

Landslides (2021) 18:681–695
DOI 10.1007/s10346-020-01485-5
Received: 9 April 2020
Accepted: 6 July 2020
Published online: 24 August 2020
© The Author(s) 2020

Ugur Ozturk · Massimiliano Pittore · Robert Behling · Sigrid Roessner · Louis Andreani · Oliver Korup

How robust are landslide susceptibility estimates?

Abstract Much of contemporary landslide research is concerned with predicting and mapping susceptibility to slope failure. Many studies rely on generalised linear models with environmental predictors that are trained with data collected from within and outside of the margins of mapped landslides. Whether and how the performance of these models depends on sample size, location, or time remains largely untested. We address this question by exploring the sensitivity of a multivariate logistic regression—one of the most widely used susceptibility models—to data sampled from different portions of landslides in two independent inventories (i.e. a historic and a multi-temporal) covering parts of the eastern rim of the Fergana Basin, Kyrgyzstan. We find that considering only areas on lower parts of landslides, and hence most likely their deposits, can improve the model performance by >10% over the reference case that uses the entire landslide areas, especially for landslides of intermediate size. Hence, using landslide toe areas may suffice for this particular model and come in useful where landslide scars are vague or hidden in this part of Central Asia. The model performance marginally varied after progressively updating and adding more landslides data through time. We conclude that landslide susceptibility estimates for the study area remain largely insensitive to changes in data over about a decade. Spatial or temporal stratified sampling contributes only minor variations to model performance. Our findings call for more extensive testing of the concept of dynamic susceptibility and its interpretation in data-driven models, especially within the broader framework of landslide risk assessment under environmental and land-use change.

Keywords Landslide susceptibility · Logistic regression · Southern Kyrgyzstan · Landslide inventory · Remote sensing

Introduction

Most modern landslide susceptibility models aim to predict and map the degree to which landscapes are prone to slope failure. A common strategy is to use stacks of environmental data and to summarise these data with respect to whether they occur within or outside of mapped landslide areas. One key assumption in this approach is that these environmental predictors are sufficient proxies for the controls on slope instability (Galli et al. 2008; Korup 2008; Schulz et al. 2018). In susceptibility studies, these causes are routinely subsumed into a number of geological, topographic, climatic, hydrologic, and land-cover variables (Dou et al. 2015; Reichenbach et al. 2018). Generalised linear models use these variables as inputs and are designed to learn their relative weights in terms of dedicated coefficients (Braun et al. 2015; Provost et al. 2017). The predictive performance and accuracy of susceptibility studies may thus depend partly on the mapping quality of these training landslides (Kalantar et al. 2018). Yet few standards are available to warrant that landslide inventories are comparable (Corominas et al. 2013; Golovko et al. 2015), and even experts

may differ in their subjective assessment and mapping of landslides (Van Den Eeckhaut et al. 2005).

Regional landslide mapping efforts have been motivated by abrupt triggers such as strong earthquakes (Tanyaş et al. 2017); widespread landslide damage (Pittore et al. 2018); expanding our knowledge base by merging existing data (Havenith et al. 2015); and efforts of creating systematic multi-temporal landslide inventories using optical satellite time series data (Behling and Roessner 2017; Behling et al. 2016). Yet, the structure and level of detail in these inventories remain highly variable. Some landslide inventories feature only scar areas, because deposits may have been eroded or otherwise difficult to trace. Still, mapped scars may reveal details about mechanisms of slope failure and reactivation (Frodella et al. 2017). Some inventories may contain only landslide deposits, especially where scarps are unrecognisable, which can often be the case of older or prehistoric landslides. Yet other inventories feature the entire area affected by landslides to understand the total direct physical impact, but may not necessarily distinguish between depletion and accumulation areas (Cruden and Varnes 1996; Pánek et al. 2019).

The effect of these different landslide mapping choices on susceptibility models has remained largely unexplored. This choice may be important to aid decisions of whether we should consider the scar, body, toe, if not the entire landslide-affected, areas to contribute most to robust susceptibility models (Lai et al. 2019). Assuming that we have only limited information about the extent of each landslide, which parts of a landslide then can and should we include in susceptibility models? Considering only landslide source areas might reduce the performance of the models in the case of retrogressive landslides. Considering instead only landslide accumulation may ignore important aspects of the initial failure mechanism (Robinson et al. 2015). For example, highly mobile debris flows travel well beyond the length of their source area, whereas slow moving, deep-seated landslides with little displacement have very similar sized source and accumulation areas (Teshebaeva et al. 2019; Yamada et al. 2013). Asymmetric landslide footprint areas often reflect flow diversions that can also influence susceptibility assessments (Golovko et al. 2017a). The fringes of landslides may be more prone to reactivation than more central portions, given that scar and deposit margins are steeper than their surroundings and more exposed to groundwater seepage (Dang et al. 2016). Hence, landslide susceptibility may also change over time. Reactivated landslides in particular may affect susceptibility assessments over years to decades (Samia et al. 2017), and acknowledging this effect in models requires continuous monitoring. Multi-temporal satellite-based monitoring of landslide occurrences allows tracking whether and how susceptibility varies over time (Behling et al. 2016; Golovko et al. 2017b), though within limits of revisit period of the satellite, overall time period of data acquisitions, sensor resolution, and cloud cover. Systematic monitoring can record, for example, precursory slope failures that

may exhaust much of the soil cover at a given site and thus reduce the susceptibility to shallow landslides. Such precursory slope failures may also indicate later occurrence of larger slope failures initiated then by rare triggering conditions such as extreme precipitation. Time series of landslide surface characteristics derivable from satellite imagery can also capture how landslide-induced losses of vegetation cover may affect the local soil hydrology and thus alter slope stability. Quantifying the above-mentioned influences on a susceptibility model could aid more objective model estimates.

In this study, we address whether and how this variability in landslide susceptibility can be captured by multivariate logistic regression, a common model used for statistical inference. Multivariate logistic regression can predict local landslide susceptibility based on the weighted combination of predictors, choice of which may change drastically the regression coefficients and the overall performance (Formetta et al. 2016). However, instead of searching for the optimal predictor set, we check how models perform if the sampled predictor values cover only parts of landslides instead of their entirety. First, we test whether it matters to use only part of the data enclosed within mapped landslide margins. The rationale behind this approach is that incomplete mapping of landslides will in many cases only afford information about environmental predictors in parts of the scar, body, or deposits of slope failures. We simulate these constraints by a stratified sampling scheme and test how this scheme affects the predictive performance of a widely used landslide susceptibility model. We also test whether and how the model performance depends on landslide size besides the sampling location. Second, we test how thus derived susceptibility estimates change over time as more data become available. We start with an initial training set of landslides and keep adding newly mapped landslides as testing data to track the performance of the logistic regression model over time. Hence, we explore whether sample size that increases with time notably improves the predictive performance of the models, and whether this warrants ongoing, resource-intensive data collection.

Study area and data

The study area is situated in southern Kyrgyzstan along the eastern rim of the Fergana Basin in the Tien Shan mountains (Fig. 1). The rising topography of the eastern rim of the Fergana Basin receives higher precipitation from incoming westerlies, whereas ongoing tectonic deformation results from the overthrusting of the Pamir block on to the southern Tien Shan (Coutand et al. 2002; Reigber et al. 2001; Strecker et al. 2003; Zubovich et al. 2016). Both elevated precipitation and tectonic activity induce landslides mostly within weakly consolidated Mesozoic and Cenozoic sediments (Roessner et al. 2005; 2006; Wetzel et al. 2000). Hence, landslides have been responsible for more than half of all natural hazard-related deaths in our study area, similar to other regions in Central Asia, including Uzbekistan, and Tajikistan (Havenith et al. 2017).

We used two landslide inventories in our analyses. The first database was compiled in a multi-temporal landslide inventory consisting of landslide objects in form of polygons which were derived in a semi-automated way from time series of optical RapidEye satellite remote sensing data covering the time period between 2009 and 2017. Landslides were detected by analyzing temporal changes in the Normalized Difference Vegetation Index

(NDVI) and relief-oriented parameters (e.g. hillslope inclination, relief position) in a rule-based approach combining pixel- and object-based analyses. Landslides are often visible as discretely appearing and gradually closing gaps in the vegetation cover as a consequence of failure and regrowth. This vegetation regrowth is usually slower compared with vegetation growth around landslide patches, while the landslide-affected areas are prone to subsequent erosion and reactivation processes. These distinct dynamics allow separating fresh landslides from seasonal changes in vegetation or cropping patterns in agricultural areas. Automated analysis of these changes also enables distinguishing permanently bare areas (e.g. urban structures, water bodies, rock outcrops) from temporarily unvegetated landslide areas. Detailed descriptions of the landslide mapping methods can be found in Behling and Roessner (2017) and Behling et al. (2014; 2016). For this study, semi-automated landslide detection was performed in a retrospective way by processing the entire time series of available images. Automatically detected landslide objects were thoroughly validated by visual inspection using high-resolution imagery. Many of these landslide locations were also visited during repeated field investigations between 2012 and 2017 (Fig. 2d). The resulting inventory for this study comprises about 8500 landslides that occurred between 2009 and 2017 within the 120 km by 150 km study area (Fig. 1).

Landslide planforms in this multi-temporal inventory vary in size between 10 m² and 2 km² (Fig. 3). Besides the outline and areal coverage, each landslide is characterised by its time period of occurrence determined by the temporal resolution of the satellite imagery contained in the time series. This reference interval consists of the time between the last suitable image acquisition before the landslide occurred and the first suitable image acquisition after the landslide failure. Hence, this inventory also includes reactivation of landslides (Fig. 2), which have previously occurred during the observation period as long as those reactivations have resulted in a significant spectral contrast detectable by the used semi-automated analysis method (Fig. 3b). The annual reactivation rate increased from 3% in 2010 to ~15% in 2012, expressed in terms of the total annual landslides area. This trend persists at similar rates thereafter. The rate is as low as 3% at the start of the monitoring, whereas reactivated landslides predating the observation period (2009–2017)—e.g. from 2007 or 2008—cannot be captured. We refer to this multi-temporal landslide inventory derived by a validated semi-automated remote sensing-based approach (Behling et al. 2016) as ‘automatic’ and to the susceptibility models based on this inventory as ‘automatic models’.

The second landslide inventory, we use here, was compiled by Havenith et al. (2015) in form of a historic inventory originally covering the whole Tien Shan. This inventory comprises various data sources, e.g. remote sensing, field investigations, reports, and previously published inventories without a time stamp (e.g. Strom and Korup 2006; Schlögel et al. 2011). To allow a direct comparison between the two inventories, we extracted some 1100 landslides from this second inventory, which fall within our study area (Fig. 1). Landslide areas range from 1500 m² to 30 km² (Fig. 3). We refer to this subset of the second inventory as ‘manual’ and the susceptibility models using this inventory as ‘manual models’.

Logistic regression is a common tool for estimating landslide susceptibility, but hinges on a carefully selected set of landslide

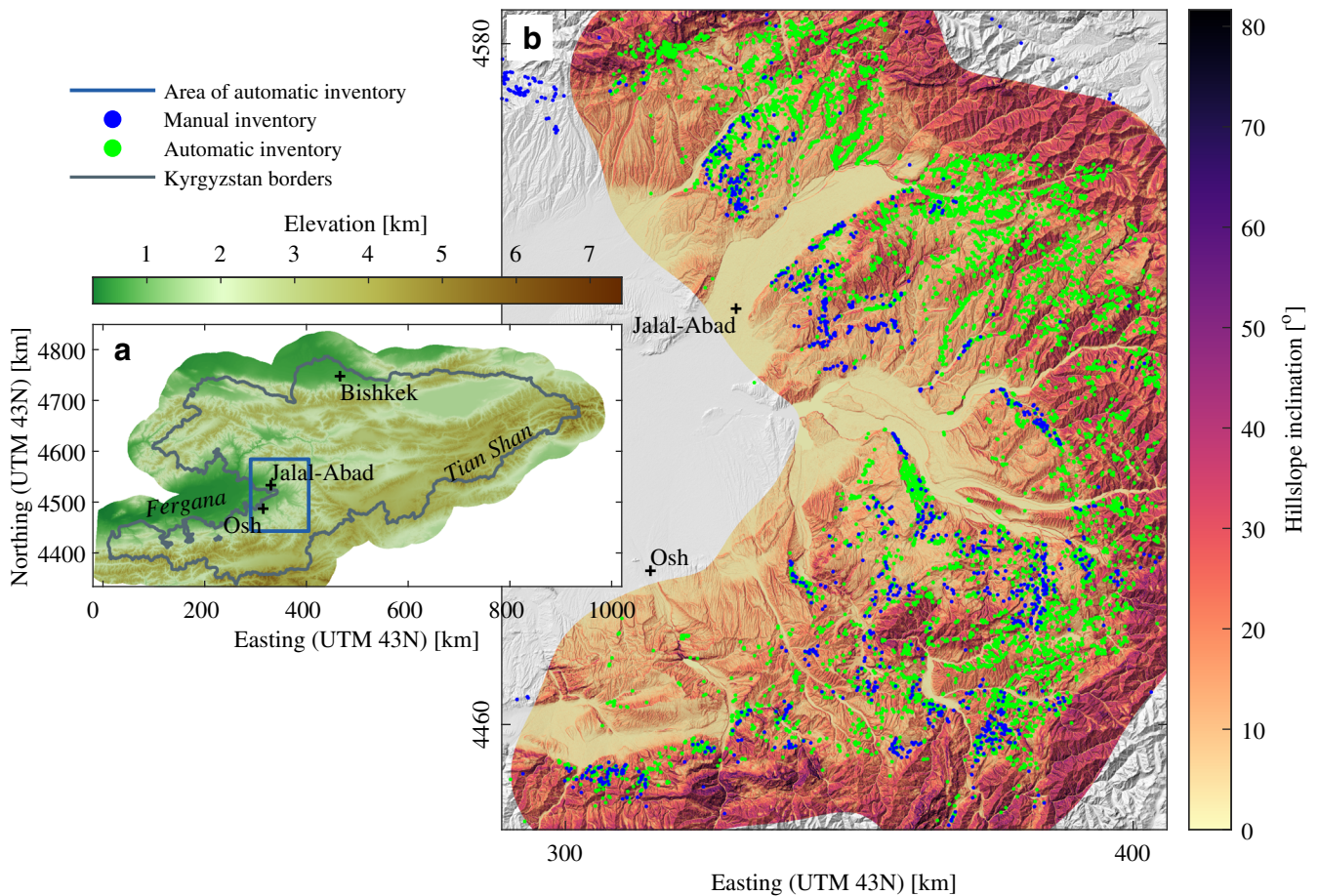


Fig. 1 Overview map of study area along the eastern rim of the Fergana Basin (a). Blue box in inset topographic map shows location of shaded hillslope inclination map (b). We use the multi-temporal inventory derived for this study from satellite remote sensing time series (automatic inventory) and the historic inventory (manual inventory) compiled by Havenith et al. (2015) as inputs to a multivariate logistic regression model of landslide susceptibility

predictors. Following recent recommendations (Liu et al. 2016; Othman et al. 2018; Reichenbach et al. 2018), we derived a set of geological, hydrological, land cover, and morphological predictors: (i) local hillslope inclination, aspect, topographic position index (TPI), and total curvature extracted using TopoToolbox (Schwanghart and Scherler 2014) from the 30-m resolution digital elevation model (DEM) of the Shuttle Radar Topographic Mission (SRTM; Farr et al. 2007); (ii) hydrological flow accumulation as a proxy for rainfall runoff, and distance from the road network based on the shortest flow direction (OpenStreetMap®; World Bank 2017), as roads can alter runoff and slope stability; (iii) satellite-based land cover indices including the NDVI, and indices of quartz and iron oxide content at the Earth surface; and (iv) nominal classes of the regional geology (geological periods; Saponaro et al. 2015) as proxies for rock types.

We used the spectral indices of quartz, iron oxide content, and NDVI in the model assuming that different soil types would have different combinations of index values. Quartz content is high in paleosols and loess sequences in which it can attain 40–65% (Gallet et al. 1996; Zheng et al. 1994). Iron oxides can help to discriminate between soil types as their iron-oxide concentration is closely related to environmental and weathering factors, such as rock mineralogy, soil moisture, temperature, or pH (Cornell and Schwertmann 2003;

Schwertmann and Taylor 1989). For the derivation of these indices, we used Sentinel-2 (S2) image data with a spatial resolution of 10 m to compute the NDVI using four Sentinel-2 L1C scenes obtained from the Copernicus Open Access Hub (<https://scihub.copernicus.eu>). The NDVI quantifies green vegetation and phenology (Rouse Jr. et al. 1974), normalising green leaf scattering in the near-infrared wavelength (S2-Band 8, 832 nm) and chlorophyll absorption in the red wavelength (Band 4, 664 nm). The spectral range, resolution, and revisit time of only 5 days of Sentinel-2 is well suited for these applications (Frampton et al. 2013; van der Meer et al. 2014). Our choice of acquisition period was also guided by preferring images with low cloud cover (<5%). Each scene was atmospherically corrected from top of atmosphere (TOA) to bottom of atmosphere (BOA) reflectance using the Sen2Cor algorithm in the Sentinel Application Platform toolbox (SNAP; Louis et al. 2016). Spectral bands were then resampled to a resolution of 20 m and a mosaic was created using ENVI® v5.1 software. We used the spectral angle mapper algorithm (SAM; Kruse et al. 1993) to measure the degree of similarity between spectra extracted from the Sentinel-2 mosaic and quartz spectra resampled to the resolution of Sentinel-2 from the United States Geological Survey (USGS) spectral library (Kokaly et al. 2017). We used the ratio between S2-Bands 4 (664 nm) and 2 (492 nm) as a proxy for iron oxides content (van der Werff and van der Meer 2015).

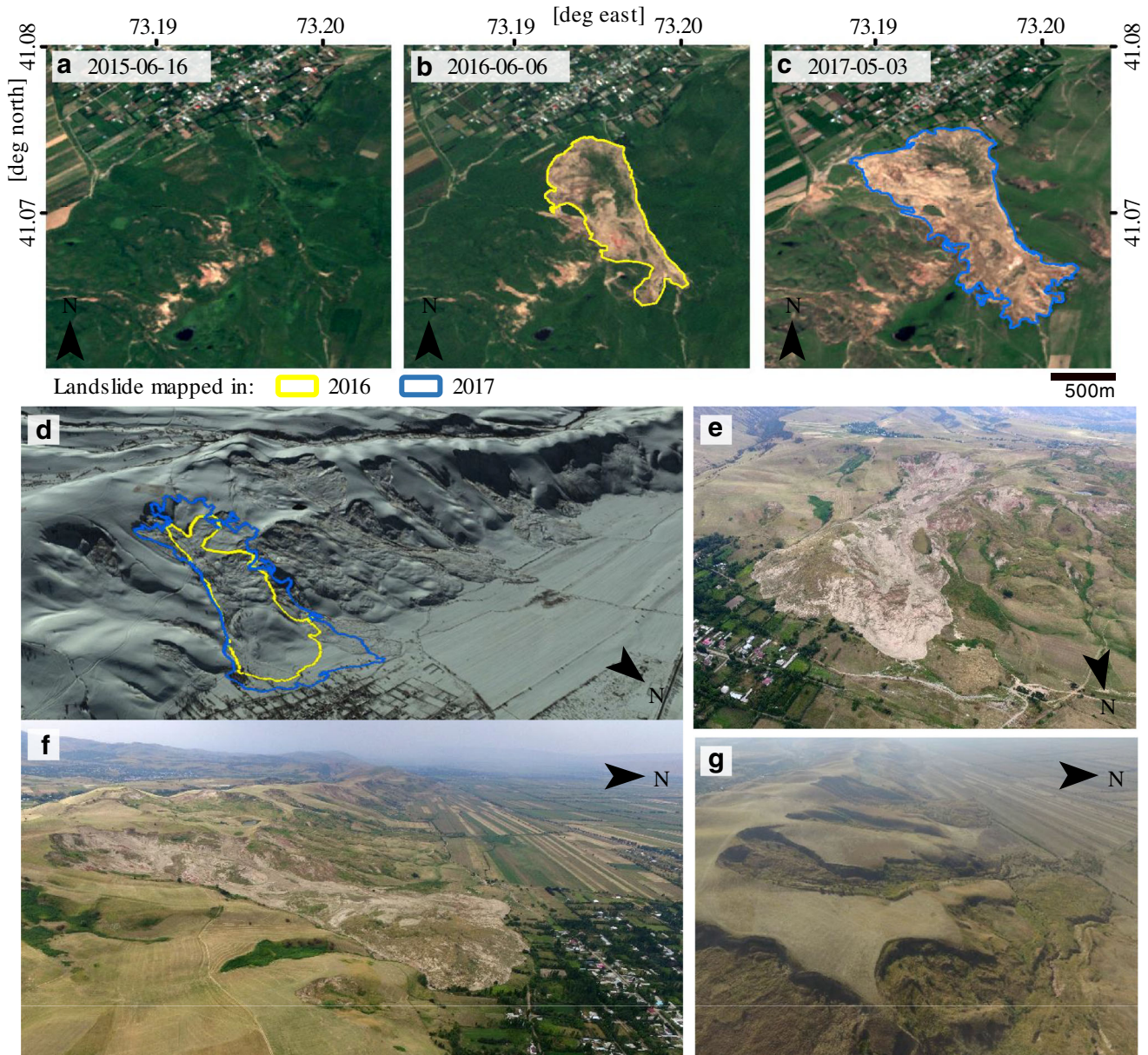


Fig. 2 Landslide sequence at the eastern rim of the Fergana Valley, Kyrgyzstan, and capture in the multi-temporal landslide inventory. (a–c) RapidEye images used for semi-automated mapping of the landslide outlines (yellow and blue polygons). (d) Mapped landslide outlines draped on Google Earth (GE) perspective view; acquisition of GE imagery 2013–11–22. (e–g) UAV images taken during field surveys in October 2016 (g) and August 2017 (e, f) show the landslide in 2016 and 2017 as part of a larger unstable hillslope; (g) older landslides to the west (predating our study) and fresh failures captured in (e, f)

We re-projected and re-sampled all data to the DEM grid resolution of 30 m for consistency (Table 1). Finally, we extracted the *Climate Hazards Group InfraRed Precipitation with Station Data* (CHIRPS version 2.0) that merge 0.05° resolution satellite with weather-station measurements (Funk et al. 2015); CHIRPS time series may help to demonstrate a possible seasonality of landslide occurrences.

Methods

We use multivariate logistic regression to estimate the probability that a given DEM grid cell is classified as a landslide. Modern susceptibility models based on random forests, support vector

machines, or artificial neural networks usually outperform logistic regression (Braun et al. 2015; Chang et al. 2019). Yet logistic regression allows a straightforward interpretation of its coefficients as the relative weights of the standardised predictors (Table 1). This opportunity for comparing the relative influence of each predictor is often lost in more sophisticated, black-box models (e.g. Martinović et al. 2016; Samia et al. 2018). Our model input consists of randomly sampled 5000 predictor values each from both within and outside of the landslide polygons separately for the two inventories. For each sample, we recorded the corresponding class label, i.e. ‘landslide’ or ‘no landslide’. We tested a number

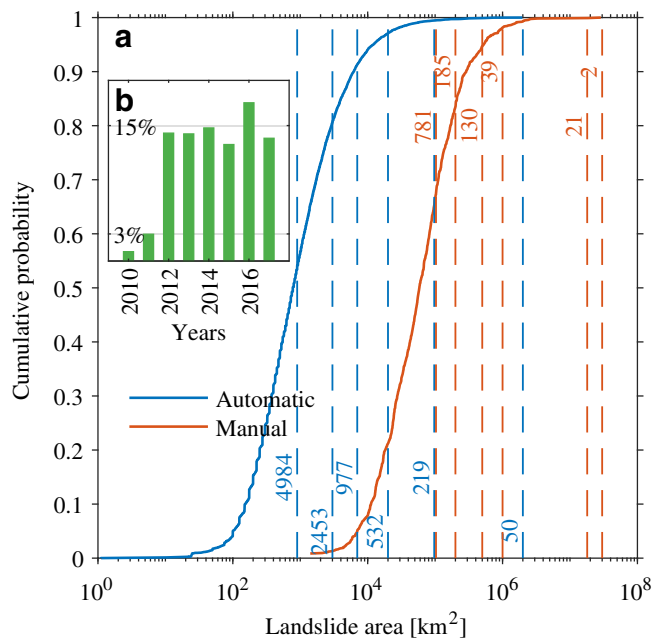


Fig. 3 (a) Empirical cumulative distribution functions of landslide footprints in the study area along the eastern rim of the Fergana Basin. Dashed lines delimit the six quantiles used for binning data in Fig. 9, and number of landslides in each bin is written on the left of the line. (b) Histogram of the annual reactivation rate of mapped landslides; percentages refer to the fraction of total landslide area reactivated

of sample variants that we divided into three groups. Two of these groups explore the sensitivity of the model to the sampling location of the predictors inside landslide margins (spatial sampling, Table 2). The third group investigates the sensitivity of the automatic model over time as increasing amount of data becomes available through newer satellite imagery (temporal sampling, Table 2). We express all performance changes with respect to a reference model that takes in all available landslide data. To estimate the predictive performance of our models, we compute

the area under the receiver operating characteristic curve (ROC-AUC) using a 10-fold cross-validation (Costache 2019). The ROC is a graphical illustration of the diagnostic ability of logistic regression. A ROC-AUC value of 0.5 identifies a purely random classifier, whereas a value of 1 is a perfect classification. We compare the average ROC-AUC of the models that of the base model and check whether the mean difference between the models exceeds the error bounds of the base model based on 10-fold cross-validation.

Spatial sampling

We test whether susceptibility to slope failure changes measurably when using predictors from either landslide scar or toe areas, while considering also landslide size. To this end, we partition the landslide inventories into two, four, and six quantiles of landslide area to maintain equal sample sizes. We then compute logistic regression models for discrete bins (quantiles) for several geomorphometric variables (Table 1). We sample predictor values according to their *normalised elevation* in the landslide to distinguish objectively between scar and toe areas (Fig. 4a), with normalised elevation ranging from 0 (landslide toe) to 1 (landslide crown).

We also test how sampling predictors from various *areal portion* and *relative centrality* affect the susceptibility model (Fig. 4b, c). An areal portion of 0 (1) means that a given cell is completely outside of (within) the landslide margins. A relative centrality of 0 (1) identifies a cell at a landslide margin (centre).

We define landslide *compactness* as the average angle between all vectors connecting the highest part of the crown with landslide grid cells and the vector defining the landslide long axis (Fig. 4d). The long axis of a landslide is defined here as the shortest path connecting the highest and lowest points within the landslide planform. We use this compactness to split the inventories into subsets of symmetric and asymmetric landslides that we discerned using different percentiles in compactness.

Landslide cells are randomly sampled 100 times for each combination of all these location metrics, leading to a total of $10 \times 10 \times$

Table 1 List of landslides predictors used in the model

Predictors	Unit	Resolution	Reprojected
Elevation	m	1" (WGS84)	30 m (UTM 43 N)
Hillslope inclination	°	30 m	-
Total curvature	m^{-2}	30 m	-
Flow accumulation	m^2	30 m	-
TPI	m	30 m	-
Aspect	°	30 m	-
Downhill distance from roads	m	30 m	-
Uphill distance from roads	m	30 m	-
Geology	-	30 m	-
NDVI	-	20 m	30 m (UTM 43 N)
Quartz index	-	20 m	30 m (UTM 43 N)
Iron oxides index	-	20 m	30 m (UTM 43 N)

Table 2 Tested sample variants, two are spatial and the last one is temporal sampling. c.v. indicates cross-validation, r.s. stands for random sampling

No.	Criteria	Bin type	Intervals	Sampling	
1	Relative elevation landslide area	Discrete	2 quantiles	$0^{-1}/_2-1$	10-fold c.v. 100 r.s.
			4 quantiles	$0^{-1}/_4-^2/_4-^3/_4-1$	
			6 quantiles	$0^{-1}/_6-^2/_6-^5/_6-1$	
2	Relative elevation areal portion	Cumulative	10 percentiles	0.1–1, 0.2–1...0.9–1	10-fold c.v. 100 r.s.
				0–1, 0.1–1...0.9–1	
2	Relative centrality compactness	Cumulative	9 percentiles	0–0.1, 0–0.2...0–1	10-fold c.v. 100 r.s.
				$0^{-1}/_{10}-1, 0^{-2}/_{10}-1...0^{-9}/_{10}-1$	
3	Temporal	Discrete	Monthly increments		10-fold c.v. 100 r.s.
				1.1.2009–31.12.2017	

$10 \times 9 \times 100 = 9 \times 10^5$ model runs. These 9×10^5 combinations are then repeated by switching crown with toe locations (Fig. 4a). By doing so, we also sample predictors in percentiles starting from the accumulation zone up to the depletion zone of each landslide. We use the Kolmogorov-Smirnov (KS) to test the statistically significant changes in model performance linked to the different sampling schemes from either crown to toe or the reverse. The KS test evaluates the difference between the empirical cumulative distribution functions of the model results for each sampling scheme.

Temporal sampling

We train a number of models at monthly steps determined by each image acquisition date t_i to investigate the effect of having more landslide data over time during a satellite-based monitoring campaign. We thus test predicted landslide susceptibility with younger landslides, i.e. those that postdate the image date t_i by less than a year, taking into account their uncertain dates. Each landslide age is in an interval defined by the date of the images showing the first landslide evidence (T_p), and the date of the previous available image (T_r). If t_i is within the age bracket of any landslide ($T_p < t_i < T_r$), we randomly place that landslide into training and (unobserved) testing data weighted by the time difference in days with respect to t_i . For example, for an image acquired on $t_i =$ April 15, and a known age bracket of a landslide between $T_p =$ April 12 and $T_r =$ April 22, we estimate a probability of $(15 - 12)(22 - 12) = 0.3$ that this landslide belongs to the training set (Fig. 5). We repeated this random placement 100 times for each t_i , together with a 10-fold cross-validation to obtain a total of 10^3 ROC-AUC estimates of model performance. We then repeat these steps by testing a total of $12 \times (2017 - 2009) = 96$ monthly updates of the model for all landslides (i.e. our reference model), thus covering nearly a decade of monitoring data.

Results

We keep the number of predictors fixed in our logistic regression models (for distributions see Fig. 7) and note that the individual weights vary considerably with the different sampling schemes. The weights fluctuate less in the automatic than in the manual model (Fig. 6). The estimated model performance in terms of the ROC-AUC is $78\% \pm 3\%$ ($68\% \pm 3\%$; 95% bootstrap interval) for the automatic (manual) model, if sampling predictor values from all grid cells within all training landslide polygons (Fig. 7). This model outcome refers to the reference model (Fig. 8a, b).

Spatial sampling

We first report the model sensitivity to sample predictors from quantiles of landslide size and relative elevation (Fig. 4a). We find that the performance of the automatic model increases $>9\%$ (from 78 to $>87\%$, Fig. 9), when sampling from the lower parts of landslides as well as from larger landslides. This trend is similar, but less distinct for the manual model, where sampling from smaller landslides slightly raises performance (from 68 to $>75\%$, Fig. 9b, d). When we use more quantiles, the manual model that samples larger landslides performs better (Fig. 9f).

Model performance changes less distinctly if we sample the predictor values cumulatively from either landslide head or toe (Fig. 10). Except for when we sample in percentiles of areal fraction and relative centrality (Figs. 4b, c and 10a, c), the ROC-AUC values generally remain unchanged for the automatic model.

Grid cells with greater overlap with landslides elevate the ROC-AUC of the automatic model by about 11% (Fig. 10a), whereas the performance of the manual models remains unchanged (Fig. 10b). The ROC-AUC also increases if including more central landslide cells in the automatic model, though hardly for the manual model (Fig. 10c, d). In all runs, the performance is on average higher for samples from the landslide accumulation zone; all the pairs are significantly different at the 5% significance level (Fig. 10e–h). The performance of the automatic model is largely insensitive to the sampling locations within landslides, and the trend is statistically insignificant (Fig. 10e, g). Nevertheless, the ROC-AUC of the manual model increases by $>10\%$, if sampling only from landslide toes (Fig. 10f, h), especially for asymmetric landslides (Fig. 10h). Sampling more towards the landslide crown consistently decreases the performance of the manual model, although the error bands overlap with those of the model using mainly toe samples.

Temporal sampling

The performance of the models changes only slightly with increasing availability of landslide data. The ROC-AUC of both the training and testing data roughly level out after the first year of landslide monitoring already when landslides affected some 0.01% of the study area (≈ 1000 cells, orange lines in Fig. 11b, c). The average ROC-AUC remains largely constant for both training and testing data until late 2016, though its uncertainty in terms of the 95% bootstrap interval marginally narrows with time. We also observe that the model performance increases above that of the

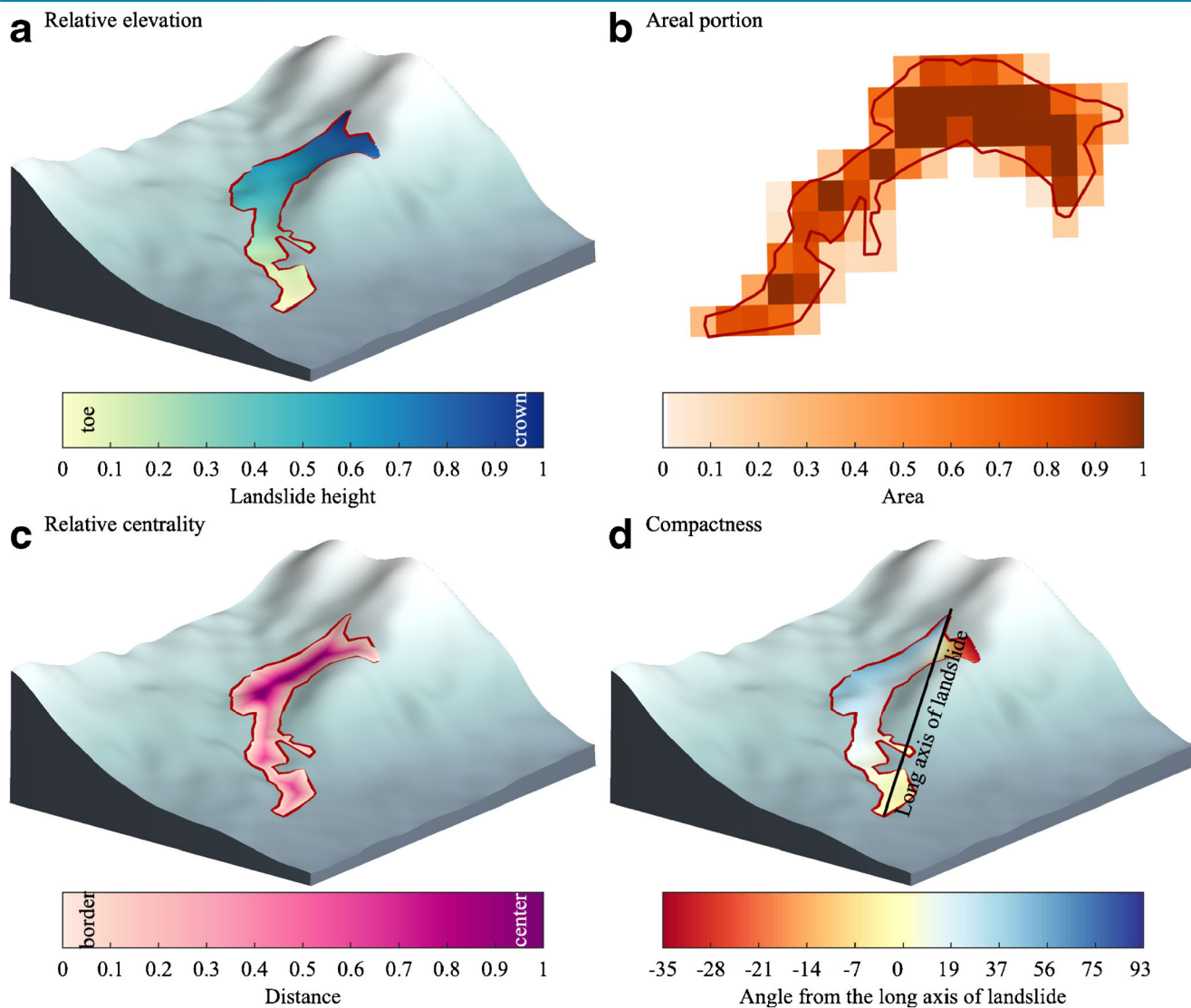


Fig. 4 Metrics of relative landslide location for simulating the quality of landslide mapping with respect to known and fixed margins. (a) Normalised elevation of grid cells within the landslide margins. (b) Fraction of grid-cell area within the landslide margins. (c) Normalised distance from landslide margins. (d) We also introduce compactness, expressed as the mean deviation of the angles between all vectors from landslide scar to grid cells and the vector defining the long axis of the landslide

training model for periods encompassing landslides that occurred within a year of t_i (blue lines in Fig. 11b, c). The error margins for the testing data from 2011 to 2014 are larger than those of other periods as the dates of those landslides are more uncertain (Fig. 11a). After 2016, landslide numbers increased threefold to fourfold compared with previous years, though most of those landslides were smaller than the DEM grid resolution of this analysis ($\leq 900 \text{ m}^2$ and shown in Fig. 11b). Upon removing these small landslides, the mean ROC-AUC increased by $\approx 3\%$ (orange lines in Fig. 11b and c) and approached that of the testing data by 2016, i.e. near the end of our monitoring period.

Discussion

Our main goal was to test how sensitive landslide susceptibility estimates are to the choice of where and when we sample predictor data from mapped landslides. Does it matter whether we include only characteristics from either landslide crowns or toes? Should

we sample across all landslide sizes evenly to achieve a predictive performance that is sufficiently representative? Do these susceptibility estimates change on a yearly or even monthly basis? What difference the choice of landslide catalog make in the final susceptibility estimate? Our results show that the performance of a simple logistic regression model, for a fixed choice of predictors, increases over the reference model when sampling predictor data from landslide-toe areas. This rather unconventional improvement could arise, while the deposits of many high mobile landslides such as debris flows might reach in scarce flat regions in the study area (Figs. 1b and 7b, c). Although the scarcity of planes could lead to spatial correlation of the predictors in those deposits, the homogeneous distribution of the landslides along the study area hinders such an impact. The improvement in terms of the ROC-AUC is about 10% and consistent for both automatic and manual models (Figs. 9 and 10). The individual predictor weights of the logistic regression remain largely unchanged in the

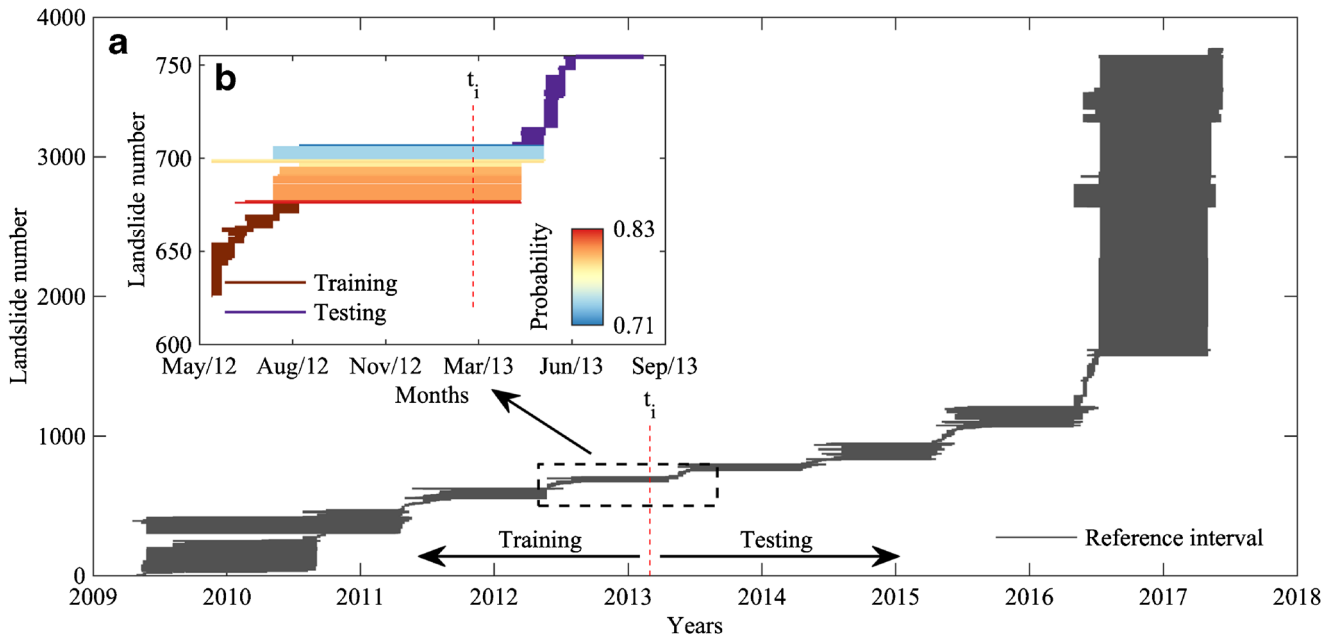


Fig. 5 Strategy of random sampling for checking updating effects on the landslide susceptibility model. (a) Estimated time intervals (dark grey horizontal lines) during which landslides >900 m² occurred, judging from the acquisition dates of satellite images (sample is from automatic inventory). Note the abundance of landslides in 2010 and 2017. Here t_i is an example date for which to determine the data into training and testing subsets (red dashed vertical line). (b) Color bar shows the probability of a landslide being sampled in the training subset based on the fraction of splitting of the age intervals of each landslide; see the ‘Study area and data’ section for more explanation

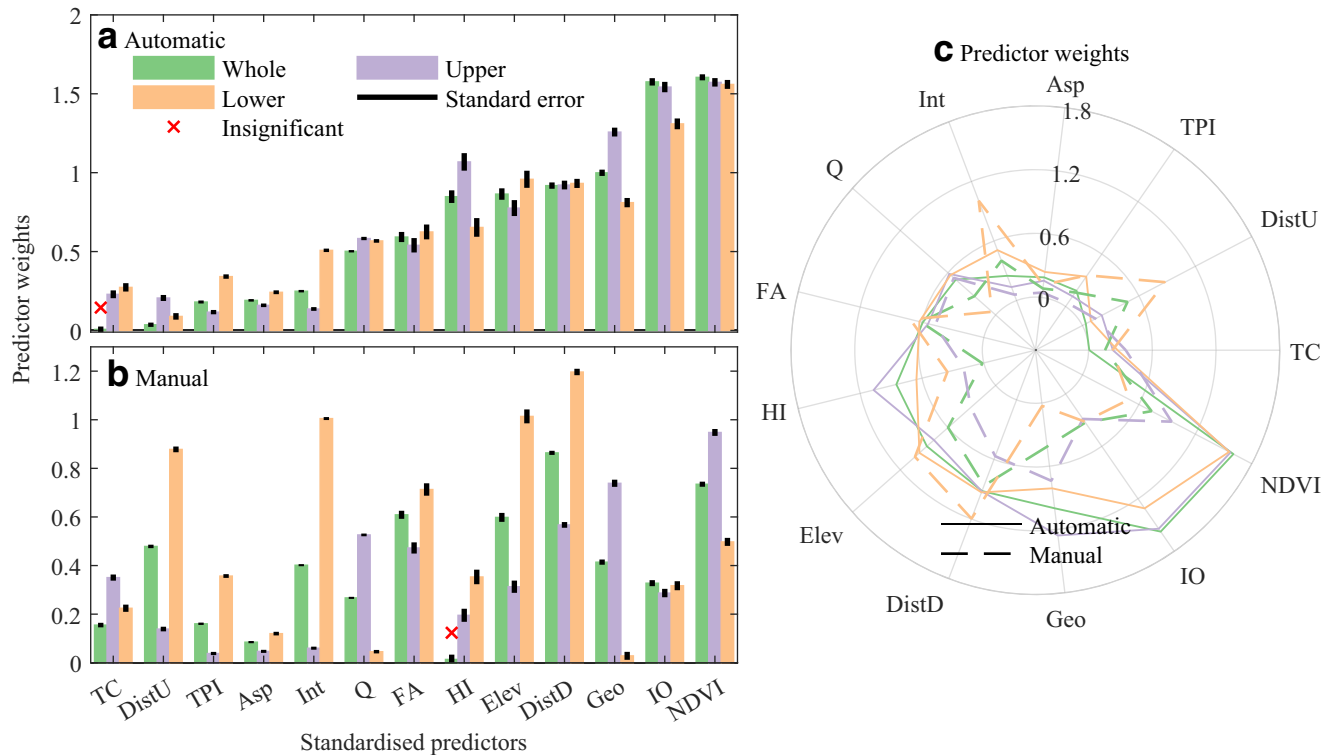


Fig. 6 Predictor weights of multivariate logistic regression models in Fig. 8 for the automatic (a) and manual (b) landslide inventories; (c) comparison between predictor weights by inventory. We split the data at the median landslide elevation to ensure balanced sample sizes, also approximate in depletion and accumulation zones. Red crosses mark inconclusive (statistically insignificant) weights. Int, interception; Elev, elevation; HI, hillslope inclination; TC, total curvature; FA, flow accumulation; TPI, topographic position index; Geo, geology; NDVI, Normalized Difference Vegetation Index; Q, quartz; IO, iron oxides; Asp, aspect; DistD, downhill distance from roads; DistU, uphill distance from roads

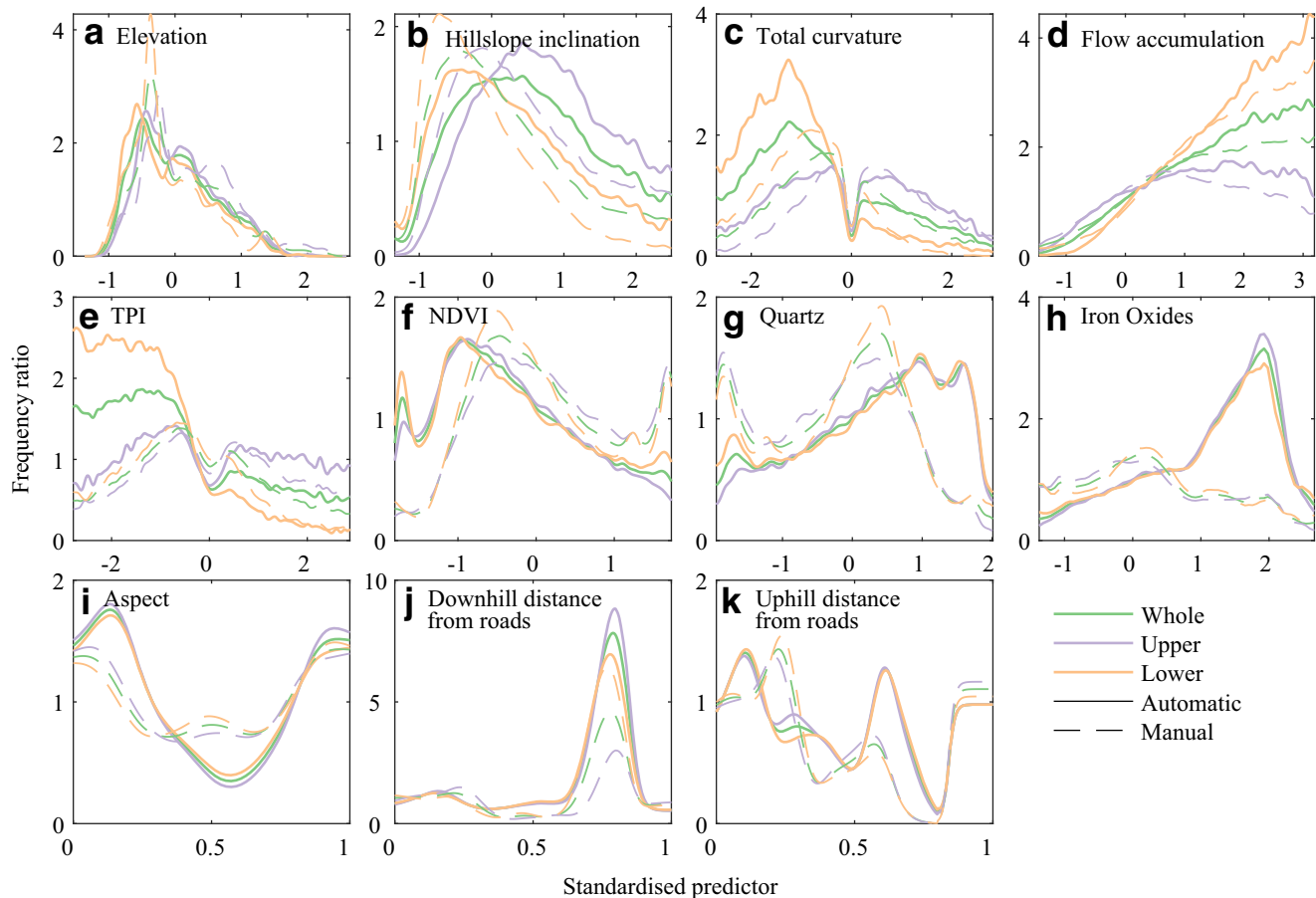


Fig. 7 Frequency ratio of the standardised predictors used in the logistic regression models in Figs. 6 and 8. Kernel densities of a predictor is divided by total density of the predictor within the study area. Smoothing bandwidth of all Kernel density estimates is 0.05

automatic model regardless of sampling strategy, but oscillate more strongly in the manual model (Fig. 6).

Involving landslide cells to the model in a way that all the landslides are represented covering the entire study should elevate model performance (Zêzere et al. 2017). This approach fundamentally aims to capture the diversity of the landslide predisposing factors. In our analyses, sampling from larger landslides ($>10^4$ m²) elevates the performance of the automatic model by $>10\%$ (Fig. 9a–e), as does considering only cells with an areal portion of 1, i.e. those being fully inside the landslide boundaries. The latter improvement is a consequence of landslide size, as nearly half of the landslides are ≤ 900 m² in the automatic inventory. Removing those smaller landslides also improves the performance of the model, most likely because they have an average area that approaches the resolution of the DEM (e.g. Golovko et al. 2017a). Sampling only central landslide portions that are well away from margins favours the influence of large landslides in the model and is consistent with the seemingly better performance of models that use only larger landslides (Fig. 9a–e). For the manual inventory, it is the sampling of preferably smaller (<10 km²) landslides that raised predictive performance, though only marginally by $>7\%$.

In only one case, where the manual model was trained by samples from very large landslides exclusively, i.e. >20 km² (2

landslides, Fig. 3), was the performance better (Fig. 9f). This improvement may come at the cost of high spatial autocorrelation between the predictors. Susceptibility models using data inputs by cell always carry the risk of this bias, since the landslide cells are treated as being statistically independent (Erener et al. 2010; Reichenbach et al. 2018). Irrespective of the different landslide size distributions in the two inventories (Fig. 3), sampling from landslides with areas between 10^4 and 10^6 m² raised the performance. In manual (automatic) inventory, about 1100 (800) are in this interval counting about 60% (30%) of the total landslide area (Fig. 3). Considering all the above-mentioned, we could synthesis that mapping small ≤ 900 m² or very large >10 km² landslides lowers the model performance (excluding the case where we suspect autocorrelation; Figs. 9 and 11). In essence, smaller landslides may fail to capture sufficiently the diversity of the factors that lead to slope instability, whereas the larger landslides tend to represent more broadly the range of values in the multivariate predictor space (Figures 3 and 9).

Increasing number of landslides that cover the entire study area should gradually elevate the landslide susceptibility estimates (Günther et al. 2014). On the contrary, the predictive performance for the automatic model (orange line in Fig. 11b, c) hardly changes as more landslide data become available over time, when we study whether and how landslide susceptibility changes through time.

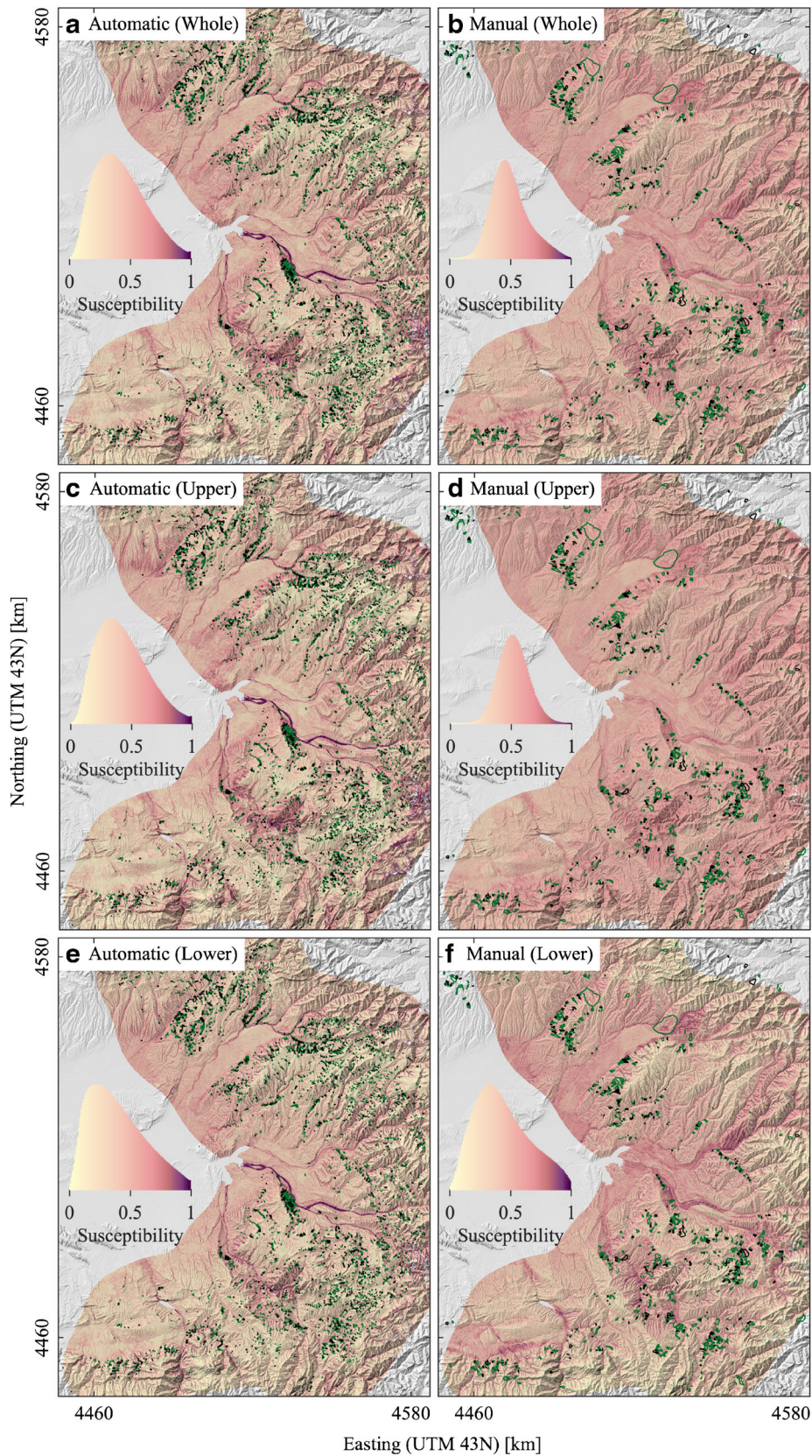


Fig. 8 Landslide susceptibility inferred from multivariate logistic regression. (a, b) The reference models trained on all available data in each inventory; (c, d) Based on the same model, though sampling only from above the median elevations of landslides; (e, f) Based on samples only from below the median elevations of landslides. Green (black) polygons are mapped landslides used to train (test) the models. See Fig. 9 for the predictive performance of these models. Density estimates show the susceptibility with the same colormap

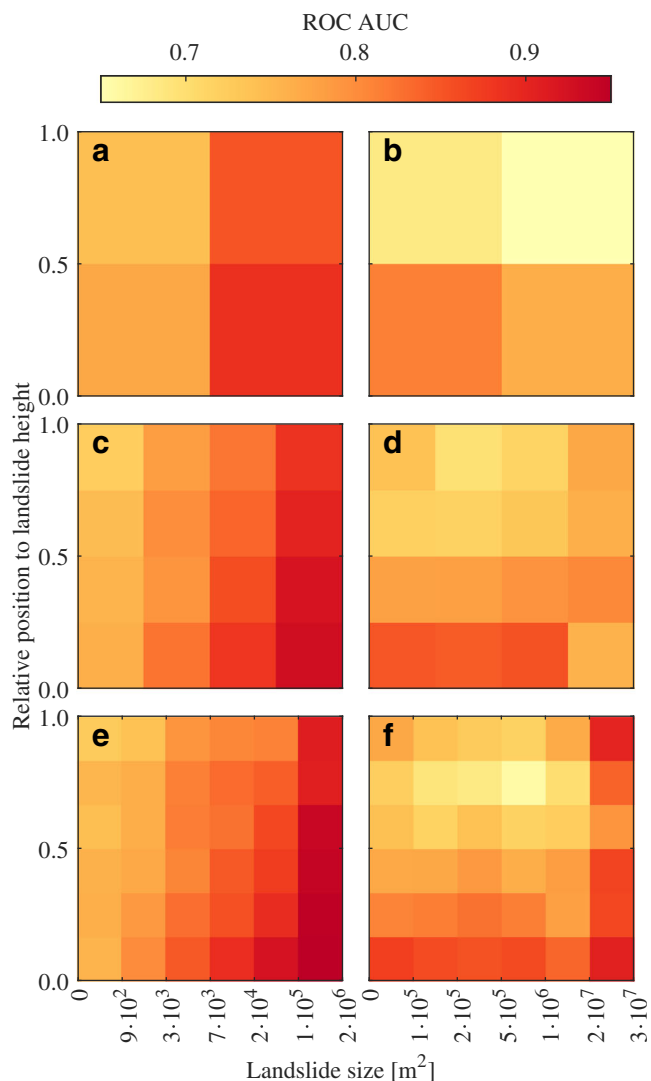


Fig. 9 Performance of landslide susceptibility model as a function of sampling data discretely from different quantiles of landslide portions and size (see Fig. 4a for the definition normalised elevation). Pixel color coded to area under the receiver operating characteristic curve (ROC-AUC); the number of landslide cells is equal for each pixel in each panel. Left (right) column shows results from the automatic (manual) model

This indicates a saturation limit, beyond which more data might not necessarily improve the model performance. In the Fergana Basin, we find that the performance stabilised roughly after 1000 landslide cells that cover 0.01% of our study area, given that these early landslides of 2009 and 2010 are somehow homogeneously distributed over the study area.

By dividing the landslide inventory to training and testing subsets at fixed dates, we can test repeatedly the performance of the model with previously unobserved sets of landslides. Our strategy differs from that in many landslide susceptibility studies that routinely extract training and testing subsets from their data by chance (e.g. Braun et al. 2015; Günther et al. 2014; Lombardo and Mai 2018). Random subsetting loses information on the timing of landslides and potential coeval triggers such as earthquakes or rainstorms. Such event-triggered landslides might tend to cluster

in certain topographic or lithologic niches leading to the overestimation of the performance of a susceptibility model (Steger et al. 2016). With temporal division of the training and testing subsets, we minimise the chance of event-triggered landslides to fall into both the subsets; thus, topographic or lithologic properties of the testing set are unknown to the trained model. This is reflected by the consistent drop of ROC-AUC (5% in Fig. 11b and 2% in Fig. 11c) between the training and testing samples between 2010 and 2017 (Fig. 11b, c), which might indicate that the diagnostic performance of a regular landslide susceptibility model for future landslides is 2–5% lower than the original model performance, which is estimated, e.g. via 10-fold cross-validation (Lombardo and Mai 2018).

The above-mentioned assumption—the model performance of testing model should be lower than the one of training model—failed in a particular period. Our model classified the testing subset unexpectedly better than the training subset between 2012 and 2016 when the testing landslides that occurred in subsequent years (blue lines in Fig. 11b, c). We suggest a few potential reasons to explain this performance mismatch. The ratio of training and testing subsets in 2012 (~5:1) increases until 2016. Such a mismatch in performance is absent, when the ratio imbalance was favouring the testing subset in the model that cover all the subsequent landslides (orange lines in Fig. 11b, c), likely ruling out the possibility that the sampling proportions distort the ROC-AUC estimate. The other reason is the reactivation of landslides that were used to train the models. We observe a fivefold increase in the reactivation rate in 2012 (Fig. 3b), which coincides with a drop in the total count and area of landslides (Figs. 5a and 11a). Besides this reactivation rate, postlandslides are claimed to increase the susceptibility for follow-up landslides in the subsequent years (Samia et al. 2017). Half of the landslides in 2012 occurred only within 300 m of an earlier landslide, which could be related to the legacy effect of the postlandslides. This distance gradually increases first to 450 m in 2013 and then to 700 m in 2014.

The reference interval to detect landslides increases towards the autumn and winter months due to cloud and snow cover (shaded blue, Fig. 11a). These uncertainties in landslide timing contribute to a higher bootstrapping variance, while estimating the model performance (shaded areas, Fig. 11b, c). We note a small decrease in the landslide area during the early summer months, most likely linked to higher rainfall and snowmelt around the rims of the Fergana Valley (Behling et al. 2014; Havenith et al. 2015).

By sampling predictor data from different portions of the landslides, we were able to alter the model performance by about 10% compared with a reference model that uses all the available data. Although the maximum ROC-AUC is achieved when sampling predictors predominantly from the toe areas of the landslides, we would like to state that the improvement in ROC-AUC does not necessarily lead to a better landslide susceptibility model (Zêzere et al. 2017). Many of the improvements remain marginal (<5%) in spatial tests indicating the statistical robustness of the susceptibility model to sampling locations. The choice of the landslide inventory seems to be the main determinant of the final susceptibility performance, via landslide area distribution and the coverage over the study area (Günther et al. 2014). Also in the temporal sampling, we witness that the model improvement after a short time period is little to none. Our temporal tests point out another issue about the correct assessment of a susceptibility

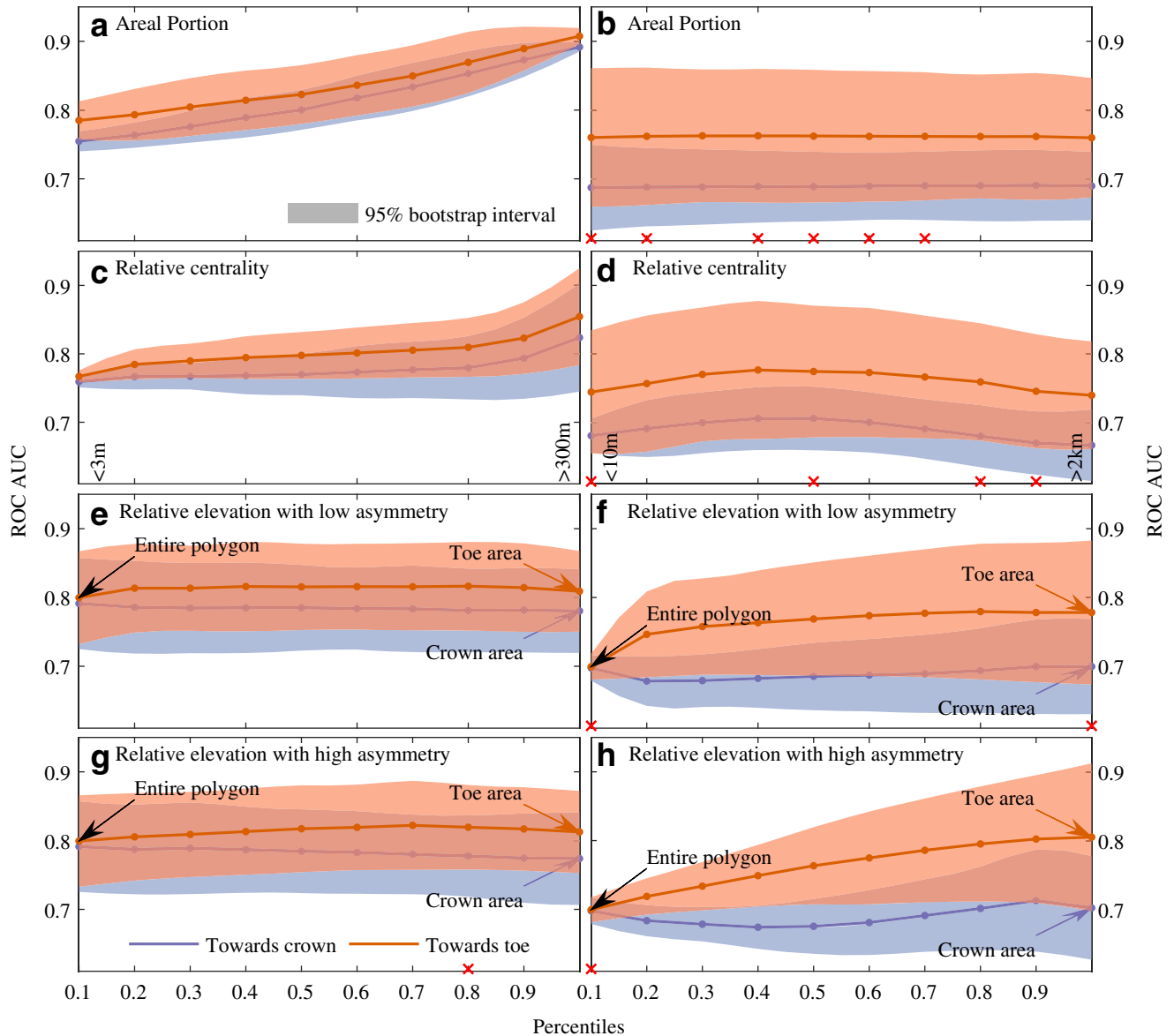


Fig. 10 Performance of a landslide susceptibility model and how it depends on sampling data from different percentiles of landslide portions. Each panel shows one metric of landslide position while varying randomly all others shown in Fig. 4, ensuring that sample size is equal. Left (right) panels show results from the automatic (manual) model. Solid lines are mean values of ROC-AUC, computed for some 5000 landslide cells. Shaded areas are 95% bootstrap intervals. Blue (orange) tones show models that sample data from above (below) a given percentile. Red crosses mark model variants that are indistinguishable judging from a KS test ($p = 0.01$). Relative landslide position is defined in Fig. 4a

model performance due to fair division of testing and training subsets of the landslide inventory (Lombardo and Mai 2018): traditional methods for this division based on chance might lead to a slight overestimation of performance.

Conclusion

Multivariate logistic regression has been widely used to predict the likelihood for a given (pixel) location to be involved in a landslide. We applied this method to a study area at the eastern rim of the Fergana Basin, Kyrgyzstan, considering a set of morphological, geological, and land cover proxies as predictors. Specifically, we tested how the model performance changes with different pixel

sampling strategies according to landslide position, shape, and size. We also explored the sensitivity of predictions as the number of landslides for training the model increases with time. Our findings suggest that, where detailed data are unavailable for this highly landslide-prone region, training the model using pixels from landslide deposits might suffice for predicting terrain susceptible to slope failure (Figs. 9 and 10). The model performs consistently better when predictors are sampled from landslide toes, regardless of different methods of landslide mapping that we tested systematically for two inventories. Landslide size distribution can alter predictions significantly. We obtain a better performance when sampling data from intermediate-sized ($10^4 - 10^6 \text{ m}^2$)

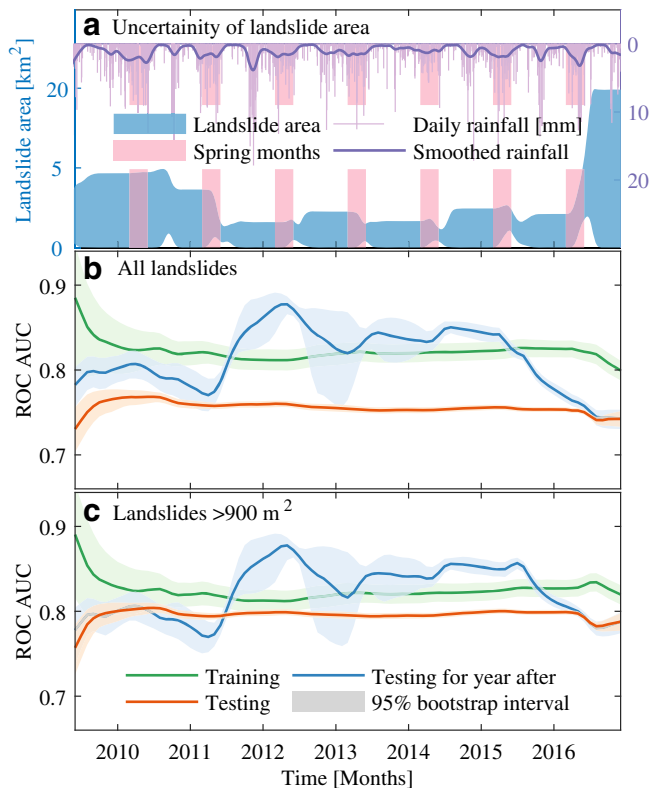


Fig. 11 Landslide susceptibility as function of time, based on the automatic inventory. (a) New total landslide area estimated from landslide-age brackets in 30-day intervals together with daily rainfall averaged over the study area. Rainfall is smoothed with a Gaussian kernel ($\sigma = 45$ days; Funk et al. 2015). (b, c) predictive performance of the model with increasing training data over time; ‘testing’ refers to new landslides that were unobserved in previous runs of the model; (b) uses all landslides, (c) excludes landslides that are smaller than a grid cell of the digital elevation model. The orange line is the predictive performance using landslides postdating t_i ; in this case, more training data become available over time, while the testing data continuously decreases. Whereas blue line is predictive performance only for landslides that occurred within a year after t_i , amount of testing data varies while training data increases over time

landslides, regardless of whether the inventory was derived manually or automatically. We infer that intermediate-sized landslides reflect the diversity of the slope instability factors better than smaller or larger landslides.

Finally, the increases in predictive performance through time are modest because the first year of observation already offered a large number of training data (Fig. 11). In this context, we find that temporal sampling could reduce biases in assessing the performance of a susceptibility model. Our results motivate revisiting the concept of static susceptibility and its interpretation in statistical models. Yet even the best performing models may fail to capture the underlying physical processes or give insight into the possible landslide release mechanism. This perspective might contrast with requirements of site-specific engineering but may be useful in regional landslide risk assessment.

Acknowledgements

We thank Hans-Balder Havenith for providing and explaining the manually derived landslide inventory, which is available as part of the Central Asia Fault Database from [esdynamics.geo.uni-](http://esdynamics.geo.uni-tuebingen.de/faults/)

[tuebingen.de/faults/](http://esdynamics.geo.uni-tuebingen.de/faults/) (Mohadjer et al. 2016). SRTM DEM can be downloaded from <http://dwtkns.com/srtm30m/> and Sentinel-2 data from <https://www.esa.int/>. We are grateful to Kevin Fleming and Fabrice Cotton for reviewing and helping us to improve the paper, and thank Vera Ozturk for helping in visualisation.

Funding information

Open Access funding provided by Projekt DEAL. Research presented in this paper has been financially supported by the Federal Ministry of Education and Research of Germany (BMBF) within the project CLIENT II – CaTeNA (FKZ o3Go878A).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Behling R, Roessner S (2017) Spatiotemporal landslide mapper for large areas using optical satellite time series data. In Matjaz Mikos, Binod Tiwari, Yueping Yin, and Kyoji Sassa, editors, *Advancing culture of living with landslides*, pages 143–152. Springer International Publishing, Cham. ISBN 978-3-319-53497-8 978-3-319-53498-5. <https://doi.org/10.1007/978-3-319-53498-5>. http://link.springer.com/10.1007/978-3-319-53498-5_17.
- Behling R, Roessner S, Segl K, Kleinschmit B, Kaufmann H (2014) Robust automated image co-registration of optical multi-sensor time series data: database generation for multi-temporal landslide detection. *Remote Sens* 6(3):2572–2600. <https://doi.org/10.3390/rs6032572>
- Behling R, Roessner S, Golovko D, Kleinschmit B (2016) Derivation of long-term spatiotemporal landslide activity—a multi-sensor time series approach. *Remote Sens Environ* 186:88–104. <https://doi.org/10.1016/j.rse.2016.07.017>
- Braun A, Fernandez-Steegeer T, Havenith H-B, Torgoev A (2015) Landslide susceptibility mapping with data mining methods—a case study from Maily-Say, Kyrgyzstan. In Giorgio Lollino, Daniele Giordan, Giovanni B. Crosta, Jordi Corominas, Rafiq Azzam, Janusz Wasowski, and Nicola Sciarra, editors, *Engineering geology for society and territory - Volume 2*, pages 995–998. Springer International Publishing, ISBN 978-3-319-09057-3
- Chang K-T, Merghadi A, Yunus AP, Pham BT, Dou J (2019) Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. *Sci Rep* 9(1):12296. <https://doi.org/10.1038/s41598-019-48773-2>
- Cornell RM, Schwertmann U (2003) *The iron oxides: structure, properties, reactions, occurrences and uses*. Wiley, 1 edition. ISBN 978-3-527-30274-1 978-3-527-60209-4. <https://doi.org/3527602097/3527602097>. <https://onlinelibrary.wiley.com/doi/book/10.1002/3527602097>
- Corominas J, van Westen C, Frattini P, Cascini L, Malet J-P, Fotopoulou S, Catani F, Van Den Eeckhaut M, Mavrouli O, Agliardi F, Pitalakis K, Winter MG, Pastor M, Ferlisi S, Tofani V, Hervás J, Smith JT (2013) Recommendations for the quantitative analysis of landslide risk. *Bull Eng Geol Environ* 1435–9537. <https://doi.org/10.1007/s10064-013-0538-8> <http://link.springer.com/10.1007/s10064-013-0538-8>
- Costache R (2019) Flash-flood Potential Index mapping using weights of evidence, decision trees models and their novel hybrid integration. *Stoch Env Res Risk A* 330(7):1375–1402. <https://doi.org/10.1007/s00477-019-01689-9>

- Coutand I, Strecker MR, Arrowsmith JR, Hillel G, Thiede RC, Korjenkov A, Omuraliev M (2002) Late Cenozoic tectonic development of the intramontane Alai Valley, (Pamir-Tien Shan region, central Asia): an example of intracontinental deformation due to the Indo-Eurasia collision: Cenozoic development of the intramontane Alai Basin. *Tectonics* 21(6):3–13–19. <https://doi.org/10.1029/2002TC001358>. <http://doi.wiley.com/10.1029/2002TC001358>
- Cruden DM, Varnes DJ (1996) Landslide types and processes. In *Landslides: investigation and mitigation*, number 247 in Special Report (National Research Council (U.S.) Transportation Research Board). 247:36–75 National Academy Press, Washington, D.C. ISBN 0-309-06208-X
- Dang K, Sassa K, Fukuoka H, Sakai N, Sato Y, Takara K, Quang LH, Loi DH, Van Tien P, Duc Ha N (2016) Mechanism of two rapid and long-runout landslides in the 16 April 2016 Kumamoto earthquake using a ring-shear apparatus and computer simulation (LS-RAPID). *Landslides* 13(6):1525–1534. <https://doi.org/10.1007/s10346-016-0748-9>
- Dou J, Bui DT, Yunus AP, Jia K, Song X, Revhaug I, Xia H, Zhu Z (2015) Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata. Japan. *PLoS One* 10(7):e0133262. <https://doi.org/10.1371/journal.pone.0133262>
- Erener A, Sebnem H, Düzgün B (2010) Improvement of statistical landslide susceptibility mapping by using spatial and global regression methods in the case of More and Romsdal (Norway). *Landslides* 7(1):55–68. <https://doi.org/10.1007/s10346-009-0188-x>
- Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L, Seal D, Shaffer S, Shimada J, Umland J, Werner M, Oskin M, Burbank D, Alsdorf D (2007) The Shuttle Radar Topography Mission. *Rev Geophys* 45(2):RG2004. <https://doi.org/10.1029/2005RG000183>. <http://doi.wiley.com/10.1029/2005RG000183>
- Formetta G, Capparelli G, Versace P (2016) Evaluating performance of simplified physically based models for shallow landslide susceptibility. *Hydrol Earth Syst Sci* 20(11):4585–4603. <https://doi.org/10.5194/hess-20-4585-2016>
- Frampton WJ, Dash J, Watmough G, Milton EJ (2013) Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J Photogramm Remote Sens* 82:83–92. <https://doi.org/10.1016/j.isprsjprs.2013.04.007>
- Frodella W, Salvatici T, Pazzi V, Morelli S, Fanti R (2017) GB-InSAR monitoring of slope deformations in a mountainous area affected by debris flow events. *Nat Hazards Earth Syst Sci* 17(10):1779–1793. <https://doi.org/10.5194/nhess-17-1779-2017>
- Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A, Michaelsen J (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci Data* 2(1):150066. <https://doi.org/10.1038/sdata.2015.66>
- Gallet S, Jahn B-m, Torii M (1996) Geochemical characterization of the Luochuan loess-paleosol sequence, China, and paleoclimatic implications. *Chem Geol* 133(1–4):67–88. [https://doi.org/10.1016/S0009-2541\(96\)00070-8](https://doi.org/10.1016/S0009-2541(96)00070-8)
- Galli M, Ardizzone F, Cardinali M, Guzzetti F, Reichenbach P (2008) Comparing landslide inventory maps. *Geomorphology* 94(3–4):268–289. <https://doi.org/10.1016/j.geomorph.2006.09.023>
- Golovko D, Roessner S, Behling R, Wetzel H-U, Kleinschmit B (2015) Development of multi-temporal landslide inventory information system for Southern Kyrgyzstan using GIS and satellite remote sensing. *Photogrammetrie - Fernerkundung - Geoinformation* 2015(2):157–172. <https://doi.org/10.1127/pfg/2015/0261>
- Golovko D, Roessner S, Behling R, Kleinschmit B (2017) Automated derivation and spatio-temporal analysis of landslide properties in southern Kyrgyzstan. *Nat Hazards* 85(3):1461–1488. <https://doi.org/10.1007/s11069-016-2636-y>
- Golovko D, Roessner S, Behling R, Wetzel H-U, Kleinschmit B (2017) Evaluation of remote-sensing-based landslide inventories for hazard assessment in southern Kyrgyzstan. *Remote Sens* 9(9):943. <https://doi.org/10.3390/rs9090943>
- Günther A, Van Den Eeckhaut M, Malet J-P, Reichenbach P, Hervás J (2014) Climate-physiographically differentiated Pan-European landslide susceptibility assessment using spatial multi-criteria evaluation and transnational landslide information. *Geomorphology* 224:69–85. <https://doi.org/10.1016/j.geomorph.2014.07.011>
- Havenith HB, Strom A, Torgoev I, Torgoev A, Lamair L, Ischuk A, Abdrakmatov K, Shan T (2015) Geohazards database: earthquakes and landslides. *Geomorphology* 249:16–31. <https://doi.org/10.1016/j.geomorph.2015.01.037>
- Havenith HB, Umaraliev R, Schlögel R, Torgoev I (2017) Past and potential future socioeconomic impacts of environmental hazards in Kyrgyzstan. *Curr Polit Econ Northern Western Asia* 26(2/3):179–215
- Kalantar B, Pradhan B, Naghibi SA, Motevalli A, Mansor S (2018) Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomatics Nat Hazards Risk* 9(1):49–69. <https://doi.org/10.1080/19475705.2017.1407368>
- Kokaly RF, Clark RN, Swayze GA, Livo KE, Hoefen TM, Pearson NC, Wise RA, Benzel WM, Lowers HA, and Driscoll RL (2017) *USGS Spectral Library: US geological survey data series 1035*. 61, 7 edition. <https://doi.org/10.3133/ds1035>
- Korup O (2008) Rock type leaves topographic signature in landslide-dominated mountain ranges. *Geophys Res Lett* 35(11):L11402. <https://doi.org/10.1029/2008GL034157>
- Kruse FA, Lefkoff AB, Boardman JW, Heidebrecht KB, Shapiro AT, Barloon PJ, Goetz AFH (1993) The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sens Environ* 44(2–3):145–163. [https://doi.org/10.1016/0034-4257\(93\)90013-N](https://doi.org/10.1016/0034-4257(93)90013-N)
- Lai J-S, Chiang S-H, Tsai F (2019) Exploring influence of sampling strategies on event-based landslide susceptibility modeling. *ISPRS Int J Geo-Inf* 8(9):397. <https://doi.org/10.3390/ijgi8090397>
- Liu CC, Luo W, Chen MC, Lin YT, Wen HL (2016) A new region-based preparatory factor for landslide susceptibility models: the total flux. *Landslides* 13(5):1049–1056. <https://doi.org/10.1007/s10346-015-0620-3>
- Lombardo L, Mai PM (2018) Presenting logistic regression-based landslide susceptibility results. *Eng Geol* 244:14–24. <https://doi.org/10.1016/j.enggeo.2018.07.019>
- Louis J, Debaecker V, Pflug B, Main-Knorn M, Bieniarz J, Mueller-Wilm U, Cadau E, Gascon F (2016) Sentinel-2 sen2cor: L2a processor for users. In *Proceedings of the Living Planet Symposium, Prague, Czech Republic*, pages 9–13
- Martinović K, Gavin K, Reale C (2016) Development of a landslide susceptibility assessment for a rail network. *Eng Geol* 215:1–9. <https://doi.org/10.1016/j.enggeo.2016.10.011>
- Mohadjer S, Ehlers TA, Bendick R, Stübner K, Strube T (2016) A quaternary fault database for central Asia. *Nat Hazards Earth Syst Sci* 16(2):529–542. <https://doi.org/10.5194/nhess-16-529-2016>
- Othman AA, Glogauer R, Andreani L, Rahnama M (2018) Improving landslide susceptibility mapping using morphometric features in the Mawat area, Kurdistan Region, NE Iraq: comparison of different statistical models. *Geomorphology* 319:147–160. <https://doi.org/10.1016/j.geomorph.2018.07.018>
- Pánek T, Břežný M, Kapustová V, Lenart J, Chalupa V (2019) Large landslides and deep-seated gravitational slope deformations in the Czech Flysch Carpathians: New LiDAR-based inventory. *Geomorphology* 106852. <https://doi.org/10.1016/j.geomorph.2019.106852> <https://linkinghub.elsevier.com/retrieve/pii/S0169555X19303332>
- Pittore M, Ozturk U, Moldobekov B, Saponaro A (2018) EMCA Landslide catalog Central Asia. <https://doi.org/10.5880/GFZ.2.6.2018.004> c:3657915. type: dataset
- Provost F, Hibert C, Malet J-P (2017) Automatic classification of endogenous landslide seismicity using the random forest supervised classifier: seismic sources automatic classification. *Geophys Res Lett* 44(1):113–120. <https://doi.org/10.1002/2016GL070709>
- Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. *Earth Sci Rev* 180:60–91. <https://doi.org/10.1016/j.earscirev.2018.03.001>
- Reigber C, Michel GW, Galas R, Angermann D, Klotz J, Chen JY, Papschev A, Arslanov R, Tzurkov VE (1910) Ishanov MC (2001) New space geodetic constraints on the distribution of deformation in Central Asia. *Earth Planet Sci Lett* 1–2:157–165. [https://doi.org/10.1016/S0012-821X\(01\)00414-9](https://doi.org/10.1016/S0012-821X(01)00414-9)
- Robinson TR, Davies TRH, Reznichenko NV, De Pascale GP (2015) The extremely long-runout Komansu rock avalanche in the Trans Alai range, Pamir Mountains, southern Kyrgyzstan. *Landslides* 12(3):523–535. <https://doi.org/10.1007/s10346-014-0492-y>
- Roessner S, Wetzel H-U, Kaufmann H, Sarnagoev A (2005) Potential of satellite remote sensing and GIS for landslide hazard assessment in southern Kyrgyzstan (Central Asia). *Nat Hazards* 35(3):395–416. <https://doi.org/10.1007/s11069-004-1799-0>
- Roessner S, Wetzel H-U, Kaufmann H, Sarnagoev A (2006) Satellite remote sensing and GIS for analysis of mass movements with potential for dam formation. *Italian J Eng Geol Environ Special* 0(2006):103–114. ISSN 2035-5688. <https://doi.org/10.4408/IJEGE.2006-01.S-14>
- Rouse JW Jr, Haas RH, Schell JA, Deering DW (1974) Monitoring vegetation systems in the Great Plains with ERTS. 1:309–317
- Samia J, Temme A, Bregt A, Wallinga J, Guzzetti F, Ardizzone F, Rossi M (2017) Do landslides follow landslides? Insights in path dependency from a multi-temporal landslide inventory. *Landslides* 14(2):547–558. <https://doi.org/10.1007/s10346-016-0739-x>
- Samia J, Temme A, Bregt AK, Wallinga J, Stuiver J, Guzzetti F, Ardizzone F, Rossi M (2018) Implementing landslide path dependency in landslide susceptibility modelling. *Landslides* 15(11):2129–2144. <https://doi.org/10.1007/s10346-018-1024-y>
- Saponaro A, Pilz M, Wieland M, Bindi D, Moldobekov B, Parolai S (2015) Landslide susceptibility analysis in data-scarce regions: the case of Kyrgyzstan. *Bull Eng Geol Environ* 74(4):1117–1136. <https://doi.org/10.1007/s10064-014-0709-2>

- Schlögel R, Torgoev I, De Marneffe C, Havenith H-B (2011) Evidence of a changing size-frequency distribution of landslides in the Kyrgyz Tien Shan, Central Asia: landslide activity in the Kyrgyz Tien Shan. *Earth Surf Process Landf* 36(12):1658–1669. <https://doi.org/10.1002/esp.2184>
- Schulz WH, Smith JB, Wang G, Jiang Y, Roering JJ (2018) Clayey landslide initiation and acceleration strongly modulated by soil swelling. *Geophys Res Lett* 45(4):1888–1896. <https://doi.org/10.1002/2017GL076807>
- Schwanghart W, Scherler D (2014) Short communication: TopoToolbox 2—MATLAB-based software for topographic analysis and modeling in Earth surface sciences. *Earth Surf Dynam* 1(20):1–27. <https://doi.org/10.5194/esurf-2-1-2014>
- Schwertmann U, Taylor RM (1989) Iron oxides. In *Minerals in soil environments*, number 1 in SSSA Book Series, pages 379–438. Soil Science Society of America, 1989. ISBN 978-0-89118-860-5. <https://doi.org/10.2136/sssabookser1.2ed.c8> <https://dl.sciencesocieties.org/publications/books/abstracts/sssabookseries/mineralsinsoile/379>
- Steger S, Brenning A, Bell R, Petschko H, Glade T (2016) Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. *Geomorphology* 262:8–23. <https://doi.org/10.1016/j.geomorph.2016.03.015>
- Strecker MR, Hillel GE, Arrowsmith JR, Coutand I (2003) Differential structural and geomorphic mountain-front evolution in an active continental collision zone: the northwest Pamir, southern Kyrgyzstan. *GSA Bull* 115(2):166–181. ISSN 0016-7606. [https://doi.org/10.1130/0016-7606\(2003\)115<0166:DSAGMF>2.0.CO;2](https://doi.org/10.1130/0016-7606(2003)115<0166:DSAGMF>2.0.CO;2)
- Strom AL, Korup O (2006) Extremely large rockslides and rock avalanches in the Tien Shan Mountains, Kyrgyzstan. *Landslides* 3(2):125–136. <https://doi.org/10.1007/s10346-005-0027-7>
- Tanyaş H, van Westen CJ, Allstadt KE, Jessee MAN, Görüm T, Jibson RW, Godt JW, Sato HP, Schmitt RG, Marc O, Hovius N (2017) Presentation and analysis of a worldwide database of earthquake-induced landslide inventories: earthquake-induced landslide inventories. *J Geophys Res Earth Surf* 122(10):1991–2015. <https://doi.org/10.1002/2017JF004236>
- Teshebaeva K, Ehtler H, Bookhagen B, Strecker M (2019) Deep-seated gravitational slope deformation (DSGSD) and slow-moving landslides in the southern Tien Shan Mountains: new insights from InSAR, tectonic and geomorphic analysis. *Earth Surf Process Landf* 44(12):2333–2348. <https://doi.org/10.1002/esp.4648>
- Van Den Eeckhaut M, Poesen J, Verstraeten G, Vanacker V, Moeyersons J, Nyssen J, van Beek LPH (2005) The effectiveness of hillshade maps and expert knowledge in mapping old deep-seated landslides. *Geomorphology* 67(3–4):351–363. <https://doi.org/10.1016/j.geomorph.2004.11.001>
- van der Meer FD, van der Werff HMA, van Ruitenbeek FJA (2014) Potential of ESA's Sentinel-2 for geological applications. *Remote Sens Environ* 148:124–133. <https://doi.org/10.1016/j.rse.2014.03.022>
- van der Werff H, van der Meer F (2015) Sentinel-2 for mapping iron absorption feature parameters. *Remote Sens* 7(10):12635–12653. <https://doi.org/10.3390/rs71012635>
- Wetzel H-U, Roessner S, Sarnagoev A (2000) Remote sensing and CIS based geological mapping for assessment of landslide hazard in Southern Kyrgyzstan (Central Asia). *WIT Trans Inf Commun Technol* 24:12. <https://doi.org/10.2495/MIS000341> ISBN: 1853128155 Publisher: WIT Press
- World Bank (2017) Kyrgyz Republic - measuring seismic risk. Working Paper AUS0000061, World Bank Group, Washington, D.C. URL <http://documents.worldbank.org/curated/en/911711517034006882/Kyrgyz-Republic-Measuring-seismic-risk>
- Yamada M, Kumagai H, Matsushi Y, Matsuzawa T (2013) Dynamic landslide processes revealed by broadband seismic records: dynamic landslide processes. *Geophys Res Lett* 40(12):2998–3002. <https://doi.org/10.1002/grl.50437>
- Zêzere JL, Pereira S, Melo R, Oliveira SC, Garcia RAC (2017) Mapping landslide susceptibility using data-driven methods. *Sci Total Environ* 589:250–267. <https://doi.org/10.1016/j.scitotenv.2017.02.188>
- Zheng H, Theng BKG, Whitton JS (1994) Mineral composition of Loess-Paleosol samples from the Loess Plateau of China and its environmental significance. *Chin J Geochem* 13(1):61–72. <https://doi.org/10.1007/BF02870857>
- Zubovich A, Schöne T, Metzger S, Mosienko O, Mukhamediev S, Sharshbaeva A, Zech C (2016) Tectonic interaction between the Pamir and Tien Shan observed by GPS: Pamir-Tien Shan interaction by GPS. *Tectonics* 35(2):283–292. <https://doi.org/10.1002/2015TC004055>

U. Ozturk (✉) · **M. Pittore** · **R. Behling** · **S. Roessner**

Helmholtz Centre Potsdam-GFZ German Research Centre for Geosciences,
14473, Potsdam, Germany
Email: oetztuerk@gfz-potsdam.de

L. Andreani

Helmholtz Centre Dresden-Rossendorf,
Helmholtz Institute Freiberg for Resource Technology,
09599, Freiberg, Germany

O. Korup

Institute of Environmental Science and Geography & Institute of Geosciences,
University of Potsdam,
14476, Potsdam, Germany