The Value of Using Multiple Hydrometeorological Variables to Predict Temporal Debris Flow Susceptibility in an Alpine Environment

Prenner, D.; Kaitna, R.; Mostbauer, K.; Hrachowitz, M.

DOI
10.1029/2018WR022985

Publication date
2018

Document Version
Final published version

Published in
Water Resources Research

Citation (APA)
Prenner, D., Kaitna, R., Mostbauer, K., & Hrachowitz, M. (2018). The Value of Using Multiple Hydrometeorological Variables to Predict Temporal Debris Flow Susceptibility in an Alpine Environment. *Water Resources Research*, *54*(9), 6822-6843. https://doi.org/10.1029/2018WR022985

Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR022985

Key Points:
- Signals of debris flow triggers are visible in regional hydrometeorological parameters
- Hydrometeorological parameters can be used to differentiate between different trigger types
- Debris flow susceptibility prediction is improved when hydrological parameters are included

Supporting Information:
- Supporting Information S1

Correspondence to:
D. Prenner,
david.prenner@boku.ac.at

Citation:
Prenner, D., Kaitna, R., Mostbauer, K., & Hrachowitz, M. (2018). The value of using multiple hydrometeorological variables to predict temporal debris flow susceptibility in an Alpine environment. *Water Resources Research*, *54*. https://doi.org/10.1029/2018WR022985

Received 21 MAR 2018
Accepted 5 SEP 2018
Accepted article online 10 SEP 2018

The Value of Using Multiple Hydrometeorological Variables to Predict Temporal Debris Flow Susceptibility in an Alpine Environment

D. Prenner[1], R. Kaitna[1], K. Mostbauer[1], and M. Hrachowitz[2]

[1]Institute of Mountain Risk Engineering, University of Natural Resources and Life Sciences, Vienna, Austria, [2]Water Resources Section, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands

Abstract Debris flows are typically triggered by rainfall-related weather conditions—including short-duration storms and long-lasting rainfall, in cold climates sometimes in connection with intensive snowmelt. Given the considerable observational uncertainties of rainfall, we tested if other hydrometeorological variables carry enough information content to compensate for these uncertainties and if the combined information of hydrologic catchment state and rainfall can be used to predict the regional temporal susceptibility for debris flow initiation. For this we carried out a probabilistic analysis of variables derived from a conceptual hydrological model for the Montafon region, Austria, where debris flows were recorded on 41 days between 1953 and 2013. Exclusively from hydrological characteristics and, importantly, neglecting precipitation itself, we quantitatively determined different trigger types for historical debris flows. Subsequently, we used four Naive Bayes classifier models, ranging from a simple rainfall-only model to a multiparameter hydrometeorological model differentiating between trigger types, to predict days susceptible for debris flow occurrence in the region. The results suggest that debris flows were triggered by convective rainstorm events on 23 days, on 12 days due to gradual soil moisture buildup in the course of long-lasting rainfall events and on six further days snowmelt played an important role. We find that the differences between the trigger types are statistically significant and that a susceptibility prediction differentiating between trigger types and including hydrological information can outperform simple rainfall-only models. This study thereby contributes to an improved understanding of the hydrometeorological impact on debris flow initiation in a mountain watershed.

1. Introduction

Debris flows are geomorphologic processes that drive mountain landscape evolution over long time scales (e.g., Stock & Dietrich, 2003) but may represent a hazard for people, settlements, and infrastructure at shorter time scales (e.g., Ballesteros-Cánovas et al., 2016). Besides earthquakes, volcanism, or dam break floods, precipitation is the most frequent trigger of debris flows. Triggering precipitation may origin from long-lasting rainfall (LLR) or short-duration storm (SDS) events that are mostly associated with convective processes (e.g., Berti et al., 1999; Church & Miles, 1987; Mostbauer et al., 2018; Stoffel et al., 2011). In addition, intense snowmelt (SM) can play a considerable role for debris flow initiation in cool regions (Cardinali et al., 2000; Church & Miles, 1987; Mostbauer et al., 2018; Stoffel et al., 2011). Notwithstanding considerable progress over the past years, reliable regional predictions of debris flows remain problematic and need to be improved to reduce human and economic losses (cf. Bogaard & Greco, 2016, 2018).

Similar to and sometimes in combination with shallow landslide initiation, many approaches to predict debris flow occurrence are based on threshold-value-concepts, which attempt to identify typical precipitation characteristics, such as total rainfall amount over a specific period, rainfall intensity, or rainfall duration, that lead to initiation of debris flows (cf. Guzzetti et al., 2007, 2008, and references therein). As pointed out by Leonarduzzi et al. (2017), a common problem of precipitation-threshold concepts is the objective definition and meaningful validation of a threshold value without considering rainfall events that did not trigger an event. Probabilistic thresholds represent an alternative approach and have been successfully applied for shallow landslides by Frattini et al. (2009), Brunetti et al. (2010), Berti et al. (2012), and Braun and Kaitna (2016).

By considering the hydrological history of a watershed, Crozier (1999) presented the Antecedent Water Status (AWS) model to account for the hydrological susceptibility of a watershed for landslide occurrence. The term

susceptibility here reflects the notion that the temporal varying hydrological catchment states may influence the critical water input needed to trigger a landslide. The idea of the AWS model was that less rainfall input is needed to trigger a landslide when antecedent soil moisture is high. The method conceptually and in a very simplified manner mimics soil storage behavior under consideration of rainfall input, evaporative fluxes, and drainage. The model considers both, antecedent water accumulation over 10 days and event-day water to derive a critical soil-water-content threshold for debris flow occurrence. Ciavolella et al. (2016) further extended this concept by applying a conceptual rainfall-runoff model to reproduce the water balance of a catchment and derive a threshold curve based on event precipitation and some metric of catchment storage. Prediction performance was slightly better compared to intensity-duration (I-D) threshold curves and characterized by a higher number of correctly predicted debris flows (true positives) but also considerably more incorrectly predicted debris flow events (false positives). Meyer et al. (2012) introduced a hydrometeorological I-D curve by considering critical water supply from two sources (SM and rainfall), which improved the performance especially for detecting debris flow initiation during spring time.

A general shortcoming of the concepts that rely on observed precipitation data is a result of the spatiotemporal heterogeneity of precipitation (Chaubey et al., 1999; Shah et al., 1996; Singh, 1997), which is commonly not captured by even a dense network of meteorological stations (e.g., Hrachowitz & Weiler, 2011). Nikolopoulos et al. (2014) investigated the propagation of rainfall measurement uncertainties into resulting I-D thresholds curves and showed that uncertainties increase with decreasing rainfall duration. Leonarduzzi et al. (2017) address this issue by deriving probabilistic rainfall thresholds for shallow landslide initiation from a novel data set of gridded rainfall (2×2-km resolution) on a daily basis. Even remote precipitation observation technologies like geostationary satellites, weather radar, or cellular communication networks sometimes face problems in detecting regional, short-duration, convective rainfall events (Devoli et al., 2015; Leijnse et al., 2007; Llasat et al., 2005; Marra et al., 2014; Salio et al., 2015) and have been available for merely a very limited time period, frequently including no or only a few debris flow events.

Although debris flow initiation itself is a highly local phenomenon, the hydrometeorological formation processes that ensure sufficient water input into the system and lead to debris flows may be clearly identifiable at larger scales. We expect that the formation processes exhibit distinct signatures that can serve as a proxy for unrepresentative precipitation observations on a regional scale but without giving detailed spatial information of the subcatchment(s) in which debris flow(s) may be initiated. For example, a LLR event is mostly based on the frontal movement of air masses of different temperatures and may also develop thunderstorms at the fronts (e.g., Ahrens, 2008; Houze, 2014; Haeckel, 2016). Effectively large-scale stratiform clouds and relatively low mean daily temperatures prevail (e.g., Rulfová & Kyselý, 2013). In contrast, SDSs origin from convective lifting of moist air (Ahrens, 2008). They occur much more localized at higher mean temperatures and show strong updraft movement rates of air masses, which can lead to larger precipitation elements (i.e., raindrops) and even hailstones (e.g., Ahrens, 2008; Houze, 2014). Orographic lifting can enhance this effect (Haeckel, 2016). For the formation of SDS, the soil moisture-atmosphere coupling plays an ambivalent role. While transpiration from wet soils bundles latent heat, the warming over dry soils generates sensible heat, which both provide the necessary energy to start a convective process (Ford et al., 2015). Combined, some or ideally all of the above factors may hold sufficient information to infer, within some limits of uncertainty, the probability of LLR or SDS occurring on a given day, even if rainfall was not explicitly recorded.

Characteristic signatures of these different meteorological processes can be found in the evolution of the hydrological state of a catchment during the days preceding these processes. A steady increase of soil moisture over several days can be expected to be a result of generally wet conditions, with prolonged periods of relatively low intensity water input due to either LLR or SM. Similarly, decreasing soil moisture over some days before a rainfall event may be interpreted as a signal for SDS. The driving force behind this soil moisture reduction is evapotranspiration (Hargreaves & Samani, 1982) Thus, high potential evapotranspiration rates, a consequence from the associated solar energy input, some days before the rainfall event, provide evidence for the presence of sufficient energy supply to start a SDS event. In contrast, a continuous decrease of potential evapotranspiration can result from clouds blocking the energy input, which rather suggests an LLR event. Eccel (2012) found that the daily minimum temperature is an adequate approximation for the dew point temperature (i.e., the temperature at which clouds form) in humid Alpine climates. Hence, the difference between maximum and minimum air temperature (temperature span) is much smaller when LLR prevails than at days where a SDS begins to form (Haeckel, 2016). Likewise, an unambiguous sign for considerable influence of SM

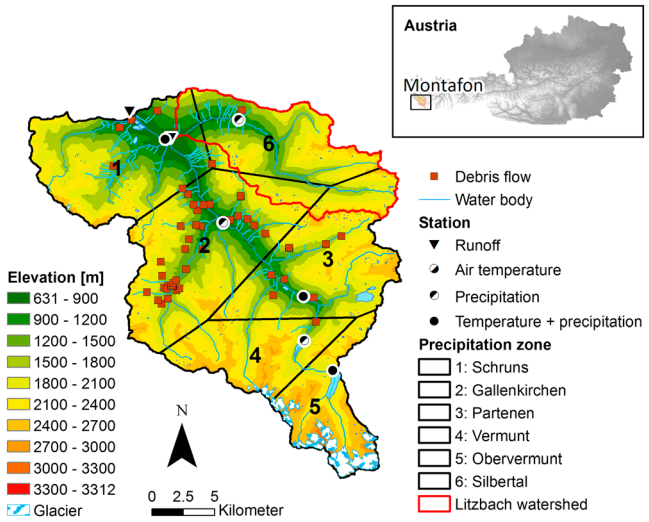

Figure 1. Overview of the study region Montafon, Austria, including elevation distribution, location of observed debris flow deposition, measurement station locations (precipitation, temperature, and runoff), the six precipitation zones (Thiessen-Polygon decomposition of precipitation stations), and the hydropower-undisturbed catchment Litzbach which was used for calibration (red boundary).

as debris flow trigger during periods of the year when a snowpack is present are prolonged periods of above-freezing temperatures that can, depending on the temperature and water content of the snowpack, lead to sufficiently intense melt to trigger debris flows or to at least contribute to trigger conditions.

This study directly follows up on recent results of Mostbauer et al. (2018) who hydrologically modeled an Alpine watershed and identified three relevant sources of water input (high-intensity precipitation, SM, and antecedent soil moisture) that are connected to debris flow occurrence with varying seasonal importance. Here we test for the Montafon region in the Austrian Alps the hypotheses that (1) debris flows in such an environment occur at distinct catchment wetness states and are triggered by water input from different weather conditions, (2) signatures of these distinct triggering processes can be found in larger-scale pattern of the hydrometeorological conditions in the days leading up to a debris flow event, without the need for actual precipitation observations and the associated uncertainties and that (3) the hydrometeorological signatures are robust enough to be used for a probabilistic regional prediction of a region's temporal susceptibility to debris flows occurrence. We use a semi-distributed, process-based hydrological rainfall-runoff model to obtain estimates of daily catchment states and fluxes of the study region. With a Bayesian analysis we identify different trigger regimes and establish criteria to classify the catchment state into different trigger types LLR, SDS and SM. A similar trigger-type definition has already been proposed by Merz and Blöschl (2003) for flood events. Finally, we use a Naive Bayes model to perform a trigger-type separated prediction of temporal debris flow susceptibility on the basis of multiple hydrometeorological variables.

2. Study Catchment

The Montafon region is located in the western part of Austria and is drained by the Ill river (Figure 1). Bounded by the discharge gauge in Vandans, the study catchment covers an area of 510 km^2 and an elevation range from 631 to 3,312 m above sea level (a.s.l). It is influenced by an oceanic climate (Hammerl, 2001), bringing moist air from the Atlantic Ocean, which results in a mean annual precipitation of 1,325 mm. The geology consists mainly of metamorphic rock like amphibolite, gneis, and schiest. About 37% of the Montafon region is covered by grassland, 31% by forest, and 25% by bare rock/sparsely vegetated areas at high elevations, including 2% glacial areas in the south. The remaining 7% represent the riparian zone adjacent to the channel system. A major part of the catchment is affected by water diversion and/or (pumped-)storage hydropower plants. Only the subcatchment of the Litzbach (100 km^2 at gauge Schruns), characterized by a nival regime (according to Mader et al., 1996) with highest discharges in June, July, and May, has undisturbed runoff conditions and is subsequently used for calibration of the hydrological model in this study. In the region, local authorities documented 78 damage-causing debris flow events since the year 1900 by known deposition location and date.

3. Methodology

The experimental setup for this study involved four steps as displayed in the flow chart of Figure 2. At first a semidistributed hydrological model was calibrated for the undisturbed Litzbach catchment and run for the Montafon catchment to generate estimates of water storage in different compartments of the system (see sections 3.2 and 4.1). We assume that the variety of hydrological state and flux variables (e.g., soil moisture, SM, and evapotranspiration) together with meteorological information of precipitation and temperature displays a holistic representation of the catchment condition on a daily basis. In step two, a Bayesian analysis of debris flow trigger probabilities conditional on the above hydrometeorological variables as well as their change over time was used to detect characteristic signatures associated with the occurrence of debris flows in the study region (see sections 3.3, 4.1, and 4.2). The derived pattern allowed a classification of the observed

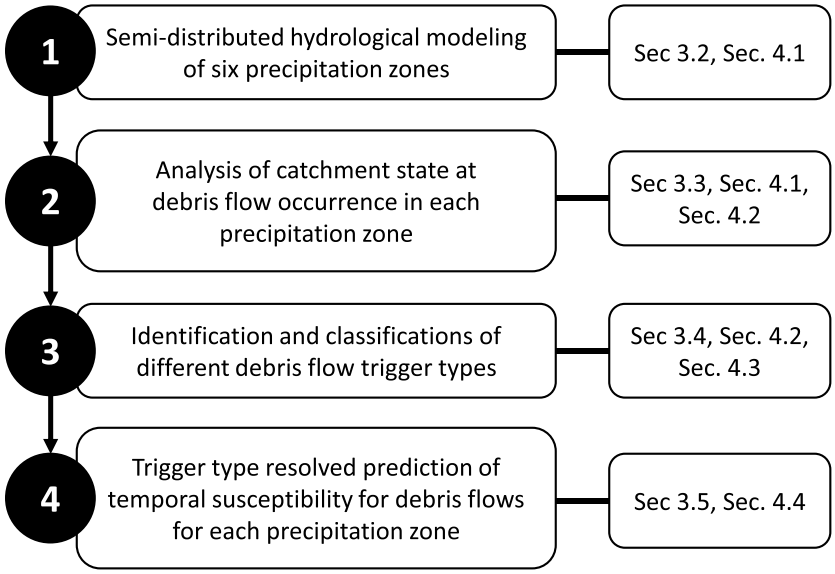

Figure 2. Flow chart of the introduced procedure and link to the relevant sections.

debris flows into the three trigger classes LLR, SDS, and intense SM, which was done in step three (see sections 3.4, 4.2, and 4.3). In the final step four, the above classification was used to predict the hydrometeorological susceptibility for debris flows occurrence on a specific day as a function of a range of hydrometeorological variables (see sections 3.5 and 4.4).

3.1. Available Data

For the analysis, we use data on a daily basis of stations operated by the Austrian Central Institute for Meteorology and Geodynamics (ZAMG), Hydrographic Service (HD), and hydropower plant company Illwerke. The model period was from 1 January 1953 to 31 December 2013. Precipitation was recorded at six stations (Figure 1), which are located in Schruns (660 m a.s.l), Gallenkirchen (860 m a.s.l), Silbertal (880 m a.s.l), Partenen (1,028 m a.s.l), Vermunt (1,735 m a.s.l), and Obervermunt (2,045 m a.s.l). Note that the stations are located on easily accessible places and some distance from the suspected initiation zones of the observed debris flows. Precipitation sums were registered at 7 a.m. each day instead of ideally at midnight, which introduces some uncertainty in the analysis as stated by Peres et al. (2018). A Thiessen-Polygon decomposition was used to delineate the areal extend of each precipitation zone, thereafter used as distributed moisture input into the model (e.g., Euser et al., 2015). Temperature measurements were available at three stations. For the HD operated stations, mean daily air temperature had to be approximated from observations at 7 a.m., 14 p.m., and 21 p.m. using the Kaemtz method (Dall'Amico & Hornsteiner, 2006). Runoff data for model calibration were available from 1 January 1976 to 31 December 2013 for the Litzbach, because this gauge was not affected by hydropower diversion. All over the study region 78 debris flow events were observed on 38 days between 1956 and 2013, without detailed information on the mechanism or the location of initiation. In our analysis, we were able to cover 41 event days since events at the same date but in different precipitation zones were counted separately. Following this division, Gallenkirchen had the most event days (21), followed by Partenen (12), Schruns (4), Silbertal (3), and Vermunt (1). No debris flow event was registered in zone Obervermunt. For defining the individual response units of the hydrological model we used the CORINE Land Cover data set from 1990 (European Environment Agency, 2014), a 10 × 10-m digital elevation model (vogis.cnv.at), 10 × 10-m height-above-nearest drainage map (Rennó et al., 2008), and a glacier distribution map of 1969 (Patzelt, 2015).

3.2. Hydrological Model

3.2.1. Model Structure

The hydrological catchment state was modeled on a daily basis with a process-based, semidistributed rainfall-runoff model (e.g., Fovet et al., 2015; Nijzink et al., 2016) written in C and run on high-performance computers of the Vienna Scientific Cluster. Figure 3 shows the model structure that includes several

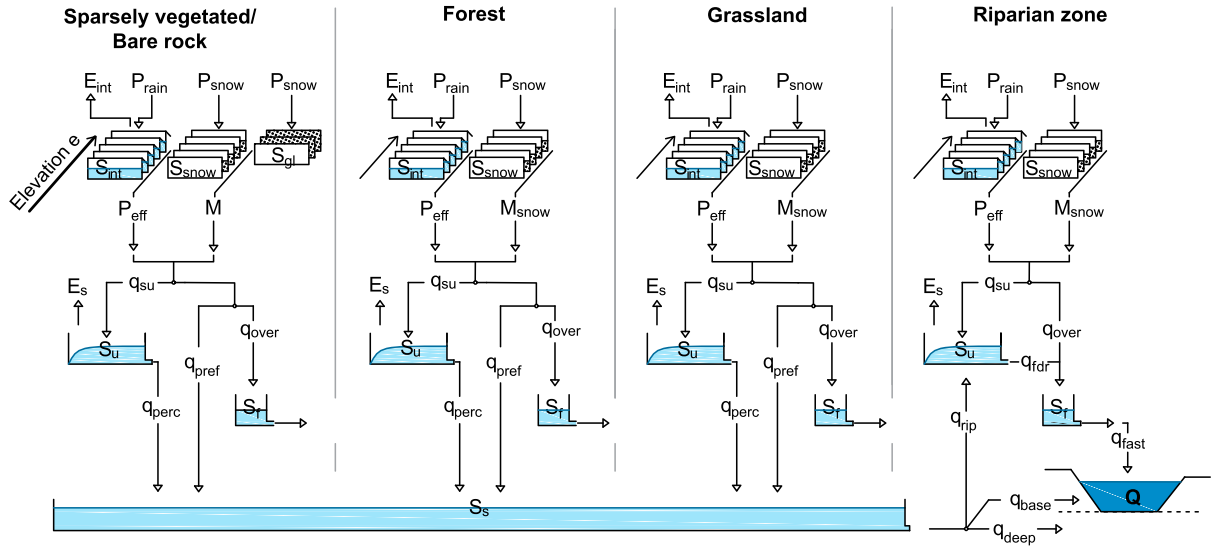

Figure 3. Structure, fluxes, and reservoirs of the hydrological rainfall-runoff model, which was applied to each precipitation zone. E_{int} = elevation resolved interception evaporation; P_{rain} = elevation corrected liquid precipitation; P_{snow} = elevation and seasonal corrected snowfall; S_{int} = elevation resolved interception reservoir; S_{snow} = elevation resolved snow reservoir; S_{gl} = elevation resolved glacier reservoir; P_{eff} = effective precipitation; M = snow M_{snow} and glacier melt; E_s = soil evapotranspiration; S_u = soil reservoir, q_{su} = flux into soil reservoir; q_{pref} = preferential flow; q_{perc} = percolation into groundwater reservoir S_s; q_{over} = excess overland flow; S_f = fast responding reservoir; q_{fdr} = fast soil reservoir drainage; q_{rip} = groundwater flux into riparian zone; q_{fast} = fast runoff component; q_{base} = baseflow from groundwater; q_{deep} = deep groundwater loss; Q = channel runoff.

storage components, representing snow and glacier, interception, unsaturated soil, fast responding, and slow responding system components individually for each precipitation zone. To ensure a balance between necessary spatial detail of the model and available data to meaningfully constrain the model parameter space (Hrachowitz & Clark, 2017), the model domain was discretized into six individual precipitation zones, which corresponds to the highest resolution of available information on precipitation (as the main driver for debris flow initiation). In addition, each precipitation zone was further subdivided into four hydrological response units to distinguish between different response dynamics of forest, grassland, sparsely vegetated/bare rock, which represents hillslope units on the one hand and riparian zones with shallow groundwater tables (and thus little unsaturated storage capacities) on the other hand (cf. Savenije & Hrachowitz, 2017). The classification into the hillslope and the riparian zones was based on the height-above-nearest drainage concept (Gharari et al., 2011), with a threshold of 3 m to distinguish between the classes (Gao et al., 2014). Further, elevation was stratified into bands of 100 m to better account for altitude dependent quantities like precipitation, temperature as well as thereof related processes like SM (Gao et al., 2017; Rolland, 2003; Sevruk, 1997). In total, the model represents 23 individual, parallel components. The modeled flux and state variables (e.g., SM and soil moisture) of the 23 components are area weighted to obtain representative values for the individual precipitation zones as well as for the entire catchment (e.g., modeled runoff, which is used for calibration). Calibration parameters were defined for each hydrological response unit individually (e.g., Gao et al., 2016; Gharari, Hrachowitz, et al., 2014), except some globally valid parameters like the freezing point and temperature lapse rates (see supporting information Table S1). All differential water balance equations are discretized using the implicit Euler scheme to minimize the impact of numerical distortions on the results (Kavetski & Clark, 2010, 2011; Kavetski et al., 2006). The equations are then solved by using the Newton-Raphson iteration approach (Ypma, 1995). A complete description of the model including all relevant constitutive relations is provided in supporting information Text S1.

3.2.2. Model Calibration

The calibration period ranged from 1 January 1978 to 31 December 2013 and 2-year warm up period preceding the calibration period. The model uses 44 calibration parameters, 6 of which are required for the error model. The uniformly distributed prior as well as the posterior distributions of all parameters can be found in Table S1 of the supporting information. For calibration, we applied a differential evolution adaptive metropolis sampler, which enabled an efficient sampling of the posterior distribution (Vrugt, 2016; Vrugt et al., 2008).

For our application, we specified 16 Markov chains for performing a random walk to detect the global optimal parameter set. The likelihood function, stating the probability that the observed runoff could be reproduced with a sampled parameter set, was determined by a standardized skew exponential power distribution combined with a Box-Cox transformation. The residual errors of the objective function are hence quantified not only by the mean and variance but also by the kurtosis and skewness of the distribution (Schoups & Vrugt, 2010). Heteroscedasticity in the input data was considered by a linear relationship between variance and increasing runoff. Similar to Gharari, Shafiei, et al. (2014), we applied soft parameter constraints (e.g., interception capacity is larger in forests than in grassland). The differential evolution adaptive metropolis sampler algorithm was run for 35,000 generations using the Vienna Scientific Cluster. The hydrological simulation of the complete Montafon region was subsequently carried out with the calibrated parameter set determined for the Litzbach subcatchment. We considered the uncertainties of the hydrological modeling by simulating the study region with 100 different model parameter sets, randomly sampled from the parameters posterior distributions.

To ensure the robustness of the posterior parameter distributions, the model's skill to simultaneously reproduce several performance criteria and catchment signatures (e.g., Euser et al., 2013; Hrachowitz et al., 2014) was evaluated postcalibration by sampling 100 model parameter sets from the posterior distributions and computing the Nash-Sutcliffe efficiency (NSE) of flows (Nash & Sutcliffe, 1970), the NSE of the logarithm of flows NSElog, the volumetric efficiency (VE; Criss & Winston, 2008), and the NSE of the flow duration curve (NSEFDC).

3.3. Hydrometeorological Trigger Probabilities

We applied Bayes' theorem (Bayes & Price, 1763) to compute trigger probabilities for the time series of hydrological model states and flux variables, meteorological input variables rainfall and temperature, and their temporal derivatives, sums, and averages:

$$P(E|x_i) = \frac{P(E) * P(x_i|E)}{P(x_i)} \tag{1}$$

Analog to Berti et al. (2012), equation (1) can be simplified to yield equation (2), which expresses the ratio between the number of occurrences of magnitude $x_i$in connection to debris flow events $N_{x_i|E}$ and the total number of occurrences N_{x_i}:

$$P(E|x_i) = \frac{N_{x_i|E}}{N_{x_i}} \tag{2}$$

The complete value range of a hydrometeorological variable over the simulation period was discretized into percentile classes (bin size five) for each precipitation zone before they are evaluated according to equation (2). For a catchment wide comparison (all six zones analyzed at once), we aggregate the items of the same percentile bins over all precipitation zones and consequently run equation (2) for each bin of the newly generated data set. This approach allowed a better comparison of different parameter value ranges across the precipitation zones (e.g., an extremely high temperature value measured at a mountain site likely displays an inconspicuous value at a valley station) as well as eased the detection of high trigger probabilities, because each class contains the same number of items N_{x_i} (only $N_{x_i|E}$ controls the probabilities then). Note that the latter mentioned feature is not necessarily true when we carry out the analysis for the temporal development of hydrometeorological variable over multiple prior days, what represents a kind of a two-dimensional analysis.

3.4. Identification of Debris Flow Trigger Type

It is hypothesized that characteristic signatures can be identified for different hydrometeorological conditions with a combined analysis of multiple hydrometeorological variables, which, in turn, allow a robust classification of the distinct debris flow trigger mechanisms LLR, SDS, and SM in absence of spatially and temporally high resolution precipitation information. To test this hypothesis, we explored a range of different observed and modeled variables that may allow a simplified identification of these hydrometeorological conditions. On the one hand, LLR events are typically associated with a gradual increase of soil moisture content

Table 1
Criteria for the Identification Whether Long-Lasting Precipitation (LLR), Short-Duration Storm (SDS), or Snowmelt (SM) Triggered a Debris Flow Event

	Criteria		Threshold percentile (threshold value)	Trigger-type class
C1	Positive 2-day soil moisture gradient	>	20–80 (0.16–4.80%/day)	LLR
C2	Negative 3-day potential evapotranspiration gradient	>	20–50 (0.03–0.12 mm/day)	LLR
C3	Event-day temperature span	≤	10–50 (2.20–8.40 °C)	LLR
C4	Negative 2-day soil moisture gradient	>	20–80 (0.16–2.16%/day)	SDS
C5	3-day mean potential evapotranspiration	>	85–95 (2.83–3.96 mm/day)	SDS
C6	Event-day temperature span	>	value from C3	SDS
C7	Event-day snowmelt	>	93–97 (5.51–25.64 mm/day)	SM

due to continuous rainfall. At the same time, it is likely that evapotranspiration rates drop as a consequence of reduced solar energy input (persistent cloud cover), elevated partial vapor pressure, and maximum temperatures close to the dew point (Ahrens, 2008; Eccel, 2012; Haeckel, 2016). On the other hand, SDS events are convective phenomena perceptible by decreasing soil moisture in the time leading up to the event, due to comparatively high energy input for evapotranspiration (Ahrens, 2008). Contrary to LLR events, the maximum temperature typically lies far away from the dew point temperature, as a result of the short formation time and duration of SDS events and the frequently associated shorter duration of significant cloud cover. Finally, intensive SM is associated with above-freezing temperatures and the presence of a snow cover.

Based on the above, we formulated three individual criteria for the respective classification of LLR and SDS and one criterion to capture SM events. The selection of the variables used as criteria was based on an iterative, exploratory approach, guided by the general, simplified pattern described above. Importantly, the actual threshold values for these criteria were not arbitrarily defined a priori but sampled from a uniform distribution within a range of respective percentiles that were selected from an explorative-iterative modeling process, guided by the outcomes of the probability analysis. An overview of the criteria C1–C7 including the respective absolute values associated with the above percentile ranges are given in Table 1.

The classification itself was done according to the following procedure: if criterion C7 was met, the trigger-type SM is assigned. Otherwise, the event was classified either as LLR or SDS, depending on which trigger met more of the respective criteria. When a clear classification was not possible, we used the event-day temperature span (C3 and C6) as key criteria. This procedure was repeated 1,000 times, each time sampled from a uniform distribution of respective parameter ranges. The trigger mechanism assigned to each debris flow event was the most frequent mechanism identified in the preceding analysis.

It is important to note that we did not use observed station precipitation for the above trigger-type identification. Instead, we tested the trigger class assignment for plausibility by comparing it to the station precipitation on the event days for the respective classes. To test whether the classified events emerge from different populations as evidence for distinct trigger regimes, we applied the nonparametric Kruskal-Wallis test (Kruskal & Wallis, 1952). Finally, we compared our classification with a qualitative description of circulation pattern and general weather conditions in Austria provided by federal research institute ZAMG, which is available back to the year 1999 (zamg.ac.at).

3.5. Determination of Temporal Debris Flow Susceptibility

We define temporal susceptibility for debris flow occurrence in a hydrometeorological and therefore temporally variable sense; that is, we assume that the hydrological history of the catchment affects the critical water input needed to trigger a debris flow. Compared to earlier work (cf. Cardinali et al., 2000; Ciavolella et al., 2016; Crozier, 1999) and building on the above classification into distinct trigger types, we go a step further and differentiate between susceptibilities for different catchment states and potentially triggering rainfall. We developed a probabilistic tool for predicting the susceptibility for each day of the study period and tested whether this method improves the prediction of days susceptible to debris flow events compared to rainfall-only models. The tool is based on the Naive Bayes classifier method, which calculates the relative probability that a certain day during the study period belongs to a certain trigger class, based on different predictor variables. The calculated probabilities cannot be interpreted as debris flow occurrence probabilities, since for a certain trigger type the method does not include the days where this trigger type prevailed but did

not lead to a debris flow. In other words, the Bayes classifier only identifies combined catchment states and rainfall events that are similar to the catchment state and rainfall on days with observed debris flows. These days have then a high hydrometeorological susceptibility for debris flow occurrence.

The Naive Bayes classifier is given by equation (3) (Heiser et al., 2015; John & Langley, 1995; Pérez et al., 2009; Tsangaratos & Ilia, 2016):

$$P(c_j|x_1,\ldots x_n) = \frac{P(c_j)*\prod_{i=1}^{N}P(x_i|c_j)}{\sum_{j=1}^{K}P(c_j)*\prod_{i=1}^{N}P(x_i|c_j)} \quad (3)$$

where $P(c_j|\, x_1, \ldots x_n)$ is the probability that a signature of the catchment state, described by N hydrometeorological predictor variables $x_1, \ldots x_n$, corresponds to the trigger class c_j. The prior probability $P(c_j)$ is the probability for a given trigger class c_j to occur, regardless of the catchment state. $P(x_i|\, c_j)$ describes the likelihood that the magnitude of predictor variable x_i was observed in connection with a debris flow event from that trigger (or nontrigger, NT) class c_j. The multiplication of single likelihood quantities relies on the *naive* assumption that the predictor variables are conditionally independent from each other. The denominator acts as a normalizing constant, which ensures that the determined probabilities for the K trigger classes integrate to unity. The highest trigger (or NT) class probability conditional on a given catchment state $P(c_j|\, x_1, \ldots x_n)$ then provides an estimate whether the prevailing hydrometeorological conditions on a certain day are similar to catchment states when debris flows occurred in the past.

In addition, to reduce the potentially high number of false positives, a probability threshold is introduced. In other words, only when the probability of a specific trigger class $P(c_j|\, x_1, \ldots x_n)$ exceeds the chosen threshold, the respective day is assigned to a certain trigger class. Otherwise, that day is classified as NT. In absence of more detailed information, a range of different thresholds between 55% and 95% were used in this study to explore and quantify the sensitivity of the results to the choice of thresholds.

Four Naive Bayes classifier models of different complexity were set up for the determination of the hydrometeorological susceptibility for debris flows. The simplest model we tested (M^{prec}) uses the observed event-day precipitation as single predictor variable to discriminate between trigger classes debris flow and NT. Model M^{met} corresponds to M^{prec} but additionally uses observed mean air temperature to account for the findings of Stoffel et al. (2011) that debris flow occurrence is connected to air temperatures in the high percentile range. In contrast, models M^{hmet2} and M^{hmet3} differentiate between multiple trigger classes instead of a single class debris flow by using additional hydrometeorological predictor variables. M^{hmet2} accounts for trigger-type classes LLR, SDS, and NT. Model M^{hmet3} additionally accounts for trigger-type class SM. To describe the characteristics for the trigger classes (which are a combination of catchment state and rainfall input) as holistically as possible, we used the following set of predictor variables for both hydrometeorological models M^{hmet2} and M^{hmet3}: (1) effective precipitation (mm/day), (2) SM (mm/day), (3) soil moisture at the beginning of the day (−), (4) mean temperature (°C), (5) temperature span (°C), (6) soil moisture gradient to the previous day (day^{-1}), (7) soil moisture gradient between the first and the second previous day (day^{-1}), (8) soil moisture gradient between the second and the third previous day (day^{-1}), (9) mean potential evapotranspiration over three prior days (mm/day), and (10) mean evapotranspiration deficit over three prior days (mm/day).

The selection of the above variables describing the catchment state was guided by an exploratory-iterative approach and by the necessary condition for using a Naive Bayes classifier approach that allows only a weak correlation between the chosen predictor variables as this may influence the classification performance (e.g., Chawla, 2009). Here the highest correlation between individual predictor variables was computed between mean temperature and soil moisture gradient with an R^2 of 0.76, followed by a R^2 of 0.42 for mean potential evapotranspiration and mean evapotranspiration deficit over three prior days, while all others did not exhibit statistically significant correlations.

All four classifier models were trained and evaluated with independent data sets following a sequential split sample procedure, similar to k-fold cross validation (Priddy & Keller, 2005). Ideally, for applying the Naive Bayes classifier method, the members of each class should be equally represented. In our case this would mean that we would have to use only 41 nonevent days, which may not represent the distribution of catchment states over our study period. This is a general problem when using machine learning approaches and

might influence the predictive performance (Chawla, 2009). As a compromise we increased the number of members in the class *nonevent days* by a factor of five (5 times 41 = 205) to better represent the distribution of catchment states.

For training, we used around half of each data set, that is, 22 catchment states of days with debris flow occurrence and 103 nonevent days. Additionally, we ensured that for model M^{hmet3} each trigger class is represented by at least four elements to enable the fit of the likelihood probability function $P(x_i|\, c_j)$. As validation data set we used the remaining 19 and 102 catchment states, respectively. All elements were randomly sampled across all precipitation zones. We repeated this procedure 1,000 times to consider uncertainties by varying the model training data sets. The training stage of the models includes on the one hand the determination of the prior probability $P(c_j)$ and on the other hand the likelihood $P(x_i|\, c_j)$.The latter expresses the probability that a magnitude x_i of predictor i is observed with class c_j, for which we fitted a normal distribution to the sampled training data. However, for precipitation we applied a general extreme value distribution (McFadden, 1978) since it allowed a much better description of the data (e.g., Bayesian information criterion 336.6 for general extreme value distribution vs. 377.5 for a normal distribution for event-day station precipitation).

The remaining 19 debris flow event days and further 103 sampled nonevent days for each of the 1,000 model realizations were in the following used to evaluate the model's predictive performance by comparing the predicted debris flow trigger (and NT) classes with actually observed debris flows in a binary way, that is, debris flow occurrence on a given day (yes/no). To quantify and summarize the model skill, receiver operating characteristic (ROC) statistics of true positive rates ($TPR = \frac{TP}{TP+FN}$, often termed *sensitivity*) and false positive rates ($FPR = 1 - TNR$, where $TNR = \frac{TN}{TN+FP}$ is the true negative rate and also known as *specificity*) were used (Fawcett, 2006). TP is the number of days for which a debris flow was predicted and a debris flow was actually observed (independent of the trigger class). FN expresses the number of false negative predictions, that is, the days for which the model failed to predict a high temporal susceptibility even though a debris flow was observed. Similarly, FP accounts for false positives; that is, a high debris flow susceptibility was predicted but no event observed, while TN are the true negatives, that is, days for which a debris flow was neither predicted nor observed. A perfect model yields unity for TPR and a value of 0 for FPR, which corresponds to an area under curve (AUC) value of 1 when plotting these variables in the 2-D parameter space. The advantage of the AUC value is that model performance can be expressed by a single scalar (Fawcett, 2006).

To test whether our susceptibility models represent an improvement for the assessment of debris flow occurrence, we compare them with a classic I-D curve approach. To allow a fair comparison, we create specific I-D curves for our study region from the elements of the training data set and test their performance on the validation data set, analogs to the procedure of evaluating the susceptibility models. These 1,000 realizations of I-D curves were created by following a percentile regression approach of using the 10th percentile threshold (see Guzzetti et al., 2007, 2008; Saito et al., 2010). For the validation we used only rainfall, what means that an observed station precipitation below 0° air temperature was set to 0 and assumed to occur as snowfall. The duration was determined by counting the number of consecutively days with a rainfall intensity higher than 1 mm/day.

4. Results

4.1. Hydrometeorological Conditions and Signatures of Catchment States

The likelihood-based calibration approach for the hydrological model resulted, in postcalibration evaluation, in comparatively robust performance metrics, with NSE = 0.78 of the best performing model (0.69/0.78; 5th/95th percentile), NSElog = 0.83 (0.77/0.85), VE = 0.76 (0.70/0.78), and NSEFDC = 0.92 (0.84/0.96), indicating a meaningful representation of the system-internal processes and thus a plausible model formulation and parameter selection. Figure 4a exemplarily shows the model results for rainfall, snowfall, SM, soil moisture, and runoff for the year 1999 (including three debris flow event days) at gauge Litzbach. Some deficits exist in mapping the SM process, which can be even more distinct in other years than shown and which is likely the consequence of the simple degree-day melt model for snow processes (Hock, 2003) and considerable uncertainties in wintertime precipitation observations, which can make up to 30% in the study region

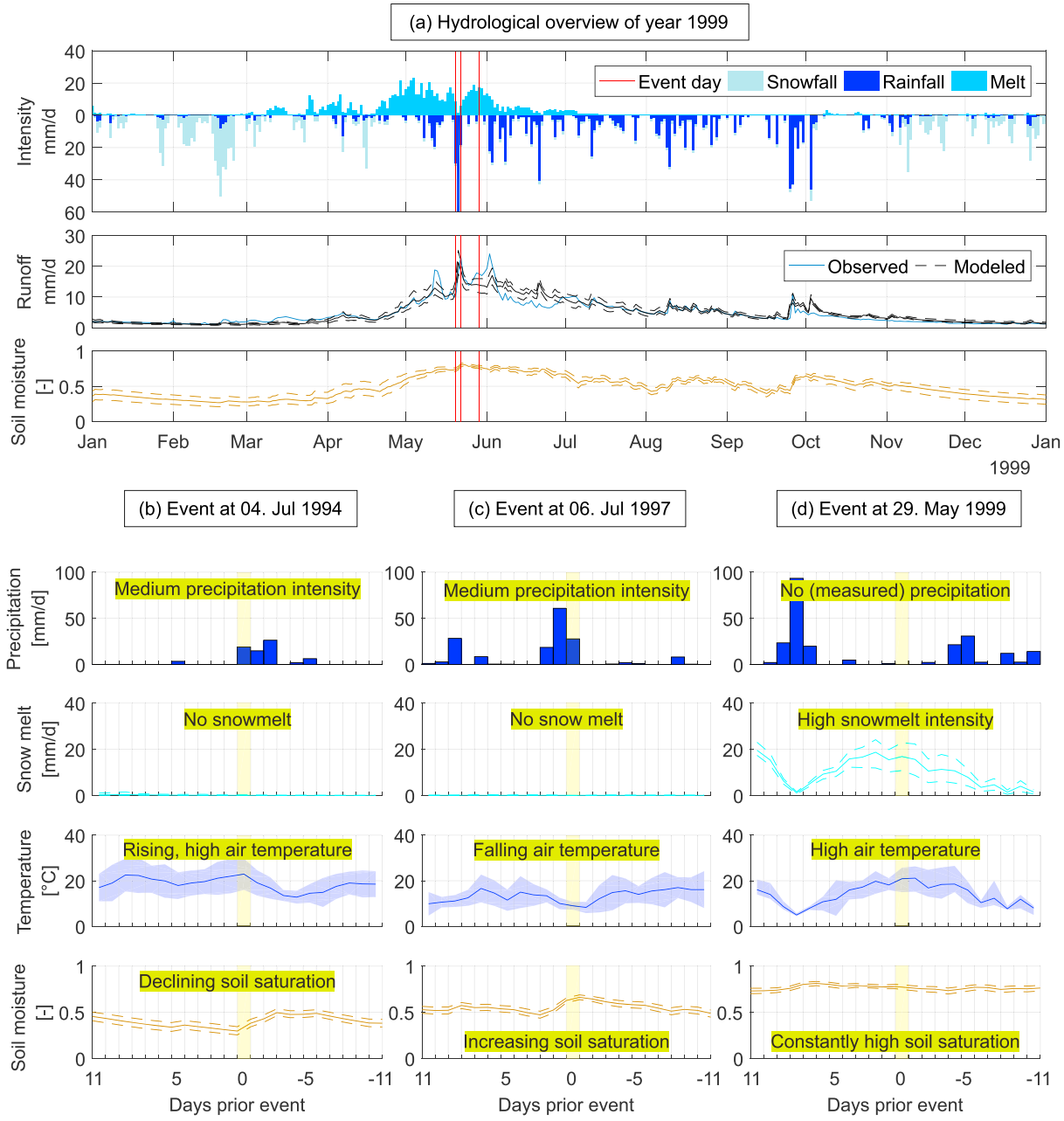

Figure 4. Example of the observed and modeled hydrometeorological variables for the year 1999 of the Litzbach, including station precipitation, snowfall, snowmelt, soil moisture, and modeled and observed runoff. The dashed lines account for the 5th and 95th percentile uncertainty bounds. The red vertical lines represent the days on which debris flows occurred (a); precipitation, snowmelt, temperature (daily minimum, maximum, and mean), and soil moisture including uncertainty bounds around the debris flow event on 4 July 1994, which may be interpreted as short-duration storm (b), on 6 July 1994, which suggests being a long-lasting rainfall (c) and on 29 May 1999, which may be triggered by intense snowmelt (d).

(Parajka et al., 2005). Nevertheless, the overall good evaluation of the model strengthens our confidence in the applicability for further analysis.

In Figures 4b–4d we document, in an illustrative example, different hydrometeorological conditions around debris flow event days, which are shaded yellow. In the first example (Figure 4b) we show the catchment state around the debris flow on 4 July 1994, where the observed precipitation reached only 19 mm on the event day. On the days before, the air temperature as well as temperature span increased while modeled soil moisture decreased due to evapotranspiration, pointing toward conditions typical for a SDS. A different

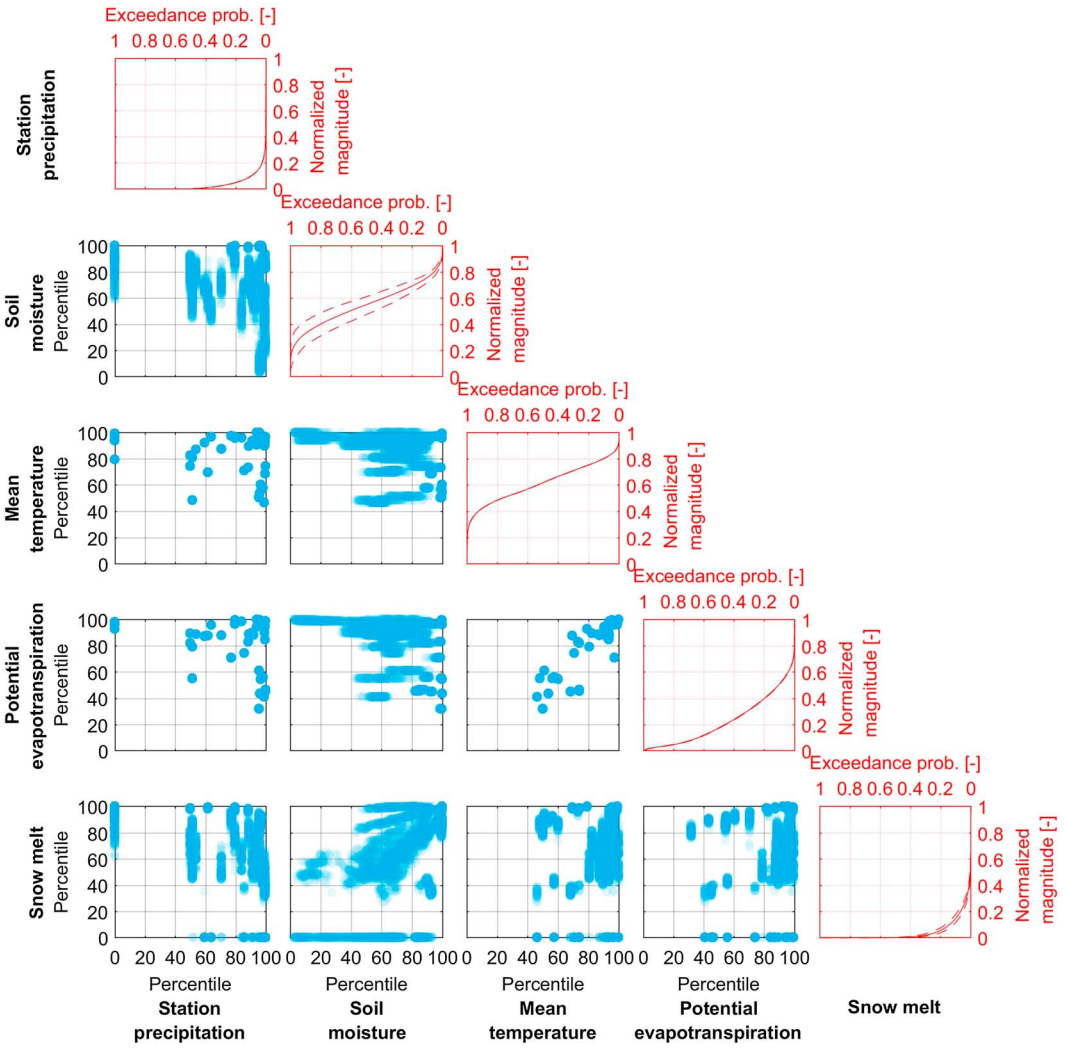

Figure 5. Selected hydrometeorological variables for the 41 event days across all precipitation zones in the Montafon region including the model uncertainties (blue shaded symbols). The red diagonal plots represent median exceedance probabilities including uncertainty bounds (5th and 95th percentiles as dashed lines) of the respective normalized parameter values at the event days.

system state emerged for the debris flow event on 6 July 1997, where a prior soil moisture increase due to continuous precipitation was measured while air temperature and temperature span continuously decreased, suggesting a LLR as trigger (Figure 4c). In the third example, the debris flow event on 29 May 1999, the role of SM is important, as continuous warm temperatures induced high SM rates and increasing soil moisture. Note that no precipitation was observed, neither on the event day itself nor on the days before.

A comprehensive overview of the catchment states on the 41 event days is displayed in the scatter matrix in Figure 5. The selected parameters include station precipitation and mean air temperature, as well as modeled soil moisture, potential evapotranspiration, and SM. Since the calculations were carried out for each precipitation zone individually, the values are plotted in the dimensionless percentile representation to allow a better comparison. Following the diagonal plots in Figure 5, representing the cumulative frequency of the scaled variables at the days of debris flow occurrence, we find that no precipitation was measured for about half of the time in the study region including four debris flow event days. For the other events, precipitation on event days spreads widely over the parameter space. For 13 event days precipitation below the 70th percentile (~2.8 mm) was observed. For some debris flow events modeled SM was 0 (rather independent of precipitation or soil moisture), and for some we find SM values above the 40th percentile. There is, similarly, a clustering of debris flows at temperatures (and related actual evapotranspiration) above the 40th

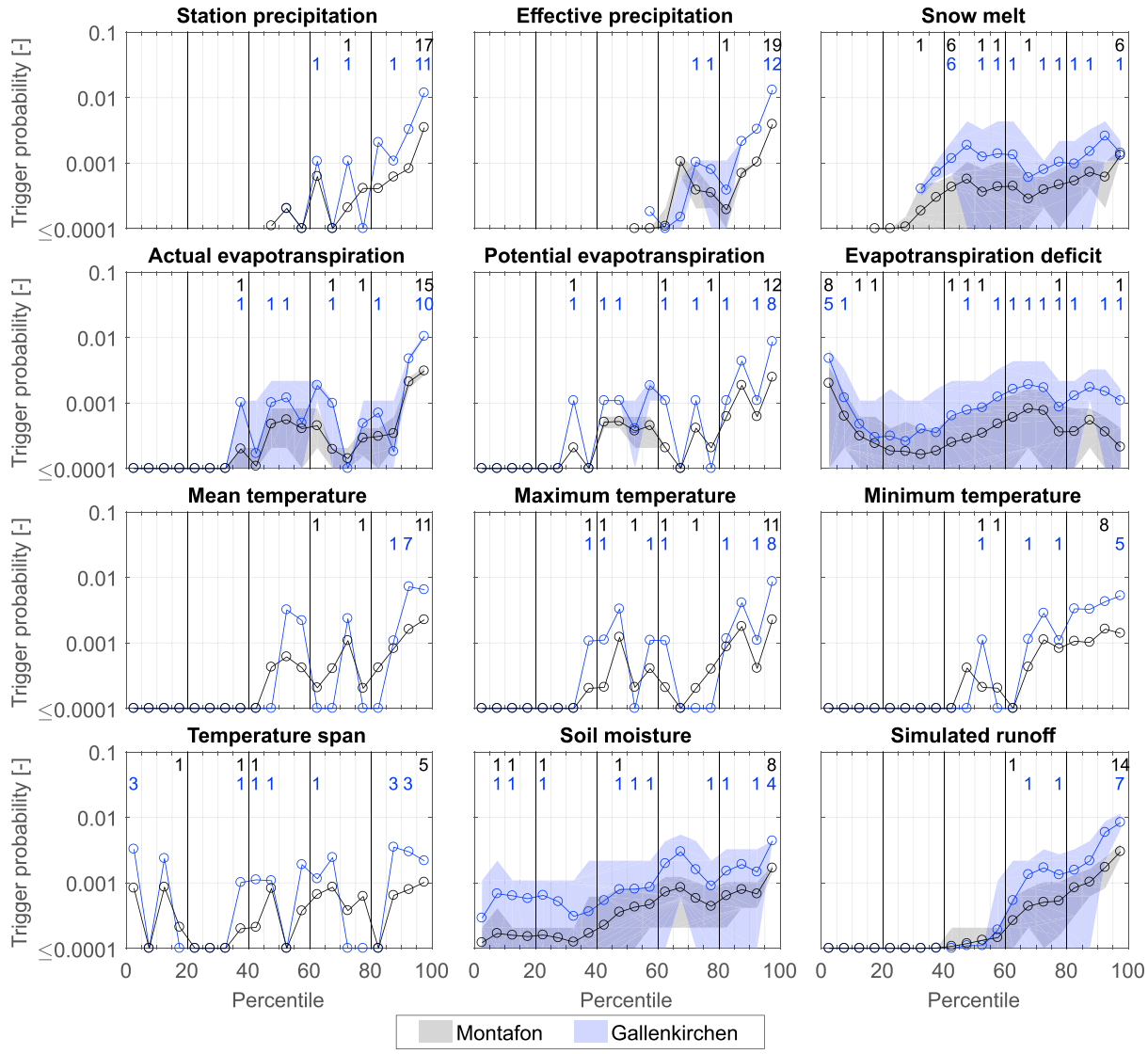

Figure 6. Trigger probabilities $P(E|x_i)$ to observe a debris flow event E conditional to magnitude x_i across the parameter's i percentile value range for selected hydrometeorological parameters. The gray lines represent the comprehensive data set from all precipitation zones on the study region Montafon, and the blue lines exemplarily display the precipitation zone Gallenkirchen, where the majority of the debris flow events occurred. The shaded areas represent the 5th and 95th percentile uncertainty bounds of the hydrological model. Variables station precipitation, air temperature (mean, maximum, and minimum) and temperature span (difference between maximum and minimum air temperature) were measured. All other variables were simulated. Effective precipitation represents elevation corrected precipitation, which is not intercepted. The evapotranspiration deficit is the difference between the potential evapotranspiration and the actual evapotranspiration. The numbers at the top of each plot designate the median number of events $N_{x_i|E}$ in a bin and are only plotted when there is on median exactly one event in a bin as well as when a bin contains on median the maximum number of events.

percentile. Soil moisture on event days widely spreads over all percentile ranges, which does not support the notion that debris flows occur only at high antecedent moisture conditions. In summary we find for most variables a wide scatter over the parameter space and no obvious grouping of data. We attribute this spread to the presence of different meteorological conditions resulting in different hydrometeorological signatures, which have to be separated before a clearer picture emerges.

4.2. Trigger Probabilities Conditional on Different Hydrometeorological Variables

Figure 6 shows the trigger probabilities $P(E|x_i)$ conditional on the magnitudes of different hydrometeorological variables on the associated event days for the region Montafon, as well as exemplarily for the precipitation zone Gallenkirchen, which experienced the highest number of debris flows. For a better interpretation of the trigger probabilities, we plotted the median number of events $N_{x_i|E}$ in a bin at the top of each plot when

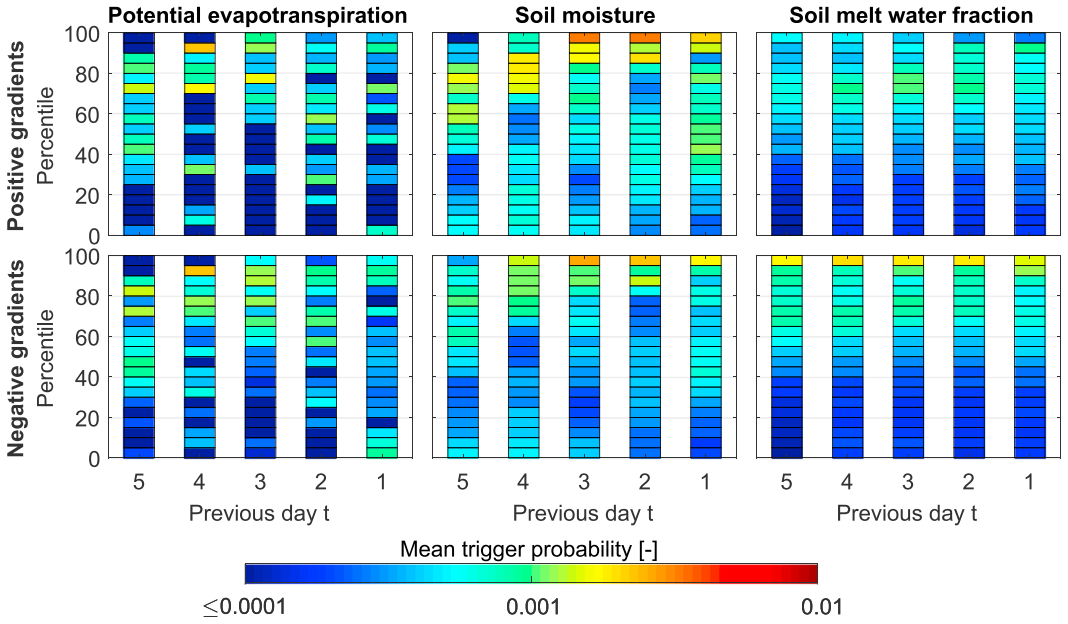

Figure 7. Mean trigger probability $P(E|x_i)$ to observe debris flow event E conditional to a gradient of variable x_i over t days. A positive gradient means that the parameter value increases over previous t days; a negative gradient means that the parameter value decreases over previous t days. (The soil melt water fraction expresses the fraction of melt water to the total water in the soil.)

there is on median exactly one event in a bin as well as when a bin contains on median the maximum number of events. We find that probabilities mostly show an increasing trend with increasing variable values but do not exclusively peak only at the maximum of the percentiles. For example, we observe two probability maxima for recorded station precipitation and effective precipitation at the percentile range 60–70 and above 95. SM and actual evapotranspiration also show two maxima around the percentiles 50 and again above 95. The evapotranspiration deficit, calculated as the difference between potential and actual evapotranspiration, shows a maximum in the lowest percentile range and peaks again at around the 70th percentile. The minimum and maximum daily temperature exhibit similar trends and peak around the 50th and above the 95th percentiles. The mean air temperature shows even three distinct peaks around percentiles 50, 70, and above 95. Alternating highs and lows are visible for the event-day temperature span (difference between event-day maximum and minimum temperature), which may indicate a low-pressure system at small differences and the disposition for convective rain at high differences. Debris flow trigger probability with respect to soil moisture content generally increases with increasing moisture and also show two elevated values in the low percentile range and around the 70th percentile, indicating a low sensitivity to increased pore pressure in the hillslope. A steady increase of trigger probabilities is observed with increasing runoff. These scattered pattern importantly show that there is no single typical catchment state when debris flows are triggered. In the following we try to go a step further and look how changes of hydrometeorological parameters prior to debris flow events may tell a more consistent story.

In Figure 7 we show trigger probabilities for debris flows conditional to gradients of potential evapotranspiration, soil moisture, and soil melt water fraction up to 5 days prior to the event (always with respect to the event day). For a better overview, we separated the analysis into the subsets of positive and negative absolute gradients. At the first sight, we obtain a rather scattered pattern with locally high probabilities in the positive as well as the negative gradient domain. For potential evapotranspiration we find increased trigger probabilities when positive and negative gradients 3 days before the event are in the high percentile range (i.e., evapotranspiration [solar input] either dramatically increases or decreases). At the same time, trigger probabilities increase when soil moisture gradients increase toward increasing positive and negative gradients. An extreme increase of soil moisture above 80th percentile 2 days before the event shows the most compact high trigger probability block, ranging from 0.1% to 0.3%. Taken together, this indicates that at least two types of catchment states start to develop around 3 days before the event occurs. Strongly decreasing

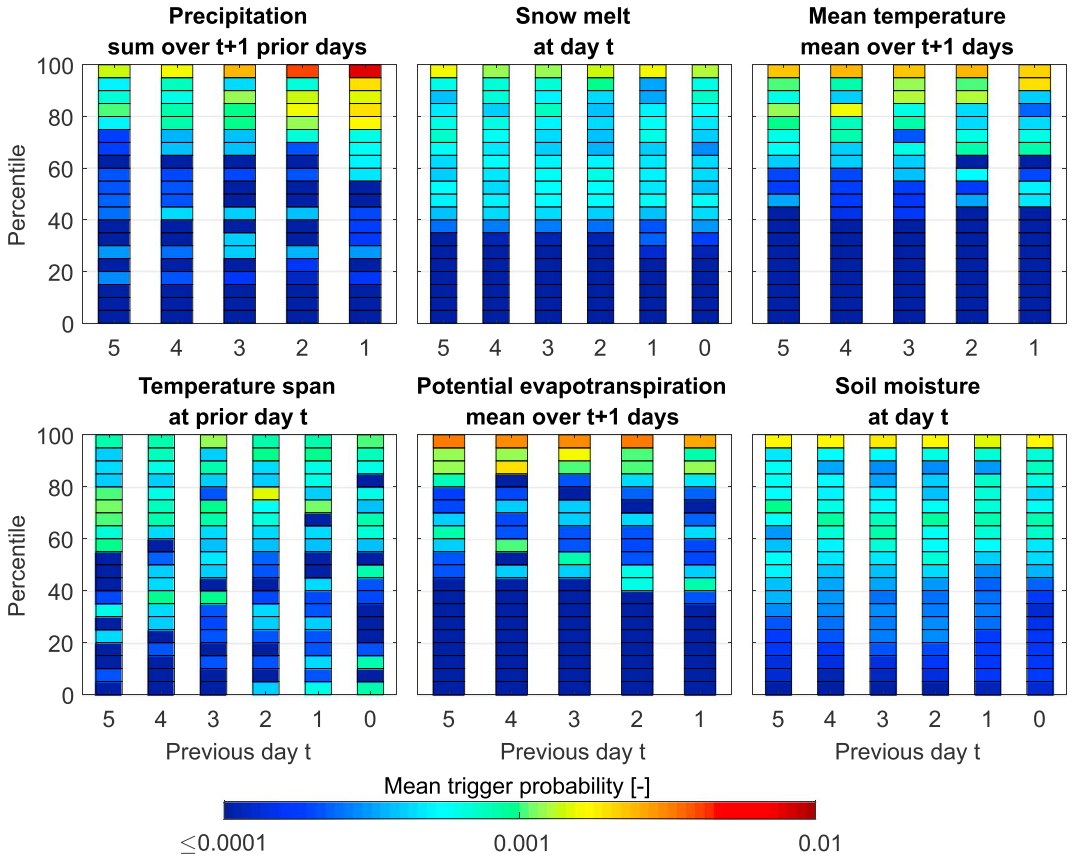

Figure 8. Mean trigger probabilities $P(E|x_i)$ to observe a debris flow event E conditional to a magnitude x_i of selected hydrometeorological parameter *i*. Parameter *i* represent either a moving sum over *t* days, a moving average over *t* days or just the parameter at previous day *t*.

evapotranspiration and increasing soil moisture point toward a LLR, while increasing evapotranspiration and decrease of soil moisture point toward a SDS. The results for gradients of soil melt water fraction (i.e., the fraction of melt water to total water in the soil) show generally higher probabilities when soil melt water fraction decreases. We interpret this with the fact that most of the debris flow events occurred outside of the SM season where the initially almost saturated soil due to melt water steadily gets drained as well as mixed by rain water. Therefore, a decreasing fraction of melt water in the soil is observable during most of the debris flow season.

In Figure 8 we show trigger probabilities $P(E|x_i)$ up to 5 days prior to the event for actual, cumulative, and average values of selected parameters. As expected, we find increased probabilities at the high percentile range of most variables. The highest mean trigger probability of almost 1% is obtained when the sum of station precipitation at the event day and one day prior exceeds the 95th percentile; that is, rainfall is the most important factor for triggering debris flows in our study region. However, also high percentile ranges of averaged mean and maximum temperatures, actual and potential evapotranspiration increase the probability. Regarding daily temperature span, we find increased probabilities in the high and in the low percentile range. In our interpretation, rainfall is the most important but not the only factor for triggering debris flows. For example, SM and soil moisture can also play a role. Interestingly, trigger probabilities for intensive SM (above the 90th percentile) are higher at prior event days than on the event day itself. This may be a key for future analysis to separate between rain-on-snow events and purely SM-triggered events, since the melt rate may be smaller due to decreased air temperatures in attendance of rain than without.

In summary, these sometimes opposing trends of increased debris flows probabilities conditional to different hydrometeorological conditions go in line with our hypothesis that different weather conditions are connected with debris flows occurrence in our study region. In the following section we use the

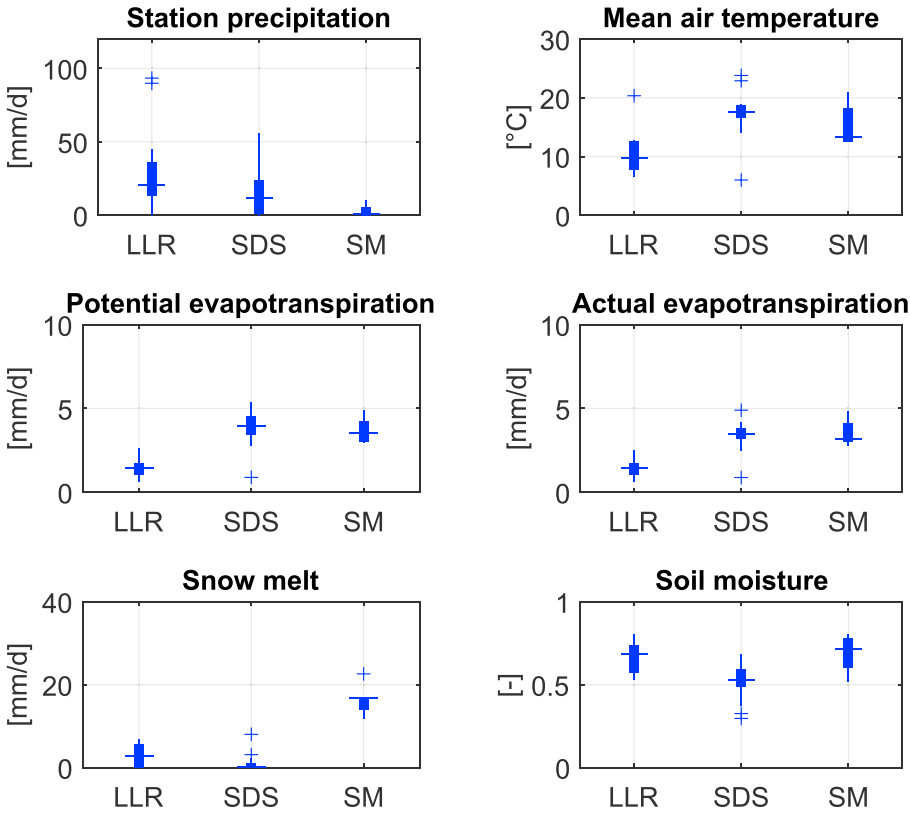

Figure 9. Distribution of event-day precipitation (which was not used as classification criteria), mean air temperature, potential evapotranspiration, snowmelt and soil moisture at event days for each trigger class (LLR = long-lasting rain; SDS = short-duration storm; SM = snowmelt). The median is marked by a minus symbol, boxes cover the 25th and 75th percentile ranges, and outliers are marked as a cross symbol.

quantitative information to derive different trigger types using a simplified hydrometeorological classification scheme.

4.3. Identification of Debris Flow Trigger Types

Based on the procedure described in section 3.4, data suggest that debris flows in the Montafon region were triggered on twelve days by LLR, on 23 days by SDS and on 6 days by SM (see Table S2 of the supporting information), consistent with the event days exemplarily shown in Figure 4. Note that using this technique allows us to identify only the most likely trigger; however, we do not identify distinct threshold values for the different criteria.

Our simplified classification based on the interpretation of the hydrometeorological parameters of all debris flow event days that were observed in this period generally reflects well the description of the general weather conditions federal research institute ZAMG (Table S2). Additionally, we used the observed event-day station precipitation to check whether the observed triggering rainfall was significantly different for our trigger classes, as we would expect that LLR events responsible for debris flow initiation to show a high value of daily rainfall and are well captured by the station network. On the other side, local SDS events might have lower daily rainfall sums (but higher intensities which we do not measure on a daily basis) and might not be registered by the station network. Our analysis shows that the median daily precipitation for the LLR class with 20.3 mm/day was significantly higher than for SDS with 12.0 mm/day (Figure 9). For the trigger class SM only 0.6 mm/day were recorded by the station network, indicating that only a very low rainfall input is needed for triggering debris flows when SM is high. We performed the Kruskal-Wallis test (Kruskal & Wallis, 1952) also for variance and skewness for each rainfall class find that the group members emerge from different populations ($p < 0.01$), strongly supporting the notion that different trigger types can be found in our study region. Three distinct regimes were also found for the mean air temperature, with the highest

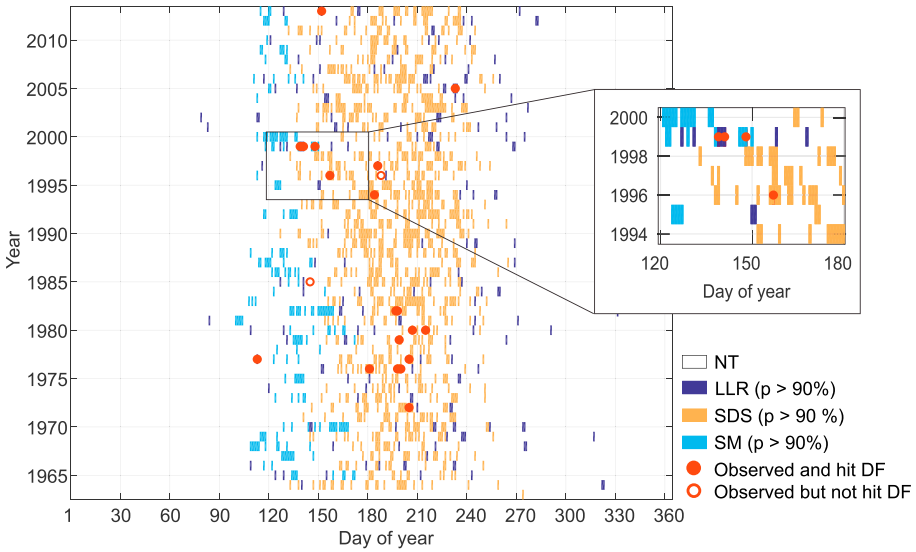

Figure 10. Daily predicted trigger classes marking a temporal debris flow susceptibility exemplarily for the precipitation zone Gallenkirchen from 1963 to 2013 based on model M^{hmet3}. A day is colored according to its trigger class c_j (NT = no trigger; LLR = long-lasting rain; SDS = short-duration storm; SM = snowmelt) limited to a probability $P(c_j|\, x_1, \ldots x_n)$ of more than 90%. Filled red circles indicate observed debris flows that were correctly predicted (true positives), and open red circles indicate observed debris flows that were not predicted (false negative).

median for the SDS trigger (17.5 °C), followed by SM (13.25 °C), and LLR (9.8 °C). Initial median soil moisture in a precipitation zone was highest when SM triggered the event (0.72), slightly drier at LLR trigger (0.68), and its lowest saturation came along with the SDS trigger (0.53).

4.4. Prediction of Temporal Hydrometeorological Susceptibility

The four Bayesian classifier models, M^{prec}, M^{met}, M^{hmet2}, and M^{hmet3}, were used to predict days with a high temporal hydrometeorological susceptibility for debris flow occurrence for the complete study period. Figure 10 exemplary visualizes the daily predicted susceptibility for the model M^{hmet3} for the subregion Gallenkirchen (which showed the highest number of observed debris flow events) between 1963 and 2013. A day is colored when a susceptibility trigger class c_j (LLR, SDS, and SM), conditional to the prevailing hydrometeorological state $x_1, \ldots x_n$yields a probability $P(c_j|\, x_1, \ldots x_n)$ of more than 90%. The results show that the predicted days of high susceptibility cluster well within the typical debris flow season between April (about day of year 90) and October (about day of year 270), as reported by many authors (e.g., Badoux et al., 2009; Stoffel et al., 2011; Szymczak et al., 2010). During summer months, days with SDS as potential trigger are dominating. Critical LLR occur more preferably offset the hot season. A high susceptibility for a SM trigger is mostly until the end of May. The distribution parameters for the likelihood $P(x_i|\, c_j)$, giving the probability that a magnitude of state variable x_i is connected to trigger class c_j, is found in Tables S3 (for M^{prec} and M^{met}) and S4 (M^{hmet2} and M^{hmet3}) in the supporting information.

The performance of the four models was evaluated based on the ROC curve formed by true positive rate (TPR) and false positive rate (FPR) computed by comparing the model predictions with the subsets of data not used for model training. Curve vertices (each defined by a TPR and FPR couple) are obtained by stepwise decreasing the probability threshold, which had to be exceeded that a predicted trigger class actually counts as final predicted class for a state. If the actual probability for a trigger class is below this threshold, it is counted to be member of the no-trigger class. The initial threshold starts at a value of 95% and ends at relative majority. A perfect model would plot in the top left corner, which means that both, event days as well as nonevent days, were recognized and predicted as such. As a measure for model performance we use the AUC number.

Figure 11 shows the ROC curves of the four classifier models. We find that the hydrometeorological models M^{hmet2} and M^{hmet3} perform better than models M^{prec} and M^{met}, which only rely on meteorological data. In addition, the uncertainty appears lower for the more complex models, which primarily comes from the sampling of nonevent days for the models training and validation data set. The model M^{hmet3}, which

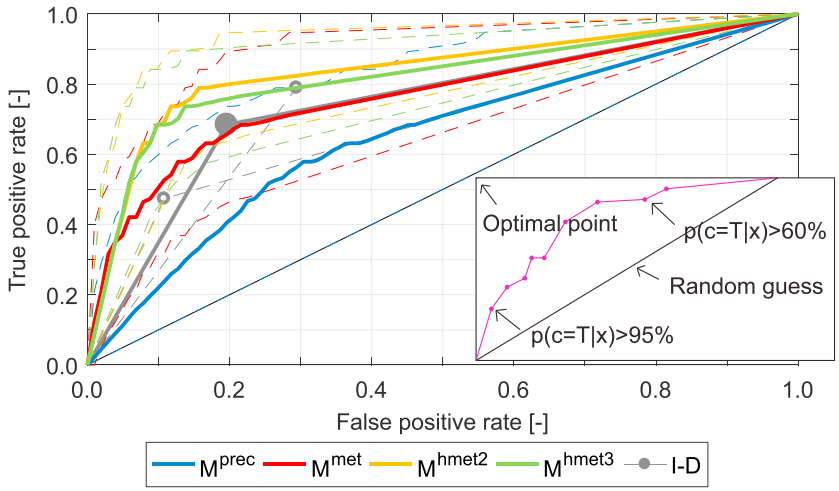

Figure 11. Prediction performance of the different classifier models M^{prec}, M^{met}, M^{hmet2}, M^{hmet3}, and specifically established intensity-duration curves I-D for the region visualized in the receiver operating characteristic space. True positives rate refers to the proportion of correctly predicted susceptibility (debris flow event actually occurred in reality); false positives rate refers to the proportion of falsely predicted susceptibility were no debris flows event observed (false alarm). Thick solid lines represent the median values, and the dashed lines show the 5th/95th uncertainty bounds of the 1,000 Monte Carlo realizations. For I-D curves median (full circle), 5th and 95th percentiles (open circles) of the false positive rate-true positive rate value couple are shown.

susceptibility assessment differentiates between triggers LLR and SDS, performs slightly better than the three trigger class model M^{hmet3} which additionally uses SM, although they use the same predictors (AUC 0.84 for M^{hmet2} vs. 0.82 for M^{hmet3}). In contrast, M^{prec} and M^{met} perform less with AUC values of 0.65 and 0.77, respectively. Also, in terms of TPR and FPR hydrometeorological models are advantageous. For a class assignment based on the relative majority, the model M^{hmet2} achieves a median TPR of 0.78 while having a FPR of 0.14. The slightly better performance of the three-class model M^{hmet2} compared to the four-class model M^{hmet3} may be connected to a limited data availability for SM triggers the training data set. A detailed summary of the performance statistics for each model is given in Table 2.

We find that AUC values for the I-D curves are in the same range as the susceptibility models M^{prec} and M^{met}. The median I-D curve performance shows a TPR rate in the range of M^{met}, but this is at expense of a sufficient FPR. Both hydrometeorological susceptibility models M^{hmet2} and M^{hmet3} show higher AUC measures. We conclude that use hydrometeorological assessment for debris flow occurrence has benefits.

Table 2
ROC Statistic Model Performance Summary of Median (5th/95th Percentile) TPR, FPR, and AUC Number for the Independent Validation Data Sets

Model	TPR	FPR	AUC
M^{prec}	0.48 (0.00/0.95)	0.25 (0.00/0.51)	0.65 (0.50/0.83)
M^{met}	0.67 (0.42/0.95)	0.17 (0.05/0.31)	0.77 (0.66/0.89)
M^{hmet2}	0.78 (0.58/0.92)	0.14 (0.07/0.23)	0.84 (0.75/0.92)
M^{hmet3}	0.73 (0.53/0.89)	0.14 (0.07/0.23)	0.82 (0.72/0.91)
I-D	0.68 (0.47/0.79)	0.20 (0.11/0.29)	0.73 (0.66/0.80)

Note. For TPR and FPR we use the relative majority as a threshold that a predicted debris flow trigger class (DF, LLR, SDS, or SM) counts for a prevailing temporal debris flow susceptibility. AUC value bases on curve vertices stemming from a varying probability threshold for a predicted trigger class that had to be exceeded to count as susceptibility beginning from 95% to relative majority. Additionally, ROC statistic of specifically (from the training data sets) generated rainfall intensity-duration curves were used as threshold between trigger and nontrigger. ROC = receiver operating characteristic; TPR = true positive rate; FPR = false positive rate; AUC = area under curve; DF = debris flow; LLR = long-lasting rainfall; SDS = short-duration storm; SM = snowmelt.

Certainly, the shape of I-D curves has a major influence on the performance for debris flow detection and is a known challenge in the community. (e.g., Gariano et al., 2015; Leonarduzzi et al., 2017). Usually an I-D threshold curve becomes more stable when more events are available for computation (Peruccacci et al., 2017). The I-D curves generated in this study were based on 22 event rainfalls only, which is a comparably small number and therefore large uncertainties remain. Nevertheless, for the initiation of intensive bedload transport, Badoux et al. (2012) report of a FPR around 40% when I-D curves were fitted in such a way, that TPR is maximized. Leonarduzzi et al. (2017) note a lack of data consistency because rain gauges are usually several kilometers away from the triggering location of debris flows. Further, they mentioned the necessity of excluding debris flow events when the trigger is different to rainfall like SM. In addition, it is common practice to exclude debris flow events from the analysis when no precipitation was measured at all, although its trigger is water related (e.g., Berti et al., 2012). Hydrometeorological susceptibility models may overcome these deficits by accessing diverse information from the catchment including the hydrological history, which may

explain their better performance. Moreover, variables like temperature, evapotranspiration, and SM, and to some extent soil moisture, are more spatially stable compared to precipitation and may give a clearer picture of the catchment state if it is prone to debris flows. Even if no station precipitation was measured (although there was one at another part of the catchment, which triggered the debris flow), hydrometeorological susceptibility models are able to compensate for that, due to pertinent signals in catchment state variables.

5. Discussion

The triggering of debris flows in Alpine regions is the result of an interplay between meteorology-hydrology and geomorphology (e.g., Borga et al., 2014; Gregoretti & Fontana, 2008; Jomelli et al., 2007; Stoffel et al., 2011). Our study focuses on the connection between weather conditions and hydrological catchment state at the time when debris flows occur. The 1-D Bayesian analysis indicates that different weather conditions that lead to different hydrological catchment states are involved for triggering debris flows in the Montafon region. It is important to emphasize that the models that we developed only predict debris flow susceptibility in a temporal, hydrometeorological sense and for a large region. We do not focus on a spatially resolved prediction of debris flow susceptibility on a hillslope or subbasin scale as demonstrated, for example, by Berenguer et al. (2015). For our study we can assume that the general geomorphic disposition, like relief gradient or shape of the subcatchments (e.g., Bertrand et al., 2013; Heiser et al., 2015) are constant over the time scale of our study period. However, other factors like limiting sediment availability, landslide activity on the hillslope, or weathering processes (e.g., Rickenmann & Zimmermann, 1993; Schraml et al., 2015) were not considered and may partially explain that the hydrometeorological susceptibility predicted by our model is high on many days that did not experience debris flows. We also have no information of the mechanism for the initiation of the debris flows in the catchment, which may range from slope failure to intensive channel erosion (Rickenmann & Zimmermann, 1993). In other words, we implicitly assume that exclusively hydrometeorological conditions control the triggering of debris flows in the study region, while in reality local geomorphic conditions are likely to be a limiting factor.

Another limitation, as in all debris flow-related studies, is the largely unobservable spatial heterogeneity of the study region. Local high rainfall or melt intensities may occur close to channel or slopes and promote debris flow initiation and are likely to remain frequently unobserved. As Beven, Almeida, et al. (2017) and Beven, Aspinall, et al. (2017) pointed out, hazard assessment is naturally connected to epistemic uncertainties, which stem from our lack of knowledge of reality. To the same extent these uncertainties can be treated as aleatory probability distributions of factors influencing the process of interest, but we have to acknowledge that we may underestimate the uncertainties associated with our analysis. For example, the uncertainties connected with a probably incomplete inventory of debris flow events in the Montafon region during the study period were not considered in our study, and might contribute to a nonoptimal performance of the susceptibility model (Gariano et al., 2015).

As outlined in section 3, we quantified and included uncertainties in the hydrological modeling and subsequent statistical analysis to a large extent by rigorously following probabilistic approaches. In addition, we tried to compensate for uncertainties stemming from the definition of the classification criteria by repeated sampling from a wide range, instead of relying on a certain threshold. While the determination of SDS and LLR is comparatively stable by relying on three criteria, the classification of SM triggered events is based on only one criterion and may indicate only certain influence on of SM, without separating between a rain-on-snow and SM-only trigger.

Both measurement deficits of precipitation, particularly of snow (Parajka et al., 2005), and the rather rudimentary described melt process by a degree-day model (Hock, 2003) may result in errors of trigger interpretation. Also the fact that the conceptual model was majorly calibrated to runoff instead of probably more relevant variables for debris flow initiation like soil moisture may affect the representation of the hydrological system. Since no soil moisture measurements were available, we solely relied on runoff measurements. Further uncertainties result from data upscaling of point measurements to a regional quantity valid for the whole precipitation zone (Beven, Almeida, et al., 2017; Beven, Aspinall, et al., 2017). Nevertheless, recorded precipitation intensity at the event days was significantly different compared among the trigger classes. This finding allows manifold conclusions. On the one hand, it acknowledges current understanding that LLR, which is a more large-scale phenomenon (and may include thunderstorms) than a locally forming SDS (Rulfová & Kyselý,

2013) and therefore enables a better capturing of the areal rainfall (Hrachowitz & Weiler, 2011). For SM triggers, we observed both rain-on-snow events and precipitation absent cases. Based on limited data, we cannot exclude that for the latter local rainfall events were just not detected. On the other hand, these different triggering rainfall pattern open the question whether precipitation as the only quantity is sufficient to predict debris flow hazard or not. This can be critical when establishing classical I-D thresholds, which are based on precipitation only.

Here we show that the hydrometeorological parameter set offers plenty of signals, which were emitted by LLR, SDS, and SM that lead to debris flows, without detailing on all possible circulation pattern. The hydrometeorological susceptibility models in Figure 11 will potentially serve as a basis for an engineering prediction tool of regional temporal debris flow susceptibility. For that forecasts of precipitation and temperature (on the time scale of one to a few days) would suffice as input for the calibrated hydrological model. However, it is to consider that when maximizing the detection rate, susceptibility models show a, for practical purposes, rather high FPR, mostly caused by predicted susceptibilities for nonevent triggering SDS. This can either mean that debris flow occurrence is controlled to a large part by other factors not considered here. Alternatively, we may have captured the characteristics for SDS trigger insufficiently, which implies the need of additional variables or a higher temporal resolution since SDS typically last only a few hours, to get a more precise representation of the catchment state.

6. Summary and Conclusions

In this study, we investigated the hydrometeorological conditions of an Alpine catchment when debris flows occur and utilized this information to capture temporal, hydrometeorological susceptibility by taking account different triggers types, namely, LLR, SDS, and SM. First, we set up a process-based, semidistributed rainfall-runoff model for the region Montafon, Austria, on a daily basis between 1953 and 2013 to gain hydrological parameters besides meteorological information of precipitation and temperature. Using these results, we calculated Bayesian trigger probabilities for various hydrometeorological parameters. To determine different hydrometeorological trigger types (LLR, SDS, or SM) for historically observed debris flow events, we formulated a set of criteria based on rational arguments and the outcomes of the Bayesian analysis. After this classification, we developed two hydrometeorological susceptibility models for debris flow occurrence using a Naive Bayes classifier. Its performance was compared with two meteorological susceptibility models as well as with simple I-D thresholds. The main finding can be summarized as follows:

1. We find that in our study region, debris flows were initiated at a wide range of meteorological and hydrological conditions, and there is no single catchment state that is typical for debris flow occurrence. Rainfall is the most important factor for triggering debris flows, but the temporal evolution of hydrometeorological variables provides insight how critical trigger conditions develop.
2. Signals from different weather conditions guided the definition of criteria that allowed to identify different trigger conditions to a given catchment state. We find that debris flows were triggered by LLR on 12 days, by SDS on 23 days, and on 6 days, SM played an important role between 1956 and 2005.
3. Observed station precipitation on the event day significantly differs depending on the prevailing trigger type. This finding may partly explain the substantial uncertainties that typically come along with I-D thresholds for debris flows triggering. Similarly, antecedent moisture conditions vary for different trigger types and may therefore, if used alone, be an unreliable indicator for temporal debris flow susceptibility in our study region.
4. Debris flow susceptibility models based on combined hydrometeorological information and a trigger-type resolved prediction, performed better than meteorological models and conventional I-D thresholds.
5. The days of predicted high temporal debris flow susceptibility fit well to the typical debris flow season in an eastern Alpine environment. The model predicts more days susceptible to debris flows, than actually occurred. We attribute this partly to the shortcomings of our modeling efforts (e.g., limited temporal and spatial resolution and catchment heterogeneity). This observation might also indicate that the local geomorphologic disposition (e.g., sediment availability) is of significant importance for debris flow initiation.

The results presented in this study provide a more in-depth view of the regional hydrometeorological conditions that lead to debris flows in an Alpine environment but excludes local geomorphological disposition that might change over time scales relevant for our analysis (e.g., sediment availability, landslide activity, or

weathering processes). Our approach includes uncertainties, makes them transparent, and reduces the dependence on precipitation as the only predictor variable for debris flow triggering. An extension of the analysis to further meteorological or hydrological variables as well as an increase of spatial and temporal resolution of the data may bring clearer pattern or more distinct signals for an upcoming debris flow initiation. The combination of real-time modeling and inclusion of forecasted temperature and precipitation may offer a tool for regional debris flow forecasting and warning.

Acknowledgments
We thank Hydrographic Services (HD) Austria and its subdivision HD Vorarlberg, Zentralanstalt für Meteorologie und Geodynamik (ZAMG) and the Voralberger Illwerke AG for supplying the climate and hydrologic data sets. These data sets are publicly available from the archives of above curators upon request: HD station data (for Gallenkirchen, Obervermunt, Partenen, Schruns, Silbertal, and Vermunt) as well as runoff data (for Litzbach and Vandans) from https://www.bmnt.gv.at/wasser/wasser-oesterreich/wasserkreislauf/hydrographische_daten/Datenbestand_Bezug.html; in German), ZAMG station data (for Partenen and Obervermunt) from https://www.zamg.ac.at/cms/en/top-menu/contact-us and from the Voralberger Illwerke via https://www.illwerke.at/. The latest digital elevation model can be downloaded from http://vogis.cnv.at/geodaten/?service=files&t=418171be46e40175bfe171edfc4097dd&path=%2FGelaendemodelle%2FHoehenmodelle. The Austrian glacier inventory can be retrieved from https://doi.pangaea.de/10.1594/PANGAEA.844983. Copernicus Corine Land Cover is accessible from https://land.copernicus.eu/pan-european/corine-land-cover/clc-1990?tab=download. The model runs were performed on the Vienna Scientific Cluster (www.vsc.ac.at), which we thankfully acknowledge. This project receives financial support from the Austrian Climate and Energy Fund and is carried out within the framework of the *ACRP* Programme. A demo script about the prediction of temporal debris flow susceptibility according to the approach presented in the manuscript is available under https://www.hydroshare.org/resource/30e5dec16c5648ee987c6d2bdd26435b/ as well as upon request from D. P. We also thank the three anonymous reviewers for their valuable comments and suggestions. There are no conflicting interests for any author to declare.

References

Ahrens, C. D. (2008). *Essentials of meteorology: An invitation to the atmosphere* (5th ed.). Belmont: Cengage Learning.

Badoux, A., Graf, C., Rhyner, J., Kuntner, R., & McArdell, B. W. (2009). A debris-flow alarm system for the Alpine Illgraben catchment: Design and performance. *Natural Hazards*, *49*(3), 517–539. https://doi.org/10.1007/s11069-008-9303-x

Badoux, A., Turowski, J. M., Mao, L., Mathys, N., & Rickenmann, D. (2012). Rainfall intensity-duration thresholds for bedload transport initiation in small Alpine watersheds. *Natural Hazards and Earth System Sciences*, *12*(10), 3091–3108. https://doi.org/10.5194/nhess-12-3091-2012

Ballesteros-Cánovas, J., Stoffel, M., Corona, C., Schraml, K., Gobiet, A., Tani, S., et al. (2016). Debris-flow risk analysis in a managed torrent based on a stochastic life-cycle performance. *Science of the Total Environment*, *557-558*, 142–153. https://doi.org/10.1016/j.scitotenv.2016.03.036

Bayes, T., & Price, R. (1763). An essay towards solving a problem in the doctrine of chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions*, *1763*, 370–418.

Berenguer, M., Sempere-Torres, D., & Hürlimann, M. (2015). Debris-flow forecasting at regional scale by combining susceptibility mapping and radar rainfall. *Natural Hazards and Earth System Sciences*, *15*(3), 587–602. https://doi.org/10.5194/nhess-15-587-2015

Berti, M., Genevois, R., Simoni, A., & Tecca, P. R. (1999). Field observations of a debris flow event in the Dolomites. *Geomorphology*, *29*(3-4), 265–274. https://doi.org/10.1016/S0169-555X(99)00018-5

Berti, M., Martina, M. L. V., Franceschini, S., Pignone, S., Simoni, A., & Pizziolo, M. (2012). Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. *Journal of Geophysical Research*, *117*, F04006. https://doi.org/10.1029/2012JF002367

Bertrand, M., Lièbault, F., & Piègay, H. (2013). Debris-flow susceptibility of upland catchments. *Natural Hazards*, *67*(2), 497–511. https://doi.org/10.1007/s11069-013-0575-4

Beven, K. J., Almeida, S., Aspinall, W. P., Bates, P. D., Blazkova, S., Borgomeo, E., et al. (2017). Epistemic uncertainties and natural hazard risk assessment. 1. A review of different natural Hazard areas. *Natural Hazards and Earth System Sciences Discussions*, 1–53. https://doi.org/10.5194/nhess-2017-250

Beven, K. J., Aspinall, W. P., Bates, P. D., Borgomeo, E., Goda, K., Hall, J. W., et al. (2017). Epistemic uncertainties and natural Hazard risk assessment. 2. What should constitute good practice? *Natural Hazards and Earth System Sciences Discussions*, 1–25. https://doi.org/10.5194/nhess-2017-251

Bogaard, T., & Greco, R. (2018). Invited perspectives: Hydrological perspectives on 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

Bogaard, T. A., & Greco, R. (2016). Landslide hydrology: From hydrology to pore pressure. *Wiley Interdisciplinary Reviews Water*, *3*(3), 439–459. https://doi.org/10.1002/wat2.1126

Borga, M., Stoffel, M., Marchi, L., Marra, F., & Jakob, M. (2014). Hydrogeomorphic response to extreme rainfall in headwater systems: Flash floods and debris flows. *Journal of Hydrology*, *518*, 194–205. https://doi.org/10.1016/j.jhydrol.2014.05.022

Braun, M., & Kaitna, R. (2016). Analysis of meteorological trigger conditions for debris flows on a daily time scale. In S. A. Makarov, J. V. Atutova, & A. I. Shekhovtsov (Eds.), *Debris flows: risks, forecast, protection: Materials of IV International Conference (Russia, Irkutsk-Arshan Village, The Republic of Buriatia)* (pp. 255–257). Irkutsk: Publishing House of Sochava Institute of Geography SB RAS.

Brunetti, M. T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., & Guzzetti, F. (2010). Rainfall thresholds for the possible occurrence of landslides in Italy. *Natural Hazards and Earth System Sciences*, *10*(3), 447–458. https://doi.org/10.5194/nhess-10-447-2010

Cardinali, M., Ardizzone, F., Galli, M., Guzzetti, F., & Reichenbach, P. (2000). Landslides triggered by rapid snow melting: The December 1996–January 1997 event in Central Italy. In *Proceedings 1st Plinius conference on Mediterranean storms* (pp. 439–448). Cosenza: Bios.

Chaubey, I., Haan, C. T., Grunwald, S., & Salisbury, J. M. (1999). Uncertainty in the model parameters due to spatial variability of rainfall. *Journal of Hydrology*, *220*(1-2), 48–61. https://doi.org/10.1016/S0022-1694(99)00063-3

Chawla, N. V. (2009). Data mining for imbalanced datasets: An overview. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 853–867). Boston, MA: Springer. https://doi.org/10.1007/978-0-387-09823-4_45

Church, M., & Miles, M. J. (1987). Meteorological antecedents to debris flow in southwestern British Columbia; some case studies. *Reviews in Engineering Geology*, *7*, 63–80. https://doi.org/10.1130/REG7-p63

Ciavolella, M., Bogaard, T., Gargano, R., & Greco, R. (2016). Is there predictive power in hydrological catchment information for regional landslide hazard assessment? *Procedia Earth and Planetary Science*, *16*, 195–203. https://doi.org/10.1016/j.proeps.2016.10.021

Criss, R. E., & Winston, W. E. (2008). Do Nash values have value? Discussion and alternate proposals. *Hydrological Processes*, *22*(14), 2723–2725. https://doi.org/10.1002/hyp.7072

Crozier, M. J. (1999). Prediction of rainfall-triggered landslides: A test of the antecedent water status model. *Earth Surface Processes and Landforms*, *24*(9), 825–833. https://doi.org/10.1002/(SICI)1096-9837(199908)24:9<825::AID-ESP14>3.0.CO;2-M

Dall'Amico, M., & Hornsteiner, M. (2006). A simple method for estimating daily and monthly mean temperatures from daily minima and maxima. *International Journal of Climatology*, *26*(13), 1929–1936. https://doi.org/10.1002/joc.1363

Devoli, G., Kleivane, I., Sund, M., Orthe, N. K., Ekker, R., Johnsen, E., & Colleuille, H. (2015). Landslide early warning system and web tools for real-time scenarios and for distribution of warning messages in Norway. In *Engineering geology for society and territory* (Vol. 2, pp. 625–629). Cham: Springer.

Eccel, E. (2012). Estimating air humidity from temperature and precipitation measures for modelling applications. *Meteorological Applications*, *19*(1), 118–128. https://doi.org/10.1002/met.258

European Environment Agency (2014). Corine Land Cover (CLC) 1990. Retrieved from https://www.eea.europa.eu/data-and-maps/data/corine-land-cover-1990-raster-3

Euser, T., Hrachowitz, M., Winsemius, H. C., & Savenije, H. H. (2015). The effect of forcing and landscape distribution on performance and consistency of model structures. *Hydrological Processes*, *29*(17), 3727–3743. https://doi.org/10.1002/hyp.10445

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., & Savenije, H. H. G. (2013). A framework to assess the realism of model structures using hydrological signatures. *Hydrology and Earth System Sciences*, *17*(5), 1893–1912. https://doi.org/10.5194/hess-17-1893-2013

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

Ford, T. W., Rapp, A. D., Quiring, S. M., & Blake, J. (2015). Soil moisture–precipitation coupling: Observations from the Oklahoma Mesonet and underlying physical mechanisms. *Hydrology and Earth System Sciences*, *19*(8), 3617–3631. https://doi.org/10.5194/hess-19-3617-2015

Fovet, O., Ruiz, L., Hrachowitz, M., Faucheux, M., & Gascuel-Odoux, C. (2015). Hydrological hysteresis and its value for assessing process consistency in catchment conceptual models. *Hydrology and Earth System Sciences*, *19*(1), 105–123. https://doi.org/10.5194/hess-19-105-2015

Frattini, P., Crosta, G., & Sosio, R. (2009). Approaches for defining thresholds and return periods for rainfall-triggered shallow landslides. *Hydrological Processes*, *23*(10), 1444–1460. https://doi.org/10.1002/hyp.7269

Gao, H., Ding, Y., Zhao, Q., Hrachowitz, M., & Savenije, H. H. G. (2017). The importance of aspect for modelling the hydrological response in a glacier catchment in Central Asia. *Hydrological Processes*, *31*(16), 2842–2859. https://doi.org/10.1002/hyp.11224

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S., & Savenije, H. H. G. (2014). Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the upper Heihe, China. *Hydrology and Earth System Sciences*, *18*(5), 1895–1915. https://doi.org/10.5194/hess-18-18952014

Gao, H., Hrachowitz, M., Sriwongsitanon, N., Fenicia, F., Gharari, S., & Savenije, H. H. G. (2016). Towards understanding the influence of vegetation and topography on model transferability. *Water Resources Research*, *52*, 7999–8022. https://doi.org/10.1002/2016WR019574

Gariano, S. L., Brunetti, M. T., Iovine, G., Melillo, M., Peruccacci, S., Terranova, O. G., et al. (2015). Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, Southern Italy. *Geomorphology*, *73*(1), 111–136. https://doi.org/10.1007/s11069-014-1129-0

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., & Savenije, H. H. G. (2014). Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration. *Hydrology and Earth System Sciences*, *18*(12), 4839–4859. https://doi.org/10.5194/hess-18-4839-2014

Gharari, S., Hrachowitz, M., Fenicia, F., & Savenije, H. H. G. (2011). Hydrological landscape classification: Investigating the performance of HAND based landscape classifications in a central European meso-scale catchment. *Hydrology and Earth System Sciences*, *15*(11), 3275–3291. https://doi.org/10.5194/hess-15-3275-2011

Gharari, S., Shafiei, M., Hrachowitz, M., Fenicia, F., Gupta, H. V., & Savenije, H. H. G. (2014). A search strategy for constraint-based parameterization of environmental systems models. *Hydrology and Earth System Sciences*, *18*(12), 4861–4870. https://doi.org/10.5194/hess-18-4861-2014

Gregoretti, C., & Fontana, G. D. (2008). The triggering of debris flow due to channel-bed failure in some Alpine headwater basins of the Dolomites: Analyses of critical runoff. *Hydrological Processes*, *22*(13), 2248–2263. https://doi.org/10.1002/hyp.6821

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in Central and Southern Europe. *Meteorology and Atmospheric Physics*, *98*(3-4), 239–267. https://doi.org/10.1007/s00703-007-0262-7

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2008). The rainfall intensity–duration control of shallow landslides and debris flows: An update. *Landslides*, *5*(1), 3–17. https://doi.org/10.1007/s10346-007-0112-1

Haeckel, H. (2016). *Meteorologie* [in german]. UTB. Stuttgart, 6th edition.

Hammerl, C. (Ed.) (2001). *Die Zentralanstalt Für Meteorologie Und Geodynamik 1851–2001: 150 Jahre Meteorologie Und Geophysik in Österreich*. Graz: Leykam.

Hargreaves, G. H., & Samani, Z. A. (1982). Estimation of potential evapotranspiration, Journal of Irrigation and Drainage Division. *Proceedings of the American Society of Civil Engineers*, *108*, 223–230.

Heiser, M., Scheidl, C., Eisl, J., Spangl, B., & Hübl, J. (2015). Process type identification in torrential catchments in the eastern Alps. *Geomorphology*, *232*, 239–247. https://doi.org/10.1016/j.geomorph.2015.01.007

Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, *282*(1-4), 104–115. https://doi.org/10.1016/S0022-1694(03)002579

Houze, R. A. (2014). *Cloud Dynamics*, (2nd ed.). New York: Academic Press.

Hrachowitz, M., & Clark, M. P. (2017). HESS opinions: The complementary merits of competing modelling philosophies in hydrology. *Hydrology and Earth System Sciences*, *21*(8), 3953–3973. https://doi.org/10.5194/hess-21-3953-2017

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., et al. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resources Research*, *50*, 7445–7469. https://doi.org/10.1002/2014WR015484

Hrachowitz, M., & Weiler, M. (2011). Uncertainty of precipitation estimates caused by sparse gauging networks in a small, mountainous watershed. *Journal of Hydrologic Engineering*, *16*(5), 460–471. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000331

John, G. H., & Langley, P. (1995). Estimating continuous distributions in Bayesian classifiers. In *Eleventh conference on uncertainty in artificial intelligence* (pp. 338–345). San Mateo: Morgan Kaufmann.

Jomelli, V., Brunstein, D., Grancher, D., & Pech, P. (2007). Is the response of hill slope debris flows to recent climate change univocal? A case study in the Massif des Ecrins (French Alps). *Climatic Change*, *85*(1–2), 119–137. https://doi.org/10.1007/s10346-007-0112-1

Kavetski, D., & Clark, M. P. (2010). Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction: Numerical daemons of hydrological modeling, 2. *Water Resources Research*, *46*, W10511. https://doi.org/10.1029/2009WR008896

Kavetski, D., & Clark, M. P. (2011). Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing. *Hydrological Processes*, *25*(4), 661–670. https://doi.org/10.1002/hyp.7899

Kavetski, D., Kuczera, G., & Franks, S. W. (2006). Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology*, *320*(1-2), 173–186. https://doi.org/10.1016/j.jhydrol.2005.07.012

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583. https://doi.org/10.2307/2280779

Leijnse, H., Uijlenhoet, R., & Stricker, J. N. M. (2007). Rainfall measurement using radio links from cellular communication networks: Rapid communication. *Water Resources Research*, *43*, W03201. https://doi.org/10.1029/2006WR005631

Leonarduzzi, E., Molnar, P., & McArdell, B. W. (2017). Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data: Rainfall thresholds for landslides in Switzerland. *Water Resources Research*, *53*, 6612–6625. https://doi.org/10.1002/2017WR021044

Llasat, M. C., Rigo, T., Ceperuelo, M., & Barrera, A. (2005). Estimation of convective precipitation: The meteorological radar versus an automatic rain gauge network. *Advances in Geosciences*, *2*, 103–109. https://doi.org/10.5194/adgeo-2-103-2005

Mader, H., Steidl, T., & Wimmer, R. (1996). *Abflussregime Österreichischer Fliessgewässer: Beitrag Zu Einer Bundesweiten Fliessgewässertypologie*, *Monographien des Umweltbundesamts* (Vol. 82). Wien: UBA.

Marra, F., Nikolopoulos, E. I., Creutin, J. D., & Borga, M. (2014). Radar rainfall estimation for the identification of debris-flow occurrence thresholds. *Journal of Hydrology*, *519*, 1607–1619. https://doi.org/10.1016/j.jhydrol.2014.09.039

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Forecasting and Travel Behavior*, 72–77.

Merz, R., & Blöschl, G. (2003). A process typology of regional floods: Process typology of regional floods. *Water Resources Research*, *39*(12), 1340. https://doi.org/10.1029/2002WR001952

Meyer, N. K., Dyrrdal, A. V., Frauenfelder, R., Etzelmüller, B., & Nadim, F. (2012). Hydrometeorological threshold conditions for debris flow initiation in Norway. *Natural Hazards and Earth System Sciences*, *12*(10), 3059–3073. https://doi.org/10.5194/nhess-12-3059-2012

Mostbauer, K., Kaitna, R., Prenner, D., & Hrachowitz, M. (2018). The temporally varying roles of rainfall, snowmelt and soil moisture for debris flow initiation in a snow-dominated system. *Hydrology and Earth System Sciences*, *22*, 3493–3513. https://doi.org/10.5194/hess-22-3493-2018

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models Part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nijzink, R. C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., et al. (2016). The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models. *Hydrology and Earth System Sciences*, *20*(3), 1151–1176. https://doi.org/10.5194/hess-20-1151-2016

Nikolopoulos, E. I., Crema, S., Marchi, L., Marra, F., Guzzetti, F., & Borga, M. (2014). Impact of uncertainty in rainfall estimation on the identification of rainfall thresholds for debris flow occurrence. *Geomorphology*, *221*, 286–297. https://doi.org/10.1016/j.geomorph.2014.06.015

Parajka, J., Merz, R., & Blöschl, G. (2005). Regional water balance components in Austria on a daily basis. *Österreichische Wasser- und Abfallwirtschaft*, *57*(3-4), 43–56. https://doi.org/10.1007/BF03165611

Patzelt, G. (2015). The Austrian Glacier Inventory GI 1, 1969, in ArcGIS (Shapefile) Format. *Pangaea*. https://doi.org/10.1594/PANGAEA.844983

Peres, D. J., Cancelliere, A., Greco, R., & Bogaard, T. A. (2018). Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds. *Natural Hazards and Earth System Sciences*, *18*(2), 633–646. https://doi.org/10.5194/nhess-18-633-2018

Pérez, A., Larrañaga, P., & Inza, I. (2009). Bayesian classifiers based on kernel density estimation: Flexible classifiers. *International Journal of Approximate Reasoning*, *50*(2), 341–362. https://doi.org/10.1016/j.ijar.2008.08.008

Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., & Guzzetti, F. (2017). Rainfall thresholds for possible landslide occurrence in Italy. *Geomorphology*, *290*, 39–57. https://doi.org/10.1016/j.geomorph.2017.03.031

Priddy, K. L., & Keller, P. E. (2005). *Artificial neural networks: An introduction*. Bellingham, Wash: SPIE Press. https://doi.org/10.1117/3.633187

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., & Waterloo, M. J. (2008). HAND, a new terrain descriptor using SRTM-DEM: Mapping Terra-Firme rainforest environments in Amazonia. *Remote Sensing of Environment*, *112*(9), 3469–3481. https://doi.org/10.1016/j.rse.2008.03.018

Rickenmann, D., & Zimmermann, M. (1993). The 1987 debris flows in Switzerland: Documentation and analysis. *Geomorphology*, *8*(2-3), 175–189. https://doi.org/10.1016/0169-555X(93)90036-2

Rolland, C. (2003). Spatial and seasonal variations of air temperature lapse rates in Alpine regions. *Journal of Climate*, *16*(7), 1032–1046. https://doi.org/10.1175/1520-0442(2003)016<1032:SASVOA>2.0.CO;2

Rulfová, Z., & Kyselý, J. (2013). Disaggregating convective and stratiform precipitation from station weather data. *Atmospheric Research*, *134*, 100–115. https://doi.org/10.1016/j.atmosres.2013.07.015

Saito, H., Nakayama, D., & Matsuyama, H. (2010). Relationship between the initiation of a shallow landslide and rainfall intensity—Duration thresholds in Japan. *Geomorphology*, *118*(1–2), 167–175. https://doi.org/10.1016/j.geomorph.2009.12.016

Salio, P., Hobouchian, M. P., García Skabar, Y., & Vila, D. (2015). Evaluation of high-resolution satellite precipitation estimates over southern South America using a dense rain gauge network. *Atmospheric Research*, *163*, 146–161. https://doi.org/10.1016/j.atmosres.2014.11.017

Savenije, H. H. G., & Hrachowitz, M. (2017). Catchments as meta-organisms—A new blueprint for hydrological modelling. *Hydrology and Earth System Sciences*, *21*(2), 1107–1116. https://doi.org/10.5194/hess-21-1107-2017

Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors: Likelihood function for parameter and predictive inference. *Water Resources Research*, *46*(10). https://doi.org/10.1029/2009WR008933

Schraml, K., Thomschitz, B., McArdell, B. W., Graf, C., & Kaitna, R. (2015). Modeling debris-flow runout patterns on two Alpine fans with different dynamic simulation models. *Natural Hazards and Earth System Sciences*, *15*(7), 1483–1492. https://doi.org/10.5194/nhess-15-1483-2015

Sevruk, B. (1997). Regional dependency of precipitation-altitude relationship in the Swiss Alps. In H. F. Diaz, M. Beniston, & R. S. Bradley (Eds.), *Climatic change at high elevation sites*, (pp. 123–137). Dordrecht, Netherlands: Springer. https://doi.org/10.1007/978-94-015-8905-5_7

Shah, S. M. S., O'Connell, P. E., & Hosking, J. R. M. (1996). Modelling the effects of spatial variability in rainfall on catchment response. 2. Experiments with distributed and lumped models. *Journal of Hydrology*, *175*(1-4), 89–111. https://doi.org/10.1016/S0022-1694(96)80007-2

Singh, V. P. (1997). Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph. *Hydrological Processes*, *11*(12), 1649–1669. https://doi.org/10.1002/(SICI)1099-1085(19971015)11:12<1649::AID-HYP495>3.0.CO;2-1

Stock, J., & Dietrich, W. E. (2003). Valley incision by debris flows: Evidence of a topographic signature: Valley incision by debris flows. *Water Resources Research*, *39*(4), 1089. https://doi.org/10.1029/2001WR001057

Stoffel, M., Bollschweiler, M., & Beniston, M. (2011). Rainfall characteristics for periglacial debris flows in the Swiss Alps: Past incidences–potential future evolutions. *Climatic Change*, *105*(1-2), 263–280. https://doi.org/10.1007/s10584-011-0036-6

Szymczak, S., Bollschweiler, M., Stoffel, M., & Dikau, R. (2010). Debris-flow activity and snow avalanches in a steep watershed of the Valais Alps (Switzerland): Dendrogeomorphic event reconstruction and identification of triggers. *Geomorphology*, *116*(1-2), 107–114. https://doi.org/10.1016/j.geomorph.2009.10.012

Tsangaratos, P., & Ilia, I. (2016). Comparison of a logistic regression and Naive Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. *Catena*, *145*, 164–179. https://doi.org/10.1016/j.catena.2016.06.004

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation: Forcing data error using MCMC sampling. *Water Resources Research*, *44*, W00B09. https://doi.org/10.1029/2007WR006720

Ypma, T. J. (1995). Historical development of the Newton–Raphson method. *SIAM Review*, *37*(4), 531–551. https://doi.org/10.1137/1037125