Author's Response:

Optimizing and validating the Gravitational Process Path model for regional debris-flow runout modelling

Jason Goetz[1], Robin Kohrs[1], Eric Parra Hormazábal[2], Manuel Bustos Morales[3], María Belén Araneda Riquelme[3], Cristián Henríquez[3], Alexander Brenning[1]

[1] Department of Geography, Friedrich Schiller University Jena, Germany
[2] Institute of Geosciences, University of Potsdam, Germany
[3] Instituto de Geografía, Pontificia Universidad Católica de Chile, Chile. Centre for Sustainable Urban Development, CEDEUS. Centro de Cambio Global UC.

Correspondence to: Jason Goetz (jason.goetz@uni-jena.de)

Dear Margreth Keiler (Editor),

Thank you for considering our manuscript for publication. We're excited to present the new version of our manuscript, which has been improved thanks to highly constructive comments of the reviewers.

Overall, we addressed Reviewer #1's main comments by providing more detail in the methods regarding the use of the AUROC. We also included a much more detailed description of the runout path and distance model components. Reviewer #2's main comments were addressed by adding new insights to the discussion related to our regional runout-model behaviour and its limitations, which are supported by our results.

Below you will find our point-by-point reply to the reviewers' comments.

Thank you,

Jason Goetz and co-authors.

Reviewer #1 Comments

General Comments

Dear Editor, dear Authors,

this a well-written and interesting paper on the automatic calibration and validation of a framework for regional debris-flow modelling. Besides the modelling of debris-flow initiation sites with a GAM, the GPP model is used for debris-flow path and runout modelling in the upper Maipo river basin, Andes of central Chile. The authors develop and present a novel approach for model optimization and validation, including several aspects like uncertainty in parameter selection, spatial transferability, and the models's sensitivity to sample size. The results are well presented and discussed, including very nice and informative figures to illustrate the findings. Most parts of section 2 (material and methods) are also well written, but I think this is the section which could be improved most by adding some more detail on some of the aspects (see specific comments below). Apart from that I think the paper is well suited for publication in NHESS. It is also really nice to see that the tools developed for this paper (as well as the data) are also made available to the public.

With best regards.

We thank this reviewer for their highly constructive comments. As shown in our following response, we will make improvements to the methods section to help clarify our approach on the use of the AUROC as a metric for runout path modelling, sampling debris-flow and non-debris-flow source areas, and how the GPP implementation of the random walk model limits runout distance. Additionally, we believe by addressing their comments on the potential multiple optimal PCM (runout distance) model solutions, we enhance our discussion and further demonstrate the suitability of our approach for regional debris flow runout modelling.

Specific Comments

Section 2.1.2

Please use a different color for debris flows and roads in Fig. 1, they are both grey and can't be distinguished very well.

We have updated the colours in this figure.

Section 2.1.3

Regarding the sampling of presence and absence of source points: how do you exactly determine the non-source points? Do you somehow guarantee that the samples are not "too close" to mapped source points? There are much more non-source than source points in your study area, how does this influence the results? This affects training as well as validation, please elaborate.

We have rephrased some sentences in this section to help clarify how non-source points were sampled and how many.

“The non-source (i.e. absence) points were determined by random sampling locations within the mapped sub-basins outside of the debris flow polygons. The resulting training and test data contained 541 source points and 541 non-source points.”

We guaranteed that the samples were not too close to source points by sampling outside of the mapped polygons. As mentioned in L.102. We used the commonly applied 1:1 sampling ratio of source and non-source points.

After denoising, you apply a sink filling algorithm to the DEM, which one?

We used the sink filling algorithm from Planchon and Darboux (2001). This citation was added.

Section 2.2.1

Regarding the rating of the random walk performance (line 160): performance was rated higher if observed debris-flow tracks were within the modelled paths. Please provide more details on how this was done exactly, e.g. did you also take the number of cells into consideration that were outside the mapped track? Otherwise you might get optimized parameters that overestimate the process area.

We accounted for the cells outside of the mapped track. The ROC curve plots the true positive rate (TPR) vs the false positive rate (FPR; Zweig and Campbell, 1993). Therefore the AUROC does consider cells inside and outside of the mapped tracks. We have added a brief description of the AUROC to the paper to help clarify this, as well as the Zweig and Campbell 1993 citation.

Regarding the random walk parameter optimization before the runout optimization (two-stage approach): in order to optimize the random walk parameters, wouldn't you also require to use some kind of friction model to limit the runout distance? This overlaps with the previous question, please explain.

The Gamma (2000) random walk model implemented in the GPP model (Wichman, 2017) does not have controls for runout distance. The flow paths will continue downslope until neighboring cells have a higher or equal elevation compared to the central cell being processed. We will add this detail to the paper.

Regarding runout distance optimization: here, you use a minimum area bounding box to measure length. What impact has the character of the debris flow path on this concept? For example, take (1) a quite short, more or less straight debris-flow path versus a (2) very long path, which runs from a hillslope into a channel with a distinct change of direction, let's say 90°? Then you get (1) a bounding box matching the real length quite well and (2) a bounding box which is almost square, strongly underestimating the runout length.

This is an excellent question. It was also brought up by Reviewer #2.

In theory, it is possible that the runout length of the minimum area bounding box can be underestimated when a debris flow makes an abrupt 90 degree change in direction. This may occur for some iterations of decreasing sliding friction coefficient (or increasing mass-to-drag ratio) past the actual optimal value. However, by visualizing our debris flow *optimization procedure for different sliding friction coefficients for all training samples, we did not observe this to be an issue.*

Additionally, to mitigate this potential issue in optimal parameter selection, we use the AUROC to break any ties in performance. Longer runout paths should have a lower AUROC. We added the following paragraph to the discussion to better explain this issue:

“In theory, it is possible that the minimum area-bounding box could contribute to 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 μ (or increasing M/D). However, we did not observe this to be an issue within our training data for our given parameter ranges of μ and M/D.”

Regarding the optimization of the 2 parameters of the PCM model (sliding friction coefficient "my" and mass-to-drag ratio "M/D"): a general problem with the PCM model calibration is, that there is some mathematical redundancy between the parameters. I.e., you can achieve the same runout length with different parameter combinations of my and M/D. How does your calibration approach handle this? Please add some information on this, because this may also have some impact on other sections of the paper, e.g. section 3.2 ("low sensitivity for a large range of parameter combinations"), section 3.5 ("no clear spatial pattern in optimal my and M/D parameter combinations across the study area"), section 4.1.2 ("we observed high variability in optimal PCM parameters").

We found that this “uniqueness problem” was not a major issue for regional optimization - the focus of our study. We were able to obtain a regionally unique PCM model solution. Additionally, when training and testing on different clusters of sampled debris flows, we observed only 5% of the repeated spatial cross-validation iterations (n = 5000) had multiple optimal solutions.

If there were ties in the PCM model, we selected the parameter set that resulted in the highest AUROC of the runout path performance. For individual events the occurrence rate of multiple solutions before tie-breaking was much higher (56%). However, after using the AUROC to break ties, the vast majority of individual events (97%) had a unique solution - we have added these results to sections 3.2 and 3.5. For the remaining cases, which still had ties, we simply selected the first record.

We will add the following to the discussion to better highlight this issue with optimizing the PCM model.

“The two-parameter PCM model has a uniqueness problem (Perla et al., 1980). Possibly infinitely many pairs of the sliding friction coefficient and mass-to-drag ratio result in the same runout distances. When optimizing individual events, we did observe this phenomenon. The majority of individual events had more than one optimal combination of parameters. Obtaining a unique solution was not an issue for the regional optimization in our study for the given grid search space. Likely this is due to having to satisfy the runout distances for a variety of hillslope conditions and lengths across the study area. The observed reduced variability in optimal solutions for larger sample sizes (Figure 15) provides some evidence for this conjecture.”

Section 2.2.2

You assessed the transferability of optimized model parameters by 5-fold spatial cross-validation. In section 2.2.1 you state that you are using a random sample of 100 debris-flow tracks for optimization. Is this the sample size you use here too? Or how is this related?

It is the same sample. We added, "Based on our random sample of 100 debris-flow tracks, ..." to help clarify this.

Section 2.2.4

To calculate the AUROC, you used 1000 samples of both debris-flow and non-debris-flow locations. How did you sample the non-debris-flow locations? Thematically similar to my question on the non-source point sampling.

We randomly sampled locations outside of debris flow polygons. We rephrased this to, "The AUROC was calculated using a sample of 1,000 debris-flow runout locations and 1,000 non-debris-flow locations outside of the debris-flow polygons".

Section 3.1

You write that areas with slightly concave profile curvature were modelled as more likely being source areas. So far plan (not profile) curvature was used, and it is also plan curvature that is shown in Fig. 5.

Thanks, this was a typo. We mean plan curvature.

Section 3.2

I think it would improve the reading of Table 2 if you would name the "third" model component "Runout distance (spatially varying friction)" instead of only "Runout distance" (like the "second" model component).

Good point! We made this change.

Section 3.4

In line 294 you write "... the modelled runout paths failed to follow the flow direction ...": is this due to a general problem of the flow path model or is this caused by errors in the DEM?

This is likely a problem of the errors in the DEM than the flow path model. We previously mentioned this in the discussion - however, we added references to works that cover these issues in more detail,

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

In line 299 you write that "these cases were related to missclassifying stream erosion ...": was the runout lemgth over- or underestimated in these cases?

Runout was underestimated for these cases (Figure 11c), likely due to the relatively gentle slope of these stream channels.

Section 4.1.2

This section (mostly) discusses the runout distance model, please also add a few sentences on the runout path model.

Thanks, we added the following interpretation of the runout path model results to the discussion,

“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.”

Technical corrections

p1, l5: fix typo in "Germany"

Corrected.

p1, l29: remove additional blank after "learning"; "source" instead of "sources"

Corrected.

p2, l30: remove additional blank after "and"

Corrected.

p2, l46/47: not sure if a comma should be used instead of a semicolon in the enumeration

Corrected.

p2, l55/56: add missing periods after "al" in three citations

Corrected.

p2, l56: remove additional blanks after "be"

Corrected.

p3, l73: "our" instead of "out"

Corrected.

p3, l77: add missing periods after "al" in three citations; Moreiras et al. 2012 and Serey et al. 2019 are missing in the references, please add

Corrected.

p3, l80: add missing period after "al" in the citation; Sepulveda et al. 2006 is missing in the references, please add

Corrected.

p3, l81: add missing period after "al" in the citation

Corrected.

p3, l83: add missing period after "al" in the citation

Corrected.

p4, l95: add period at the end of the table description

Corrected.

p5, l115: add missing "the" in "with ___ remaining set"

Corrected.

p6, l136: the PCM model was developed by three authors, so it isn't "Perla's" model, please rephrase

We made this change.

p9, l208: throughout the text you use a hypen in "debris-flow", here you write "non-debris flow"; should this be changed?

We rephrased this to, "locations outside of the debris-flow polygons"

p9, l214: add the missing "a", the package is called "Rsagacmd"

Corrected.

p11, l243: "mass-to-drag ratio" not "mass-to-drag-ratio"

Corrected.

p14, l278: "towards a threshold of 0.5" instead of "thresholds"

Corrected.

p14, l279: There's quite a break between the two sentences, I had to read it twice to realize that "The resulting runout prediction map ..." was meant to be that with a threshold of 0.7. Maybe it would help to start a new paragraph here or to reformulate the sentence to something like "The runout prediction map resulting from the best threshold ..."

Thanks for the recommendation. We rephrased it to, "The runout prediction map resulting from the best threshold".

p18, l314, Figure 11: "... runout path (a), ..." "... relative error (b), actual runout length error (c), and ..."

Corrected.

p22, l264: "source conditions to spatially" instead of "source conditions spatially"

Corrected.

p22, l373: "parameters of the PCM model" instead of "parameters the PCM model"

Corrected.

p24, **References**: please add the missing references and also have a look at the formatting - there are many references in which the author's first names are not shortened to the initials

Corrected.

References

Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., & Reichenbach, P. (2002). Impact of mapping errors on the reliability of landslide hazard maps. Natural hazards and earth system sciences, 2(1/2), 3-14.

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.

Planchon, O., & Darboux, F. (2001). A fast, simple and versatile algorithm to fill the depressions of digital elevation models. Catena, 46(2-3), 159-176.

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.

Reviewer #2 Comments

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 is 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 potentially 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 added 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. These are excellent questions that enabled us to better understand the pattern of PCM model performance across grid search space.

Regarding parameter insensitivity. For our study, we do not 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. We visualized the bounding *box positions and relative errors for different sliding friction coefficient iterations for the entire debris flow training sample to confirm this.*

We agree that it is important to mention this potential limitation, so we have added the following paragraph to the discussion and removed our previous statements (l. 380-381 and l. 407) on this issue.

"In theory, it is possible that the minimum area-bounding box could contribute to 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 μ (or increasing M/D). However, we did not observe this to be an issue within our training data for our given parameter ranges of μ and M/D."

Alternatively, this observed 'insensitivity' of runout distance performance across the sliding friction coefficients μ is likely due to decreasing hillslope gradient being one of the main controls of runout in the PCM model. Similar to Perla et al (1980), we used tanθ > μ to interpret μ as the slope angle where deceleration and deposition of runout begins, and found that best performing μ values were similar to the slope angles at or near the stopping positions of the training sample of debris flows. We have added the following paragraph to the discussion to explain this, as well describing the equation tanθ > μ in our methods.

"Although we obtained a unique regional model solution, runout-distance relative errors were only slightly higher than the best performer for combinations of μ and M/D across a band in grid-search space of lower μ values (Figure 6). We believe this model performance insensitivity to sliding friction angles is due to decrease in slope being one of the main factors controlling runout distance. The lowest relative errors tended to occur when the slope condition for deceleration was in the range of 2.3° to 11.3°, and the optimal being 6° (μ > 0.04, μ > 0.20 and μ > 0.11; Eq. 6; Figure 6). This range matches well to the slope values observed at or near the stopping locations of the debris flows used to train the PCM model. These results are also very similar to modelled observations of debris-flow runout in the semi-arid Andes by Mergili et al. (2012). They observed that best PCM runout modelling results for individual events occurred when deposition began at slope angles ranging from 2.6° to 14.0° (μ > 0.045 and μ > 0.25). Additionally, this range fits within other global observations of debris-flow deposition occurring on slopes smaller than 6° to 17° (Hungr et al, 1984; Ikeya, 1989, Rickenmann and Zimmermann, 1993; Bathurst et al. 1997; Lorente et al., 2003). The insensitivity across M/D values (Figure 6) is due to many possible solutions for obtaining similar runout distances with different combinations of μ and M/D (i.e. the uniqueness problem). The range of μ values resulting in low relative error only slightly increased with higher M/D value, which indicates that μ had a much stronger role than M/D in governing runout distances."

Thank you for suggesting other approaches to estimating runout distance. We are satisfied with our approach to quickly estimate runout length for our study area. As we confirmed by visualizing optimization iterations for all sampled debris flows, we found that the bounding-box approach did well at approximating runout length for model optimization. Additionally, 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 added, "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 added 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 added 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."

We also added equations and a more in-depth description of the PCM model to this methods section.

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 added this interpretation to the discussion.

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 added 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 the hillslope angle was used to help with the interpretation, so we added this to the paper.

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

Thanks. We clarified 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 added 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 added 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 also added 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 added 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 added the following points to our conclusion.

- *We demonstrated that the combination of the statistical prediction (GAM) of source areas and our regional optimization of the GPP runout model (random walk and PCM) performed well at generalizing runout patterns across the upper Maipo river basin.*

- *In addition to high model performance, the transparency and interpretability of the GAM provided further confidence in the prediction of source areas by illustrating regionally geomorphically plausible modelled behavior.*

- *The optimized runout model parameters sets were consistently similar within grid search space when assessing transferability using spatial cross-validation. We believe this strong transferability of our runout model was due to the hillslope gradient of the deposition area being one of the major controls of runout distance in the PCM model.*

- *The regionally optimized runout model also resulted in geomorphically plausible results, with best performing μ and M/D combinations occurring when simulated debris-flow deposition and termination occurred on slopes less than 11°.*

- *Although obtaining unique PCM parameter solutions for individual events can be an issue, we were able to obtain a unique PCM model solution for our regional model. In general, we found unique regional-optimal PCM model solutions were more prone with larger sample sizes, as well as higher model performance 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 changed 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

Bathurst, J. C., Burton, A., and Ward, T. J. (1997). Debris Flow Run-Out and Landslide Sediment Delivery Model Tests, Journal of Hydraulic Engineering, 123, 410–419.

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

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.

Hungr, O., Morgan, G. C., and Kellerhals, R. (1984). Quantitative analysis of debris torrent hazards for design of remedial measures, Can. Geotech. J., 21, 663–677.

Ikeya, H.: Debris flow and its countermeasures in Japan (1989) Bull Eng Geol Env, 40, 15–33.

Lorente, A., Beguería, S., Bathurst, J. C., and García-Ruiz, J. M. (2003). Debris flow characteristics and relationships in the Central Spanish Pyrenees, Nat. Hazards Earth Syst. Sci., 3, 683–691.

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.

Rickenmann, D. and Zimmermann, M. (1993). The 1987 debris flows in Switzerland: documentation and analysis, Geomorphology, 8, 175–189.

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.