Landslides (2020) 17:1881–1896
DOI 10.1007/s10346-020-01400-y
Received: 20 December 2019
Accepted: 20 March 2020
Published online: 15 April 2020

Adrian Wicki · Peter Lehmann · Christian Hauck · Sonia I. Seneviratne · Peter Waldner · Manfred Stähli

Assessing the potential of soil moisture measurements for regional landslide early warning

Abstract In mountainous terrain, rainfall-induced landslides pose a serious risk to people and infrastructure. Regional landslide early warning systems (LEWS) have proven to be a cost-efficient tool to inform the public about the imminent landslide danger. While most operational LEWS are based on rainfall exceedance thresholds only, recent studies have demonstrated an improvement of the forecast quality after the inclusion of soil hydrological information. In this study, the potential of in situ soil moisture measurements for regional landslide early warning is assessed. For the first time, a comprehensive soil moisture measurement database was compiled for Switzerland and compared with a national landslide database (Swiss flood and landslide damage database, WSL). The time series were homogenized and normalized to represent saturation values. From ensembles of sensors, the mean and standard deviation saturation were calculated and infiltration events were delimited, characterized, and classified as landslide-triggering or non-triggering based on the occurrence of landslides within a specified forecast distance. A logistic regression function was applied to model the landslide activity based on the infiltration event characteristics and several models were analysed and compared with receiver operating characteristics (ROC). A strong distance dependence becomes apparent showing a forecast goodness decrease with increasing distance between water content measurement site and landslide, and a better forecast goodness for long-lasting as opposed to short-duration precipitation events. While most variability can be explained by the two event properties *antecedent saturation* and *change of saturation during an infiltration event*, event properties that describe antecedent conditions are more important for long-lasting as opposed to short-duration precipitation events that can be better explained by properties describing event dynamics. Overall, the analysis demonstrated that in situ soil moisture data effectively contains specific information useful for landslide early warning.

Keywords Landslide early warning · Shallow landslides · Soil moisture monitoring · Hydrological thresholds · Temporal soil moisture variability

Introduction

Landslides pose a serious risk to people and infrastructure in mountainous terrain. Globally, over 55,000 fatalities were reported in the period of 2004 to 2016 (Froude and Petley 2018). In Switzerland, landslides caused the death of 74 people in the period from 1946 to 2015 (Badoux et al. 2016) and damages to houses and infrastructure of 520 million Euros for the period from 1972 to 2007 (Hilker et al. 2009). The failure of a slope is related to factors acting on various spatial and temporal scales. For rainfall-triggered landslides, the availability of unconsolidated material, the topography, the vegetation cover, and the hydrological preconditions (pre-wetting of the hillslope) determine the susceptibility of a slope to slide and are referred to as "cause factors" (Bogaard and Greco 2016). During a rainfall or snow melt event, the complex patterns of infiltration, redistribution, and drainage of water may lead to a rise in the groundwater level and soil saturation (trigger factors). This can cause a critical increase of pore water pressure and loss of matric suction to a point at which the slope eventually fails (Terzaghi 1943; Bogaard and Greco 2016).

Landslide early warning systems (LEWS) are tools to measure and analyse short-term hydrometeorological variations (e.g. prewetting of the slope, infiltration during storms) and to identify periods of imminent elevated landslide danger. They allow decision-makers to alert people at risk and to move them to safety and have proven to be a cost-efficient and effective element of integral risk management (Stähli et al. 2015). LEWS are applied on different spatial scales and typically cover single slopes or catchments (local scale) or areas of a few to several thousand square kilometres with similar climatic and physiographic characteristics (regional scale). Where no local or regional LEWS is available, global models can be applied (Guzzetti et al. 2007). Many operational LEWS exist around the world, reviews of which are given for local (Michoud et al. 2013) and regional LEWS (Bell et al. 2010; Piciullo et al. 2018; Guzzetti et al. 2020), for systems applied to a range of processes and scales (Stähli et al. 2015), and more specifically for LEWS in Europe (Guzzetti et al. 2007) and the USA (Baum and Godt 2010).

Many LEWS are based on triggering rainfall exceedance threshold models which empirically relate the occurrence of landslides to rainfall event characteristics such as intensity, duration, total amounts, or combinations thereof (Guzzetti et al. 2007, 2008; Segoni et al. 2018a). Most prominently, triggering event thresholds are formulated by intensity-duration relationships (e.g. Caine 1980; Crosta and Frattini 2001). In recognition of the predisposing effect of the antecedent wetness conditions, cumulative rainfall amounts prior to the triggering event have been explicitly incorporated into exceedance thresholds (Chleborad 2000; e.g. Aleotti 2004; Wieczorek and Glade 2005; Martelloni et al. 2012) and antecedent precipitation indices were developed that estimate the water balance or mimic soil moisture dynamics (e.g. Crozier 1999; Glade 2000; Godt et al. 2006; Ponziani et al. 2012). To further determine the spatial extent of increased landslide activity, Kirschbaum and Stanley (2018) have combined satellite-based rainfall estimates with a landslide susceptibility map. In Switzerland, Rickli et al. (2008) analysed the occurrence of shallow landslides during a large-scale precipitation event in 2005. They found a large range of intensity-duration precipitation thresholds that vary across different geological settings. In a study by Leonarduzzi et al. (2017), countrywide rainfall thresholds for shallow landslides were developed for Switzerland. The study of Leonarduzzi et al. (2017) is based on a comparison of a 2×2 km gridded daily precipitation dataset from interpolated rain gauge measurements (RhiresD, Swiss Federal Office of Meteorology and Climatology MeteoSwiss) with a national landslide database (Swiss flood and

landslide damage database, Swiss Federal Research Institute WSL). The best predictive power was found for the intensity-duration rainfall threshold (i.e. consideration of antecedent rainfall did not improve the predictive power).

While rainfall measurements are available in increasing spatial resolution, using rainfall thresholds and rainfall-based indices for landslide early warning bears specific limitations. Conceptually, rainfall characteristics represent only proxies for variations in critical hydrological properties in the ground (e.g. Reichenbach et al. 1998; Seneviratne et al. 2010; Stähli et al. 2015). Furthermore, due to the incorporation of regional climatological and geological characteristics, rainfall intensity and duration may vary over few orders of magnitude on a country or a global level (Guzzetti et al. 2008; Baum and Godt 2010). Antecedent water conditions and the prewetting of the slope at depth cannot be reflected fully (Godt et al. 2009) and critical antecedent rainfall amounts can vary considerably between different soils and even throughout the year (Aleotti 2004; Baum and Godt 2010; Napolitano et al. 2016). Finally, snow melt cannot be accounted for by pure rainfall thresholds (Cardinali et al. 2000; Martelloni et al. 2013).

Model approaches that enable an assessment of the current hydrological state or directly predict the triggering factors might be able to overcome these limitations. A challenge for such approaches is, however, the often high spatial variability of the hydrological properties. Nevertheless, the need for new hydrological-based thresholds was stated by Berne et al. (2013) or Devoli (2017). Bogaard and Greco (2018) further postulated that both false and missed alarms could be significantly reduced if the antecedent wetness state is included by direct measurements of soil water content, catchment storage, or groundwater levels. In fact, this could be demonstrated in recent studies: At a railway line in Seattle (WA, USA), Mirus et al. (2018b) reported an improvement of the forecast quality after replacing an antecedent cumulative rainfall threshold by an antecedent saturation component. The saturation is derived from five volumetric water content probes installed in a steep hillslope (Smith et al. 2017). Similarly, an improvement of the forecast goodness could be demonstrated for a landslide prone site near the City of Portland (OR, USA), where the antecedent saturation is calculated from 11 volumetric water content probes installed in a soil pit (Mirus et al. 2018a). At a landslide-prone slope in Cervinanra (Campania, Italy), a significant improvement of the threshold performance was reported after rainfall characteristics were normalized by the measured soil moisture state from suction and soil water content sensors (Comegna et al. 2016; Greco and Bogaard 2016). Segoni et al. (2018b) reported a significant improvement of the forecast quality of a rainfall-based LEWS in Emilia Romagna (Italy) by replacing the antecedent rainfall component with a modelled soil moisture value averaged over specific territorial units. Satellite-derived soil moisture estimates were used for landslide early warning in the Italian regions of Umbria (Brocca et al. 2016) and Emilia Romagna (Zhuo et al. 2019), as well as the San Francisco Bay Area (CA, USA; Thomas et al. 2019). While the use of satellite-based soil moisture products allows for issuing spatially distributed landslide early warnings, limitations persist due to the coarse spatial resolution. Finally, streamflow measurements as a proxy for catchment water storage were found to be able to identify landslide activity in the Tiber Basin, Italy (Reichenbach et al. 1998), and in the North Shore Mountains, Canada (Jakob and Weatherly 2003).

The hydrological state of the unsaturated zone is commonly described with the volumetric soil water content (VWC), Θ (m^3/m^3), which is usually defined as the ratio of a volume of water V_w (m^3) to a given volume of soil V_T (m^3), with

$$\theta = V_W / V_T \quad (1)$$

Electromagnetic sensor networks have been widely used for indirect estimation of the in situ VWC at the point scale. Along two to four parallel electrodes or pairs of rings, the electromagnetic sensor generates an electric field. Electromagnetic properties are measured (e.g. travel time, charge time) which can be related to the dielectric permittivity of the surrounding material (Babaeian et al. 2019). The large dielectric permittivity differences between water (~ 80) and solids (~ 4) or air (1) allow then the estimation of VWC using a specific calibration function (Looyenga 1965; Topp et al. 1980). Two main types of sensors are used to infer permittivity of the soil. In time domain reflectometry (TDR), the propagation velocity of a step voltage pulse is measured along a pair of electrodes at measurement frequencies above 500 MHz. Capacitance sensors on the other hand measure the impedance of the surrounding material and typically operate at lower frequencies (5–150 MHz) (Robinson et al. 2008). At frequencies below 500 MHz and in clay-rich soils, the derived permittivity is further dependent on measurement frequency, temperature, texture, and salinity. Consequently, site-specific calibration is needed particularly for capacitance-based sensors and clay-rich soils (Kelleners et al. 2005; Chen and Or 2006, Mittelbach et al. 2012).

Analysis of measured soil water content often involves combining multiple sensors and normalizing VWC values in order to spatially integrate the point information and generalize local scale measurements (e.g. Mittelbach and Seneviratne 2012). Mirus et al. (2018a, b) normalized the measured VWC (θ) values by the porosity at each soil moisture probe which was assumed to be equal to the maximum measured VWC value θ_{Max} of each probe. Saturated conditions were identified by positive pore-water pressures measured from co-located tensiometers. Antecedent saturation conditions were then calculated by the average saturation of all sensors at all depths and locations over specified time periods (1 day, 9 days, 10 days, and 15 days) prior to rainfall events. At a steep forested slope in Switzerland, Lehmann et al. (2013) analysed the relationship between the average of the VWC of a number of sensors placed in a slope and the spatial standard deviation between the sensors. Critical hydrological conditions for landslide triggering were identified as periods of high average and low standard deviation VWC values that sustained during several hours. At a slope in southern Italy, Greco and Bogaard (2016) normalized precipitation event characteristics (rainfall depth, intensity, and duration) by hydraulic properties of the ground (characteristic infiltration depth, infiltration rate, and duration) that were estimated based on the measured moisture content of the ground. This resulted in non-dimensional variables that take into account hydrological and morphological characteristics as well as initial hydrological conditions of a slope.

While the predictive power of soil water information at hillslope scale was confirmed in previous studies, no information about water content directly at the hillslopes prone to failure was available in most cases. With an increasing number of soil moisture measurements within operational networks, this data source

becomes more and more available. However, soil properties have a high spatial variability and the distance between soil moisture station and landslide may often be too large to have any local predictive value. In the present study, we aim to (i) analyse the information content of existing soil moisture measurements on critical hydrological conditions for landslide triggering, (ii) assess the representativeness of point measurements at predominantly flat locations for regional landslide activity, and (iii) identify statistical properties that best anticipate the imminent landslide danger. Here, we use all available volumetric soil water content measurements in Switzerland, motivated by the fact that the degree of saturation has been identified as key parameter controlling soil strength (e.g. Springman et al. 2003; Lu et al. 2010).

Data base

Soil moisture data

For this study, a comprehensive soil moisture database using data from existing soil monitoring networks was compiled. For the first time, all known operational soil moisture monitoring networks of Switzerland and additional project-based time series which contain at least two soil moisture sensors at different depths were combined. In total, the dataset included 35 sites and 284 soil moisture sensors (Fig. 1, Table 1), most of which are located at flat open-land locations predominantly on the Swiss Plateau. All sites include multiple sensors that were installed at different depths of up to 150 cm. Concurrent soil water tension measurements at the same temporal resolution were only available for three sites within the FOEN pilot project and two LWF sites. Therefore, only soil water content measurements were considered. The individual sensor networks are described in more detail below.

In 2008, the SwissSMEX (Swiss Soil Moisture Experiment) network was established by ETH Zurich, Agroscope, and MeteoSwiss, to monitor soil moisture on a long-term basis for climate research applications as well as for meteorological and seasonal forecasting (Mittelbach and Seneviratne 2012; https://iac.ethz.ch/group/land-climate-dynamics/research/swisssmex.html). The sensors installed include TDR probes (TRIME-EZ and TRIME-IT, IMKO GmbH) and capacitance-based sensors (10HS, METER Group). Measurements are recorded every 10 min and sensors are installed in profiles of up to 150 cm depth (Mittelbach et al. 2011). In this study, data from 13 grassland sites were used that are mainly located on the Swiss Plateau.

The Long-Term Forest Ecosystem Research Programme (LWF) of the Swiss Federal Research Institute WSL includes the investigation of effects on forest ecosystem processes based on measurements on about 20 selected intensive monitoring plots in Switzerland (Schaub et al. 2011) and Europe (Ferretti and Schaub 2014). In this study, the soil moisture measurements from 10 forested plots in Switzerland were included. Soil moisture has been measured since 2008 or later using capacitance sensors (EC-5, METER Group) at several depths down to 80 cm with typically up to three replicates. Measurements are recorded every 10 or 60 min. SMEX-Veg includes three sites with installations of both SMEX and EC-5 sensors.

As part of the Soil Moisture in Mountainous Terrain (SOMOMOUNT) framework, a soil moisture monitoring network was installed in 2013 by the University of Fribourg. Along an altitudinal gradient from the Jura Mountains to the western Swiss Alps at mid to high altitudes, six sites were equipped with hybrid capacitance/TDR sensors (SMT100, TRUEBNER GmbH), TDR sensors (TRIME-PICO64, IMKO GmbH), and capacitance sensors (PR2/6, Delta-T Devices Ltd). Measurements are taken at several depths down to 100 cm and at measurement intervals up to every 10 min (Pellet and Hauck 2017).

Furthermore, the collected database includes sites from cantonal authorities which were established with the aim of soil conservation in agriculture and construction. In the canton of Lucerne, two sites are equipped with TDR sensors (TRIME-PICO64, IMKO GmbH) down to 60 cm depth (60 min resolution) (Umwelt und Energie Kanton Luzern 2019). In the canton of Uri, TDR sensors (TRIME-EZ, IMKO GmbH) are installed down to 60 cm depth (10 min resolution) (Geisser et al. 2011). Furthermore, time series of soil moisture measurements are included from a pilot project of the Federal Office for the Environment FOEN. From 2015 to 2016, TDR measurements (TRIME-PICO64, IMKO GmbH) were conducted at three agricultural sites in several depths down to 60 cm (60 min resolution) (Loretz and Ruckstuhl 2017).

Landslide data

The landslide dataset has been extracted from the Swiss flood and landslide damage database (WSL, Hilker et al. 2009). Information on damage caused by flood, debris flow, landslide, and rockfall events has been recorded since 1972 (rockfall since 2002) and is based on event descriptions in newspaper and magazine articles. Each database entry includes the x-y-coordinates and timing of the landslide, an event description as well as a classification of the triggering rainfall event duration. A triggering event is classified as long-lasting if the triggering rainfall intensity sustained 0.5 mm/h for at least 6 h which typically can be related to frontal systems. Triggering events that lasted shorter than 6 h are classified as short-duration and they are mostly connected to convective storms. This landslide database was chosen because it represents a nation-wide inventory that was established with a consistent methodology.

For this study, a subset of the dataset was compiled that contains landslides that were recorded from 2008 to 2018 only. Further, deep-seated and human-induced landslides (e.g. due to pipe breaks, slips at artificial road embankments) were excluded when indicated in the event description. This amounts to a total of 452 landslide records that were included in the analysis (Fig. 1). The hour of triggering was provided for 25% of the landslide records only. For 35% of the records, the timing was described in general terms such as “in the morning” or “at night”, in which case, the timing was assumed (e.g. 09:00 a.m. for “in the morning”). For the rest of the landslide records (40% of the dataset), only the date of occurrence was provided and thus, we set the time of occurrence to 12:00 p.m.

Methodology

Data preparation

The VWC time series are of different data quality and we applied a quality control scheme. First, VWC values outside the 0.0–1.0 m^3/m^3 range were removed automatically. Furthermore, each time series was checked visually and apparent outliers and periods of unusual VWC variation were removed manually. In a next step, measurements during periods of frozen conditions were removed

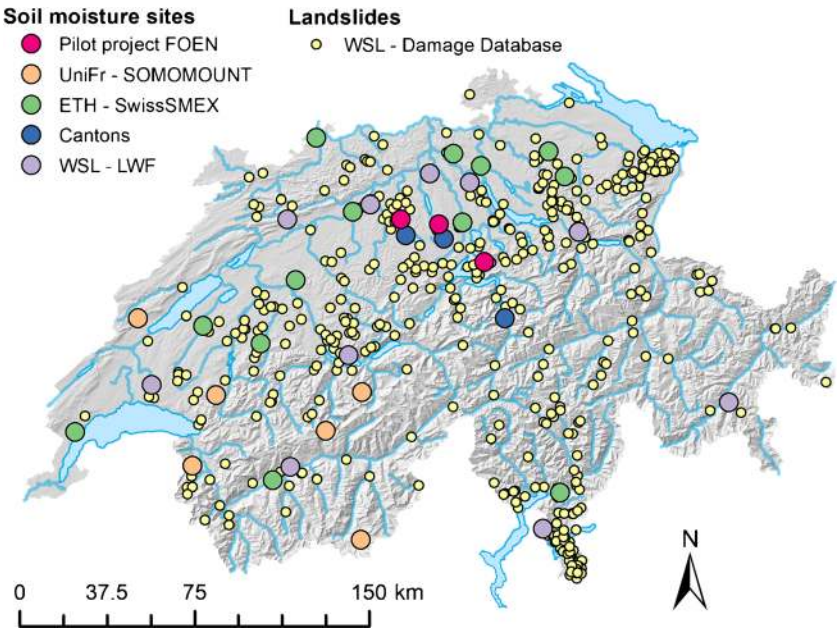

Fig. 1 Map of Switzerland showing all soil moisture measurement sites used in this study (large filled circles) and recorded shallow landslide events from 2008 to 2018 (yellow dots)

as they are not relevant for the present analysis. Finally, all timestamps were synchronized.

The VWC values of each time series were then normalized by the minimum and maximum value to account for the spatial variation of soil properties and the uncertainties in the absolute VWC values resulting from the uncertainty in the sensor calibration due to the limited knowledge of the effective local soil conditions in the measurement volume around the sensor (Fig. 2b). To reduce the effect of outliers at the saturated end, the 99.9 percentile rank was used as a proxy for the maximum. The normalized VWC, θ_{norm} (%), was calculated with

$$\theta_{\mathrm{norm}} = \frac{\theta - \theta_{\mathrm{Min}}}{\theta_{99.9\%} - \theta_{\mathrm{Min}}} \quad (2)$$

where θ is the measured, θ_{Min} the minimum, and $\theta_{99.9\%}$ the 99.9 percentile rank of VWC of the individual sensor time series. The θ_{norm} thus corresponds more or less to the often used term *water saturation S* (%). Finally, all time series were aggregated to hourly values by calculating the hourly average.

Next, ensembles of sensors were built and the ensemble mean and ensemble standard deviation were calculated (Fig. 2b-d). The standard ensemble of sensors contains all available sensors at a specific measurement site. From this, two sub-ensembles were drawn containing all sensors down to 30 cm depth and all sensors below 30 cm depth. A fourth and a fifth set of ensembles, respectively, contain a subset of two and three sensors per site, each at different depths within the uppermost 100 cm of soil. For these two ensembles, each sensor was attributed a weight for the mean and standard deviation calculation according to the representing depth interval of the sensor (see ESM 1 for weight calculation details). Only sensors up to 100 cm depth were used since most sites contain no deeper sensors. The sensors to include were selected by the quality of the records (sensors that contain apparent trends, changes in the temporal variability or strong noise were excluded), by the length of records (sensors with large gaps or that stopped operating were excluded), and also such that they cover a wide depth distribution.

Finally, onsets and endings of infiltration events were assigned for each ensemble mean time series (Figs. 2c and 3a). Here, an infiltration event was defined as the continuous increase of soil

Table 1 Number of sites and sensors as well as the period of operation per monitoring network

Operator	Network	Number of sites	Number of sensors	First record	Last record
ETH Zurich	SwissSMEX	13	137	06.08.2008	Ongoing
WSL	LWF	10	75	19.08.2008	Ongoing
University of Fribourg	SOMOMOUNT	6	45	17.07.2013	Ongoing
Canton of Luzern	Cantonal Network	2	6	22.09.2010	Ongoing
Canton of Uri	Cantonal Network	1	3	08.04.2008	Ongoing
Federal Office for the Environment	Project based	3	18	15.04.2015	24.11.2016

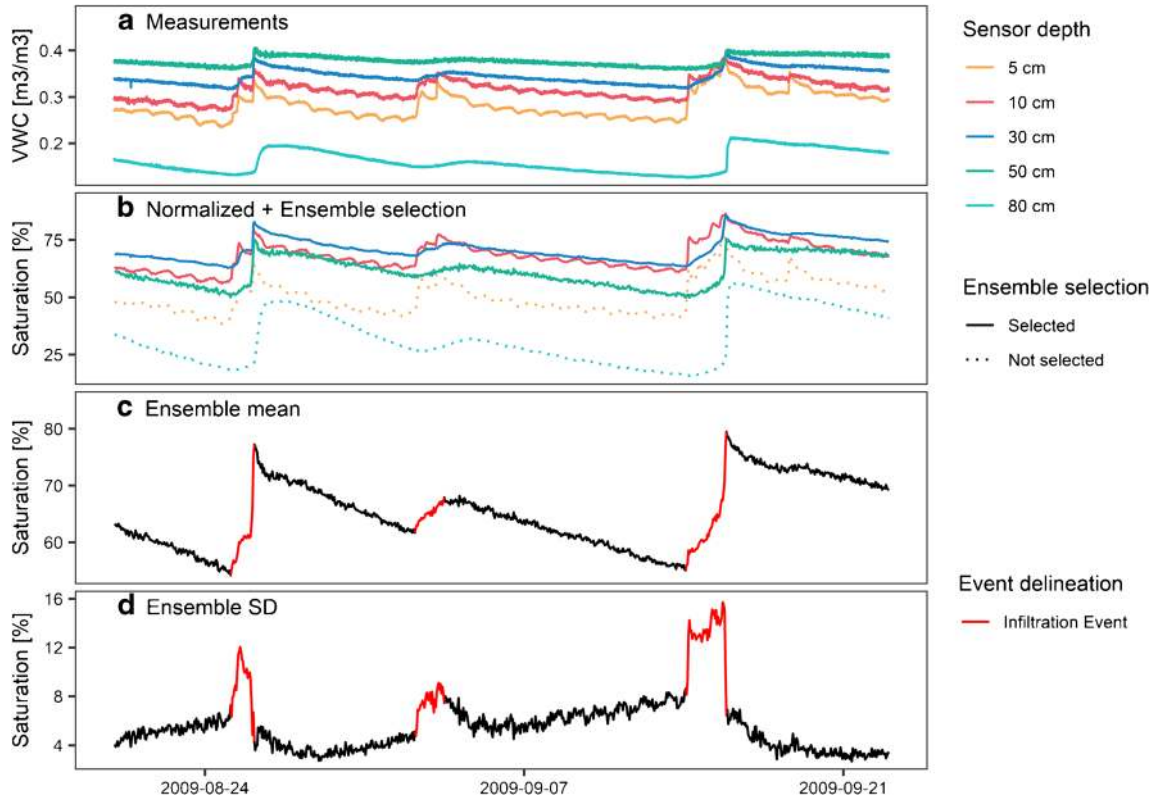

Fig. 2 Sample time series extract of the Cadenazzo site (SwissSMEX, ETH) showing the measured volumetric water content values at different depths (**a**), the normalized saturation values (**b**), the ensemble mean (**c**), and the ensemble standard deviation saturation (**d**). In (**b**), sensors which were included in the ensemble are drawn in solid lines; in (**c**) and (**d**), infiltration events are highlighted in red

moisture due to infiltration processes. Only events with at least 2% saturation increase (4% saturation increase for noisy datasets) and a minimum time lag of 3 h to the next one were considered. They were detected automatically according to the following procedure.

1. Data points of saturation increase are identified as such if they are followed by a minimum 0.4% saturation increase over 1 h, 3 h, or 6 h (0.8% saturation increase for noisy datasets).
2. Data points of saturation increase which are less than 3-h apart are grouped to continuous periods of saturation increase.
3. The minimum and maximum saturation points of each period of saturation increase are identified as the onset and end of an infiltration event, respectively.
4. Infiltration events of less than a total of 2% saturation increase are removed (4% saturation increase for noisy datasets) to omit periods of VWC variation due to other factors than water infiltration (measurement noise, temperature effects, etc.).

The used minimum saturation increase values of steps 1 and 4 were determined by visual inspection of the identified infiltration events, and in 2, the value of 3 h was chosen to capture as many individual events as possible. Different parameters resulted mainly in changing the number of small events and did not impact the overall forecast goodness.

Each infiltration event was then characterized by the following event properties (Fig. 3a): Antecedent saturation (saturation at the onset), saturation change from the onset to the end, event duration, infiltration rate (saturation change divided by event duration), maximum 3-h infiltration rate (maximum saturation change over 3 h), and the 2-week preceding mean and maximum saturation. Furthermore, at the standard deviation time series, the antecedent standard deviation and the standard deviation change were characterized (Fig. 3b).

Infiltration events were then classified as landslide triggering or landslide non-triggering, based on the occurrence or non-occurrence of a landslide during the infiltration event within a specified distance (hereinafter referred to as the forecast distance, Fig. 4). The selection of the forecast distance is a critical step as soil moisture measurements and landslides are expected to better correlate if they are close to each other than if their distance is large. To assess the model sensitivity, the forecast distance was varied from 5 to 40 km in equal steps of 5 km. At such forecast distances, a landslide may occur outside the infiltration event duration due to the onward movement of a storm cell. Furthermore, the timing information of the landslide records may bear considerable uncertainties as outlined before. Therefore, the time window for landslide classification was extended to 12 h prior to the onset until 24 h after the end of an infiltration event.

Logistic regression modelling

A multiple logistic regression model was applied to assess the predictive power of specific infiltration event properties for the occurrence of landslides. We chose this model because it accounts

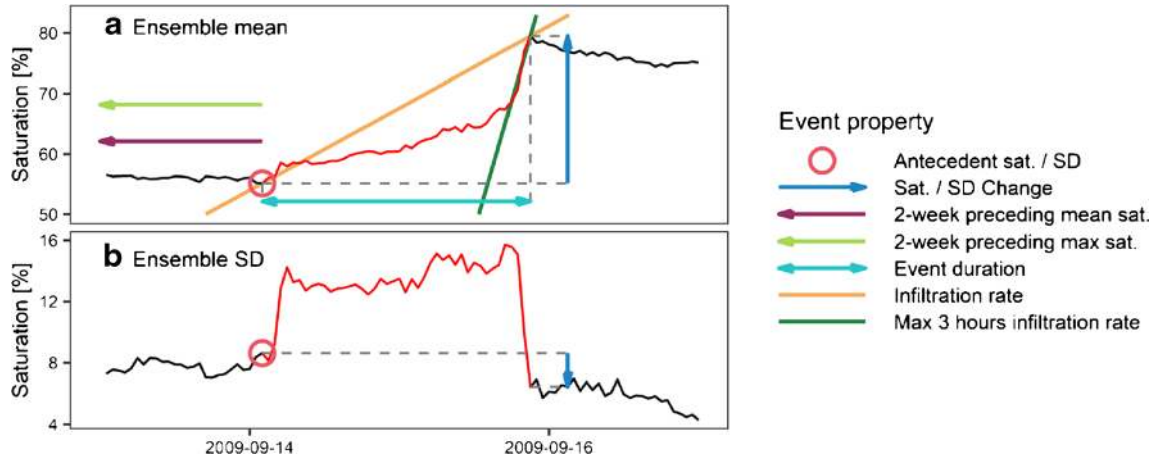

Fig. 3 Sample time series extract of the Cadenazzo site (SwissSMEX, ETH) showing a specific infiltration event and corresponding event properties at the ensemble mean (**a**) and the ensemble standard deviation time series (**b**)

for binary output data, multiple input variables, and non-linearity. Yet, it is a rather robust and transparent modelling approach as opposed to other more non-linear statistical models. In the logistic regression model, the probability $p(X)$ that landslides are observed during an infiltration event was expressed as a function of X:

$$\mathrm{p}(X) = \Pr(\text{landslide triggering} = \text{Yes}|X) \tag{3}$$

where X is a set of infiltration event properties. The value $p(X)$ ranges between 0.0 and 1.0, indicating a low (0.0) to high (1.0) probability of landslide occurrence.

For a set of n infiltration event properties, the following logistic function was fit to the observations:

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n \tag{4}$$

where the left term is referred to as the *log odds* and the fitting parameters β_0, …, β_n were fit with the maximum likelihood method (iteratively reweighted least squares). Rewriting Eq. 4 solves for the probability of landslide occurrence $p(X)$ as follows:

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+\ldots+\beta_n X_n}}{1+e^{\beta_0+\beta_1 X_1+\ldots+\beta_n X_n}} \tag{5}$$

The resulting landslide probabilities were then reclassified into binary landslide triggering classes by applying a threshold between 0.0 and 1.0. If the model is not able to perfectly separate the two

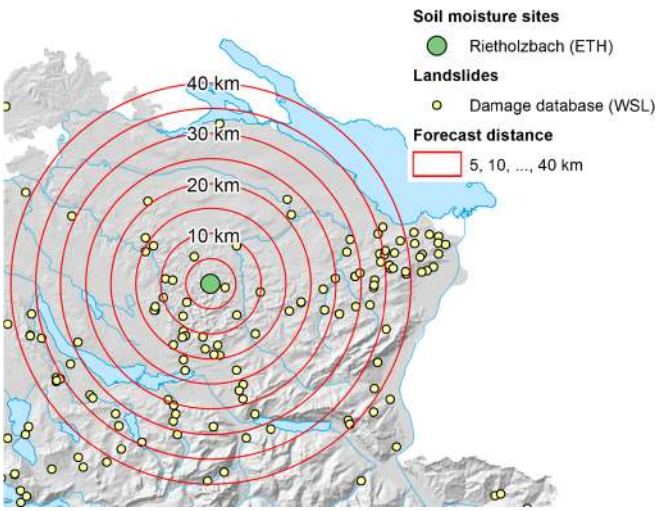

Fig. 4 Extract of the map of Switzerland showing the measurement site Rietholzbach (SwissSMEX, ETH, green point) and the corresponding forecast distances (red circles) varying from 5 to 40 km. Recorded landslide events (Swiss flood and landslide damage database, WSL) are denoted in yellow

classes compared with the observations, applying a specific threshold will always be a trade-off between maximizing the number of true alarms and minimizing the number of false alarms. To explore the trade-off space (or the overall potential of a model), the threshold was thus varied 5000 times in equal increments between 0.0 and 1.0.

Furthermore to the above, a 5-fold cross-validation (CV) scheme was applied to test potential overfitting of the model. In this approach, the dataset was split into five random and equally sized folds of infiltration events. The model was fit to a set of four folds (training dataset) and predictions were made on the fifth fold (validation dataset) using the fitting parameters of the training dataset. The procedure was conducted five times for all training-set-validation-set combinations and the probabilities were reclassified to binary landslide triggering classes. The 5-fold CV scheme was chosen because it represents a good compromise between a low variation in between the folds and using low computational power (James et al. 2013). In the following, the first approach is referred to as the *full dataset approach*, the latter as the *validation set approach*.

Model evaluation

To assess the ability of a model fit and a particular threshold to separate landslide-triggering from non-triggering infiltration events, receiver operating characteristic (ROC) analysis was performed according to Fawcett (2006). First, a two-by-two confusion matrix was constructed between the observed and the modelled landslide-triggering classification. Four instances were distinguished: true-positive count (*TP*, also known as true alarms), false-positive count (*FP*, false alarms), true-negative count (*TN*, true non-alarms), and false-negative count (*FN*, missed alarms). From the confusion matrix, the two ROC metrics (i) true-positive rate, $TPR = TP/(TP + FN)$, and (ii) false-positive rate, $FPR = FP/(FP + TN)$, were calculated. For a model fit and a specific threshold, the TPR corresponds to the fraction of the observed landslide-triggering events that were correctly classified as positive, and the FPR denotes the fraction of the observed non-triggering events which were incorrectly classified as landslide triggering. The two metrics can be plotted in the ROC space with the FPR on the *x*-axis and the TPR on the *y*-axis. A perfect model fit would result in a point with the coordinates (0.0, 1.0); points towards the left bottom corner are considered conservative (less false alarms at the expense of fewer true alarms) and points towards the top right corner are considered optimistic (more true alarms at the expense of more false alarms). Points along the (0.0, 0.0) to (1.0, 1.0) line are equivalent to random classifying.

Since the reclassification threshold was varied 5000 times per model fit, the ROC space contains 5000 points too. The resulting ROC points were then sorted and connected from the bottom left to the top right to produce the ROC curve (cumulative curve). In analogy to the above, a model with a high ability to separate landslide triggering from non-triggering events would plot an ROC curve that bulges towards the top left corner. ROC curves with a bulge in the bottom left corner are conservative and ROC curves that bulge in the top right corner are considered optimistic. The forecast goodness can further be expressed by calculating the area under the ROC curve (AUC) which ranges from 0.0 to 1.0. The AUC value corresponds to the probability of a classifier to rank positive instances higher than negative instances. An AUC value of 1.0 corresponds to a perfect ability for ranking positive instances higher than negative instances and 0.5 corresponds to random guessing. The AUC value is thus a measure for the general performance of a classifier and was used in this study to compare different model fits. In the validation set approach, the ROC metrics and curves were calculated for each of the five validation sets. They were then averaged at each threshold (threshold averaging).

Results

General model performance

The classification procedure and model output are visualized on Fig. 5 for the 15-km forecast distance and a sample time period from June to July 2014 at the SwissSMEX site Plaffeien (Canton Fribourg, pre-Alps), when a series of rainfall events caused many landslide events in all of Switzerland but particularly in the region of Fribourg. During this period, several infiltration events (*I1*-*I9*) led to a continuous increase of the mean saturation (Fig. 5a). While the mean saturation increased continuously, the standard deviation saturation increased until a critical mean saturation point was exceeded during *I5*, after which the standard deviation dropped immediately (Fig. 5b). It was during the infiltration event *I5* that the first landslide events (*L1* and *L2*) were recorded (Fig. 5c). Four more infiltration events followed with *I8* and *I9* being accompanied by more intense landslide activity (*L3* and *L4*). Correspondingly, *I5*, *I8*, and *I9* were classified as landslide triggering. The corresponding landslide-triggering probabilities from the logistic regression model are shown for each infiltration event in Fig. 5e. The triggering probability remained close to 0.00 during the first few infiltration events. It increased significantly during *I5* and reached a maximum of 0.07 during *I8*. While all landslide-triggering infiltration events showed relatively high landslide triggering probabilities, some infiltration events classified as non-triggering yielded a high triggering probability too. This uncertainty can be related to (i) landslides that were not recorded and missing in the landslide database, (ii) a too short forecast distance (and thus missing of landslide events), (iii) inhomogeneities between the different measurement sites, (iv) an inadequate representation of the triggering processes by the used infiltration event properties, or (v) non-representativeness of the soil moisture measurement site for local soil wetness conditions at the landslide location. A detailed assessment of the forecast quality and an analysis of the main model drivers will be given in the following sections.

ROC curves by forecast distance for the same model set-up (i.e. using all event properties, all infiltration events and all available sensors) by forecast distance and of the full dataset and the validation set approach are shown in Fig. 6. The number of infiltration events and AUC values by forecast distance are shown in Table 2. From the ROC curves and AUC values, a strong distance dependence is visible whereas the forecast quality increases with decreasing forecast distance for full dataset approach (AUC = 0.72 at 40 km to AUC = 0.83 at 5 km). At the same time, forecast quality differences between the full and validation set approach increase towards short forecast distances (AUC = 0.83 for full dataset vs. AUC = 0.74 for validation set at 5 km) indicating overfitting of the model. This can be explained by the lower number of landslide-triggering events at short forecast distances

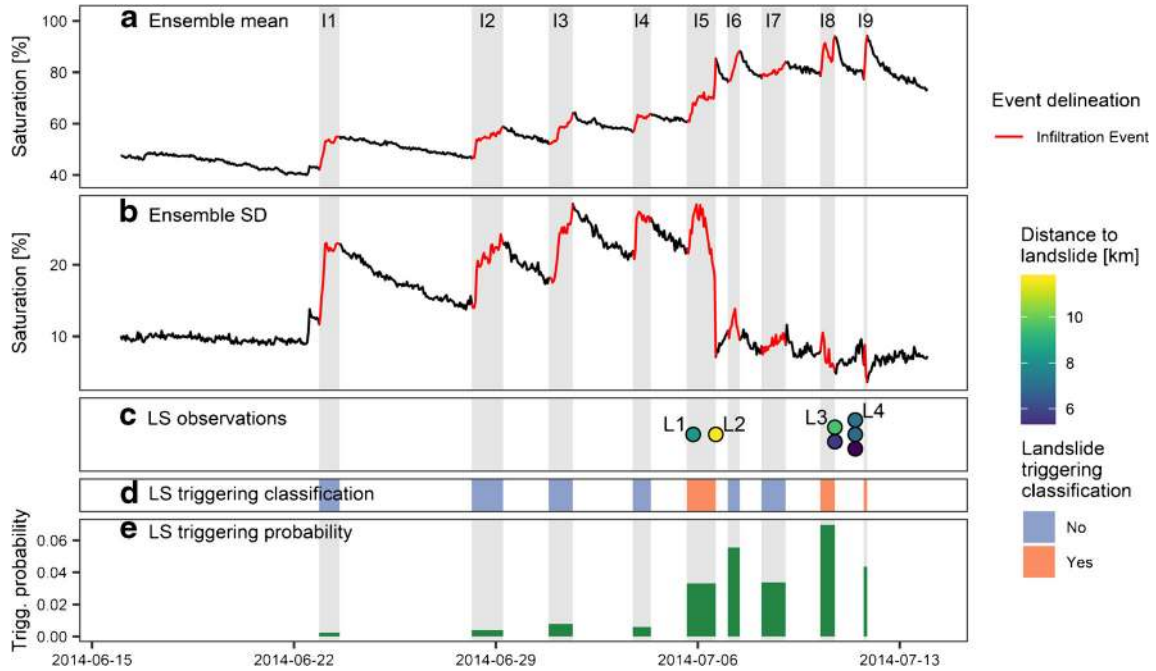

Fig. 5 Sample time series extract of the Plaffeien site (SwissSMEX, ETH) showing the ensemble mean (**a**) and the ensemble standard deviation (**b**) including the identified infiltration events in red, the observed landslide events within 15-km distance from the site showing the distance to the landslide (**c**), the corresponding landslide-triggering classification (**d**), and the landslide-triggering probability from the logistic regression model fit using all infiltration event properties (**e**)

(17 landslide-triggering events at 5 km vs. 424 landslide-triggering events at 40 km). For this dataset, an optimal compromise between a good forecast quality and a robust model fit can be identified at 10 to 15 km forecast distance (see AUC values in Table 2). Furthermore, the shape of the ROC curves indicates a better performance at the conservative range; i.e., the model is better at avoiding many false alarms at the expense of fewer true alarms.

Model drivers

To assess the forecast skill of individual infiltration event properties, the logistic regression model was fit including individual event properties only. Mean AUC values (blue bars) and the range of AUC values across all forecast distances (error bars) of individual event properties are shown in Fig. 7b. The best predictive power was found for the antecedent saturation (mean AUC = 0.65) and the two-week preceding maximum saturation (mean AUC = 0.64) which both describe predisposing conditions, as well as the saturation change (mean AUC = 0.65) and the 3-h maximum infiltration rate (mean AUC = 0.63) which describe more event dynamics; yet the predictive power of each individual properties is significantly lower compared with if all event properties are used (mean AUC = 0.76, Fig. 7a). All other event properties yield low to almost no predictive power by themselves (i.e. AUC values close to 0.5).

Furthermore, three model fits were obtained that each contain a subset of event properties consisting of a predisposing factor (antecedent saturation, 2-week preceding maximum saturation) and factors that describe event dynamics (saturation change, maximum 3-h infiltration rate). The best model fit using subsets of event properties was achieved by combining the antecedent saturation with the saturation change (mean AUC = 0.75, Fig. 7c). It yields almost the same forecast goodness as using all event properties. Calculation of the variation of inflation factor (VIF) and analysis of correlation between individual event properties revealed that no significant collinearity exists between the two event properties (not shown). In the interest of reducing the model complexity and understanding the driving forces in more detail, the model fit using antecedent saturation and saturation change only is explored in more detail in the following. It is referred to as the *baseline model* hereinafter.

To better understand the model drivers and parameters of the baseline model, Fig. 8a shows all landslide-triggering events (coloured dots) underlain by the number of landslide non-triggering events on a regular grid (blue tiles) in the antecedent saturation vs. saturation change space and for the 15-km forecast distance. The colour of the dots corresponds to the landslide-triggering rainfall type as recorded in the landslide database (*long-lasting*, *short-duration*, or *unknown*). A clustering of the non-triggering infiltration events is apparent between 60 and 75% antecedent saturation and between 2 and 10% saturation change. The group of the landslide-triggering events appears more scattered; however, a clustering is visible at high antecedent saturation or at high saturation change values. Since the total range of values is limited by a maximum of 100% saturation (sum of antecedent saturation and saturation change), this corresponds to a triangle in the antecedent saturation vs. saturation change space (Fig. 8a, b). The colouring of the points indicates that landslides due to long-lasting rainfall events cluster more in the

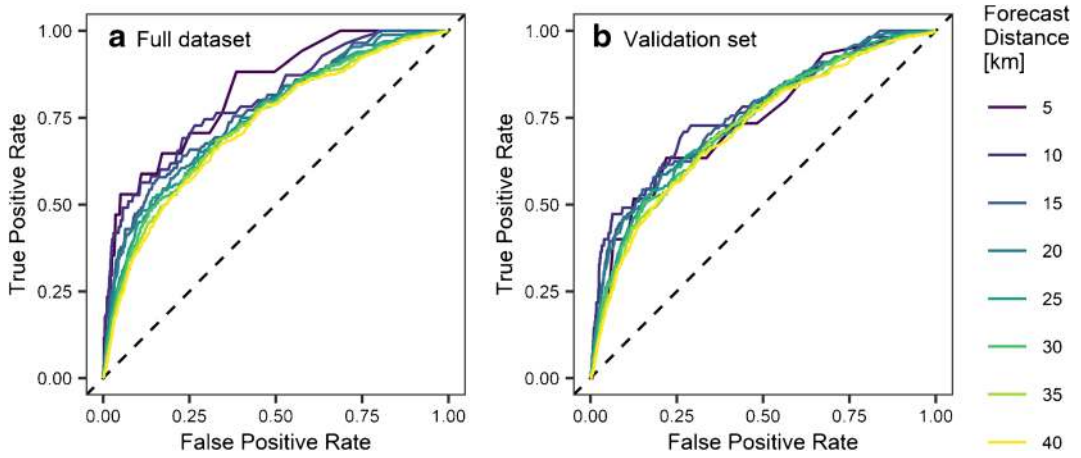

Fig. 6 ROC curves from the logistic regression model fit using all infiltration event properties based on the full dataset (**a**) and the validation set approach (**b**). The different line colours represent different forecast distances

bottom right corner while events due to short-duration rainfall are scattered more in the upper part of the plot.

The corresponding model fit is shown in Fig. 8b which depicts the triggering probability as a function of antecedent saturation and saturation change. The dotted and dashed lines show isolines of triggering probability for the 0.01–0.09 and 0.1–0.5 range, respectively. Note that applied thresholds always run parallel to the isolines of triggering probability. It can be observed that both model parameters are positive, i.e. with increased antecedent saturation or saturation change, the landslide-triggering probability increases too. This can be related to increased pore water pressures and decreased matric suction due to high infiltration amounts. Furthermore, the triggering probability increases more for one unit change of the saturation change event property as opposed to one unit change of the antecedent saturation which we relate to the different distribution of values for both parameters. While the non-triggering saturation change values cluster at the low end between 2 and 10%, the triggering values scatter in a higher band and range of values. Opposingly, non-triggering antecedent saturation values scatter in an intermediate band of values (60 to 75%) while triggering values scatter over the entire range with an apparent cluster at the high value end. The limited value range at the high value end of the antecedent saturation limits the ability of the model to distinguish between landslide triggering and non-triggering conditions for this property.

From Fig. 8a, it becomes apparent that landslide-triggering events due to long-lasting rainfall can be better distinguished by the antecedent saturation property while landslide-triggering events due to short-duration precipitation events are more variable in the saturation change space. To test for different model drivers, the logistic regression model was fit using triggering events due to the individual rainfall types only. Generally, the forecast quality increases if triggering events due to long-lasting rainfall events are considered only (mean AUC = 0.84, Fig. 7d). Conversely, the forecast quality decreases significantly if landslide triggering events due to short-duration rainfall events are included only (mean AUC = 0.71, Fig. 7g). Further to that, different model drivers become apparent for the different rainfall types (Fig. 7e, h). While the landslide events due to long-lasting rainfall can best be explained by the predisposing factors antecedent saturation (mean AUC = 0.73) and 2-week preceding maximum saturation (mean AUC = 0.71), landslide events due to short-duration rainfall events are best explained by the event properties of 3-h maximum infiltration rate (mean AUC = 0.67) and saturation change (mean

Table 2 Number of landslide-triggering and non-triggering infiltration events by forecast distance as well as AUC values for the full and validation set approach using all infiltration events and all available sensors

Forecast distance (km)	*N* infiltration events LS triggering	LS non-triggering	Ratio (%)	AUC Full dataset	Validation set
5	17	11,226	0.2	0.83	0.74
10	55	11,188	0.5	0.80	0.77
15	114	11,129	1.0	0.77	0.76
20	172	11,071	1.6	0.76	0.75
25	229	11,014	2.1	0.75	0.74
30	288	10,955	2.6	0.74	0.73
35	359	10,884	3.2	0.73	0.72
40	424	10,819	3.9	0.72	0.71

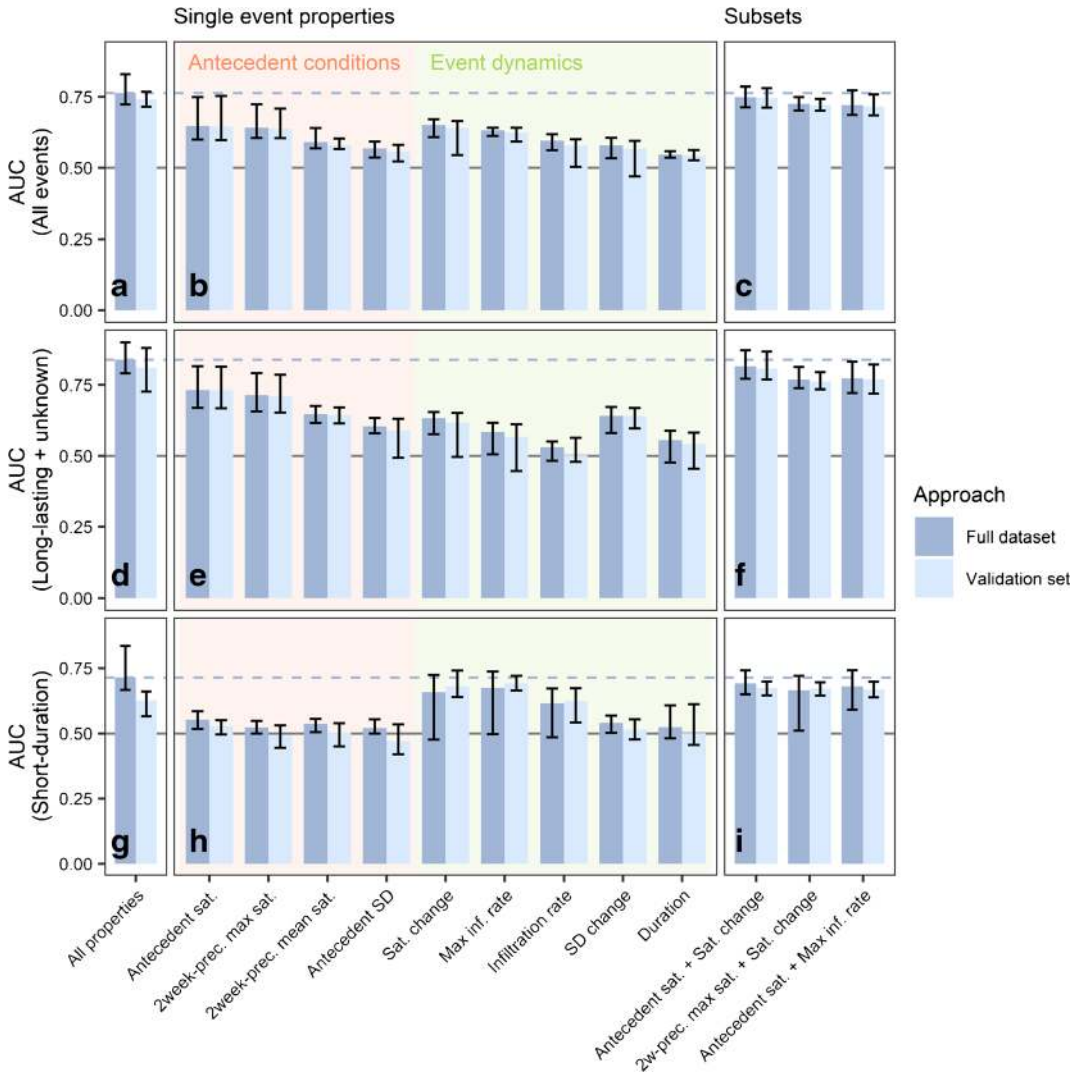

Fig. 7 AUC values from the logistic regression model fit using all infiltration event properties (**a**, **d**, **g**), individual infiltration event properties only (**b**, **e**, **h**), and subsets of infiltration event properties (**c**, **f**, **i**), including all landslide events (**a**-**c**), landslide events due to long-lasting and unknown rainfall events only (**d**-**f**), and due to short-duration rainfall events only (**g**-**i**). The blue bars show the mean and the black error bars show the range of AUC values across all forecast distances

AUC = 0.66) which both are a measure for event dynamics. Nevertheless, for both rainfall types, the antecedent saturation vs. saturation change model (i.e. the baseline model) still yields the best predictive performance amongst all model fits that are based on a subset of two event properties only (Fig. 7f, i, the three best combinations are shown only). We relate the overall better performance of the model fits based on long-lasting rainfall events to a better spatial representation of point scale soil moisture measurements for the regional landslide triggering conditions during long-lasting as opposed to short-duration rainfall events that can be spatially more variable. The different model drivers may indicate different triggering mechanisms between the two rainfall types where landslides due to long-lasting rainfall are triggered after continuous increase of the saturation as opposed to landslides due to short-duration rainfall that are triggered mainly due to high infiltration rates over a limited period of time.

Effect of homogenization by subgrouping and ensemble standardization

The baseline model was fit to different subgroups of infiltration events as well as to subgroups of VWC measurement site properties to test an improvement of the forecast quality due to homogenizing the infiltration event set. Grouping criteria included the season of occurrence, the site characteristics slope, vegetation cover, and texture expressed as the average sand content, as well as the regional landslide disposition expressed by the geographical location of the site (Fig. 9b–f).Generally, an improvement is to be expected if infiltration events of similar antecedent saturation vs. saturation change characteristics are grouped. However, an apparent improvement can also result if overfitting occurs due to a lower number of landslide-triggering events which can be identified as large AUC differences between the full dataset and the validation set approach.

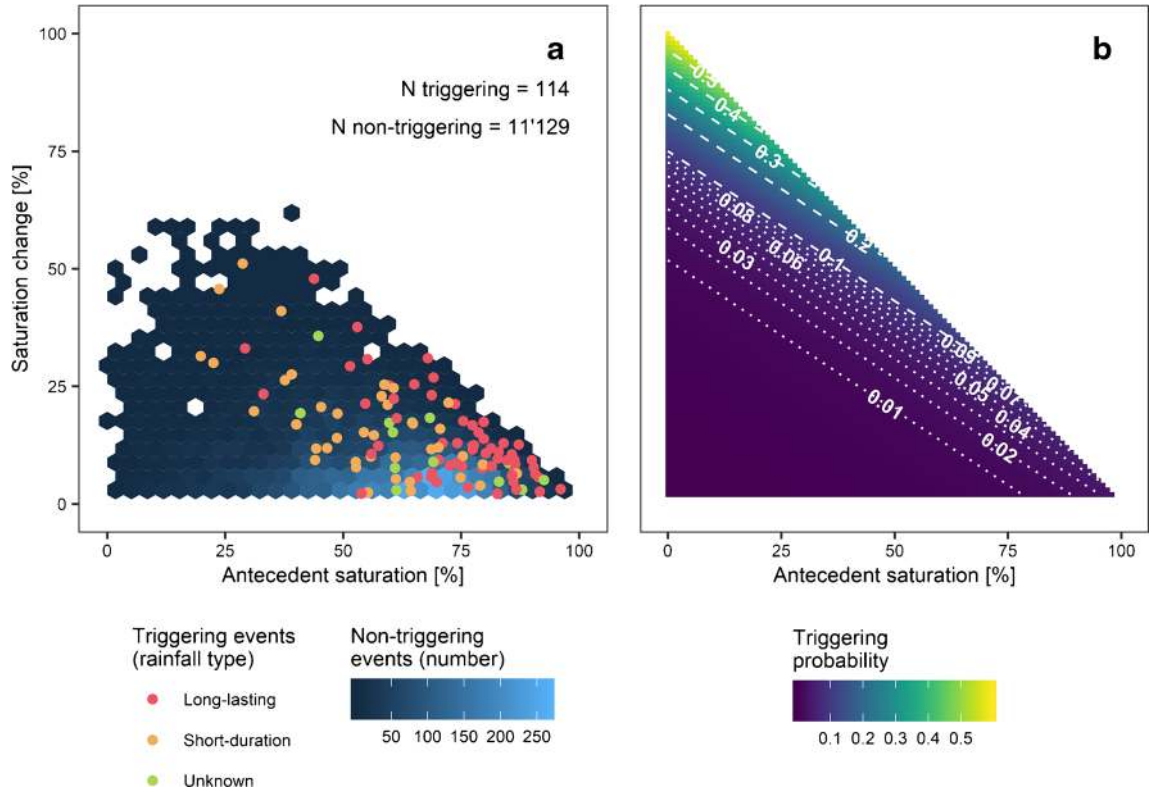

Fig. 8 **a** Individual landslide-triggering infiltration events at the 15-km forecast distance (coloured points) underlain by the density of landslide non-triggering infiltration events (blue tiles) in the antecedent saturation vs. saturation change space. The point colours indicate the landslide triggering rainfall type. **b** The triggering probability of the corresponding logistic regression model fit of the baseline model (i.e. using antecedent saturation and saturation change only). The white lines indicate isolines of equal landslide-triggering probability

While some subgroups show a considerable improvement of the forecast goodness, others show either a decreasing AUC or large discrepancies to the validation set approach. The largest forecast quality increase can be observed for the ≤ 30% as well as the > 50% sand split groups in the texture class which is accompanied by a larger spread across all forecast distances. Smaller improvements could be observed for the summer months April, May, and June (AMJ) and July, August, and September (JAS) as well as flat and open-land measurement sites (note that in the dataset, most flat sites are open-land sites and vice versa, correlation not shown). The decrease of the forecast quality for the winter months January, February, and March (JFM) and October, November, and December (OND) can be attributed to overfitting due to a low number of landslide-triggering infiltration events whereas the decrease of the forecast quality for hillslope and forested sites might be related to increased inhomogeneities within the subgroup or a decreased representativeness for the locations of landslide occurrence. No significant forecast quality increases or decreases can be found for the geography split which further shows problems of overfitting (Alpine sites).

Furthermore, the effect of homogenizing the different measurement set-ups is tested by fitting the baseline model to infiltration events derived from more homogenized ensembles of sensors (Fig. 10). Slight increases in the forecast quality can be observed if shallow sensors (≤ 30 cm depth) are considered only (mean AUC = 0.77) or if three sensors at different depths are used per site (3-layered profile, mean AUC = 0.75). While the first indicates that the highest information content is stored in the uppermost sensors, the latter demonstrates the potential of homogenizing different measurement set-ups from different monitoring networks. The forecast quality decreases if only deep-seated sensors (> 30 cm depth) are considered (mean AUC = 0.70).

Discussion

Representativeness of point measurements for regional landslide activity

The representativeness of point scale soil moisture measurements for landslide activity at regional scale can be discussed with regard to the spatial and temporal variability of soil moisture. Most prominently, our study could demonstrate a strong distance dependency of the forecast quality. While the forecast goodness decreases with distance, some forecast skill remains even at the 40-km forecast distance. We explain the distance dependency at the regional scale to be predominantly driven by meteorological factors, in a first order by regional scale variations of the precipitation characteristics such as phase, intensity, or duration that determine the amount of infiltrating water, and secondarily by variations of the irradiance, wind, or humidity that drive the drying up of the soil column (e.g. Crow et al. 2012). Mittelbach and Seneviratne (2012) also found coherent soil moisture dynamics across the SwissSMEX sites showing that relative anomalies in

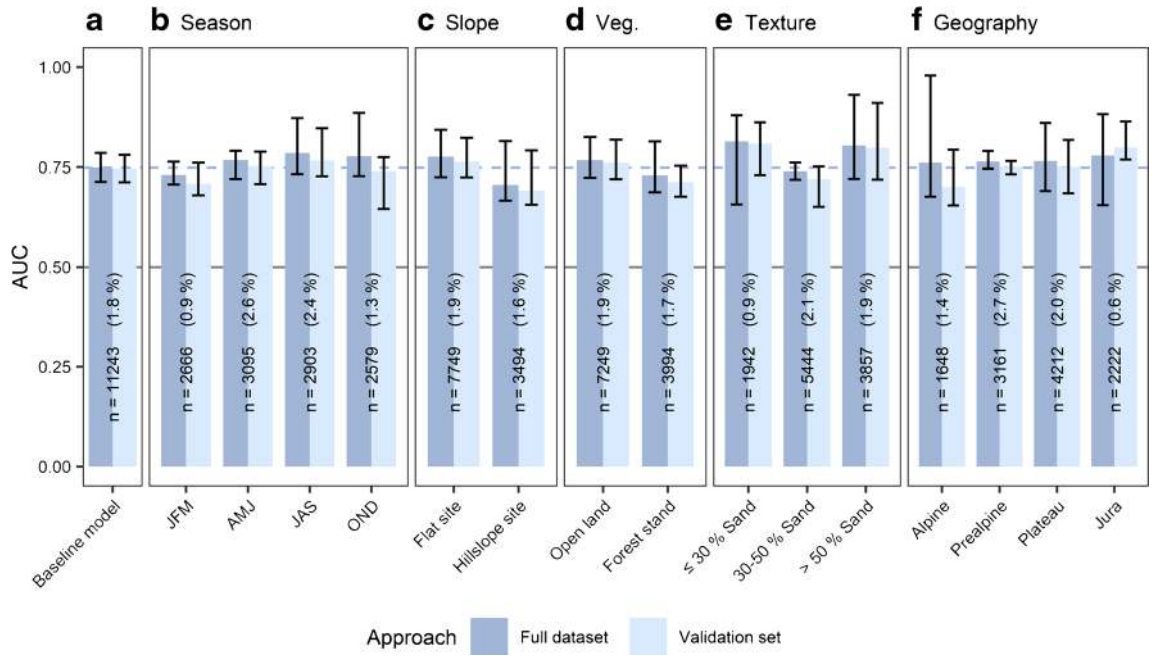

Fig. 9 **a** AUC values from the logistic regression model fit of the baseline model (i.e. using antecedent saturation and saturation change only) using all infiltration events in comparison with model fits based on subsets of infiltration events grouped by season (**b**), the site characteristics slope (**c**), vegetation (**d**), and sand fraction (**e**), and by geography (**f**). The blue bars show the mean and the black error bars show the range of AUC values across all forecast distances. The numbers denote the number of infiltration events and the percentage number shows the average rate of landslide triggering events

soil moisture were generally consistent within much of Switzerland despite variations in absolute soil moisture content. In addition, Seneviratne et al. (2012) identified high correlation in hydrological conditions in the Thur river basin in Northeastern Switzerland, within a radius of ca. 40 km. Furthermore, several studies report a decreasing spatial soil moisture variability during saturated conditions (e.g. Famiglietti et al. 1998; Western et al. 2003; Rosenbaum et al. 2012). Since shallow landslides predominantly occur during saturated conditions, this may explain why the forecast skill extends even to a distance of 40 km. Considerably better forecast skill was found for landslides due to long-lasting rainfall events than due to events of short duration. While the first goes along with a more long-term pre-wetting of the soil and can be characterized by specific antecedent saturation conditions, the latter may be driven by the triggering event only and thus, the event dynamics are more important too. We explain the overall worse performance of landslides triggered by short-duration rainfall events by the connection to convective storms and thus by a higher spatial variability.

Next to meteorology and climate, the spatial variability of soil moisture is influenced by factors related to topography, vegetation, and soil properties that typically vary from the point to watershed scale (Crow et al. 2012). Testing for these factors was attempted by grouping the infiltration events by specific site characteristics under the assumption that subgroups of similar sites with homogeneous infiltration event characteristics will have a better forecast skill. Significant improvement could be achieved for flat and open-land measurement sites whereas sloped and forested sites showed a worse performance. Since both site factors are correlated in the dataset (most flat sites are open-land sites and most sloped sites are forested), it is not possible to determine which explanatory factor is driving the forecast quality differences. The presence of a topographic gradient or a forest cover may both account for inhomogeneous soil moisture regimes and thus reduce the spatial representativeness of point measurements: Increased spatial soil moisture variability is reported on slopes particularly during wet conditions due to lateral flow, convergence, or divergence of overland flow and during dry conditions due to irradiance differences from aspect and slope variations (e.g. Western et al. 1999). The presence of a forest may greatly impact the infiltration process by variations in the hydraulic conductivity due to root activity and by throughfall. During dry conditions, shading and plant water uptake impact evapotranspiration rates (Famiglietti et al. 1998; Atchley and Maxwell 2011). However, a better forecast quality for landslide prediction was reported if sloped sites are used as the gravity-dominated regime makes it easier to identify critically saturated conditions at hillslopes (Thomas et al. 2019). Finally, spatial soil moisture variability is reported for soils of different texture, structure, porosity, and macroporosity. A higher sand content was previously reported to increase drainage and decrease storage capacity at the satellite footprint scale (Panciera 2009). In this study, significant increase of the forecast goodness could be identified for the high and the low sand content groups which demonstrates the importance of accounting for local to regional variations in the soil properties.

On a temporal scale, it could be shown that summer months perform better than winter months. Since precipitation is evenly distributed over the year in most areas of Switzerland, this leads to high soil moisture values throughout the year with a distinct dry period and increased temporal variability during the summer

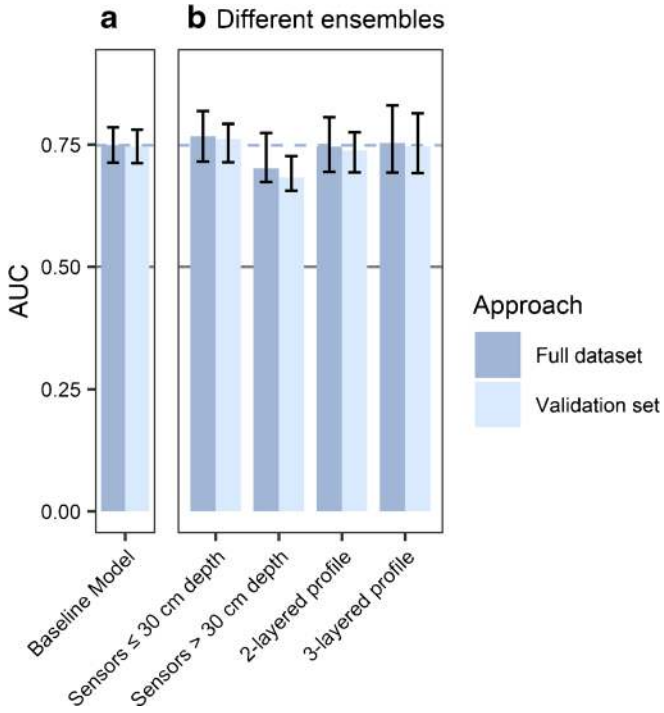

Fig. 10 **a** AUC values from the logistic regression model fit of the baseline model (i.e. using antecedent saturation and saturation change only) using all sensors in comparison with model fits based on different ensembles of sensors (**b**). The blue bars show the mean and the black error bars show the range of AUC values across all forecast distances

months (Pellet and Hauck 2017). We associate the better forecast quality of the summer months with the higher variation of soil moisture values due to the elevated drying of the soils. During this regime, the model can better distinguish critically saturated conditions. Furthermore, in winter and spring, infiltration not only depends on precipitation patterns but also on its form (snow/rain) and snow melt (snow distribution and radiation) which may increase the spatial variability.

Soil moisture as a proxy for landslide occurrence

As previously stated, soil moisture measurements provide a direct measurement of infiltration, storage and exfiltration processes as opposed to the characterization of precipitation events only. However, it still remains a proxy for the actual landslide triggering mechanism which is the localized increase of pore water pressure and decrease of matric suction. In the following, the performance of soil moisture as a proxy will be discussed in more detail.

It was found that the combination of the two event properties antecedent saturation and saturation change can best separate landslide triggering from non-triggering events. While the first property describes the medium-term hydrological preconditioning of the ground, the latter is a measure for the short-term event dynamics. This is in line with recent propositions of a cause-trigger-framework (Bogaard and Greco 2016, 2018) that is based on the combination the two process domains of hydrological prewetting (cause factors) and meteorological triggering (trigger factors). While the combination of the two factor domains is beneficial for distinguishing landslide triggering from non-triggering events, the respective importance varies with the landslide triggering rainfall type. During long-lasting rainfall events, the main driver for increased landslide probability is the subsequent saturation increase in the entire soil column to nearly saturated conditions prior to the triggering rainfall event. For short-duration rainfall events in contrast, it is rather the saturation increase and rate of saturation increase during the triggering rainfall event that drive the triggering probability. This points to different infiltration regimes and different landslide triggering mechanisms.

The good performance of the two event properties can further be related to the standardization of the volumetric soil moisture time series. It has been shown previously for the SwissSMEX dataset that the spatial variability of the mean soil moisture is greater than the spatial variability due to temporal dynamics (Mittelbach and Seneviratne 2012). Therefore, using standardized soil moisture states can effectively reduce spatial variability due to site characteristics. Conversely, the two event properties of duration and infiltration rate further incorporate a timing component which is not standardized for and which might cause larger spatial variability and thus a worse model performance. The poor forecast goodness of event properties related to the standard deviation time series might be related to the considerable differences in the measurement set-ups both the in number of sensors and installation depths. Additionally, the worse performance of the 2-week antecedent mean or maximum saturation can be attributed to a rather indirect indication of the current wetness state of the ground.

Finally, the potential of soil moisture as a proxy for landslide occurrence is limited by the value distribution. As opposed to precipitation amounts, intensity, or duration, both the antecedent saturation and the saturation change have an upper physical limit (i.e. when all the pores are filled with water), which is also referred to as a bounded distribution (Western et al. 2003). Furthermore, the distribution of non-triggering infiltration events is clustered around 60 to 75% antecedent saturation which has been reported as the potential range of the field capacity (Assouline and Or 2014). In this state, which is typically reached several

hours to days after maximum saturation, all gravitational water has drained and the rate of water content change within the soil column becomes small. The prevalence of this state further limits the range of values at saturated and near-saturated conditions and thus the potential of volumetric soil moisture measurements to distinguish critical from non-critical hydrological conditions. A better proxy in this respect could be the matric potential which during fully saturated conditions can further measure different levels of porewater pressure. Also, it would serve as a direct measure for the loss of matric suction.

Methodological limitations

Specific limitations arise from the use of existing soil moisture monitoring data. It was shown that the dataset size limits the robustness of the model fit due to a low number of measurement sites and short time series. In our study, an optimal compromise between robustness of the model and the forecast quality could be achieved at a forecast distance of 10 to 15 km which could probably be at shorter distance if more data would be available.

However, if longer time series are used, data quality issues might arise that are particularly connected to long-term monitoring using soil moisture probes. In the used dataset, drifts, trends, or variability changes in the soil moisture time series were observed for single sensors. These changes can be attributed to actual processes in the ground but could simply be related to technical issues such as the loosening of the sensor contact to the ground or sensor defects. Such changes introduce inhomogeneities that significantly impact the normalized soil moisture calculation and thus the forecast goodness. Further inhomogeneities can result if defective sensors are replaced by disturbing the soil column or if sensors are not replaced at exactly the same location.

Furthermore, our standardization approach that uses the minimum and 99.9 percentile rank of a sensor time series overstates the temporal soil moisture variability for sensors that never reached fully saturated conditions or conditions near the residual water content. This may occur particularly for deep-seated sensors and sensors with a short period of record. An improvement in this respect could be the standardization by measurements of porosity and texture (to estimate saturated and residual water content) at each sensor location. However, this is only possible with destructive methods, i.e. only after the end of the measurement time series of a sensor.

Finally, the analysis is limited by the composition of the landslide dataset. In this study, landslide records are based on newspaper articles. Consequently, small events with less relevance to the public or events that caused no damage to infrastructure or buildings are missing which can introduce many false negatives in the ROC analysis. Further uncertainty was introduced by partly imprecise timing and location information and an incomplete process description and thus the probability of a misclassification as shallow landslides. Finally, no conclusion on the event magnitude can be drawn since the exact number of landslides triggered is not always stated in the database.

Implications for an operational LEWS

From the above, implications can be formulated for the use of soil moisture monitoring data in a regional LEWS. While a high density of monitoring sites is beneficial, the density should ideally cover the spatial variability of precipitation events and soil properties, and it should focus on areas with a high landslide susceptibility as suggested by other studies (e.g. Baum and Godt 2010). Furthermore, measurement sites should not be influenced by local-scale hydrological conditions (e.g. groundwater interactions, hillslope flow); a homogeneous measurement set-up (e.g. depth of installation, number of sensors) is advantageous and measuring porosity at the installation depth would help to calculate normalized soil moisture. Open questions remain about the representativeness of sloped and forested sites, the value of modelled soil moisture, as well as the potential of other soil wetness measurements such as the matric potential.

Conclusions

This study could demonstrate that in-situ soil moisture data effectively contains specific information on the regional landslide activity. The forecast goodness is strongly dependent on the distance between the network location and hillslopes with landslide activity, and the goodness increases with decreasing distances between measurement sites and landslides. At the same time, a shorter forecast distance reduces the number of observed landslide events and thus the robustness of the model. In the present study, a good trade-off between forecast goodness and robustness of the model was found for a forecast distance of 10 to 15 km. Furthermore, the statistical model performs significantly better for landslides due to large-scale long-lasting precipitation events as opposed to landslides due short-duration rainfall events that are connected to more local-scale convective storms.

It could be shown that normalizing the VWC values and integrating the measurements at several depths are useful. Furthermore, homogenization of the dataset by grouping the infiltration event set by site characteristics and standardizing the ensembles of sensors could partially improve the forecast goodness. In this respect, grouping by soil texture led to the largest forecast improvement.

The focus on infiltration events proved to be a good unit of reference for the statistical analysis and it could be shown that quantifying the saturation at the onset and the change of saturation during an infiltration event explains most of the variability. However, characterizing the antecedent saturation conditions is more important for landslides triggered by long-lasting rainfall events whereas describing event dynamics (e.g. saturation change, infiltration rate) is more important for landslides triggered by short-duration rainfall.

The findings imply that the density of the measurement networks strongly impacts the forecast quality of a LEWS and that an optimal design of such a network should consider the spatial variability of meteorological events and soil properties. Questions remain about the ideal measurement set-up, potential locations of soil moisture measurements (flat or inclined sites), or whether a similar forecast goodness could be achieved by modelling approaches only.

Acknowledgements

We thank Umwelt und Energie (Kanton Luzern), Amt für Umweltschutz (Kanton Uri), Amt für Umwelt (Kanton Solothurn), Monitron AG, inNET Monitoring AG, and Meteotest AG for providing soil moisture data. We thank Heidi Mittelbach, Karl Schroff, Irene Lehner, Dominik Michel and Martin Hirschi as well as Elisabeth Graf Pannatier, Markus Schlegel, Benno Staub, Oliver Schramm, Roger Köchli, Karl Schroff, Peter Jakob and Simpal Kumar, and further Cécile Pellet for installation, maintenance and storage of the VWC measurements on the SMEX, the LWF, and the SOMOMOUNT plots, respectively. We thank the two

anonymous reviewers for their useful and constructive comments towards improving the manuscript.

Funding information

This research project was financially supported by the Swiss National Science Foundation (project number 175785) and is part of the programme *Climate Change Impacts on Alpine Mass Movements* of the Swiss Federal Research Institute WSL.

References

Aleotti P (2004) A warning system for rainfall-induced shallow failures. Eng Geol 73:247–265. https://doi.org/10.1016/j.enggeo.2004.01.007

Assouline S, Or D (2014) The concept of field capacity revisited: defining intrinsic static and dynamic criteria for soil internal drainage dynamics. Water Resour Res 50:4787–4802. https://doi.org/10.1002/2014WR015475

Atchley AL, Maxwell RM (2011) Influences of subsurface heterogeneity and vegetation cover on soil moisture, surface temperature and evapotranspiration at hillslope scales. Hydrogeol J 19:289–305. https://doi.org/10.1007/s10040-010-0690-1

Babaeian E, Sadeghi M, Jones SB, Montzka C, Vereecken H, Tuller M (2019) Ground, proximal and satellite remote sensing of soil moisture. Rev Geophys 2018:RG000618. https://doi.org/10.1029/2018RG000618

Badoux A, Andres N, Techel F, Hegg C (2016) Natural hazard fatalities in Switzerland from 1946 to 2015. Nat Hazards Earth Syst Sci 16:2747–2768. https://doi.org/10.5194/nhess-16-2747-2016

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

Bell R, Mayer J, Pohl J, Greiving S, Glade T (2010) Zur Bedeutung eines integrativen Frühwarnsystems für gravitative Massenbewegungen. In: Bell R, Mayer J, Pohl J, Greiving S, Glade T (eds) Integrative Frühwarnsysteme für gravitative Massenbewegungen (ILEWS). Monitoring, Modellierung, Implementierung. Klartext Verlag, Essen

Berne A, Ferrari A, Frank F, Gruner U, Haeberle J, Hählen N, Huggel Ch, Kos A, Lehmann P, McArdell B, Or D, Sättele M, Schleiss M, Springman SM, Stähli M, Tobler D, Van Herwijnen A (2013) Strategies towards design of next-generation early warning systems (EWS) for rapid mass movements

Bogaard T, Greco R (2018) Invited perspectives: hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: proposing hydro-meteorological thresholds. Nat Hazards Earth Syst Sci 18:31–39. https://doi.org/10.5194/nhess-18-31-2018

Bogaard TA, Greco R (2016) Landslide hydrology: from hydrology to pore pressure. Wiley Interdiscip Rev Water 3:439–459. https://doi.org/10.1002/wat2.1126

Brocca L, Ciabatta L, Moramarco T, Ponziani F, Berni N, Wagner W (2016) Use of satellite soil moisture products for the operational mitigation of landslides risk in Central Italy. In: Satellite Soil Moisture Retrieval. Elsevier, pp 231–247

Caine N (1980) The rainfall intensity - duration control of shallow landslides and debris flows. Geogr Ann A, Phys Geogr 62:23–27. https://doi.org/10.1080/04353676.1980.11879996

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. Proc EGS Plinius Conf held Maratea 439–448

Chen Y, Or D (2006) Geometrical factors and interfacial processes affecting complex dielectric permittivity of partially saturated porous media. Water Resour Res 42:1–9. https://doi.org/10.1029/2005WR004744

Chleborad AF (2000) Preliminary method for anticipating the occurrence of rainfall-induced landslides in Seattle, Washington. 30. Doi: open-file report, vol. 00-469

Comegna L, Damiano E, Greco R, Guida A, Olivares L, Picarelli L (2016) Field hydrological monitoring of a sloping shallow pyroclastic deposit. Can Geotech J 53:1125–1137. https://doi.org/10.1139/cgj-2015-0344

Crosta GB, Frattini P (2001) Rainfall thresholds for triggering soil slips and debris flow. Proc 2nd EGS Plinius Conf Mediterr storms 463–487

Crow WT, Berg AA, Cosh MH, Loew A, Mohanty BP, Panciera R, de Rosnay P, Ryu D, Walker JP (2012) Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. Rev Geophys 50:1–20. https://doi.org/10.1029/2011RG000372

Crozier MJ (1999) Prediction of rainfall-triggered landslides: a test of the antecedent water status model. Earth Surf Process Landf 24:825–833. https://doi.org/10.1002/(SICI)1096-9837(199908)24:9<825::AID-ESP14>3.0.CO;2-M

Devoli G (2017) Workshop "Regional early warning systems for rainfall- and snowmelt-induced landslides. Need for an international forum?" summary report

Famiglietti JS, Rudnicki JW, Rodell M (1998) Variability in surface moisture content along a hillslope transect: Rattlesnake Hill, Texas. J Hydrol 210:259–281. https://doi.org/10.1016/S0022-1694(98)00187-5

Fawcett T (2006) An introduction to ROC analysis. Pattern Recogn Lett 27:861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Ferretti M, Schaub M (2014) Monitoring European forests - detecting and understanding changes. For Ecol Manag 311:1–2. https://doi.org/10.1016/j.foreco.2013.08.027

Froude MJ, Petley DN (2018) Global fatal landslide occurrence from 2004 to 2016. Nat Hazards Earth Syst Sci 18:2161–2181. https://doi.org/10.5194/nhess-18-2161-2018

Geisser H, Bohne J, Kaiser C (2011) Bodenmessstation Erstfeld (Pfaffenmatt) Jahresbericht 2010. Altdorf

Glade T (2000) Modelling landslide-triggering rainfalls in different regions of New Zealand - the soil water status model. Zeitschrift für Geomorphol Suppl.-Bd.:63–84

Godt JW, Baum RL, Chleborad AF (2006) Rainfall characteristics for shallow landsliding in Seattle, Washington, USA. Earth Surf Process Landf 31:97–110. https://doi.org/10.1002/esp.1237

Godt JW, Baum RL, Lu N (2009) Landsliding in partially saturated materials. Geophys Res Lett 36:1–5. https://doi.org/10.1029/2008GL035996

Greco R, Bogaard T (2016) The influence of non-linear hydraulic behavior of slope soil covers on rainfall intensity-duration thresholds. In: Aversa S, Cascini L, Picarelli L, Scavia C (eds) Landslides and Engineered Slopes. Experience, Theory and Practice. CRC Press, London, pp 1021–1025. https://doi.org/10.1201/9781315375007

Guzzetti F, Peruccacci S, Rossi M, Stark CP (2007) Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorog Atmos Phys 98:239–267. https://doi.org/10.1007/s00703-007-0262-7

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

Guzzetti F, Gariano SL, Peruccacci S, Brunetti MT, Marchesini I, Rossi M, Melillo M (2020) Geographical landslide early warning systems. Earth Sci Rev 200:102973. https://doi.org/10.1016/j.earscirev.2019.102973

Hilker N, Badoux A, Hegg C (2009) The Swiss flood and landslide damage database 1972–2007. Nat Hazards Earth Syst Sci 9:913–925. https://doi.org/10.5194/nhess-9-913-2009

Jakob M, Weatherly H (2003) A hydroclimatic threshold for landslide initiation on the North Shore Mountains of Vancouver, British Columbia. Geomorphology 54:137–156. https://doi.org/10.1016/S0169-555X(02)00339-2

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer New York, New York

Kelleners TJ, Robinson DA, Shouse PJ, Ayars JE, Skaggs TH (2005) Frequency dependence of the complex permittivity and its impact on dielectric sensor calibration in soils. Soil Sci Soc Am J 69:67–76

Kirschbaum D, Stanley T (2018) Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness. Earths Future 6:505–523. https://doi.org/10.1002/2017EF000715

Lehmann P, Gambazzi F, Suski B, Baron L, Askarinejad A, Springman SM, Holliger K, Or D (2013) Evolution of soil wetting patterns preceding a hydrologically induced landslide inferred from electrical resistivity survey and point measurements of volumetric water content and pore water pressure. Water Resour Res 49:7992–8004. https://doi.org/10.1002/2013WR014560

Leonarduzzi E, Molnar P, McArdell BW (2017) Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data. Water Resour Res 53:6612–6625. https://doi.org/10.1002/2017WR021044

Looyenga H (1965) Dielectric constants of heterogeneous mixtures. Physica 31:401–406. https://doi.org/10.1016/0031-8914(65)90045-5
Loretz C, Ruckstuhl C (2017) Bodenfeuchtemonitoring mit Informationssystem - Pilotprogramm Anpassung an den Klimawandel : Schlussbericht
Lu N, Godt JW, Wu DT (2010) A closed-form equation for effective stress in unsaturated soil. Water Resour Res 46:1–14. https://doi.org/10.1029/2009wr008646
Martelloni G, Segoni S, Fanti R, Catani F (2012) Rainfall thresholds for the forecasting of landslide occurrence at regional scale. Landslides 9:485–495. https://doi.org/10.1007/s10346-011-0308-2
Martelloni G, Segoni S, Lagomarsino D, Fanti R, Catani F (2013) Snow accumulation/melting model (SAMM) for integrated use in regional scale landslide early warning systems. Hydrol Earth Syst Sci 17:1229–1240. https://doi.org/10.5194/hess-17-1229-2013
Michoud C, Bazin S, Blikra LH, Derron MH, Jaboyedoff M (2013) Experiences from site-specific landslide early warning systems. Nat Hazards Earth Syst Sci 13:2659–2673. https://doi.org/10.5194/nhess-13-2659-2013
Mirus B, Morphew M, Smith J (2018a) Developing hydro-meteorological thresholds for shallow landslide initiation and early warning. Water 10:1274. https://doi.org/10.3390/w10091274
Mirus BB, Becker RE, Baum RL, Smith JB (2018b) Integrating real-time subsurface hydrologic monitoring with empirical rainfall thresholds to improve landslide early warning. Landslides 15:1–11. https://doi.org/10.1007/s10346-018-0995-z
Mittelbach H, Casini F, Lehner I, Teuling AJ, Seneviratne SI (2011) Soil moisture monitoring for climate research: evaluation of a low-cost sensor in the framework of the Swiss soil moisture experiment (SwissSMEX) campaign. J Geophys Res Atmos 116:1–11. https://doi.org/10.1029/2010JD014907
Mittelbach H, Seneviratne SI (2012) A new perspective on the spatio-temporal variability of soil moisture: temporal dynamics versus time-invariant contributions. Hydrol Earth Syst Sci 16:2169–2179. https://doi.org/10.5194/hess-16-2169-2012
Mittelbach H, Lehner I, Seneviratne SI (2012) Comparison of four soil moisture sensor types under field conditions in Switzerland. J Hydrol 430–431:39–49. https://doi.org/10.1016/j.jhydrol.2012.01.041
Napolitano E, Fusco F, Baum RL, Godt JW, De Vita P (2016) Effect of antecedent-hydrological conditions on rainfall triggering of debris flows in ash-fall pyroclastic mantled slopes of Campania (southern Italy). Landslides 13:967–983. https://doi.org/10.1007/s10346-015-0647-5
Panciera R (2009) Effect of land surface heterogeneity on satellite near surface soil moisture observations. University of Melbourne, Parkville
Pellet C, Hauck C (2017) Monitoring soil moisture from middle to high elevation in Switzerland: set-up and first results from the SOMOMOUNT network. Hydrol Earth Syst Sci 21:3199–3220. https://doi.org/10.5194/hess-21-3199-2017
Piciullo L, Calvello M, Cepeda JM (2018) Territorial early warning systems for rainfall-induced landslides. Earth Sci Rev 179:228–247. https://doi.org/10.1016/j.earscirev.2018.02.013
Ponziani F, Pandolfo C, Stelluti M, Berni N, Brocca L, Moramarco T (2012) Assessment of rainfall thresholds and soil moisture modeling for operational hydrogeological risk prevention in the Umbria region (Central Italy). Landslides 9:229–237. https://doi.org/10.1007/s10346-011-0287-3
Reichenbach P, Cardinali M, De Vita P, Guzzetti F (1998) Regional hydrological thresholds for landslides and floods in the Tiber River Basin (Central Italy). Environ Geol 35:146–159. https://doi.org/10.1007/s002540050301
Rickli C, Raetzo H, McArdel B, Presler J (2008) Ereignisanalyse Hochwasser 2005, Teil 2 - Analyse von Prozessen, Massnahmen und Gefahrengrundlagen. In: Bezzola GR, Hegg C (eds) Umwelt-Wissen. Bundesamt für Umwelt BAFU, Eidgenössische Forschungsanstalt WSL, p 429
Robinson DA, Campbell CS, Hopmans JW, Hornbuckle BK, Jones SB, Knight R, Ogden F, Selker J, Wendroth O (2008) Soil moisture measurement for ecological and hydrological watershed-scale observatories: a review. Vadose Zone J 7:358–389. https://doi.org/10.2136/vzj2007.0143
Rosenbaum U, Bogena HR, Herbst M, Huisman JA, Peterson TJ, Weuthen A, Western AW, Vereecken H (2012) Seasonal and event dynamics of spatial soil moisture patterns at the small catchment scale. Water Resour Res 48:1–22. https://doi.org/10.1029/2011WR011518
Schaub M, Dobbertin M, Kräuchi N, Dobbertin MK (2011) Preface-long-term ecosystem research: understanding the present to shape the future. Environ Monit Assess 174:1–2. https://doi.org/10.1007/s10661-010-1756-1
Segoni S, Piciullo L, Gariano SL (2018a) A review of the recent literature on rainfall thresholds for landslide occurrence. Landslides 15:1483–1501. https://doi.org/10.1007/s10346-018-0966-4
Segoni S, Rosi A, Lagomarsino D, Fanti R, Casagli N (2018b) Brief communication: using averaged soil moisture estimates to improve the performances of a regional-scale landslide early warning system. Nat Hazards Earth Syst Sci 18:807–812. https://doi.org/10.5194/nhess-18-807-2018
Seneviratne SI, Corti T, Davin EL, Hirschi M, Jaeger EB, Lehner I, Orlowsky B, Teuling AJ (2010) Investigating soil moisture-climate interactions in a changing climate: a review. Earth Sci Rev 99:125–161. https://doi.org/10.1016/j.earscirev.2010.02.004
Seneviratne SI, Lehner I, Gurtz J, Teuling AJ, Lang H, Moser U, Grebner D, Menzel L, Schroff K, Vitvar T, Zappa M (2012) Swiss prealpine Rietholzbach research catchment and lysimeter: 32 year time series and 2003 drought event. Water Resour Res 48:1–20. https://doi.org/10.1029/2011WR011749
Smith JB, Baum RL, Mirus BB, Michel AR, Stark B (2017) Results of hydrologic monitoring on landslide-prone coastal bluffs near Mukilteo. Washington. 1–33
Springman SM, Jommi C, Teysseire P (2003) Instabilities on moraine slopes induced by loss of suction: a case history. Geotechnique 53:3–10. https://doi.org/10.1680/geot.2003.53.1.3
Stähli M, Sättele M, Huggel C, McArdell BW, Lehmann P, Van Herwijnen A, Berne A, Schleiss M, Ferrari A, Kos A, Or D, Springman SM (2015) Monitoring and prediction in early warning systems for rapid mass movements. Nat Hazards Earth Syst Sci 15:905–917. https://doi.org/10.5194/nhess-15-905-2015
Terzaghi K (1943) Theoretical soil mechanics. John Wiley & Sons, Inc., Hoboken
Thomas MA, Collins BD, Mirus BB (2019) Assessing the feasibility of satellite-based thresholds for hydrologically driven landsliding. Water Resour res 9006–9023. doi: https://doi.org/10.1029/2019WR025577
Topp GC, Davis JL, Annan AP (1980) Electromagnetic determination of soil water content: measurements in coaxial transmission lines. Water Resour Res 16:574–582. https://doi.org/10.1029/WR016i003p00574
Umwelt und Energie Kanton Luzern (2019) Bodenfeuchte: Messnetz Kanton Luzern. In: uwe.lu.ch. https://uwe.lu.ch/themen/bodenschutz/bodenfeuchte. Accessed 1 Oct 2019
Western AW, Grayson RB, Blöschl G, Willgoose GR, McMahon TA (1999) Observed spatial organization of soil moisture and its relation to terrain indices. Water Resour Res 35:797–810. https://doi.org/10.1029/1998WR900065
Western AW, Grayson RB, Blöschl G, Wilson DJ (2003) Spatial variability of soil moisture and its implications for scaling. In: Pachepsky Y, Radcliffe DE, Selim HM (eds) Scaling methods in soil physics. CRC Press, Boca Raton, p 456
Wieczorek GF, Glade T (2005) Climatic factors influencing occurrence of debris flow. In: Jakob M, Hungr O (eds) Debris-flow hazards and related phenomena. Springer, Berlin Heidelberg, Berlin, Heidelberg
Zhuo L, Dai Q, Han D, et al (2019) Evaluation of remotely sensed soil moisture for landslide hazard assessment. IEEE J-STARS 12:162–173. https://doi.org/10.1109/JSTARS.2018.2883361

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10346-020-01400-y) contains supplementary material, which is available to authorized users.

A. Wicki (✉) · **P. Waldner** · **M. Stähli**
Swiss Federal Institute for Forest,
Snow and Landscape Research WSL,
Zürcherstrasse 111, 8903, Birmensdorf, Switzerland
Email: adrian.wicki@wsl.ch

P. Lehmann
Institute of Biogeochemistry and Pollutant Dynamics,
ETH Zurich,
Universitätsstrasse 16, 8092, Zürich, Switzerland

C. Hauck
Department of Geosciences,
University of Fribourg,
Chemin du Musée 4, 1700, Fribourg, Switzerland

S. I. Seneviratne
Institute for Atmospheric and Climate Science,
ETH Zurich,
Universitätsstrasse 16, 8092, Zürich, Switzerland