This is one of the most interesting and insightful studies I have read over the past few years. First, I would like to give a high respect to the heavy workload and computation efforts that the authors endeavored to produce such a meticulously designed and solid results. To be honest, it took me a couple of days to really finish reading through this excellent work. Very impressive.
As one of the most advanced techniques, cellular automation becomes increasingly popular in modeling soil erosion processes. In the Introduction part, the authors did a good job in reviewing and summarizing the state-of-the-art of cellular automation in the field of soil erosion modeling. This therefore makes it quite easy to comprehend the knowledge gaps in current studies, and in turn not difficult to understand the novelty of this study. Yet, I reserve some concerns that the authors may consider to include in the revised manuscript.
Thank you very much for your encouraging words and thank you very much for your comments and suggestions for change, which helps a lot to improve our manuscript. We are addressing each of your comments below. We are particularly grateful to you for taking two whole days in order to read thoroughly: thank you again!

- The opening of the Abstract seems a bit too long, and key research questions or knowledge gaps come in a bit too late. It would be better if the authors could specify the research question of this study more explicitly, and also save more writing space to present the key findings in the following part. Apart from the methods, the results and observations are a bit too general. More specific data or quantitative descriptions would be more helpful to underscore the novelty.
- Thank you for your suggestions. However, respectfully, we disagree regarding the opening of the Abstract. Given the current dominance of wholly empirical RUSLE-GIS approaches in soil erosion modelling, we submit that it is necessary to point out the "need for erosion models, necessarily process-focused, which are able to reliably represent rates and extents of soil erosion under unprecedented circumstances."
  Again, with respect, we disagree regarding the need to specify the research question of this study more explicitly. On line 15 of the abstract we state the research objective: „This study explores the use of Structure-from-Motion photogrammetry as a means to calibrate and evaluate this model". We suggest that this is sufficiently explicit. We have however inserted line breaks in the rather monolithic Abstract, to hopefully make it easier to spot the research objective during a quick scan of the Abstract.
  It is true that we wait until line 120, at the end of the Introduction, to state the research objective in the main body of the text: „The aims of this study were first to use several multi-objective functions to calibrate and evaluate a process-focused soil erosion model (RillGrow), and then to evaluate these objective functions in terms of information gained from each function." We chose to do this because it seems essential to us to summarise the context for this research before stating the research objective.
  Also, we are afraid that we are unclear regarding your suggestion that our results and observations "are a bit too general". Surely any research necessarily focuses on the specific to make inferences about the general? Put another way: we can't test everything, so we have to choose a test case and then extrapolate out from that.
  Finally, (and we again ask this with all due respect!) what exactly do you mean by "More specific data or quantitative descriptions would be more helpful to underscore the novelty"?

- In general, the Results part is a bit too long and saturated with figures. I would suggest remove some figures into the supplementary, or selectively show some of the subfigures. I believe, there must have been far more figures plotted out during the entire modelling and analyzing. The authors must have already tried a lot to reduce the number of illustrations. Yet, still, as a piece of a regular research article, too many figures and too lengthy results somehow might make the readers feel overwhelmed. Is it possible to include a table listing out the major performances of different model runs under the three different approaches? Or, the authors may even consider to develop a sort of "graphical abstract" or "conceptual diagram" to summarize the research questions, critical methods and key findings? So as to help the authors develop a "holistic" comprehension over this study?

- Indeed, the result descriptions are long. However, we also performed various comparisons. However, we will reconsider each description of our findings to evaluate whether it could be shortened and/or moved to the supplement.
  Unfortunately, a table summarizing the results and describing the performance would also be too long and would not bring more clarity. For follow-up work, we would like to refer to our provided raw data including python scripts to run the analysis (including a short description on how to use them) and thus being able to have an even more detailed insight.
  The graphical abstract / conceptual diagram is a great idea and will be provided.

- In addition, there are always bits and pieces of discussion mixed in the Result part. This on the one hand makes the Results part quite lengthy; on the other hand, in the current state, the Discussion part is more into "limitations and future implications", but short in in-depths explanations, coherent arguments and discussions with other peer studies (most of which is actually scattering in Results).
- We check the manuscript and revise it accordingly to make the results and discussion more coherent.

- L580 to L595 in the Discussion section is actually a review over currently available models. They should be moved to the Introduction part, to better specify the knowledge gaps of current studies.
- We move the section accordingly.

- Some of the results were described in present tense. Should they be in past tense? For instance, L485 to L515 in subsection 3.2.
- Thank you for noticing. We will revise this.

- Although the authors mentioned the specific subprocesses, such as raindrop detachment, splash transport, flow transport and flow detachment (mostly derived from Kinnell 2001), the potential effects or impacts of these subprocesses were not adequately discussed. For instance, the selectivity of runoff over eroding time in carrying soil particles of different sizes. This may partly contribute to the unmatched temporal variations of sediment yield.
- Thanks! This is indeed something that we wished to focus on more when writing the paper. However, the paper is already rather long (as you yourself point out). Lengthening the paper to include this extra discussion would not be a good idea. Also, we have learned a good deal from this work and, as a result, the RillGrow model is now being modified to better represent some of these erosional subprocesses (splash redistribution in particular). Therefore we suggest that discussion of the effects of strengths and weaknesses in RillGrow's representation of these subprocesses, and their implications, would be better kept back for a follow-up paper.
- Furthermore, this paper focuses primarily on modelling the development of rills. During the simulations, the grain composition of the sediment was measured and not

only the grain composition but also the disintegration (by ultrasound) of these particles was evaluated (methodology described in Kubinova et al., 2021). Based on our results so far, which have been carried out on steep slopes where erosion rills have formed, the grain size distribution of the eroded soil does not differ significantly from that of the original soil throughout the simulation. The generally accepted concept of selective erosion was not confirmed in these experiments, and the use of a model could help to refine erosion processes in the future. However, this "sub-topic" is beyond the scope of this manuscript.

- L615, the statement on "erosion model calibration might use EC-based measures only, and even possible without using sediment yield" is somewhat a bit bold, I think. That the relatively smaller errors of EC-based approaches were valid in this study, at least to some extent, was because the soil surface was prepared with compaction and heavy bulk density. Some minor changes in EC, such as the settling of soil surface after wetting, the removal and in turn runout of depositional sediment over time, the periodic initiations of different rills and rejuvenated eroding surface, and the progressive equilibrium of runoff and sediment in the intervals of rill development, may trigger major changes in sediment yield.
- Indeed, the removal and in turn runout of depositional sediment over time, the periodic initiations of different rills and rejuvenated eroding surface, and the progressive equilibrium of runoff and sediment in the intervals of rill development, can trigger major changes in sediment yield, which however should be measurable by the DoDs as these processes result in direct changes of the soil surface height. However, with the settling, this is a different picture. This needs to be considered differently. Here, we already investigate, how approximations of settling processes might help to disentangle this process from erosion or at least indicate when both processes are happening at the same time (Epple et al., moderate revision, Soil and Tillage Research). In this case sediment yield measurement is needed if indeed erosion is to be measured. We made our statement less bold in the revised manuscript.

Overall, this is a well-written manuscript offering a huge amount of information and new thoughts. Yet, I think it would be even better appreciated, if the authors may consider to trim and condense it a bit shorter (even just for the sake of less APC 😊).
Thank you for your comment. We shorten the manuscript accordingly.

References:

Beven, K.: Towards a methodology for testing models as hypotheses in the inexact sciences, Proc. R. Soc. A Math. Phys. Eng. Sci., 475(2224), doi:10.1098/rspa.2018.0862, 2019.

Beven, K. J.: Environmental Modelling: An Uncertain Future, Routledge, Oxon., 2009.

Beven, K. J.: Rainfall-Runoff Modelling, 2nd ed., John Wiley & Sons, Chichester., 2012.

Beven, K. J.: On hypothesis testing in hydrology: Why falsification of models is still a really good idea, WIREs Water, 5, e1278, doi:10.1002/wat2.1278, 2018.

Beven, K. J. and Young, P.: A guide to good practice in modeling semantics for authors and referees, Water Resour. Res., 49(8), 5092–5098, doi:10.1002/wrcr.20393, 2013.

Brazier, R. E., Beven, K. J., Anthony, S. G. and Rowan, J. S.: Implications of model uncertainty for the mapping of hillslope-scale soil erosion predictions, Earth Surf. Process. Landforms, 26, 1333–1352, 2001.

Cândido, B. M., Quinton, J. N., James, M. R., Silva, M. L. N., de Carvalho, T. S., de Lima, W., Beniaich, A. and Eltner, A.: High-resolution monitoring of diffuse (sheet or interrill) erosion using structure-from-motion, Geoderma, 375(May), 114477, doi:10.1016/j.geoderma.2020.114477, 2020.

Fischer, F. K., Kistler, M., Brandhuber, R., Maier, H., Treisch, M. and Auerswald, K.: Validation of official erosion modelling based on high-resolution radar rain data by aerial photo erosion classification, Earth Surf. Process. Landforms, 43(1), 187–194, doi:10.1002/esp.4216, 2018.

Jetten, V., Govers, G. and Hessel, R.: Erosion models: Quality of spatial predictions, Hydrol. Process., 17(5), 887–900, doi:10.1002/hyp.1168, 2003.

Kaiser, A., Erhardt, A., Eltner, A.: Addressing uncertainties in interpreting soil surface changes by multi-temporal high resolution topography data across scales. Land Degradation & Development, 29(8), 2264-2277, doi: 10.1002/ldr.2967, 2018

Klemeš, V.: Operational testing of hydrological simulation models, Hydrol. Sci. J., 31(1), 13–24, doi:10.1080/02626668609491024, 1986.

Kubínová, R., Neumann, M., Kavka, P.: Aggregate and Particle Size Distribution of the Soil Sediment Eroded on Steep Artificial Slopes. Appl. Sci, 11, 4427. doi: 10.3390/app11104427l, 2021.

Onnen, N., Eltner, A., Heckrath, G., Van Oost, K. (2020): Monitoring soil surface roughness under growing winter wheat with low altitude UAV sensing. Earth Surface Processes and Landforms, 45(14), 3747-3759

Oreskes, N.: Evaluation (not validation) of quantitative models, Environ. Health Perspect., 106(6), 1453–1460, doi:10.1289/ehp.98106s61453, 1998.

Oreskes, N., Shrader-Frechette, K. and Belitz, K.: Verification, validation, and confirmation of numerical models in the Earth Sciences, Science (80-.)., 263, 641–646, doi:10.1126/science.263.5147.641, 1994.

Saggau, P., Kuhwald, M., Hamer, W. B. and Duttmann, R.: Are compacted tramlines underestimated features in soil erosion modeling? A catchment-scale analysis using a process-based soil erosion model, L. Degrad. Dev., 33(3), 452–469, doi:10.1002/ldr.4161, 2022.

Takken, I., Beuselinck, L., Nachtergaele, J., Govers, G., Poesen, J. and Degraer, G.: Spatial evaluation of a physically-based distributed erosion model (LISEM), Catena, 37(3–4), 431–447, doi:10.1016/S0341-8162(99)00031-4, 1999.

Vigiak, O., Sterk, G., Romanowicz, R. J. and Beven, K. J.: A semi-empirical model to assess uncertainty of spatial patterns of erosion, Catena, 66(3), 198–210, doi:10.1016/j.catena.2006.01.004, 2006.

Warren, S. D., Mitasova, H., Hohmann, M. G., Landsberger, S., Iskander, F. Y., Ruzycki, T. S. and Senseman, G. M.: Validation of a 3-D enhancement of the Universal Soil Loss Equation for prediction of soil erosion and sediment deposition, Catena, 64(2–3), 281–296, doi:10.1016/j.catena.2005.08.010, 2005.