

“Integrated statistical modelling of spatial landslide probability” by Martin Mergili and Hone-Jay Chu, Nat. Hazards Earth Syst. Sci. Discuss., 3, 5677–5715, 2015

Response to the comments of Reviewer #1

We would like to thank the reviewer for her or his constructive remarks.

The article presents a new approach in order to assess spatial landslide probability at a regional scale. The procedure allows for release areas assessment and estimation of the impact probability in the propagation zone, by deriving statistics from an inventory of events. The approach is statistical by nature and thus allows for a good characterization of both the release and the impact probabilities, in a rigorous way. The approach and the discussion on the issue of zonal probability are worth publishing. I have however some concerns about the release area assessment. The paper is well written and the figures are of good quality.

General comments:

- The paper is well written and of high quality. The structure is fine.
- The figures are of very good quality and help understanding the text.

Thank You, we are very glad to hear this!

- There are plenty of abbreviations or acronyms, and truth is, it pretty hard to keep them all in mind. Table 1 with the various probabilities definitions is fine, but would it be a possibility to list them all somewhere? Or should you reduce the number?

It is certainly true that an additional table with all the variables shortly explained would facilitate reading. Therefore we will add an additional table (Table 2) to a revised manuscript.

- The approach is interesting, seems fairly robust, and is worth publishing.

Thank You!

- However, I have 2 concerns. The first one is the fact that you try to encompass all types of gravitational mass movements, as you say by the end of the Introduction. It is known that the triggering factors and the propagation behavior differ considerably from a phenomenon to another. I would agree to your approach when you consider them separately, both for the release and the impact probabilities. In your case study however, you seem to focus on shallow landslides or debris flow, which is not so clear. You may be more specific on what contains your inventory map.

Yes, this comment is very reasonable. We wanted to express that the chosen approach can be useful for many types of mass movements, but of course, the parameterization has to be adapted to the type of mass movement. In the present article, most landslides are rather shallow, developing into debris flows. We will improve the explanation of this aspect in a revised manuscript.

- My second concern is about the predictors used for the assessment of the release areas. Using only local slope and aspect as predictors for shallow landslides and debris flows is rather poor. This results in a not so relevant map of Pr. The use of geological maps and landuse information, may they exist, should be considered or at least discussed, as well as stability indexes or flow accumulation data, which can be processed on the DEM. This part is regrettable as it also impacts the other results. Moreover, you argue in favor of specific meteorological conditions related to a single event, which is not wrong, but you don't question your selection of predictors.

Indeed, slope and aspect are quite few predictors. However, we did some preliminary tests which showed that other factors such as upslope contributing area, topographic index or curvature showed a very poor capacity in predicting the distribution of the observed landslides. Land cover is highly problematic as variations in land cover are largely caused by the landslides themselves and are therefore not independent. A geological map could be valuable, but is currently not available to us at an appropriate scale. In the meantime we have identified the average slope between a given pixel and the next downslope river as a useful predictor which could be used in addition to those employed in the discussion paper. However, this parameter probably strongly correlates with the local slope, so that we have to be cautious in applying it. A more detailed discussion with regard to the selection of models and predictors would be included in the revised manuscript (please see also response to Reviewer #2).

Specific comments:

- P.5683 l.9: you may provide some examples of predictors to help the reader understand.

In a revised manuscript we would add some examples of predictor maps.

- P.5685 l.3-5: you may specify which routing you consider here: a single flow algorithm (D8) or your random walks?

We consider random walk routing, this aspect will be clarified in a revised manuscript.

- P.5688 l.15: it would be appreciable to have 1-2 sentences to describe the properties / behavior of your random walks.

We will briefly characterize the random walk characteristics in a revised manuscript. However, entering a higher level of detail would require a large amount of additional explanations which are out of scope of this article.

- P.5689 l.20: it should be Fig. 7d and not 7c.

Thank You for looking so carefully, this error will be corrected in the revised manuscript.

- P.5690 & Fig. 7e: I don't really get what represents Figure 7e. Is it here the upslope contributing area or an aggregate of zone with random sizes as defined at p.5684?

It is some kind of upslope contributing area, but based on the outcome of the random walk routing procedure and not on an ordinary single flow direction routing procedure. This means that the area includes the release pixels of all random walks impacting the considered pixel. We will attempt to better explain this issue in a revised manuscript.

- P.5690 l.18-19: Is it PRZ and σ PRZ or PL σ PL ??

It is indeed P_L and σ_{PL} , this error will be corrected in a revised manuscript.

- P.5691 I.7: is it PI or P*I ?

It is P_i^* , this error would be corrected in a revised manuscript.

- Table 1: is Pi the probability starting from all pixels, even with a release probability of zero ?

Yes, also the pixels with release probability zero are included as release pixel. This means that the area of the release zone (and therefore the zonal release probability) is computed including those pixels with zero release probability. An alternative would be to exclude those pixels with zero release probability both from the area of the release zone and from the impact probability. This issue would be discussed shortly in a revised manuscript.

- Table 2: the arrangement is a bit confusing. First, the elements in the MEA column should be spread in the rows, right? Then, can you put the content of the Description and Components columns vertically aligned in the middle? That would be easier to understand they are not related to a specific row, but to the whole section (eg. 1A-D).

Yes, the letters in the MEA column should be spread among all rows. In a revised manuscript, also the content of the Description and Components columns would be aligned according to the suggestion.

- Figure 1: A part of the line is missing after “Landslide inventory”.

In our opinion the figure is complete, we could not identify a missing line.

- Figure 9: the choice of the colors is not optimal here. We don't see the red on red and the yellow on blue and green.

Yes, this is true. In a revised manuscript we would use a colour scheme facilitating the interpretation of Fig. 9.

Response to the comments of Reviewer #2

We would like to thank the reviewer for her or his constructive remarks. Below, we address each comment in full detail.

The article by Mergili & Chu proposes a novel approach to combined modeling of landslide release and deposition probabilities. This is, in principle, a very interesting research direction that has the potential to enhance the utility of landslide susceptibility maps. The proposed approach is based on numerically combining and spatially aggregating release and impact probabilities. Unfortunately the combination of probability distributions of random variables requires a thorough probabilistic treatment that goes beyond the computational steps proposed in this article. In the comments below I am trying to point out some specific points where a thorough treatment of probability distributions is required.

Overall, considering the methodological issues, I recommend to reject it in its present form. From my perspective resubmission would require a thorough revision of the probabilistic model.

We partly agree with the reviewer's general comments, even though we think that not all details of the criticism are fully justified. Please see the responses below for further details.

Detailed comments:

1. In numerous occasions probabilities are averaged or their maximum is taken in order to obtain some overall probability. No justification is given, and I cannot see how these operations can constitute valid probability estimates. In particular, $P(A \text{ or } B)$ is not $\max(P(A), P(B))$ or $(P(A)+P(B))/2$, but in general $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$; if I understand correctly, the authors sometimes use maximum or averaging to calculate the probability of OR combinations of random variables / events. Examples of where averaging, maximum or multiplication seem to be incorrect or at least lack a mathematical justification: P5685L21, P5685L25, P5686L16; also Eq. (6) and (7) not justified/derived; P5686L1 what is a “simple overlay” of probabilities?

This is, actually, a good point. There is a number of aspects involved:

a. The maximum of P_{IR} is used to compute the impact probability P_i^* for a specific pixel. The reviewer is absolutely right that this way of computation is questionable from a statistical point of view. We therefore suggest to either omit this part from a revised manuscript (it is not essential for the final outcome or the general message) or – as proposed below – rephrase to „impact susceptibility“ and discuss the limitations.

b. The average of P_{IR} is employed to compute P_i for the derivation of P_L . This, in our opinion, is correct in principle even though it represents an approximation connected to some limitations which are, however, addressed in the discussion paper. In a revised manuscript we would attempt to improve the explanation, maybe as follows: let us consider a given pixel in the study area, which is characterized by a value of P_{RZ} for its upslope contribution area, depending on Z and on the pixel-based values of P_R in that area. The same pixel is further characterized by several values of P_{IR} , each relating to a possible landslide release from one pixel in the upslope contribution area. Working with the concept of the zonal release probability, (i) we have to assume that the release of landslides is equally probable for each pixel in the upslope contribution area, whilst (ii) the probability that landsliding occurs at all in this area is given by P_{RZ} . This means that the probability that a landslide reaches our considered pixel could be approximated by $P_{RZ} * \text{average of } P_{IR}$.

c. P_L is computed as the maximum of the pixel-based release probability P_R and $P_{RZ} * P_i$. The reviewer is right that this is not correct from a statistical point of view: the idea behind was to avoid that pixels where P_R is higher than $P_{RZ} * P_i$ are assigned $P_{RZ} * P_i$ as the final result. However, strictly keeping to the concept of the zonal release probability (with all its limitations), we would skip this step in a revised manuscript. This would not change the result anyway as $P_L = P_{RZ} * P_i$ for almost all pixels.

d. We agree with the reviewer that the concepts involved (due to a number of generalizations related to the zonal release probability) might not fully justify the use of the term „probability“, even though we are only talking about spatial probabilities. Instead, we would rephrase all outcomes to „susceptibilities“ (e.g., zonal release susceptibility etc.) and include the susceptibility/probability issue in the discussion. This would most likely better reflect the characteristics of the results.

2. Section 2.4 – “zonal release probability” is not well defined, and the proposed algorithm lacks mathematical justification

We have read several times through the explanation of the zonal release probability, and in our opinion it is clearly defined and the concept is sound (even though it is connected to a number of limitations which are, however, discussed – therefore we suggest to rename it to „zonal release susceptibility“). We would require more specific criticism in order to respond in more detail to this

comment. In a revised manuscript we would attempt to point out the scope and limitations of the concept in a clearer way.

3. P5680L10-11 “by combining the two newly developed open source software tools” – The authors’ phrasing of the proposed approach suggests that they consider it a merely computational step; it is important to realize that this computational step implements a stochastic model, which should therefore be firmly based on probability calculus.

If we understand correctly, the reviewer would like us to better point out that the combination of the two tools is more than just a technical issue – we agree with that and would do so in a revised manuscript.

4. Other than mentioning the `r. landslides.statistics` function in GRASS, the article does not provide any information whatsoever on the type of model used for spatially predicting landslide susceptibility, e.g. logistic regression or weights of evidence. The use of slope and aspect as the only predictors is rather unusual as many authors also include upslope contributing area, lithology or land use / land cover in their models. Especially upslope contributing area (in combination with slope angle) is meaningful from a physical perspective (compare physically-based slope stability models such as SHALSTAB).

The type of model used is explained in Section 2.3 „Zonal release probability“. We have chosen a very simple and easily reproducible approach which (as required for the general conceptual framework) yields spatial probabilities in the range 0–1. We could certainly compare the outcome of this approach to the outcomes of other approaches such as logistic regression, weight of evidence, neural networks etc. However, our article does not focus on statistical modelling of landslide release, so that, in our opinion, this would be out of scope.

Indeed, slope and aspect are quite few predictors. However, we did some preliminary tests which showed that other factors such as upslope contributing area, topographic index or curvature showed a very poor capacity in predicting the distribution of the observed landslides. Land cover is highly problematic as variations in land cover are largely caused by the landslides themselves and are therefore not independent. A geological map could be valuable, but is currently not available to us at an appropriate scale.

In the meantime we have identified the average slope between a given pixel and the next downslope river as a useful predictor which could be used in addition to those employed in the discussion paper. However, this parameter probably strongly correlates with the local slope, so that we have to be cautious in applying it.

In summary, there is certainly a significant potential for refining and extending the statistical modelling of the release. However, modelling landslide release is not the primary scope of the present article (there is a huge bulk on literature on this topic) so that we do not want to blow up this part too much. However, a more detailed discussion with regard to the selection of models and predictors would be included in the revised manuscript.