Response to Pedro Batista – Epple et al. In Review at SOIL

The authors thank Pedro Batista for his extensive and highly regarded comments. In the following, we offer suggestions on how to improve the manuscript and answer the referee's observations:

R = Referee's comment

A = Authors' response

R: For instance, I had a very difficult time reading the paper. Sentences are often long, inadequately punctuated, and disconnected between themselves. The exaggerated number of superfluous citations makes the text convoluted, and demonstrates an excessive reliance on previous review papers (e.g. Pandey et al., 2016 gets cited over 20 times). Ultimately, there is a lot of repetitive information without sufficient reflection and interpretation. I suggest summarizing each section more concisely and then focusing on your actual hypothesis.

A: We will adapt the manuscript with these insightful suggestions. We will shorten paragraphs, rephrase our statements with more purpose, and focus on our hypothesis. Furthermore, we ask a native speaker for an extensive proof reading.

R: In particular, there could be a section where you explain why, after analysing models, processes, and new measurement techniques, you corroborated your initial hypothesis about models not keeping up with the data. I would also recommend trying to answer why models cannot keep up with the data. Is it a software issue, a programming issue, a process description issue? Or is it a problem with the modellers and the science? I think these questions are crucial for advancing the discussion and bridging the gap between models and data.

A: We would like to add more ideas about this to individual paragraphs and include another paragraph, summarising our discoveries on this topic to the conclusion. Regarding the models' shortcomings we see a big challenge in the process description, its implementation by the different models, as well as in the software itself. We would like to add the following paragraphs to our conclusion:

"While models vary according to scale, resolution, process description, parameterization, and result, there are certain models which are better at implementing or make use of new and improved data. As established in literature, many process-based soil erosion models were developed for the field or even the catchment scale. By ignoring the smaller scales, the small-scale processes preparing the surface for the next level processes are neglected. While researchers are already using high resolution data in process-based soil erosion models, it is difficult to evaluate whether the models can actually handle these resolutions correctly. A change in resolution is accompanied by a change in the prevailing processes. However, the choice for the "right" resolution, scale, and number of input parameters depends on the research question. It varies already by, for example, looking, at a whole catchment or just a sub-catchment, and therefore represents a decision that has to be made individually. The various models developed for different sites cope with the various soil erosion processes differently well. One of the processes that all these models struggle with is the spatial and temporal initiation of rills. While researchers, due to high resolution soil surface changes, improve their understanding on these processes, implementing this knowledge must be a future goal. Modellers and developers would need to adapt models accordingly. Evaluating the scale dependent boundaries of processes, researchers should strive to include the initial development of rills and enable cross-scale modelling.

Most models have not yet been developed for so many and so much data and therefore cannot make use of the available parameters. There is huge potential in remote sensing to further develop process *descriptions, assess parameters as topography, roughness or flow velocity with high temporal and spatial resolution, or to work across scale. Techniques with low cost, low time expenditure, and high resolution, show potential to gain adequate data from the micro to the macro scale. Of further interest are [machine learning] ML approaches and tracing techniques. They pave the way to respond to different processes on different scales (splash-, sheet-, rill-, gully erosion, transport, and deposition). ML and automated assessment systems, could even offer opportunities on a completely new level, enabling the development of fully automated modelling approaches in the future. Process based soil erosion models need to be further developed and improved in terms of both process understanding and programming. In conclusion even though the data has been extensively developed over recent years, in many cases the models are not able to keep up with the data anymore. Therefore it is necessary to properly develop and adapt these models.“*

R: On a more specific topic, I would like to strongly encourage the authors to reflect upon their notions of models and realism. A good starting point would be Chapter 2 (A philosophical diversion) of Environmental modelling: Un uncertain future? (2009), by Keith Beven.

A: We would like to elaborate on this topic at the start of chapter 2.1 as follows.

"Models as simplifications of reality can by definition never fully represent the processes of the real world. They can only approximate reality to a certain degree and so we strive to find a balance between modelling and reality. Depending on what the models focus on, they feature different advantages and limitations. With increasing knowledge, computing power, and observational technologies these limitations constantly shift towards models that are more and more realistic. Nevertheless, models are always limited regarding realism. Where to exactly draw this line is more of a philosophical question and differs from model to model."

R: I see a lot of potential here and I look forward to seeing the authors exploring it accordingly. Hence, I would be glad to review this again if i) the manuscript goes through a thorough language and style revision; ii) repetitive or superfluous content is removed and the main findings from previous research are concisely summarised; iii) the authors deepen their discussion and make justice to their interesting and novel research idea.

A: We will adapt our manuscript according to the many helpful comments and the conditions for a renewed proofreading. We will ask a native speaker to comment on the vocabulary, grammar and general clarity of the written English. Furthermore, we will revisit the subchapters and paragraphs, restructure them, concisely summarise the information, check the content for redundant parts, and discuss our results more in-depth in respect to our main hypothesis. Especially in chapter 3, we would like to expand on the topics we discuss and phrase our results clearly.

We would also like to thank Pedro Batista for his extensive comments and like to include all minor revisions. In a revised version, we will correct any spelling and punctuation mistakes, properly connect paragraphs, and rephrase any content errors. In the following, we would like to react to some of the referees' comments made within the manuscript:

R (line 51-52): I suggest introducing the topic of process-oriented erosion models before pointing out their limitations.

(line 53-55) Could empirical models be subject to the same limitations? If I might add, I think a major limitation of process-based erosion models is that, in fact, they often have an empirical basis. This is discussed extensively in Parsons (2019), which you refer to below.

A: Thank you for pointing out this structural issue. We would like to briefly introduce process-based soil erosion models in the introduction and place the limitations common to many of the models in chapter 2.1. It is true that process-based and empirical soil erosion models have certain limiting factors in common, however we would like to focus on the process-based models. In order to clearly distinguish between the two model types, we would like to add a few sentences about their differences.

R (line 64): I found this weakly introduced. What data? Why would models not be able to keep up with the data? This becomes clearer below, but I suggest to improve the description of your hypothesis here.

A: A valid observation. We would like to change this part and justify our hypothesis in the introduction as follows:

"Increasing computing power is constantly pushing the limits of data collection and processing and new opportunities to assess and use soil data and understand soil erosion processes become available. The continuous development and improvement of measurement techniques leads to spatially and temporally highly resolved information at different scales (Li et al., 2017).

This raises the question which of the formerly limiting factors, both in terms of model capabilities and data availability may have become obsolete. Can new and modernized measurement techniques, novel data on soil and soil erosion, a new range on resolutions or new achievements regarding computing power help to further develop these models adequately? Most process-based models demand measured soil data (e.g. texture, water retention function, hydraulic conductivity or aggregate stability) most of which are not available in a high resolution. However, recent data assessment techniques especially regarding photogrammetric methods offer new data on soil and soil erosion, and with this a broad range of model development as well as model ensemble strategies. Furthermore, we regard cross-process as well as cross-scale understanding of soil erosion on different spatial and temporal scales as a vital step towards improving process-based soil erosion modelling. To fully understand and achieve an impact reduction by applying adapted management strategies, models need to integrate the current understanding of soil erosion processes from splash to gully erosion (Parsons, 2019; Li and Fang, 2016). For future model development we find it crucial to determine whether the present assessment techniques are contributing to the improvement of process-based soil erosion models. Most models do not include high resolution soil or surface input data and work only on specific scales, which begs the question whether models can keep up with the data."

R (line 156): I think process-based models contribute to process understanding, whilst empirical models do not. Now, how accurate or realistic is this understanding might be up for debate.

A: A fair argument. We would like to phrase our ideas less absolute and discuss the relationship between models and reality in a few sentences.

R (line 157): This is not entirely true. Process-oriented models often have an empirical basis, developed upon a limited number of site-specific experiments. This is discussed by Parsons (2019) and Govers (2011), from your reference list.

A: We acknowledge that this argument is poorly phrased and will change it accordingly.

R (line 192): I suggest to rethink how to present this data. The table was a bit confusing with the missing/limitations terminology. Maybe just say what the models can do, not what they cannot do. The second can be deduced from the first, no?

(lines 494-495): I found this part of the manuscript a bit disappointing. The model comparison table is confusing and contains several mistakes.

A: We like the suggested structure of the table and will change accordingly. We plan on removing the limitations and instead naming the limitations common in most models in the text itself. Furthermore, we acknowledge that the focus lies too much on the table and therefore we will relocate it to the supplement. Finally, we plan on adapting the table according to the referees' feedback and double checking every model and their right to exist. The limiting factors common in many of the models will instead be listed in chapter 2.1:

"Next to their capabilities, the authors also list challenges and shortcomings of the models. The following difficulties, present a selection which can be found in many models and are sorted by how frequently they are mentioned:

- *High data demand*
- *Process description of gully erosion*
- *Process description of rill initiation (and rill structure)*
- *Sediment connectivity*
- *High computing demand*
- *Assumption of homogenous input parameter*
- *Site-specific calibration*
- *Temporal unchanging soil and surface input parameters*
- *The risk of equifinality"*

R: (line 211): Not sure you need to keep referencing this and other review papers to make simple points, like this one. I have the impression that in general you are relying a bit too much on these reviews. I would like to hear more about what you have to say.

A: In the course of restructuring and revising the various paragraphs, we will include our own conclusions more often.

R: (line 231/figure 2): Sediment tracing and sediment fingerprinting are techniques. REE, soil magnetic parameters, and fallout radionuclides are properties used in these techniques.

A: We will change our figure according to your suggestions and use the precise wordings.

R (lines 319-320): This definition is incorrect. https://en.wikipedia.org/wiki/Equifinality

A: Thank you for pointing out this error. This cause-result statement, as we now see, has led to a false overall statement. We will change it in the text and separate those statements from one another.

R (lines 450-451): I don't know if I agree. Take a look at this: https://kwarc.info/teaching/TDM/Borges.pdf Perhaps what is needed is a balance between the quality of relevant process descriptions, the uncertainty in the required input data, and the model structures. Keeping up with the data, as you put it, might help us achieve this balance.

A: Thank you for your input. You make a valid point. We have therefore rephrased the paragraph as follows.

"Finding a way between models covering the relevant processes accurately while only being as complex as they need to be, represents a current challenge. Future model development needs to balance these aspects of model structures, processes, and uncertainties regarding the input parameters and the models' outcome. Keeping up with the data, by using new, improved data assessments for more accurate parameterization and a better validation of the processes and results could be the way to achieve balanced process-based soil erosion models."

R (line 513): Is user friendliness really the goal? I would focus more on reproducible code than user friendly GUIs.

A: This is well argued. User-friendliness is a useful addition but not the overall goal of model improvement. We would like to revisit our conclusion altogether, correcting this statement and adding more of the aspects discussed above.