Mergili et al., r.randomwalk v1.0, a multi-functional conceptual tool for mass movement routing

Response to Reviewer #1 (Jörg Robl)

We would like to thank the reviewer for his constructive remaks really helping to improve the paper. Below, we address each comment in full detail. Our responses are written in blue font.

The manuscript was revised accordingly. In addition, we have repeated all computations with an updated, bug-fixed version of r.randomwalk. As a consequence, some of the results have slightly changed, but the general findings remain the same, compared to the discussion paper. The new version is referred to as version 1.1, therefore "r.randomwalk v1.0" is changed to "r.randomwalk v1" in the title of the revised manuscript. All changes are highlighted in yellow colour in the revised manuscript.

Mergili et al., present a new open source tool to describe a variety of rapid mass movements by applying a random walk approach. Such codes have been and still are frequently used to assist natural hazard projects. In contrast to other codes, the presented tool offers a variety of additional options: the code allows different types of break criteria, account for the uncertainties, performs advanced statistics, allows the back-calculation of a set of observed mass movements and provides nice visualizations of the results. To achieve this functionality the code is integrated in the GRASS GIS framework and uses R.

Although not fully convinced by the random walk approach itself, the presented tool unifies the benefits of different random walk approaches presented by others in a free and open GIS and add additional functionality. Hence, this code will be a valuable tool for the natural hazard community. The functions and methods are well described and the test areas differ significantly in terms of occurring rapid mass movements and model requirements, so that a wide range of model functions are successfully applied.

To be short, the test cases are very well chosen. The figures are informative and of high quality. The code and some datasets for testing are freely available from the website of the first author. The installation requires the compilation of GRASS GIS from source and additional hacking to get the r.randomwalk code correctly installed (at least on my rather old CentOS6.5 installation). This seems to be an obstacle and I wonder if it is possible to use the add-on system of GRASS GIS for releases in the future. The code itself worked flawlessly for the first test case and it is by far the most flexible and advanced random walk code for flow routing I have tested during the last ten years. Overall, I suggest publishing this study in GMD after minor revisions.

The reviewer is absolutely right that an easier way to make the code working is highly desirable. It is definitely envisaged to release the tool as an official add-on to GRASS GIS. However, some more bug-fixing and testing is required to do so. The code is steadily improved and a new bug-fixed and slightly extended version, compared to the one referred to in the Discussion paper, has already been released. The new version is referred to as version 1.1, therefore "r.randomwalk v1.0" is changed to "r.randomwalk v1" in the title of the revised manuscript.

General Remarks: Every random walk approach requires defining several “unphysical” parameters like Rmax (user-defined maximum vertical run-up height) or Lseg, Lctrl (to avoid uneven flow paths) and several others. These parameters may lead to flow directions that are directed backwards or in a circular way if they are not set appropriately. Physical-based models that describe the motion of fluid by a depth averaged form of the Navier-Stokes-equations and an appropriate flow resistance law require less parameters and are in my opinion better constrained (although uncertainties are still large). The biggest advantage of a random walk approach over physical based models may be the computational performance of random walk codes that allow “Monte Carlo Simulations” and the exploration of the crucial parameter space. However, additional parameters to perform a random walk approach (see above) increase the parameter space dramatically. Sophisticated numerical methods (e.g. AMR: adaptive mesh refinement) in concert with modern and fast computers reduce to computational costs for physical based models and parameter studies can also be performed. Despite of that, these models provide us with flow height, velocity and momentum, travel time and other parameter that are important for mitigation strategies against natural hazards. Thus, I expect that random walk approaches describing natural hazards will vanish in the next decade(s) but in the meantime the code of Mergili et al. will probably the best choice.

Future will show whether physically-based models will “outcompete” conceptual ones. We are not so sure about that, but we acknowledge that this issue is an important point of discussion (which would, however, go beyond the scope of the present article). It is absolutely true that the parameters for r.randomwalk are unphysical and have to be optimized using a number of test cases. However, the parameters fed into physically-based models are often unphysical, too. Even though it is pretended that they are physical, in practice they are optimized in terms of predicting the observation in the best possible way. Therefore, they represent statistically or heuristically derived rather than really physical parameters. In this sense, conceptual models are used in a more “honest” way as it is admitted that the parameters are unphysical … Nevertheless, physically-based models are highly useful for many purposes which conceptual models cannot serve for in an appropriate way. We have currently running a dissertation project where we test whether conceptual or physically-based models better predict observed deposition areas, using comparable evaluation strategies.

After reading the manuscript several times I have the impression that the separation in chapters “3 Test sites and model parameterization” and “4 Results” is somehow confusing and follows the design layout of a classical natural hazard study. Maybe I’m wrong but I think it would be less confusing by merging chapter 3 and chapter 4, call it for example “3 Test cases” and stitch the model setup, the parametrization and the results of each test site together that will get the subheading 3.1, 3.2,3.3. However, this is just a suggestion.

Thank You for this comment. We have followed the suggestion and merged the description, parameterization and results for each of the test sites. Indeed, this makes reading of the paper easier. The name of the new chapter is “Test cases and results”. The section for each case study is split into two sub-sections named “Area description and model parameterization” and “Results”.

There are many terms and abbreviations that require an exact definition. Some of them are defined in the context where they are used the first time; some are not defined at all (common knowledge?). Maybe the authors should implement one additional paragraph defining the most important terms (impact indicator score, release indicator score, impact indicator index, exposure indicator scores, impact probability raster map PI). All these terms and even worse their abbreviations together with a couple of dozen of variables are really confusing (at least for me). In this context I encourage the authors to implement a new table 1 summarizing all abbreviations and variables with a short explanation (maybe instead of an additional paragraph?)

We have added a new Table where all the key variables are explained.

Please implement one or two sentences on the computational performance of your model. How long will it take to perform each of the test cases on standard computer hardware?

r.randomwalk automatically returns the computational time. The computational times associated to some of the test runs are addressed in the revised manuscript.

Line by Line Remarks

P8194 - L10: parameter settings ! parameter sets?

Yes, “parameter settings” was changed to “parameter sets”.

P8195 – L5-L13: I do not fully agree with your statements here (especially the first sentence). Modelling rapid mass movements with numerical tools is always affected by large uncertainties related to initial parameters and fluid properties. However, I do not see why random walk models with empirical break criteria are “better” constrained than physically based models and should therefore applied in regions with sparse input data. Instead of calibrating parameters of flow resistance laws for physical based models, break criteria have to be calibrated (and a bunch of other parameters) for random walk models. In my opinion the main advantage of models based on a random walk approach may be that they cost less in terms of computational time and allow therefore (a) modelling many rapid mass movements on valley scale and (b) the exploration of a large variety of parameter sets without employing a computer cluster.

Besides the issue of computational times, which is certainly important, there is in our opinion an additional point making conceptual (random walk or multiple flow direction) models preferable particularly for large areas:

A comprehensive set of literature is available with regard to published angles of reach of various types of mass flows. Such relationships allow to extract statistical relationships along with uncertainty measures in a very transparent and reproducible way. Trimming physically-based models in a way that they correspond to the observation, forward calculations with the back-calculated geotechnical and fluid parameters may also yield useful results, but the parameters do not necessarily have a direct physical meaning then. Even though r.randomwalk parameters such as Lseg, Lctrl, Rmax, fbeta and fd, which also have to be optimized by back-calculation, are unphysical, their role in influencing the model results is quite straightforward and it is, in our opinion, better justifiable to use them particularly for broad-scale studies than geotechnical or fluid dynamics parameters.

P8197-L2-4: “The parallelization procedure is implemented at the python level (analogous to the way described in Mergili et al., 2014) and serves for two purposes” First, I think that you should try to explain your parallelization approach in a few sentences. Model development is in the heart of this journal (although the approach has been already published in Mergili et al., 2014.).

Second: The main reason for parallelization is the reduction of the computational time for Monte Carlo simulations, I assume.

You are right. We have added a brief explanation of the parallelization procedure and emphasized that the main reason for its implementation is the reduction of computational time.

P8197-L5-6: This statement is not clear to me. Maybe you can rephrase it.

We have rephrased this statement in the following way: “Analyses with multiple random subsets of the release areas or coordinates. In each model run, one subset is used for back-calculating the probability density function (PDF) of the angle of reach, the other subset is employed for validating the distribution of the impact probability derived with this PDF against the observed deposition areas.”

P8200 – L15: IIS: was not explained before

IIS is now explained in the newly added Table 1.

P8200 – L16: parameter sensitivity tool AIMEC: Please explain the functionality of this tool and also the abbreviation.

AIMEC stands for “Automated Indicator based Model Evaluation and Comparison”. The full name of the model is given in the revised manuscript.

P8200: L21.24: This is not clear to me. What do you mean with model development in this context?

The term “model development” was used for “parameter optimization” which, indeed, was not a good choice. The term “parameter optimization” is used in the revised manuscript.

P8200: L27 - : This is a cool feature!

Thank You! ☺

P8201: L7 - : IF not defined before.

IF is now explained in the newly added Table 1.

P8202: ROC – although this is common statistics, it should at least be explained by writing out the full term once.

The full name of the ROC (Receiver Operating Characteristics) procedure is given in the revised manuscript along with the brief explanation of the method.

P8202: L8: Maybe all these abbreviations (e.g. AUCROC) are frequently used by the natural hazard community. However, too many abbreviations hamper reading this paper and I wonder if all of them are really required for this study.

P8203-L16: using an ROC Plot ! using a ROC Plot

“an ROC Plot” is correct (at least as long as the three letters R, O and C are pronounced separately).

P8203-L5-: “We assume that the values of nruns, Rmax, f_, fd , Lctrl, Lseg and the pixel size applied to the Acheron Rock Avalanche (see Sect. 3.1) are valid also for this study area.” Well –

I just wonder how you can assume that. Does this mean that we can assume this for all types of rapid mass movements worldwide ;-)? In this case I wonder why we need nruns, Rmax, f_, fd , Lctrl, Lseg as free parameters.

You are correct that this issue requires a more thorough investigation. Whilst, in our opinion, such an investigation is out of scope of this paper, it is the topic of ongoing research. Therefore, the formulation has been changed to “We employ the values of Lseg, Lctrl, Rmax, fbeta and fd resulting from the optimization procedure for the Acheron Rock Avalanche“, and address the issue in the discussion in a more appropriate way, compared to the discussion paper.

P8203-L14: How did you identify” the Gaussian distribution as the most suitable type of distribution for this purpose”? Please explain in one or two sentence(s).

This was done by just testing and visually comparing the Gaussian against the Log-normal distribution – it is briefly explained in the revised manuscript. However, in the updated version (with the tangent of the angle of reach employed), it is the log-normal distribution which suits best.

P8208-L19-L22: I’m not sure if it is a good idea to use degrees for statistics. In the light of the non-linear relationship between percent and degrees - what will happen by averaging slope in degree and what can we learn from the STD in degree? This seems to be dangerous for rough flow paths with alternating high and low gradient flow segments?

This is a good point. We have changed the r.randomwalk code in the way that it uses the tangent of the slope instead of the slope itself, recomputed the pdfs and the resulting impact probabilities, and updated the text accordingly.

8213 – L26: “Krenn et al., submitted”: Paper is missing in the list of references.

Thank You for the remark, this conference paper has been added to the reference list.

Table 2: What are the numbers in the brackets after the formulas?

These numbers represent the Equation numbers.

Mergili et al., r.randomwalk v1.0, a multi-functional conceptual tool for mass movement routing

Response to Reviewer #2

We would like to thank the reviewer for her/his remark.

I find the paper of Mergili et al. very valuable for both the scientific quality and the possibility to apply the model freely as open-source.

We are glad to hear that the reviewer finds the paper valuable! The manuscript was modified in some points according to the remarks of the other reviewer. In addition, we have repeated all computations with an updated, bug-fixed version of r.randomwalk. As a consequence, some of the results have slightly changed, but the general findings remain the same, compared to the original manuscript. The new version is referred to as version 1.1, therefore “r.randomwalk v1.0” is changed to “r.randomwalk v1” in the title of the revised manuscript. All changes are highlighted in yellow colour in the revised manuscript.