06 period using data from all collocated IMERG and Stage IV grid cells in the study region.

To assess the importance of record length in error model parameter estimation, a regional CSGD model

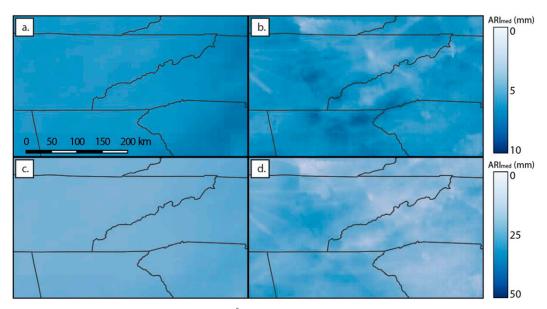


FIG. 6. Median of the conditional CSGD for $A\hat{R}I_t = 10$ mm generated by (a) the regional error model and (b) the localized error model. Median of the conditional CSGD for $A\hat{R}I_t = 50$ mm generated by (c) the regional model and (d) the localized model. The regional error model produces a nearly uniform median (i.e., expected values of ARI_t) for all pixels while the localized model's characterization of error exhibits heterogeneity, including radar artifacts indicative of beam blockage. A conditional CSGD median below $A\hat{R}I_t$ reflects IMERG's positive systematic bias.

trained using the entire 2002–18 period was compared against one trained using only the 2002–06 period (lines in Fig. 5a). These two models are nearly indistinguishable. Additional models were fit using 5-yr subsets from 2003 to 2018 (i.e., 2003–07, 2004–08, etc.; shaded areas of Fig. 5a). Only modest differences from the 2002–06 and the 2002–18 models are visible. These results suggest that relatively short data records can produce satisfactory parameter estimates.

One-hundred additional CSGD error models were fitted that were intended to assess the model variability that would result if spatially incomplete rainfall data were available. Each error model was trained on ARI data (for 2002-06) from 10 pixels randomly selected from the study region. The spread of the uncertainty estimates from these 100 models (shaded areas in Fig. 5b) are very similar to those produced by the 2002– 06 regional model (lines in Fig. 5b), and nearly identical for IMERG-observed ARI below 30 mm. Thus, although the regional CSGD model utilizes all 8640 collocated Stage IV and IMERG pixels in the study region, such a model can be effectively approximated using data from far fewer pixels. The 2002–06 regional model with all pixels is used in the evaluation of probabilistic LHASA (section 4b).

Localized CSGD models fitted for every single pixel in the domain yielded uncertainty estimates (shaded areas in Fig. 5c) that can vary widely from the regional model (lines in Fig. 5c). This highlights that substantial

variation in IMERG-observed ARI error properties can exist throughout the study region (Fig. 6) which is ignored in the regional model. While radar beam patterns and umbrellas are visible in the Localized model outputs in Fig. 6, the regional model smooths out such artifacts of radar estimation, demonstrating one benefit of a regional approach. The regional approach also reduces sampling error because it uses a larger sample of precipitation estimates during model calibration. The implicit downscaling of IMERG uncertainty in the CSGD model training process results in localized models that reflect the very specific relationships between precipitation in individual Stage IV grid cells and the IMERG estimates covering those grid cells. For this reason, radar artifacts indicating beam blockage are visible in the localized model results in Fig. 6. Regardless, we show in section 4b that once uncertainty estimates are integrated into the LHASA framework, the lack of spatial variability in the regional error model approach is of virtually no consequence.

b. LHASA model performance

Fifteen landslides were reported in the study region on 5 May 2003. For this day, deterministic LHASA issues nowcasts encompassing all landslides (Fig. 7a). The same is true for probabilistic LHASA using the regional CSGD error model (Fig. 7b). The spatial distributions of nowcasts differ, however. Deterministic LHASA issues high hazard nowcasts throughout the study area, including a large part of the north where no landslides

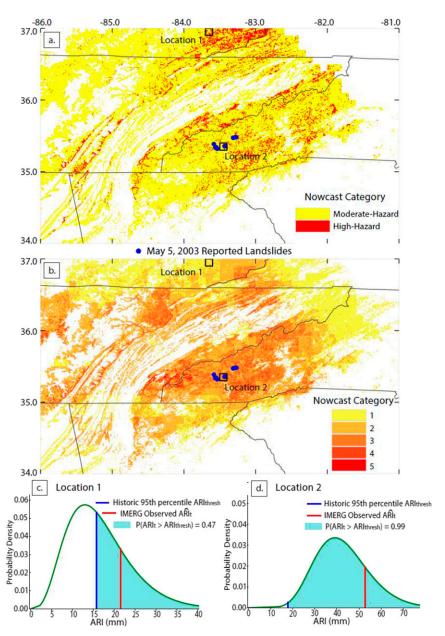


FIG. 7. (a) Deterministic LHASA model output and (b) probabilistic LHASA model (using regional CSGD model) output on 5 May 2003. Locations 1 and 2 are also shown. Conditional CSGD PDFs generated by probabilistic LHASA to describe uncertainty surrounding IMERG-observed ARI on 5 May 2003 at (c) location 1 and (d) location 2.

were reported. Probabilistic LHASA, on the other hand, issues category 3–5 nowcasts in the central part of the study area where most landslides were reported; nowcast categories were generally low (1–3) farther away from reported landslides.

Conditional CSGDs for two 0.1° IMERG pixels demonstrate how deterministic and probabilistic LHASA differ in their evaluation of IMERG-observed $A\hat{R}I_t$ (Figs. 7c,d; locations shown in Figs. 7a,b). IMERG-observed $A\hat{R}I_t$ for

both locations is above ARI_{thresh}. Deterministic LHASA therefore issues moderate hazard (high hazard) nowcasts for all areas within the two pixels that have moderate to high (very high) static susceptibility, in accordance with Fig. 4a. Probabilistic LHASA, on the other hand, estimates the probability that the true ARI_t is above ARI_{thresh} by calculating $P(ARI_t > ARI_{thresh})$. For location 1, $P(ARI_t > ARI_{thresh}) = 0.47$, while for location 2, $P(ARI_t > ARI_{thresh}) = 0.99$ (Figs. 7c,d). For this reason,

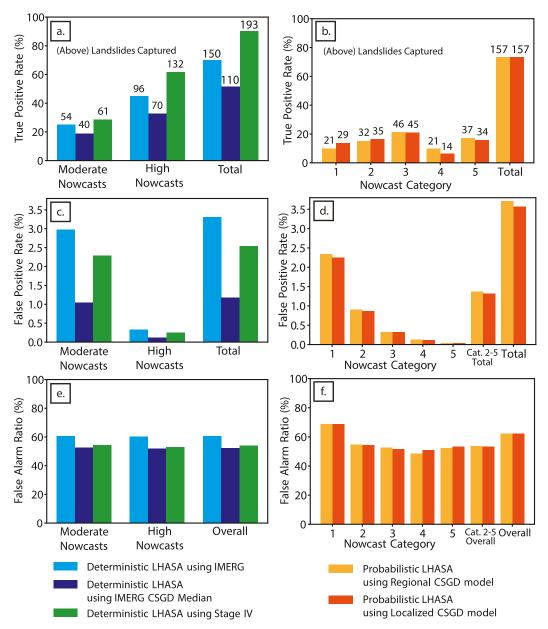


FIG. 8. Evaluation statistics TPR, FPR, and FAR of (a),(c),(e) deterministic LHASA models and (b),(d),(f) probabilistic LHASA models over 2002–18. The total number of landslides in the 2002–18 study period is 214.

probabilistic LHASA issues higher category nowcasts for location 2, in accordance with Fig. 4b.

Over the 2002–18 study period deterministic (probabilistic) LHASA using IMERG captures 150 (157) of 214 reported landslides (Figs. 8a,b). TPR is fairly evenly dispersed for probabilistic LHASA among all nowcast categories (Fig. 8b). FPR for category 5 nowcasts from probabilistic LHASA is an order of magnitude lower than the FPR for high hazard nowcasts in deterministic LHASA (Figs. 8c,d). With the exception of category 1, FAR for all probabilistic LHASA nowcast categories

is less than 55%. This is lower than the FAR of both moderate and high hazard nowcasts from deterministic LHASA (both have FAR of 60%), meaning that probabilistic LHASA category ≥ 2 nowcasts are less likely to be issued than moderate or high hazard nowcasts by deterministic LHASA when conditions do not actually produce landsliding in the region (Figs. 8e,f). The relatively high FPR and FAR for category 1 nowcasts from probabilistic LHASA are expected since these nowcasts are issued when there exists a nonzero but low probability of ARI exceeding ARI_{thresh}.

Deterministic LHASA will never issue a nowcast when the $A\hat{R}I_t$ is less than ARI_{thresh} , which is problematic given the prevalence of large random errors in SMPPs, particularly during extreme rainfall events. In contrast, probabilistic LHASA evaluates $P(ARI_t > ARI_{thresh})$ to determine the probability of hazardous ARI. Though $P(ARI_t > ARI_{thresh})$ may not be high when $A\hat{R}I_t$ is below ARI_{thresh} , probabilistic LHASA can still generate a category 1 or 2 nowcast.

Differences between probabilistic LHASA using the regional and localized CSGD models are negligible (Figs. 8b,d,f). Probabilistic LHASA captures the same number of landslides regardless of which CSGD error model is used, while the localized CSGD error model within probabilistic LHASA produces a slightly lower total FPR (Fig. 8d). Even though the two models generate different $P(ARI_t > ARI_{thresh})$ estimates, resulting in slight differences in which nowcast category is issued, each model is able to broadly identify cases when ARI_t is associated with greater uncertainty and when ARI_t has a nonnegligible probability of exceeding ARI_{thresh} .

Deterministic LHASA using the CSGD median performs poorly, capturing only 110 out of 214 reported landslides (Fig. 8a). Recall from section 3b(2) that the median or mean from the CSGD error model removes systematic bias, but does not reflect random error. This effectively eliminates any high values of ARI, resulting in fewer nowcasts and fewer landslides captured. Although FPR and FAR are low in this case relative to other models (Figs. 8c,e), this is outweighed by an unsatisfactory TPR.

Deterministic LHASA using Stage IV captures 193 out of 214 landslides, with lower FAR and FPR than deterministic LHASA using IMERG. This represents the maximum performance achievable when LHASA is forced with the best available precipitation dataset. In terms of TPR, probabilistic LHASA is closer to this "optimal" performance than either of the IMERG-based deterministic models. Because of the conceptual differences in nowcast categories between the deterministic and probabilistic versions of LHASA, it is difficult to directly compare FAR and FPR between the two versions.

c. Spatial variation in nowcasts

Nowcast rates over the study region, calculated as the percentage of days on which nowcasts are issued for 2002–18, are shown in Fig. 9. Nowcast rates for all models are zero in areas of low or very low static susceptibility (see Fig. 2). The nowcast rate for deterministic LHASA using IMERG (Stage IV) is 6%–8% (4%–6%) for most of the study region in pixels with moderate or higher static susceptibility, which is logical since

 $A\hat{R}I_t$ exceeds ARI_{thresh} on 5% of days. Probabilistic LHASA using the regional CSGD model has higher now-cast rates in the mountainous terrain in the center and northeast of the study region and lower rates in the less-steep northwest (Fig. 9c). Probabilistic LHASA using the localized CSGD model increases nowcast rates along the edges of the mountainous terrain and decreases nowcast rates in the center, reflecting differing levels of uncertainty in IMERG estimates in these two regions (Fig. 9d).

In both cases, nowcast rates from probabilistic LHASA exhibit more spatial variability than deterministic LHASA. While all moderate to very high susceptibility areas within an 0.1° IMERG grid cell will have the same nowcast rate in deterministic LHASA, the nowcast rate in probabilistic LHASA will vary according to static susceptibility (Fig. 9). For instance, if IMERG-observed $A\hat{R}I_t$ in a 0.1° pixel barely exceeds ARI_{thresh}, deterministic LHASA will issue nowcasts for all areas in that pixel with moderate to very high susceptibility, as per Fig. 4a. Probabilistic LHASA, on the other hand, calculates $P(ARI_t > ARI_{thresh})$ to determine which static susceptibility level merits a nowcast. A probability $P(ARI_t > ARI_{thresh}) = 0.2$, for example, would result in nowcasts being issued only for high and very high static susceptibility pixels, as per Fig. 4b.

When category 1 nowcasts, which imply relatively modest landslide hazard, are excluded, the nowcast rate of probabilistic LHASA drops by approximately 50% (Figs. 9e,f). Results show that when probabilistic LHASA issues nowcasts to an area but deterministic LHASA does not, these nowcasts are moderate hazard (category 1) and are issued because probabilistic LHASA predicts a low but nonzero probability of landslide hazard (see nowcasts issued in the north east of study region in Figs. 7a,b).

5. Discussion

a. CSGD error models with limited ground-truth data

CSGD model training results (section 4a) demonstrate that relatively little ground-reference data is needed to approximate the regional SMPP error model trained on abundant ground data. Specifically, 10 randomly selected pixels in the study area are found to be sufficient to approximate a regional CSGD error model of ARI trained over all 8640 pixels (Fig. 5b). This extends the applicability of the regional CSGD error model incorporated in probabilistic LHASA to datalimited regions with few ground reference records, and indicates that SMPP uncertainty can still be accounted for in the LHASA model in data-limited regions provided that at least some amount of ground

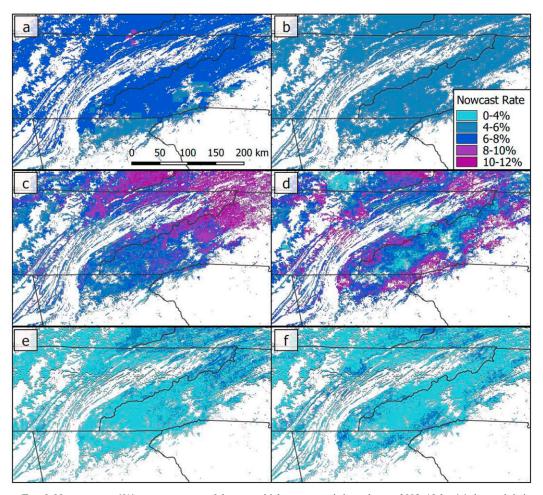


FIG. 9. Nowcast rate (%), or percentage of days on which a nowcast is issued, over 2002–18 for (a) deterministic LHASA using IMERG, (b) deterministic LHASA using Stage IV, (c) probabilistic LHASA using regional CSGD error model, and (d) probabilistic LHASA using localized CSGD error model. Also shown is the nowcast rate considering only category 2–5 nowcasts for (e) probabilistic LHASA with regional error model and (f) probabilistic LHASA with localized error model. Note the significant decrease in nowcast rate for probabilistic LHASA when category 1 nowcasts are excluded.

truth precipitation such as multiyear rain gauge records are available.

Localized CSGD error models trained on individual IMERG and Stage IV pixels reveal that there can be a wide range of variation in the uncertainty around IMERG-based ARI across the study region (Fig. 5c). Such variation could imply that the regional approach may not adequately characterize the uncertainties associated with IMERG-based ARI at any particular location, but could also stem at least in part from sampling error. True regional variability would argue in favor of localized error modeling; regional variability resulting from sampling error (e.g., insufficient record length to sample the extreme tail of the local precipitation distribution) would argue in favor of a regional approach. The consequence of differences

between regional and localized CSGD error models is not readily apparent in comparing the CSGD model results (i.e., Fig. 5c), but is best evaluated through the resulting performance of probabilistic LHASA, which is discussed below. More work would be needed to verify whether or not the conclusions regarding localized versus regional error models of 7-day ARI would pertain to error models of precipitation at shorter time steps (e.g., daily, hourly, etc.).

Even though regional and localized error models generate varying estimates of ARI uncertainty (Fig. 5c), their incorporation into the probabilistic LHASA model produces nearly identical results: improved TPR relative to deterministic LHASA and lower FAR in all but the lowest nowcast category (Fig. 8). This indicates that the mere inclusion of IMERG uncertainty in LHASA is

more critical than whether this uncertainty is estimated based on broader regional error characteristics or localized error characteristics.

b. Deterministic and probabilistic LHASA model performance

Probabilistic LHASA outperforms deterministic LHASA and provides greater gradation in landslide hazard using a continuous nowcast scale. Since probabilistic LHASA models using localized and regional CSGD error models perform similarly (section 5a), discussion of probabilistic LHASA in this section refers to the results using the regional error model.

The superior performance of deterministic LHASA using Stage IV confirms that accurate precipitation data can significantly improve LHASA's performance (Figs. 8a,c,e). The Stage IV-based model only misses 10% (21 out of 214) landslides in the study period. These missed landslides are attributable to some combination of deficiencies in Stage IV, the LHASA model structure, the susceptibility map, or the landslide inventory. Deterministic LHASA using IMERG, in contrast misses 64 out of 214 (30%). This is improved somewhat by incorporating the IMERG uncertainty information via probabilistic LHASA (57 out of 214 landslides missed; 27%). This modest improvement in TPR is accompanied by reductions in FAR and FPR associated with higher hazard (category 2–5) nowcasts and more realistic spatial distributions of nowcasts.

While probabilistic LHASA improves total TPR and landslides captured relative to its deterministic counterpart (Figs. 8a,b), comparing the distribution of TPR across nowcasts is less straightforward since the two model frameworks use different nowcast categories. For deterministic LHASA, TPR for high hazard nowcasts is higher than for moderate hazard nowcasts, and thus can be said to have higher skill. On the other hand, TPR for probabilistic LHASA is more evenly dispersed across nowcast categories. Though this suggests equal skill in all nowcast categories, the FAR and FPR indicate far fewer instances of category 3-5 nowcasts being issued in error. While moderate hazard and high hazard nowcasts from deterministic LHASA are equally likely to be a false alarm (Fig. 8e), the FAR for probabilistic LHASA nowcasts decreases above category 1 such that higher category nowcasts are less likely to be false alarms (Fig. 8f). When category 1 nowcasts are excluded, the total FPR of probabilistic LHASA is half that of deterministic LHASA (Fig. 8d). Probabilistic LHASA category 4 and 5 nowcasts have much lower FPRs and FARs than high hazard nowcasts from deterministic LHASA, which are important for stakeholders to establish how confident they can be in a nowcast. Probabilistic LHASA's ability to issue low category nowcasts even when $A\hat{R}I_t < ARI_{thresh}$ could be particularly beneficial in regions with high random error (such as mountainous terrain with finescale heterogeneities) or where satellite observations frequently underestimate heavy precipitation events (Derin et al. 2019).

Eighty-six percent of reported landslides in the study area occur in moderate and high static susceptibility pixels (Table 1). This suggests that it would be reasonable for moderate and high static susceptibility areas (from SI = 3 to SI = 4) to merit high hazard nowcasts under certain conditions, which is not possible in deterministic LHASA (Fig. 4a). The deterministic LHASA decision tree only permits high hazard nowcasts to be issued in areas with SI = 5 (Fig. 4a). The deterministic LHASA framework could be modified to allow issuance of high hazard nowcasts in areas with SI = 4 by greatly increasing the associated FPR and FAR, or by introducing some moderating factor to decide whether a High susceptibility pixel is issued a moderate or high hazard nowcast. Probabilistic LHASA allows the nowcast category to vary based on the probability of hazardous ARI (Fig. 4b). For example, a category 4 nowcast can be issued to areas with SI = 4 under the circumstance that $P(ARI_t > ARI_{thresh})$ is greater than 0.85. Thus, probabilistic LHASA is able to differentiate between two locations with the same static susceptibility and for which $ARI_t > ARI_{thresh}$, but which have different probabilities of experiencing hazardous precipitation (Figs. 7c,d).

The poor performance of deterministic LHASA using the CSGD-median of IMERG reveals that removing systematic bias in SMPPs can actually significantly worsen model performance, if the potential for random error is neglected. These results emphasize the importance of considering the range of random error present in SMPP observations in addition to systematic bias when accounting for SMPP uncertainty, especially for models of environmental processes driven by extreme precipitation (such as floods and landslides).

c. Spatial nowcast variations

The largest differences in nowcast rates between deterministic and probabilistic LHASA occur in the steepest terrain (Fig. 9), where uncertainty-prone extreme precipitation events occur most frequently. Probabilistic LHASA's incorporation of the relatively wide range of error surrounding extreme events results in more frequent nowcasts. It is also noteworthy that probabilistic LHASA does not uniformly increase the nowcast rate across the entire study region, but actually decreases the nowcast rate in the lower-elevation northwest, which experiences positive systematic bias from IMERG (Fig. 6) and receives less extreme rainfall. Even though the

evaluation statistics for regional and localized CSGD models are very similar (Fig. 8), the spatial patterns of nowcast rate for these two models across the study area are distinct (Figs. 9c,d). Although probabilistic LHASA is issuing more nowcasts than deterministic LHASA (Figs. 9a,c,d), the two models' nearly identical overall FAR implies that probabilistic nowcasts are more likely to be issued on days when conditions actually produce landsliding in the region. Further, if a user chose to only consider category 2–5 nowcasts from probabilistic LHASA, the nowcast rate and associated FAR and FPR drop significantly (Figs. 8d,f, 9e,f).

6. Summary and conclusions

Satellite multisensor precipitation products (SMPPs) offer promise for environmental modeling in regions that lack ground-based sources of precipitation information. This promise has gone largely unfulfilled, however, mainly due to the high uncertainty of satellite-based precipitation estimates. This study demonstrates how accounting for SMPP uncertainty can improve predictions from a landslide hazard model over the mountainous southeastern United States. Historical records of Integrated Multisatellite Retrievals for GPM (IMERG) and ground-based precipitation are used to train an error model that characterizes uncertainty in current and antecedent rainfall. A probabilistic version of NASA's Landslide Hazard Assessment for Situational Awareness (LHASA) model is developed that can translate this rainfall uncertainty into near-realtime landslide nowcasts. The hazard nowcast scale is continuous, though we transform it into a simple fivecategory scale. Run retrospectively using IMERG, the probabilistic version of LHASA performs well in the study region, capturing more landslides than the existing deterministic version with lower false positive rates and false alarm ratios in high hazard nowcast categories. Using the error model to generate only the "best guess" (e.g., median) antecedent rainfall as input for the deterministic LHASA model leads to worse model performance, highlighting that removal of systematic bias is not sufficient for addressing SMPP uncertainty in modeling. Instead, results demonstrate the need to account for the range of SMPP random error.

Data-limited regions have few available ground records on which to train SMPP error models. We show that while ground truth precipitation data are indeed necessary for error model training, such data need not be ubiquitous and a few ground truth records can go a long way. Nearly identical improvements in landslide hazard model performance can be obtained by using a regionalized parameter training scheme instead of a data-intensive localized training scheme, and this regionalized scheme can be robust with relatively few training data. These findings suggest accounting for SMPP uncertainty within a land-slide hazard model is viable even in data-limited regions.

Though the probabilistic LHASA model framework was developed over a limited region in the United States, its performance is promising for application to other regions. The probabilistic nowcast scale, which can be subdivided into discrete categories indicating varied levels of landslide hazard based on factors such as number of desired categories or a desired trade-off between true positive rate and false positive rate, provides a flexible way to adapt landslide hazard nowcast categories to stakeholder requirements. Though the nowcast categories in probabilistic LHASA are presented here as equally subdivided portions of a continuous scale (Fig. 4b), this scale could be divided in any number of ways. This flexibility could be used to optimize specific performance metrics such as TPR or FPR. A stakeholder could, for example, use four categories instead of five, and specify that the highest category have a minimum TPR of 0.25 or a maximum FPR of 0.001. This is consistent with the original goal of the LHASA framework—to provide a flexible model for providing situational awareness of landslide hazard that can be fine-tuned as required (Kirschbaum et al. 2015b; Kirschbaum and Stanley 2018).

Accounting for SMPP uncertainty would allow the LHASA model to better utilize SMPPs globally and in data-limited regions. Since precipitation forecasts such as those from the Global Forecasting System (Saha et al. 2006) are generally produced as ensembles, this methodology could be used to incorporate probabilistic precipitation forecasts as inputs into the probabilistic LHASA model to generate landslide hazard forecasts. We anticipate that other environmental prediction models such as the Global Flood Monitoring System (Wu et al. 2012) and the Global Land Data Assimilation System (GLDAS; Fang et al. 2009) could likely benefit from deeper consideration of SMPP uncertainty and error—both systematic and random.

In this study, we found it necessary to modify (albeit modestly) the operational LHASA framework to accommodate IMERG uncertainty estimates, as opposed to inserting those estimates directly into the existing version. Probabilistic precipitation estimates are likely to become more common in the near future. To realize the full benefits of probabilistic representations of precipitation, we believe that it will be necessary to seriously consider how existing environmental models can be adapted to accommodate such estimates, or how new models could be developed that are explicitly designed around such estimates. We acknowledge that this may be very challenging for more complex physically based models.

Given the centrality of precipitation (and thus, precipitation errors) to so many phenomena, however, relatively simple models that make effective use of precipitation uncertainty could prove useful. In addition, as our experience with LHASA has shown, it may be necessary to rethink how outputs from such models would be produced and interpreted. This study shows that quantifying satellite precipitation uncertainty and converting SMPPs from an unknown to a known source of uncertainty can significantly improve model performance, even if it remains impossible to derive the true precipitation from SMPP observations.

Acknowledgments. S.H. Hartke was supported by the NASA Earth and Space Science Fellowship Program (Grant 80NSSC18K1321) and the Arthur H. Frazier Fellowship at University of Wisconsin-Madison. D.B. Wright, D.B. Kirschbaum, T.A. Stanley, and Z. Li were supported by the NASA Precipitation Measurement Mission Program (Grant NNX16AH72G). We thank Ana Barros for sharing rain gage data from the region, which we used to confirm the accuracy of the Stage IV dataset. We also thank the organizers of the 12th International Precipitation Conference for their contributions towards publication fees. The generous support by NSF (Conference Grant EAR-1928724) and NASA (Conference Grant 80NSSC19K0726) to organize IPC12 and produce the IPC12 Special Collection of papers is gratefully acknowledged.

REFERENCES

- AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai, 2011: Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *J. Geophys. Res.*, 116, D02115, https://doi.org/10.1029/2010JD014741.
- Barros, A. P., and Coauthors, 2014: NASA GPM-Ground Validation Integrated Precipitation and Hydrology Experiment 2014 Science Plan. Duke University Tech. Rep., 64 pp., https://doi.org/10.7924/ GRCCOXMR
- Boyle, J., 2014: From the archives: Impact of Frances, Ivan lingers years later. Asheville Citizen-Times, 6 September, accessed 1 April 2018, https://www.citizen-times.com/story/news/local/2014/09/06/ hurricanes-frances-ivan-impact-lingers-years-later/15217637/.
- Brunetti, M. T., M. Melillo, S. Peruccacci, L. Ciabatta, and L. Brocca, 2018: How far are we from the use of satellite rainfall products in landslide forecasting? *Remote Sens. Environ.*, 210, 65–75, https://doi.org/10.1016/j.rse.2018.03.016.
- Calvello, M., R. N. D'Orsi, L. Piciullo, N. M. Paes, M. A. Magalhaes, R. Coelho, and W. A. Lacerda, 2015: The community-based alert and alarm system for rainfall induced landslides in Rio de Janeiro, Brazil. *Landslide Processes*, G. Lollino et al., Eds., Vol. 2, *Engineering Geology for Society and Territory*, Springer, 653–657, https://doi.org/10.1007/978-3-319-09057-3_109.
- Carter, A., 2018: "Shouldn't you have a right to know?"

 Daughter of landslide victim says warnings needed.

- News & Observer, 4 June, https://www.newsobserver.com/latest-news/article212367984.html.
- Cloke, H. L., and F. Pappenberger, 2009: Ensemble flood fore-casting: A review. *J. Hydrol.*, **375**, 613–626, https://doi.org/10.1016/j.jhydrol.2009.06.005.
- Dai, F. C., C. F. Lee, and Y. Y. Ngai, 2002: Landslide risk assessment and management: An overview. *Eng. Geol.*, **64**, 65–87, https://doi.org/10.1016/S0013-7952(01)00093-X.
- Derin, Y., and Coauthors, 2019: Evaluation of GPM-era global satellite precipitation products over multiple complex terrain regions. *Remote Sens.*, **11**, 2936, https://doi.org/10.3390/rs11242936.
- Devoli, G., I. Kleivane, M. Sund, N. K. Orthe, R. Ekker, E. Johnsen, and H. Colleuille, 2015: Landslide early warning system and web tools for real-time scenarios and for distribution of warning messages in Norway. *Landslide Processes*, G. Lollino et al., Eds., Vol. 2, *Engineering Geology for Society* and Territory, Springer, 625–629, https://doi.org/10.1007/978-3-319-09057-3_104.
- Dilley, M., R. S. Chen, U. Deichmann, A. L. Lerner-Lam, and M. Arnold, 2005: Natural Disaster Hotspots: A Global Risk Analysis. World Bank Publications, 132 pp., https://doi.org/ 10.1596/0-8213-5930-4.
- Doom, J., 2018: Landslide kills two in North Carolina. ABC News, 31 May, https://abcnews.go.com/US/landslide-kills-north-carolina/ story?id=55553200.
- Fang, H., H. K. Beaudoing, M. Rodell, W. L. Teng, and B. E. Vollmer, 2009: Global Land Data Assimilation System (GLDAS) products, services and application from NASA Hydrology Data and Information Services Center (HDISC). ASPRS 2009 Annual Conf., Baltimore, MD, American Society for Photogrammetry and Remote Sensing, 9 pp., https://www.asprs.org/a/publications/ proceedings/baltimore09/0020.pdf.
- Farahmand, A., and A. AghaKouchak, 2013: A satellite-based global landslide model. *Nat. Hazards Earth Syst. Sci.*, 13, 1259– 1267, https://doi.org/10.5194/nhess-13-1259-2013.
- Ferraro, R. R., and Coauthors, 2013: An evaluation of microwave land surface emissivities over the continental United States to benefit GPM-Era precipitation algorithms. *IEEE Trans. Geosci. Remote Sens.*, **51**, 378–398, https://doi.org/10.1109/TGRS.2012.2199121.
- Froude, M. J., and D. N. Petley, 2018: Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.*, **18**, 2161–2181, https://doi.org/10.5194/nhess-18-2161-2018.
- Gebregiorgis, A. S., and F. Hossain, 2013: Understanding the dependence of satellite rainfall uncertainty on topography and climate for hydrologic model simulation. *IEEE Trans. Geosci. Remote Sens.*, **51**, 704–718, https://doi.org/10.1109/TGRS.2012.2196282.
- —, and —, 2014: Making satellite precipitation data work for the developing world. *IEEE Geosci. Remote Sens. Mag.*, **2**, 24–36, https://doi.org/10.1109/MGRS.2014.2317561.
- Guzzetti, F., S. Peruccacci, M. Rossi, and C. P. Stark, 2007: Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteor. Atmos. Phys.*, **98**, 239–267, https://doi.org/10.1007/s00703-007-0262-7.
- Hong, Y., R. Adler, and G. Huffman, 2006: Evaluation of the potential of NASA multi-satellite precipitation analysis in global landslide hazard assessment. *Geophys. Res. Lett.*, 33, L22402, https://doi.org/10.1029/2006GL028010.
- —, and —, 2007: Use of satellite remote sensing data in the mapping of global landslide susceptibility. *Nat. Hazards*, **43**, 245–256, https://doi.org/10.1007/s11069-006-9104-z.

- Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.*, 8, 38–55, https://doi.org/10.1175/JHM560.1.
- ——, D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, and P. Xie, 2015: NASA Global Precipitation Measurement Integrated Multi-satellitE Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 4.5, 30 pp., http://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.5.pdf.
- —, —, and E. J. Nelkin, 2017: Integrated Multi-satellitE Retrievals for GPM (IMERG) technical documentation. NASA Tech. Doc., 54 pp., https://pmm.nasa.gov/sites/default/ files/document_files/IMERG_technical_doc_3_22_17.pdf.
- Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. J. Hydrometeor., 5, 487–503, https://doi.org/10.1175/ 1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.
- Kidd, C., and G. Huffman, 2011: Global precipitation measurement. *Meteor. Appl.*, **18**, 334–353, https://doi.org/10.1002/met.284.
- —, and V. Levizzani, 2011: Status of satellite precipitation retrievals. *Hydrol. Earth Syst. Sci.*, 15, 1109–1116, https://doi.org/10.5194/hess-15-1109-2011.
- —, A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum, 2017: So, how much of the Earth's surface is covered by rain gauges? *Bull. Amer. Meteor. Soc.*, 98, 69–78, https://doi.org/10.1175/BAMS-D-14-00283.1.
- Kirschbaum, D. B., and T. Stanley, 2018: Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness. *Earth's Future*, 6, 505–523, https://doi.org/10.1002/2017EF000715.
- —, R. Adler, Y. Hong, S. Hill, and A. Lerner-Lam, 2010: A global landslide catalog for hazard applications: Method, results, and limitations. *Nat. Hazards*, **52**, 561–575, https://doi.org/10.1007/s11069-009-9401-4.
- —, T. Stanley, and Y. Zhou, 2015a: Spatial and temporal analysis of a global landslide catalog. *Geomorphology*, **249**, 4–15, https://doi.org/10.1016/j.geomorph.2015.03.016.
- —, —, and J. Simmons, 2015b: A dynamic landslide hazard assessment system for Central America and Hispaniola. *Hazards Earth Syst. Sci*, 15, 2257–2272, https://doi.org/10.5194/ nhess-15-2257-2015.
- Kirstetter, P. E., N. Karbalaee, K. Hsu, and Y. Hong, 2018: Probabilistic precipitation rate estimates with space-based infrared sensors. *Quart. J. Roy. Meteor. Soc.*, 144, 191–205, https://doi.org/10.1002/qj.3243.
- Krøgli, I. K., G. Devoli, H. Colleuille, S. Boje, M. Sund, and I. K. Engen, 2018: The Norwegian forecasting and warning service for rainfall-and snowmelt-induced landslides. *Hazards Earth Syst. Sci*, 18, 1427–1450, https://doi.org/10.5194/nhess-18-1427-2018.
- Lin, Y., 2011: GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data, version 1.0. UCAR/NCAR Earth Observing Laboratory, accessed 1 February 2018, https://doi.org/10.5065/D6PG1QDD.
- Maggioni, V., R. H. Reichle, and E. N. Anagnostou, 2011: The effect of satellite rainfall error modeling on soil moisture prediction uncertainty. *J. Hydrometeor.*, 12, 413–428, https:// doi.org/10.1175/2011JHM1355.1.
- —, M. R. P. Sapiano, R. F. Adler, Y. Tian, and G. J. Huffman, 2014: An error model for uncertainty quantification in high-time-resolution precipitation products. *J. Hydrometeor.*, **15**, 1274–1292, https://doi.org/10.1175/JHM-D-13-0112.1.

- —, —, and —, 2016: Estimating uncertainties in highresolution satellite precipitation products: Systematic or random error? *J. Hydrometeor.*, 17, 1119–1129, https://doi.org/ 10.1175/JHM-D-15-0094.1.
- Mahoney, K., and Coauthors, 2016: Understanding the role of atmospheric rivers in heavy precipitation in the southeast United States. *Mon. Wea. Rev.*, **144**, 1617–1632, https://doi.org/10.1175/MWR-D-15-0279.1.
- Moore, B. J., K. M. Mahoney, E. M. Sukovich, R. Cifelli, and T. M. Hamill, 2015: Climatology and environmental characteristics of extreme precipitation events in the southeastern United States. *Mon. Wea. Rev.*, 143, 718–741, https://doi.org/10.1175/MWR-D-14-00065.1.
- Nikolopoulos, E. I., E. N. Anagnostou, F. Hossain, M. Gebremichael, and M. Borga, 2010: Understanding the scale relationships of uncertainty propagation of satellite rainfall through a distributed hydrologic model. *J. Hydrometeor.*, 11, 520–532, https://doi.org/ 10.1175/2009JHM1169.1.
- —, E. Destro, V. Maggioni, F. Marra, and M. Borga, 2017: Satellite rainfall estimates for debris flow prediction: An evaluation based on rainfall accumulation–duration thresholds. *J. Hydrometeor.*, 18, 2207–2214, https://doi.org/10.1175/ JHM-D-17-0052.1.
- Osanai, N., I. Takeshi, S. I. Kazumasa, K. I. Shinichi Kojima, and T. Noro, 2010: Japanese early-warning for debris flows and slope failures using rainfall indices with radial basis function network. *Landslides*, **7**, 325–338, https://doi.org/10.1007/s10346-010-0229-5.
- Petley, D., 2012: Global patterns of loss of life from landslides. *Geology*, 40, 927–930, https://doi.org/10.1130/G33217.1.
- —, S. A. Dunning, and N. J. Rosser, 2005: The analysis of global landslide risk through the creation of a database of worldwide landslide fatalities. *Landslide Risk Management*, O. Hungr et al., Eds., CRC Press, 367–374.
- Piciullo, L., S. L. Gariano, M. Melillo, M. T. Brunetti, S. Peruccacci, F. Guzzetti, and M. Calvello, 2017: Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. *Landslides*, 14, 995–1008, https://doi.org/10.1007/s10346-016-0750-2.
- Rossi, M., and Coauthors, 2012: SANF: National warning system for rainfall-induced landslides in Italy. *Landslides and Engineered Slopes: Protecting Society through Improved Understanding*, E. Eberhardt et al., Eds., CRC Press, 1895–1899, https://doi.org/10.13140/2.1.4857.9527.
- Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. J. Climate, 19, 3483–3517, https://doi.org/10.1175/JCLI3812.1.
- Sassa, K., P. Canuti, and Y. Yin, Eds., 2014: Methods of Landslide Studies. Vol. 2, Landslide Science for a Safer Geoenvironment, Springer, 851 pp., https://doi.org/10.1007/978-3-319-05050-8.
- Scheuerer, M., and T. M. Hamill, 2015: Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. *Mon. Wea. Rev.*, 143, 4578–4596, https:// doi.org/10.1175/MWR-D-15-0061.1.
- Shige, S., S. Kida, H. Ashiwake, T. Kubota, and K. Aonashi, 2013: Improvement of TMI rain retrievals in mountainous areas. J. Appl. Meteor. Climatol., 52, 242–254, https://doi.org/10.1175/ JAMC-D-12-074.1.
- Sorooshian, S., K. L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046, https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2.

- Stanley, T., and D. B. Kirschbaum, 2017: A heuristic approach to global landslide susceptibility mapping. *Nat. Hazards*, 87, 145– 164, https://doi.org/10.1007/s11069-017-2757-y.
- Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K. L. Hsu, 2018: A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.*, 56, 79–107, https://doi.org/10.1002/2017RG000574.
- Tan, J., W. Petersen, and A. Tokay, 2016: A novel approach to identify sources of errors in IMERG for GPM ground validation. J. Hydrometeor., 17, 2477–2491, https://doi.org/ 10.1175/JHM-D-16-0079.1.
- Tang, L., and F. Hossain, 2009: Transfer of satellite rainfall error from gaged to ungaged locations: How realistic will it be for the global precipitation mission? *Geophys. Res. Lett.*, 36, L10405, https://doi.org/10.1029/2009GL037965.
- —, and —, 2012: Investigating the similarity of satellite rainfall error metrics as a function of Köppen climate classification. *Atmos. Res.*, **104-105**, 182–192, https://doi.org/ 10.1016/j.atmosres.2011.10.006.
- Tapiador, F. J., and Coauthors, 2012: Global precipitation measurement: Methods, datasets and applications. *Atmos. Res.*, 104–105, 70–97, https://doi.org/10.1016/j.atmosres.2011.10.021.
- Tian, Y., and C. D. Peters-Lidard, 2010: A global map of uncertainties in satellite-based precipitation measurements. *Geophys. Res. Lett.*, 37, L24407, https://doi.org/10.1029/2010GL046008.
- —, and Coauthors, 2009: Component analysis of errors in satellite-based precipitation estimates. J. Geophys. Res., 114, D24101, https://doi.org/10.1029/2009JD011949.

- —, G. J. Huffman, R. F. Adler, L. Tang, M. Sapiano, V. Maggioni, and H. Wu, 2013: Modeling errors in daily precipitation measurements: Additive or multiplicative? *Geophys. Res. Lett.*, 40, 2060–2065, https://doi.org/10.1002/grl.50320.
- White, J. A., and D. I. Singham, 2012: Slope stability assessment using stochastic rainfall simulation. *Procedia Comput. Sci.*, **9**, 699–706, https://doi.org/10.1016/j.procs.2012.04.075.
- Wooten, R. M., A. C. Witt, C. F. Miniat, T. C. Hales, and J. L. Aldred, 2016: Frequency and magnitude of selected historical landslide events in the southern appalachian highlands of North Carolina and Virginia: Relationships to rainfall, geological and ecohydrological controls, and effects. *Natural Disturbances and Historic Range of Variation*, C. Greenberg and B. Collins, Eds., Managing Forest Ecosystems Series, Vol 32, Springer, 203–262, https://doi.org/10.1007/978-3-319-21527-3_9.
- Wright, D. B., 2018: Rainfall information for global flood modeling. Global Flood Hazard: Applications in Modeling, Mapping, and Forecasting, G. Schumann et al., Eds., John Wiley & Sons, 19–42.
- —, D. B. Kirschbaum, and S. Yatheendradas, 2017: Satellite precipitation characterization, error modeling, and error correction using censored shifted gamma distributions. J. Hydrometeor., 18, 2801–2815, https://doi.org/10.1175/JHM-D-17-0060.1.
- Wu, H., R. F. Adler, Y. Hong, Y. Tian, and F. Policelli, 2012: Evaluation of global flood detection using satellite-based rainfall and a hydrologic model. *J. Hydrometeor.*, 13, 1268– 1284, https://doi.org/10.1175/JHM-D-11-087.1.