Natural Hazards
and Earth System
Sciences
Discussions

Estimating global landslide susceptibility and its uncertainty through ensemble modelling

Anne Felsberg[1], Jean Poesen[1,2], Michel Bechtold[1], Matthias Vanmaercke[1], and Gabriëlle J. M. De Lannoy[1]

[1]KU Leuven, Department of Earth and Environmental Sciences, Belgium
[2]Maria-Curie Sklodowska University, Faculty of Earth Sciences and Spatial Management, Lublin, Poland

Correspondence: Anne Felsberg (anne.felsberg@kuleuven.be)

Abstract. This study assesses global landslide susceptibility (LSS) at the coarse 36-km spatial resolution of global satellite soil moisture observations, to prepare for a subsequent combination of a global LSS map with dynamic soil moisture estimates for landslide modelling. Global LSS estimation intrinsically contains uncertainty, arising from errors in the underlying data, the spatial mismatch between landslide events and predictor information, and large-scale model generalizations. For a reliable

5 uncertainty assessment, this study combines methods from the landslide community with common practices in meteorological modelling to create an ensemble of global LSS maps. The predictive LSS models are obtained from a mixed effects logistic regression, associating hydrologically triggered landslide data from the Global Landslide Catalog (GLC) with predictor variables from the Catchment land surface modeling system (incl. input parameters of soil (hydrological) properties and resulting climatological statistics of water budget estimates), geomorphological and lithological data. Road network density is introduced as

10 a random effect to mitigate potential landslide inventory bias. We use a blocked random cross validation to assess the *model uncertainty* that propagates into the LSS maps. To account for other uncertainty sources, such as *input uncertainty*, we also perturb the predictor variables and obtain an ensemble of LSS maps. The perturbations are optimized so that the *total predicted uncertainty* fits the observed discrepancy between the ensemble average LSS and the landslide presence or absence from the GLC. We find that the most reliable *total uncertainty* estimates are obtained through the inclusion of a topography-dependent

15 perturbation between 15% and 20% to the predictor variables. The areas with the largest LSS uncertainty coincide with moderate ensemble average LSS. The spatial patterns of the average LSS agree well with previous global studies and yield areas under the Receiver Operation Characteristic between 0.63 and 0.9 for independent regional to continental landslide inventories.

1 Introduction

Mitigating landslide impacts requires a good understanding of the spatial and temporal patterns of landslide occurrence. The

20 spatial likelihood is referred to as landslide susceptibility (LSS) and plays a crucial role in risk assessment and land use planning (Guzzetti et al., 2005; Crozier, 2013; Reichenbach et al., 2018). LSS maps derived from environmental conditions are a fundamental tool for informing local population, city planners and decision makers both on the immanent landslide likelihood, but also about secondary effects such as major sediment sources (Crozier, 2013; Maes et al., 2017; Broeckx et al.,

2020). Due to their generalizing nature, LSS models are however prone to uncertainty (Petschko et al., 2014). A large number of

25 LSS models exists, but most focus on local to regional scales and typically lack thorough validation or uncertainty assessment (Reichenbach et al., 2018). Recent advances in computational power and data availability have fostered the development of larger scale LSS maps at continental level (for example Europe: Wilde et al. (2018) and Van Den Eeckhaut et al. (2012), Africa: Broeckx et al. (2018)) or at the global scale (for example Nadim et al. (2006); Hong et al. (2007); Lin et al. (2017); Stanley and Kirschbaum (2017)). While information about the uncertainty would be essential to know how reliable these predictions are,

30 only Broeckx et al. (2018) provide such a measure and only to a limited degree. The intrinsic uncertainty of LSS may become more relevant at larger (global) scales and coarser spatial resolution due to necessary generalizations, higher chances of errors in the underlying data or the increased spatial mismatch between landslide events and predictor information. When estimating LSS both globally and at a coarse spatial resolution to facilitate a subsequent combination with, for example, satellite soil moisture products from Soil Moisture Ocean Salinity (SMOS) or Soil Moisture Active Passive (SMAP) as used by Felsberg

35 et al. (2021), a reliable uncertainty assessment becomes even more crucial.

Uncertainty is typically grouped according to its origin into *model uncertainty* (here: 'How correct are the equations that we use to predict LSS?') and *input uncertainty* (here: 'How correct is the input to these equations?'). *Model uncertainty* stems from heuristic choices that are necessary in the process of model creation, including the choice of the statistical modelling approach, the selection of predictor variables, training data sampling and training data quality (see for example Steger et al.

40 (2015); Pourghasemi and Rossi (2016); Zêzere et al. (2017); Depicker et al. (2020); Lima et al. (2021)). In order to estimate these model-intrinsic errors for a chosen modelling approach, cross validation (CV) is a widely used method where data is divided into a number of subsets, that are subsequently used for training and testing of the model. How to best sample the CV subsets to retrieve realistic uncertainty estimates is in itself a field of research. For LSS maps, random sampling is most common (see for example Broeckx et al. (2018)), while spatial sampling is used less often for an additional uncertainty estimate

45 (see for example Steger et al. (2020) or Depicker et al. (2020)). However, these are known to respectively strongly under- and possibly overestimate the *model uncertainty*, and hybrid methods such as blocked random CV (B-CV) have been suggested to result in the most reliable uncertainty estimates (Roberts et al., 2017). CV leads to multiple LSS model equations (one per CV subset) and the standard deviation of the resulting LSS values gives an indication of the associated *model uncertainty* as shown by Broeckx et al. (2018) for Africa.

50 *Input uncertainty* results from errors in the environmental data as well as from possible variability within a modelling unit such as a grid cell or catchment. More specifically, coarser input data might be less representative for local events, such as landslides. To assess how *input uncertainty* propagates into the *total prediction uncertainty*, ensemble simulations can be used. Meteorologists, for example, simulate the weather based on a distribution of starting conditions and retrieve an ensemble of equally possible predictions (ensemble members). Instead of only one deterministic weather forecast, they use the ensemble

55 average prediction that has been found to perform better than their deterministic counterpart (Kalnay et al., 2006). Note that this is not necessarily the case for individual ensemble members which can and often will perform worse than the deterministic prediction. The uncertainty of the final ensemble average prediction can then be estimated by the variance or standard deviation among the ensemble members.

The *total ensemble uncertainty*, resulting from the combination of these methods that account for *model* and *input uncer-*
tainty respectively, is assumed to be reliable if it matches the observed 'actual' *total uncertainty*. The latter is estimated by
comparing the predicted average LSS against the observed presence and absence of landslides. The gap between this observed
and the predicted *total uncertainty* can then be closed by tuning the magnitude of the ensemble input perturbations. Note that
this implies that the perturbations might in the end not purely capture the *input uncertainty*, but actually compensate for other
sources of uncertainty as well that are not specifically addressed.

In this study, we combine CV and an ensemble approach to create global LSS maps with a reliable *total uncertainty* (full
ensemble standard deviation). We create multiple LSS equations as part of CV ('weak model constraint'), and subsequently
perturb the selected predictor variables (input of the LSS model equations) to retrieve a 'full ensemble' of possible LSS values.
Specifically, we focus on hydrologically triggered landslides based on the Global Landslide Catalog (GLC). We use a mixed
effects logistic regression (MELR) relying on the strong generalizing capabilities of logistic regression, while mitigating the
potential reporting bias of landslide *presences* in the GLC with road network density as a random effect. To limit biases from
unreliable and confounding definitions of landslide *absence* grid cells for the model creation, we introduce a novel approach
based on a 'characteristic distance' between landslides. After having taken these steps to limit the introduced uncertainty, the
B-CV is used to introduce *model uncertainty* and we further add (and tune) ensemble perturbations to the selected predictor
variables to obtain a reliable *total ensemble uncertainty*. This LSS assessment is carried out on the 36-km Equal-Area Scalable
Earth version 2 (EASEv2) grid, which is is also used for satellite soil moisture products from SMOS or SMAP. Producing
spatial LSS estimates at this resolution facilitates a subsequent combination with the satellite-based temporally dynamic data,
as well as the development of computationally intense ensemble approaches. To our knowledge, no framework has earlier been
developed for the assessment of the *total uncertainty* of LSS predictions.

2 Data

2.1 Landslide data

A first step in creating our LSS models is the creation of suitable training datasets (see Fig. 1). We use reported hydrologically
triggered landslide occurrences from the most recent version of the GLC (https://landslides.nasa.gov/viewer, accessed 8th
February 2021). The GLC is a landslide inventory based on media reports (Kirschbaum et al., 2010, 2015), which has been
supplemented with the citizen science based Landslide Reporter Catalog (LRC) data (Juang et al., 2019), see Stanley et al.
(2021) for details. Any reference to the GLC hereafter refers to this combined data product. Despite known English-language
and economic biases (Kirschbaum et al., 2010, 2015), the GLC covers all continents and landslide hotspots. It has already been
used for the creation of two global LSS maps (Stanley and Kirschbaum, 2017; Lin et al., 2017) and was used to train the newest
version of the Landslide Hazard Assessment for Situational Awareness (LHASA) model version 2.0 (Stanley et al., 2021).

For this study, we use 12515 landslides reported mainly between January 2007 and November 2020. Multiple landslides
within the same 36-km EASEv2 grid cell are aggregated into one 'landslide location', resulting in N_{LS}=3757 landslide pres-
ence grid cells (orange grid cells, Fig. A1). This reduces the landslide presence reporting bias of the GLC and also excludes

Natural Hazards
and Earth System
Sciences
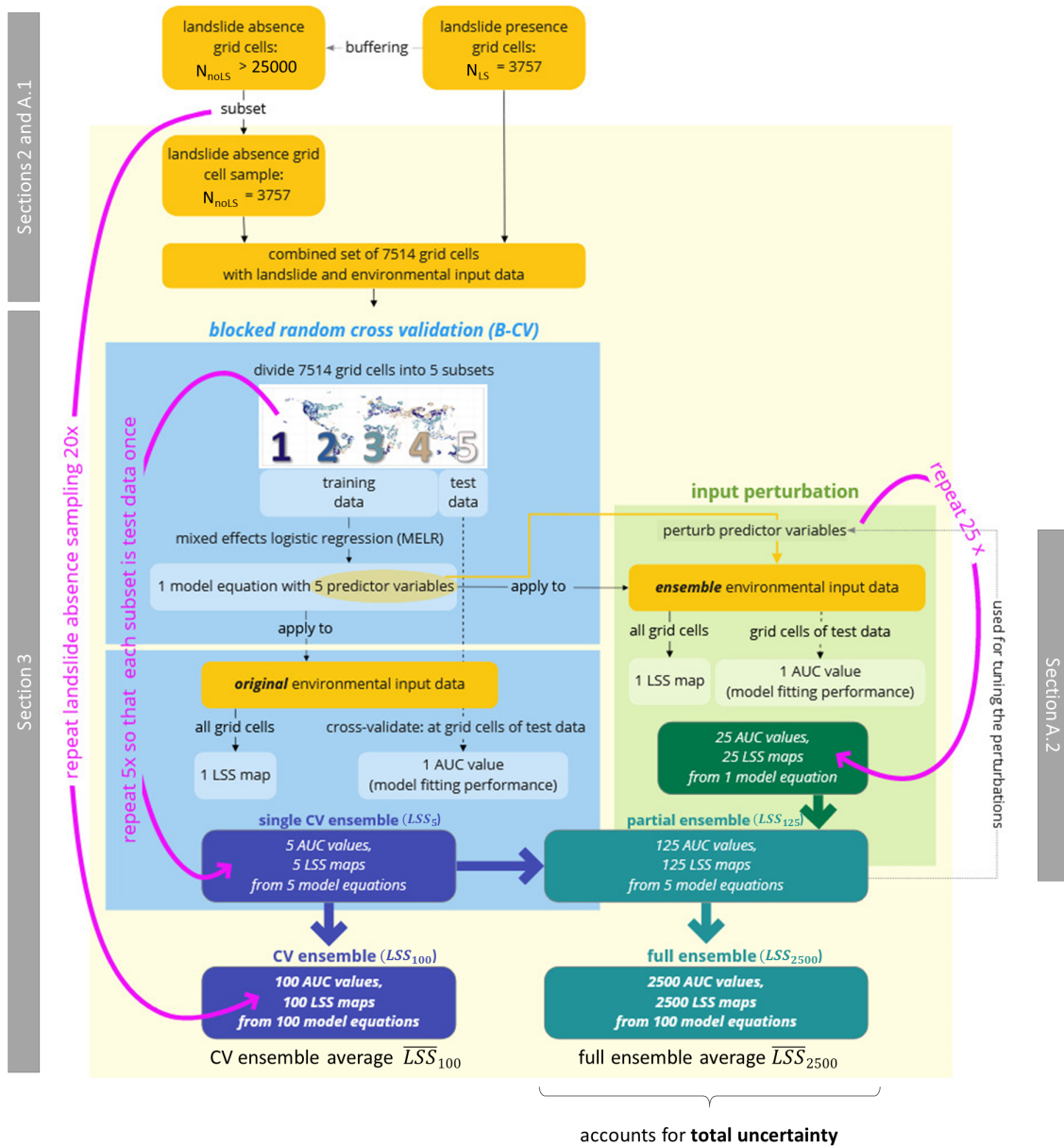Discussions

Open Access

EGU

Figure 1. Schematic of the setup for LSS assessment used in this study and subsections that describe the corresponding methods in more detail. 'Ensemble' refers to a collection of LSS maps. In the course of this study, we refer to different subsets of the full ensemble (LSS_{2500}), namely the ensemble from one single B-CV application (LSS_5), when adding input perturbations to it (LSS_{125}) or when repeating the underlying landslide absence subsampling (LSS_{100}).

any temporal aspect of landslide occurrence. To address the remaining landslide presence bias originating from more landslide reporting in frequently accessed areas, we use data on the average road network density (including highways and all types of roads, ranging from primary to local roads) provided by the Global Roads Inventory Project (GRIP) (Meijer et al., 2018).

95   The creation of realistic statistical LSS models and uncertainty estimates depends on the knowledge of both landslide presences and absences (Roberts et al., 2017; Steger and Glade, 2017; Knevels et al., 2020; Lucchese et al., 2021). Usually, an absence grid cell is simply defined as one without a recorded landslide. For local modelling, this might work when complete and reliable landslide inventories are available. For large or remote areas, however, no reported landslide does not necessarily mean that the site never experienced one. Terrain features show a certain amount of spatial autocorrelation indicating that

100   locations in proximity of a known landslide are generally prone to instability as well. It should therefore be avoided to use grid cells too close to known landslide locations as absence reference (Brenning, 2005). On the other hand, absence grid cells sampled very far from the reported landslide locations, in so-called 'trivial' or easily classifiable areas (for example flat areas), might result in an underrepresentation of stable areas in the vicinity of the known landslide locations (Steger and Glade, 2017). Additionally, it might confound the selection process of geomorphologically meaningful predictor variables and lead to an

105   overoptimistic conception of the resulting LSS map's quality (Steger and Glade, 2017; Lucchese et al., 2021).

In this study, we therefore adopt a sampling strategy as used in earlier LSS assessments (Van Den Eeckhaut et al., 2012; Lin et al., 2017; Zhu et al., 2017; Nowicki Jessee et al., 2018; Knevels et al., 2020; Lucchese et al., 2021), where reliable absence grid cells are defined between a minimum (buffer) and maximum radius around known landslide locations. As a measure of spatial autocorrelation we derive the 'characteristic distance' between two landslides from the GLC (for details

110   see Appendix A1). We use this characteristic distance of 221.43 km (appr. 6 grid cells) as the buffer radius, and 2.5 times this distance (appr. 15 grid cells) as maximum radius. Absence grid cells are hence selected from grid cells 7 to 15 around a landslide occurrence (blue grid cells in Fig. A1). This definition still results in more than six times more absence grid cells ($N_{noLS} > 25000$) than landslide presence grid cells (N_{LS}=3757). We therefore sample from the absence grid cells with a 1:1 ratio (N_{LS}:N_{noLS}) as is commonly done, for example by Brenning (2005), Steger and Glade (2017), Nowicki Jessee et al.

115   (2018), Depicker et al. (2020), Knevels et al. (2020), Lin et al. (2021) and Lucchese et al. (2021). LSS models are subsequently constructed based on data from 7514 grid cells, as illustrated in Fig. 1.

2.2   Environmental data

The 77 predictor variables considered in this study are listed in Table 1 and were selected based on earlier reviews on the most common predictors used for LSS maps (Pourghasemi and Rossi, 2016; Reichenbach et al., 2018). In statistical LSS models,

120   these predictor variables act as proxies for one or multiple processes underlying a landslide (Whiteley et al., 2019). Since LSS in its definition is referring to the spatial likelihood of landslides, we only consider predictor variables that are (quasi) static in time.

To better represent processes underlying hydrologically triggered landslides, we include hydrological climatological statistics of soil moisture in different layers, soil surface temperature, runoff, evaporation and snow depth as possible predictor

125   variables. These climatological statistics include the range (here defined as the difference between percentiles 1 and 99), inter-

Natural Hazards
and Earth System
Sciences

Discussions

Table 1. Environmental predictor variables used in this study, alongside their data source, original spatial resolution and methods used for aggregation to the 36 km EASEv2 grid (if applicable). Predictor variables that are part of the CLSM parameter set or output do not require any spatial aggregation. Long-term climatological statistics of all hydrological variables comprise the range (here: difference between 1st and 99th percentile), inter-quartile range, mean, median, 99th percentile and maximum between 1990 and 2020. MERRA-2 precipitation is used as input for the calculations of the hydrological climatological statistics and has been interpolated to the 36 km EASEv2 grid as part of the simulation process. Units are given for the original data, but are removed through the rescaling of the data to the interval (0,1) (see text).

Predictor variables	Data source	Original spatial resolution	Aggregation method
slope (mean, maximum) [°]	Verdin et al. (2007) provided by USGS	30"	mean and maximum
elevation (mean, standard deviation) [m a. s. l.]	CLSM parameters: details in Verdin (2013)	EASEv2, 36 km	-
depth to bedrock [m]	CLSM parameters: details in Dirmeyer et al. (2006) as part of GSWP-2[a]	EASEv2, 36 km	-
percentage of gravel (0-30 cm) [vol%]	CLSM parameters details in De Lannoy et al. (2014) based on STATSGO2[b] and HWSD1.21[c]	EASEv2, 36 km	-
percentage of clay (0-30 cm and 0-100 cm) [w%]			
percentage of sand (0-30 cm and 0-100 cm) [w%]			
porosity (0-30 cm and 0-100 cm) [m³/m³]			
wilting point divided by porosity (0-30 cm and 0-100 cm) [-]			
compound topographic index, CTI (mean, maximum) = ln(specific catchment area/tan(slope)) [log(m)]	CLSM parameters: details in Verdin (2013)	EASEv2, 36 km	-
land fraction within grid cell	CLSM parameters	EASEv2, 36 km	-
fraction covered by each of 13 lithological classes [-]: *metamorphic rocks, mixed sedimentary rocks, siliclastic sedimentary rocks, basic plutonic rocks, acid plutonic rocks, basic volcanic rocks, intermediate volcanic rocks, carbonate sedimentary rocks, unconsolidated sediments, intermediate plutonic rocks, pyroclastics, evaporites, acid volcanic rocks*	GLiM created by Hartmann and Moosdorf (2012)	polygons	areal fraction within 36-km EASEv2 grid cell
peak ground acceleration, PGA [m/s²] due to earthquakes expected with a return period of 475 years (i.e. 10% exceedance probability in 50 years)	GSHM[d] created by GSHAP[e] (Giardini et al., 2003)	1°	nearest neighbour
rainfall climatological statistics [mm]	MERRA-2 (Bosilovich, 2015)	0.625° lon x 0.5° lat	bilinear interpolation
surface soil moisture climatological statistics (0-5 cm) [m³/m³]	CLSM output	EASEv2, 36 km	-
root zone soil moisture climatological statistics (0-100 cm) [m³/m³]			
profile soil moisture climatological statistics (0-100 cm) [m³/m³]			
land surface temperature climatological statistics [K]			
runoff climatological statistics [mm]			
evaporation climatological statistics [mm]			
snow depth climatological statistics [mm]			

[a] Second Global Soil Wetness Project; [b] State Soil Geographic project; [c] Harmonized World Soil Databank version 1.21; [d] Global Seismic Hazard Map; [e] Global Seismic Hazard Assessment Project;

quartile range, mean, median, percentile 99 and maximum within the time period 1990-2020, derived from 36-km simulations with the Catchment Land Surface Model (CLSM) (Koster et al., 2000; Reichle et al., 2019), forced with Modern-Era Ret-

Natural Hazards
and Earth System
Sciences
Discussions

Open Access

EGU

rospective analysis for Research and Applications, Version 2 (MERRA-2) meteorological data, as in Felsberg et al. (2021). MERRA-2 rainfall input climatological statistics are used to accompany the above-mentioned hydrological ones.

130   Most other predictor variables are part of the 36-km input parameters to the CLSM. Of these, elevation and Compound Topographic Index (CTI) stem from the same underlying Shuttle Radar Topography Mission (SRTM) data as the morphological information on slope from the United States Geological Survey (USGS) (Verdin et al., 2007).

We use lithological information from the Global Lithological Map (GLiM) (Hartmann and Moosdorf, 2012) aggregated to the fraction of a grid cell covered by each of the 13 lithological classes (we exclude the classes 'water', 'ice and glacier', and
135 'no data'). This produces a dataset with 13 fields, each with a continuous fraction estimate. Peak ground acceleration (PGA) (Giardini et al., 2003), the likely level of ground motion from earthquakes, can be seen as a proxy for the fracturation and weakening that lithologies have undergone due to seismic and tectonic activity (Lin et al., 2017; Vanmaercke et al., 2017; Broeckx et al., 2018). Details on the aggregation methods are given in Table 1.

3   Model construction and evaluation

140 ### 3.1   Mixed effects logistic regression (MELR) for model development

In this study, we create a statistical LSS model using MELR (Zuur, 2009), as previously also employed by Steger et al. (2017), and at national scale by Lin et al. (2021) and Lima et al. (2021). Logistic regression is the most commonly used approach for statistical LSS mapping (Reichenbach et al., 2018), and is associated with strong generalizing capabilities (Brenning, 2005) - a necessity when working at the global scale. Within such an approach, the LSS, here defined as the probability of landslide
145 presence within a grid cell, $P(Y = 1)$, is given by:

$$P(Y = 1) = \frac{exp(\alpha + \sum_{i=1}^{n} \beta_i x_i)}{1 + exp(\alpha + \sum_{i=1}^{n} \beta_i x_i)} \qquad (1)$$

with α [-] the intercept, x_i [-] the independent predictor variable, β_i [-] the associated coefficient and n the number of predictor variables. A one unit change in the predictor variable x_i results in a muliplicative change in the odds of landslide presence by $exp(\beta_i)$. Positive (negative) β-values hence indicate an increase (decrease) in LSS with an in increase in the
150 predictor variable. In this study, we work with rescaled predictor variables (between their global minimum and maximum) to detach the magnitude of the β-values from the magnitude of the predictor variable. This facilitates subsequent interpretation.

We employ a stepwise forward technique to select five predictor variables, corresponding well to the commonly used number of predictor variables for LSS assessment at the global scale (Nadim et al., 2006; Stanley and Kirschbaum, 2017; Lin et al., 2017; Reichenbach et al., 2018). Based on the Akaike information criterion (AIC), a measure that is proportional to the sum of
155 squared errors and allows for comparison between non-nested models, we determine the best performing univariate MELR, i.e. the first predictor variable. The AIC comparison is subsequently repeated for multivariate MELR with one additional predictor variable at a time. This stepwise forward selection also allows to exclude correlated predictor variables ($r > 0.7$, following for example Dormann et al. (2013)), so that largely independent predictor variables are used in the logistic regression. Analysis

Natural Hazards
and Earth System
Sciences
Discussions

EGU
Open Access

of the generated models using the Variance Inflation Factor (VIF) proved that this approach indeed successfully prevented a
160   logistic regression model construction based on predictor variables that are too strongly intercorrelated.

The mixed effects approach allows us to include a so-called 'random effect', here the random intercept α, for which we use
the average road network density stratified into 6 groups (divided by the global quintile thresholds). The underlying assumption
here is that the representativeness of the landslide data from the GLC varies with the road network density of the region. We
recognize that road network density may also serve as a proxy for human interference or likelihood of slope cutting and may
165   hence be included as a predictor variable, as was argued by Stanley and Kirschbaum (2017). Instead, we opt for the inclusion
as a random effect in order to favour the selection of natural, physically valid predictor variables while allowing for locations
without roads to also receive a high predicted LSS, as was put forward by Steger et al. (2017) for forested areas. While they
Steger et al. (2017) found results to be very similar between these two options for small biases in the landslide inventory, they
conclude a clear underrepresentation in the underreported areas for stronger biases. We use the glmer function from the lme4
170   package (Bates et al., 2015) to create MELR models in R version 4.0.3 (R Core Team, 2020).

3.2   Cross validation (CV) and input perturbations for reliable uncertainty estimation

The inclusion of random effects in a regression model results in unbiased model parameter estimates, but it does not inform
about the uncertainty of the predictions (Roberts et al., 2017). Similar to a Brier Score (BS) (Wilks, 2011), a measure of the
'actual' total uncertainty can be obtained by comparing the predicted average LSS ($\overline{LSS}$) against landslide observations from
175   the GLC (o) at different grid cells i ($i = 1, ..., N$):

$$BS = \frac{1}{N} \sum_{i=1}^{N} (\overline{LSS} - o)_i^2 \qquad (2)$$

with o being 1 for landslide presence and 0 for absence grid cells. The aim is to design an LSS model setup so that the predicted
total ensemble uncertainty, quantified by the ensemble variance or spread σ_{LSS}^2 matches this 'actual' uncertainty.

In this study, the predicted *total ensemble uncertainty* results from the combination of CV techniques and input ensemble
180   perturbations. For CV, data is separated into five subsets, which subsequently are used for training and testing the model with
the hold-one-out technique, as illustrated in Fig. 1. We employ a blocked random CV (B-CV), as recommended by Roberts et al.
(2017), which we found to indeed yield most realistic error estimates in comparison to random or spatial sampling (not shown).
One subset consists of 20 randomly sampled 'blocks', i.e. small groups, of the 7514 grid cells selected for model creation. We
group the grid cells into a total of 100 blocks according to climatological conditions within 10 predefined regions (roughly two
185   per continent), independent of landslide absence or presence. Within these regions, we mimic typical climatological zonations
(for example that of Köppen) through k-means clustering (Lloyd, 1982) of 30-year average soil surface temperature and rainfall
(see Table 1), dividing each region into 10 blocks. Next, the MELR is iteratively trained on 4 subsets and the model fitting
performance is tested against the 5th, i.e. the hold-one-out subset, until each of these has served as a test-subset once (see
Fig. 1). This results in 5 different model equations and corresponding LSS maps. By repeating the absence sampling 20 times,

Natural Hazards
and Earth System
Sciences
Discussions

EGU

Open Access

190   we obtain a total of 100 LSS maps (referred to as CV ensemble or LSS_{100}, see Fig. 1) that allow for calculations of an ensemble average LSS ($\overline{LSS}_{100}$), as well as a standard deviation ($\sigma_{LSS_{100}}$).

For the input ensemble perturbations, we apply a retrieved model equation to slightly perturbed predictor variable values and obtain an additional LSS map realization. In total, 25 repetitions of this process are conducted, resulting in an ensemble of 25 LSS maps per model equation (see Fig. 1). In combination with the 5 model equations and 20 repetitions for the CV

195   ensemble, this results in a total amount of 2500 LSS maps (referred to as full ensemble or LSS_{2500}) with corresponding average ($\overline{LSS}_{2500}$) and standard deviation ($\sigma_{LSS_{2500}}$), which is representative of the *total prediction uncertainty*.

The perturbations to the predictor variables are randomly sampled from a normal distribution with the mean being the original value of the grid cell and a standard deviation that is referred to as the perturbation magnitude. The latter is tuned, so that the resulting total ensemble spread (including the spread originating from CV) matches the observed actual uncertainty as

200   retrieved by Equation 2. For details of the tuning process, see Appendix A2. We apply the same perturbation magnitude to all predictor variables, but have it increase proportional to the topographic complexity of a location from 0.15 to 0.2. We use the variation of elevation within a grid cell as a measure of said topographic complexity and find this perturbation scaling to yield better results than a globally constant perturbation magnitude. Note that these perturbations do not linearly propagate into the final LSS, as we are working with a logistic regression (see Equation 1) and locations of largest perturbation do not necessarily

205   coincide with large resulting ensemble uncertainty.

3.3   Evaluation

To quantify how well the predicted average LSS ($\overline{LSS}_{2500}$) represents the observed landslide presences and absences, a BS can be used (see Equation 2). Alternatively, the LSS accuracy can be quantified using the Receiver Operation Characteristic (ROC), commonly used as evaluation for categorical response values such as landslide presence or absence (Reichenbach et al.,

210   2018). For the ROC, the true positive rate is displayed against the false positive rate for different possible thresholds in the continuous probability (here: $\overline{LSS}$) that is predicted. We summarize its information by the area under the ROC curve (AUC). A value of 1 indicates a perfect representation of the LSS, while a value of 0.5 indicates that the model does not perform better than random guessing. As part of the CV model construction, we assess the model *fitting* performance for each retrieved model equation against landslide presences within the test subset (see Fig. 1). This evaluation is conducted both for the LSS resulting

215   from the original predictor variable values and the perturbed ones, resulting in a total of 100 AUC values for the CV ensemble (LSS_{100}) and 2500 AUC values for the full ensemble (LSS_{2500}). Note that the performance of individual perturbed ensemble members can be worse than their counterpart based on the original predictor variable values.

The ensemble averages, in contrast, are assumed to outperform deterministic predictions, i.e. have a higher accuracy (Kalnay et al., 2006). To test this assumption we additionally assess model *predictive* performance of ensemble average LSS ($\overline{LSS}_{100}$

220   and $\overline{LSS}_{2500}$), corresponding ensemble members and one fully deterministic reference MELR equation (based on neither CV nor input perturbations) against independent validation data: from i) quarterly reports issued by the Russian Federation (FSBIH, 2018) with $N_{LS} = 56$ aggregated from 183 observations, ii) an inventory for Africa by Broeckx et al. (2018) with $N_{LS} = 649$ aggregated from 18053 observations and iii) FraneItalia, a catalog of recent landslides in Italy (Calvello and Pecoraro, 2020)
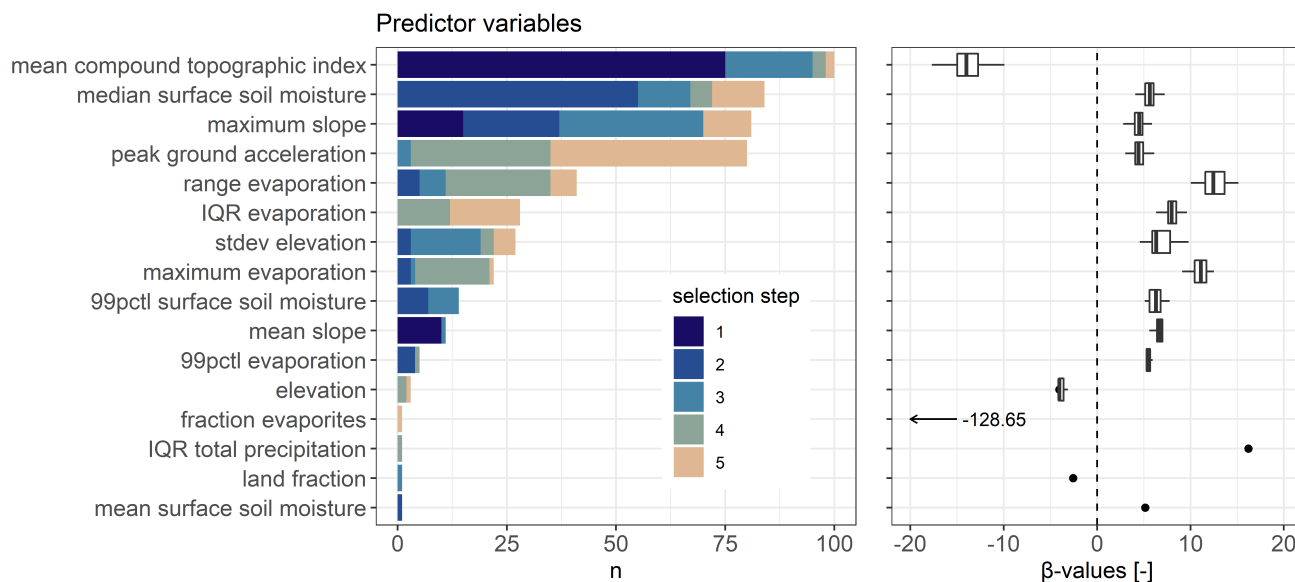
Figure 2. (Left) Frequency of selected predictor variables and (right) corresponding β-values. The 5 best predictor variables (out of 77, see Table 1) are determined using stepwise forward selection for each MELR model equation (n=100). Colors indicate at which selection step (1-5) the predictor variable was selected at. Boxplots base on the n values of the left panel, independent of the selection step. Where $n = 1$, boxplots are replaced by a point.

with $N_{LS} = 309$ aggregated from 5438 observations. These validation inventories cover different climatic zones and hence
225   landslide regimes, stem from (mostly) non-English speaking regions (Africa, Russia, Italy) and include less populated areas
(Africa, Russia). Predictions from an LSS model created based on the GLC might here hence be less reliable. With Italy being
a hot-spot of landslide occurrence within Europe, we are moreover able to assess whether the coarse spatial resolution hinders
realistic regional assessment within smaller, potentially very susceptible areas.

4   Results

230   ### 4.1   LSS model structure

To assess the global LSS, we create 100 MELR model equations using B-CV (see Fig. 1). Each one of these 100 models has a
different set of 5 predictor variables with associated β-values and intercept α. The landslide absence data, used to train these
models, differ for each of the 20 repetitions and subsequently the definitions of the subsets for B-CV vary slightly as well. All
100 models result in LSS maps with very high AUC values above 0.8, with a median of 0.92 for the corresponding test data.
235   Figure 2 (left panel) shows which predictor variables were selected how often and during which step of the selection process.
The right panel shows boxplots of the β-values retrieved for each predictor variable (see Equation 1). Values of the intercept,

which is part of all models, vary with road network density as part of the MELR and mostly have and average close to zero (not shown). The first selected predictor variable was always related to the slope, i.e. either the mean CTI within the grid cell, the maximum slope or the mean slope. CTI, also known as topographic wetness index, was selected as part of all 100 models. It is

240 inversely proportional to slope (see Table 1), which is in line with the negative β-values, i.e. an expected decrease in LSS with increasing CTI. As a second selection, if not again a slope-measure (slope max or standard deviation of the elevation i.e. local relief), we mostly find variables related to the climatologic conditions (median surface soil moisture, range of evaporation, maximum evaporation or surface soil moisture). All of these variables are modeled with positive β-values, i.e. the higher the predictor variable, the larger the odds of a landslide presence and hence the LSS.

245 The areal fraction of evaporites within the grid cell is the only lithological class that was selected, and only in the final selection step. The very unrealistic β-value associated with this predictor (-128.65) suggests that this selection is possibly a statistical artefact. PGA, treated as a proxy for lithologic weakening due to regular seismic activity, is dominantly selected in the later variable selection steps, but still part of about 80% of the models.

4.2 Evaluation of ensemble LSS

250 Based on these 100 model equations and the input parameter perturbations, we obtain the full ensemble average LSS ($\overline{LSS}_{2500}$) and standard deviation ($\sigma_{LSS_{2500}}$) as shown in Fig. 3 (based on 2500 LSS values per grid cell, see Fig. 1). The highest $\overline{LSS}_{2500}$ can be found in the large mountain ranges on all continents as well as coastal areas (especially the islands in South-East Asia). Very flat areas or planes, such as central northern Canada, Siberia, the Tibetan plateau, the Sinai peninsula, the Sahara as well as central Australia have very low $\overline{LSS}_{2500}$. Intermediate $\overline{LSS}_{2500}$ values are found in the northern Rocky Mountains towards

255 Alaska as well as the Kolyma Range in Russia, at the north-eastern shores of South America and the western shores of Africa, along the East African Rift, Scandinavia and India. Areas with intermediate $\overline{LSS}_{2500}$ exhibit large $\sigma_{LSS_{2500}}$ while those with very high or low $\overline{LSS}_{2500}$ typically have smaller $\sigma_{LSS_{2500}}$ associated. This parabolic behaviour is summarized in the density scatter plot of Fig. 4a).

Even though we quantify the uncertainty with a $\sigma_{LSS_{2500}}$ statistic, the distributions around $\overline{LSS}_{2500}$ within one grid cell are

260 mostly non-gaussian, as illustrated for 20 randomly sampled landslide presence and absence grid cells in Fig. 5. Most displayed landslide presence (absence) grid cells have LSS distributions ranging at the upper (lower) end of the interval (0,1). Grid cells 1, 7 and 18, however, exhibit an opposite tendency to their landslide observation.

The ROC curves for ensemble average $\overline{LSS}_{2500}$ are shown in Fig. 6, with the curve for Russia being closest to the upper left corner (AUC: 0.90), that for Africa being a little further from this optimum (AUC: 0.82) and Italy relatively close to the

265 1-1 line (AUC: 0.63). The $\overline{LSS}_{2500}$ map hence performs very well over Russia and Africa, while showing some difficulties to capture the patterns for Italy.

4.3 Impact of input perturbations

The additional inclusion of perturbations to the predictor variables alters the ensemble standard deviations (σ_{LSS}), while only slightly changing the ensemble averages ($\overline{LSS}$) compared to the ensemble statistics obtained by the CV techniques alone. The
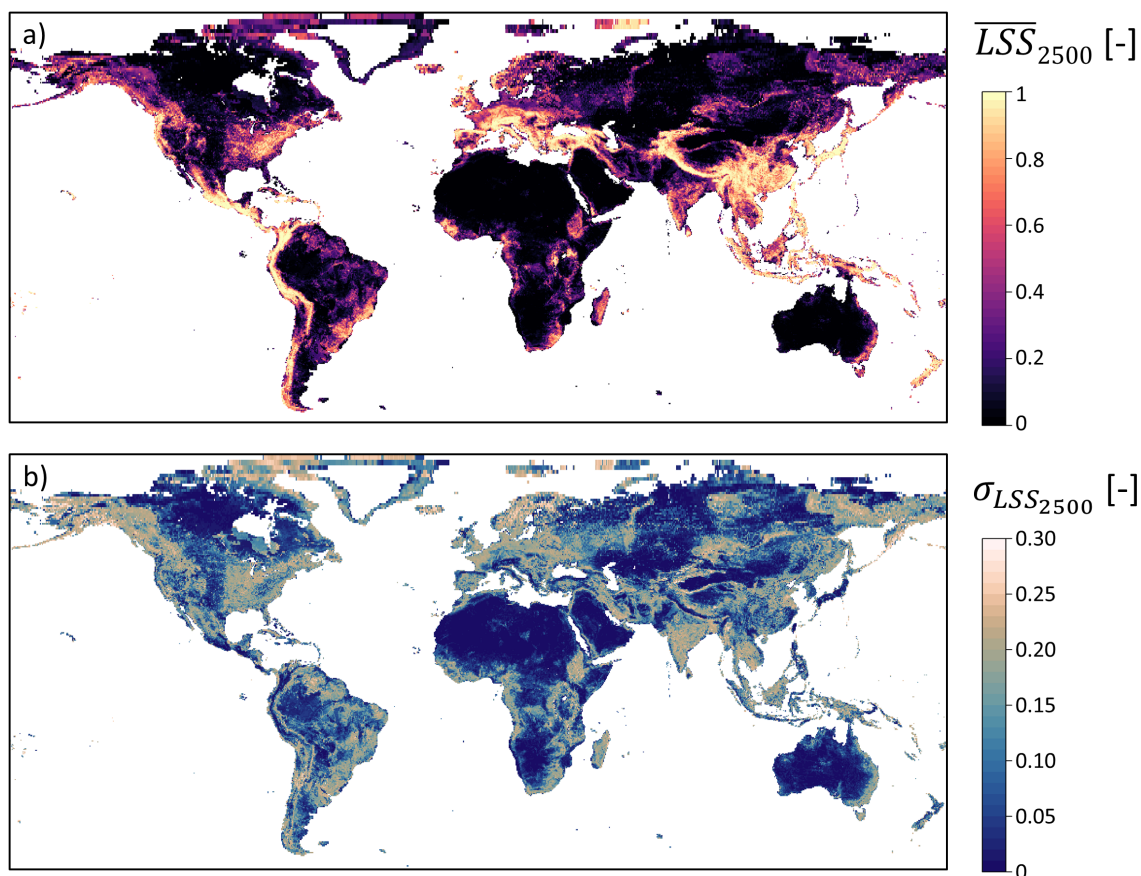
Figure 3. a) Ensemble average LSS ($\overline{LSS}_{2500}$) and b) standard deviation ($\sigma_{LSS_{2500}}$) at 36-km resolution. White areas denote missing values (water bodies, ice). Seemingly larger grid cells in the North are characteristic of the EASEv2 grid projection.

parabolic behaviour of σ_{LSS} with $\overline{LSS}$ (as shown in Fig. 4a and b) is hence amplified. Figure 4c) and d) compare results for the full ensemble (LSS_{2500}) against those of the CV ensemble (LSS_{100}) and show only slightly smaller $\overline{LSS}_{2500}$ in comparison to $\overline{LSS}_{100}$, except for very small $\overline{LSS}$ (< 0.1). By contrast, the standard deviation $\sigma_{LSS_{2500}}$ is larger than $\sigma_{LSS_{100}}$ for nearly all locations, as was intended by the additional predictor variable perturbation. The differences between $\sigma_{LSS_{2500}}$ and $\sigma_{LSS_{100}}$ are the least for the very high and low $\sigma_{LSS_{100}}$.

Figure 7 shows boxplots of the AUC values for individual members of the CV ensemble (LSS_{100}) and the full ensemble (LSS_{2500}) compared against the according CV test subsets, as well as the independent validation inventories. Note that LSS_{100} is a subset of LSS_{2500}. The median AUC value is lower for LSS_{2500} than for LSS_{100} for all reference data. Note though that despite this shift, a number of the LSS_{2500} ensemble members also perform better than any of those from LSS_{100}. As stated before, the intention is not for the individual ensemble members to have the best prediction, but rather the ensemble average
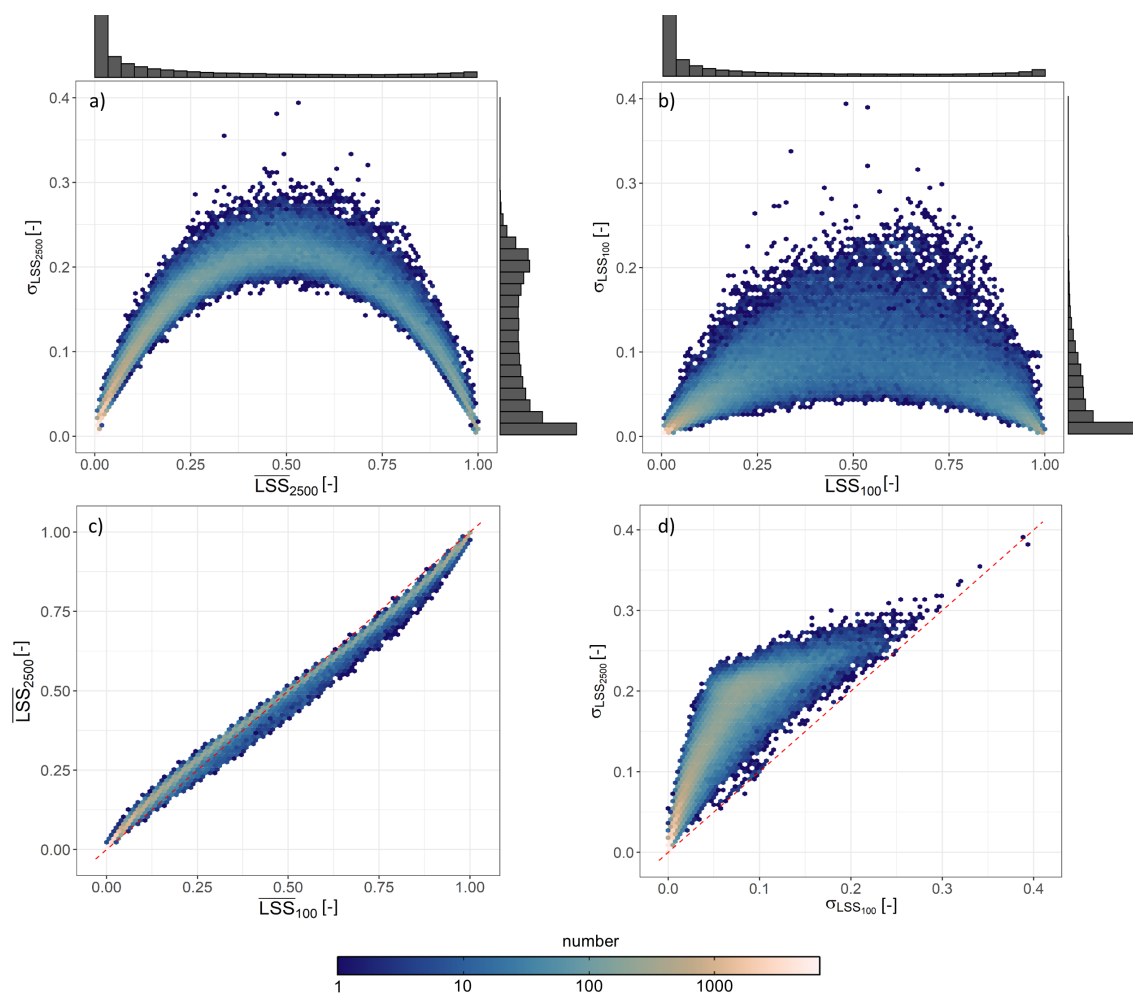
Natural Hazards
and Earth System
Sciences
Discussions
Open Access
EGU

Figure 4. (Top) Ensemble standard deviation LSS (σ_{LSS}) versus ensemble average ($\overline{LSS}$) of a) the full ensemble (LSS_{2500}) and b) CV ensemble (LSS_{100}) with the corresponding marginal distributions. Please note that all marginal distributions contain values of the complete set of 112573 'land' grid cells for which LSS is estimated and are merely scaled by their peak. (Bottom) Comparison of the c) ensemble average and d) standard deviation of LSS_{2500} and LSS_{100}. The 1-1 line (red, dashed) is shown as reference.

280   $\overline{LSS}$. We find AUC values for these $\overline{LSS}_{2500}$ and $\overline{LSS}_{100}$ to be practically the same, since there are only minor difference between the two (Fig. 4c)
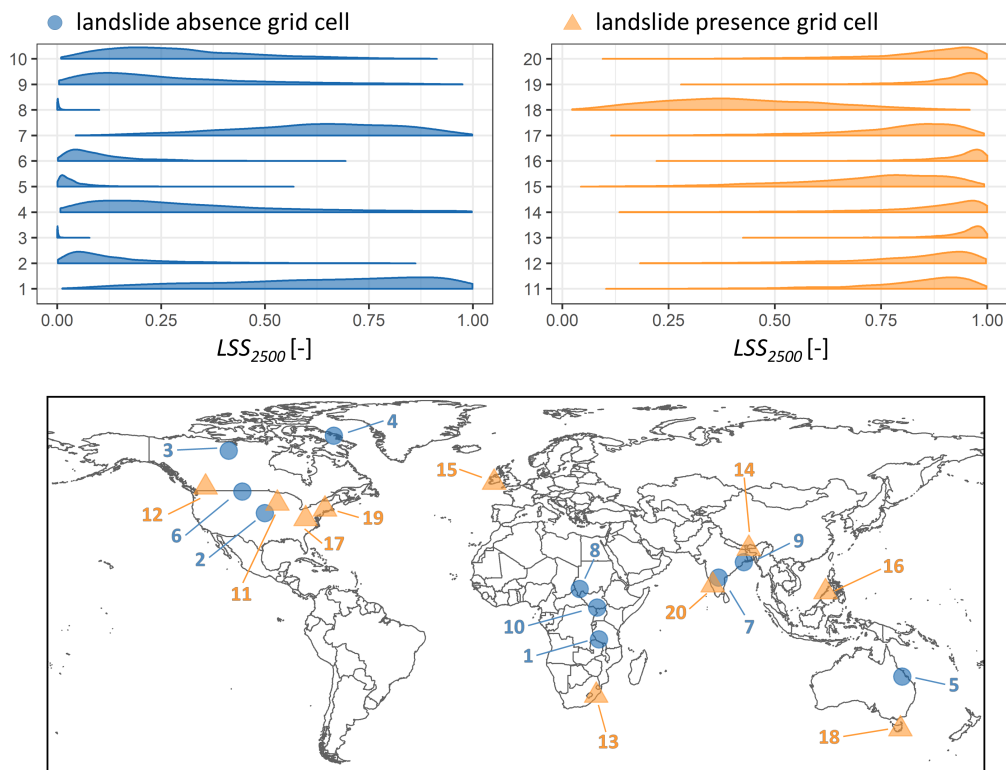
Figure 5. Distribution of ensemble member LSS values (LSS_{2500}, see Fig. 1) within sample grid cells for select landslide presence (light orange triangle on map) and absence (blue circle on map). Please note that the distributions (top) all contain 2500 LSS ensemble members and are merely scaled by their peak to avoid overlaying (large peak) or invisible (small peak, but wide distribution) curves.

5 Discussion

5.1 Selected predictor variables

For the global LSS prediction of this study, the mean CTI per grid cell is the most important predictor variable. Mean and
285   maximum slope within a grid cell are selected less often as the first predictor variable, but one of the two is still included
in nearly every MELR model. We attribute the primary importance of CTI to the fact that our model is trained with data
from hydrologicall triggered landslides (Kirschbaum et al., 2010, 2015), which do not uniquely occur on strong slopes. CTI
intrinsically contains information on the potential hydrological conditions of the site (through the catchment area) as well as
its slope. In line with our study, Emberson et al. (2021) found that the CTI is a strong predictor of rainfall-induced landslides
290   for a number of inventories in the tropics and subtropics. Earlier global LSS maps by Nadim et al. (2006), Hong et al. (2007)
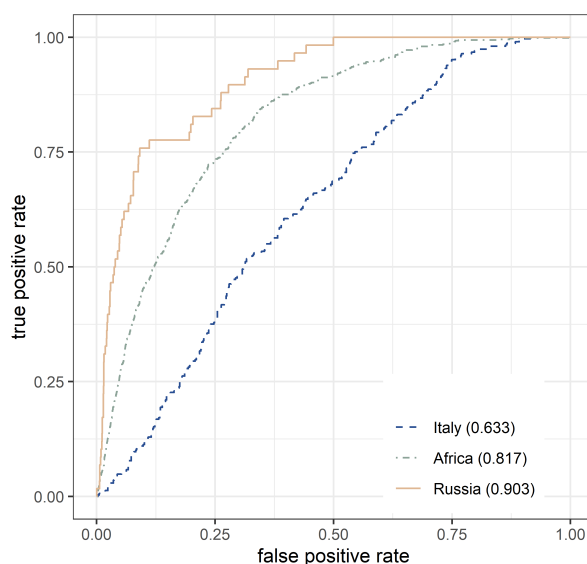
Figure 6. ROC curves of full ensemble average LSS ($\overline{LSS}_{2500}$) for validation inventories from Russia, Africa and Italy. Corresponding AUC values are denoted in brackets.

and Stanley and Kirschbaum (2017) primarily used slope information, while Lin et al. (2017) use relative relief. The latter is comparable to the standard deviation of elevation, which is selected in more than 25% of the models of our study.

The positive connection of LSS to the long-term median surface soil moisture in the MELR equations is logical for hydro-logically triggered landslides. Surface soil moisture is closely related to rainfall and the high correlation between both prevents
295 that both are selected during one model creation (see sect. 3.2). The preference for median surface soil moisture over average rainfall is probably due to the less extreme values in soil moisture (quasi-normal distribution) compared to the highly non-Gaussian distribution of rainfall. The positive relation of LSS to the (inter-quartile) range of evaporation accounts for regions with strong seasonality in rainfall and in the associated evaporation over wet soils. In earlier global LSS maps, Nadim et al. (2006) and Lin et al. (2017) also included information on the soil moisture (soil moisture index by Willmott and Feddema
300 (1992)), with the latter finding soil moisture as the most important predictor variable.

We did not find significant contributions of lithological predictor variables. For Africa, Broeckx et al. (2018) found a (limited) impact of the presence of unconsolidated sediments and siliclastic sedimentary rocks on LSS. Stanley et al. (2021) found the lithology (regrouped from GLiM) to be the least important factor. While local lithology plays a vital role for landslide occurrence, the large data uncertainty and often very broad definitions (as for example elaborated by Campforts et al. (2020) in
305 a different context) hinder meaningful contributions to LSS assessment, even for smaller scale studies. This might also explain why, instead, PGA was favoured as a proxy for structural weakening during the variable selection. The one-time selection of the fraction of land within a grid cell, with a negative β-value assigned, reflects that coastal or shore areas with a low land
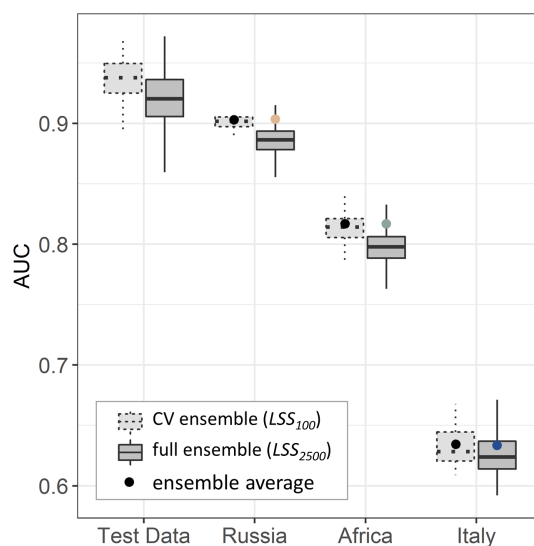
Figure 7. Distribution of AUC for model fitting performance (test data) and model prediction performance (based on independent validation inventories from Russia, Africa and Italy). Results are shown for CV ensemble members (LSS_{100}) and full ensemble members (LSS_{2500}, including CV ensemble members). AUC values for ensemble averages are displayed as points (black: $\overline{LSS}_{100}$, coloured: $\overline{LSS}_{2500}$). The latter correspond to the ROC curves shown in Fig. 6.

fraction are more prone to landslides (higher LSS). Overall, the selected predictor variables, as well as the assigned β-values are in line with general geomorphologic understanding and previous studies.

5.2 Full ensemble results

The spatial patterns of ensemble average LSS ($\overline{LSS}_{2500}$, see Fig. 3) agree well with those of the categorical LSS maps by Stanley and Kirschbaum (2017) at 1 km resolution and Lin et al. (2017) at $0.5°$ resolution (see Fig. 8a and b), with high LSS in strongly mountainous areas, and low LSS for very flat and dry areas. Figure 8 c) and d) show the distribution of the continuous 36-km $\overline{LSS}_{2500}$ per LSS class of these two reference maps. In comparison to both, we find a larger area covered by high $\overline{LSS}_{2500}$ for example in the Eastern United States, Latin America, Mediterranean Europe, India, South-East Asia and New Zealand. At the same time, $\overline{LSS}_{2500}$ shows much less variation than the map by Stanley and Kirschbaum (2017) within large deserts (Sahara, Sinai peninsula and central Australia). This might be a result of the coarser spatial resolution, but is also attributable to the fact that $\overline{LSS}_{2500}$ is strongly governed by hydrological predictor variables apart from the typical geomorphological ones. With a very large proportion of the lowest LSS class, Lin et al. (2017) have even less variation within these areas than $\overline{LSS}_{2500}$.

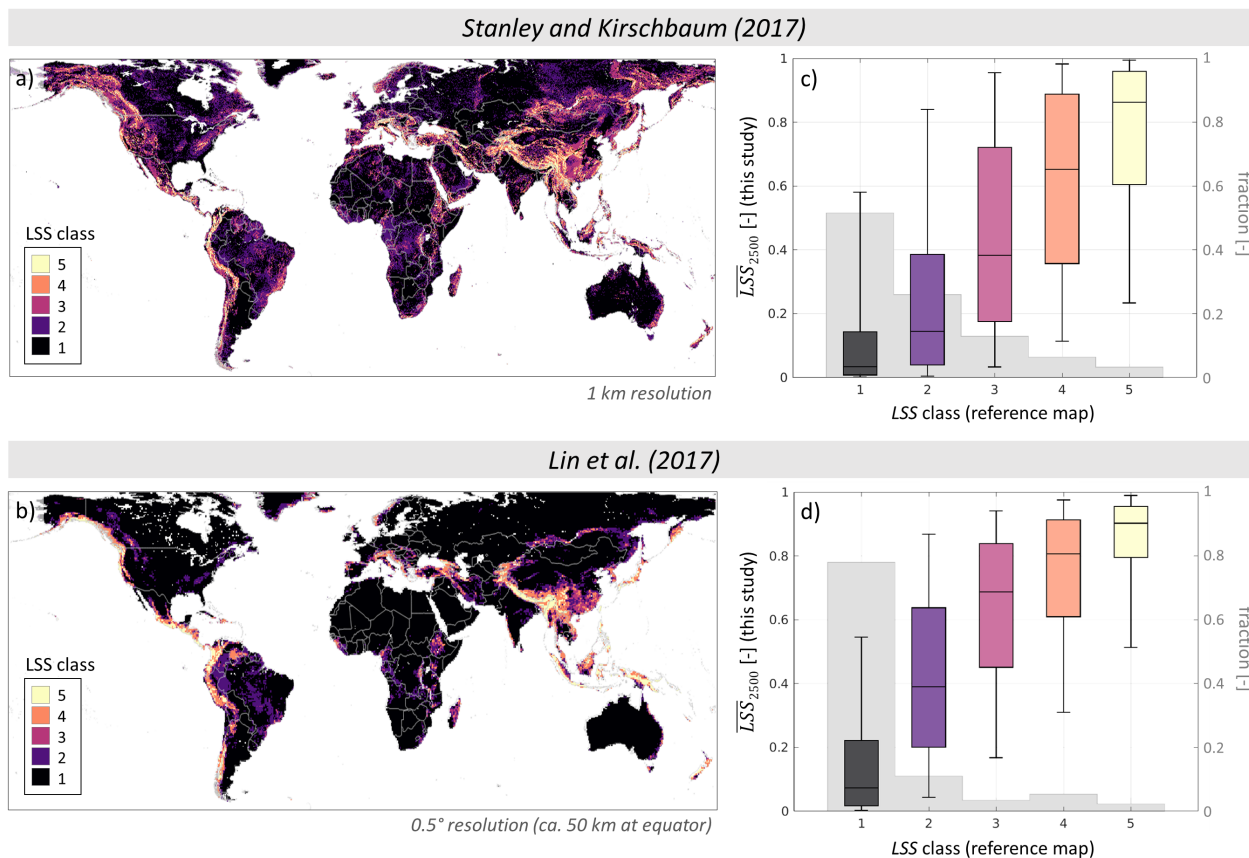**Figure 8.** Comparison of $\overline{LSS}_{2500}$ against global categorical LSS maps by (top) Stanley and Kirschbaum (2017) and (bottom) Lin et al. (2017). (a,b) spatial distributions; (c,d) boxplots of $\overline{LSS}_{2500}$ values: c) from nearest 36-km grid cell and d) the average of the 6 nearest grid cells.Boxplots are underlayn with the fractions of the reference map LSS classes (grey). Note that both reference maps start off from continuous LSS values but use very different thresholds for the class definitions: Lin et al. (2017) set breakpoints at [0.4,0.6,0.7,0.9], following Guzzetti et al. (2006) and Van Den Eeckhaut et al. (2012); Stanley and Kirschbaum (2017) at [0.11,0.49,0.67,0.75], defined so that each category contains twice as many grid cells as the next highest.

These realistic spatial distributions of $\overline{LSS}_{2500}$ are supported by the AUC values calculated for this ensemble average (Fig. 7). The lower AUC value for Africa can be attributed to the fact that the inventory comprises also very old landslides from very different climatic conditions. For Italy, $\overline{LSS}_{2500}$ lacks detail to distinguish between the various levels of high susceptibility: most $\overline{LSS}_{2500}$ values are close to 1, i.e. there is little spatial variation. In general though, these AUC values are

325   in line with those of Stanley and Kirschbaum (2017), who reported AUC between 0.6 and 0.9, and and Lin et al. (2017) of around 0.9, respectively.

Figure 5 shows that the distributions of LSS ensemble members within one grid cell could have a very wide range. Even though in this figure we only selected locations within English-speaking countries and excluded unreliable absence grid cells (see sect. 2.1), it is still possible that an 'absence' could experience a landslide, even if none has been reported in the GLC.

330 A prominent example of this are absence grid cells 1 and 7, located in the East African Rift and India, respectively. Both grid cells have no reported landslide, but very wide LSS distributions, with relatively high LSS values. This discrepancy between prediction and observation could indicate the need for further research in this location.

Calculating the ensemble standard deviation of these distributions ($\sigma_{LSS_{2500}}$) is a good measure of *total prediction uncertainty* associated with a $\overline{LSS}_{2500}$ for one grid cell. The $\sigma_{LSS_{2500}}$ is typically small for distributions at either end of the LSS

335 interval (0,1), resulting in the parabolic pattern as displayed in Fig. 4a-b). This pattern has also been found for local assessments (Guzzetti et al., 2006; Depicker et al., 2020)) and seems to be hold for Broeckx et al. (2018) over Africa as well (visual comparison of two maps). The reasons for this $\overline{LSS}_{2500}$-$\sigma_{LSS_{2500}}$ relationship are twofold: (i) The classification algorithm works best for extreme environmental conditions, such as very steep slope or completely flat areas and has a strongly nonlinear behaviour (logistic regression), and (ii) the predictions are limited to the interval (0,1), restraining the opportunity for devia-

340 tions at the extremes to one side. A comparison of $\sigma_{LSS_{2500}}$ with independent global estimates is currently not possible for lack of uncertainty estimates (Nadim et al., 2006; Hong et al., 2007; Stanley and Kirschbaum, 2017; Lin et al., 2017). However, a comparison with the standard deviations retrieved during the process of random CV for the continental LSS map of Africa by Broeckx et al. (2018) (i.e. not accounting for the *total uncertainty*) reveals that the patterns are very similar, but with less (more) variation in $\sigma_{LSS_{2500}}$ for the very arid (humid) regions.

345 **5.3 Impact of input perturbations**

In this study, we add predictor variable perturbations to the CV approach in order to obtain a more reliable estimate of the *total prediction uncertainty* from the resulting full ensemble. By design, the zero-mean input perturbation does only marginally affects the ensemble $\overline{LSS}$, i.e. $\overline{LSS}_{2500}$ is practically the same as the CV ensemble average $\overline{LSS}_{100}$ (see Fig. 4). Slightly increased (decreased) $\overline{LSS}_{2500}$ at the lower (upper) limits can be attributed to the resampling of predictor variable values

350 if they exceed the definition interval of rescaled predictor variables (0,1). Overall, this introduced bias remains rather small. That the individual ensemble member LSS maps (based on perturbed variables) have a lower median AUC values is, however, realistic: the model equations are created with the original variables so that they are optimally combined into an LSS prediction. Any change of these variables naturally deteriorates the outcome. This is, however, no lack in quality of the ensemble, but rather a side effect. We do not use the individual ensemble members but their average as an LSS prediction, for which we find

355 practically unchanged accuracy between CV ensemble and full ensemble. For a reference, we also retrieve one deterministic MELR equation (resulting in f(CTI, maximum slope, median surface soil moisture, range of evaporation, PGA), not shown). Again, we retrieve nearly identical AUC values (Russia: 0.90, Italy: 0.63, Africa 0.82). The finding of Kalnay et al. (2006) that the introduction of ensembles increases the accuracy of the prediction does not hold for our LSS modelling. This is probably due to the non-linear characteristics of logistic regression and LSS being static.

360   By tuning the predictor variable perturbations to match the *total ensemble prediction uncertainty* to the observed 'actual' uncertainty, we are able to provide statistically reliable uncertainty estimates for the predicted average LSS, even in places where landslide observations are unavailable. As stated before, this optimized spread is introduced to the input variables, but does not actually reflect the input errors only: it also compensates for other uncertainty sources that are not specifically addressed. This does, however, not include most uncertainties introduced by heuristic decisions along the way, such as the

365 choice of spatial resolution, the choice of the statistical model, the definition of a random effect variable, the definition of landslide presence and absence grid cells (characteristic distance) and the preselection of predictor variables to include in the study. These would require dedicated analyses to assess their introduced uncertainty (as for example conducted by Depicker et al. (2020)). Future research could also explore the additional information from the GLC, such as landslide sizes or types instead of reducing the data to landslide presence and absence.

370 **6 Conclusions**

This study presents the first global landslide susceptibility (LSS) map directly developed to be compatible with satellite soil moisture products retrieved from passive microwave sensors, i.e. at a spatial resolution of 36 km. The novel method of combining blocked random CV (B-CV) and predictor variable perturbations results in a reliable assessment of the associated *total prediction uncertainty:* For each grid cell, we estimate 2500 individual LSS values ('full ensemble') that are summarized by the

375 ensemble average LSS ($\overline{LSS}$) and standard deviation (σ_{LSS}, i.e. the uncertainty). Together, these LSS statistics can provide unprecedented information for subsequent global probabilistic spatio-temporal landslide modeling and have the potential to discern areas that deserve more attention for landslide detection.

    A mixed effects logistic regression (MELR) is used as the model structure to relate environmental predictor variables to spatial landslide likelihoods. The objectively selected predictor variables are mainly related to slope and hydrology, in line

380 with the expectations for hydrologically triggered landslides. The odds of landslide occurrence were found to (i) decrease with increasing Compound Topographic Index (CTI), which depends on the ratio of catchment area and slope and (ii) increase with increasing slope, peak ground acceleration (PGA), surface soil moisture (median and 99th percentile) or range of evaporation. The map of the full ensemble $\overline{LSS}$ reproduces global patterns of LSS as presented in previous global studies well. The performance assessment yields area under the ROC curve (AUC) values of 0.9, 0.8 and 0.6 for independent data from Russia, Africa

385 and Italy, respectively. The latter emphasizes that the coarse spatial resolution might not be suited for detailed assessments within small regions. The uncertainty σ_{LSS} is largest for intermediate $\overline{LSS}$. High predicted LSS at (reliable) landslide absence grid cells might furthermore indicate regions that could benefit from future landslide detection and research.

    For the ensemble perturbations of the selected predictor variables we use a perturbation magnitude of 15% to 20%, linearly proportional to the variation of elevation within a grid cell. The magnitude is chosen to match the *total predicted ensemble*

390 *uncertainty* with the observed actual uncertainty relative to data from the Global Landslide Catalog (GLC). Adding these perturbations does not linearly propagate into the ensemble spread due to the nature of logistic regression. It increases the ensemble spread for locations of intermediate $\overline{LSS}$ while having negligible impact where $\overline{LSS}$ is close to its lower or upper

limit. The ensemble $\overline{LSS}$ remains practically unchanged, regardless of the input perturbations. This is also the case when

comparing $\overline{LSS}$ to that of pure B-CV ensembles, leaving the accuracy of these average predictions virtually unchanged. In

395   short, these novel methods explicitly focus on the uncertainty quantification. The availability of global reliable uncertainty

estimates is an unprecedented new contribution to the suite of global LSS maps, and it will support stochastic landslide hazard

modeling.

Code and data availability.   For most of the landslide and environmental predictor data we refer the reader to the provided sources. Climato-

logical statistics of hydrological parameters can be obtained in netCDF format by contacting the authors, as is the case for the R code for the

400   complete setup of model creation. The resulting full LSS ensemble is available as a netCDF file as well and will be publicly available after

the acceptance of this paper.

Appendix A

A1   Landslide absence sampling
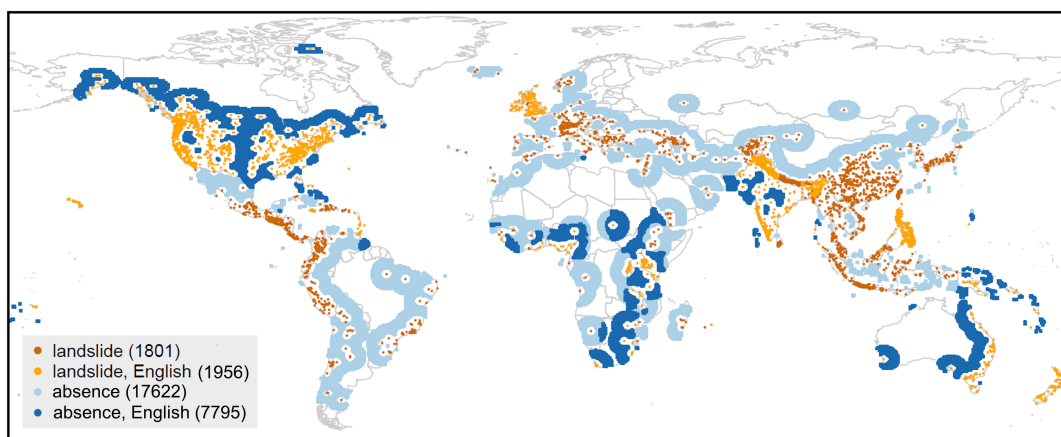
Figure A1. Spatial distribution of landslide presence (shade of orange) and absence (shade of blue) grid cells at 36-km resolution, for English

speaking countries (light orange and dark blue) and non-English speaking countries (dark orange and light blue). White indicates grid cells

that are excluded during the model creation process (buffer and maximum radius around landslide location, see sect. 2.1). The numbers are

the sum of each subgroup of grid cells.

Figure A1 shows the $N_{LS}=3757$ landslide locations based on data from the GLC aggregated to the 36-km EASEv2 (sec-

405   tion 2.1). Landslide absence grid cells are sampled between a minimum (buffer) and maximum distance around known landslide
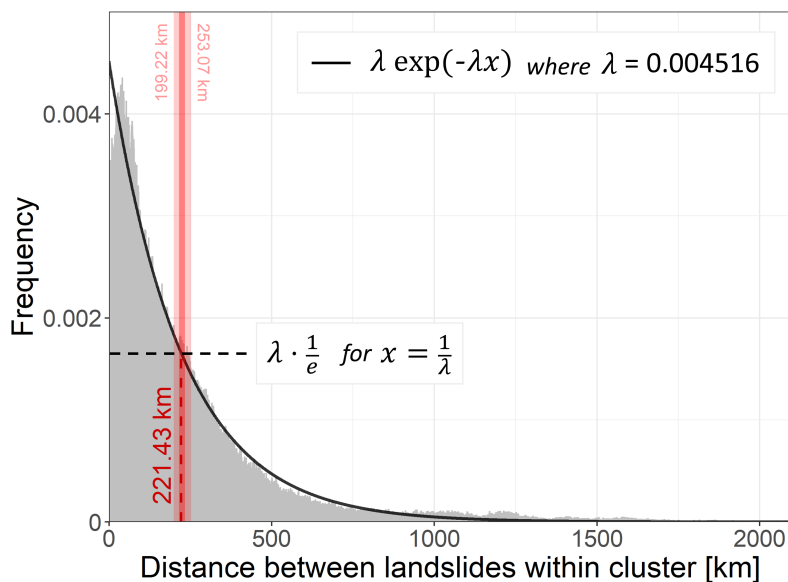
Figure A2. Histogram of distances [km] between landslides within a k-means cluster (for 100 clusters across the globe) of the GLC (grey) and Poisson exponential fit (black line) to retrieve the characteristic landslide distance (red). The red dashed line indicates median characteristic landslide distance from 50 repetitions of the k-means clustering, with the smallest and largest characteristic distance indicated by the light red bar and numbers at top.

locations (N_{noLS}=25417). These distances can be based on either heuristic choices (Van Den Eeckhaut et al., 2012; Lin et al., 2017; Knevels et al., 2020) or empirical approaches (Zhu et al., 2017; Nowicki Jessee et al., 2018; Lucchese et al., 2021).

For our global study, we set a buffer based on the probability for any two landslide locations from the GLC to be reported within a specific distance interval for 100 spatially defined clusters (k-means-clustering (Lloyd, 1982) on latitude and lon-
410 gitude). Figure A2 shows that the frequency of encountering two landslide locations decreases for larger distances and can be characterized by a Poisson exponential fit. In line with the definition of autocorrelation length (Gaspari and Cohn, 1999), we define the 'characteristic distance' between two landslides as the distance where the probability to meet another landslide drops by 1/e. We use this characteristic distance of 221.43 km or appr. 6 36-km grid cells (median of characteristic distances retrieved for 50 repetitions of the clustering) as a buffer around landslide locations. The maximum distance around a landslide
415 is subsequently defined as 2.5 times this characteristic distance (553.58 km, appr. 15 grid cells), borrowing from the data as-similation community where 2.5 times the autocorrelation length is a measure for absence of correlation (Gaspari and Cohn, 1999; De Lannoy, 2006; De Lannoy et al., 2010).

Landslide absence grid cells are hence selected from grid cells 7 to 15 grid cells around a landslide location (blue grid cells in Fig. A1). These distances are inevitably much larger than those found in literature for finer-scale studies, because
420 autocorrelation lengths are scale-dependent and the retrieved characteristic distance is influenced by the spatial extent, or the definition of the clusters in our case.

A2 Input perturbation and optimization

For a reliable assessment, the *total ensemble prediction uncertainty* of the obtained ensemble average $\overline{LSS}$ map ideally should match the observed actual uncertainty. The first can be defined for a single location by the standard deviation (σ) among the

425 LSS ensemble members (LSS_i, with $i = 1, ..., N_{ens}$), as also displayed in sect. 4 and Fig. 3. Similarly, it is possible to assess the according variance (σ^2), referred to as ensemble spread (*ensp*):

$$ensp = \frac{1}{N_{ens}} \sum_{i=1}^{N_{ens}} (LSS_i - \overline{LSS})^2 \tag{A1}$$

Similar to a Brier Score (Wilks, 2011), the observed actual uncertainty at a single location is defined as the difference between $\overline{LSS}$ and the aggregated landslide observations from the GLC (o), referred to as the ensemble skill (*ensk*):

430 $$ensk = (\overline{LSS} - o)^2 \tag{A2}$$

where o is 1 (0) in case of a landslide presence (absence) grid cell. The smaller *ensk*, the closer the predicted $\overline{LSS}$ to the observation.

The optimization of the uncertainty estimates entails tuning of *ensp* to match *ensk*. In this study, this is done by varying the perturbation magnitude that is added to the input variables (see sect. 3.2). Talagrand et al. (1997) defined spread-skill

435 relationships that allow to verify the statistical consistency between the assumed uncertainty (chosen perturbation) and the actual 'observed' uncertainty based on the ergodicity principle. Over a large number of realizations, i.e. for large enough ensembles, $< ensk - ensp > \rightarrow 0$ or

$$\frac{< ensk >}{< ensp >} \rightarrow 1 \Leftrightarrow log\left(\frac{< ensk >}{< ensp >}\right) \rightarrow 0 \tag{A3}$$

where $< . >$ denotes the average. In most hydrological or meteorological applications, this is the temporal average within

440 one grid cell. As this is not applicable for the static LSS data, we consider (i) spatial averages $< ensk > / < ensp >$ per LSS interval as well as (ii) the distribution of individual $ensk/ensp$ per grid cell. Both should only be performed over grid cells with reliable information about landslide presence or absence (see Appendix A1). Note that this definition of $< ensk >$ corresponds to the definition of the Brier Score as given in sect. 3.2.

We tested various magnitudes of perturbations to the rescaled predictor variables either by using (i) a globally constant

445 standard deviation or (ii) a standard deviation proportional to the topographic complexity (i.e. the variation within a grid cell, here the standard deviation of elevation). A range of possible perturbation options was tested for a partial ensemble (LSS_{125}, i.e. no repetition of landslide absence sampling as illustrated in Fig. 1). Figure A3 shows $log(< ensk_{LSS_{125}} > / < ensp_{LSS_{125}} >)$ for 10 intervals of $\overline{LSS}_{125}$ and two examples of constant and linear perturbations. Adding any of the four perturbations brings $log(< ensk_{LSS_{125}} > / < ensp_{LSS_{125}} >)$ values closer to zero, i.e. improves the spread-skill relationship,

450 compared to results without a perturbation (LSS_5, single CV ensemble). Linear perturbations introduce larger spread in areas of higher $\overline{LSS}_{125}$ resulting in $log(< ensk_{LSS_{125}} > / < ensp_{LSS_{125}} >)$ closer to zero than constant perturbations, and are therefore preferred here.
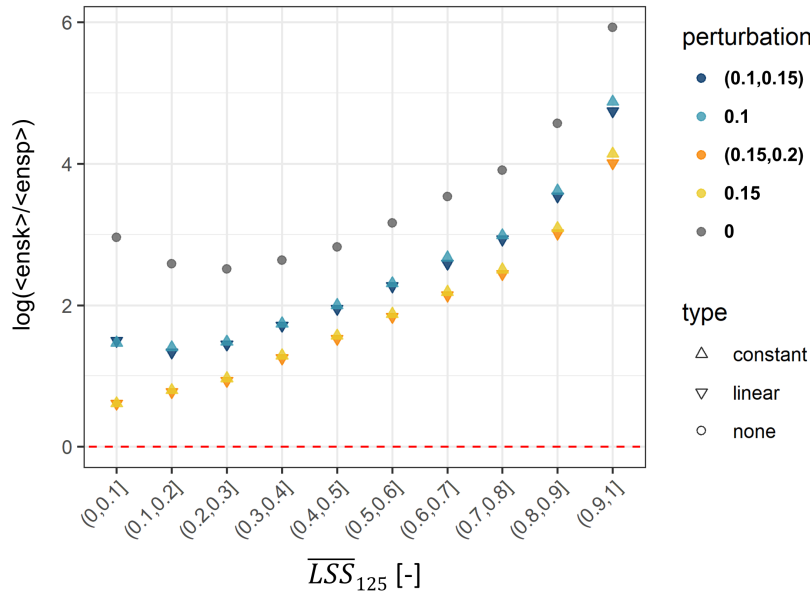
Figure A3. Spread-skill relationship $log(<ensk>/<ensp>)$, stratified per ensemble average LSS ($\overline{LSS}_{125}$). The optimum of 0 indicated by red dashed line. Shapes indicate the type and colours the magnitude (constant) and interval (linear) of perturbation.

We futher analyze the distribution of individual *ensp* and *ensk* across all grid cells in Fig. A4 (top), stratified for landslide presence and absence. Ideally, *ensp* versus *ensk* should stay close to the 1-1 line. Adding a perturbation to the predictor variables

455   (Fig. A4 c in comparison to a) nudges the distribution in this direction, but fails to do so for large *ensk*: a large *ensk* results from a large difference between $\overline{LSS}_{125}$ and landslide observation (o), and often coincides with very small *ensp*. This can be attributed in part to the incompleteness of the GLC (missing observations in a very susceptible area) and the coarse spatial resolution of this study (one very susceptible location surrounded by dominantly non-susceptible area within grid cell). Note also that the logistic regression (see Equation 1) does not linearly propagate the perturbations of predictor variables into the

460   resulting LSS values, especially not at the edges of the definition interval (0,1). Accepting this tail of the distribution as an unavoidable characteristic, we further analyze the histogram of grid cell wise $log(ensk/ensp)$ as displayed in Fig. A4 c) and d). An optimal perturbation would result in median $log(ensk/ensp)$ close to zero and a small inter-quartile range (IQR). We therefore define the optimal perturbation for a minimum Euclidean distance (d) between the point ($median|IQR$) and ($0|0$), averaged over the distribution of observed landslide presences and absences ($o = 0, 1$):

465   $$\bar{d} = \frac{1}{2} \sum_{o=0,1} (median^2 + IQR^2)_o \qquad (A4)$$

The $\bar{d}$ for a range of possible linear perturbation options for LSS_{125} is summarized in Fig. A4e). The optimal perturbation (smallest $\bar{d}$) scales the applied standard deviation according to topographic complexity, represented by the standard deviation

of elevation within a grid cell, between $(0.15, 0.2)$. Fine tuning of the standard deviation is left for future work, but could involve other variables or transformations thereof or different amounts of perturbations per predictor variable.
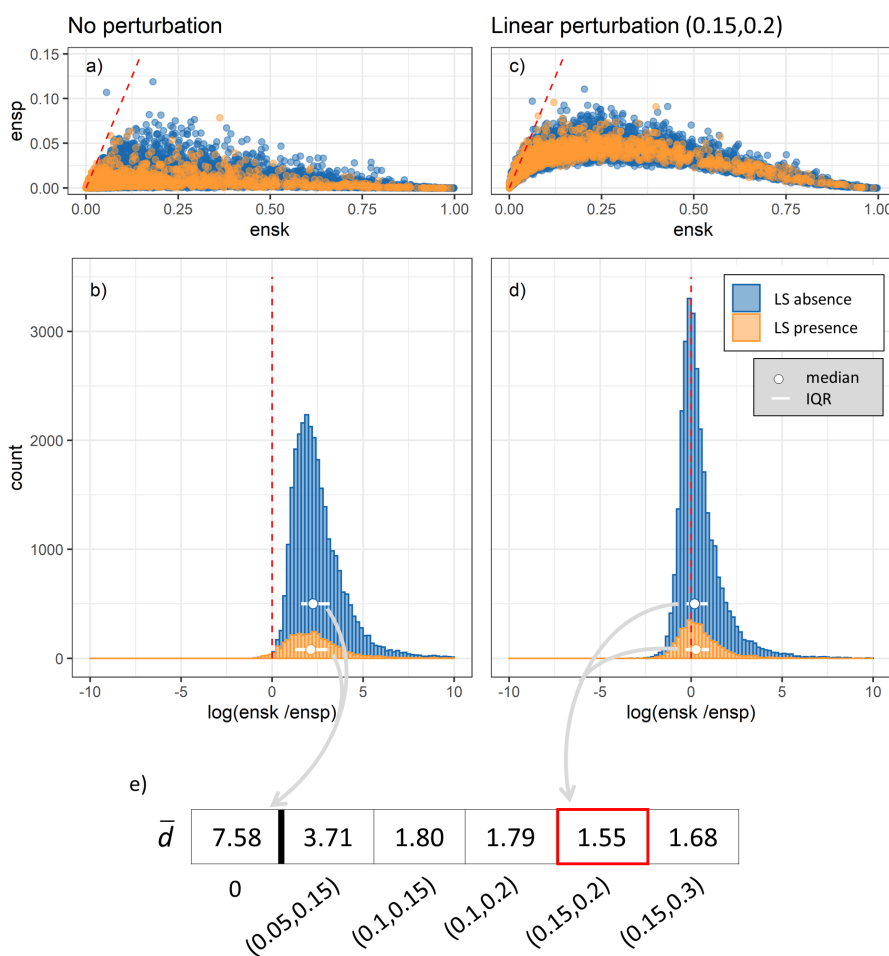

Figure A4. Spread-skill relationship per grid cell with the optimum indicated by the red dashed lines: (top; a,c) scatter plots of *ensk* against *ensp*, (middle; b,d) histograms of $log(ensk/ensp)$, stratified for landslide presence and absence (between buffer and maximum distance). (Bottom; e) summary of the average Euclidean distance $\bar{d}$ for all applied linear perturbations with the optimum framed in red. Shown are results for a-b) without perturbation of predictor variables (LSS_5), and c-d) for linear perturbation of predictor variables within the interval $[0.15, 0.2]$ (LSS_{125}). In other words, (a-b) account for *model uncertainty* alone whereas (c-d) account for the *total uncertainty* (see Fig. 1).

Natural Hazards
and Earth System
Sciences

Open Access

EGU

Discussions

480 **Abbreviations**

N_{LS} number of landslide locations, i.e. landslide presence grid cells

N_{noLS} number of landslide absence grid cells

ensk ensemble skill

ensp ensemble spread

485 **AIC** Akaike information criterion

AUC area under the ROC curve

B-CV blocked random CV

BS Brier Score

CLSM Catchment Land Surface Model

490 **CTI** Compound Topographic Index

CV cross validation

EASEv2 Equal-Area Scalable Earth version 2

GLC Global Landslide Catalog

GLiM Global Lithological Map

495 **GRIP** Global Roads Inventory Project

GSHAP Global Seismic Hazard Assessment Project

GSHM Global Seismic Hazard Map

GSWP-2 Second Global Soil Wetness Project

HWSD1.21 Harmonized World Soil Databank version 1.21

500 **IQR** inter-quartile range

LHASA Landslide Hazard Assessment for Situational Awareness

LRC Landslide Reporter Catalog

LSS landslide susceptibility

MELR mixed effects logistic regression

505 **MERRA-2** Modern-Era Retrospective analysis for Research and Applications, Version 2

PGA peak ground acceleration

R-CV random CV

ROC Receiver Operation Characteristic

SMAP Soil Moisture Active Passive

510 **SMOS** Soil Moisture Ocean Salinity

SRTM Shuttle Radar Topography Mission

STATSGO2 State Soil Geographic project

USGS United States Geological Survey

VIF Variance Inflation Factor

References

515 Bates, D., Mächler, M., Bolker, B., and Walker, S.: Fitting Linear Mixed-Effects Models Using lme4, Journal of Statistical Software, 67, 1–48, https://doi.org/10.18637/jss.v067.i01, 2015.

Bosilovich, M. G. L.: MERRA-2: File Specification, Tech. rep., 2015.

Brenning, A.: Spatial Prediction Models for Landslide Hazards: Review, Comparison and Evaluation, Natural Hazards and Earth System
520 Science, 5, 853–862, 2005.

Broeckx, J., Vanmaercke, M., Duchateau, R., and Poesen, J.: A Data-Based Landslide Susceptibility Map of Africa, Earth-Science Reviews, 185, 102–121, https://doi.org/10.1016/j.earscirev.2018.05.002, 2018.

Broeckx, J., Rossi, M., Lijnen, K., Campforts, B., Poesen, J., and Vanmaercke, M.: Landslide Mobilization Rates: A Global Analysis and Model, Earth-Science Reviews, 201, 102 972, https://doi.org/10.1016/j.earscirev.2019.102972, 2020.

525 Calvello, M. and Pecoraro, G.: FraneItalia: A Catalog of Recent Italian Landslides (Version 2.0), 2, https://doi.org/10.17632/zygb8jygrw.2, 2020.

Campforts, B., Vanacker, V., Herman, F., Vanmaercke, M., Schwanghart, W., Tenorio, G. E., Willems, P., and Govers, G.: Parameterization of River Incision Models Requires Accounting for Environmental Heterogeneity: Insights from the Tropical Andes, Earth Surface Dynamics, 8, 447–470, https://doi.org/10.5194/esurf-8-447-2020, 2020.

530 Crozier, M.: 7.26 Mass-Movement Hazards and Risks, in: Treatise on Geomorphology, pp. 249–258, Elsevier, https://doi.org/10.1016/B978-0-12-374739-6.00175-5, 2013.

De Lannoy, G. J. M.: Assimilation of Soil Moisture Observations into a Spatially Distributed Hydrologic Model, Ph.D. thesis, Ghent University, 2006.

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Arsenault, K. R., Verhoest, N. E. C., and Pauwels, V. R. N.: Satellite-Scale
535 Snow Water Equivalent Assimilation into a High-Resolution Land Surface Model, Journal of Hydrometeorology, 11, 352–369, https://doi.org/10.1175/2009JHM1192.1, 2010.

De Lannoy, G. J. M. D., Koster, R. D., Reichle, R. H., Mahanama, S. P. P., and Liu, Q.: An Updated Treatment of Soil Texture and Associated Hydraulic Properties in a Global Land Modeling System, Journal of Advances in Modeling Earth Systems, 6, 957–979, https://doi.org/10.1002/2014MS000330, 2014.

540 Depicker, A., Jacobs, L., Delvaux, D., Havenith, H.-B., Maki Mateso, J.-C., Govers, G., and Dewitte, O.: The Added Value of a Regional Landslide Susceptibility Assessment: The Western Branch of the East African Rift, Geomorphology, 353, 106 886, https://doi.org/10.1016/j.geomorph.2019.106886, 2020.

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki, N.: GSWP-2: Multimodel Analysis and Implications for Our Perception of the Land Surface, Bulletin of the American Meteorological Society, 87, 1381–1398, https://doi.org/10.1175/BAMS-87-10-1381, 2006.

545 Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S.: Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance, Ecography, 36, 27–46, https://doi.org/10.1111/j.1600-0587.2012.07348.x, 2013.

Emberson, R., Kirschbaum, D., Amatya, P., Tanyas, H., and Marc, O.: Insights from the Topographic Characteristics of a Large Global
550 Catalog of Rainfall-Induced Landslide Event Inventories, Preprint, Landslides and Debris Flows Hazards, https://doi.org/10.5194/nhess-2021-250, 2021.

Felsberg, A., De Lannoy, G. J. M., Girotto, M., Poesen, J., Reichle, R. H., and Stanley, T.: Global Soil Water Estimates as Landslide Predictor: The Effectiveness of SMOS, SMAP, and GRACE Observations, Land Surface Simulations, and Data Assimilation, Journal of Hydrometeorology, 22, 1065–1084, https://doi.org/10.1175/JHM-D-20-0228.1, 2021.

555 FSBIH: Federal State Budgetary Institution "Hydrospetzgeologiya": Archive of Quarter Annual Reports of Exogenous Geological Processes on Territories of the Russian Federation, http://geomonitoring.ru/arxiv.html, 2018.

Gaspari, G. and Cohn, S. E.: Construction of Correlation Functions in Two and Three Dimensions, p. 57, 1999.

Giardini, D., Grünthal, G., Shedlock, K., and Zhang, P.: The GSHAP Global Seismic Hazard Map, Lee, W., Kanamori, H., Jennings, P. and Kisslinger, C. (eds.): International Handbook of Earthquake & Engineering Seismology, International Geophysics Series 81 B, Academic 560 Press, Amsterdam, pp. 1233–1239, 2003.

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., and Ardizzone, F.: Probabilistic Landslide Hazard Assessment at the Basin Scale, Geomorphology, 72, 272–299, https://doi.org/10.1016/j.geomorph.2005.06.002, 2005.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., and Galli, M.: Estimating the Quality of Landslide Susceptibility Models, Geomorphology, 81, 166–184, https://doi.org/10.1016/j.geomorph.2006.04.007, 2006.

565 Hartmann, J. and Moosdorf, N.: The New Global Lithological Map Database GLiM: A Representation of Rock Properties at the Earth Surface, Geochemistry, Geophysics, Geosystems, 13, Q12 004, https://doi.org/10.1029/2012GC004370, 2012.

Hong, Y., Adler, R., and Huffman, G.: Use of Satellite Remote Sensing Data in the Mapping of Global Landslide Susceptibility, Natural Hazards; Dordrecht, 43, 245–256, https://doi.org/http://dx.doi.org/10.1007/s11069-006-9104-z, 2007.

Juang, C. S., Stanley, T. A., and Kirschbaum, D. B.: Using Citizen Science to Expand the Global Map of Landslides: Introducing the 570 Cooperative Open Online Landslide Repository (COOLR), PLOS ONE, 14, e0218 657, https://doi.org/10.1371/journal.pone.0218657, 2019.

Kalnay, E., Hunt, B., Ott, E., and Szunyogh, I.: Ensemble Forecasting and Data Assimilation: Two Problems with the Same Solution?, in: Predictability of Weather and Climate, pp. 157–180, Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9780511617652.008, 2006.

575 Kirschbaum, D., Stanley, T., and Zhou, Y.: Spatial and Temporal Analysis of a Global Landslide Catalog, Geomorphology, 249, 4–15, https://doi.org/10.1016/j.geomorph.2015.03.016, 2015.

Kirschbaum, D. B., Adler, R., Hong, Y., Hill, S., and Lerner-Lam, A.: A Global Landslide Catalog for Hazard Applications: Method, Results, and Limitations, Natural Hazards, 52, 561–575, https://doi.org/10.1007/s11069-009-9401-4, 2010.

Knevels, R., Petschko, H., Proske, H., Leopold, P., Maraun, D., and Brenning, A.: Event-Based Landslide Modeling in the Styrian Basin, 580 Austria: Accounting for Time-Varying Rainfall and Land Cover, Geosciences, 10, 217, https://doi.org/10.3390/geosciences10060217, 2020.

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., and Kumar, P.: A Catchment-Based Approach to Modeling Land Surface Processes in a General Circulation Model: 1. Model Structure, Journal of Geophysical Research: Atmospheres, 105, 24 809–24 822, https://doi.org/10.1029/2000JD900327, 2000.

585 Lima, P., Steger, S., and Glade, T.: Counteracting Flawed Landslide Data in Statistically Based Landslide Susceptibility Modelling for Very Large Areas: A National-Scale Assessment for Austria, Landslides, https://doi.org/10.1007/s10346-021-01693-7, 2021.

Lin, L., Lin, Q., and Wang, Y.: Landslide Susceptibility Mapping on a Global Scale Using the Method of Logistic Regression, Natural Hazards and Earth System Sciences, 17, 1411–1424, https://doi.org/10.5194/nhess-17-1411-2017, 2017.

Natural Hazards
and Earth System
Sciences
Discussions

Lin, Q., Lima, P., Steger, S., Glade, T., Jiang, T., Zhang, J., Liu, T., and Wang, Y.: National-Scale Data-Driven Rainfall Induced
Landslide Susceptibility Mapping for China by Accounting for Incomplete Landslide Data, Geoscience Frontiers, 12, 101 248,
https://doi.org/10.1016/j.gsf.2021.101248, 2021.

Lloyd, S.: Least Squares Quantization in PCM, IEEE Transactions on Information Theory, 28, 129–137,
https://doi.org/10.1109/TIT.1982.1056489, 1982.

Lucchese, L. V., de Oliveira, G. G., and Pedrollo, O. C.: Investigation of the Influence of Nonoccurrence Sampling on Landslide Susceptibility
Assessment Using Artificial Neural Networks, CATENA, 198, 105 067, https://doi.org/10.1016/j.catena.2020.105067, 2021.

Maes, J., Kervyn, M., de Hontheim, A., Dewitte, O., Jacobs, L., Mertens, K., Vanmaercke, M., Vranken, L., and Poesen, J.: Land-
slide Risk Reduction Measures: A Review of Practices and Challenges for the Tropics, Progress in Physical Geography, 41, 191–221,
https://doi.org/10.1177/0309133316689344, 2017.

Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J., and Schipper, A. M.: Global Patterns of Current and Future Road Infrastructure,
Environmental Research Letters, 13, 064 006, https://doi.org/10.1088/1748-9326/aabd42, 2018.

Nadim, F., Kjekstad, O., Peduzzi, P., Herold, C., and Jaedicke, C.: Global Landslide and Avalanche Hotspots, Landslides; Dordrecht, 3,
159–173, https://doi.org/http://dx.doi.org/10.1007/s10346-006-0036-1, 2006.

Nowicki Jessee, M. A., Hamburger, M. W., Allstadt, K., Wald, D. J., Robeson, S. M., Tanyas, H., Hearne, M., and Thompson, E. M.: A Global
Empirical Model for Near-Real-Time Assessment of Seismically Induced Landslides, Journal of Geophysical Research: Earth Surface,
123, 1835–1859, https://doi.org/10.1029/2017JF004494, 2018.

Petschko, H., Brenning, A., Bell, R., Goetz, J., and Glade, T.: Assessing the Quality of Landslide Susceptibility Maps – Case Study Lower
Austria, Nat. Hazards Earth Syst. Sci., 14, 95–118, https://doi.org/10.5194/nhess-14-95-2014, 2014.

Pourghasemi, H. R. and Rossi, M.: Landslide Susceptibility Modeling in a Landslide Prone Area in Mazandarn Province, North of
Iran: A Comparison between GLM, GAM, MARS, and M-AHP Methods, Theoretical and Applied Climatology, 130, 609–633,
https://doi.org/10.1007/s00704-016-1919-2, 2016.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https:
//www.R-project.org/, 2020.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F.: A Review of Statistically-Based Landslide Susceptibility Models,
Earth-Science Reviews, 180, 60–91, https://doi.org/10.1016/j.earscirev.2018.03.001, 2018.

Reichle, R. H., Liu, Q., Koster, R. D., Crow, W. T., De Lannoy, G. J. M., Kimball, J. S., Ardizzone, J. V., Bosch, D., Colliander, A., Cosh,
M., Kolassa, J., Mahanama, S. P., Prueger, J., Starks, P., and Walker, J. P.: Version 4 of the SMAP Level-4 Soil Moisture Algorithm and
Data Product, Journal of Advances in Modeling Earth Systems, 11, 3106–3130, https://doi.org/10.1029/2019MS001729, 2019.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller,
W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F.: Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical,
or Phylogenetic Structure, Ecography, 40, 913–929, https://doi.org/10.1111/ecog.02881, 2017.

Stanley, T. and Kirschbaum, D. B.: A Heuristic Approach to Global Landslide Susceptibility Mapping, Natural Hazards, 87, 145–164,
https://doi.org/10.1007/s11069-017-2757-y, 2017.

Stanley, T. A., Kirschbaum, D. B., Benz, G., Emberson, R. A., Amatya, P. M., Medwedeff, W., and Clark, M. K.: Data-Driven Landslide
Nowcasting at the Global Scale, Frontiers in Earth Science, 9, https://doi.org/10.3389/feart.2021.640043, 2021.

Steger, S. and Glade, T.: The Challenge of "Trivial Areas" in Statistical Landslide Susceptibility Modelling, in: Advancing Culture of Living
with Landslides, pp. 803–808, Springer, Cham, 2017.

Natural Hazards
and Earth System
Sciences
Discussions

Steger, S., Bell, R., Petschko, H., and Glade, T.: Evaluating the Effect of Modelling Methods and Landslide Inventories Used for Statistical Susceptibility Modelling, in: Engineering Geology for Society and Territory - Volume 2, pp. 201–204, Springer, Cham, https://doi.org/10.1007/978-3-319-09057-3_27, 2015.

630 Steger, S., Brenning, A., Bell, R., and Glade, T.: The Influence of Systematically Incomplete Shallow Landslide Inventories on Statistical Susceptibility Models and Suggestions for Improvements, Landslides, 14, 1767–1781, https://doi.org/10.1007/s10346-017-0820-0, 2017.

Steger, S., Schmaltz, E., and Glade, T.: The (f)Utility to Account for Pre-Failure Topography in Data-Driven Landslide Susceptibility Modelling, Geomorphology, 354, 107 041, https://doi.org/10.1016/j.geomorph.2020.107041, 2020.

Talagrand, O., Vautard, R., and Strauss, B.: Evaluation of Probabilistic Prediction Systems, 1997.

635 Van Den Eeckhaut, M., Hervás, J., Jaedicke, C., Malet, J.-P., Montanarella, L., and Nadim, F.: Statistical Modelling of Europe-Wide Landslide Susceptibility Using Limited Landslide Inventory Data, Landslides, 9, 357–369, https://doi.org/10.1007/s10346-011-0299-z, 2012.

Vanmaercke, M., Ardizzone, F., Rossi, M., and Guzzetti, F.: Exploring the Effects of Seismicity on Landslides and Catchment Sediment Yield: An Italian Case Study, Geomorphology, 278, 171–183, https://doi.org/10.1016/j.geomorph.2016.11.010, 2017.

Verdin, K.: Final Report High Resolution Topographic Analysis for GMAO's Catchment LSM, Tech. rep., Global Modeling and Assimilation
640 Office, NASA/Goddard Space Flight Center, Greenbelt, MD 201771, 2013.

Verdin, K. L., Godt, J., Funk, C., Pedreros, D., Worstell, B., and Verdin, J.: Development of a Global Slope Dataset for Estimation of Landslide Occurrence Resulting from Earthquakes, Open-File Report 2007-1188, Colorado: U.S. Geological Survey, Reston, Virginia, 2007.

Whiteley, J. S., Chambers, J. E., Uhlemann, S., Wilkinson, P. B., and Kendall, J. M.: Geophysical Monitoring of Moisture-Induced Landslides:
645 A Review, Reviews of Geophysics, 57, 106–145, https://doi.org/10.1029/2018RG000603, 2019.

Wilde, M., Günther, A., Reichenbach, P., Malet, J.-P., and Hervás, J.: Pan-European Landslide Susceptibility Mapping: ELSUS Version 2, Journal of Maps, 14, 97–104, https://doi.org/10.1080/17445647.2018.1432511, 2018.

Wilks, D. S.: Chapter 8 - Forecast Verification, in: International Geophysics, edited by Wilks, D. S., vol. 100 of *Statistical Methods in the Atmospheric Sciences*, pp. 301–394, Academic Press, https://doi.org/10.1016/B978-0-12-385022-5.00008-7, 2011.

650 Willmott, C. J. and Feddema, J. J.: A More Rational Climatic Moisture Index*, The Professional Geographer, 44, 84–88, https://doi.org/10.1111/j.0033-0124.1992.00084.x, 1992.

Zêzere, J. L., Pereira, S., Melo, R., Oliveira, S. C., and Garcia, R. A. C.: Mapping Landslide Susceptibility Using Data-Driven Methods, Science of The Total Environment, 589, 250–267, https://doi.org/10.1016/j.scitotenv.2017.02.188, 2017.

Zhu, J., Baise, L. G., and Thompson, E. M.: An Updated Geospatial Liquefaction Model for Global ApplicationAn Updated
655 Geospatial Liquefaction Model for Global Application, Bulletin of the Seismological Society of America, 107, 1365–1385, https://doi.org/10.1785/0120160198, 2017.

Zuur, A. F., ed.: Mixed Effects Models and Extensions in Ecology with R, Statistics for Biology and Health, Springer, New York, NY, 2009.