For clarity and improved visualisation, the reviewer comments are shown from here on in black. The authors' replies are in blue font below each of the reviewers' statements. The changes in the revised manuscript are displayed in green.

Reviewer #1

General Comments

The paper is excellent in terms of organization, writing, and content. The results and figures are convincing, and I am particularly excited about how the authors utilize changes in model structure to address sediment connectivity, which is timely and quite important in terms of advancing watershed sediment simulations. Overall, I have very minor comments regarding some clarifications in a few instances and I believe that a few statements made by the authors should be relaxed a bit. Additionally, I would suggest including a brief paragraph at the end of the discussion regarding limitations of the WaTEM/SEDEM approach, and how we can further move to improve the spatial and temporal resolution of sediment connectivity simulations.

Many thanks for the time and effort put into reviewing our manuscript. We are pleased to know it was appreciated. Below we respond to all your specific comments. We will also include a paragraph in the discussion further highlighting the limitations of the WaTEM/SEDEM approach and some thoughts on how to improve sediment connectivity simulations.

Specific Comments:

L 32: Perhaps you can also mention that the paper is, for the first time to my knowledge, advancing tools to assess connectivity by quantifying structural uncertainty within the sediment simulations (not referring to *structural* connectivity here, by the way, just how there are inherent uncertainties within how the model is configured to predict fluxes/loads).

Thanks for pointing this out. We will highlight how the quantification of structural model uncertainty is a novel aspect of the paper.

L 159-167: It would be helpful within the text to tell readers the temporal resolution of the model. It seems like it's yearly according to the RUSLE equations but could be clarified.

Apologies for this omission. Indeed, the model is operating in a yearly time-step. This will be clarified in the revised manuscript.

L 180: Is this of the individual pixel or along the slope length?

The slope gradient and the transport capacity are calculated per individual pixel.

L 188: Does this include bank erosion?

We will include an explicit mention to bank erosion here, as this process is also not simulated by the model.

L 188-190: If this is a yearly model then perhaps this statement can be slightly relaxed... For example if the system is known to not aggrade or degrade over longer-term (decadal) timescales then instream erosion and deposition are approximately in equilibrium and so I would not be as concerned with the instream.

We completely agree with your comment and that our statement could be somewhat relaxed. We would like to rephrase to:

"Since WaTEM/SEDEM does not represent gully and bank erosion or in-stream erosion and deposition processes, any comparison between modelled sediment yields and catchment-outlet sediment loads must be interpreted with caution. However, in catchments where rill and interrill can be assumed to be the main erosion processes, and assuming a state of fluvial quasi-equilibrium due to the temporal resolution of the analysis and the model, the outlet sediment loads should be at least comparable to the model outputs, even if not fully commensurable".

L 205: How did you decide 1200? Is this enough? Sometimes people will utilize 100,000 monte carlo simulations. I'm not saying that you need to run the model for more realizations, just a bit more justification please.

In all honesty we do not have a strict justification for the number of iterations in the Monte Carlo simulation. We understand 1200 iterations was enough to explore the parameter space for the purpose of the sensitivity analysis. If we wanted to sample the parameter space exhaustively in a rejectionist approach, then we would likely need a higher number of simulations considering the number of parameters.

L 205-208: Right, the typical approach is to calibrate the model and along the way assess sensitivity/uncertainty such that sensitivity/uncertainty of the model is addressed within solution spaces that are plausibly behavioral. I'm not rejecting your approach by any means,

but perhaps some additional acknowledgement of the traditional approach and how you are slightly deviating here could be helpful to readers. Some readers might question why you present realizations that will not adequately describe the sediment load/flux in the system. L 208: *model assumptions* – I would clarify you are making assumptions about the structure of the model, so quantifying structural uncertainty.

Precisely – with the sensitivity analysis we did not aim to identify the behavioural parameter space, but rather to understand how the model responds to the different structural assumptions. We will make the following changes in this paragraph:

"Our model application consists of a global all-at-a-time sensitivity analysis, as described by Pianosi et al. (2016). That is, we performed a Monte Carlo simulation to explore the variability of the whole parameter space, and all input factors were sampled simultaneously for each model realisation (n = 1200). The framework is similar to an uncertainty analysis, except in this case we did not focus on locating the parameter space which produced behavioural model realisations. Instead, we concentrated on apportioning sources of uncertainty to different model input factors, aiming to rank their contribution to the variability of the response surface (see Pianosi et al., 2016 for a review on sensitivity analysis). This should allow us to identify parameters and model assumptions that have a greater impact on the manner with which WaTEM/SEDEM describes sediment connectivity in the Baldegg catchment. In particular, the analysis of different assumptions about the structure of the model should provide a connectivity assessment based on the apportionment of the structural uncertainty withing the simulations. To the best of our knowledge, this is the first time model structural error is incorporated into sediment connectivity research."

L 238: can you clarify why a value for *Pcon* wouldn't be applied everywhere in the catchment, but instead for just the forest and buffer strips? What if there is disconnectivity from microtopography in the roadside ditches, for example? Again – I'm not asking for additional analyses, just a sentence or two for clarification and that you *might* parameterize this other places in the watershed if you had overt reason to.

This is a great point, thanks for bringing it to our attention. The parcel connectivity parameter was originally developed to represent the extent with which water and sediment transport is reduced at parcel borders in case the downslope patch is composed of forests or grasslands. We will include this clarification in the text. We completely agree that the parameter could be

incorporated in other places, and we will mention this in the model limitations/improvements section you suggested.

L 255-259: can you please add a sentence that details the difference between scenario two and three? The way I understand it is that in scenario three sediment deposition does not occur on the road or in swales/ditches along side the road, but deposition can still occur downstream, for example in between the road and the stream network. In scenario two sediments are automatically connected to the stream, correct?

That is absolutely correct, we will include this clarification in the text:

"For this scenario, deposition will never occur on road cells, however sediments can still be deposited on downstream patches, before reaching the stream network".

L 270: Again I might suggest using the word *structural* uncertainty of the model.

Thanks, we will include a mention to structural uncertainty here:

L 275-276: This was a bit confusing to me. L 306: What is the mean squared error in relation to? The yearly predicted sediment load and the yearly average sediment load from the rating curves? Please clarify.

Apologies for this confusion. The random forest analysis was used to predict the WaTEM/SEDEM simulations of hillslope sediment yield, based on the parameter values that were sampled for each iteration. The mean squared error is calculated from the RFA predictions and the WaTEM/SEDEM simulations. The increase in error due to the absence of a variable is used to rank its importance. We will clarify these points in the methods and the results.

L 323: Fig. 6g – what about interrill erosion?

Thanks for noticing this, we change it to rill/interrill erosion.

L 323: Fig 6b,c; L 333: Is it worth showing land use for all the details here?¨

Yes, that is a good idea, thanks. We will add an identification on the landuse in figures 6b,c.

L 374: Perhaps you can say in the caption that the short-cut generally overlaps the IQR better than the other 2 scenarios... this could help readers quickly interpret the figure.

Thanks, we will add that to the caption.

L 385: Out of bound percentage – is this a fraction or a percentage what is presented in the table?

Apologies for this mistake, indeed we were presenting the fraction. This will be corrected accordingly.

L 388: It would be great if we could see at this same time scale how SEDEM was performing… but I think this is just a limitation of the model since it runs at a yearly scale, correct?

Yes, exactly. For this reason, we only made comparisons with the average yearly loads.

L 424: Perhaps also the rating curve is underestimating the load, as you previously mentioned? Which would improve the performance of simulations with respect to the short cutting, correct?

We are not sure this would be the case here, as the curves are probably underestimating the actual loads for all streams – not just for the Hohibach.

L 451: I believe Mahoney et al., 2018 talks about importance of road networks a bit in the USA.

Thanks, we will include this reference here.

L 463-465: I think this last sentence should be relaxed a bit... quantifying all of the sources of uncertainty due to observation data, model input data, model output data, parameter uncertainty, etc. etc. is quite the undertaking. In fact, in my opinion, it might be an impossible task. Does this invalidate the use of models, however? In my opinion, no, it does not. We can still discern important information from models even though we don't account for 100% of uncertainties. It ultimately will depend on what questions we are trying to answer with the model and what the model is attempting to do, which can be equally as important as quantifying certain uncertainties in my opinion.

We completely agree that accounting for all uncertainty is simply impossible. What we were trying to convey here is that soil erosion and numerical connectivity models are highly uncertain. This uncertainty stems from multiple unknows about the modelled phenomenon, the input data, and the forcing data. Our opinion is that neglecting such uncertainty makes it very

difficult to provide meaningful insight based on the modelling. In any case, we see how our statement might have been too strict here, and we will rephrase considering your comment.

L 465-466: It would be nice if a paragraph on limitations of the modeling approach and future opportunities could be included. For example, while RUSLE is relatively easily implemented and approachable, it would be nice if the RUSLE approach was a bit more physically based. Additionally the RUSLE approach limits the temporal resolution of the model, so seeing event- and seasonal-scale connectivity seems a bit limited. Furthermore, the advanced geospatial data that facilitates this novel connectivity modeling is wonderful, and can help to elucidate hotspots of connectivity. Additionally there is recent sentiment to move towards high-temporal resolution models to quantify hot-moments of connectivity. The yearly timescale inherent to the RUSLE approach perhaps is underserving this sentiment.

We completely agree. We think a model like WaTEM can be useful to explore structural connectivity patterns, but much more dynamic models are needed to quantify these hot-moments of connectivity, and to get a better grasp of the functional connectivity of the system. As we previously stated, we will include the paragraph you suggested.

L 466: I'd suggest perhaps emphasizing that exploring structural uncertainties in the model framework - and not just parameter uncertainties, as is the traditional method - allowed for advanced understanding of connectivity processes. This type of approach in my opinion is quite underserved in modeling work and should be considered in the future where high-resolution geospatial data is available.

Thanks for pointing this out throughout the manuscript. We will emphasize here and in the remainder of the manuscript the relevance of quantifying structural uncertainties.

Technical Comments:

L 312: I'm not sure if the different colors are helpful here, maybe consider symbols?

The colours are just there to keep the theme of the graphs – the facet titles identify the scenarios. If its ok, we would like to keep them.

L 324: typo

Thanks, corrected.

L 364: typo, confusing

We will rephrase to: “The comparison between WaTEM/SEDEM simulations and the sediment loads revealed a larger overlap between the latter and the results from the ‘road-as-shortcuts’ scenario (Figure 7).”