Replies to referee 1 and referee 2 for the manuscript titled:

A landslide runout model for sediment transport, landscape evolution and hazard assessment applications

Contents

RC1 review replies

Thank you for the comments, particularly your questions about MWR regional performance. We developed MWR to fit a number of needs, including those of sediment budget evaluations, runout hazard assessments and landscape evolution studies. As this is the first paper for MWR, model description and evaluation took up significant space, and we decided to demonstrate the first two applications. Section 5 shows MWR's skill for replicating observed sediment transport and topographic change caused by landslides. Section 6.2 shows how MWR can be used to create runout hazard maps.

As you note, the 2 km^2 watershed is not large enough to demonstrate a regional application and we leave such work for future studies (see last paragraph of concluding statements). As this still needs demonstration, we downplayed our claim that the model is faster than other models (lines 784 to 788). We gave an example of model speed in Section 5.3 (Line 529 to 530) and included a statement in the concluding statement (791 to 792).

Replies to line comments: line numbers in reply refer to the tracked changes version of the paper.

Line 91: It's unclear what the two parameters are.

Thank you, we list the parameters and direct the reader to section 2, where they are described (line 90)

Line 98: The locations should be named, at least approximately.

Thank you, we added text to direct the reader to Figure 1 and Section 5, where they are described.

Line 145: It's not clear what the "attributes at node n" are. A few examples would be nice to have.

Thank you, examples are now listed at lines 136 and 303

Line 221: The assumption that the deposit slope and the underlying topographic slope are unified is confusing, especially since the model indicates that the slope changes with each deposition in Line 303.

Thank you for the comment. That's correct, the terrain and slope does evolve as the model progresses. However, for the purpose of deriving a rule that permits deposition to vary as a function of the critical slope parameter, we assume a deposition geometry. During the last revision and this one, we refinde this section (line 211 to 251).

Line 277: It seems field calibration is necessary for the model, which could make regional applications nearly impossible without extensive field campaigns. Could the values be approximated using globally available data?

Yes, parameter values can be inferred from globally available data if that data shows where debris flows stop (for Sc) and how thick their frontal lobes are at the point of deposition (qc). We added text in section 6.1 (lines 706 to 718) and to the conclusion (lines 790 to 799) to reflect this. For accurate, site-specific estimates of sediment transport or topographic change caused by the runout, field calibration will be necessary.

Line 282: What are the "regolith attributes"?

Thank you for the question, we now list example attributes at line 303

Lines 304–306: The suggested workaround seems like a simple functioning solution. I appreciate the statement.

Thank you for the comment. This workaround is now described at lines 321-327

Lines 314–315: In the Algorithm tables, repeating the parameter meanings for easy reading would be helpful.

Thank you for the comment. We decided to delete the Algorithm tables and rely on the text descriptions.

Line 369: The paper is dense with equations and parameters. Adding more acronyms, such as" DoD", could make it hard to follow. It would be better to minimize their use.

Thank you for the comment. Since we repeatedly refer to the DEM of Difference, we decided to use an acronym. To help the reader keep track of all of the parameters we have added a notation table (line 860)

Lines 381–387: It's still unclear how "A minimum flow thickness" is determined.

Thank you, in the revision, we describe how to estimate the range of potential q_c values from the range of minimum observed thickness of debris flow termini in the field at all of the validation sites (lines 518 to 524)

Line 405: Could the authors suggest a range for "N_cycles"? Is it around 2000, as mentioned in line 422?

Thank you for the suggestion, we added text to indicate why a Markov chain of 2000 repetitions was adequate for our applications as well as what we intend to do in the future to improve the calibration algorithm (lines 378 to 380). To help reduce the number of parameters in the manuscript, we no longer mention "ncycles".

Lines 440–448: It seems only two of the three options for specifying the initial landslide have been used in testing. Clarification would be helpful.

Thank you for the comment. We rewrote this section to be more concise and indicate the scenario in which each option might be of use (lines 391 to 408). We don't test option 1 in Section 6 because it's use is demonstrated in the Section 5.

Table 1: These data may not be available for most real-world applications without detailed field campaigns. Could there be a case study showing reliance on global data or remotely sensed data? In the case of hazard assessment, we don't know where the landslides will actually be in a region. As anticipated in the real-world application in Figure 12, there may be several potential sources that could be activated. For this tool to work, one needs to visit the entire region of interest and get detailed data, as depicted in Table 1, which makes the tool practically unusable beyond very small catchments such as the one presented in Figure 12.

Table 1 reports a combination of attributes that inform calibration as well as basic statistics of the landslides (like total mobilized volume). We discuss and report how we can infer some of the model parameters such as parameter k from such data (See appendix). These inferences were intended for application of the model for sediment transport and topographic change prediction. For cases when detailed erosion and deposition are not needed and the model objective is just runout extent, MWR can be calibrated using only Sc and qc inferred from remote sensing data sets. We hope we made this point clear in the paper. We added text in section 6.1 (lines 706 to 718) and to the conclusion (lines 790 to

799) to reflect this. For accurate, site-specific estimates of sediment transport or topographic change caused by the runout, field calibration will be necessary.

Line 557: "m3" should be formatted as m^3.

Thank you, this has been corrected at lines 503.

Lines 614 and 629: It's impressive that the model can replicate depositional behaviour with high organic debris content. This is a strong point of the tool.

Thank you.

Line 624: "DEMS" should be DEMs.

Thank you for catching this, we removed this sentence from the paper.

Figure 7: Including the original landslide (source and deposition) polygon in all subplots for comparison would be helpful. This applies to Figures 8, 10, and 11 as well.

Thank you for the suggestion. In the revision, the figure numbers have changed but we show the observed source material location in all figures. We show the observed deposition location in Figure 6. The observed deposition and runout location can be inferred in figures 8 and 11.

Lines 691–692: Re-running the models and comparing results with the original, calibrated results is a great idea.

Thank you.

Line 704: This suggests the model is highly sensitive to accurate parameter representations; am I correct? This comment of mine again links to the regional applicability of the tool.

For sediment transport or topographic change assessments, that is correct, it does suggest the model is highly sensitive to accurate parameter representations. This is not unlike any other runout model. However, we note in section 6.1 that "In regions where landslide processes are relatively uniform (like the Olympic Mountain site), calibration to one landslide might be sufficient to predict the depositional patterns of another. At sites like the Cascade Mountain and Black Hills sites, which consisted of a diverse range of landslide processes including small, confined debris flows to large, unconfined debris avalanches, MWR may need to be calibrated to each type of landslide and predictive applications might involve applying the appropriate parameter set based on landslide type" (lines 694-699). For applications in which only the runout extent is needed, the model is less sensitive to parameter values. We added text in section 6.1 (lines 706 to 718) and to the conclusion (lines 790 to 799) to reflect this.

Lines 730–731: It would be nice to see a regional test application that includes multiple types of landslides in different catchments.

Thank you for the comment, we agree and will attempt this in future studies. We provide several examples from western USA in this study. Since this is the first paper that described MWR, adding a regional application would go beyond the typical size of a journal paper. Our paper is already long and we left this for future research. We added some text in the conclusion of our paper, stating that future studies will need to focus on regional applications to test the applicability of our model for regional hazard mapping.

Line 764: Does "option (3)" refer to the options from section 3.2 on page 17, or is it site (3)?

It refers to option (3) of the MassWastingRunout probability utility. We rewrote this sentence to be more clear (line 732 to 733)

Figure 12 illustrates a real-world application of the tool, demonstrating its potential for practical utility. Such an application is indeed an achievement. However, the application area of approximately 2km by 2km is relatively small and does not fully represent the broader/regional applicability that could showcase the tool's full potential. It would be highly beneficial if the next version of the manuscript could include a case study encompassing multiple catchments, where landslides follow diverse paths instead of converging into a single deposition zone. This expansion would not only highlight the tool's versatility but also its adaptability to various geological and environmental conditions, thereby reinforcing its value for regional landscape and hazard assessment studies. Such a case would also be nice to see in the testing phase of the developed tool.

Thank you for the comment. As noted in the reply, at this point, the focus of the paper and our work has been model performance in terms of ability to replicate observed runout phenomenon. In future work, we will apply the model to larger catchments and will look for terrains that include multiple flow paths and drainages. Thank you for the suggestion.

RC2 review replies

Thank you for the detailed and critical comments. Replies to each of your comments are listed below.

This reviewer repeatedly brings up several points in their review, all of which we truly agree. We feel that there's some philosophical differences in the approaches we took and the reviewer wants to see. We appreciate this opinion, and it is likely that a more sophisticated, process-based model can be superior to our approach when there is detailed information on debris flows and large amount of computing power available. However the point we are making is that with two parameters inferred from some field observations, our model can perform predictions with some skill. We feel that our model will have place in the ecosystem of runout models, and perhaps it can be used in a tiered-fashion along with other more sophisticated models for hazard predictions.

One of the points the reviewer brought up repeatedly is the fact that flow depth of the debris flow is not realistically calculated. We agree that our model does not represent the flow of sediment-water mixture and its properties. Instead the flow depth variable represents the volume of outgoing sediment from a cell divided by the cell area. As such, it grows rapidly as the debris flow moves downstream. Because the scour of debris was related to this "flow depth" via the shear stress equation, this depth needed to constrained to a field-inferred value to avoid unrealistic scour estimates. This field-inferred value comes from the field marks of debris flow. However, when we are calculating the cumulative volume along the debris flow, Qs, the flow depth is used as it (without any limit), obeying the conservation of mass principle. This choice of the model was made to develop as simple of a model as possible for regional applications. We are not claiming to have developed the "ultimate" debris flow model, this is just part of a research exercise to see if we can build a simpler model than other debris flow models such that it can be used for larger space and time scales.

The second issue the reviewer repeatedly brings up is the selection of grid size. We used a 10m of grid size across all sites because we wanted our final model parameters and results comparable across the different study sites. Selecting a 10m cell size was largely due to the availability of DEMs for differencing at each site, and partly the fact that in our field sites capturing debris flow behavior at 10m resolution was sufficient for regional hazard mapping applications. The cell size might have an influence on the final model parameters and model performance. This is also another issue that will need to be addressed in future research. By using the same cell size we are able to identify where and why the model doesn't perform well rather than simply showing that we can make the model work by calibration for different cell sizes. We think this finding is more helpful to potential users or others who are developing similar models. Finally, thank you for helping us find the variables that were not listed in the Notation section. We have added the missing variables to the Notation section or removed ones that were not necessary.

Line comments: line numbers in reply refer to the tracked changes version of the paper.

Line 20: mass flow is driven by topography but not the rules

Thank you, this sentence has been rewritten (lines 20-22)

Line 23: This sentence is not clear without further context

Thank you, this sentence has been rewritten (lines 25-26)

Line 70: If this is the FLATmodel reference than this contains to (2).

Thank you, the Medina et al., 2008 model has been moved to group (2) (line 67).

Line 101: It would be good to include a scale bar - at least approximately, so that you can estimate the width of the landslide. This is important because then you can see whether the model resolution of 10 m is sufficient for the selected landslides.

Thank you for the suggestion, I updated the figure to include a scale bar in each image and noted the scale bars in the caption (line 105)

Line 113: That's a strange sentence.Wy do models include references?

Thank you for catching the odd sentence, I corrected it. (lines 117 to 118)

Line 137: The use of t for iteration is unfavorable because t stands for time in almost all time-dependent models - which is not the case here.

Thank you for the comment, we chose t because i was already used to indicate node ID.

Line 153: As I understand it, this belongs in the main text and not in the figure caption.

qc is not yet explained and it is also missing in the table under 8.0 Notation.

t is also missing in the table.

Thank you, this text has been deleted. Variable t, which represents model iteration, is described in the text but I added a clarification in the caption (lines 158-162).

Line 155: n is not in the table

Thank you, I replaced n with q_o in the illustration to match caption and text description (lines 164-167).

Since n is no longer referred to in the figure or text it will not be added to the table.

Line 157: main text not figure caption

Thank you, this has been removed (lines 164-167).

Line 169: What does the , mean?

Thank you, I deleted the "," that was a typo (line 177)

Line 179: How was this observed (experiments) and how is the exponent a related to linear-viscous shear?

Thank you, I added more context to this statement to better communicate intended meaning (lines 183 to 185)

Line 187: How are the boundary conditions defined?

In this context, boundary conditions refer to the flux in and out of a single grid cell, but I changed to "terrain slope and convergence", to be more specific, as these control flux into a node (lines 196)

Line 200: not in the table

Thank you for the comment, this is a variable that is determined by the model so it is not included in Table 1. It is described in the Notation section (line 860).

Line 200: not in the table

Thank you for the comment, Sc is described in the Notation section (line 860) and an Sc value for each model run is shown in Figure 7 (line 539).

Line 211: Sn and n not in the table

should it be S or Si?

Thank you for the comment. Sn should be S. I removed Sn and changed to S and updated the figure (Figure 4, line 223).

Line 216: not in the table

Thank you, I added to notation section at line 860.

Line 216: not in the table there is only Ap,i

Thank you, we only need to label Ap,i and Ap has been deleted (line 229).

Line 219: not in the table

Thank you, this should be S. I changed to S and deleted Sn from the text (lines 223 to 246)

Line 224: hard to read from line 224 to line 238. I understand what you are doing but is it really necessary to break the process down to individual nodes?

In order to clearly describe the logic used to determine Na, we believe it is necessary to break the process down.

Line 252: with what? shear stress?

Thank you, changed to the non-linearity of h_e with shear stress (line 265)

Line 259: As I wrote in my last review, this is a pretty wild assumption

Thank you for the comment. We felt this was appropriate to keep the model parsimonious as possible while obtaining realistic results.

Line 273: an this is also a wild assumption

Thank you, our last reply applies to this comment as well.

Line 290: Which attributes are meant here?

Thank you for the comment, the attributes are now clarified in the text (line 303)

Line 290: Does this not result in all attributes tending towards the respective mean value in the course of the simulation?

Yes, the attribute value at a node would be the average (weighted by volume) attribute value of the initial regolith minus the eroded depth and the attribute values of all deposited material.

Line 294: not in the table

Thank you, added to Notation table (line 860)

Line 297: This section is difficult to read and I wonder if the equations are really necessary. It is quite sufficient to describe how this was solved. In my opinion, it is unnecessarily complicated.

Thank you, these equations would help the users to track the model code if they want to modify it.

Line 301: not in the table

Thank you, added to table (line 860)

Line 310: not entirely clear what is happening here

Thank you for the question, this describes a "work-around", or solution for dealing with pits or flat topography

Line 320: that are ...

Added text that hopefully clarifies what the parameters are (lines 325 to 337)

Line 323: is this really required?

Yes, distinguishing calibration iteration t and t+1 allows a clear description of the algorithm logic.

Line 323: random?

No, the jump distance is a function of the value of the L metric at time t and previous L metric, at time t-1. I added text to let the reader know a description of the jump size is provided later in the text (line 332).

Line 325: not in table - is it important?

Thank you for the comment, no, this parameter does not need to be specified with a letter, deleted (line 334).

Line 326: in which way? by a factor of?

Thank you for the question I clarified this statement at line 343.

Line 331: table description: omega metric - this is not helpful

Thanks, I improved the Notation description (line 860).

Line 331: explain

Thank you, I added text to direct reader to explanation (line 349)

Line 332: in total für the entire domain?

yes, added text to clarify (line 350)

Line 334: mean in cross-section? How is this computed?

Thank you, this is described in detail in the following paragraphs as written.

Line 339: I agree that this is a useful metric, but it doesn't take into account how large the maximum flow depth is - which the model can't do. However, it makes a difference whether the flow depth is 0.5 m or 5 m.

We are not using flow depth for calibration. We use extent, cumulative sediment export volume, and local change in landscape elevations. We clarified this by more clearly stating what Qs in the model (lines 366 to 369).. *Line 348:*

thank you for checking, these variables have been added to the notation table (line 860)

Line 351: not in the table

Thank you, added at line 860

Line 351: not found in the table

Thank you, added to the notation table (line 860).

Line 352:

Thank you, added to notation table (line 860)

Line 365: repeated

Thank you, corrected at line 385.

Line 386: I'm not sure if number_of is a valid mathematical operator. If you need this equation at all, you could probably solve it with sign() and a sum formula.

Thank you for the comment. We changed to the operator "#" (lines 405 to 408). The equation helps explain how probability of aggradation or erosion is determined.

Line 398: should that be Ds; at least not in the Table

Thank you, corrected (line 417)

Line 400: It's really good to see that the suggestion from the last review has been taken up. But I can't understand why this really rough resolution was expected here. The model domain has just 15 x 50 cells and the initial release area only covers 15 cells - why? It should be easy to calculate the whole thing with a resolution of 1m, 150 x 500 cells - especially because the model should have a very good performance

Thank you for the comment, we could have run the model at 1m resolution; however MWR is intended for watershed-scale applications and we suspect most DEMs will be 10m or coarser. Here we show that using a coarse DEM still produces reasonable erosion and depositional patterns.

Line 400 units missing (m?):

added units (line 419)

Line 403: Is the DEM of difference really exaggerated here or is the height model and the difference in height only shown in color?

Added text to figure to clarify that it is only the 3-D visualization that is exagerated. The shading is to scale. Line 420 to 423.

Line 405: I'm just wondering what effect the resolution has on lateral spreading.

Grid resolution affects the behavior of the multiflow direction algorithm and modeled spread would change unless the model is re-calibrated. I added a sentence to clarify this (lines 454 to 456).

Line 416: The problem, however, is that the flow depth is not meaningful in this model.

It is meaningful in the sense that it increases and decreases with entrainment or erosion but not meaningful in the sense that it does not reflect flow velocity and varies as a function of grid-size. I added a sentence at lines 200 to 202 to help clarify this.

Line 431: This is also a critical point because both speed and depth are derived and then also form the basis for erosion/deposition.

Thank you, we agree and acknowledge flow depths may not be correct but that flow does change as a function of flux (e.g., an increase in influx causes higher depth, lines 200 to 202). By applying the flow depth constraint in equation 3, the model works and produces realistic runout patterns.

Line 433: I'm not sure if this statement will make experts in the field of natural hazard management happy.

Thank you for the comment. It perhaps is an unusual approach, as many runout models are designed with the intention to accurately model flow depth and momentum. But that is not the intent of MWR. It is developed to predict the final runout extent, sediment transport and topographic change without necessarily correctly modeling the processes that created that final outcome.

Line 447: there are notes on the time schedule. Was it really a block release as shown in the model?

Yes, based on field and remote sensing interpretation of the initial landslide scar and runout path, the landslide body failed and mobilized in one instance (versus a slower, lower-flow-rate retrogressive failure).

Line 449: see comment above

We also believe that this landslide source material failed and mobilized in one instance.

Line 457: Block release oder hydrograph? The latter is not implemented but might be a better choice?

A debris flow hydrograph is not necessary because mobilization of the landslide body via the landslide polygon and algorithm 1 and then algorithm 2 control the order that material from the landslide is released into the channel.

Line 460: release area length? For me, the word landslide encompasses the entire process from release to deposition.

Thank you, changed to "initial landslide body" and clarified in text at 482 to 483

Line 460: comment above

Changed to "initial landslide body" and clarified in text at 482 to 483

Line 466: How was the maximum grain size determined if there are arrow bars? How many samples?

Grain size samples are described in the supplementary material. I added a sentence to the text to direct the reader there if they have a similar question (lines 509 to 510).

Line 468: not in the table

Thank you, added to the notation table (line 860)

Line 471: Why so coarse?

A justification for the 10-m grid size is provided later at line 572 to 573.

Line 473: Eq.

Thank you, corrected at line 495.

Line 498: repeated?

corrected (line 523)

Line 504: Unfortunately, this says very little about the model performance. How does the model scale with an increasing number of compute nodes? With a 1m resolution and a linear increase in computing time with nodes, this approach would already require 150 - 600 hours. However, due to the iterations in Algorithm 2, I can imagine a much less favorable relationship between nodes and computing time.

Thanks for the comment. this may be true but we also expect that, given that MWR relies on a simple, slope-based approximation of the shallow water equations, if the code proves to be slow, we will refactor the code to improve speed. In the conclusion, a statement has been added that notes that we will evaluate regional MWR applications in future studies.

Line 512: There are to many plots. In the print out, labels are tiny. Lines in (a), (b), (c) ... are hard to see.

I increased the axis label size (line 538)

Line 512: why \theta - relic of the old version?

corrected, line 538

Line 512: No axis labeling

axis labeling for histograms is described in the caption, line 544

Line 512: No axis labeling of the vertical axis

added a vertical axis label, line 538

Line 512: Unlike described in the figure caption, the horizontal axis is labeled with qSc here. What does this mean now?

Corrected to be q_c, qsc was the old version, line 538

Line 543: but then, why not computing with a better resolution?

We wanted to use a consistent grid resolution, see reply below.

Line 545: I cannot agree with this general statement. It depends on the scale, and a 10 m DEM is clearly too coarse, especially for small landslides.

Thank you for the comment, part of the reason for using a consistent 10-m DEM is that it allows for apples-to-apples comparison of model performance across sites. We are able to identify where and why the model doesn't perform well rather than simply show that we can make the model can be made to perform well.

Line 551: I think this part belongs in the discussion section on model limitations

Thank you, we reiterate this finding in the concluding statements (lines 775 to 766).

Line 552: which applies to almost all rapid mass movements

This sentence is misleading. We changed the sentence to state "where flow momentum is the primary control on runout extent" (line 579). You are correct, momentum does control all rapid mass movement but it is often not the primary control, as demonstrated at the 6 field sites in this study where most of the flow path varied as a function of the total mobilized volume and generally followed parallel to slope.

Line 573: This figure is rather chaotic, mainly because the maps are not uniform and small diagrams have been inserted. These are almost impossible to read in the printout.

Qs is not in the table. I understand that this characteristic (sediment transport pattern?) can be determined from the model solutions, but how were the actual values determined here? Perhaps I have overlooked it?

I increased the fontsize of some of the Qs labels, added text to direct the reader to where Qs is explained, and clipped off whitespace from the edge of the figure so that it can be made as large as possible (line 601).

Line 573: Why the unit (m) in capital letters

changed to m (line 599).

Line 573: Please explain more clearly how you arrive at reproducible values here.

The figure caption now clarifies that the plots show both observed and modeled Qs and directs the reader to how Qs is computed (lines 600 to 604).

Line 573: Vertical axis labels are missing

added text to caption that notes that y-label is indicated on the plot (line 602).

Line 578: Perhaps the legend could be placed a little more "lovingly".

moved to a better spot (line 608).

Line 578: I do not understand this illustration. The authors themselves write that the "temporal" development has no meaning in the model, only the final pattern with erosion / deposition and extent. But now a "temporal" development figure is shown and also explained?

The point of the illustration is to show how the runout material is released from the landslide polygon (Algorithm 1) and demonstrate MWR response to topography (flow around ridge) as well as highlight where the model may not perform well, as shown by the model's inability to replicate slope-perpendicular flow over the bench. The caption has been re-written and a statement added to the text to emphasize this point (line 609 to 617).

Line 578: not reproduced in the model

Those apparent runout scars are actually roads. They are now labeled and mentioned in the caption so other readers will not misinterpret the image (line 616).

Line 582: main text not figure caption

Thank you, deleted (see caption at lines 613 to 614)

Line 589: not in the table

Thank you, added to notation section (line 860)

Line 589: not in the table

Thank you, added to notation section (line 860)

Line 590: not in the table

Thank you, added to notation section(line 860)

Line 595: Wouldn't it be logical to use stream power or ksn instead? I think that should come out-of-the-box with Landlab.

Thank you for the comment. We found these metrics were most responsive to differences in runout topography

Line 603: Again, if you can see that the resolution does not sufficiently describe the terrain - why not use a higher resolution? I just don't get it.

Part of the reason for using a consistent 10-m DEM is that it allows for apples-to-apples comparison of model performance across sites. We are able to identify where and why the model doesn't perform well rather than simply show that we can make the model can be made to perform well.

Line 608: I don't think that this is a good approach.

It may not be, but it is the approach that almost all runout models use to compensate for processes not included in the model or a DEM that doesn't perfectly represent real-world conditions.

Line 614: unit missing

Thank you, added (line 651)

Line 641: However, these models do not promise that a calibration is valid for other catchments.

We do not promise that a calibration to one site is valid for another but speculate that if landslide processes are relatively uniform for a specific region (part of a watershed underlain by the same geologic unit perhaps), then a single parameterization may work for that region or if the model objective is runout extent, acceptable results may be achieved with a less rigorous calibration

Line 647: Of course, because entrainment is a very complicated process that is not yet well understood, even for "relatively simple" snow avalanches. Estimating the thickness of the regolith alone is hardly feasible for large areas and even within this layer the spatial variation is huge.

Your comment relates to the point we are trying to make: Since only a few of the processes that control the behavior of a debris flow are physically described in detailed-mechanistic models and since accurate representation the model domain (e.g., regolith thickness) is generally not feasible, the predictive ability of the detailed mechanistic models is still reliant on parameterization. If the end goal of the model is prediction of the runout extent or topographic change caused by the runout, then MWR, which is also a runout model that relies on parameterization to produce accurate prediction, is a suitable model.

Line 651: (m) not (M)

Thank you, done (line 687).

Line 651: That is honest and I find it very honorable. Unfortunately, it also shows that in some cases the promised calibration fails completely.

We don't want to promise more than can be delivered and this figure helps guide discussion on when a new calibration is necessary versus when using the calibration of one landslide to predict runout at another is appropriate

Line 672: this has to be proven. Wouldn't it have made sense to test the code in these areas first?

In hindsight, probably, but calibration and evaluation of MWR will be an ongoing process, larger than a small section in this paper.

Line 684: That's just way too much now. Why does this part still have to be presented in this article?

This section (Section 6.2, line 724) is relatively short and demonstrates both practical uses of the model and an additional advantage of using the calibration utility so we would like to include it.

Line 688: not in the table

Thank you, added to notation section (line 860)

Line 701: Eq.

Thank you, done (line 748)

Line 757: There are no references to this section and the table is incomplete.

The Notation table has been updated to include all variables, parameters and indices and is referenced at lines 138 to 139A number of the variables were described in a section of the supporting information. That section has been moved to an appendix to help link with the notation (Lines 799 to 827) table to that text.