AGU100 ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR025577

Assessing the Feasibility of Satellite-Based Thresholds for Hydrologically Driven Landsliding

Matthew A. Thomas[1], **Brian D. Collins**[2], **and Benjamin B. Mirus**[1]

[1]U.S. Geological Survey Geologic Hazards Science Center, Golden, CO, USA, [2]U.S. Geological Survey Geology, Minerals, Energy, and Geophysics Science Center, Menlo Park, CA, USA

Key Points:
- We compared hydrometeorological thresholds for landslide initiation defined with ground- versus satellite-based information
- Ground-based thresholds show superior performance because in situ soil moisture data better reflect gravity-dominated subsurface flow
- Satellite-based thresholds produce more false alarms but may merit consideration where in situ soil moisture data are nonexistent

Supporting Information:
- Supporting Information S1.
- Figure S1
- Figure S2
- Figure S3
- Figure S4
- Figure S5
- Figure S6
- Figure S7
- Figure S8
- Figure S9
- Figure S10
- Figure S11
- Figure S12
- Figure S13
- Figure S14

Correspondence to:
M. A. Thomas,
matthewthomas@usgs.gov

Citation:
Thomas, M. A., Collins, B. D., & Mirus, B. B. (2019). Assessing the feasibility of satellite-based thresholds for hydrologically driven landsliding. *Water Resources Research*, *55*, 9006–9023. https://doi.org/10.1029/2019WR025577

Received 15 MAY 2019
Accepted 26 SEP 2019
Published online 13 NOV 2019

Abstract Elevated soil moisture and heavy precipitation contribute to landslides worldwide. These environmental variables are now being resolved with satellites at spatiotemporal scales that could offer new perspectives on the development of landslide warning systems. However, the application of these data to hydrometeorological thresholds (which account for antecedent soil moisture and rainfall) first needs to be evaluated with respect to proven, direct measurement-based thresholds that use rain gages and in situ soil moisture sensors. Here we compare ground-based hydrologic data to overlapping satellite-based data before, during, and after a recent season of widespread shallow landsliding in the San Francisco Bay Area (California, USA). We then explore how the remotely sensed information could be used to empirically define hypothetical thresholds for shallow landsliding. We find that the ground-based thresholds developed with a single monitoring station show superior performance because the in situ soil saturation data better reflect the gravity-dominated subsurface flow conditions that are characteristic of hillslopes during the rainy season. Although the satellite-based thresholds can identify most of the landslide days, they include a greater number of false alarms due to overestimates of soil moisture between major storm events. To avoid the type of false alarms that are characteristic of our satellite-based thresholds, further postprocessing of the near-surface hydrologic response data should be integrated into satellite-based model outputs to better reflect gravity-dominated drainage. Our results encourage further deployment of ground stations in landslide-prone terrain and cautious exploration of satellite-based hydrometeorological thresholds where in situ networks are nonexistent.

Plain Language Summary Soil wetness and rainfall contribute to landslides across the world. Using soil moisture sensors and rain gages, these environmental conditions have been monitored at numerous points across the Earth's surface to define threshold conditions, above which landsliding should be expected for a localized area. Satellite-based technologies also deliver estimates of soil wetness and rainfall, potentially offering an approach to develop thresholds as part of landslide warning systems over larger spatial scales. To evaluate the potential for using satellite-based measurements for landslide warning, we compare the accuracy of landslide thresholds defined with ground- versus satellite-based soil wetness and rainfall information. We find that the satellite-based data overpredict soil wetness during the time of year when landslides are most likely to occur, resulting in thresholds that also overpredict the potential for landslides relative to thresholds informed by direct measurements on the ground. Our results encourage the installation of more ground-based monitoring stations in landslide-prone settings and the cautious use of satellite-based data when more direct measurements are not available.

1. Introduction

The potential for satellite-based rainfall data to inform global landslide hazard assessment has been recognized for more than a decade (Hong et al., 2006). Despite the multitude of studies focused on comparisons of ground- and satellite-based rainfall for locations throughout the world, few have evaluated criteria important to landslide initiation potential (e.g., Brunetti et al., 2018; Kirschbaum et al., 2009, 2012; Rossi et al., 2017). This is owed, in part, to the relatively coarse resolution of satellite-based rainfall products. More recent satellite-based rainfall missions, including the National Aeronautics and Space Administration (NASA) Global Precipitation Measurement (GPM), now provide rainfall information at higher spatiotemporal resolutions (~9-km grid cell size, 30-min time steps; 6-hr minimum latency) that are increasingly relevant for landslide science (Hou et al., 2014; Kirschbaum et al., 2017). These advancements facilitated the first

operational system for satellite-based assessment of rainfall-triggered landslides across the globe (Kirschbaum & Stanley, 2018), which relies on a 7-day antecedent rainfall index, updated every 3 hr, to evaluate hazard potential.

Whereas the incorporation of satellite-based rainfall into hazard assessment tools shows promise, rainfall is an indirect proxy for the subsurface hydrologic state variables (i.e., soil moisture and pore-water pressure) that cause slope failure (Terzhagi, 1943). The hydrometeorological framework for identifying thresholds for landslide initiation, which favors the incorporation of the hydrologic processes that cause and trigger slope failure (e.g., Bogaard & Greco, 2018), has been shown to improve tools that are traditionally developed with only rainfall-based criteria (e.g., Mirus, Becker, et al., 2018; Segoni et al., 2018). However, when we extend our view of thresholds to include those that also account for subsurface hydrology, the potential for data availability issues grows considerably. Despite the relatively coarse spatial resolution of satellite-based soil moisture data products, studies have demonstrated their potential to add value to site- (Brocca et al., 2012; Ray et al., 2010; Ray & Jacobs, 2007) and regional-scale (Brocca et al., 2016) landslide studies. Newer data sources, such as the NASA Soil Moisture Active Passive (SMAP) mission, now deliver estimates of soil moisture at spatiotemporal scales (~9-km grid cell size, 3-hr time steps, up to 1-m depth using remote sensing data and a land surface model; 3-day minimum latency; Reichle, De Lannoy, Liu, Ardizzone, et al., 2017,Reichle, De Lannoy, Liu, Koster, et al., 2017) that are meaningful for hazard science. These data have already proven useful for improving predictions of flooding (Crow et al., 2017, 2018) and evaluating regional patterns in landslide concentration (Bessette-Kirton et al., 2019).

Satellite-based rainfall and soil moisture products need to be evaluated carefully in settings where landslides actually occur to test their potential suitability for applications in warning systems. Furthermore, given the disparity between the availability of ground-based rainfall versus soil moisture information, it is paramount to understand the trade-offs between developing spatially variable hydrometeorological thresholds with accurate (but less widespread) ground-based data versus globally comprehensive (but likely less accurate) satellite-based information. In our study, we explore (1) how satellite-based rainfall and soil saturation estimates compare to in situ measurements at spatiotemporal scales relevant to landslide forecasting and (2) how that information might be used to inform spatially variable hydrometeorological thresholds. The study area where we apply these questions is the ~1,000-km^2 East Bay Hills of the San Francisco (SF) Bay Area in California (USA; Figures 1a and 1b), which is an extensively studied landslide-prone region (e.g., Cannon & Ellen, 1985; Coe & Godt, 2001; Collins et al., 2012; Keefer et al., 1987; Nilsen & Turner, 1975; Pike & Sobieszczyk, 2008; Taylor & Brabb, 1972; Wieczorek et al., 1988; Wilson & Jayko, 1997) that hosts overlapping ground- and satellite-based rainfall and soil moisture data coverage for areas of recent, widespread landsliding.

2. Study Region

The East Bay Hills comprise a diverse assemblage of lithologies, including metamorphic, sedimentary, and volcanic rocks that record the area's transition from a deep-marine subduction margin to shallow-marine and terrestrial environments associated with the evolution of the San Andreas fault system. Local components of compression within a network of active transform faults are uplifting the region (Sloan, 2006). Whereas total relief in the region is low (667 m), the predominantly grass- and shrub-dominated landscape (Figure 1c) exhibits steep topography (mean = 17° ± 10° excluding ~20% of the land surface that is characterized by low-lying alluvial deposits, reservoirs, and other areas ≤1°; U.S. Geological Survey [USGS] 2018). This terrain is characteristic of landslide-prone settings and includes many areas with slopes that are ≥20°, which account for ~30% of the study region. A Mediterranean climate delivers ~90% of the rainfall between 1 November and 30 April, with the most powerful rainstorms typically arriving in January and February (PRISM 2018).

The SF Bay Area is subject to high levels of landslide hazard and risk (Crovelli & Coe, 2009). Major region-impacting storm events occur on the decadal time scale (e.g., Ellen & Wieczorek, 1988; Godt, 1999), and individual storms are capable of destroying homes in any given year (e.g., Collins & Corbett 2019). During a sequence of rainstorms in January and February 2017, approximately 9,000 landslides initiated in the East Bay Hills region (Collins et al., 2018). Most of the slope failures were shallow translational slides (Casadei et al., 2003), but many transitioned into debris flows. Based on mapping of high-resolution satellite

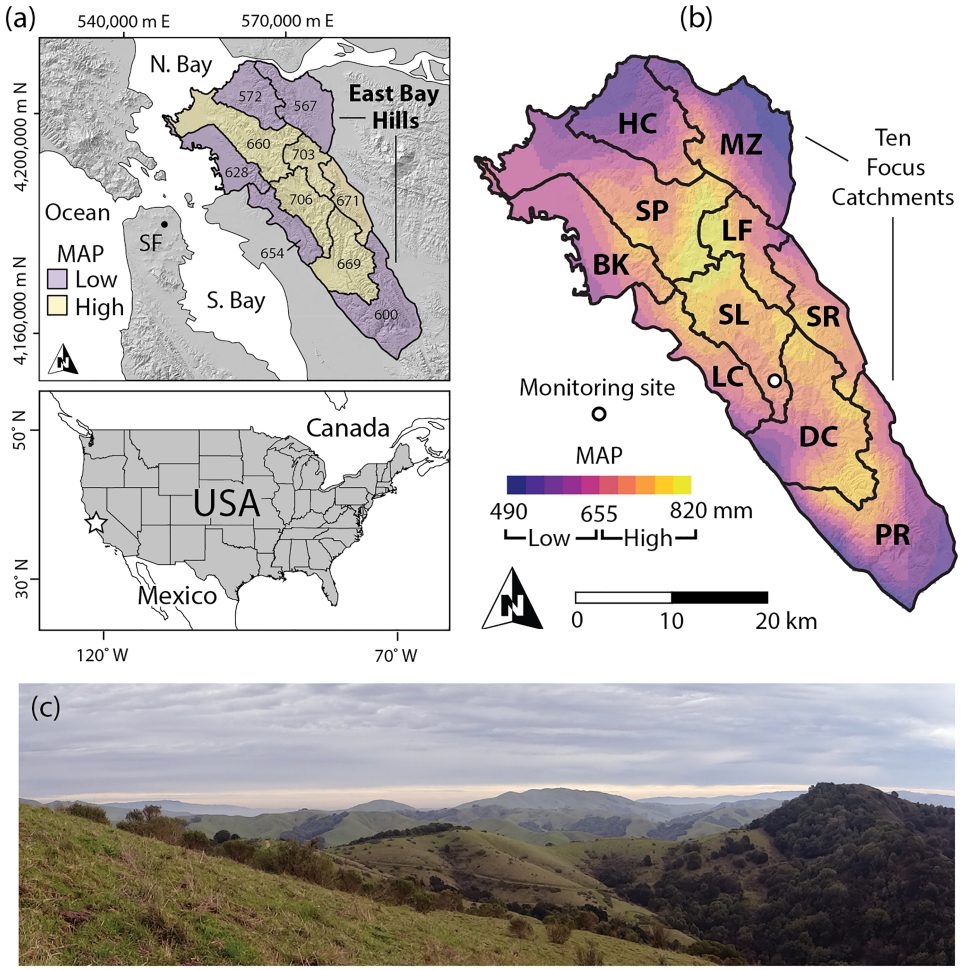

Figure 1. (a) Location of the East Bay Hills (EBH) region of the San Francisco (SF) Bay Area in California, USA. Yellow and purple polygons respectively refer to the 10 focus catchments with high and low mean annual precipitation (MAP; reported in millimeters). Coordinates in degrees and meters correspond to the WGS84 and NAD83 UTM (Zone 10) projections, respectively. (b) Topographic hillshade (USGS 2018) for the 10 focus catchments with an overlay of MAP at 800-m horizontal resolution (PRISM 2018). Location of the BALT1 monitoring site shown with white circle. Catchment names coincide with abbreviations listed in Table 1. (c) Southeast looking view of the EBH from BALT1.

imagery, the overall landslide concentration for the East Bay Hills region was ~8.5 landslides/km^2. Landslides occurred in all of the focus catchments of our study area (Figure 1b) with catchments on the west side (BK and LC) of the topographic high of the hills (running northwest to southeast along catchments SP, SL, DC, and PR) having a lower total number of landslides (i.e., tens of landslides in BK and LC versus hundreds of landslides in each of the other regions), likely as a result of the built-up development on the more urbanized (SF Bay) side of the topographic divide. Landslides generally occurred on slopes ranging from 15° to 35° and throughout all directional aspects. Although many of these landslides occurred in undeveloped open spaces, ~15% affected structures and infrastructure, including homes and roads. Unlike other widespread landsliding events in the SF Bay Area that occurred during El Niño events, the 2016–2017 landsliding occurred during a weak La Niña year with above-average rainfall exceeding one standard deviation of the mean.

Elevated soil moisture and positive pore-water pressure development are recognized as key factors for the initiation of shallow landslides in the SF Bay Area (Keefer et al., 1987), where failure surfaces are often at (or near) the soil-weathered bedrock interface (Coe & Godt, 2001). Water Year (WY; defined as beginning 1 October and ending 30 September) 2016–2017 is an important benchmark for landslide warning and threshold research in the East Bay Hills (e.g., Cannon, 1988; Cannon & Ellen, 1985; Keefer et al., 1987;

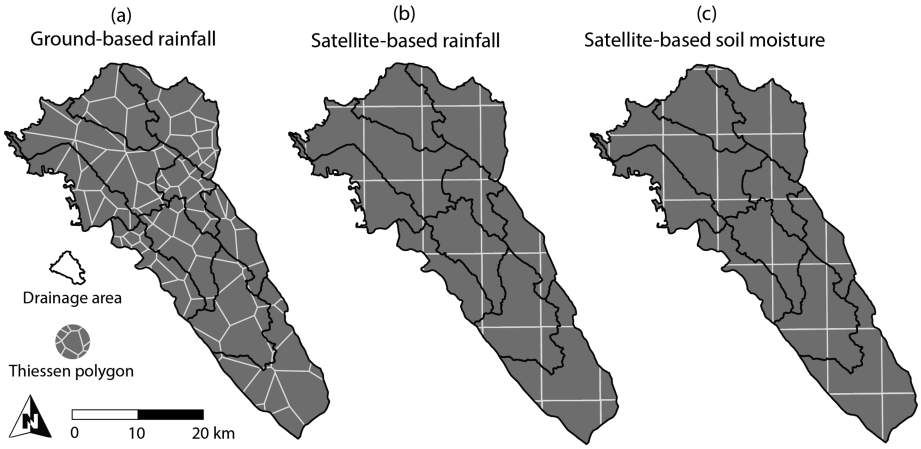

Figure 2. Thiessen polygons intersected with the 10 focus catchments for the (a) ground-based rainfall, (b) satellite-based rainfall, and (c) satellite-based soil moisture data sets. In (a), each ground-based polygon center corresponds to a rain gage. In (b) and (c), the satellite-based polygons appear grid-like because the underlying data product is distributed in a quasi-rectangular geographic projection.

Thomas, Mirus, & Collins, 2018; Thomas, Mirus, Collins, Lu, et al., 2018; Wilson, 2004; Wilson & Jayko, 1997; Wilson et al., 1993) because a monitoring station (Collins et al., 2012) recorded subsurface soil saturation and pore-water pressure dynamics before, during, and after a season of widespread shallow landsliding.

3. Methods

3.1. Ground- and Satellite-Based Data Sets

We compiled ground-based rainfall information for WY 2016–2017 across 10 focus catchments (Figure 1b) from the Alameda County Public Works Agency (F. Codd, personal communication, May 2018), Alameda County Water District (J. Gekov, personal communication, May 2018), California Data Exchange Center (2018), and Weather Underground Personal Weather Station Network (Wunderground, 2018). Wunderground is a commercial weather service that hosts a network of personal weather stations that undergo quality control processes, including checks for the data falling outside a normal range, changing at an abnormal rate, and whether or not the observations agree with nearby stations. The hourly (or less) tipping-bucket rain gages that we used are all within 5 km of the East Bay Hills and provide data for at least 90% of the WY, 95% of the rainy season (i.e., November through April), and 100% of four probable multiday landslide periods that occurred in January and February 2017. The data set contains 96 rain gage records (Table S1) with an average map-view spacing of 2.25 km (Figure 2a).

Ground-based volumetric soil-water content data were collected by a shallow landslide monitoring station (termed "BALT1") that is operated by the USGS (Collins et al., 2012) and is located within the San Leandro (SL) catchment (Figure 1b). The unchanneled topographic hollow (e.g., Sidle et al., 2018) where the instrumentation is sited is characteristic of many that experience shallow landslides during winter rainstorms. Two vertical arrays record subsurface volumetric soil-water content (25- and 70-cm depth; 30- and 140-cm depth) at 10-min intervals. A tipping-bucket rain gage is also installed at the site. We calculated the average soil moisture for each array and then used the two resultant time series to calculate an overall average. We converted the volumetric soil-water content time series to soil saturation based on observed porosity values for the two sensor arrays (i.e., 0.43 and 0.34; Thomas et al., 2017).

We acquired WY 2016–2017 satellite-based rainfall and volumetric soil-water content data from the NASA GPM-based IMERG (Integrated Multi-Satellite Retrievals; Huffman et al., 2018) and SMAP L4_SM (Reichle, De Lannoy, et al., 2018) repositories. The IMERG algorithm intercalibrates, merges, and interpolates passive microwave sensor data, infrared imagery, and conventional rain gage data. The GPM mission has produced 30-min rainfall estimates at ~9-km grid cell resolution since 2014. The L4_SM

algorithm assimilates remotely sensed surface soil moisture estimates based on 36-km brightness temperatures from an L-band radiometer (Chan et al., 2014) into the Goddard Earth Observing System, Version 5 (GEOS-5), Land Data Assimilation System (see De Lannoy & Reichle, 2016a, 2016b; Reichle et al., 2014), including the Catchment land surface model (Ducharne et al., 2000; Koster et al., 2000). The Catchment land surface model solves a water and energy balance forced with conventional and satellite-based meteorological observations including precipitation, downward shortwave/longwave radiation, wind speed, near-surface air temperature, specific humidity, and air pressure. The modeling process transfers water between the surface (0–5 cm), the "root zone" (0–100 cm), and the water table for a network of catchment boundaries (Ducharne et al., 2000; Koster et al., 2000). An analytical distribution of topographic indices and an empirical function is used to define the initial configuration of the local water table. The distribution of soil moisture above the water table is set as an equilibrium profile whose shape is calculated based on soil-hydraulic properties (De Lannoy et al., 2014). The rise and fall of the water table is controlled by "root zone excess" (i.e., the amount of water, per unit area, by which the root zone is out of equilibrium; Koster et al., 2000). The root zone excess term considers surface runoff, infiltration, and evapotranspiration when recalculating the position of the water table for each time step (Reichle, De Lannoy, Liu, Ardizzone, et al., 2017,Reichle, De Lannoy, Liu, Koster, et al., 2017).

The SMAP mission has produced 3-hr estimates of volumetric water content at the surface (0–5 cm) and modeled values for the near surface (0–100 cm) at ~9-km resolution since 2015 (Reichle, De Lannoy, Liu, Ardizzone, et al., 2017,Reichle, De Lannoy, Liu, Koster, et al., 2017). We refer to 100-cm SMAP soil moisture data as "satellite-based" because it is a modeled product that assimilates global 0- to 5-cm surface brightness observations, as well as other space-borne remote sensing observations including land cover, topography, and vegetation height (Reichle, Liu, et al., 2018). The soil depth for the BALT1 shallow landslide monitoring station is ~110 cm, which is consistent with our field mapping of recent landsliding in the area (Collins et al., 2018). Therefore, we utilized the 100-cm satellite-based soil moisture estimates (Reichle, De Lannoy, et al., 2018) for our analyses. To provide a normalized comparison of soil wetness, we transformed the volumetric water content information to soil saturation based on the porosity values used in the Catchment land surface model.

We converted and resampled the ground- and satellite-based rainfall and soil saturation information to the minimum common temporal resolution (i.e., 3-hr intervals, local time). The various spatial resolutions of the ground-based rainfall, satellite-based rainfall, and satellite-based soil saturation data are divided among Thiessen (1911) polygons (Figure 2), a straightforward approach wherein the percent contributing area of each polygon within a catchment is used to estimate mean areal rainfall and soil saturation in that catchment.

The spatiotemporal resolution of the ground- and satellite-based data facilitated a regional assessment of landslide-relevant hydrologic characteristics across climatological- and forecast-relevant time scales. We first evaluated basic trends in annual rainfall, including the percent difference between each catchment. The seasonal comparisons, for areas of high ($\geq$660 mm) and low (<660 mm) mean annual precipitation (MAP; see Figures 1a and 1b; PRISM 2018), include multimonth periods before (October–December), during (January–February), and after (March–September) widespread rainfall-induced landsliding. We aimed to evaluate our data at a temporal scale that is consistent with landslide warning systems that have been deployed in the United States (Baum & Godt, 2010). Therefore, for the hypothetical landslide thresholds that we developed for this study, we also tested the degree to which rainy-day (i.e., >0.254 mm, 0.01 in.) accumulations for the ground- and satellite-based data sets are correlated. The simple coefficient of determination (r^2) tests are binned by MAP and season.

Although spatially variable estimates of soil moisture for the East Bay Hills are available from the SMAP mission, we have a single ground station (BALT1) that monitors in situ conditions that are relevant to rainfall-induced shallow landsliding. Therefore, we focused our most direct comparison of ground- and satellite-based soil saturation on the SL catchment (Figure 1b) and later used the station measurements as a proxy for threshold comparisons across the nine remaining catchments. We evaluated the WY 2016–2017 soil saturation time series for the SL catchment with r^2 and root mean square error tests binned at the seasonal level. For comparison of our results to a WY in which no extensive landsliding occurred in our study region (i.e., WY 2017–2018), we also examined satellite-based soil saturation from these same data sources (i.e., the BALT1 site and NASA SMAP data; see the supporting information). Despite only having one ground-based site for evaluation, our experience with monitoring in the region indicates that our instrumentation captures

signals that are applicable to identifying conditions required for landsliding over the broad landscape of the East Bay Hills (Collins et al., 2012, 2018).

3.2. Objective Definition of Hydrometeorological Thresholds

In addition to comparing landslide-relevant hydrologic characteristics gleaned from satellite-based data to those obtained from in situ sensors, we also assessed how this information can affect landslide thresholds that are developed for use in warning systems. For this study, we compare hypothetical empirical thresholds defined by antecedent soil saturation (s_0 [%]) on the x-axis and 24-hr rainfall (r_{24} [mm]) on the y-axis of a two-dimensional threshold evaluation space. We selected these metrics to reflect a hydrometeorological threshold space (e.g., Bogaard & Greco, 2018) that includes both "cause" (i.e., soil saturation conditions that predispose the hillslope to failure) and "trigger" (i.e., the rainfall associated with the timing of landslide initiation) components. We focus on a 24-hr accumulation level to place our results within the context of practical rainfall forecasting constraints in the United States (Novak et al., 2014). These thresholds are based on a sequence of storms in WY 2016–2017, as opposed to longer (multiyear) observational records, and, therefore, should not be considered validated landslide thresholds for the East Bay Hills. However, the definition of our ground- and satellite-based thresholds are objective in that they are developed with the same criteria, providing a comparison for storms where overlapping ground- and satellite-based data coverage exists for a period of widespread landsliding in the region.

Our most straightforward comparison of landslide thresholds is for the SL catchment, which includes both ground- and satellite-based observations. We used these data to assemble s_0 and r_{24} data pairs throughout WY 2016–2017. Observations available to us suggest that widespread landsliding likely occurred over the course of four storms in January and February 2017 (Collins et al., 2018). Elevated positive pore-water pressures were also observed at BALT1 during these periods. Whereas we know (from media reports and onsite observations made by the authors) that landsliding occurring during several of these storms, the exact timing of widespread landsliding in each catchment is unknown. It is highly likely that landsliding occurred in each catchment during multiple storms, and for this study, we assume landsliding was coincident with the most intense rainfall during each of these four storm periods. Therefore, we designated the maximum r_{24} (and corresponding s_0) as "landslide days" during the four multiday periods in January and February 2017. We assigned the rainy-day s_0 and r_{24} pairs outside of these periods, which we evaluated at 12-hr increments, as "null cases." The landslide days and null cases distinguish between periods when landsliding is likely and unlikely, respectively.

We then used receiver-operator characteristic analysis (Fawcett, 2006; Swets, 1988) and two skill statistics (threat score, TS [Equation (1)], and radial distance, RD [Equation (2)]) to quantitatively evaluate landslide thresholds across the ground- and satellite-based inputs. At the core of this kind of analysis is a "confusion matrix" that is calculated for a given threshold. The confusion matrix reflects the number of true positives (TP; correctly predicted landslide days), false positives (FP; incorrectly predicted landslide days), true negatives (TN; correctly predicted null cases), and false negatives (FN; incorrectly predicted null cases). We followed prior precedent (Mirus, Morphew, et al., 2018) and selected a simple bilinear (as opposed to a linear or more complex power or high-degree polynomial) functional form because the landslide events tend to group in the upper-right corner of the threshold space. For this study, we optimized the TS and RD for 1×10^6 combinations of s_0 and r_{24} (i.e., 0% to 100% saturation at 0.1% levels and 0- to 100-mm rainfall at 0.1-mm levels, respectively) in the threshold domain. We refer to the optimal s_0 and r_{24} pair, which defines the bottom left-hand corner of the bilinear threshold, as the critical s_0 and r_{24}. The TS is defined as follows:

$$\mathrm{TS} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \tag{1}$$

where TP (−), FN (−), and FP (−) are the numbers of true positives, false negatives, and false positives, respectively. A TS ranges from 0 to 1, where 1 is a perfect score. The RD is defined as follows:

$$\mathrm{RD} = \sqrt{\left(\frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}\right)^2 + \left(\left(\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}\right) - 1\right)^2} \tag{2}$$

where TN (−) is the number of true negatives. The RD ranges from 0 to 1, where 0 is a perfect score. The TS results in a "pessimistic" threshold that is averse to false results, whereas the RD results in an "optimistic" or

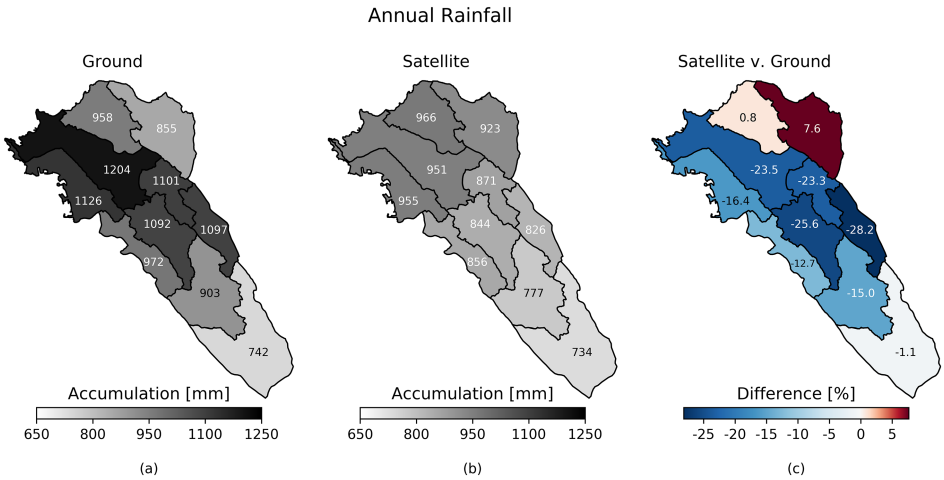

Figure 3. Cumulative rainfall for Water Year 2016–2017 across the 10 focus catchments for the (a) ground- and (b) satellite-based data sets. (c) Percent difference between satellite- and ground-based annual rainfall, with cooler and warmer colors respectively indicating underprediction and overprediction by the satellite-based product.

more balanced threshold (Mirus, Morphew, et al., 2018; Postance et al., 2018; Staley et al., 2013). We optimized four empirical bilinear landslide thresholds with ground- and satellite-based data for the SL catchment to evaluate how the data input types affect the threshold definition.

Using the in situ soil saturation monitoring station (i.e., within the SL catchment; Figure 1b), which is likely correlative with other adjacent catchments in the region (i.e., LC and DC), we formulated threshold comparisons for catchments across the East Bay Hills. Whereas we would prefer one (or more) monitoring stations recording in situ near-surface hydrologic response in landslide-prone areas of each catchment, these data types are currently only collected for a handful of sites across the country (e.g., USGS 2019). The BALT1 monitoring station was sited with the intention of being representative of the general regional, topographic, geologic, hydrologic, and vegetative characteristics of areas in in the East Bay Hills that are prone to shallow landsliding (Collins et al., 2012). Specifically, the BALT1 site was chosen because it (1) showed evidence of landsliding in at least two previous major storm events, (2) is a colluvial-filled hollow with at least a 30° slope, (3) has coarse-grained soil that commonly mobilizes into debris flows, (4) has a soil profile depth of ~1 m, (5) is dominated by grassland vegetative cover, and (6) does not show signs of anthropogenic disturbance other than cattle grazing, which is widespread in landslide-prone areas of the East Bay Hills. These are characteristics representative of the 2017 storm-induced landslides in all of our study area catchments. Therefore, we extended the site-specific (BALT1) soil saturation observations to the nine remaining catchments and paired them with more readily available catchment-specific rain gage data (Figure 2a). With this approach, we test an information trade-off between ground-based antecedent soil saturation values from a single site and the more spatially comprehensive satellite-based information. In total, we optimized 40 catchment-specific thresholds (after building a landslide day/null case data set for each catchment) and defined four East Bay Hills average regional thresholds.

4. Results

4.1. Spatiotemporal Characteristics of Ground- and Satellite-Based Data Sets

Comparisons of annual rainfall across the 10 focus catchments (Figure 3), which we calculated by area-averaging of data within each catchment, indicate general similarities but also reveal systematic differences between the ground- and satellite-based information. Both data sets show higher annual rainfall in the northern catchments, with the lowest accumulations in the southern catchments. However, the satellite-based totals (Figure 3b) report a narrower range (734–966 mm) compared to the ground-based totals (742–1,204 mm; Figure 3a). These ranges drive catchment-level differences of up to 28% (Figure 3c). The satellite-based product generally underpredicts the annual totals, particularly in areas of high MAP (Figures 1a and 1b; PRISM 2018).

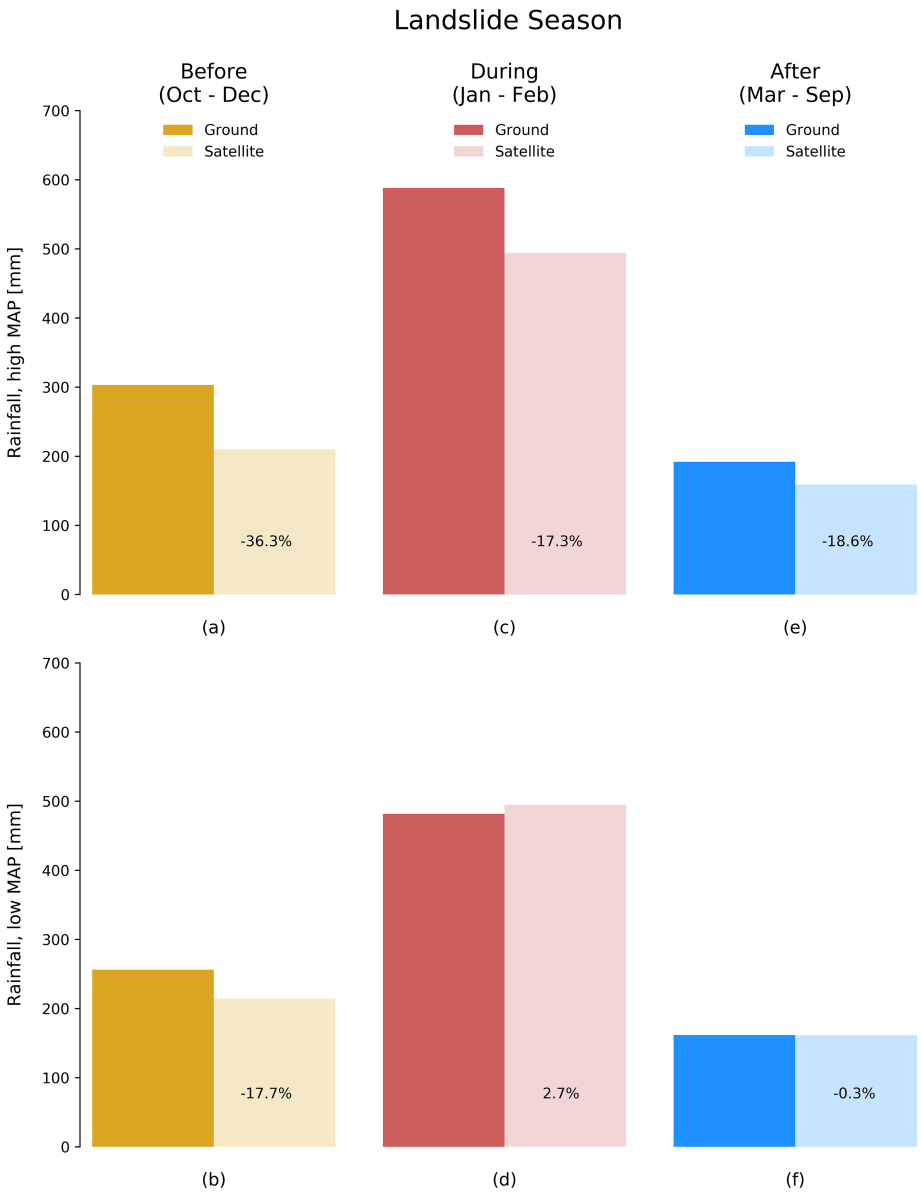

Figure 4. Ground- versus satellite-based rainfall, averaged for catchments for (top) high and (bottom) low mean annual precipitation (MAP; see Figures 1a and 1b), (a, b) before, (c, d) during, and (e, f) after the landslide season. Percentages indicate the difference between the satellite- and ground-based rainfall accumulations.

Despite the narrow 2-month (i.e., January–February) window for the landslide season, storms during this time typically produce at least one third of the total annual rainfall (PRISM 2018). The ground- and satellite-based information captures this temporal pattern, with landslide season totals accounting for ~50% of the accumulation in WY 2016–2017 (Figure 4). Similar to the annual totals, satellite-based seasonal accumulations are generally underpredicted in areas of high MAP. The seasonal totals indicate that the greatest discrepancies between the ground- and satellite-based sources (up to 36% underprediction; Figure 4a) occurred before the landslide season began.

Rainy-day accumulations binned across MAP zones and seasons for the ground- and satellite-based products show a notable degree of scatter (Figure 5). Many of the points lie along the *x*-axis (i.e., ground-based rainfall) between 0 and 10 mm, indicating that smaller rainfall accumulations are consistently underpredicted by the satellite-based data. Three of the four rainiest days during the landslide season are also underpredicted by the

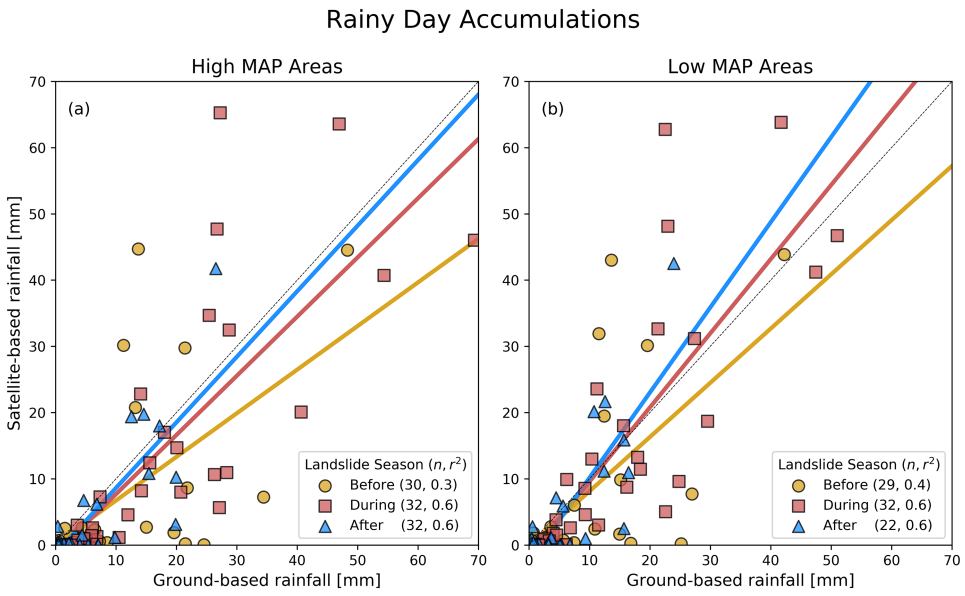

Figure 5. Rainy-day accumulations for ground- versus satellite-based data sets before, during, and after the landslide season, averaged for areas of (a) high and (b) low mean annual precipitation (MAP; see Figures 1a and 1b), with sample size (n) and the coefficient of determination (r^2) for a linear regression provided in parentheses. Dashed line indicates 1:1 correspondence (i.e., perfect fit).

satellite data (up to 68% and 47% for the high and low MAP areas, respectively). The poorest rainy-day r^2 values (0.3–0.4) occur before the landslide season begins, a trend identical to the seasonal percent differences. During and after the landslide season, the rainy-day correlations show moderate improvement ($r^2 = 0.6$).

There is strong temporal correspondence ($r^2 = 0.8$) between the ground- and satellite-based soil saturation records at the annual time scale as revealed by the WY 2016–2017 data (Figure 6a). The periods of wetting and drying (i.e., the rising and falling limbs) in these time series are well aligned, although the magnitude of change is often different. Before the landslide season (Figure 6b), soil saturation values range from ~30% to 70%. A mid-October 2016 storm induced a spike in soil saturation at the monitoring site during this time (Figure 6a). The satellite-based peak in mid-October is roughly one half of that from the in situ monitoring data. Despite this difference, by the onset of the landslide season (Figure 6c), the soil saturation values for both data sets are nearly identical. During the landslide season, soil saturation ranges from ~60% to 90%. Four storms generate sharp soil saturation peaks in the ground-based record. Each time, the ground-based soil saturation appears to return (or nearly return) to its prestorm level within a few days. In contrast, the satellite-based data show four lower amplitude and asymmetric pulses with very little desaturation. Instead, the overall soil saturation increases throughout the 2-month period, resulting in the largest difference between the two data sets. After the landslide season (Figure 6d), soil saturation values range from ~40% to 70%. The impacts of a few rainstorms are visible in both records, but a drying trend dominates the response in both. At the end of the WY, the ground- and satellite-based soil saturation values nearly match. Comparisons of the ground- and satellite-based data across the three seasons results in r^2 values of 0.8, 0.3, and 1.0, respectively (Figures 6b–6d). Unlike the rainfall data, the poorest soil saturation correlation (and highest root mean square error, 12.9%) occurs during, rather than before, the landslide season. Although correlations are good before and after the landslide season, the very low correlation during the landslide season is reflective of an overall increasing soil moisture trend in the SMAP data.

4.2. Spatially Variable Hydrometeorological Thresholds

The most direct comparison of hydrometeorological thresholds that we developed with ground- versus satellite-based information is for the SL catchment because this is where our subsurface monitoring instrumentation is located (Figure 1). The overall pattern of the s_0 (antecedent soil saturation) and r_{24} (24-hr rainfall) pairs for the two threshold spaces (ground- and satellite-based) is similar for the TS metric (Figures 7a

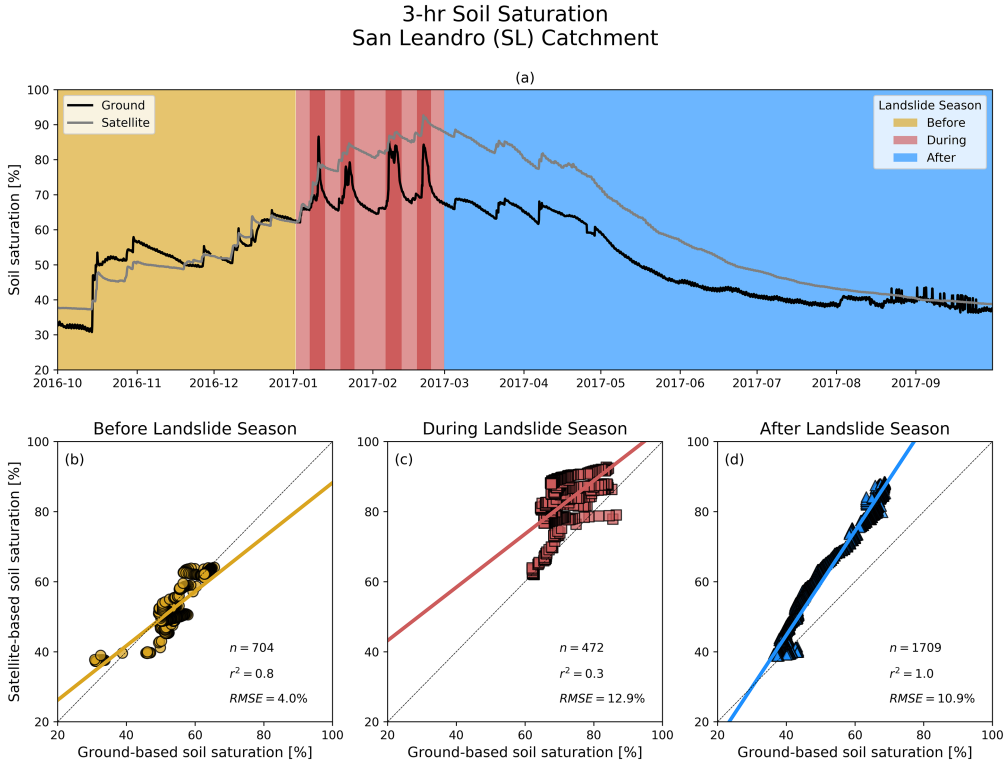

Figure 6. (a) Ground- versus satellite-based 3-hr soil saturation for the San Leandro (SL; see Figure 1b) catchment (b) before, (c) during, and (d) after the landslide season. The four dark red vertical bars in (a) correspond to the most likely time windows for which landsliding was observed. Symbols n, r^2, and *RMSE* respectively indicate the sample size, coefficient of determination for a linear regression, and root mean square error. Dashed line indicates 1:1 correspondence (i.e., perfect fit).

and 7b) and RD metric (Figures 7c and 7d). Both data sets exhibit low to intermediate s_0 and r_{24} values before the landslide season, recording the hydrologic "wetting-up" for the upcoming landslide season. During the landslide season, the s_0 and r_{24} values are at their highest, coincident with the arrival of the most powerful storms. The after-landslide season data generally cluster on either side of the before-landslide season data, recording the climatological transition from cool/wet to warm/dry conditions, as well as a departure from elevated hazard potential. Based on our field observations, we plot four landslide-generating s_0 and r_{24} pairs in the ground- and satellite-based data sets (Figures 7a–7d) for each of the January and February 2017 storms in which landsliding occurred. Taken together, the four ground-based s_0 landslide days exhibit a narrower range than do the satellite-based s_0 events, whereas the r_{24} ranges for both sources are nearly the same.

The SL ground-based s_0 and r_{24} landslide days are more distinguished from the null cases than in the satellite-based data (Figures 7a and 7b). Correspondingly, the ground-based optimizations via TS and RD result in thresholds that capture all four landslide days without erroneously including any of the null cases (Figures 7a and 7c). Therefore, both optimization metrics produce identical thresholds with perfect scores for the ground-based data. The satellite-based data result in thresholds with less than perfect scores (mean TS = 0.505; RD = 0.041) because the four landslide days are not as well isolated from the null cases (Figures 7b and 7d). The satellite-based threshold that we optimized with the TS (a statistic averse to false alarms) results in a narrower threshold space, missing one of the four landslide days to avoid incorporating additional null cases. In contrast, the satellite-based threshold that we optimized by the RD captures all four landslide days but at the expense of including more null cases, particularly some false alarms not just during but also well after the landslide season.

The spatially variable landslide thresholds that we developed for each of our catchments (Figure 8) reflect hydroclimatic patterns in the region, as well as differential sensitives to the metric by which we optimized

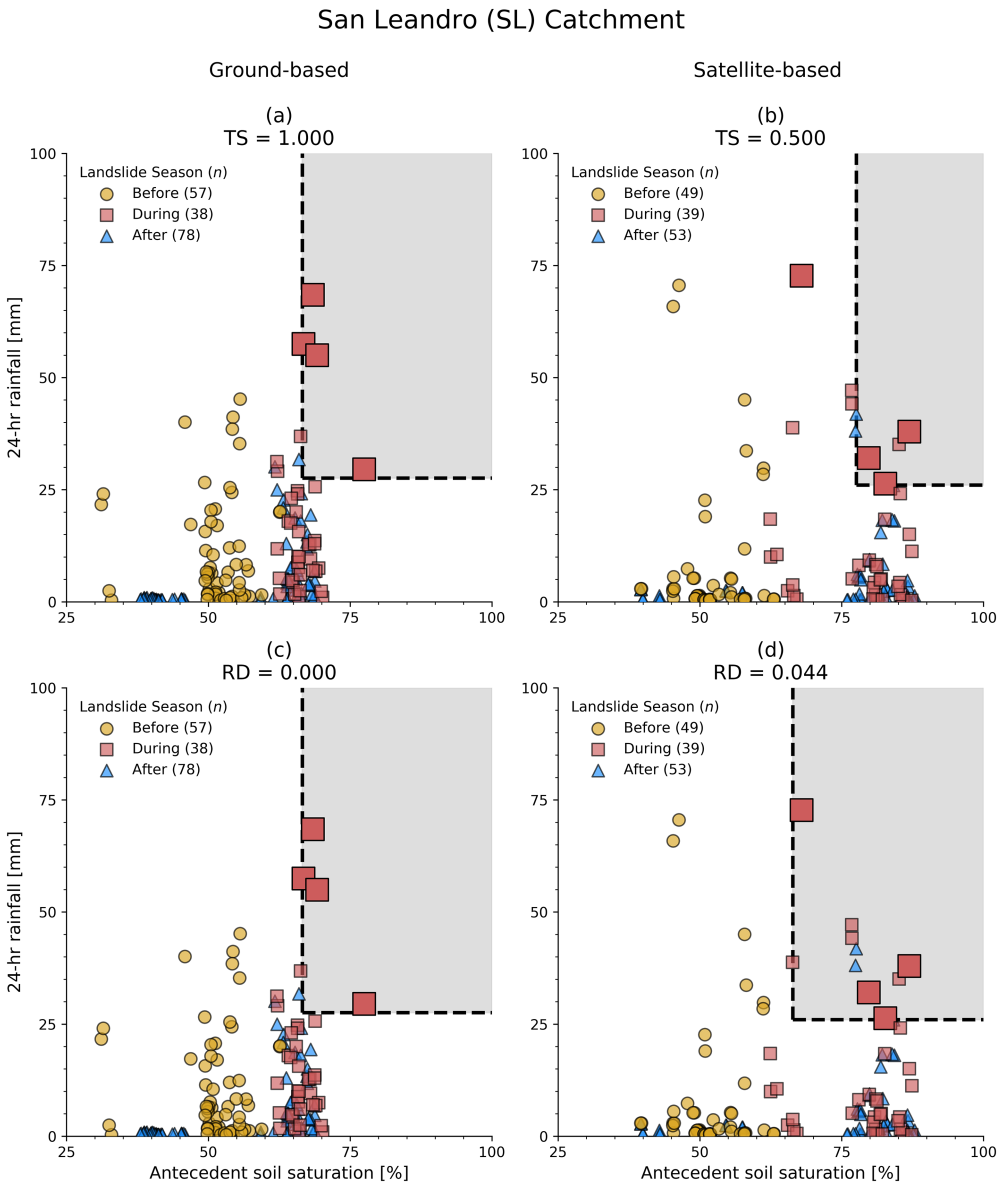

Figure 7. Empirical bilinear ground- and satellite-based landslide thresholds for the San Leandro (SL; see Figure 1b) catchment optimized by (a, b) threat score (TS) and (c, d) radial distance (RD). Large and small symbols respectively indicate landslide days and null cases.

the soil saturation and rainfall data. When we optimized the ground-based landslide thresholds by the TS and RD by catchment (Figures 8a and 8c), the critical s_0 and r_{24} (i.e., the bottom left corner of the bilinear threshold) ranged from 61.5% to 66.6% (mean = 65.3%; Table 1) and from 24.4 to 37.3 mm (mean = 29.4 mm; Table 1), respectively. The relatively narrow s_0 range for the ground-based thresholds (Figures 7a and 7c) suggests that the timing of the r_{24} data is similar, lending some confidence to our extension of the site-specific soil saturation data from BALT1 across the 10 focus catchments. The wide r_{24} range reveals more realistic spatial patterns in rainfall totals compared to the site-specific s_0 data that we used in this study. For example, the threshold for the San Pablo (SP) catchment (Figures 8a and 8c), which is located in the wetter northern hills, requires ~40% more rainfall compared to the Don Castro Reservoir (DC) catchment, which is located in the drier southern hills (i.e., 37.3 versus 28.1 mm, respectively). The satellite-based critical s_0 values that we optimized by TS range from 66.0% to 81.6% (mean = 74.3%; Table 1), and the critical r_{24} values range from 18.4 to 28.1 mm (mean = 24.9 mm; Table

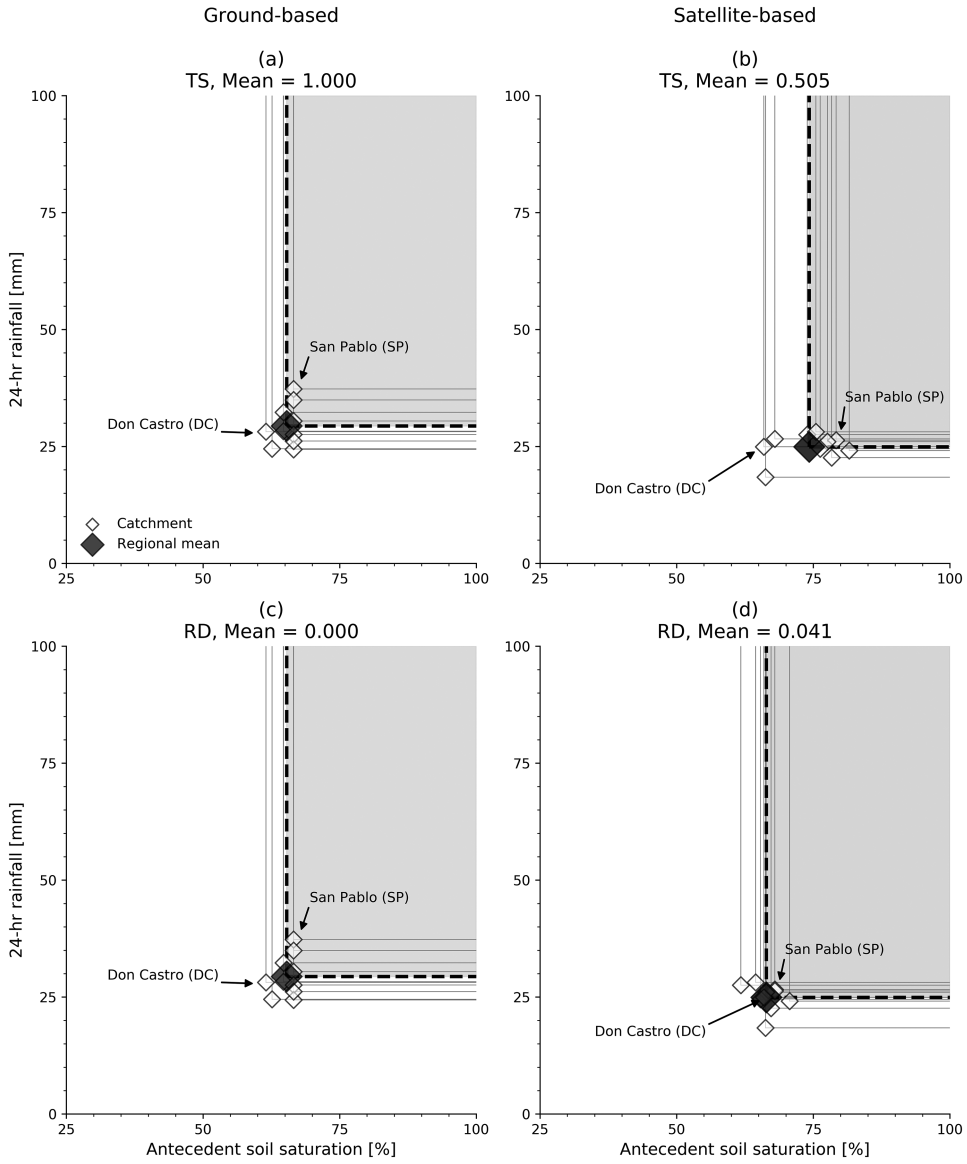

Figure 8. Ground- and satellite-based landslide thresholds for the 10 focus catchments (Figures 1a and 1b) optimized by (a, b) threat score (TS) and (c, d) radial distance (RD). White diamonds indicate the bottom left-hand corner of the bilinear threshold for each catchment and represent critical antecedent soil saturation and 24-hr rainfall levels. Black diamond indicates mean of all catchment-specific critical threshold levels for the East Bay Hills study region.

1). When we optimized the RD (Figures 8b and 8d), the critical s_0 ranges from 61.8 to 70.7 (mean = 66.4%; Table 1) and the critical r_{24} range remains the same. These ranges reflect higher satellite-based critical s_0 values and lower satellite-based critical r_{24} values.

5. Discussion

5.1. Rainfall Characteristics

Our analysis of the NASA GPM-based IMERG data across the 10 focus catchments (Figures 1a and 1b) reveals that the satellite-based product can reproduce the overarching spatiotemporal rainfall patterns we would expect to see in the East Bay Hills of the SF Bay Area. The annual satellite-based rainfall totals are generally higher in the northern hills than in the southern hills, a pattern that is consistent with the ground-based observations (Figures 3a and 3b). The relative proportions of rainfall before, during, and after the

Table 1
Summary of Threshold Characteristics for Each Catchment

Catchment[a]	Ground based			Satellite based		
	$s_{0\text{-}c}$	$r_{24\text{-}c}$	TS, RD	$s_{0\text{-}c}$	$r_{24\text{-}c}$	TS, RD
Hercules (HC)	64.7	32.3	1, 0	73.9, 61.8	27.5	0.600, 0.033
Martinez (MZ)	64.7	28.3	1, 0	68.0, 68.0	26.6	0.444, 0.033
San Pablo (SP)	66.6	37.3	1, 0	79.2, 68.0	26.3	0.600, 0.033
Berkeley (BK)	66.6	34.9	1, 0	75.5, 64.5	28.1	0.600, 0.034
Lafayette (LF)	66.6	30.4	1, 0	81.6, 70.7	24.1	0.429, 0.050
San Leandro (SL)	66.6	27.6	1, 0	77.6, 66.4	26.0	0.500, 0.044
Lake Chabot (LC)	66.6	24.4	1, 0	76.3, 65.4	24.5	0.500, 0.043
San Ramon (SR)	66.6	26.2	1, 0	66.3, 66.3	18.4	0.333, 0.058
Don Castro (DC)	61.5	28.1	1, 0	66.0, 66.0	24.9	0.444, 0.037
Pleasanton Ridge (PR)	62.6	24.5	1, 0	78.4, 67.3	22.6	0.600, 0.043
Regional Mean	65.3	29.4	1, 0	74.3, 66.4	24.9	0.505, 0.041

Note. $s_{0\text{-}c}$ = critical antecedent soil saturation (%); $r_{24\text{-}c}$ = critical 24-hr rainfall (mm); TS = threat score (−); RD = radial distance (−).
[a]Catchments are shown in Figures 1a and 1b.

landslide season are also similar to the ground-based data (Figure 4). However, the annual rainfall for 80% of the catchments is underpredicted by the satellite-based data. The smooth transitioning and narrow range of spatially variable rainfall totals in the satellite data (Figure 3b), likely a result of kilometer-scale averaging and calibration effects, limit the ability of the satellite-based thresholds to resolve catchment-level differences in hazard potential (Figures 8b and 8d). The satellite-based rainfall underestimation trend is also visible in the seasonal comparisons (Figure 4). With the exception of the landslide season for low MAP catchments, the satellite-based rainfall totals appear low, suggesting that the use of an antecedent rainfall metric to identify the onset of the landslide season in our study region would produce a late start date. The marginally better performance for low MAP areas, which exhibit lower-lying topography, confirms that satellite-based rainfall products tend to show sensitivity to more complex topography that is susceptible to orographic effects (e.g., Brunetti et al., 2018; Derin & Yilmaz, 2014; Monsieurs et al., 2018; Rossi et al., 2017; Zambrano-Bigiarini et al., 2017).

Some of the best rainy-day correlations between the ground- and satellite-based products in our study correspond to the landslide season (Figure 5). However, the high degree of scatter that is apparent in some of the daily comparisons throughout WY 2016–2017 suggests that even lower rainfall accumulation windows (e.g., 3 or 6 hr), which would be more characteristic of convective, high-intensity/short-duration rainstorms, may not be resolvable in satellite-based thresholds (see also Rossi et al., 2017). This may be particularly important in watersheds impacted by wildfire, where short bursts of high-intensity rainfall are capable of triggering deadly and destructive debris flows (Staley et al., 2017). Based on our comparisons of longer (i.e., monthly and annual) rainfall levels, we may expect multiday accumulation periods to exhibit slightly lower percent differences than those based on the 24-hr comparisons. However, if multiday accumulation periods were used for thresholds in an operational context, they would have to be balanced against the drop-off in rainfall forecasting skill for periods exceeding 24 hr (Novak et al., 2014).

5.2. Near-Surface Hydrologic Response

In the months leading up to the landslide season, when the soils are relatively dry, the near surface is characterized by capillary-dominated flow processes (Collins et al., 2012; Keefer et al., 1987; Wilson & Jayko, 1997). Here the ground- and satellite-based soil saturation data show good correlation with nearly identical values by the start of the landslide season (Figures 6a and 6b). During the WY 2016–2017 landslide season, after each of the four major storms, the ground-based soil saturation returns (or nearly returns) to its pre-storm level within days, approximating the field capacity (i.e., the amount of water that a soil can hold against free drainage). This gravity-dominated response reflects a seasonal transition from locally controlled to topographically controlled soil moisture states (Grayson et al., 1997). In contrast, the satellite-based soil saturation data do not return to the field capacity after each storm but rather show an increasing soil saturation trend in response to storms throughout and even following the landslide season (Figures 6a and 6c), implying that hazard potential could extend well into April. During the landslide season, the ground- versus satellite-based soil saturation data exhibit the weakest correlation (Figure 6c). The return to higher correlations between the two data sets after the landslide season (Figure 6d), when the soil is drying out, suggests that the satellite-based soil saturation states may better reflect the capillary- as opposed to gravity-dominated subsurface flow observations.

We do not evaluate the absolute accuracy of the SMAP data but rather whether or not it is useful for capturing the antecedent conditions relevant to shallow landsliding. By investigating the analytics used to calculate the satellite-based data, we find compelling scientific evidence for important discrepancies between the data sets. The satellite-based soil saturation data provide a catchment-integrated view of soil saturation with snapshots of soil moisture that are redefined through time. They do not account for hillslope-scale variability in soil depth, vadose zone thickness, and, perhaps most importantly, local topographic slope. In contrast, the

in situ monitoring data are recording soil saturation dynamics high in the catchment among the steepest (and most landslide-prone) slopes, where the near surface drains more quickly, "resetting" the soil saturation to lower levels between each storm. This difference between the ground- and satellite-based soil saturation data (Figure 6), which is most exacerbated during (and just after) the landslide season, is a critical detail reflected in our comparisons of threshold performance (Figure 7). Rather than the s_0 values across the four January and February storms that resemble the field capacity, the satellite-based s_0 is progressively higher from one storm to the next (Figure 6a). The upward trend in satellite-based soil saturation leads to r_{24} values that are paired with a wider range of s_0 values than the ground-based data during and after the landslide season, indicating that the s_0 and r_{24} landslide days are not as readily distinguishable from the null cases. The satellite-based thresholds, therefore, host false positives and correspondingly lower predictive capabilities (Table 1). Although the satellite-based thresholds consistently show lower performance than do the ground-based thresholds, their TS and RD values (mean = 0.505 and 0.041, respectively) suggest these hydrometeorological thresholds may be more useful than rainfall-only thresholds (e.g., the maximum TS for Seattle Area rainfall thresholds reported in Tables 1–3 of Scheevel et al., 2017, is ~0.1).

Adjusting the satellite-based data with relatively straightforward transformations does not markedly improve the performance of our satellite-based thresholds. For example, when we applied a simple vertical offset to the WY 2016–2017 satellite-based data (Figure S10), we found that the soil saturation is still overpredicted. Our comparison of the ground- and satellite-based data for a recent non-landslide year (i.e., WY 2017–2018) further verifies the tendency of the SMAP data to overpredict seasonal soil moisture (Figure S11). Here soil moisture is overpredicted for nearly the entire WY following the first major storm of the season, whether a vertical correction is included or left out. This introduces the potential for producing false positives if used for warning system purposes. We also see that the satellite-based surface (0–5 cm) soil moisture product exhibits higher-amplitude fluctuations than does the near-surface (0–100 cm) product due to the increased effects of evapotranspiration but retains an overall wetting trend during the landslide season (Figure S12). When we optimized our thresholds with this surface soil moisture data, the TS remained the same and the RD improved only marginally (Figure S13). Whereas the surface product reveals more apparent wetting/drying trends than does the near-surface product, it does not reflect the correct soil moisture processes at landslide-relevant depths. Hybrid thresholds, which pair satellite-based soil moisture with ground-based rainfall (and vice versa), produce threshold performance scores in between the purely ground- and satellite-based thresholds (Figure S14). Both hybrid approaches incorporate false positives; however, the threshold that pairs satellite-based rainfall with ground-based soil moisture performs better. These findings confirm that satellite-derived soil moisture is missing key subsurface hydrologic elements needed to capture hillslope drainage following storms.

5.3. Directions for Further Investigation

One possible research direction to address the inability of the land surface model to capture hillslope hydrologic response for periods of gravity-dominated flow would be to use the equilibrium profile from the water balance as an initial condition for physics-based simulations of variably saturated flow that are forced with ground-based rainfall (e.g., Yatheendradas et al., 2019). Our experience using vertical (i.e., infiltration-focused) boundary-value problems to simulate pore-water pressure development at the soil/weathered bedrock interface (i.e., Thomas, Mirus, & Collins, 2018; Thomas, Mirus, Collins, Lu, et al., 2018) suggests that multidimensional simulations may be critical to accommodate the possible impacts of unsaturated zone preferential flow (e.g., Gerke, 2006; Germann & Hensel, 2006; Nimmo, 2012) and lateral saturated flow within unchanneled catchments (e.g., Mirus et al., 2007). Another possibility would be to explore further downscaling the SMAP L4_SM data with higher-resolution vegetation, soil, topographic, surface temperature, and hydroclimatic information (e.g., Abbaszadeh et al., 2018). However, the progressive increase for the satellite-based soil moisture throughout the landslide season would still be problematic for any static downscaling techniques.

Our work highlights the value of in situ measurements for understanding hillslope hydrologic response dynamics relevant to landsliding. These network types are relatively inexpensive compared to the installation of satellite-based networks. Like rain gages, however, it can be difficult to know the spatial applicability of a single sensor measuring in situ soil moisture. Climatic, topographic, and geologic heterogeneities pose challenges to establishing monitoring stations that aim to be physically and statistically representative of

near-surface hydrologic response over large areas. An important step towards this goal is an explicit set of siting criteria (e.g., regionally relevant soil type, soil depth, bedrock type, vegetation, topography, and landslide history), as applied to the shallow landslide monitoring station whose data we used in this study (Collins et al., 2012). Our results suggest that a well-sited landslide monitoring station may be applicable for regional-scale applications and encourages more widespread deployment of near-surface hydrologic monitoring stations to inform hydrometeorological threshold development.

As the collection of overlapping ground- and satellite-based hydrologic data coincides with future, landslide-relevant hydrologic events, it will be important to expand upon the analyses presented here with additional work to explore how ground-based thresholds perform compared to their satellite-based counterparts using more sophisticated analysis techniques (e.g., applying time-derivative plots and variable time scales of rainfall accumulation) across interregional scales for multiyear periods. These kinds of results will broaden understanding beyond our conclusions based on a single (but highly important) year of ground- and satellite-based data that corresponded with widespread shallow landsliding.

6. Conclusion

We found that hydrometeorological (i.e., 24-hr rainfall versus antecedent soil saturation) thresholds optimized with soil moisture dynamics from a single but carefully sited monitoring station in the SF Bay Area (California, USA) outperform those based on more spatially comprehensive satellite-based information. A nearly constant upward trend in the satellite-based soil saturation during the landslide season, as opposed to a punctuated wet up and subsequent dry down to the field capacity after each storm, results in landslide days that do not distinguish themselves from the null cases as captured by ground-based instrumentation. The satellite-based thresholds appear to be more sensitive to the particular skill statistics used for optimization than to catchment-level variability in rainfall and soil saturation, although their overall performance suggests that these hydrometeorological thresholds may be more useful than rainfall-only thresholds. At the regional level, the ground- and satellite-based threshold values were nearly indistinguishable but with more false alarms caused by overestimated soil moisture levels associated with the satellite-based data. In areas with nonexistent in situ monitoring networks, cautious implementation of a satellite-based warning system may be possible, although our findings suggest that warnings may be conservative and negatively impact the predictive capabilities of thresholds for rainfall-induced landsliding. The paired collection of ground- and satellite-based hydrologic observations associated with future widespread landsliding events will be critical to further evaluate the performance of satellite-based thresholds.

Acknowledgments
Rex Baum, Dennis Staley, Thom Bogaard, Dalia Kirschbaum, and an anonymous reviewer provided constructive comments on an earlier version of this work. We appreciate insights regarding computational techniques for threshold optimization provided by Sam Johnstone, productive discussions concerning seasonal soil moisture dynamics with Rex Baum, and catchment shapefiles supplied by Skye Corbett. In situ soil moisture data are available in a USGS Data Release (Thomas et al., 2017; https://doi.org/10.5066/F7M0449D). Satellite-based soil moisture data are available in the NASA SMAP L4_SM repository (Reichle et al. 2018a; https://doi.org/10.5067/60HB8VIP2T8W). Rain gage data are available from the sources listed in Table S1. Satellite-based rainfall data are available in the NASA GPM-based IMERG repository (Huffman et al., 2018; https://pmm.nasa.gov/data-access/downloads/gpm). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Abbaszadeh, P., Moradkhani, H., & Zhan, X. (2018). Downscaling SMAP radiometer soil moisture over the CONUS using an ensemble learning method. *Water Resources Research*, *55*, 324–344. https://doi.org/10.1029/2018WR023354

Baum, R. L., & Godt, J. W. (2010). Early warning of rainfall-induced shallow landslides and debris flows in the USA. *Landslides*, *7*(3), 259–272. https://doi.org/10.1007/s10346-009-0177-0

Bessette-Kirton, E. K., Cerovski-Darriau, C., Schulz, W. H., Coe, J. A., Kean, J. W., Godt, J. W., et al. (2019). Landslides triggered by Hurricane Maria: An assessment of an extreme event in Puerto Rico. *GSA Today*, *29*(6), 4–10. https://doi.org/10.1130/GSATG383A.1

Bogaard, T., & Greco, R. (2018). Invited perspectives. A hydrological look to precipitation intensity duration thresholds for landslide initiation: Proposing hydro-meteorological thresholds. *Natural Hazards and Earth System Sciences*, *18*(1), 31–39. https://doi.org/10.5194/nhess-18-31-2018

Brocca, L., Ciabatta, L., Moramarco, T., Ponziani, F., Berni, N., & Wagner, W. (2016). Use of satellite soil moisture products for the operational mitigation of landslide risk in central Italy. In P. K. Srivastava, G. P. Petropoulos, & Y. H. Kerr (Eds.), *Satellite soil moisture retrieval*, (pp. 231–247). Amsterdam, Netherlands: Elsevier. https://doi.org/10.1016/B978-0-12-803388-3.00012-7

Brocca, L., Ponziani, F., Moramarco, T., Melone, F., Berni, N., & Wagner, W. (2012). Improving landslide forecasting using ASCAT-derived soil moisture data: A case study of the Torgiovannetto Landslide in central Italy. *Remote Sensing*, *4*(5), 1232–1244. http://www.mdpi.com/2072-4292/4/5/1232

Brunetti, M. T., Melillo, M., Peruccacci, S., Ciabatta, L., & Brocca, L. (2018). How far are we from the use of satellite rainfall products in landslide forecasting? *Remote Sensing of Environment*, *210*, 65–75. https://doi.org/10.1016/j.rse.2018.03.016

California Data Exchange Center (2018). Precipitation, "Real-time precipitation", available from: https://cdec.water.ca.gov/snow_rain.html (accessed 30 May 2018).

Cannon, S. H. (1988). Regional rainfall-threshold conditions for abundant debris-flow activity. In S. D. Ellen, & G. F. Wieczorek (Eds.), *Landslides, floods, and marine effects of the storm of January 3–5, 1982, in the San Francisco Bay region, California*, (pp. 35–42). Reston, VA: U.S. Geological Survey. https://doi.org/10.3133/pp1434

Cannon, S. H., & Ellen, S. D. (1985). Rainfall conditions for abundant debris avalanches, San Francisco Bay region, California. *California Geology*, *38*(12), 267–272.

Casadei, M., Dietrich, W. E., & Miller, N. L. (2003). Testing a model for predicting the timing and location of shallow landslide initiation in soil-mantled landscapes. *Earth Surface Processes and Landforms*, *28*(9), 925–950. https://doi.org/10.1002/esp.470

Chan, S., Njoku, E. G., & Colliander, A. (2014). *Soil Moisture Active Passive (SMAP), Algorithm Theoretical Basis Document, Level 1C Radiometer Data Product, Revision A*. Pasadena, CA: Jet Propulsion Laboratory.

Coe, J. A., & Godt, J. W. (2001). Debris flows triggered by the El Nino rainstorm of February 2–3, 1998, Walpert Ridge and vicinity, Alameda County, California (Miscellaneous Field Studies Map 2384). Reston, VA: U.S. Geological Survey. https://doi.org/10.3133/mf2384

Collins, B. D., Corbett, S., Thomas, M. A., Mirus, B. B., & Cerovski-Darriau, C. (2018). A "typical" landslide distribution from above-average winter storms in the San Francisco Bay Area: A new landslide inventory from the East Bay region. Paper presented at the American Geophysical Union Fall 2018 Meeting, Washington, DC.

Collins, B. D. & Corbett, S. C. (2019). *Terrestrial lidar data of the February 14, 2019 Sausalito Boulevard Landslide, Sausalito, California (Data Series 1112). Geological Survey*. Reston, VA: U.S. https://doi.org/10.3133/ds1112

Collins, B. D., Stock, J. D., Foster, K. A., Whitman, M. P. W., & Knepprath, N. E. (2012). Monitoring the subsurface hydrologic response for precipitation-induced shallow landsliding in the San Francisco Bay area, California, U.S.A. In E. Eberhardt, C. Froese, K. Turner, & S. Leroueil (Eds.), *Landslides and engineering slopes: Protecting society through improved understanding*, (pp. 1249–1255). London, UK: Taylor and Francis Group.

Crovelli, R. A., & Coe, J. A. (2009). Probabilistic estimation of numbers and costs of future landslides in the San Francisco Bay region. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, *3*(4), 206–223. https://doi.org/10.1080/17499510802713123

Crow, W. T., Chen, F., Reichle, R. H., & Liu, Q. (2017). L band microwave remote sensing and land data assimilation improve the representation of prestorm soil moisture conditions for hydrologic forecasting. *Geophysical Research Letters*, *44*, 5495–5503. https://doi.org/10.1002/2017GL073642

Crow, W. T., Chen, F., Reichle, R. H., Xia, Y., & Liu, Q. (2018). Exploiting soil moisture, precipitation, and streamflow observations to evaluate soil moisture/runoff coupling in land surface models. *Geophysical Research Letters*, *45*, 4869–4878. https://doi.org/10.1029/2018GL077193

De Lannoy, G. J. M., Koster, R. D., Reichle, R. H., Mahanama, S. P. P., & Liu, Q. (2014). An updated treatment of soil texture and associated hydraulic properties in a global land modeling system. *Journal of Advances in Modeling Earth Systems*, *6*, 957–979. https://doi.org/10.1002/2014MS000330

De Lannoy, G. J. M., & Reichle, R. H. (2016a). Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 catchment land surface model for soil moisture estimation. *Journal of Hydrometeorology*, *17*(2), 669–691. https://doi.org/10.1175/JHM-D-15-0037.1

De Lannoy, G. J. M., & Reichle, R. H. (2016b). Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model. *Hydrology and Earth System Sciences*, *20*(12), 4895–4911. https://doi.org/10.5194/hess-20-4895-201

Derin, Y., & Yilmaz, K. K. (2014). Evaluation of multiple satellite-based precipitation products over complex topography. *Journal of Hydrometeorology*, *15*(4), 1498–1516. https://doi.org/10.1175/JHM-D-13-0191.1

Ducharne, A., Koster, R. D., Suarez, M. J., Stieglitz, M., & Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model: 2. Parameter estimation and model demonstration. *Journal of Geophysical Research*, *105*(D20), 24823–24838. https://doi.org/10.1029/2000JD900328

Ellen, S. D., & Wieczorek, G. F. (Eds) (1988). *Landslides, floods, and marine effects of the storm of January 3–5, 1982*, in the San Francisco Bay region, California (Professional Paper 1434). Reston, VA: U.S. Geological Survey. https://doi.org/10.3133/pp1434

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Gerke, H. H. (2006). Preferential flow descriptions for structured soils. *Journal of Plant Nutrition and Soil Science*, *169*(3), 382–400. https://doi.org/10.1002/jpln.200521955

Germann, P. F., & Hensel, D. (2006). Poiseuille flow geometry inferred from velocities of wetting fronts in soils. *Vadose Zone Journal*, *5*(3), 867–876. https://doi.org/10.2136/vzj2005.0080

Godt, J. W. (Ed) (1999). *Maps showing locations of damaging landslides caused by El Niño rainstorms, winter season 1997–98, San Francisco Bay region, California* (Miscellaneous Field Studies Maps MF-2325-A-J). Reston, VA: U.S. Geological Survey.

Grayson, R. B., Western, A. W., & Chiew, F. H. S. (1997). Preferred states in spatial soil patterns: Local and nonlocal controls. *Water Resources Research*, *33*(12), 2897–2908. https://doi.org/10.1029/97WR02174

Hong, Y., Adler, R., & Huffman, G. (2006). Evaluation of the potential of NASA multi-satellite precipitation analysis in global landslide hazard assessment. *Geophysical Research Letters*, *33*, L22402. https://doi.org/10.1029/2006GL028010

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, D. D., Kojima, M., et al. (2014). The Global Precipitation Measurement mission. *Bulletin of the American Meteorological Society*, *95*(5), 701–722. https://doi.org/10.1175/BAMS-D-13-00164.1

Huffman, D., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R, Xie, P. (2018). Integrated Multi-satellitE Retrievals for GPM (IMERG), version 5b "NASA's Precipitation Processing Center", available from: https://pmm.nasa.gov/data-access/downloads/gpm (accessed 19 April 2018).

Keefer, D. K., Wilson, R. C., Mark, R. K., Brabb, E. E., Brown, W. M., Ellen, S. D., et al. (1987). Real-time landslide warning during heavy rainfall. *Science*, *238*(4829), 921–925. https://doi.org/10.1126/science.238.4829.921

Kirschbaum, D. B., Adler, R., Hong, Y., Kumar, S., Peters-Lidard, C., & Lerner-Lam, A. (2012). Advances in landslide nowcasting: Evaluation of a global and regional modeling approach. *Environmental Earth Sciences*, *66*(6), 1683–1696. https://doi.org/10.1007/s12665-011-0990-3

Kirschbaum, D. B., Adler, R., Hong, Y., & Lerner-Lam, A. (2009). Evaluation of a preliminary satellite-based landslide hazard algorithm using global landslide inventories. *Natural Hazards and Earth System Sciences*, *9*(3), 673–686. https://doi.org/10.5194/nhess-9-673-2009

Kirschbaum, D. B., Huffman, G. J., Adler, R. F., Braun, S., Garrett, K., Jones, E., et al. (2017). NASA's remotely sensed precipitation: A reservoir for applications users. *Bulletin of the American Meteorological Society*, *98*(6), 1169–1184. https://doi.org/10.1175/BAMS-D-15-00296.1

Kirschbaum, D. B., & Stanley, T. (2018). Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness. *Earth's Future*, *6*, 505–523. https://doi.org/10.1002/2017EF000715

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., & Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *Journal of Geophysical Research*, *105*, 24809–24822. https://doi.org/doi:10.1029/2000JD900327

Mirus, B. B., Becker, R. E., Baum, R. L., & Smith, J. B. (2018). Integrating real-time subsurface hydrologic monitoring with empirical rainfall thresholds to improve landslide early warning. *Landslides*, *15*(10), 1909–1919. https://doi.org/10.1007/s10346-018-0995-z

Mirus, B. B., Ebel, B. A., Loague, K., & Wemple, B. C. (2007). Simulated effect of a forest road on near-surface hydrologic response: Redux. *Earth Surface Processes and Landforms*, *32*(1), 126–142. https://doi.org/1002/esp.1387, https://doi.org/10.1002/esp.1387

Mirus, B. B., Morphew, M. D., & Smith, J. B. (2018). Developing hydro-meteorological thresholds for shallow landslide initiation and early warning. *Water*, *10*(9), 1274. https://doi.org/10.3390/w10091274

Monsieurs, E., Kirschbaum, D. B., Tan, J., Maki Mateso, J. C., Jacobs, L., Plisnier, P. D., et al. (2018). Evaluating TMPA rainfall over the sparsely gauged East African Rift. *Journal of Hydrometeorology*, *19*(9), 1507–1528. https://doi.org/10.1175/JHM-D-18-0103.1

Nilsen, T. H., & Turner, B. L. (1975). *The influence of rainfall and ancient landslide deposits on recent landslides (1950–1971) in urban areas of Contra Costa County, California (Bulletin 1388)*. Reston, VA: U.S. Geological Survey. https://doi.org/10.3133/b1388

Nimmo, J. R. (2012). Preferential flow occurs in unsaturated conditions. *Hydrological Processes*, *26*(5), 786–789. https://doi.org/10.1002/hyp.8380

Novak, D. R., Bailey, C., Brills, K. F., Burke, P., Hogsett, W. A., Rausch, R., & Schichtel, M. (2014). Precipitation and temperature forecast performance at the Weather Prediction Center. *Weather and Forecasting*, *29*(3), 489–504. https://doi.org/10.1175/WAF-D-13-00066.1

Pike, R. J., & Sobieszczyk, S. (2008). Soil slip/debris flow localized by site attributes and wind-driven rain in the San Francisco Bay region storm of January 1982. *Geomorphology*, *94*(3-4), 290–313. https://doi.org/10.1016/j.geomorph.2006.09.024

Postance, B., Hillier, J., Dijkstra, T., & Dixon, N. (2018). Comparing threshold definition techniques for rainfall-induced landslides: A national assessment using radar rainfall. *Earth Surface Processes and Landforms*, *43*(2), 553–560. https://doi.org/10.1002/esp.4202

PRISM Climate Group, Oregon State University (2018). 30-Year normals, "30-yr normal precipitation: Annual", available from: http://www.prism.oregonstate.edu/normals/ (accessed 17 August 2018).

Ray, R. L., & Jacobs, J. M. (2007). Relationships among remotely sensed soil moisture, precipitation and landslide events. *Natural Hazards*, *43*(2), 211–222. https://doi.org/10.1007/s11069-006-9095-9

Ray, R. L., Jacobs, J. M., & Cosh, M. H. (2010). Landslide susceptibility mapping using downscaled AMSR-E soil moisture: A case study from Cleveland Corral, California, US. *Remote Sensing of Environment*, *114*(11), 2624–2636. https://doi.org/10.1016/j.rse.2010.05.033

Reichle, R., De Lannoy, G., Koster, R. D., Crow, W. T., Kimball, J. S., & Liu, Q. (2018). SMAP L4 Global 3-hourly 9 km EASE-Grid Surface and Root Zone Soil Moisture analysis update, version 4, "NASA National Snow and Ice Data Center Distributed Active Archive Center", available from: https://doi.org/10.5067/60HB8VIP2T8W (accessed 05 October 2018).

Reichle, R. H., De Lannoy, G. J. M., Liu, Q., Ardizzone, J. V., Colliander, A., Conaty, A., et al. (2017). Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture product using in situ measurements. *Journal of Hydrometeorology*, *18*(10), 2621–2645. https://doi.org/10.1175/JHM-D-17-0063.1

Reichle, R. H., De Lannoy, G. J. M., Liu, Q., Koster, R. D., Kimball, J. S., Crow, W., et al. (2017). Global assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture product using assimilation. *Journal of Hydrometeorology*, *18*(12), 3217–3237. https://doi.org/10.1175/JHM-D-17-0130.1

Reichle, R. H., Koster, R., De Lannoy, G. J. M., Crow, W., & Kimball, J. (2014). *Algorithm Theoretical Basis Document, Soil Moisture Active Passive (SMAP), Level 4 Surface and Root Zone Soil Moisture (L4_SM) Data Product*. Greenbelt, MD: National Aeronautics and Space Administration Goddard Space Flight Center.

Reichle, R. H., Liu, Q. Koster, R., Ardizzone, J. V., Colliander, A., Crow, W. T. Gabrielle, et al. (2018). *Technical Report Series on Global Modeling and Data Assimilation, Volume 52. Soil Moisture Active Passive (SMAP) Project Assessment Report for Version 4 of the L4_SM Data Product (TM–2018-104606)*. Greenbelt, MD: National Aeronautics and Space Administration.

Rossi, M., Luciani, S., Valigi, D., Kirschbaum, D., Brunetti, M. T., Peruccacci, S., & Guzzetti, F. (2017). Statistical approaches for the definition of landslide rainfall thresholds and their uncertainty using rain gauge and satellite data. *Geomorphology*, *285*, 16–27. https://doi.org/10.1016/j.geomorph.2017.02.001

Scheevel, C. R., Baum, R. L., Mirus, B. B., & Smith, J. B. (2017). *Precipitation thresholds for landslide occurrence near Seattle, Mukilteo, and Everett, Washington* (Open-File Report 2017-1039). Reston, VA: U.S. Geological Survey. https://doi.org/10.3133/ofr20171039

Segoni, S., Rosi, A., Lagomarsino, D., Fanti, R., & Casagli, N. (2018). Brief communication: Using averaged soil moisture estimates to improve the performances of a regional-scale landslide early warning system. *Natural Hazards and Earth System Sciences*, *18*(3), 807–812. https://doi.org/10.5194/nhess-18-807-2018

Sidle, R. C., Gomi, T., & Tsukamoto, Y. (2018). Discovery of zero-order basins as an important link for progress in hydrogeomorphology. *Hydrological Processes*, *32*(19), 3059–3065. https://doi.org/10.1002/hyp.13246

Sloan, D. (2006). *Geology of the San Francisco Bay region*, (p. 335). Berkeley, CA: University of California Press.

Staley, D. M., Kean, J. W., Cannon, S. H., Schmidt, K. M., & Laber, J. L. (2013). Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern California. *Landslides*, *10*(5), 547–562. https://doi.org/10.1007/s10346-012-0341-9

Staley, D. M., Negri, J. A., Kean, J. W., Laber, J. L., Tillery, A. C., & Youberg, A. M. (2017). Prediction of spatially explicit rainfall-intensity duration thresholds for post-fire debris-flow generation in the western United States. *Geomorphology*, *278*(1), 149–162. https://doi.org/10.1016/j.geomorph.2016.10.019

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285–1293. https://doi.org/0.1126/science.3287615, https://doi.org/10.1126/science.3287615

Taylor, F. A., & Brabb, E. E. (1972). *Maps showing distribution and cost by counties of structurally damaging landslides in the San Francisco Bay region, California, winter of 1968–69* (Miscellaneous Field Studies Map 327). Reston, VA: U.S. Geological Survey. https://doi.org/10.3133/mf327

Terzhagi, K. (1943). *Theoretical soil mechanics*. New York, NY: Wiley. https://doi.org/10.1002/9780470172766

Thiessen, A. H. (1911). Precipitation averages for large areas. *Monthly Weather Review, 39*, 1082-1089. https://doi.org/10.1175/1520-0493(1911)39<1082b:PAFLA>2.0.CO;2

Thomas, M. A., Collins, B. D., Stock, J. D., Corbett, S. C., Schmidt, K. M., Reid, M. E., et al. (2017). Field data used to support numerical simulations of variably-saturated flow focused on variability in soil-water retention properties for the U.S. Geological Survey Bay Area Landslide Type (BALT) Site #1 in the East Bay region of California, USA (U.S. Geological Survey data release). Reston, VA: U.S. Geological Survey. https://doi.org/10.5066/F7M0449D

Thomas, M. A., Mirus, B. B., & Collins, B. D. (2018). Identifying physics-based thresholds for rainfall-induced landsliding. *Geophysical Research Letters*, *45*, 9651–9661. https://doi.org/10.1029/2018GL079662

Thomas, M. A., Mirus, B. B., Collins, B. D., Lu, N., & Godt, J. W. (2018). Variability in soil-water retention properties and implications for physics-based simulation of landslide early warning criteria. *Landslides*, *15*(7), 1265–1277. https://doi.org/10.1007/s10346-018-0950-z

U.S. Geological Survey (2018). The National Map, "1/3 arc-seconds digital elevation model", available from: https://www.usgs.gov/core-science-systems/national-geospatial-program/national-map (accessed 19 November 2018).

U.S. Geological Survey (2019). Landslide hazards, "Monitoring", available from: https://www.usgs.gov/natural-hazards/landslide-hazards/monitoring (accessed 10 April 2019).

Wieczorek, G. F., Harp, E. L., & Mark, R. K. (1988). Debris flows and other landslides in San Mateo, Santa Cruz, Contra Costa, Alameda, Napa, Solano, Sonoma, Lake and Yolo Counties, and factors influencing debris-flow distribution. In S. D. Ellen, & G. F. Wieczorek (Eds.), *Landslides, floods, and marine effects of the storm of January 3–5, 1982, in the San Francisco Bay region*, (pp. 133–161). Reston, VA: U.S. Geological Survey. https://doi.org/10.3133/pp1434

Wilson, R. C. (2004). The rise and fall of a debris-flow warning system for the San Francisco Bay region, California. In T. Glade, M. Anderson, & M. J. Crozier (Eds.), *Landslide hazard and risk*, (pp. 493–516). New York, New York: Wiley. https://doi.org/10.1002/9780470012659.ch17

Wilson, R. C. & Jayko, A. S. (1997). Preliminary maps showing rainfall thresholds for debris flow activity, San Francisco Bay Region, California (Open File Report 97-745F). Reston, VA: U.S. Geological Survey. https://doi.org/10.3133/ofr97745F

Wilson, R. C., Mark, R. K., & Barbato, G. (1993). Operation of a real-time warning system for debris flow in the San Francisco Bay area, California. In H. W. Shen, S. T. Su, & F. Wen (Eds.), *Hydraulic engineering 1993: Proceedings of the 1993 conference*, (Vol. 2, pp. 1908–1913). San Francisco, California: American Society of Civil Engineers.

Wunderground (2018). Personal Weather Station Network, "Rainfall", available from: https://www.wunderground.com/weatherstation/overview.asp ().

Yatheendradas, S., Kirschbaum, D. B., Nearing, G., Vrugt, J. A., Baum, R. L., Wooten, R., et al. (2019). Bayesian analysis of the impact of rainfall data product on simulated slope failure for North Carolina locations. *Computational Geosciences*. https://doi.org/10.1007/s10596-018-9804-y

Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., & Ribbe, L. (2017). Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. *Hydrology and Earth System Sciences*, *21*(2), 1295–1320. https://doi.org/10.5194/hess-21-1295-2017