

Color scheme

- Reviewer comments: blue
- Answers to reviewer (i.e. author comments: AC): black

RC: Dear Authors, dear Editor,

In the paper, the authors demonstrate a statistical data-driven model (generalised additive mixed model) for modelling landslide triggering conditions at a regional scale, i.e., in South Tyrol. The content of the paper is comprehensive, generally well written and interesting for the landslide community. However, before final acceptance, the manuscript needs to be improved by additional clarifications and revisions, which are listed below.

AC: We thank Reviewer #2 for the valuable feedback and the constructive suggestions. We are pleased that our manuscript was described as “*comprehensive, generally well written and interesting for the landslide community.*” Please find our point-to-point replies below.

RC: - Line 114: one of the goals of the paper is to identify critical seasonal precipitation conditions using GAMM, but throughout the paper these critical parameters are mentioned only very sporadically, so they are not apparent to the reader. I propose to create a table listing all critical parameters, both triggering and preparatory, along with thresholds and metrics from ROC.

AC: The reviewer is correct that the identification of critical season-dependent precipitation conditions for shallow landslides was one of the main goals of this research. The parameters used in the model are already highlighted within several parts of the methods section (i.e. from a more general description within Section 4 and Figure 3 to a more detailed technical elaboration within Section 4.2, Section 4.3 and Table 1).

As described in the results section (Line 491f), the highest landslide probability scores “*were predicted for situations in which high P and high T occur simultaneously*”. This continuous relationship is visualized within Figure 8a in the form of a prediction surface. Within our model setup, the amount of precipitation required to exceed a threshold is always dependent on a combination of three variables, namely short-term precipitation T, antecedent preparatory precipitation P and seasonality (DOY). This relationship is visualized within Figure 9 and the Animation GIF (supplementary material). From our viewpoint, the creation of a concise table showing the critical amount of precipitation required to induce a landslide is not feasible, particularly due to the non-trivial interplay of the three variables T, P and DOY (i.e. a large number of T-P-DOY combinations refer to a threshold). This is why we would prefer to use plots (i.e. probability spaces) and the Animation to visualize the elaborated critical conditions.

We agree that adding a table for the thresholds and metrics from ROC will further enhance transparency. Thus, we will add the following Table 2 to Section 5.2:

Table 1 Probability thresholds and associated ROC-based metrics.

Threshold name	Probability	True positive rate (TPR)	True negative rate (TNR)
TPR95	0.04	95% (552 out of 581 presences)	40% (900 out of 2251 absences)
Optimal	0.13	79% (460 out of 581 presences)	81% (1823 out of 2251 absences)
TNR95	0.41	56% (325 out of 581 presences)	95% (2138 out of 2251 absences)

RC: -Line 165: it is not enough to provide only a web link as a reference for the IdroGeo platform. Please add some paper reference(s) (e.g Iadanza et al., 2021)

AC: Thank you for your suggestions. We will add the reference Iadanza et al., 2021 to the respective section in the manuscript:

Iadanza, C., Trigila, A., Starace, P., Dragoni, A., Biondo, T., and Roccisano, M.: IdroGEO: A Collaborative Web Mapping Application Based on REST API Services and Open Data on Landslides and Floods in Italy, ISPRS International Journal of Geo-Information, 10, 89, <https://doi.org/10.3390/ijgi10020089>, 2021.

RC: -Line 176: please explain which precipitation product is considered? It is clear that you have considered rainfall, but once you also mention snow and snowmelt, but it is not clear what exactly you have considered in the models under the term precipitation.

AC: As pointed out in the manuscript (Line 176f), the “*gridded fields of daily precipitation for South Tyrol were extracted from the 1980-2018 dataset produced by Crespi et al. (2021)*”. This data was obtained by (Line 178f) “*(...) interpolating the rain-gauge daily records from a quality-checked and homogenized archive including around 80 station sites from the weather station network of South Tyrol*”. Most rain gauges in South Tyrol are heated so that measured precipitation stems from both liquid and solid precipitation. This is why we opted to use the more general term “precipitation” instead of “rainfall” when describing our analyses and results within the paper. To enhance transparency, we will add the following sentence to this section:

“Precipitation measurements are primarily related to rainfall, but in Winter, the amount of precipitation recorded may also include snowfall.”

As mentioned in the discussion (Line 631f), the topic of snow melting directly refers to our interpretation of the seasonal effect, since we assume that “*the elaborated seasonal variation is primarily attributable to effects related to vegetation cover, and to a lower degree also to temperature and snow melting, rather than to delayed response of a hillslope to antecedent precipitation.*”

RC: -Line 243: missing reference to the criteria for landslides on the basis of which you made the selection

AC: The criteria used to select suitable “shallow landslide data” were not based on a specific publication, but defined specifically for the purpose of this analysis in collaboration with the entity that manages/collects the South Tyrolean landslide data (cf. acknowledgments: Provincial Office for Geology). The following description highlights the adopted criteria (Line 248f):

“Prior to the analyses, suitable landslide records were selected based on the filter criteria “movement-type”, “material-type”, “movement-cause” and “date-availability”. Only translational and rotational movement-types associated with the assigned material-type “earth” and “debris” were included. In this context, we explicitly removed slide-types associated to deep-seated gravitational slope deformations. Furthermore, only landslides with the assigned causes “short-intense precipitation” or “prolonged precipitation” were selected. The resulting 1822 entries were then filtered according to an additional temporal criterion: only entries with reliable information on the day of occurrence, and from January 2000 to the end of 2020 were selected, resulting in a sample of 676 landslide records.”

Within the revised manuscript, we will add new text to the results section 5.1 to highlight the associated numbers (as requested by reviewer 1). This may further enhance traceability of the landslide data selection procedure.

“The initial 11420 IFFI points were reduced to 2714 entries by first filtering translational and rotational slide-types and by excluding deep-seated movements. Most of these entries ($n = 2319$) were selected according to the subsequent material type filter, i.e. only “earth” or “debris” slides were considered. Further data subsampling according to the assigned movement cause “precipitation-induced” led to a subsample consisting of 1822 landslides. Out of these 1822 entries, 676 landslides were associated with reliable day information while occurring between January 2000 and the end of 2020.”

RC: -Line 255-260: the number of training landslides for presence and absence days are unclear. Please improve this part with additional information and also edit Fig. 1a, where you have to distinguish training and validation landslides, for example with different colours.

AC: Within a previous NHSS publication, we elaborated advantages of k-fold cross-validation (CV) and k-fold spatial cross-validation (SCV) in comparison to conventional hold-out validation:

Steger, S., Brenning, A., Bell, R., Glade, T., 2016a. The propagation of inventory-based positional errors into statistical landslide susceptibility models. Nat. Hazards Earth Syst. Sci. 16, 2729–2745. <https://doi.org/10.5194/nhess-16-2729-2016>.

In this paper we highlighted that *“In contrast to single hold-out validation, CV and SCV are not based on one single split of the training and test sample (e.g. 80 % for calibration and 20 % for validation), but on a repeated partitioning of the original sample into k subsamples. In each iteration, a performance measure (e.g. AUROC) is estimated for one of the k subsamples, while*

the remaining (k-1) subsamples are combined into a training set that is used to calibrate the model. Thus, validation results that are based on CV and SCV are not dependent on one specific sample split.” This is why “CV as well as SCV allow that all available data can be used to validate and to calibrate the final models.”

We therefore did not distinguish between training and validation landslides in our text or in the Figure, since each landslide is in fact both, a training landslide (within certain data partitions) and a validation landslide (within other data partitions). The ratio of training to test samples also varies between the different applied CV procedures (e.g. Leave-one-of-10-clusters out: 9:1; Leave-one-of-25-clusters out: 24:1; Leave-one-month-out: 11:1). However, we would prefer not to further expand on the description of the cross-validation procedure, as these techniques are established in the data-driven modeling community and we have already highlighted their benefits in the mentioned NHESS publication.

However, as described in the manuscript, 47 independent “IRPI landslide records” (cf. Section 3.1 and Section 4.1) were used for additional cross-checks. In this context, using the wording “cross-check” instead of “validation” was done on purpose, since we consider the comparison of our modelling results with landslide data from a very different data source of reduced explanatory power due to differences in e.g. the underlying definition of the process under investigation (i.e. shallow landslide) or differences in positional mapping uncertainty. Within the revised paper, we will add the following sentence to the description of the validation methods (Section 4.5):

“Finally, the three independent IFFI entries from the year 2021 and the 47 IRPI landslide records were used to cross-check the model predictions.”

RC: -Line 425: please explain why you removed 95 attributes

AC: We did not remove 95 attributes from the analyses, but 95 landslides. Within the paper we described that (Line 425f) *“Despite having the label “precipitation-induced” in their attributes, 95 of the initial 676 landslide presence observations (14%) were removed from the precipitation filter.”*

To further enhance transparency, we will rephrase this section to: *“After preliminary landslide data filtering, an additional 95 landslides were removed from the precipitation filter, because the available precipitation data provided no evidence that these recorded slope instabilities were primarily induced by precipitation.”*

RC: -Line 602: the discussion is too extensive; some parts are only informative and have nothing to do with the main goal of the paper, which is to decipher the seasonal effect of triggering and preparatory precipitation. I suggest to focus on the interpretation of the results, you could also refer to the new table in line 114 and focus on the critical parameters and their thresholds.

AC: We will shorten the discussion and remove discussion points that do not directly relate to the conducted research. We propose the following changes:

6.1 Antecedent precipitation and seasonal effects

This subchapter discusses season-dependent effects of dynamic precipitation conditions on slope stability, their potential causes while also highlighting challenges in causal model interpretation. We relate our work to the results of Luna and Korup 2022 that recently investigated seasonality in landslide occurrence using data-driven modelling. This discussion section directly relates to the main goal of the paper, which is why we would like to keep it as it is.

6.2 Mixed effects modelling and data sampling design

This subchapter discusses implications of input data quality issues (e.g. representativeness of landslide data) on the modelling results. It suggests methodical procedures that can be employed to achieve less biased results, which may be of particular interest to those in the scientific community seeking to replicate our approach.

In the interest of conciseness, we will remove the final section and associated references, which delves into broader methodological concerns, such as the choice between statistical models and more flexible "black box" algorithms. Furthermore, we will rename this section "Research design and representativeness of input data" to better convey its main focus.

6.3 Thresholds that relate to performance metrics

This section emphasizes the importance of creating thresholds in addition to continuous prediction surfaces to improve the practical applicability and interpretability of the results. In the revised version, we will eliminate the parts of the text that extend the discussion to landslide susceptibility studies to increase conciseness.

6.4 Multi-perspective model validation

This section focuses on the benefits of validating results from multiple perspectives. In line with this topic, we will remove sections that stray from this core topic. Specifically, we will delete the part that discusses the possibility of expanding the model into the spatial domain and the text that suggests the potential model's application in a multi-hazard context.

RC: You also need to better demonstrate the quality of the data, which you point out in the abstract but do not present and discuss enough in the paper.

AC: In the abstract we mention that "*the discussion illustrates why the quality of input data, study design and model transparency are crucial for landslide prediction using advanced data-driven techniques*". The revised discussion section 6.2 will now be fully dedicated to this topic, as described in the AC to the previous RC.

RC: - Line 750-755: in line with the introduction of this paper where you state: "reliable decision support tools" in this context you concluded that this approach needs further improvement. Please be more specific and give us a concrete evaluation of the applied approach as a promising tool for landslide early warning.

AC: In the revised manuscript, details on “further model improvements” will directly appear in the final discussion section and not in the conclusion (where we only mention that the validation process helped us to “*shed light on entry points for further model improvements*”).

The final discussion section will read as follows: “(...) *This in turn provided inspiration for further hypotheses to test and entry points for model improvement. For example, the systematically lower model performance on higher altitudes (Fig. 12c) and at steep terrain (Fig. 12b) indicates that in high alpine terrain and on steep slopes, landslide occurrence might be associated with different season-specific precipitation amounts, compared to lower lying and flatter areas. This in turn suggests that further model improvements may be achieved by including spatially explicit variables that directly describe landslide predisposition. The current model was specified only for landslide-prone terrain and thus disregards effects of spatially varying landslide predisposition by design (Crozier, 1989).*”

In the revised paper (Section 6.3), we will add an example on how our model may be exploited for the purpose of early warning. The new text will read as follows:

“The developed approach may be used in the context of early warning by exploiting the day-specific predictions, as shown in Figure 9. This involves evaluating whether observed or expected precipitation amounts are likely to exceed a particular threshold. To illustrate, consider a September day (Fig. 9d), where the optimal threshold will not be exceeded if 100 mm of precipitation were accumulated within the past 28 days (x-axis) and 25 mm of precipitation were expected to fall within the upcoming 2 days (y-axis). In contrast, the threshold will be exceeded if the same amount of short-term precipitation (i.e., 25 mm) falls after a particularly wet period (i.e., 200 mm within the past 28 days). Knowing that 81% of past landslides surpassed this threshold, while 79% of “wet” days without slope instability occurred below it, may further ease interpretation.”
