Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., author comment AC2
https://doi.org/10.5194/nhess-2021-22-AC2, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

Jason Goetz et al.

Author comment on "Optimizing and validating the Gravitational Process Path model for regional debris-flow runout modelling" by Jason Goetz et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-22-AC2, 2021

General comments

The paper presents an approach to optimize the parameters of the Gravitational Process Path model for regional debris-flow runout modelling. It addresses the evaluation of the source areas as well as of the runout path and its length. The approach is illustrated with a case study in the upper Maipo river basin in the Andes of Santiago, Chile. The method and the sensitivity analyses are interesting and add value to the field of regional debris flow modelling. The paper is well written, and the figures are of high quality. I recommend publishing it after consideration of two main concerns I have about performance metrics that may require additional work.

We would also like to thank this reviewer for providing highly constructive comments. We believe the first concern regarding the use of the AUROC as a metric for runout path can be addressed by providing a more explicit description of how the AUROC is computed, as well as by providing a more detailed interpretation of the random walk optimization in the discussion. We addressed the second issue regarding runout distance based on the minimum area bounding box length by providing a discussion of how this metric may impact the results. Overall, we believe by addressing these concerns and the specific comments, we are able to provide an even more valuable contribution to the debris flow modelling community.

Main concerns

I have two main concerns about the performance metrics of the runout distance and runout path:
AUROC for the path: you process the AUROC as defined by: "Model performance was rated higher if the random walk model contained observed debris-flow tracks within its simulated paths" (2.2.1). The problem here is that there is no "false positive" in your approach, and thus the model is not penalized for over-predicting. The approach is correct for the source areas but not for the runout path. As we can see in Fig. 2, the extent of the modelled debris flow is much larger than the observed one, but the AUROC is almost = 1. It means that your model needs to spread as widely as possible to have a good score. I get the difficulty of comparing potential events to a single observed event, but you might then use another contingency table score that does not have false positives. Using a ROC-type score is misleading here if there is no false positive.

Thanks for bringing up this concern. The AUROC does account for false positives. The receiver operating characteristic curve (ROC), from which we calculate the area under the

curve (AUROC), is a plot of true positive vs. false positive rates (Zweig and Campbell 1993). We did not explicitly state this in the original manuscript, so we will put a brief description of the AUROC into the methods section. Our results also indicated that the path optimization does not always favour maximum lateral spread. This is illustrated by the variety of exponent-of-divergence values in the parameter selection frequency plot (Figure 12a). We didn't make this point clear in the paper, so we will add it, as well as the following interpretation of the optimization of the random walk model:

"The best-performing regional random-walk parameters allowed for maximum lateral spreading of the runout path given the range of parameters for optimization. Individual events tended to also optimize for high lateral spreading, but not as strongly as the regional model. We believe this high lateral spreading may be due to the location of the observed debris flows relative to simulated paths and the quality of the DEM. A large proportion of the observed debris-flow tracks were located at the fringe of the most frequent simulated paths. Thus, a higher slope threshold and exponent of divergence are required to capture these fringe debris flows. Additionally, the surface of DEMs with resolutions greater than 20 m can be too general to capture minor gullies that may have high flow accumulation (Blahut et al 2010b). The 12.5 m resolution ALOS DEM used in this study is derived from downsampled SRTM data, and would likely contain some of the topographic generalizations of the original DEM (~ 30 m spatial resolution). Despite potential issues with DEM quality, similarly to Horton et al. (2013), we illustrated valuable results can still be achieved."

Relative error for the runout distance: your approach of using a bounding box on the median frequency (2.2.1) to quantify the runout distance is interesting, but I have an issue with it. Most observed debris flows will likely propagate to the valley-bottom, where they might meet the main river. My problem is that when you model the debris flow propagation with small friction values, it is likely to reach the main river and continue perpendicularly, thus not increasing the bounding box for some iterations of the parameter values. We can see such behaviour in your Fig. 2. There is, therefore, a discontinuity as too long propagations are less penalized than too short ones. I believe this might play a role in the results of Fig. 6, where the runout length error remains low for a large range of sliding friction coefficients. It might provide a misleading impression of insensitivity. Or is it the case that most propagations reach a flatter area where they quickly stop anyway? Although an approach based on actual length (for example, defined by a D8) might better represent the difference in runout distance, it might not be trivial to use the median frequency criteria. What about using the median length of all random walk runs for one setting, provided it's a piece of information you can get? This problem should be at least discussed and considered in the interpretation of the sensitivity analyses. Then, interpretation such as in l. 380-381 ("This may indicate that the combination of random walk and the process based PCM model dictates a general runout pattern that is insensitive to values within a broad and nearly optimal range of physically reasonable parameters") might not be stated this way. Same for l. 407 ("we observed a general insensitivity in runout distance performance of the PCM model to a range of parameters").

Thanks. Excellent questions.

Regarding parameter insensitivity. We will add the following paragraph to the discussion highlighting that the bounding box approach may have some influence on the sensitivity of parameter performance, and remove our previous statements (l. 380-381 and l. 407) on this issue.

"Although we obtained a unique regional model solution, runout-distance relative errors were only slightly higher than the best performer for pairs of sliding-friction coefficients and mass-to-drag ratios across a band in grid-search space of lower sliding-friction coefficients. This insensitivity of performance to different combinations of PCM model

parameters may be due to the uniqueness problem. Our approach using the minimum area-bounding box could also contribute to this observed parameter insensitivity. Abrupt changes in flow perpendicular to the initial flow direction, such as a flow meeting a channel, may only slightly increase the length of the bounding box for several iterations of decreasing sliding-friction coefficient (or increasing mass-to-drag ratio). However, our approach of breaking parameter ties using the AUROC ensures that we select the parameter set that best fits the runout extent in addition to distance – slides with longer runout distances should also have a lower AUROC performance"

We don't believe there is a general issue of longer slides being less penalized than short ones. This seems to only occur when the flow path is nearly perpendicular to the initial flow direction, which is not the case for all the mapped debris-flows tracks used for regional-model training.

Thank you for suggesting other approaches to estimating runout distance. We are generally satisfied with our approach to quickly estimate runout length for our study area. Much of the work in this paper was developing an open-source framework to optimize process-based models for runout simulation that can be adapted by others. We highly encourage and look forward to seeing future applications of this approach modify this framework, such as trying different performance metrics, to best suit particular applications.

Specific comments

Figure 1: The caption should be a bit more comprehensive, explaining, for example, the random sample.

Thanks, we will add, "The selection of these debris flow polygons was based on a random sample" to the caption.

Section 2.1.3: It might be useful to describe the fundamental principles of the AUROC in 1 sentence.

Here, we will add the following, "The receiver operating characteristic (ROC) is a plot of the true positive rate versus the false positive rate. AUROC values range from 0.5 (random discrimination between classes) and 1.0 (a perfect classifier)."

Section 2.2: Please provide more details about the models and their parameters. For example, mention the random component in the iterative simulations of the random walk and give more information about the persistence factor and the exponent of divergence. As they are key parameters for the rest of the paper, adding a few sentences to describe them and 1-2 equations would be beneficial for the readers.

We agree that a better description of the random walk model (Gamma 2000) can help improve the reader's interpretation of the results. We therefore will add the following to Section 2.2:

"Flow path is determined using a 3×3 window that first controls the path of a central cell by considering only neighboring cells with lower elevation. If the neighbouring cells are below the slope threshold, the neighboring cell with the steepest descent is selected; otherwise, neighbours are assigned transition probabilities based on slope. These probabilities are adjusted using the exponent of divergence and the persistence factor. A higher exponent of divergence will result in more even probabilities across the neighbouring cells, allowing for a higher likelihood of not selecting the steepest descent path. The persistence factor considers the previous flow direction in weighting the probabilities. A higher persistence factor increases the probability that the selected neighbor will follow the direction of the previous cell. Based on these transition

probabilities, a pseudo-random number generator selects a cell to define the flow path (see Gamma 2000; and Wichman 2017 for a more detailed description). With this random-walk implementation, the flow path stops when the neighboring cells have a higher or equal elevation compared to the central cell."

Section 2.2.1: Please mention that you do an exhaustive grid search.

Added.

Section 2.2.4: You have chosen 1000 "non-debris flow locations". However, could these be excluded to be potential source areas for future events? Could they become source areas under certain triggering conditions?

This is a general challenge in selecting non-debris flow locations. Future work could focus on improving methods for identifying these locations.

Figure 5: It would deserve some more interpretation. For example, what can explain the role of elevation in debris flow conditioning? Why is the slope angle contribution decreasing after a certain threshold? What about the plan curvature?

The relationship between elevation and debris flow activity is complex. In the upper Maipo river basin elevation can be a proxy for vegetation, snow cover duration, terrain ruggedness, permafrost and glacial bodies, and geology. It is therefore difficult to discern any direct relationships between elevation and likelihood of being debris source areas. However, we suspect that lower elevations are predicted to be less prone to be source areas due to increased vegetation cover and less rugged terrain. The decrease observed at the highest elevations may relate to permafrost and glacial bodies holding potentially mobilized sediment (e.g. Sattler et al. 2011).

We observed a decrease in likelihood of source areas occurring at high slope angles (e.g. ~ >45°). These steep slopes can be associated with steep rock faces that are more likely sources of rock falls than debris flows (Loye et al 2009).

Slightly concave plan curvature of the slopes (relative to the DEM) are associated with being more likely source areas.

We will add this interpretation to the results.

Section 3.3 & Figure 9: Is the runout frequency relative to a single source? How are they combined when different propagations overlap? Please add some clarifications.

As computed from the GPP model, the runout frequencies are the total times a cell is traversed from all source areas (Wichmann 2017). We will add the following to section 2.2.

"This is a cumulative frequency based on simulations from all source areas"

Section 3.4, l. 299: "these cases were related to misclassifying stream erosion": can you identify such information from satellite imagery?

Through expert interpretation of DEM derived hillslope angles and very high resolution satellite imagery (0.50 m) we are confident in our ability to identify such information. We didn't make it clear in the paper that hillslope angle was used to help with the interpretation, so we will add this to the paper.

Section 3.4, l. 309-311: not so clear; please clarify.

Thanks. We will clarify this section by rephrasing it to:

"The optimization of the runout model avoided overfitting to debris-flow tracks of a certain magnitude and general terrain conditions. That is, we did not observe a strong correlation between runout distance performance to length of observed debris flow (ρ = -0.36), starting elevations (-0.21), catchment area (0.11) or hillslope angle (0.29) of source points used for model training."

Figure 12a: You do not mention plot 12a in the text, i.e., the slope threshold values in the grid of other parameters.

We will add the following to the results to describe the simulated path behaviour of the individual events:

"Most individual events optimized runout paths with parameter sets leading to high lateral spreading. The optimal-path parameters for most of the individual events had a 40° slope threshold, high exponent of divergence and low persistence values (Figure 12a). By individually examining the optimal simulated paths for each training event, we observed that 60% of the observed debris-flow tracks did occur within the most frequent simulated paths. The other 40% of events were typically located on the fringes of the most frequent paths."

Section 3.6 & Figure 14: You might mention again here that these scores are processed on the test data.

We will add that we used spatial cross-validation to assess the performance in the figure caption.

"Figure 14. Comparison of runout path (a) and distance (b) performances for different model training samples sizes assessed using spatial cross-validation. The error bars indicate the standard deviation in performances."

We will also add some clarification of this in the methods (Section 2.2.3).

"Spatial cross-validation was applied to data sets of varying training sample sizes using the random sample of 100 debris flows used for model optimization"

Section 4.2: The ability to optimize the runout path and the runout distance separately is related to the fact that the random walk mainly controls the path/spreading, and the PCM controls the runout distance. The influences of these algorithms are quite distinct.

Thank you. This is a really valid point to remind the readers in the discussion. We will add the following to Section 4.2.:

"The modular framework of the GPP model provides the ability to optimize two distinct runout components, the runout path including lateral spreading and the runout distance. In our study, we used the random walk and PCM components of the GPP model to simulate spatial extent of runout."

Conclusion: Should contain some more results of your study.

We will add the following points to our conclusion.

- *The combination of the statistical learning for source area prediction and regional optimization of the random walk and PCM model components of the GPP runout model performed well at generalizing runout patterns across the upper Maipo river basin.*

- *In addition to its strong performance, the transparency and interpretability of the GAM provided further user confidence in predicting debris flow source areas.*
- *Unique regional-optimal PCM model solutions were more prone with larger sample sizes, as well as higher model performances and lower uncertainties.*

Technical corrections
l. 5: "y" is missing in Germany

Corrected.

l. 73: "our" instead of "out"

Corrected.

l.186: "We explored *for* such spatial…" ?

Corrected.

l. 378: what do you mean by "ambiguous events"?

We meant to refer to uncertainties in mapping debris flows.

We will change this sentence to, "Poorly individually optimized events could be attributed to locally poor DEM quality (Horton et al 2013) and mapping uncertainties (Ardizzone et al 2002)."

l. 386: "very *a* specific problem"

Corrected.

References

Gamma, P.: Dfwalk – Murgang-Simulationsmodell zur Gefahrenzonierung, Geographica Bernensia, G66, 2000.

Horton, P., Jaboyedoff, M., Rudaz, B. E. A., & Zimmermann, M. (2013). Flow-R, a model for susceptibility mapping of debris flows and other gravitational hazards at a regional scale. Natural hazards and earth system sciences, 13(4), 869-885.

Loye, A., Jaboyedoff, M., & Pedrazzini, A. (2009). Identification of potential rockfall source areas at a regional scale using a DEM-based geomorphometric analysis. Natural Hazards and Earth System Sciences, 9(5), 1643-1653.

Sattler, K., Keiler, M., Zischg, A., & Schrott, L. (2011). On the connection between debris flow activity and permafrost degradation: a case study from the Schnalstal, South Tyrolean Alps, Italy. Permafrost and Periglacial Processes, 22(3), 254-265.

Wichmann, V. (2017). The Gravitational Process Path (GPP) model (v1. 0)–a GIS-based simulation framework for gravitational processes. Geoscientific Model Development, 10(9), 3309-3327.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical chemistry, 39(4), 561-577.