Probabilistic landslide ensemble prediction systems: Lessons to be learned from hydrology

Ekrem Canli[1], Martin Mergili[1,2], Benni Thiebes[1], Thomas Glade[1]

[1] Department of Geography and Regional Research, University of Vienna, Universitätsstr. 7, 1010 Vienna, Austria

[2] Institute of Applied Geology, University of Natural Resources and Life Sciences, Gregor-Mendel-Straße 33, 1180 Vienna, Austria

Correspondence to: Ekrem Canli (ekrem.canli@univie.ac.at)

Abstract. Landslide forecasting and early warning has a long tradition in landslide research and is primarily carried out based on empirical and statistical approaches, e.g. landslide-triggering rainfall thresholds. In the last decade, flood forecasting started the operational mode of so called ensemble prediction systems followingthe success of the use of ensembles for weather forecasting. These probabilistic approaches acknowledge the presence of unavoidable variability and uncertainty when larger areas are considered and explicitly introduce them into the model results. Now that high detail numerical weather predictions and high-performance computing are getting more common, physically-based landslide forecasting for larger areas becomes feasible, and the landslide research community could benefit from the experiences that have been reported from flood forecasting using ensemble predictions. This paper reviews and summarizes concepts of ensemble prediction in hydrology and discusses how these could facilitate improved landslide forecasting. In addition, a prototype landslide forecasting system utilizing the physically-based TRIGRS model is presented to highlight how such forecasting systems could be implemented. The paper concludes with a discussion of challenges related to parameter variability and uncertainty, calibration and validation, and computational concerns.

Keywords: ensemble prediction systems, probabilistic forecasting, landslide early warning

1. Introduction

Landslide prediction at the regional scale is a hot topic within the scientific community as the time-varying aspects of landslide susceptibilities, hazards and even risks are crucial for emergency response planning and protecting public safety (Baum et al., 2010; Glade and Crozier, 2015). Further, the number of landslides is assumed to increase due to global change (Crozier, 2010; Gariano et al., 2017; Papathoma-Köhle and Glade, 2013). This calls for increased efforts for the development of early warning procedures with the aim of issuing timely warnings of an upcoming hazardous event to temporarily reduce the exposure of vulnerable persons or infrastructure (Thiebes and Glade, 2016). In this paper, we explicitly distinguish between prediction or forecasting systems and early warning systems although the terms are frequently used synonymously within the landslide community. We regard landslide forecasts as predictions of landslide occurrences for a specific time and space, e.g. the likely slope failure locations for an expected rainfall event. Landslide early warnings expand the predictions

of forecasting systems and must also cover dissemination and response strategies (UNEP, 2012). Warnings can be considered as calls for the public to take protective action, and the time scale of a warning depends on the associated weather event (Stensrud et al., 2009).

The forecasting of rainfall-triggered natural hazards with a rapid onset such as landslides and flash floods greatly benefit from numerical weather predictions (NWP), in particular from rainfall nowcasting and short-term rainfall forecasting that offer rainfall predictions several hours ahead. While it is evident that processes with potentially very short response time require more efforts for timely early warning than just real-time measurement of rainfall, real forecasting initiatives are scarce especially in the landslide community (Tiranti et al., 2017). At present, only two countries operate nation-wide landslide early warning systems: Italy (Rossi et al., 2012) and Norway (Devoli et al., 2015, 2018).

The reasons for the few large area applications of landslide early warning systems are manifold. One important reason might be that losses from landslides are perceived as mainly private and localized economic losses and thus, only few public resources have been allocated to develop sound spatial landslide early warning systems (Baum and Godt, 2010). Consequently, long monitoring time series, which are indispensable for sound and reliable early warning systems (such as available e.g. for floods, storms, etc.), are commonly not available. Another reason might be the complexity of single landslide detachments: the same landslide triggering event does not necessarily cause other landslides as the time between propagation stage and the collapse phase may vary significantly based on differences in local conditions (Greco and Pagano, 2017).

Present regional landslide forecasting and early warning systems are primarily based on empirical-statistical relationships between landslide occurrences and their associated rainfall events. Guzzetti et al. (2007) give an overview of rainfall and climate variables used in the literature for the definition of rainfall thresholds for the initiation of landslides, however, such empirical-statistical approaches disregard the physical relationship between rainfall occurrence and the mechanisms leading to landslides, neglecting local environmental conditions and the role of the hydrological processes occurring along slopes (Reichenbach et al, 1998; Bogaard and Greco, 2017). Attempts to relate landslide-triggering thresholds to weather and other physically-based characteristics can be very challenging given the quality of currently available data (Peres et al., 2017).

Physically-based based forecasting approaches put a strong focus on modeling the actual landslide initiation; however, such approaches are not commonly used for landslide forecasting and early warning systems. A reason for the negligence of such process-based might be related to concerns about data quality and parameter uncertainties when dealing with large areas. However, recently reported methodological advances with respect to the treatment of uncertainties for operational flash flood forecasting systems (Hapuarachchi et al., 2011; Liu et al., 2012; Yu et al., 2015) could also facilitate the implementation of physically-based landslide forecasting and early warning systems.

With this paper, we want to identify a gap in the prevalent landslide forecasting methods. Although ensemble predictions and the explicit integration of uncertainties in forecasts are now widely used in the fields of meteorology and hydrology, such activities are not yet common for landslide forecasting. We therefore think that the landslide community could benefit from the experiences of the neighboring disciplines and that our paper can provide a starting point for increased efforts into this

direction. The overall aim of this paper is to form a basis for discussion on how probabilistic landslide forecasting and early warning systems could be implemented. To this end, we provide a review on how probabilistic modeling methods and in particular ensemble predictions are applied for hydrological forecasts, and how these deal with uncertainties. Moreover, we highlight challenges and limitations for the calibration of models focusing on extreme events such as landslides. An important novelty of our paper consists in the presentation of a landslide ensemble forecasting framework utilizing the physically-based landslide model TRIGRS which we implemented within an open-source environment. In our case study, the geotechnical parameters are treated probabilistically to highlight how ensemble prediction for landslides could be implemented as operational systems. In addition, we present suggestions on how probabilistic landslide forecasts can be visualized in a way that stakeholders can base their decisions on. We conclude the paper a discussion of concerns about and challenges of probabilistic landslide forecasting approaches utilizing physically-based models that we hope will ultimately lead to increased efforts into this research direction.

1.1 Probabilistic forecasting in hydrology and ties to landslide research

When considering ensemble prediction systems (EPS), one should clarify what is expressed with the term *ensemble* and why EPS should be used at all since it is virtually unused in the landslide community. In the *Guidelines on Ensemble Prediction Systems and Forecasting* issued by the World Meteorological Organization (WMO, 2012), EPS are defined as "NWP systems that allow us to estimate the uncertainty in a weather forecast as well as the most likely outcome. Instead of running the NWP model once (i.e. a deterministic forecast), the model is run many times from very slightly different initial conditions. Often the model physics is also slightly perturbed, and some ensembles use more than one model within the ensemble (multi-model EPS) or the same model but with different combinations of physical parameterization schemes (multi-physics EPS). […] The range of different solutions in the forecast allows us to assess the uncertainty in the forecast, and how confident we should be in a deterministic forecast. […] The EPS is designed to sample the probability distribution function (pdf) of the forecast, and is often used to produce probability forecasts – to assess the probability that certain outcomes will occur" (WMO, 2012).

Krzysztofowicz (2001) argues that forecasts should be stated in probabilistic, rather than deterministic, terms and that this "has been argued from common sense and decision-theoretic perspectives for almost a century" (Krzysztofowicz, 2001). But still, by the new millennium, most operational hydrological forecasting systems relied on deterministic forecasts and there was a strong emphasis on finding the *best* estimates rather than quantifying the predictive uncertainty (Krzysztofowicz, 2001). However, those times have been overcome a decade later (Cloke and Pappenberger, 2009). From a scientific and historical perspective, landslide prediction has strong roots in empirical-statistical threshold based approaches (Wieczorek and Glade, 2005; Guzzetti et al., 2007). This stands valid until today, since most operational landslide forecasting and early warning systems rely on the relationship between rainfall and landslide occurrence, thus representing only a simplification of the underlying physical processes. Bogaard and Greco (2017) critically analyze the role of rainfall thresholds for shallow

3

landslides and debris flows from a hydro-meteorological point of view. It is beyond the scope of this paper to provide a review of past and present landslide early warning systems which can be found elsewhere (e.g. Baum and Godt, 2010; Alfieri et al. 2012a; Thiebes 2012; and Thiebes and Glade 2016). Guidelines on how landslide early warning systems can be implemented have been put forward by e.g. Intrieri et al. (2013).

One reason why landslide forecasting is seemingly more challenging than flood forecasting can be attributed to the spatial and temporal predictability of landslide processes. The spatial occurrence of floods is topographically foreseeable and controllable, but is more difficult to assess for landslides in distributed modeling due to their localized nature (Alfieri et al., 2012a; Canli et al., 2017). Despite considerable measurement uncertainties in phases of high flow, the prediction domain in flooding, which is usually streamflow, is more straightforward and can be measured more or less accurately over a long time.

In the past 15 years, a mindset of adapting probabilistic concepts to account for inherent uncertainties has taken over in the hydrologic community and the move towards EPS in flood forecasting represents the state of the art in forecasting science, following on the success of the use of ensembles for weather forecasting (Cloke and Pappenberger, 2009). Unfortunately, initiatives such as the Hydrological Ensemble Prediction Experiment (HEPEX) were not fostered in the landslide community to date. The general aims of this ongoing bottom-up initiative are to investigate how to produce, communicate and use hydrologic ensemble forecasts in a multidisciplinary approach (Schaake et al., 2007). One reason for the absence of such cooperative efforts might be the political, and therefore also financial, situation that led to the advancement of ensemble predictions in hydrology. Many international bodies demonstrated their interest in EPS which led to this superior position of hydrological prediction. This is even more so the case when taking into account transboundary floods that are typically more severe in their magnitude, affect larger areas and cause more damage and overall losses (Thielen et al., 2009). 1.2 Benefits and types of probabilistic approaches

Generally speaking, in an ensemble forecast small changes (perturbations) are made to the model parameters and then the model is re-run with these slightly perturbed starting conditions. If the different model realizations (*ensemble members*) are similar to each other, the forecasting confidence is rather high. Contrary, if they all develop differently, the confidence is much lower (WMO, 2012). By considering the proportion of the ensemble members that predict an event, e.g. a storm or a landslide, we can make an estimate of how likely the event is.

The term *ensemble prediction* for environmental applications was coined in the field of meteorology, thus describing the application of NWP systems, but it is used in different ways in neighboring disciplines. The atmospheric component is consistently described as weather ensemble input, yet the same applies to how observations of the land surface are incorporated into distributed forecasting models. In the data assimilation stage, ensembles of plausible land surface state observations (e.g. initial streamflow, soil moisture, snowpack) are created. Using multiple feasible parameter sets for each model or for each model run will realistically increase the spread of possible outcomes, yet it is more objective in terms of considered input parameters that were not directly observed (Schaake et al., 2007). Thus, the term ensemble prediction may be used in any instance of multi-parametric or multi-model data input that is used for forecasting the target variable.

In landslide research, a number of attempts have been reported that explicitly address ensemble techniques as a means of overcoming limitations from purely deterministic approaches or by increasing the predictive performance of statistically based landslide susceptibility mapping (e.g. Arnone et al., 2014; Cho, 2007; Haneberg, 2004; Melchiorre and Frattini, 2012; Rubio et al., 2004). None of them, however, incorporate ensemble techniques in real-time applications. Pradhan et al. (2017)

5   used an ensemble approach to evaluate the output of a physically-based model for a statistical machine learning model in varying hydrological conditions. Their ensemble model is based on a maximum entropy model that creates and combines multiple models to improve modeling results. However, their distributed output does not predict when or exactly where landsliding will occur, but yields a classified map with information where landslide occurrence can be expected over the long-term. Thus, their presented ensemble approach indicates landslide susceptibility that may be applicable for spatial

10   planning. While the term *ensemble* is by no means used a lot in landslide studies, it seems that it is predominantly used by the statistical landslide susceptibility modeling community (e.g. Lee and Oh, 2012; Althuwaynee et al., 2014a; 2014b). It is, however, not used in a way to address uncertainties in a forecasting model (Bartholmes and Todini, 2005; Vincendon et al., 2011). In a very promising approach, Chen et al. (2016) coupled a deterministic model with probabilistically treated geotechnical parameters with rainfall input from an operational multi-scale and multi-member NWP system (GRAPES) to

15   forecast spatial landslide occurrences with their ensemble prediction model (GRAPES-Landslide).

While there are not many landslide studies using or addressing ensemble techniques, there has been quite some work done on probabilistic landslide hazard analysis in the recent past. Lari et al. (2014) proposed a probabilistic approach expressing hazard as a function of landslide destructive power where landslide intensity (in terms of displacement rate) is considered rather than their magnitude. Haneberg (2004), Park et al. (2013), Raia et al. (2014), Lee and Park (2016) and Zhang et al.

20   (2016) treat soil properties at regional scale applications in a probabilistic way by randomly selecting variables from a given probability density function, mostly by means of Monte Carlo (MC) simulation. Salciarini et al. (2017) tried to enhance those approaches by considering geostatistical methods to provide the spatial distribution of soil properties and by using the Point Estimate Method (PEM) as a computationally more efficient method compared to MC simulation. But still, none of those probabilistic approaches explicitly aimed at spatial real-time early warning systems. The research of Schmidt et al.

25   (2008) represents a remarkable exception: they proposed a coupled regional forecasting system in New Zealand based on multiple process-based models (NWP, soil hydrology, slope stability). Unfortunately, a continuation of this research was not further pursued.

1.3 The hydrological equivalent of rainfall-induced shallow landslides: the case of flash floods

30   One major difference between flood and landslide early warning is the available lead time. While the lead time in larger river basins is sufficiently long to provide warnings for potentially hazardous situations from river flooding, shallow landslides, in the case of first time failures, generally occur suddenly and spatially unforeseeable in a specific area susceptible to

landsliding. As opposing to regular floods, however, flash floods can indeed be considered as an appropriate counterpart to rainfall-induced shallow landslide occurrence. Flash floods are, similar to shallow landslides, characterized by the superior importance of small-scale extreme precipitation events and their rapid onset, which leaves only little response time. It is therefore appropriate to examine how flash flood forecasting is performed and how it is applicable to landslide forecasting.

5   What makes landslide forecasting particularly challenging is the evolutionary sequence of the process. Greco and Pagano (2017) distinguish between three stages of a typical predictive system's architecture: I) the predisposing stage, II) the triggering and propagation stage, and III) the collapse stage. While in flood forecasting applications stage II and II are hardly distinguishable from each other, for rainfall-induced landslides this is not necessarily the case. While the predisposing stage (I) is determined by e.g. increasing pore water pressure due to a varying length of rainfall input that

10   worsens the slope stability conditions, the triggering and propagation stage (II) spans from first local slope failures until the formation of associated slip surfaces. The collapse phase (III) ultimately consists of the mobilization of the entire mass leading to the actual failure. However, the time between stages (II) and (III) may vary significantly based on differences in landslide types, local geomorphology, soil, vegetation, etc. and spans from a couple of minutes (e.g. flow slides in slopes covered with shallow coarse-grained soils) to years (e.g. earth flows in slopes of fine grained soils) (Greco and Pagano,

15   2017). Even when spatially distributed process-based landslide predictions are performed in relatively homogeneous regions, this time offset still prevails and makes landslide modeling in any context a challenging task. Therefore, warnings should generally be issued during indications of stage (II) since the lead time of stage (III) might be too short given the rapid kinematic characterization of the post-failure behavior, as recent disastrous examples in Italy have shown (Greco and Pagano, 2017).

20   Flood forecasting systems relying only on rainfall observations do not allow for a sufficiently long lead time for warnings. Extending this forecasting lead time further than the watershed response times requires the use of quantitative precipitation forecasts (QPF) from NWP (Vincendon et al., 2011). Additionally, models to represent hydrologic and hydraulic processes within a catchment to determine how rainfall-runoff accumulates are required (Hapuarachchi et al., 2011). With regard to producing quantitative precipitation estimates (QPE) in real-time, research has gone into blending multiple sources of

25   information (radar, satellite and gauged data) to increase the accuracy of QPEs. This process is generally referred to as *data assimilation* and is considered as increasingly important for improving hydrological predictions (Reichle, 2008). For predicting flash floods, however, lead times of several hours are desirable and thus high resolution QPFs with 1–6-hour lead times are generated. In recent years, the spatial (<5 km) and temporal (<1 h) resolutions of NWP model rainfall forecasts have significantly improved, while the combination of such NWP model forecasts with blends of the advected patterns of

30   recent radar, satellite and gauged rainfall data additionally increased the accuracy of nowcasting products (Hapuarachchi et al., 2011). Based on such high-resolution NWP model forecasts, probabilistic ensemble prediction systems have aided in exploring and quantifying uncertainties. Numerous studies have used those probabilistic precipitation forecasts to drive hydrological models (Vincendon et al., 2011; Bartholmes and Todini, 2005; Siccardi et al., 2005; Thielen et al., 2009). The application of such convective-permitting ensemble NWP is computationally very demanding and still in its infancy with

respect to flash flood prediction (Alfieri et al., 2012b). However, a further reduction of the spatial uncertainties of high-resolution rainfall fields is highly desirable, given the fact that rainfall is still considered as the most uncertain parameter in hydrological forecasting systems (Hapuarachchi et al., 2011; Alfieri et al., 2012b).

1.4 Many sizes fit all: the concept of equifinality

5   The concept of equifinality is deeply rooted in the hydrological community. It expresses an acceptance that many sets of parameters may provide equally acceptable forecasts (Beven, 1996; Beven and Freer, 2001; Collier, 2007). The concept of equifinality revolves around the rejection of the concept of *the* optimal model in favor of multiple possibilities for producing acceptable simulations (Beven and Freer, 2001). This concept is based on the understanding of physical theory and relates to the plethora of interactions among the components of a system whose resulting representations may be equally acceptable.

10   Research generally follows a working paradigm that should lead to realistic representations of the real processes and characteristics. This idea of identifying a single optimal representation of reality is very distinct in environmental sciences. A major problem arises from the scale discrepancy between sampling and distributed modeling where the use of global parameters undoubtedly leads to errors in predicting local responses at points with unique characteristics (Beven and Freer, 2001).

15   Geomorphological systems can be considered as transient, inheriting remnants of past and present processes. Environmental systems can exhibit certain degrees of chaotic behavior which results in an inability to express the trajectory of their development based on present-day evidence alone. Therefore, equifinality should not be considered as an indication of a poorly developed methodology, but as something inherent in geomorphological systems (Beven, 1996). However, it should most certainly *not* serve as a loophole for an inadequate methodology or model setup. A practical consequence of this

20   equifinality may lead to a more robust approach to testing the viability of different model setups with the aim to reject some, but to retain many of the offered solutions (Beven, 1996). Similarities and differences in model results should ultimately lead to an improved process understanding and, hence, predictive models with a higher sensitivity and specificity.

1.5 Calibration and validation

A model is an abstraction and a simplification of reality, hence the need for assessing its validity. Model validation provides

25   legitimacy in terms of arguments and methods (Oreskes et al., 1994). However, model validation is difficult when the most interesting events are rare, which is generally the case for flash floods or landslides. Also, calibration might be difficult for certain variables, or where suitable observations are not available. The WMO (2012) suggests that direct model output (DMO) from ensembles, although not ideal, still provides valuable information (WMO, 2012). The probabilistic forecasts with a DMO might not be as sharp (e.g. larger ensemble spread), but they still offer an estimate of the uncertainties and thus

30   pose an advantage over purely deterministic forecasts. But even where measurements of modeling parameters are available,

it has often shown that those parameters cannot be assumed constant in space or time, which makes calibration even more difficult. Additionally, the scale of measurement generally differs significantly from the scale at which the applied model requires "effective" parameter values to be specified (Beven, 1996).

Deterministic models for landslide prediction synthesize the interaction between hydrology, topography, vegetation and soil mechanics in order to physically understand and predict the location and timing of landslide triggering. These models usually contain a hydraulic and a slope stability component with different degrees of simplification (Formetta et al., 2016). In most cases, the target variable is the slope safety factor (FoS), which is useful as it enables decision makers to take actions when it falls short of a certain threshold (the slope is unstable with. FoS < 1.0, but higher thresholds are used in practice). In contrast to FoS, which at least in theory has a clear physical meaning, the definition of thresholds ets tricky when it comes to probabilities of events. When talking about the probability of events occurring, these events must be defined:

- The threshold value to be exceeded
- The exact time or time period to which the forecast refers
- The exact location or area to which the forecast applies
- Uncertainties and their role in the modeling process

With regard to those questions and as a starting point, the FoS is a suitable variable for probabilistic forecasting. Yet it has two major flaws: a) it represents the ratio of resisting forces to driving forces that is commonly not directly measured in the field and cannot be directly monitored, and b) landslide events are rare and their future location of occurrence remains unknown until they occur. This makes landslide calibration a challenging task. And there are limitations of model calibration in the case of rare events. Commonly, calibration will improve the reliability of forecasts (i.e. the match of the target variable or forecast probabilities to frequency of observations of the event) but reduce the resolution of the forecast (the ability to discriminate whether an event will occur or not). Consequently, calibration will improve forecasts of common events, but will also lead to the underprediction of more extreme events (WMO 2012). The WMO (2012) argues that this is the case for rare events, since the statistical distributions are trained to the more common occurrences. For rare events, hence, calibration cannot be expected to provide significant improvement over the raw forecasts. Considering the issue of equifinality. changing the calibration period or the goodness-of-fit measure results in an altered ranking of parameter sets to fit the observations. Consequently, there is no single parameter set (or model structure) that serves as the *characteristic* parameter input for any given area, but there is a certain degree of model equifinality involved when reproducing observations with model predictions (Beven, 1996). Therefore, given the issues with multiple (interacting) parameter values, measurement scales, spatial and temporal heterogeneity or the model structure, there can never be a single set of parameter values for the calibration process that represents an optimum for the study area, but calibration can contribute to the reduction of range in the possible parameter space.

Before being used for predictions, calibrated parameter sets have to be exposed to independent event data, in order to further evaluate the validity – or, more appropriately, the empirical adequacy – of the results obtained with the calibrated set of

parameters. In practice, few (if any) models are entirely confirmed by observational data, and few are completely refuted (Oreskes et al., 1994).

As a result, this is a field where probabilistic model output really shines, as it expresses the entire model spread with its inherent uncertainties not in absolute terms, but shows the relative performance of a model with respect to observational data. Many decision makers and practitioners in all kind of earth science related fields still favor absolute model output, especially in areas where public policy and public safety is at stake. Unfortunately, certainty is an illusion and ultimately the reason for modeling: the lack of full access, either in time or space, to the phenomena of interest (Oreskes et al., 1994).

2. Methods

With a case study application to a larger study area in Austria (approx. 1366 km²) we put forward a physically-based landslide forecasting framework using an ensemble prediction approach. As this case study is intended as a proof-of-concept and not as an operational forecasting system, we chose a simplified ensemble modeling approach in which only the geotechnical parameters are treated probabilistically; notably the precipitation input is considered a fixed input and consists of a historic rainfall event. However, from a technical point of view, real rainfall forecasts could easily be integrated. In addition, we present how probabilistic landslide forecasts could be visualized in order to support the decision-making process and early warning.

2.1 TRIGRS model

Physically-based models used to be attributed to local scale applications (e.g. Corominas et al., 2014; van Westen et al., 2008) because of their computational requirements and data constraints. This has clearly shifted in the last couple of years and by now, physically-based models can be quite commonly found to evaluate rainfall-induced landslide susceptibility at the regional scale (Ciurleo et al., 2017; Park et al., 2017; Thiebes et al., 2017). The majority is based on the infinite-slope model which only requires few input parameters to be suitable at a regional scale. Increasing the number of physical parameters in a model comes at the cost of introducing even more parameters, while the available data for calibration does not increase at the same time and could lead to problematic overparameterization (Beven, 1996). Even the simplest infinite-slope stability models generally require a higher level of parameterization than can be justified by available data. However, there are some general features of hillslope hydrology that are relevant to slope instability that can be considered to a certain degree by infinite-slope models: vertical infiltration, dependence of infiltration on initial soil moisture conditions, varying time scales for infiltration and lateral flow (Baum et al., 2010). As a result, TRIGRS (transient rainfall infiltration and grid-based regional slope-stability analysis; refer to Baum et al. (2008) for details), which we use in this case study, offers a good trade-off between model complexity and flexibility while we acknowledge the availability of other dynamic, physically-based models that were applied at a regional scale, such as STARWARS/PROBSTAB (Kuriakose et al., 2009) or

9

r.slope.stability (Mergili et al., 2014a). Raia et al. (2014) with their TRIGRS_P model and Salciarini et al. (2017) with their PG_TRIGRS model have already attempted a probabilistic TRIGRS derivative in the recent past that gave us the confidence to use TRIGRS in an automated probabilistic approach.

TRIGRS was specifically developed for modeling the potential occurrences of shallow landslides by incorporating transient pressure response to rainfall and downward infiltration processes (Baum et al., 2008). It uses the classical infinite slope stability approach, based on the Mohr-Coulomb model with effective cohesion c' (N/m²), effective angle of internal friction φ' (degree), and specific weight of the soil γ (N/m³) as geotechnical parameters. Initial soil conditions are assumed either saturated or tension-saturated. TRIGRS computes transient pore-pressure changes to find analytical solutions to partial differential equations, representing one-dimensional vertical flow in isotropic, homogeneous materials due to infiltration from rainfall events with durations ranging from hours to a few days. It uses a generalized version of Iverson's (2000) infiltration model solution to the boundary problem posed by Richard's equation. This solution assesses the effects of transient rainfall on the timing and location of landslides by modeling the pore water pressure of a steady component and a transient component (Liao et al., 2011). However, the model is limited by its distributed one-dimensional modeling approach with non-interacting grid cells and its simplified soil-water characteristic curve (Baum et al., 2010). However, recently Tran et al. (2018) presented a case study in which TRIGRS is combined with other models to be able to analyze slope stability in 3D. The entire theoretical basis together with all model related assumptions and equations can be found in Baum et al. (2008, 2010). TRIGRS computes a FoS for each grid cell based on an infinite-slope model. It allows for the implementation of spatially varying raster input (e.g. rainfall, property zones, soil depth, infiltration, etc.) to account for horizontal heterogeneity. The FoS can generally be referred to as the ratio of resisting forces (the resisting basal Coulomb friction) over driving forces (the gravitationally induced downslope basal driving stress) on the potential failure surface, with a FoS < 1.0 indicating slope instability and a FoS ≥ 1 slope stability respectively.

2.2 Model Setup

The probabilistic modeling setup is realized entirely in an open source framework. This was done not only to make it as easily reproducible as possible, but also because it offered the largest flexibility. TRIGRS, which itself is open-source, is operated by providing input text files. These input files are used to specify the numerical values of the input parameters, the location of the input raster files in the file system, and all other relevant grids to be considered (e.g. spatially distributed rainfall maps, different property zones to subdivide the study area in homogenous regions, spatially distributed soil depth maps, etc.). We used the Python programming language in a script for all string formatting procedures that receives its data from an initialization file. That python script is also used for parsing the raw input into variables usable for TRIGRS. User provided arguments in this initialization file hold the number of property zones needed, the rainfall duration pattern, number of time-steps and all variables that are used for the probabilistic treatment of parameters, such as min/max values for soil

depth, effective cohesion and effective friction angle as well as the number of model runs. The most recent rainfall input can be automatically imported by predefined naming conventions.

We used the GDAL package (GDAL Development Team, 2017) for reading and writing raster files and the NumPy package (van der Walt et al., 2011) for all raster calculations in the python script. Based on the number of predefined model runs, for each run a single deterministic output is generated based on the selected input parameters derived randomly from a normal distribution. We computed 25 model runs for each hour which resulted in 25 equally probable model results based on the different input parameters. After the first of the 25 model runs, a file is created and updated after each iteration. This probability of failure (PoF) grid file tracks the number of model runs that resulted in a FoS < 1.0 (unstable cell) for any given raster cell. Thus, the PoF represents the number of simulations that resulted in instability in relation to the total number of simulations.

All used variables, deterministic model outputs (i.e. the FoS maps) and the probabilistic model output (i.e. the PoF map) are parsed through to R (R Core Team, 2017). In R, all piped arguments from the python script are used for producing ready-to-use maps (packages: rgdal, Bivand et al., 2017; sp, Pebesma and Bivand, 2005) or to visualize performance measures such as ROC plots (package: ROCR, Sing et al., 2005). The entire procedure from importing raw data to producing usable maps is fully automated within an executable file that may be initiated every hour. This open code structure is flexible enough to enable the direct implementation of the most recent available data (rainfall data, soil moisture data, etc.) with minimal effort and thus makes it a useful tool in considering data assimilation techniques.

2.3 Data and model parameterization

The main data used in this case study consists of a digital terrain model (DTM), geotechnical parameter and soil depth ranges and rainfall data. In addition, landslide inventory data was used for a qualitative validation of the simulation results. Information on infrastructure from the OpenStreetMap (OSM) database was used for the calculation of building exposures. Table 1 lists the utilized data and their properties.

Table 1: Data used in the case study.

We used a bare earth LiDAR-based DTM, i.e. the vegetation and most infrastructure features such as houses are filtered. The raster cell size of the DTM to derive all model relevant topographical parameters used in this case study was 10 meters. This cell size allows for a sufficiently high representation of surface topography without losing too much information through surface aggregation and smoothing.

As rainfall input, a historic three hour long rainfall event was used. This specific rainfall event occurred in June 2009 and was a period of strong precipitation within an approximately 10 hour long rainfall that covered the study area in its entirety. Precipitation was the strongest in the central and western parts of the study and reached a total of up to 59 mm while hourly intensities reached maximum values of 25 mm. The historic rainfall event was used to facilitate our proof of concept; for an

operational landslide forecasting and early warning service it would be necessary to replace the file with actual rainfall forecasts by meteorological services. The spatially distributed rainfall raster maps representing hourly rainfall are based on an automated geostatistical interpolation (the methodology is described in detail in Canli et al., 2017). The selection of hourly rainfall input as well as the decision to choose a three-hour timeframe was made arbitrarily as for the study area there

5   are no published information available on the hydrological response of landslides to rainfall. The spatial resolution of 1 km for the rainfall input was resampled to match the cell size of the DTM, which is a prerequisite of TRIGRS.

Information on the spatial distribution of landslides was used for a qualitative validation of the modeling results for the following reasons: a) for Lower Austria a very comprehensive and spatially accurate landslide inventory based on high-resolution airborne LiDAR based DTM mapping exists (Petschko et al., 2015), however, it does not contain any temporal

10 information; b) the Building Ground Registry (BGR) is the most comprehensive source of reported damage causing landslides in Austria, however, its spatial and temporal accuracy is insufficient for physically-based model calibration and validation. As a result, it is not directly possible to quantitatively validate the model outputs.

For the visualization of our probabilistic results and for the calculation of building exposures, the OSM database was used. OSM covers almost the entirety of existing buildings in Austria and is based on official Austrian administrative data, which

15 stands under an open government data (OGD) license. Building exposure were computed by a simple spatial join that assigned each building the highest PoF value within 25 m. This value, while arbitrarily chosen, further accounts for spatial uncertainties since TRIGRS models only the location of landslide initiation.

Model parameterization over large areas is a difficult task given the poor spatial comprehension of the spatial organization of involved geotechnical and hydraulic input parameters. Tofani et al. (2017) performed 59 site investigations to parameterize

20 their distributed slope stability model; this amount of in situ soil samplings with associated lab measurements is exceptional and a solid foundation for determining the parameter probability density functions. Although Tofani et al. (2017) ultimately used the median value for each lithological class, the boxplots suggested normal to lognormal parameter distributions. This is a common observation and might be a result of the central limit theorem, which indicates that lumping data from many different sources (i.e. different in situ soil sampling sites in this case) tends to result in a normal or lognormal distribution

25 (Wang et al., 2015). Based on these findings, plausible geotechnical parameter ranges in this study were selected from published sources and a normal distribution was assumed. We used a simple MC simulation of multiple randomly chosen parameter sets within a predefined parameter range and within a single model structure as the basis for incorporating the inherent parameter uncertainties.

Parameters that were considered in a probabilistic way are soil depth, c'and φ' (Fig. 2); rainfall and elevation data (i.e. the

30 DTM) were utilized as fixed input. We assumed fully saturated conditions (volumetric water content θ = 40%) and slope-parallel groundwater flow for the sake of simplicity and given the absence of appropriate initial water conditions. Using all this information, it was possible to conduct a spatially distributed probabilistic assessment of the FoS, expressed as the probability of failure (PoF). As TRIGRS is capable of calculating the increase in pore water pressure within the soil, the

result is a distributed representation of the decrease in shear strength until slope failure (FoS < 1.0) is reached at a certain depth.

Figure 1: Probabilistically derived modeling parameters based on random sampling from a normally distributed **function**. Jittering dots (to
5   prevent overplotting) indicate individual samples within a plausible parameter range.

3. Study Area

The Rhenodanubian Flysch Zone (RDF) in the federal state of Lower Austria stretches over approx. 130 km in a SW–NE striking direction. The study area is limited to this geological zone in order to keep the subsurface as homogeneous as possible (Fig. 1). The Cretaceous–early Tertiary RDF is located in the northern foothills of the East Alps, in between the
10   Molasse basin to the North and the Northern Calcareous Alps to the South. The RDF is a paleogeographic-tectonic unit as part of the oceanic Penninic zone that was to a large part eliminated in the subduction process involved in the Alpine orogeny (Hesse, 2011). Flysch materials in the RDF are typically deeply weathered and mainly consist of alterations of pelitic layers (clayey shales, silty shales, marls) and sandstones. Physiographically, the RDF can be characterized as a low mountain region with a highly undulating terrain. It is exceptionally prone to landsliding, exhibiting around five landslides
15   per km^2 (Petschko et al., 2014). Heavy rainfall events (exceeding 100 mm per day) as well as rapid snowmelt are considered to be the main triggering factors for slope failures in the region (Schwenk, 1992, Schweigl and Hervás, 2009).

Figure 2: (A) Location of Lower Austria in Austria. (B) Location of the Rhenodanubian Flyschzone in Lower Austria (DTM: CC BY 3.0 AT–Federal state of Lower Austria); (C) Typical earth slide in Lower Austria after a heavy rainfall event in May 2014
20   **(Photograph by K. Gokesch).**

4. Results

Fig. 3 shows the results for a selection of four model iterations out of the total of 25 based on a spatially distributed hourly rainfall input over three hours. Each ensemble member was initialized with probabilistically derived geotechnical parameters that are displayed on each map. EPS representations that show a range of individual ensemble member are referred to as
25   *postage stamp maps* by the WMO (2012) and allow forecasters to view the scenarios in each member forecast. The results for our specific rainfall event indicate quite significant changes across individual members, but also partly high similarities although parameters change drastically between some of the members. For example, a depth of 2.5 m, an effective cohesion of 13.4 Nm^{-2} and an effective friction angle of 35 degree in one of the deterministic outputs reveal almost the identical FoS distribution compared to a depth of 2.0 m, an effective cohesion of 5.4 Nm^{-2} and an effective friction angle of 22.7 degree.

Figure 3: Selection of individual model iterations (postage stamp map). Each ensemble member was initialized with altered geotechnical parameters within a plausible range to account for variability and spatial uncertainty, and for a specific rainfall event. FoS values < 1 indicate slope instability.

5 By using a probabilistic representation, this variability and uncertainty of model outputs is accounted for. Here, the probability is estimated as a proportion of the ensemble members that predict an event to occur (FoS < 1.0) at a specific raster cell. For example, a probability between 0.75 and 1.0 means that a specific raster cell, under varying input parameter settings, was modeled as unstable in 75% to 100% of all model runs. To provide additional information and to support different actors responsible to manage landslide hazards, the PoF is underlain with building polygons and roads for a direct

10 exposure visualization of the elements at risk towards landslides (Fig. 4). High building exposure along the river is a modeling artifact introduced by the steep retaining wall and the associated sudden and steep decline in slope angle. For operational landslide forecasting, these areas would need to be excluded to avoid false alarms. The results of the PoF map suggest quite a narrow ensemble spread, which could be an expression of equifinality. The narrow ensemble spread can be considered as some kind of spatial confidence buffer that gives some reliance that under varying input data the location of

15 possible slope failure is modeled quite consistently at the more or less same location.

Figure 4: Probability of Failure depicted as a proportion of the ensemble members that predict an event to occur (FoS < 1.0). Building exposure to current slope failure predictions adds an additional information layer for decision makers. Buildings and roads are imported from the OSM database (© OpenStreetMap contributors).

20 In this case study, only a qualitative validation could be performed due to landslide inventory data insufficiencies. Qualitative validation by visual comparison (Fig. 5) indicates, for this specific time and under the given rainfall input, that there is an agreement between some of the landslide initiation points and areas of high failure probability. It must, however, be noted that the triggering conditions of the mapped landslides are unknown and therefore no complete agreement with the model results for our specific rainfall event can be expected.

25

Figure 5: Probability of failure map detail for a specific rainfall event. Known historic landslide initiation points (ellipses) partly overlap with current slope stability conditions. However, high spatial resolution, and therefore a high degree of spatial discontinuity, poses a risk for missing many real landslide events in an early warning situation.

30 **5. Discussion**

The presented probabilistic framework for landslide forecasting does not eliminate uncertainty, but it explicitly introduces it into the model results. Introducing uncertainties may seem detrimental to the ultimate goal of predictive modeling: to be a positionally and temporally accurate mitigation tool. However, we argue that a probabilistic approach is a viable alternative to deterministic landslide forecasting models as they allow forecasters to make informed decisions based on the uncertainties

and limitations involved with model predictions. In contrast, a purely deterministic model output, in particular highly detailed map representations of slope failure probability, could suggest a certainty that simply is not achievable in landslide modeling. Although a probabilistic approach depicts spatial parameter variability and uncertainty much better than any purely deterministic result, there are still unaccounted uncertainties involved with respect to actual slope failure prediction,

5   e.g. related to limitations of the modeling approach, quality and uncertainties of data treated in a non-probabilistic way, or the limited number of model iterations that might not be sufficient to fully explore the parameter space.

The narrow ensemble spread highlighted in our PoF maps (i.e. Fig. 3 and 5) underlines the important impact of slope angle, and thus DTM quality, on model outputs. Neves Seefelder et al. (2016) and Zieher et al. (2017) identified slope angle as one of the most sensitive modeling parameters in TRIGRS, which is not surprising since slope failures are in general associated

10   with higher slope angles (Liao et al., 2011). Even under greatly varying geotechnical or hydraulic input parameter settings, the same slope segments experience the highest likelihood of slope failure. As a consequence, slope failure probability will ultimately vary spatially based on slope, and temporally based on rainfall input, unless the study area includes different geological units with greatly varying geotechnical parameters appropriately reflected in the available data.

The great variability between the ensemble members (Fig. 4) raises the question whether extensive model calibration for

15   deriving a single model output is the ideal solution. Purely deterministic forecasts suppress information and limit the possibility to judge about uncertainty. They generally pretend to be based on an optimal set of input parameter settings. Only recently, empirical approaches, such as the commonly used rainfall thresholds in landslide forecasting and early warning applications, have started to incorporate estimates of uncertainty by defining rainfall thresholds at different exceedance probabilities (e.g. Melillo et al., 2016, Piciullo et al., 2017). Yet, they rely on very good landslide event catalogues and

20   knowledge of the triggering conditions. However, reported landslides and their triggering conditions are often affected by large errors (Peres et al., 2017). Gariano et al. (2015) found that an underestimation of only 1% in the number of considered landslides can result in a significant decrease in the performance of a threshold-based landslide prediction system. Additionally, rainfall thresholds represent a simplification of the underlying physical processes by establishing no more than a statistical relationship between rainfall and landslide occurrence (Bogaard and Greco, 2017). Both deterministic and

25   empirical approaches may create the illusion of certainty in a user's mind, which could easily lead to wrong conclusions and poor decision-making.

We argue that hiding the predictive uncertainty behind the façade of a precise estimate is wrong and careless. Concerns about the acceptance of probabilities in decision making turned out to be unwarranted (Krzysztofowicz, 2001). Based on our observations we found that many published landslide studies dealing with physically-based hindcasting applications rely too

30   strongly on purely number based validation outputs. Deterministic results are taken as given when the modelers achieve *satisfactory* results based on the model validation, without defining what the criteria are for this satisfaction or when this state of satisfaction is reached. Beven (1996) argues that this is generally owed to relativism, when there is a need to adopt less stringent criteria of acceptability or to acknowledge that it is not possible to predict all the observations all the time (with common arguments ranging from scale issues, spatial heterogeneity, uncertainty in model structure or process understanding,

etc.). In general, probabilistic approaches should be prioritized since they allow not only for the incorporation of parametric uncertainties, but also facilitate the geomorphic plausibility control in the absence of proper data for model calibration and validation. Also for ensemble predictions narrowing down uncertainties is a good first step, but the ultimate goal should be to determine and explore the differences between model predictions (Challinor et al., 2014). In any case, high quality data are necessary but are often scarce in landslide research. The potential of local-scale studies to draw conclusions for a larger area applications (e.g. Bordoni et al., 2015) remains to be a very important field of study in the near future. In this regard, data assimilation might be a key factor for producing accurate model predictions while reducing those inherent uncertainties. Liu et al. (2012) give an in-depth review on the current state of data assimilation applications in both hydrologic research and operational practices that are in many parts valid for landslide prediction too. While there are a few adaptive systems in landslide early warning based on empirical thresholds, such as the SIGMA early warning system in Italy (Martelloni et al., 2012; Segoni et al., 2017), there are not yet any that use physically-based predictions with blends of most recent QPEs or other independent observations. For extreme events, this might be key if the probability of extreme floods or landslides occurring is continuously and objectively evaluated and updated in real-time, especially when it comes to assimilating new observations from multiple sources across a range of spatiotemporal scales (Liu et al., 2012).

While we explicitly acknowledge the achievements and the success of deterministic or empiric landslide now- and forecasting approaches, we argue that the implementation of EPS approaches utilizing physically-based landslide prediction models could facilitate future landslide early warning systems. In particular highly detailed convection permitting NWP would be ideal to run a physically-based landslide forecasting service. Such high resolution (<3km) NWP services are presently available for Germany (COSMO-DE; Baldauf et al., 2011, Gebhardt et al., 2011), France (AROME; Seity et al., 2011), the UK (MOGREPS-UK; Golding et al., 2016) and the US (HRRR; Ikeda et al., 2013). However, there might be concerns whether the operation of physically-based EPS modeling approaches is feasible for large areas, in particular related to concerns regarding parameter uncertainties and variability, model calibration, and the computational burden. In the following, we want to provide a discussion of these issues.

An important limiting factor for the application of physically-based landslide models to large areas is related to **parameter variability and uncertainty, and issues with the parameterization of the respective models.** Current practices for geotechnical parameterization in physically-based landslide modeling include the application of averaged values from in situ measurements (e.g. Thiebes et al., 2014, Tofani et al., 2017, Zieher et al., 2017) or using values from existing databases, lookup tables or other published/unpublished sources (e.g. Schmidt et al., 2008, Kuriakose et al., 2009, Mergili et al., 2014b). In the landslide research community, probabilistic treatment of input parameters for regional model application has seen a rise only in the last couple of years (e.g. Mergili et al., 2014a; Raia et al., 2013; Neves Seefelder et al., 2017). Probabilistic approaches allow for a more thorough consideration of uncertainties and inherent variability of model specific parameters. Spatially varying parameters (both geotechnical and hydraulic) are usually represented as univariate distributions

of random variables based on an underlying probability density function and statistical characteristics (Fan et al., 2016). Friction angle and cohesion are commonly considered as such varying variables that are treated in a probabilistic way for model parameterization (e.g. Park et al., 2013, Chen and Zhang, 2014, Raia et al., 2014, Salciarini et al., 2017). Interestingly, in hydrological streamflow prediction the parameter uncertainty of the hydraulic model is often neglected in favor of a deterministic parameter input. This is explained by the superior proportion of total estimation uncertainty introduced by the weather predictions alone, which blurs the streamflow variability that the meteorological input data cannot explain (Alfieri et al., 2012b). Measuring geotechnical and hydrological parameters for large areas is difficult, time-consuming, and expensive. Therefore, applying spatially distributed physically-based models with spatially variable geotechnical parameters is not straightforward and it is impossible to find an approach that is universally accepted (Tofani et al., 2017). Even if there is a sufficiently large amount of measured values available for one, some or even all parameter values in a model up to the point that it is possible to specify distributions and covariances for the parameter values, there remain some methodological obstacles. For example, there is no guarantee that values measured at one scale will reflect the effective values required in the model to achieve satisfactory predictions of observed variables (Beven and Freer, 2001). For larger areas (e.g. scales >1:25,000), there are several factors that cause spatial variation of, for example, soil water content, topography, differences in soil depth, -type and -texture, vegetation characteristics, as well as rainfall patterns. Additionally, spatially varying soil and hydraulic properties are influenced by interrelated soil formation processes (such as weathering processes, biological perturbations, atmospheric interactions) (Fan et al., 2016), and thus making selective in situ soil sampling a tricky task when performed at a larger scale. Small area (e.g. scales <1:10,000) variability usually lacks a spatial organization, hence its representation as stochastic process. The larger the area, however, the more soil forming processes manifest a persistent deterministic signature due to the predetermined geology, topography, climate, etc. (Seyfried and Wilcox, 1995, Fan et al., 2016). Neves Seefelder et al. (2016) suggest applying rather broad ranges of parameters for physically-based approaches to be on the "safe side" as they yield results comparable in quality to those derived with best-fit narrow ranges. By acknowledging the fact that geotechnical and hydrological parameters – when working on larger areas – are highly variable, uncertain and often poorly known, narrow parameter ranges or even singular combinations of parameters come with the risk of being off target (Neves Seefelder et al., 2016). This basically implies that, when working on a regional scale and beyond, an actual parameterization with in situ measured samples might not be purposeful. The use of literature values could mean enormous savings in time and money spent, yet this clearly needs further research to evaluate whether there is and to what degree the benefits of actual sampled in situ data are compared to just utilizing literature values in broad ranges when modeling for larger areas.

Like flood events, landslide types with a rapid onset can be considered as an extreme events. Hereby, extreme does not necessarily refer to the size of the displaced landslide volumes – also small landslides might be considered as extreme in terms of potential consequences. While it is possible to continuously monitor and forecast regular streamflow, extreme events are scarce which makes model calibration and, consequently, forecasting a real challenge. We argue that this is even

more so the case for landslides as there are no directly observable target variables to be monitored at a regional scale. Landslide models can only be calibrated on a case by case basis. Shallow landslides are one of the most common landslide types (van Asch et al., 1999). While they occur quite in abundance when looking at their spatial distribution, they are typically low-frequency events. And most of them do occur in so called 'low-risk' environments as defined by Klimeš et al.

5 (2017): low annual frequency of landslides; the majority of the landslides are of small size and are low impact events. Due to the scarcity of such extreme events, Collier (2007) argues that such events may lie outside of what model calibration is capable of providing for forecasting approaches. Therefore, it is very difficult to calibrate a model for future use, as it can be only continually evaluated in the light of the most recent data (Oreskes et al., 1994, Challinor et al., 2014). It is beyond the scope of the present paper to propose new approaches for the calibration and validation of regional landslide forecasting

10 models. However, we think that, at least to some extent, the experiences of continuous local scale monitoring system might be able to facilitate the calibration process for regional scale applications; however, additional research is required.

Landslides are per se extreme events with no common events attributed to them as they only occur under exceptional circumstances given the environmental interactions involved. This raises the question if any averaged model output, and that is by definition every model output based on model calibration from past events, will ever be able to precisely forecast

15 extreme events at the regional scale. The sensitivity of the model must be lowered in a way that much larger areas of slope failure need to be forecasted to catch a few real extreme events at the cost of significantly raising the number of false alerts. In addition, it has been reported that it is not always possible to develop sensitive models due to data constrains (e.g. Gioia et al., 2015; Alvioli and Baum, 2016). Creating sensitive models that are able to forecast all events is especially difficult when engineering conservatism comes into play in decision making, thus leaving probabilistic forecasting attempts in an inferior

20 state over purely deterministic approaches. This is a known issue (e.g. Baum et al., 2010) in a way that FoS computations usually are more likely to identify areas prone to slope failure during a given rainfall event rather than predicting exact locations of specific landslides. A term such as landslide susceptibility forecasting seems more appropriate in that case. Our results in the Flysch Zone of Lower Austria seem to point in that direction so far. This is definitely an issue that needs far more in-depth research in the future. In addition, the technical specifications of the modeling approach for slope stability

25 analysis at a regional scale have to be kept in mind. The most commonly applied modeling approach relies on the infinite-slope stability model which reduces the landslide geometry to a slope-parallel layer of infinite length and width. This leads to very pronounced patterns of the factor of safety, whereas modeling approaches that introduce more complex landslide geometries produce smoother results since the effects of neighboring pixels are averaged out. Whether complex approaches such as r.slope.stability (Mergili et al., 2014a), Scoops3d (Reid et al., 2015) or approaches based on slip circles or ellipsoids

30 (Xie et al., 2003; 2004; 2006) are able to outperform the infinite slope stability model depends on the settings, notably the landslide geometries. In theory, the infinite slope stability model is suitable for shallow landslides with length-to-depth ratios above 18-25 (Griffiths et al., 2011; Milledge et al., 2012). Consequently, parameters representing the landslide geometry assumed by the model (i.e. slope angle and depth) are highly sensitive (Zieher et al., 2017). This means that the underlying

model itself already performs some sort of averaging too since the precise landslide geometry cannot be adequately resolved in the infinite-slope stability model.

Traditionally, physically-based approaches for modeling rainfall-induced shallow landslides were suggested to be applied to small study areas while statistical based approaches were recommended for regional susceptibility assessments (e.g. van Westen et al., 2006; Corominas et al., 2014). One reason usually mentioned is the poor comprehension of the spatial organization of the geotechnical and hydraulic input parameters (e.g. Tofani et al., 2017, Park et al., 2013). However, as outlined above, it does not make too much difference whether the underlying study area is 50 km² or 5000 km² investigated at a scale of 1:1,000 or 1:25,000 — the model is still influenced by errors or uncertainties from the input parameters given the fact how input parameters are derived, and one should not believe that uncertainties disappear (or even decrease) with more data and more measurements.

Another major limitation used to be the computational costs involved when carrying out physically-based modeling at a regional scale. However, in recent years computational power has become available at reasonable costs, and the area size, associated with a high-resolution DTM, steadily increased over time for physically-based applications and today exceeds thousands of square kilometers (e.g. Tofani et al., 2017, Alvioli and Baum, 2016). Recent landslide model development is aiming towards featuring multithreading and parallelization. Since high resolution DTM are available in many parts of the world, the computational demands increased significantly, especially when applied in a dynamic and time-dependent modeling framework. Parallelization has great potential in grid-based landslide modeling, especially for the time-consuming hydraulic model components, for several reasons: in case of TRIGRS, for example, which is a coupled slope stability and hydraulic model, only excess water from infiltration is directed to the neighboring cells which makes it the only variable that relies on explicit neighborhood relations. This needs to be done only once, however. Vertical groundwater flow and one-dimensional slope stability in a two-dimensional array of non-interacting columns can subsequently be computed independently for each cell, which is a prime example for parallelization purposes (Baum et al., 2010, Alvioli and Baum, 2016). Besides TRIGRS v2.1, which received its parallel implementation by Alvioli and Baum (2016) only recently, other models for physically-based landslide applications are using a parallelized module: NewAge-JGrass (Formetta et al., 2016) or r.slope.stability (Mergili et al., 2014a). In our case study, the computational time for one model iteration is about 45 minutes, which is far too long for computing a large set of different ensemble members in an operational real-time application. We did not yet use the parallel implementation of TRIGRS on our regular commodity machine (3.40GHz quad-processor equipped with 32GB RAM), but Alvioli and Baum (2016) reported that parallel computation on a multi-core node already led to a significant speedup compared to a single-node local machine. When applied to a HPC (High-Performance Computing) cluster, they were even able to reduce running time from one day to one hour. This allows the exploration of new possibilities in how landslide forecasting can be approached in the future. While HPC applications are common in meteorological (Bauer et al., 2015) and flood forecasting (Shi et al., 2015), there are only few landslide related studies (Mulligan and Take, 2017; Shute et al., 2017; Song et al., 2017); however, none aiming specifically at landslide forecasting.

19

This opens up possibilities to encompass fine tuning of input parameters by means of multiple model runs, probabilistic applications and, first and foremost, real-time applications with a continuous consideration of antecedent and forecasted rainfall information (Alvioli and Baum, 2016).

6. Conclusions

5   The adoption of EPS approaches in meteorological and hydrological modeling has led to the implementation of operational forecasting system over the last years. Therein, parameter uncertainties are explicitly integrated in the forecasts and the results are made available to decision makers. However, similar approaches have not yet been reported from the landslide research community. By identifying this gap, i.e. the lack of landslide EPS forecasting systems, this paper aims to foster the discussion on this topic which hopefully facilitates additional research activities into this direction.

10   Our case study implementation of a landslide forecasting system, although simplified due to the limited number of parameters treated in a probabilistic way, highlights how a physically-based model could be operated for larger areas, and how uncertainties can be dealt with. Still, there are concerns regarding such large area model applications, in particular related to parameter variability and uncertainty, model calibration and validation, and the feasibility with respect to the computational burden. We argue that these concerns are to some extent unwarranted and can be overcome with the

15   probabilistic approaches and additional model development. However, the calibration of forecasting models focusing on extreme events such as landslides remains a major challenge and needs further research and high quality data.

20   **Compliance with ethical standards**

The authors declare that they have no conflict of interest.

References

Alfieri, L., Salamon, P., Pappenberger, F., Wetterhall, F. and Thielen, J.: Operational early warning systems for water-related hazards in Europe, Environmental Science & Policy, 21, 35–49, doi:10.1016/j.envsci.2012.01.008, 2012a.

25   Alfieri, L., Thielen, J. and Pappenberger, F.: Ensemble hydro-meteorological simulation for flash flood early detection in southern Switzerland, Journal of Hydrology, 424–425, 143–153, doi:10.1016/j.jhydrol.2011.12.038, 2012b.

Althuwaynee, O. F., Pradhan, B., Park, H.-J. and Lee, J. H.: A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping, CATENA, 114, 21–36, doi:10.1016/j.catena.2013.10.011, 2014a.

Althuwaynee, O. F., Pradhan, B., Park, H.-J. and Lee, J. H.: A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping, Landslides, 11(6), 1063–1078, doi:10.1007/s10346-014-0466-0, 2014b.

Alvioli, M. and Baum, R. L.: Parallelization of the TRIGRS model for rainfall-induced landslides using the message passing interface, Environmental Modelling & Software, 81, 122–135, doi:10.1016/j.envsoft.2016.04.002, 2016.

Arnone, E., Dialynas, Y. G., Noto, L. V. and Bras, R. L.: Parameter Uncertainty in Shallow Rainfall-triggered Landslide Modeling at Basin Scale: A Probabilistic Approach, Procedia Earth Planet. Sci., 9, 101–111, doi:10.1016/j.proeps.2014.06.003, 2014.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. and Reinhardt, T.: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities, Monthly Weather Review, 139(12), 3887–3905, doi:10.1175/MWR-D-10-05013.1, 2011.

Bartholmes, J. and Todini, E.: Coupling meteorological and hydrological models for flood forecasting, Hydrology and Earth System Sciences, 9(4), 333–346, doi:10.5194/hess-9-333-2005, 2005.

Bauer, P., Thorpe, A. and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525(7567), 47–55, doi:10.1038/nature14956, 2015.

Baum, R. L., Savage, W. Z. and Godt, J. W.: TRIGRS - A Fortran Program for Transient Rainfall Infiltration and Grid-Based Regional Slope-Stability Analysis, Version 2.0, 2008.

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

Baum, R. L., Godt, J. W. and Savage, W. Z.: Estimating the timing and location of shallow rainfall-induced landslides using a model for transient, unsaturated infiltration, Journal of Geophysical Research, 115(F3), doi:10.1029/2009JF001321, 2010.

Beven, K.: Equifinality and Uncertainty in Geomorphological Modelling, in: The Scientific Nature of Geomorphology: Proceedings of the 27th Binghamton Symposium in Geomorphology, held 27-29 September, 1996, John Wiley & Sons, 27, 289–313, 1996.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of hydrology, 249(1), 11–29, doi:10.1016/S0022-1694(01)00421-8, 2001.

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R. and Rouault, E.: rgdal: Geospatial Data Abstraction Library. Available at: https://cran.r-project.org/package=rgdal, last access: 30 November 2017.

Bogaard, T. and Greco, R.: Invited perspectives. A hydrological look to precipitation intensity duration thresholds for landslide initiation: proposing hydro-meteorological thresholds, Natural Hazards and Earth System Sciences Discussions, 1–17, doi:10.5194/nhess-2017-241, 2017.

Bordoni, M., Meisina, C., Valentino, R., Bittelli, M. and Chersich, S.: Site-specific to local-scale shallow landslides triggering zones assessment using TRIGRS, Natural Hazards and Earth System Science, 15(5), 1025–1050, doi:10.5194/nhess-15-1025-2015, 2015.

Canli, E., Loigge, B. and Glade, T.: Spatially distributed rainfall information and its potential for regional landslide early warning systems, Natural Hazards, doi:10.1007/s11069-017-2953-9, 2017.

Challinor, A., Martre, P., Asseng, S., Thornton, P. and Ewert, F.: Making the most of climate impacts ensembles, Nature Climate Change, 4(2), 77–80, doi:10.1038/nclimate2117, 2014.

Chen, H. X. and Zhang, L. M.: A physically-based distributed cell model for predicting regional rainfall-induced shallow slope failures, Engineering Geology, 176, 79–92, doi:10.1016/j.enggeo.2014.04.011, 2014.

Chen, Y., Chen, D., Li, Z., Wu, Y. and Huang, J.: An ensemble prediction model for rainfall-induced landslides and its preliminary application, Chinese Journal of Atmospheric Sciences, 40(3), 515–527, doi:10.3878/j.issn.1006-9895.1503.15120, 2016.

Cho, S. E.: Effects of spatial variability of soil properties on slope stability, Eng. Geol., 92(3–4), 97–109, 2007.

Ciurleo, M., Cascini, L. and Calvello, M.: A comparison of statistical and deterministic methods for shallow landslide susceptibility zoning in clayey soils, Eng. Geol., 223, 71–81, doi:10.1016/j.enggeo.2017.04.023, 2017.

Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, Journal of Hydrology, 375(3–4), 613–626, doi:10.1016/j.jhydrol.2009.06.005, 2009.

Collier, C. G.: Flash flood forecasting: What are the limits of predictability?, Quarterly Journal of the Royal Meteorological Society, 133(622), 3–23, doi:10.1002/qj.29, 2007.

Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J.-P., Fotopoulou, S., Catani, F., van ven Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J. and Smith, J. T.: Recommendations for the quantitative analysis of landslide risk, Bulletin of Engineering Geology and the Environment, 73(2), 209–263, doi:10.1007/s10064-013-0538-8, 2014.

Crozier, M. J.: Deciphering the effect of climate change on landslide activity: A review, Geomorphology, 124(3–4), 260–267, doi:10.1016/j.geomorph.2010.04.009, 2010.

Devoli, G., Kleivane, I., Sund, M., Orthe, N., Ekker, R. and Johnsen, E.: 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 - Volume 2, pp. 625–629, Springer, Cham., 2015.

Devoli, G., Tiranti, D., Cremonini, R., Sund, M. and Boje, S.: Comparison of landslide forecasting services in Piedmont (Italy) and Norway, illustrated by events in late spring 2013, Nat. Hazards Earth Syst. Sci., 18(5), 1351–1372, doi:10.5194/nhess-18-1351-2018, 2018.

Fan, L., Lehmann, P. and Or, D.: Effects of soil spatial variability at the hillslope and catchment scales on characteristics of rainfall-induced landslides, Water Resources Research, 52(3), 1781–1799, doi:10.1002/2015WR017758, 2016.

Formetta, G., Capparelli, G. and Versace, P.: Evaluating performance of simplified physically based models for shallow landslide susceptibility, Hydrology and Earth System Sciences, 20(11), 4585–4603, doi:10.5194/hess-20-4585-2016, 2016.

Gariano, S. L., Brunetti, M. T., Iovine, G., Melillo, M., Peruccacci, S., Terranova, O., Vennari, C. and Guzzetti, F.: Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy, Geomorphology, 228, 653–665, doi:10.1016/j.geomorph.2014.10.019, 2015.

Gariano, S. L., Rianna, G., Petrucci, O. and Guzzetti, F.: Assessing future changes in the occurrence of rainfall-induced landslides at a regional scale, Science of The Total Environment, 596–597, 417–426, doi:10.1016/j.scitotenv.2017.03.103, 2017.

GDAL Development Team: GDAL - Geospatial Data Abstraction Library, Version 2.1.3, Open Source Geospatial Foundation. Available at: http://www.gdal.org, last access: 30 November 2017.

Gebhardt, C., Theis, S. E., Paulat, M. and Ben Bouallègue, Z.: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries, Atmospheric Research, 100(2–3), 168–177, doi:10.1016/j.atmosres.2010.12.008, 2011.

Glade, T. and Crozier, M.J.: A review of scale dependency in landslide hazard and risk analysis, edited by T. Glade, M. Anderson and M. Crozier: Landslide hazard and risk, Wiley, Chichester 75–138, 2015.

Golding, B., Roberts, N., Leoncini, G., Mylne, K. and Swinbank, R.: MOGREPS-UK Convection-Permitting Ensemble Products for Surface Water Flood Forecasting: Rationale and First Results, Journal of Hydrometeorology, 17(5), 1383–1406, doi:10.1175/JHM-D-15-0083.1, 2016.

Griffiths, D. V., Huang, J. and deWolfe, G. F.: Numerical and analytical observations on long and infinite slopes, Int. J. Numer. Anal. Methods Geomech., 35(5), 569–585, doi:10.1002/nag.909, 2011.

Greco, R. and Pagano, L.: Basic features of the predictive tools of early warning systems for water-related natural hazards: examples for shallow landslides, Natural Hazards and Earth System Sciences Discussions, 1–31, doi:10.5194/nhess-2017-269, 2017.

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

Haneberg, W. C.: A rational probabilistic method for spatially distributed landslide hazard assessment, Environmental & Engineering Geoscience, 10(1), 27–43, doi:10.2113/10.1.27, 2004.

Hapuarachchi, H. A. P., Wang, Q. J. and Pagano, T. C.: A review of advances in flash flood forecasting, Hydrological Processes, 25(18), 2771–2784, doi:10.1002/hyp.8040, 2011.

Hesse, R.: Rhenodanubian Flyschzone, Bavarian Alps, in Geological Field Trips in Central Western Europe: Fragile Earth International Conference, Munich, September 2011, edited by S. Carena, A. M. Friedrich, and B. Lammerer, Geological Society of America, Boulder, USA, 51–73, 2011.

Ikeda, K., Steiner, M., Pinto, J. and Alexander, C.: Evaluation of Cold-Season Precipitation Forecasts Generated by the Hourly Updating High-Resolution Rapid Refresh Model, Weather and Forecasting, 28(4), 921–939, doi:10.1175/WAF-D-12-00085.1, 2013.

Intrieri, E., Gigli, G., Casagli, N. and Nadim, F.: Brief communication" Landslide Early Warning System: toolbox and general concepts", Nat Hazards Earth Syst Sci, 13, 85–90, 2013.

Iverson, R. M.: Landslide triggering by rain infiltration, Water Resources Research, 36(7), 1897–1910, doi:10.1029/2000WR900090, 2000.

Klimeš, J., Stemberk, J., Blahut, J., Krejčí, V., Krejčí, O., Hartvich, F. and Kycl, P.: Challenges for landslide hazard and risk management in "low-risk" regions, Czech Republic–landslide occurrences and related costs (IPL project no. 197), Landslides, 14(2), 771–780, doi:10.1007/s10346-017-0798-7, 2017.

Krzysztofowicz, R.: The case for probabilistic forecasting in hydrology, Journal of Hydrology, 249(1–4), 2–9, doi:10.1016/S0022-1694(01)00420-6, 2001.

Kuriakose, S. L., van Beek, L. P. H. and van Westen, C. J.: Parameterizing a physically based shallow landslide model in a data poor region, Earth Surface Processes and Landforms, 34(6), 867–881, doi:10.1002/esp.1794, 2009.

Lari, S., Frattini, P. and Crosta, G. B.: A probabilistic approach for landslide hazard analysis, Engineering Geology, 182, 3–14, doi:10.1016/j.enggeo.2014.07.015, 2014.

Lee, S. and Oh, H.-J.: Ensemble-Based Landslide Susceptibility Maps in Jinbu Area, Korea, in Terrigenous Mass Movements, edited by B. Pradhan and M. Buchroithner, Springer Berlin Heidelberg, 193–220, 2012.

Lee, J. H. and Park, H. J.: Assessment of shallow landslide susceptibility using the transient infiltration flow model and GIS-based probabilistic approach, Landslides, 13(5), 885–903, doi:10.1007/s10346-015-0646-6, 2016.

Liao, Z., Hong, Y., Kirschbaum, D., Adler, R. F., Gourley, J. J. and Wooten, R.: Evaluation of TRIGRS (transient rainfall infiltration and grid-based regional slope-stability analysis)'s predictive skill for hurricane-triggered landslides: a case study in Macon County, North Carolina, Natural Hazards, 58(1), 325–339, doi:10.1007/s11069-010-9670-y, 2011.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O. and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrology and Earth System Sciences, 16(10), 3863–3887, doi:10.5194/hess-16-3863-2012, 2012.

Martelloni, G., Segoni, S., Fanti, R. and Catani, F.: Rainfall thresholds for the forecasting of landslide occurrence at regional scale, Landslides, 9(4), 485–495, doi:10.1007/s10346-011-0308-2, 2012.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L. and Guzzetti, F.: Rainfall thresholds for the possible landslide occurrence in Sicily (Southern Italy) based on the automatic reconstruction of rainfall events, Landslides, 13(1), 165–172, doi:10.1007/s10346-015-0630-1, 2016.

Melchiorre, C. and Frattini, P.: Modelling probability of rainfall-induced shallow landslides in a changing climate, Otta, Central Norway, Clim. Change, 113(2), 413–436, doi:10.1007/s10584-011-0325-0, 2012.

Mergili, M., Marchesini, I., Alvioli, M., Metz, M., Schneider-Muntau, B., Rossi, M. and Guzzetti, F.: A strategy for GIS-based 3-D slope stability modelling over large areas, Geoscientific Model Development, 7(6), 2969–2982, doi:10.5194/gmd-7-2969-2014, 2014a.

Mergili, M., Marchesini, I., Rossi, M., Guzzetti, F. and Fellin, W.: Spatially distributed three-dimensional slope stability modelling in a raster GIS, Geomorphology, 206, 178–195, doi:10.1016/j.geomorph.2013.10.008, 2014b.

Milledge, D. G., Griffiths, D. V., Lane, S. N. and Warburton, J.: Limits on the validity of infinite length assumptions for modelling shallow landslides: testing the infinite length assumption in landslide models, Earth Surf. Process. Landf., 37(11), 1158–1166, doi:10.1002/esp.3235, 2012.

Mulligan, R. and Take, A.: Momentum transfer during landslide tsunami wave generation, vol. 19, p. 11065. [online] Available from: http://adsabs.harvard.edu/abs/2017EGUGA..1911065M (Accessed 25 May 2018), 2017.

Neves Seefelder, C., Koide, S. and Mergili, M.: Does parameterization influence the performance of slope stability model results? A case study in Rio de Janeiro, Brazil, Landslides, doi:10.1007/s10346-016-0783-6, 2016.

Oreskes, N., Shrader-Frechette, K., Belitz, K. and others: Verification, validation, and confirmation of numerical models in the earth sciences, Science, 263(5147), 641–646, doi:10.1126/science.263.5147.641, 1994.

Pagano, T. C., Wood, A. W., Ramos, M.-H., Cloke, H. L., Pappenberger, F., Clark, M. P., Cranston, M., Kavetski, D., Mathevet, T., Sorooshian, S. and Verkade, J. S.: Challenges of Operational River Forecasting, Journal of Hydrometeorology, 15(4), 1692–1707, doi:10.1175/JHM-D-13-0188.1, 2014.

Papathoma-Köhle M. and Glade T.: The role of vegetation cover change for landslide hazard and risk, edited by G. Renaud, K. Sudmeier-Rieux and M. Estrella: The Role of Ecosystems in Disaster Risk Reduction. UNU-Press, Tokyo, 293–320, 2013.

Park, H. J., Lee, J. H. and Woo, I.: Assessment of rainfall-induced shallow landslide susceptibility using a GIS-based probabilistic approach, Engineering Geology, 161, 1–15, doi:10.1016/j.enggeo.2013.04.011, 2013.

Pebesma, E. J. and Bivand, R. S.: Classes and methods for spatial data in R, R News, 5(2), 9–13, 2005.

Peres, D. J., Cancelliere, A., Greco, R. and Bogaard, T. A.: Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds, Natural Hazards and Earth System Sciences Discussions, 1–28, doi:10.5194/nhess-2017-328, 2017.

Petschko, H., Brenning, A., Bell, R., Goetz, J. and Glade, T.: Assessing the quality of landslide susceptibility maps – case study Lower Austria, Natural Hazards and Earth System Science, 14(1), 95–118, doi:10.5194/nhess-14-95-2014, 2014.

Petschko, H., Bell, R. and Glade, T.: Effectiveness of visually analyzing LiDAR DTM derivatives for earth and debris slide inventory mapping for statistical susceptibility modeling, Landslides, 65, doi:10.1007/s10346-015-0622-1, 2015.

Piciullo, L., Gariano, S. L., Melillo, M., Brunetti, M. T., Peruccacci, S., Guzzetti, F. and Calvello, M.: Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides, Landslides, 14(3), 995–1008, doi:10.1007/s10346-016-0750-2, 2017.

Pradhan, A. M. S., Kang, H.-S., Lee, J.-S. and Kim, Y.-T.: An ensemble landslide hazard model incorporating rainfall threshold for Mt. Umyeon, South Korea, Bulletin of Engineering Geology and the Environment, doi:10.1007/s10064-017-1055-y, 2017.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org, last access: 30 November 2017.

Raia, S., Alvioli, M., Rossi, M., Baum, R. L., Godt, J. W. and Guzzetti, F.: Improving predictive power of physically based rainfall-induced shallow landslide models: a probabilistic approach, Geoscientific Model Development, 7(2), 495–514, doi:10.5194/gmd-7-495-2014, 2014.

Reichenbach, P., Cardinali, M., De Vita, P. and Guzzetti, F.: Regional hydrological thresholds for landslides and floods in the Tiber River Basin (central Italy), Environmental Geology, 35(2–3), 146–159, doi:10.1007/s002540050301, 1998.

Reichle, R. H.: Data assimilation methods in the Earth sciences, Advances in Water Resources, 31(11), 1411–1418, doi:10.1016/j.advwatres.2008.01.001, 2008.

Richwien, W., Lesny, K., 2004. Bodenmechanisches Praktikum, Auswahl und Anwendung von bodenmechanischen Laborversuchen. 11 ed., Verlag Gluckauf GmbH, Essen.

Rossi, M., Peruccacci, S., Brunetti, M. T., Marchesini, I., Luciani, S., Ardizzone, F., Balducci, V., Bianchi, C., Cardinali, M., Fiorucci, F., Mondini, A. C., Reichenbach, P., Salvati, P., Santangelo, M. A., Bartolini, D., Gariano, S. L., Palladino, M., Vessia, G., Viero, A., Antronico, L., Borselli, L., Deganutti, A. M., Iovine, G., Luino, F., Parise, M., Polemio, M., Guzzetti, F., Luciani, S., Fiorucci, F., Mondini, Santangelo, N. and Tonellid, G.: SANF: National warning system for rainfall-induced landslides in Italy, in Proceedings of the 11th International & 2nd North American Symposium on Landslides, vol. 2, edited by E. Eberhardt, C. R. Froese, A. K. Turner, and S. Leroueil, pp. 1895–1899, Taylor & Francis, London., 2012.

Rubio, E., Hall, J. W. and Anderson, M. G.: Uncertainty analysis in a slope hydrology and stability model using probabilistic and imprecise information, Comput. Geotech., 31(7), 529–536, doi:10.1016/j.compgeo.2004.09.002, 2004.

Salciarini, D., Fanelli, G. and Tamagnini, C.: A probabilistic model for rainfall–induced shallow landslide prediction at the regional scale, Landslides, doi:10.1007/s10346-017-0812-0, 2017.

Schaake, J. C., Hamill, T. M., Buizza, R. and Clark, M.: HEPEX: The Hydrological Ensemble Prediction Experiment, Bulletin of the American Meteorological Society, 88(10), 1541–1547, doi:10.1175/BAMS-88-10-1541, 2007.

Schmidt, J., Turek, G., Clark, M. P., Uddstrom, M. and Dymond, J. R.: Probabilistic forecasting of shallow, rainfall-triggered landslides using real-time numerical weather predictions, Natural Hazards and Earth System Science, 8(2), 349–357, 2008.

Schweigl, J. and Hervás, J.: Landslide mapping in Austria, European Commission Joint Research Centre, Institute for Environment and Sustainability, Italy. Available at: https://esdac.jrc.ec.europa.eu/ESDB_Archive/eusoils_docs/Images/EUR23785EN.pdf, last access: 30 November 2017, 2009.

Schwenk, H.: Massenbewegungen in Niederösterreich 1953-1990, Jahrbuch der Geologischen Bundesanstalt, 135(2), 597–660, 1992.

Segoni, S., Rosi, A., Lagomarsino, D., Fanti, R. and Casagli, N.: Brief communication: Using punctual soil moisture estimates to improve the performances of a regional scale landslide early warning system, Natural Hazards and Earth System Sciences Discussions, 1–9, doi:10.5194/nhess-2017-361, 2017.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C. and Masson, V.: The AROME-France Convective-Scale Operational Model, Monthly Weather Review, 139(3), 976–991, doi:10.1175/2010MWR3425.1, 2011.

Seyfried, M. S. and Wilcox, B. P.: Scale and the Nature of Spatial Variability: Field Examples Having Implications for Hydrologic Modeling, Water Resources Research, 31(1), 173–184, doi:10.1029/94WR02025, 1995.

Shi, H., Li, T., Liu, R., Chen, J., Li, J., Zhang, A. and Wang, G.: A service-oriented architecture for ensemble flood forecast from numerical weather prediction, Journal of Hydrology, 527, 933–942, doi:10.1016/j.jhydrol.2015.05.056, 2015.

Shute, J., Carriere, L., Duffy, D., Hoy, E., Peters, J., Shen, Y. and Kirschbaum, D.: The Benefits and Complexities of Operating Geographic Information Systems (GIS) in a High Performance Computing (HPC) Environment, AGU Fall Meet. Abstr., 31 [online] Available from: http://adsabs.harvard.edu/abs/2017AGUFMIN31B0072S (Accessed 25 May 2018), 2017.

Siccardi, F., Boni, G., Ferraris, L. and Rudari, R.: A hydrometeorological approach for probabilistic flood forecast, J. Geophys. Res. Atmospheres, 110(D5), doi:10.1029/2004JD005314, 2005.

Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T.: ROCR: visualizing classifier performance in R, Bioinformatics, 21(20), 7881, doi:10.1093/bioinformatics/bti623, 2005.

Smoltczyk, U., 2001. Grundbau-Taschenbuch. Teil 1: Geotechnische Grundlagen. 6 ed., Ernst & Sohn, Berlin.

Song, Y., Huang, D. and Zeng, B.: GPU-based parallel computation for discontinuous deformation analysis (DDA) method and its application to modelling earthquake-induced landslide, Comput. Geotech., 86, 80–94, doi:10.1016/j.compgeo.2017.01.001, 2017.

Stensrud, D. J., Wicker, L. J., Kelleher, K. E., Xue, M., Foster, M. P., Schaefer, J. T., Schneider, R. S., Benjamin, S. G., Weygandt, S. S., Ferree, J. T. and Tuell, J. P.: Convective-Scale Warn-on-Forecast System: A Vision for 2020, Bulletin of the American Meteorological Society, 90(10), 1487–1499, doi:10.1175/2009BAMS2795.1, 2009.

Thiebes, B.: Landslide Analysis and Early Warning Systems. Local and Regional Case Study in the Swabian Alb, Germany, Springer Berlin Heidelberg, Germany, 2012.

Thiebes, B., Bell, R., Glade, T., Jäger, S., Mayer, J., Anderson, M. and Holcombe, L.: Integration of a limit-equilibrium model into a landslide early warning system, Landslides, 11(5), 859–875, doi:10.1007/s10346-013-0416-2, 2014.

Thiebes, B. and Glade, T.: Landslide Early Warning Systems–fundamental concepts and innovative applications, in Landslides and Engineered Slopes. Experience, Theory and Practice - Proceedings of the 12th International Symposium on Landslides, June 12-19, 2016, 3, edited by S. Aversa, L. Cascini, L. Picarelli, and C. Scavia, CRC Press, Napoli, 1903–1911, 2016.

Thiebes, B., Bai, S., Xi, Y., Glade, T. and Bell, R.: Combining landslide susceptibility maps and rainfall thresholds using a matrix approach, , 17, 2017.

Thielen, J., Bartholmes, J., Ramos, M.-H. and Roo, A. de: The European flood alert system–Part 1: concept and development, Hydrology and Earth System Sciences, 13(2), 125–140, doi:10.5194/hess-13-125-2009, 2009.

Tiranti, D., Devoli, G., Cremonini, R., Sund, M. and Boje, S.: Regional landslide forecasting in Piemonte (Italy) and in Norway: experiences from 2013 late spring, Natural Hazards and Earth System Sciences Discussions, 1–30, doi:10.5194/nhess-2017-411, 2017.

Tofani, V., Bicocchi, G., Rossi, G., Segoni, S., D'Ambrosio, M., Casagli, N. and Catani, F.: Soil characterization for shallow landslides modeling: a case study in the Northern Apennines (Central Italy), Landslides, doi:10.1007/s10346-017-0809-8, 2017.

Tran, T. V., Alvioli, M., Lee, G. and An, H. U.: Three-dimensional, time-dependent modeling of rainfall-induced landslides over a digital landscape: a case study, Landslides, 15(6), 1071–1084, doi:10.1007/s10346-017-0931-7, 2018.

Turke, H., 1999. Statik im Erdbau. 3 ed., Ernst & Sohn, Berlin.

5 UNEP: Early Warning Systems: A State of the Art Analysis and Future Directions, Division of Early Warning and Assessment (DEWA), United Nations Environment Programme (UNEP), Nairobi, Kenia. Available at: http://na.unep.net/siouxfalls/publications/early_warning.pdf, last access: 30 November 2017, 63 pp., 2012.

Vincendon, B., Ducrocq, V., Nuissier, O. and Vié, B.: Perturbation of convection-permitting NWP forecasts for flash-flood ensemble forecasting, Natural Hazards and Earth System Science, 11(5), 1529–1544, doi:10.5194/nhess-11-1529-2011, 2011.

10 van Asch, T. W., Buma, J. and van Beek, L. P. H.: A view on some hydrological triggering systems in landslides, Geomorphology, 30(1), 25–32, doi:10.1016/S0169-555X(99)00042-2, 1999.

van der Walt, S., Colbert, S. C. and Varoquaux, G.: The NumPy array: a structure for efficient numerical computation, Computing in Science & Engineering, 13(2), 22–30, doi:10.1109/MCSE.2011.37, 2011.

van Westen, C. J., van Asch, T. and Soeters, R.: Landslide hazard and risk zonation–why is it still so difficult?, Bulletin of
15 Engineering Geology and the Environment, 65(2), 167–184, doi:10.1007/s10064-005-0023-0, 2006.

van Westen, C. J., Castellanos, E. and Kuriakose, S. L.: Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview, Engineering Geology, 102(3–4), 112–131, doi:10.1016/j.enggeo.2008.03.010, 2008.

Wang, Y., Zhao, T. and Cao, Z.: Site-specific probability distribution of geotechnical properties, Computers and Geotechnics, 70, 159–168, doi:10.1016/j.compgeo.2015.08.002, 2015.

20 Wieczorek, G. F. and Glade, T.: Climatic factors influencing occurrence of debris flows, in Debris-flow Hazards and Related Phenomena, edited by M. Jakob and O. Hungr, Springer Berlin Heidelberg, 325–362, 2005.

WMO: Guidelines on Ensemble Prediction Systems and Forecasting, World Meteorological Organization, WMO-No. 1091, Geneva. Available at: http://www.wmo.int/pages/prog/www/Documents/1091_en.pdf, last access: 30 November 2017, 23 pp., 2012.

25 Xie, M., Esaki, T., Zhou, G. and Mitani, Y.: Three-dimensional stability evaluation of landslides and a sliding process simulation using a new geographic information systems component, Environ. Geol., 43(5), 503–512, doi:10.1007/s00254-002-0655-3, 2003.

Xie, M., Esaki, T. and Zhou, G.: GIS-Based Probabilistic Mapping of Landslide Hazard Using a Three-Dimensional Deterministic Model, Nat. Hazards, 33(2), 265–282, doi:10.1023/B:NHAZ.0000037036.01850.0d, 2004.

30 Xie, M., Esaki, T., Qiu, C. and Wang, C.: Geographical information system-based computational implementation and application of spatial three-dimensional slope stability analysis, Comput. Geotech., 33(4), 260–274, doi:10.1016/j.compgeo.2006.07.003, 2006.

Yu, W., Nakakita, E., Kim, S. and Yamaguchi, K.: Improvement of rainfall and flood forecasts by blending ensemble NWP rainfall with radar prediction considering orographic rainfall, Journal of Hydrology, 531, 494–507,
35 doi:10.1016/j.jhydrol.2015.04.055, 2015.

Zhang, S., Zhao, L., Delgado-Tellez, R. and Bao, H.: A physics-based probabilistic forecasting model for rainfall-induced shallow landslides at regional scale, Natural Hazards and Earth System Sciences Discussions, 1–17, doi:10.5194/nhess-2016-348, 2016.

Zieher, T., Rutzinger, M., Schneider-Muntau, B., Perzl, F., Leidinger, D., Formayer, H. and Geitner, C.: Sensitivity analysis
40 and calibration of a dynamic physically based slope stability model, Natural Hazards and Earth System Sciences, 17(6), 971–992, doi:10.5194/nhess-17-971-2017, 2017.
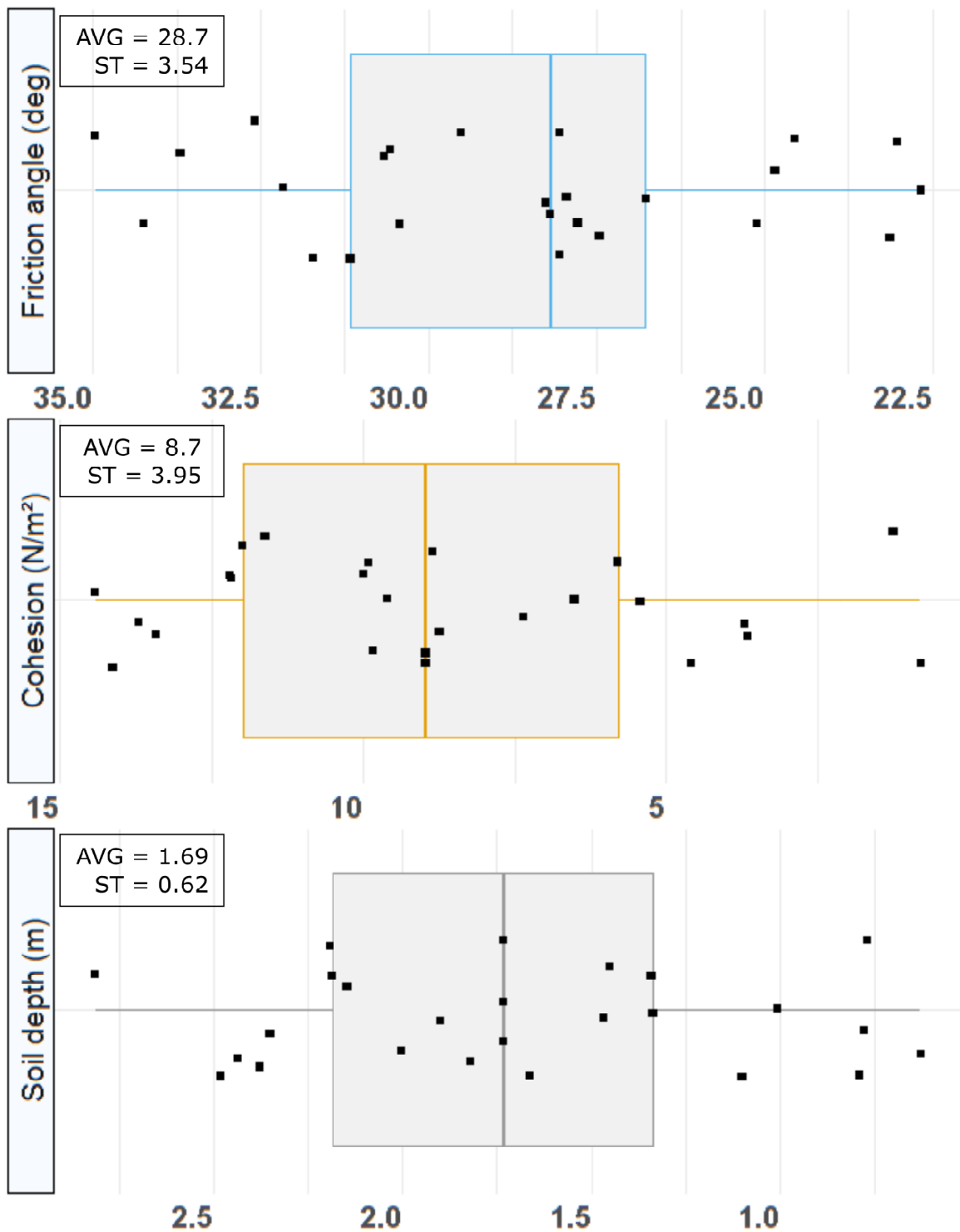
26

Figure 1: Probabilistically derived modeling parameters based on random sampling from a normally distributed function. Jittering dots (to prevent overplotting) indicate individual samples within a plausible parameter range.
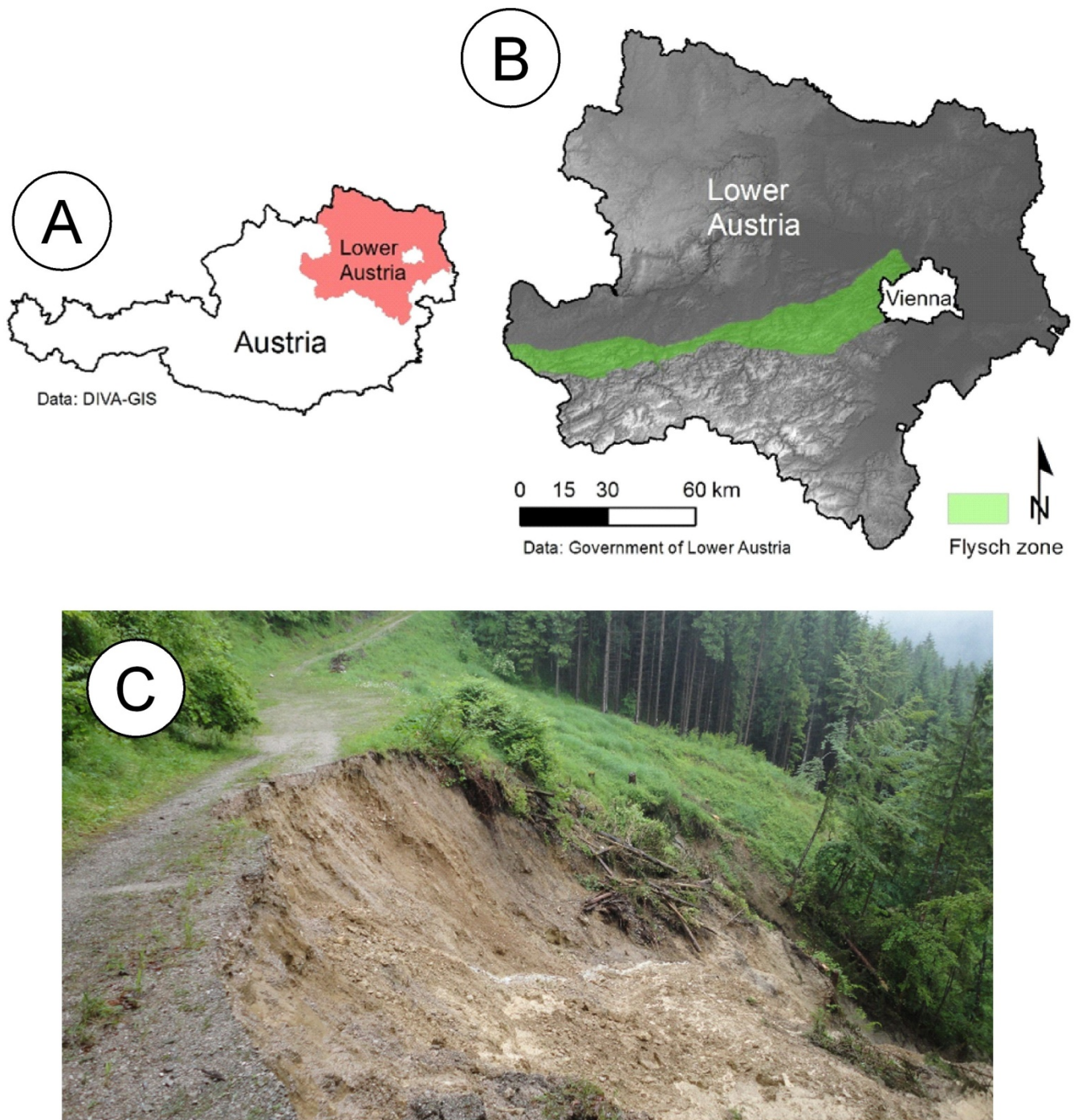
Figure 2: (A) Location of Lower Austria in Austria. (B) Location of the Rhenodanubian Flyschzone in Lower Austria (DTM: CC BY 3.0 AT–Federal state of Lower Austria); (C) Typical earth slide in Lower Austria after a heavy rainfall event in May 2014 (Photograph by K. Gokesch).
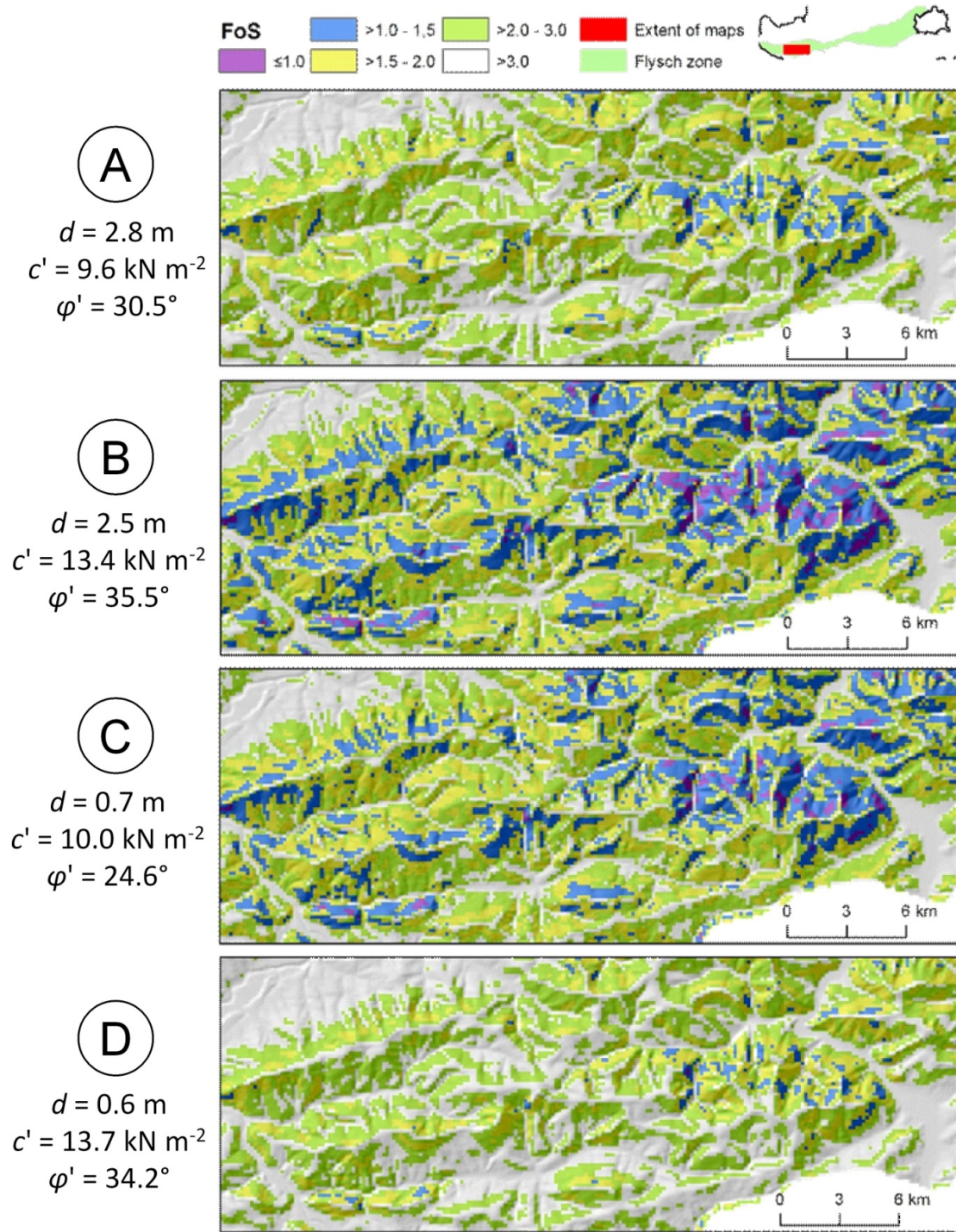
5

Figure 3: Selection of individual model iterations (postage stamp map). Each ensemble member was initialized with altered geotechnical parameters within a plausible range to account for variability and spatial uncertainty, and for a specific rainfall event. FoS values < 1 indicate slope instability.
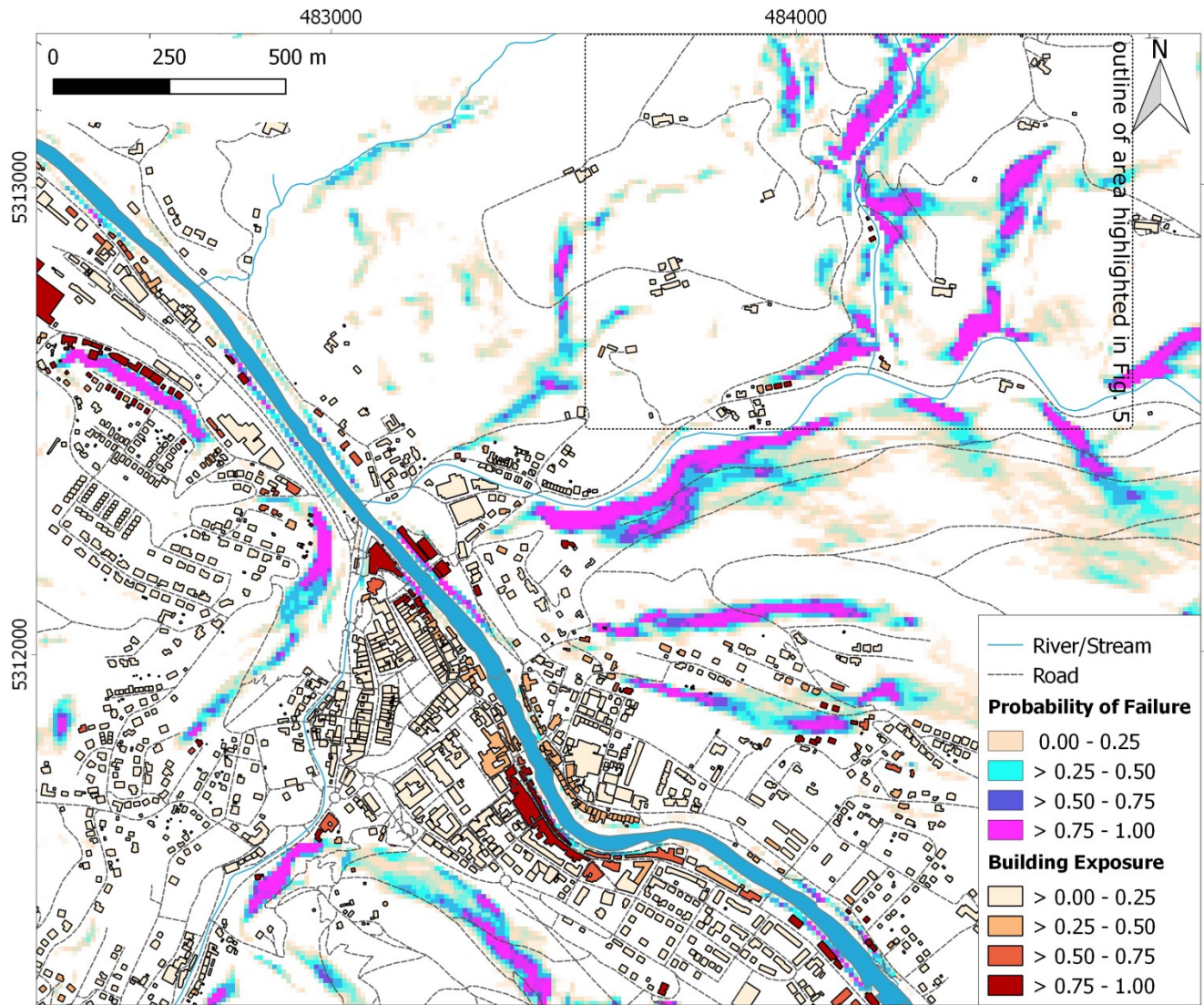
Figure 1: Probability of Failure depicted as a proportion of the ensemble members that predict an event to occur (FoS < 1.0). Building exposure to current slope failure predictions adds an additional information layer for decision makers. Buildings and roads are imported from the freely accessible OpenStreetMap (OSM) database (© OpenStreetMap contributors).
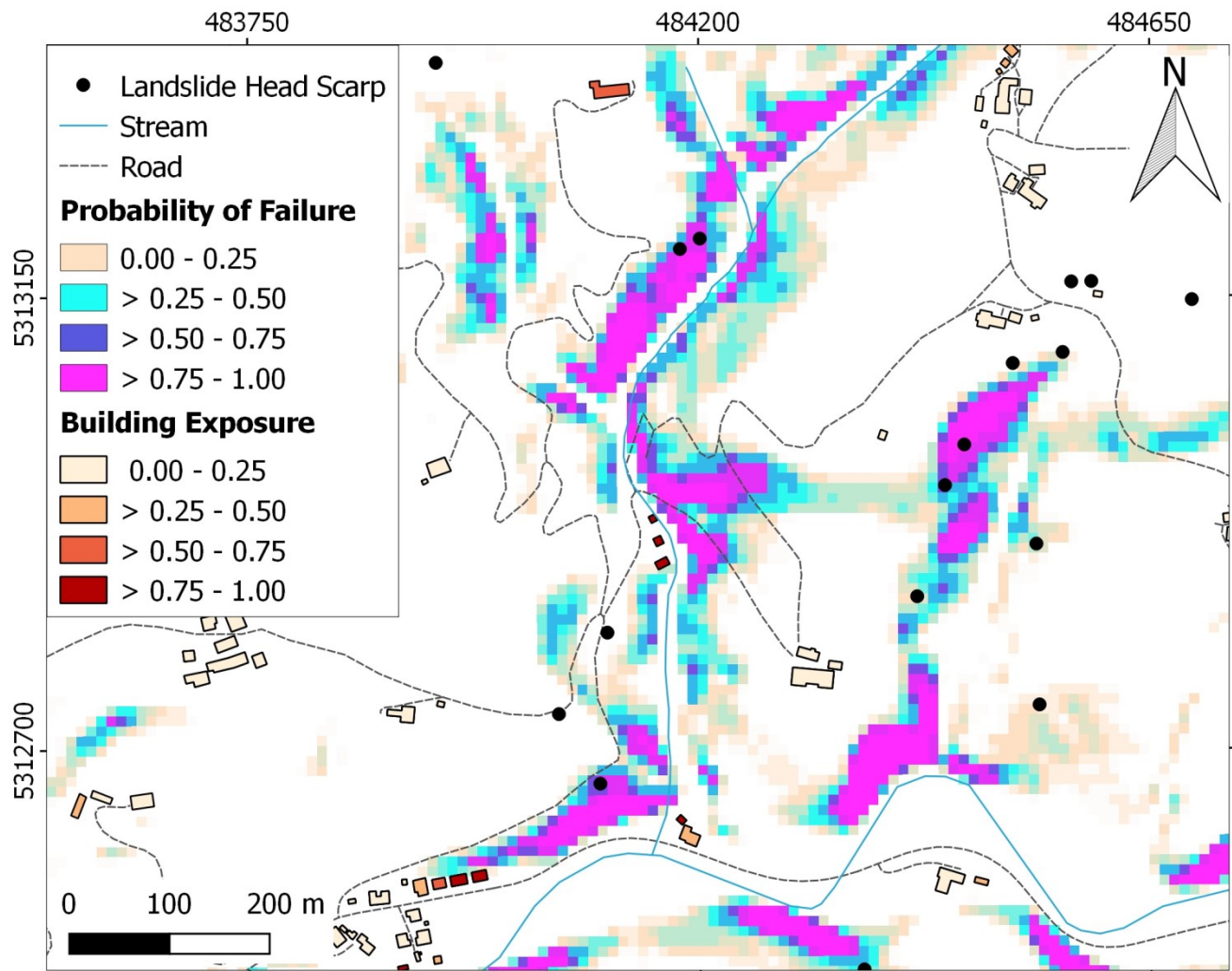
5

Figure 2: Probability of failure map detail for a specific rainfall event. Known historic landslide initiation points (ellipses) partly overlap with current slope stability conditions. However, high spatial resolution, and therefore a high degree of spatial discontinuity, poses a risk for missing many real landslide events in an early warning situation.

5

31

Table 1: Data used in the case study.

Name	Format	Description	Source
DTM	Grid data	LiDAR-based bare earth model with 10m resolution	AT–Federal state of Lower Austria
Geotechnical parameters	Excel table	Published typical parameter values (cohesion, internal friction)	Richwien and Lesny 2004; Smoltczyk 2001; Turke 1999
Rainfall data	Grid data	Historic 3 hour rainfall event with X m resolution	Canli et al., 2017
Landslide data	Point and polygon data	LiDAR-based mapping and damage information on landslides in Lower Austria	Petschko et al. 2015; Building Ground Regristry (BGR)
Buildings and roads	Polygon data		OSM database