Dear Genevieve and referees,

Apologies for the delay in our response and many thanks for your patience. Below we answer to all your questions, comments, and suggestions. 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. Line numbers refer to the revised tracked manuscript.

Please note that while editing figure 5 following the referee comments, there were some changes in the mean squared error (MSE) values calculated by the random forest. This is because the randomness of the tree building. We have now used a set seed to ensure the reproducible random objects. Importantly, these changes in the MSE values were slight and did not at all affect the interpretation of the results.

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.

Thanks again for reviewing our paper, we highly appreciated the input. Below we respond to all your specific comments. Moreover, we have included 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 highlighted how the quantification of structural model uncertainty is an important and novel aspect of our work, throughout the text (L23, L36, L237-240, L319)

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, as we now mention in lines 188-188.

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 (L195).

L 188: Does this include bank erosion?

We included an explicit mention to bank erosion here, as this process is also not simulated by the model (L210).

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, thanks for pointing this out. We rephrased 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 are the main overland erosion processes, and assuming a state of long term fluvial quasi-equilibrium, the outlet sediment loads should be at least comparable to the model outputs, even if not fully commensurable". (L212-216)

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 model realisations 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 made 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 the analysis model structural error is incorporated into sediment connectivity research." (L228-240)

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 clarified this in the text (L273-274). 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 now state:

"For this scenario, deposition will never occur on road cells, however sediments can still be deposited on downstream patches, before reaching the stream network". (L299-300)

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

Thanks, we will included a mention to structural uncertainty here (L310).

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 (L315-317, L349-351).

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

Thanks for noticing this, we included interrill erosion in the figure legend.

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

This was indeed missing, thanks for noticing. We included some text in the figure showing where the arable land and the forest are located (Fig 6b).

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 included this in the caption of figure 7.

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 has been corrected in Table 4.

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 explain this pattern, as the curves are probably underestimating the actual loads for all streams – not just for the Höhibach. In any case, we explained now how underestimation would probably improve the performance of the shortcut simulations (L436).

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

Thanks, we included the reference (L511).

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 impossible. What we were trying to convey here is that soil erosion and numerical connectivity models are highly uncertain. This uncertainty stems from multiple unknowns about the modelled phenomenon, the input data, and the forcing data. Our opinion is that neglecting such uncertainty makes it 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 rephrased to:

"As recent studies have again demonstrated, investigating the uncertainty in model structures, parameter estimation, and observational testing data is crucial for advancing soil erosion modelling research (Benaud et al., 2021; Eekhout et al., 2021; Schürz et al., 2020)" (L526-530)

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. This is now discussed in lines 531-540.

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. We emphasized the relevance of quantifying structural uncertainties throughout the manuscript

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 it's ok, we would like to keep them. We did include the symbols now, however (Figure 5).

L 324: typo

Thanks, corrected.

L 364: typo, confusing

We rephrased to: "The comparison between WaTEM/SEDEM simulations and the tributary sediment loads revealed a larger overlap between the latter and the results from the 'road as shortcuts' scenario (Figure 7)" (L410-412).

Reviewer #2

General comments

This paper provides a modelling approach to analyse the effect of linear features on sediment connectivity in a mesoscale catchment. The authors put a lot of work into adapting an existing model such that it is able to account for these effects. I appreciate this effort and I think that this work is important for improving the understanding of sediment connectivity in agricultural catchments. However, I have some major points of criticism which are the following ones:

We highly appreciate the time and effort put into reviewing our manuscript. We specifically appreciate the attention to detail and the discussion about what conclusions can be drawn from the study. We address all your comments below.

The authors state themselves that a comparison between their model results and the measuremed sediment loads in the river should be performed with upmost caution. However, in the results and in the discussion they still make strong conclusions based on exactly such a comparison. The authors should therefore reformulate the discussion such that it reflects this uncertainty better.

We apologise if this were conveyed: we tried to be very cautious about the conclusions drawn simply from the comparisons with the sediment loads. Overall, the importance of linear features for regulating sediment connectivity has been extensively documented by field observations and modelling approaches in Central Switzerland (Alder et al., 2015; Ledermann et al., 2010; Remund et al., 2021). Results from our study corroborate these observations, both due to the sensitivity analysis and the comparison with the sediment loads, which we understand were cautious (or even overly cautions considering some of the comments from reviewer #1). As we state in the manuscript, the comparison between outlet sediment loads and model outputs provide an estimate of the plausibility of the simulations.

In any case, we reformulated the sentences you mentioned in the specific comments in order to be more careful about any potentially overreaching conclusions. This is explained below in our reply to your specific comments.

The comparison between the model results for the different scenarios and the measured sediment loads does not make sense for me in the way it is done currently. In their current analysis, the authors ignored important factors influencing the sediment connectivity in the catchment: Firstly, as I understand it from the manuscript, they treated all grassland areas like arable land areas in the model. Since around a third of the catchment area is covered by grassland, this leads to a large overestimation of sediment loads.

Here we believe there has been a misunderstanding, which is entirely our fault. Indeed, we did not separate grasslands from croplands (reasons why are explained in the specific replies). However, we considered this uncertainty in the parameterisation of the CP factor, which had the lower limits of their “prior” distribution stretched to include typical values for permanent grasslands in Switzerland. This is now explicitly stated in lines 245-247. Similarly, the Ktc high parameter was sampled from 0 to 200 m to cover a wide parameter space. Hence, some model realisations will indeed overestimate the erosion rates, whilst others might underestimate them. Since we did not focus on calculating exact erosion rates for the catchment, but rather on assessing the influence of linear features and landscape patchiness in sediment connectivity, we do not see any major limitations with our approach. Particularly since all this uncertainty is incorporated into the model outputs.

Secondly, the authors assumed a two-meter grass buffer strip around all agricultural plots. The authors state that they don't know the real width of buffer strips in the field and that they therefore use a value of two meters for testing the sensitivity of the model. Even though the sensitivity analysis showed that the buffer width had a large influence on the model results, the comparison between modelled and measured sediment loads is only done for the two-meter scenario. Also the conclusions are only drawn based on this scenario. Even though the authors state that the two-meter scenario is more realistic than the scenario without a buffer, this value contradicts the values reported by other studies (e.g. Alder, 2015; Remund, 2021) and the legal requirements. Although this point is addressed in the discussion, I am missing a proper justification why the measurements are only compared to the two-meter scenario.

We assumed a 2 m width to test the sensitivity of the model to the presence of buffer strips. The value was pragmatically chosen based on the spatial resolution of the model input data – apologies for not mentioning this in the methods. In addition, we understood it was better having one fixed value for testing such sensitivity, than having multiple values which would anyway be uncertain. To clarify, our sensitivity analysis did not demonstrate the model was sensitive do the buffer *width*, as we only tested one value. We think testing additional values goes beyond of the scope of our research.

Now, if the buffer strips alongside roads in the catchment (which are at least 0.5 m wide according to legislation, but highly variable in width) are completely ineffective, then the scenario without grass strips would possibly be the most appropriate representation of the system – at least considering the capacity of model outputs to mimic the data. However, the strips would still be there in the fields, and this was our rationale when we stated that the scenario with strips "more closely represent[s] the actual structure of the agricultural fields in the Baldegg catchment". As all results are summarised in Table 3 of the manuscript, and the scenario without strips shows the same pattern for the different road connectivity assumptions as the scenario with the strips – leading to the same conclusions – we thought it would be interesting, for concision and clarity, to focus the discussion on the latter. In any case, we have now included the data for the scenarios without grass strips into Figure 7. As you can see, this does not change the interpretation of our results.

Thirdly, in their "shortcut" scenario, the authors assume that all roads and farm tracks are drained with shortcuts. (At least, this is how I understand it from the manuscript.) I expect that a major part of the roads in the Baldegg catchment is actually not drained by shortcuts. Therefore, it makes sense to use the current "shortcut" scenario in a sensitivity analysis, but not as a realistic scenario. Compared to reality, the current "shortcut" scenario is expected overestimate the real sediment loads. Consequently, even though the "shortcut" scenario is most similar to the measurements, this is possibly simply caused by a bias in the model input. Even though I agree that roads and shortcuts are in fact important for sediment transport, I think this cannot be concluded from the current analysis. To state that "roads behave as conduits for sediment transport in the catchment", as it is done in L474f, it is inevitable that the scenarios are revised such that they reflect the reality in the catchment better. (At least for the first point mentioned. Second and third point may also be discussed.)

We completely agree that in reality only a certain portion of the roads will drain the sediments directly to the stream network (note that we never used the word "realistic" to describe any of the scenarios). However, as we explained above, we used quite a wide range of parameter values and different scenarios to examine *how things could happen* in the catchment. Still, the model was only able to provide reasonably comparable results to the outlet sediment loads if we assumed that roads behave as conduits for sediment transport. And this is considering the large uncertainties in both the sediment rating curves and the model outputs. These findings, in combination with the results from the sensitivity analysis, and the multiple studies that report similar patterns for other sites in Switzerland (as reviewed in the discussion), should allow us to state the following in the conclusion (L529-538):

"Our results demonstrated that assumptions about road connectivity were by far the most important factor for modelling sediment transfer in the Baldegg catchment. Moreover, the comparison between extensive model simulations and sediment rating curve calculations indicated that roads and hydraulic shortcuts are likely to behave as conduits for sediment transport in the catchment. Hence, representing road connectivity is crucial for modelling sediment transfer from hillslope to water courses in this agricultural catchment of the Swiss Plateau, and potentially in other areas with a dense road drainage system. Moreover, our results further highlighted the effects of linear structures and landscape patchiness on sediment connectivity."

In addition, if the modelled sediment loads were highly overestimated, as you have hypothesised above due to a potential bias in the CP parameterisation of grasslands/croplands, then it would not make sense that the shortcut scenario showed a better fit with the sediment rating curve calculations. In this case, the shortcut scenario would exhibit much higher sediment loads than the measurement-based estimates, as both erosion rates and sediment connectivity would have been overestimated.

The authors state that the catchment is representative for the Swiss plateau. However, they do not further elaborate on this. Other studies, however, rather suggest that the catchment has a low shortcut connectivity compared to other catchments in the Swiss plateau (see comment to L98). The authors should improve on putting their analysis in the right context.

Thanks for pointing this out. We removed any mentions in the manuscript about the catchment being representative for the Swiss Plateau. Indeed, that cannot be affirmed, and it is not our objective to provide a representative case study for this part of Switzerland.

Several specific comments (see below) should be addressed to improve clarity and reproducibility of the study. Additionally, the manuscript should also receive some revisions regarding language and correct spelling.

From our understanding, the comments on reproducibility are mostly related to software and package versions, and to methodological details regarding data preparation. These were addressed accordingly, as we explain in the replies to your specific comments below. In addition, we have corrected the typos and spelling errors you highlighted. We would be glad to send the manuscript over to a native speaker for review if the editor deems necessary.

Specific comments

L36: Talking about a "continuous displacement of small amounts" is wrong here. The discplacement varies strongly between events and years, as you also state below.

We rephrased to: "Rainfall events on sloped surfaces continuously displace soil particles, which are transported downslope as sediments" (L40).

L47: Rephrase.

We would appreciate more guidance here about what needs rephrasing and why.

L56: You should also add the most recent publications here, e.g.: https://doi.org/10.1016/j.catena.2021.105290

Thanks, that was a very appropriate reference and it has been included to the text (L61).

L66: "assuming they are able to explicitly take connectivity into account": Difficult to understand. Please write this more clearly.

Apologies if that was not clear. We rephrased to: "These usually rely on high-resolution process-based models, assuming they are able to represent connectivity dynamics" (L72).

L73: with a size of few square kilometres

Thanks, we updated the text accordingly (L79).

L79: You state above that one major issue of erosion models is the uncertainty of input data. Then you state that you used a high resolution dataset (2x2m DEM). However, for whole Switzerland, a 0.5x0.5m DEM is freely available in the same quality as the 2x2m DEM. You still used the 2x2m DEM. Why did you not use the higher resolution model?

In such high resolutions (0.5 m), the influence of the microtopography becomes much more prominent, and we understood this conceptual model would not be the most appropriate for handling such features. We would like to highlight however that the 2 m resolution we are working with is much higher than what is usually employed in erosion modelling research, in particular at catchment/mesoscale (see Borrelli et al., 2021).

L98: Here you state that the Baldegg catchment is *patch*y and *representative* for the Swiss Plateau. Below, you only elaborate on the patchiness, but not at all on the representativeness. Either elaborate on the representativeness below or use another word here. Schönenberger et al. investigated two catchments in proximity or even inside your catchment. Compared to the distribution in the Swiss plateau, these catchments however rather seem to have a low shortcut connectivity. This indicates in my opinion that also your catchment is rather on the lower side with respect to shortcut connectivity.

We agree that we did not elaborate on the representativeness of the catchment and ultimately this is not our goal here. Hence, we removed any reference to the catchment being representative for the Swiss Plateau.

L103: In Figure 1c, you use the term "infrastructure", here you use the term settlements. It is unclear how these two terms differ and what the term infrastructure means. Please use consistent terms here. Additionally, how did you treat the areas of roads? Did you include them into the settlement area? Or are they included in the agricultural land/forest area?

Apologies for this inconsistency. Roads and settlements areas were calculated as part of the infrastructure, as we now state in line 114.

L117: Are tile drainage only located in water accumulation zones? (What are water accumulation zones? Are the determined based on topographic index, slope?)

By definition tile drainage is found where there is excess water. Upon reflection we found this information superfluous, and we will remove it from the revised manuscript.

L120: How did you determine these slopes? Which elevation model? The maximal slopes are strongly depending on the model used. You are referring to Figure 1b. However, the slope is not visible in Figure 1b.

Figure 1b is referred for altitude "higher altitudes are found in the eastern and western sides of the catchment (Figure 1b)" in line 134. Slope was calculated with the same DEM used for the model.

L122: "in this case formed by the retreat of the Reuss Glacier in the south to north direction (~18,000 years BP)" -> Not important. Consider removing.

We think this information contributes to the description of the study site and would appreciate if we could keep it.

L127: MeteoSwiss -> Please add reference.

Apologies for this lapse, the reference has been included.

L132: What is approximately? Provide the range of numbers of grab samples taken.

Apologies, we have corrected to "on average 275 grab samples were taken from each tributary" (L147)

L133: What is "opportunistical sampling"?

This has been removed and further information on the sampling was provided (L144-149).

L136: Why 2020? Above you write that you only sampled till 2019.

Apologies, we meant until the end of 2019 (L152).

L144: What is k? I guess the covariate ID. This should be written explicitly. Additionally, in contrast to Vigiak & Bende-Michl, you are only using the first five covariates, but not the long-term trend covariates (6 and 7). Why?

Thanks for noticing this. Now we explicitly mention that k is the covariate identity (L164). We did not use the long-term trend covariates because they were either not significant or did not improve the models, which makes sense considering the timescale of our analysis.

L149: First column of table: Remove the word "is". Also the word "water" is not really needed. Second column of the table: This is not the quadratic term of Q_i, but of $x_{2,i}$.

Many thanks again for noticing these errors, they have been corrected in Table 1.

L156: You are addressing the variance in sediment concentrations extensively. However, you are not addressing the uncertainty in daily discharge at all. Why?

Essentially because the uncertainty in the sediment load calculations is much larger than for water discharge. We have on average 275 measured sediment concentrations per tributary, which need to be extrapolated for 10 years. On the other hand, daily discharge measurements are available.

L165: Shortly explain why you only focused on water erosion.

We chose to focus on the water erosion instead of tillage because the latter is not relevant for the type of connectivity processes we are investigating (L184-187)

L193: “usually implemented” -> Rephrase. (It is either implemented or not. Possibly, you could write something like that this version is often used.) Provide references where this version is used.

Thanks, this was rephrased to “WaTEM/SEDEM is implemented as a user-friendly GUI […]” (L218).

L199: R version? L200: SAGA version?

Versions are now supplied in the reference list and code.

L201: The code does not contain any information on the versions of the packages used. To make sure that the code can still be used in the future, you should at least provide information on the package versions used. Consider also using tools like packrat, checkpoint or docker. To make the code useful, you should also provide examples of input files.

Thanks for the package recommendations. We included information on package versions in the code – apologies for this missing information. As we stated in the data availability section, we will upload the model input data to Envidat if the manuscript is accepted.

L219: It remains unclear how you derived the land cover map. The reference is not shown in the reference list. Therefore, I don’t understand if you used a vector dataset that you rasterized yourself or if you used a raster dataset provided by Swisstopo. What does the resolution “1:25:000” mean in the latter case? If you used vector datasets: How did you deal with point and line features, e.g. roads or hedges? Did you assume widths for roads and hedges?

We are terribly sorry for missing Swisstopo in the reference list. Indeed, we rasterised the land cover vector data (Swiss Map Vector 25 BETA, scale 1:25,000) to a 2 m resolution. The roads were firstly converted from lines to polygons with a buffer, considering their width (which is informed in the vector map). Hedges and tree lines were already represented as polygons in the vector map. We made it clearer in the manuscript that we rasterised the vector map ourselves (L2250).

L220: The statement that spatially distributed crop statistics are unavailable is wrong. There is a plot-resolution crop dataset from the canton of Lucerne available freely for the whole canton (and accordingly for the whole catchment) for the year 2019. Why was this dataset not used? (CP and K_{TC} depend on the crop and you reported them to be the most sensitive model parameters.) Lavrieux et al. state that one third of the agricultural area consists of permanent grassland. Therefore, I expect this decision to have a large influence on your results, leading to an overestimation of erosion. You should address this point at least statistically for each of the five subcatchments analysed and for the full catchment. (e.g. look at the fractions of grassland and reduce the estimated amount of sediment load accordingly). Alternatively, you could also do a spatially explicit analysis.

In all honesty we were not aware of the availability of such data for the canton of Lucerne. We inquired the canton about it, and the geodata would indeed be available. However, this would cost 4260 CHF.

In any case, as we stated before, we apologise for this misunderstanding regarding the grasslands. As we explain in the revised manuscript (L255-257), "The minimum CP values were particularly reduced to include typical values for permanent grasslands in Switzerland (~0.01) (Schmidt et al., 2018b)". This lower limit of the CP parameter is also analogous to the lowest value recommended for permanent grasslands in Europe (0.01-0.08) (Panagos et al., 2015). Similarly, the Ktc high parameter was sampled from 0 to 200 m to sample a wide parameter space. Hence, some model realisations might overestimate the erosion rates in the catchment, whilst other realisations will likely underestimate them. In either case, what we can see with the model results is that whatever values we sample for the CP or Ktc parameters, the assumptions about connectivity have the highest impact on the model results. In summary, we have already addressed the land use issue 'statistically', as you put it.

L226: Why did you use the 2x2m resolution DEM? (see comment to L79) How did you process the DEM, e.g. sink filling?

The choice of DEM resolution is explained in the comment to L79. The DEM was sink filled by the Wang & Liu (2006) method, which is implemented in SAGA.

L227: In the K_{TC} column, consider indicating the land use classes belonging to "high" and "low" (e.g. in brackets). This would make it much easier to read.

Thanks, we included the information in brackets (Table 2).

L230: Specify that this relates to the maximum CP factor.

This is clear now as we state the minimum value was selected based on the permanent grasslands (L255-257).

L233f: You talk about forests and grass buffer strips in the land cover map. However, the reader is missing were you explain the derivation of forests and grass buffer strips. Consider stating that you are explaining this below.

We now state we will explain the grass buffer strips below (L267).

L246: How much wider? Refer to the article in the corresponding legislation directly, instead of Alder et al..

Sorry for the missing information. We rephrased to: "The extent of the buffer-strips in reality is quite variable, and generally wider at forest and river vicinities (3 – 6 m), as required by law in Switzerland (Alder et al., 2015)." (L281-282)

We would like to keep the reference to Alder et al., which provides this and other information about grass buffer strips in Switzerland in English.

L249: What were your assumptions on buffer strips along hedges? Did you also use a 2m buffer? Or a buffer corresponding to the legal requirements? How did you treat tree lines?

We used a 2 m buffer along the hedges and treelines, which were rasterised from the land cover map mentioned previously.

L250: As mentioned in the comment to L219, I don't really understand how you derived the road areas.

The road widths are provided in the land cover vector data (Swiss Map Vector 25 BETA). We used these widths to perform a buffer around the road lines. Next the polygons were converted into a 2m resolution raster (same resolution as the DEM).

L252: What is "infrastructure"? Was this also derived from the land cover map?

Infrastructure includes roads and settlements. This is now informed in the text (L114)

L253: If roads act as sinks, why is this related to field drainages? From the text, I don't understand where you assume the sediments to be trapped. On the road? In the drainage system?

In sludge collectors? The scenario seems to make sense for me, but you should specify your assumptions more clearly.

We rephrased to "This represents a scenario in which roadside ditches and the road drainage system trap most sediments and partly diverge runoff to wastewater treatment plants" (L290-291).

L256: Did you assume here that all roads are acting as shortcut? Or only a part of the roads?

We assumed all roads behave as shortcuts. We now mention how this could be improved if the location of the shortcuts and the extent with which they are connected to surface waters were known, as in Schönenberger and Stamm (2021) (537-540).

L262f: How many directions?

Information on the multiple flow direction algorithm implemented in SAGA can be found in Freeman (1991) and Quinn et al. (1991). It allows for divergent flow paths, differently to the typical D8 approach.

Freeman, G.T. (1991): Calculating catchment area with divergent flow based on a regular grid. Computers and Geosciences, 17:413-22.

Quinn, P.F., Beven, K.J., Chevallier, P. & Planchon, O. (1991): The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. Hydrological Processes, 5:59-79.

L266-268: In my opinion, some (or probably all) of the additional packages (e.g. "doParallel", "foreach") are not worth mentioning here, as they are only used to speed up the calculation process, but not important for reproducibility of your work.

Thanks, we removed these references.

L278: What does RFA stand for?

Apologies, RFA stands for random forest analysis (L314).

L280: Version?

Information on package versions is provided in the reference list.

L299: Do I understand correctly that Mg yr^{-1} means tons per year? Probably write tons instead.

The megagram (Mg) is equal to a metric ton. It is common approach to express erosion and sediment transport rates with such unit.

L299: For me, the differences between GS and NGS scenarios are not well visible in the plots. Consider making this better visible, e.g. by adding a moving average per category or something similar.

We think a moving average would not the best choice here, as the idea with the scatter plots is also to display the spread of the response surface.

L317-L318: How did you quantify this for the whole catchment? Just by eye? If yes, you should provide the respective plots (e.g. in the appendix). Otherwise, can you provide a quantitative assessment?

Figure 6b, which we refer to, provides an example of the mentioned depositional patterns. In addition, increased within-field deposition rates are expressed quantitatively in Table 3.

L324: In Figures 6b, 6c, 6e, 6f, and 6g, arrows indicating the flow direction would help to understand the plot better.

Thanks for the suggestion, we included arrows indicating the flow direction (Figure 6).

L358: In my opinion, Table 3 would be understandable easier if you would write 25%, 50%, and 75% instead of Q1, Q2, Q3.

Thanks, but in this case, we would like to keep the quartile numbers in the table header.

L366f: As I understand from L244ff, you say that you don't really know what the real widths of buffer strips are in the catchment. It therefore makes sense to me that you use a fixed width of 2m and use it for testing the model sensitivity by running two scenarios – one with and one without the buffer strip. However, in L366 you now state that the 2m scenario is more realistic than the "no buffer" scenario. I agree that a 2m scenario is probably the more realistic scenario along forests. However, I expect the effect of a grass buffer along forests to be small as sediments are trapped by forests anyways. Along roads, I doubt that the 2m scenario is more realistic than the “no buffer” scenario, since the legal requirement is only 0.5m (as you write in L430). (However, I might be wrong with these doubts.) Since I expect the buffer width along roads to be much more important for your model results than the ones along forests, I think you should report both scenarios here, or give a clear explanation on why you think or why you can show that the 2m scenario is more realistic.

Thanks for these considerations. As we stated previously, if the buffer strips alongside roads in the catchment, which are at least 0.5 m wide according to legislation, but highly variable in width, are completely ineffective, then the scenario without grass strips would possibly be the most appropriate representation of the system – at least considering the capacity of the model to mimic the outlet data. However, the strips would still be there, and this was our rationale to state that the scenario with strips “more closely represent[s] the actual structure of the agricultural fields in the Baldegg catchment”. In any case, we have now included the data for the scenarios without grass strips in Figure 7.

L379-382: In the “Ron” stream, the 95% prediction interval seems much narrower than in the other rivers (Figure 8). Therefore, the observed values are mostly outside of this interval and the out of bound percentage is much higher than for the other streams. Can you explain this?

The interval is narrower because the model fit was better, and the residuals were lower. This led to a lower proportion of the observed values being outside the prediction interval.

L389: Consider using the same y axis limits for all plots. Like this, it is difficult to see the differences between the streams. At least the zero line should be visible in all plots.

We standardised the y axis limits (Figure 8). The zero line is not visible because of the log scale.

L422: Which physiographical statistics did you analyse? Please provide details. I guess you did not analyse crop types (e.g. fraction of arable land (without grassland) on the total catchment area)? Could this also be a reason for the difference?

That is a good point, thanks for bringing it to our attention. We now mention the characteristics we analysed (e.g. stream and road density), and explain that crop types might contribute to explain these differences (L477-478).

L443: Where do you show in your study that model resolution is important for your results? I don’t think that you can conclude this from your study.

We are simply stating that the model spatial resolution needs to be sufficiently fine to represent connectivity features and processes. If we were using 30 m resolution data, we would not even be able to perform this study. For instance, we would not be able to represent the roads or the grass strips.

L444-445: In L285 you state "Hence, modelled hillslope yields and suspended loads are not fully commensurable, and we did not focus on a rejectionist framework for model testing." Here, you state that "soil redistribution rates and patterns are intrinsically linked to linear features". The strength of the latter statement does not really fit to the caution you demand in the first statement. Therefore, you should reformulate this sentence.

We partially agree with this comment. As we explained above, our conclusions are not solely based on the comparison between model outputs and the catchment sediment loads. We rephrased, referring to field-based studies that should allow us to state (L500-507):

"In a wider context, our study has demonstrated how structural sediment connectivity patterns can be investigated with a conceptual model as WaTEM/SEDEM, provided that model resolution is sufficiently fine to represent relevant features and processes. In agricultural catchments of the Swiss Plateau and likely in other patchy landscapes, soil redistribution rates and patterns are intrinsically linked to linear features (Alder et al., 2015; Ledermann et al., 2010; Prasuhn, 2020; Remund et al., 2021). Hence, in order to provide relevant system descriptions, soil erosion models applied under similar conditions must be able to represent linear features and landscape patchiness".

L472-475: I very much like this part. In contrast to the discussion, I feel that here the strength of statements fits together with what you did in your work and with the related uncertainties.

Thanks, we appreciated your constructive criticism throughout the text, which helped us improve our paper.

L478: the effects -> the potential effects

Upon consideration, we do not think the effects of linear features on sediment connectivity are potential – they have been observed in multiple field studies, as the ones we cite in the discussion. These effects have been highlighted by our work, due to the sensitivity of the model to the road connectivity assumptions and the presence of grass buffer strips.

L484: Would you not rather recommend a proper validation of you model before upscaling it?

This is precisely what we tried to recommend, sorry if it was not clear. We updated the text accordingly (L559-560).

Technical corrections

L11: "In particular": Seems to be the wrong transitional phrase. Please rewrite.

“In particular” is a transitional phrase used to illustrate or explain an idea, which was our goal here.

L13: “grass-buffer-strips”: Is that the correct term/correct spelling? (Revise in whole manuscript.)

Thanks for noticing. We checked and this should not be hyphenated, we corrected it throughout the manuscript.

L31: increase -> increases

Thanks, changed to increases.

L44: infra-structure -> infrastructure

Thanks, corrected through the text.

L75: weekly -> weakly

Thanks, corrected.

L89: Baldegg Lake -> Lake Baldegg (revise in whole manuscript; see for example https://doi.org/10.1039/D0EM00317D)

L97: of the canton of Lucerne. (revise in whole manuscript)

L111: field-blocks -> field blocks (Revise in whole manuscript.)

All hyphenations and place names were revised.

L114: Consider just writing km^{-1}

We think km km^{-2} is more intuitive and we would like to keep this.

L120: Elevation -> The elevation

Updated as suggested.

L141: water discharge values -> discharges

Updated as suggested.

L194: Make a reference from this. No URL directly in the text. Check if there's a permanent identifier/URL.

We removed the URL.

L221: Wrong table referenced.

Thanks, updated to Table 2.

L235-238: Difficult to read. Rephrase.

We would appreciate more guidance here. What did you find difficult to understand?

L292: This can be easily visualised -> This is shown

Updated to "this is illustrated".

L257: Is it a hydrological or hydraulic shortcut?

Corrected to hydraulic throughout the manuscript, thanks.

L306: increased in -> by or to

Thanks, changed to "by".

L372: In Figure 7, you write "short-cut", while in the whole manuscript you wrote "shortcut". Please adapt.

Thanks, this was corrected in Figure 7.

References

Alder, S., Prasuhn, V., Liniger, H., Herweg, K., Hurni, H., Candinas, A. and Gujer, H. U.: A high-resolution map of direct and indirect connectivity of erosion risk areas to surface waters in Switzerland-A risk assessment tool for planning and policy-making, Land use policy, 48, 236–249, doi:10.1016/j.landusepol.2015.06.001, 2015.

Borrelli, P., Alewell, C., Alvarez, P., Alexandre, J., Anache, A., Baartman, J., Ballabio, C., Bezak, N., Biddoccu, M., Cerdà, A., Chalise, D., Chen, S., Chen, W., Maria, A., Girolamo, D., Desta, G., Deumlich, D., Diodato, N., Efthimiou, N., Erpul, G., Fiener, P., Freppaz, M., Gentile, F., Gericke, A., Haregeweyn, N., Hu, B., Jeanneau, A., Kaffas, K., Kiani-harchegani, M., Lizaga, I., Li, C., Lombardo, L., López-vicente, M., Lucas-borja, M. E., Märker, M., Matthews, F., Miao, C., Modugno, S., Möller, M., Naipal, V., Nearing, M., Owusu, S., Panday, D., Patault, E., Valeriu, C., Poggio, L., Portes, R., Quijano, L., Reza, M., Renima, M., Francesco, G., Rodrigo-comino, J., Saia, S., Nazari, A., Schillaci, C., Syrris, V., Soo, H., Noses, D., Tarso, P., Teng, H., Thapa, R., Vantas, K., Vieira, D., Yang, J. E., Yin, S., Antonio, D., Zhao, G. and Panagos, P.: Science of the Total Environment Soil erosion modelling : A global review and statistical analysis, , 780, doi:10.1016/j.scitotenv.2021.146494, 2021.

Ledermann, T., Herweg, K., Liniger, H. P., Schneider, F., Hurni, H. and Prasuhn, V.: Applying erosion damage mapping to assess and quantify off-site effects of soil erosion in Switzerland, L. Degrad. Dev., 21, 353–366, 2010.

Panagos, P., Borrelli, P., Meusburger, K., Alewell, C., Lugato, E. and Montanarella, L.: Estimating the soil erosion cover-management factor at the European scale, Land use policy, 48, 38–50, doi:10.1016/j.landusepol.2015.05.021, 2015.

Remund, D., Liebisch, F., Liniger, H. P., Heinimann, A. and Prasuhn, V.: The origin of sediment and particulate phosphorus inputs into water bodies in the Swiss Midlands – A twenty-year field study of soil erosion, Catena, 203(March), 105290, doi:10.1016/j.catena.2021.105290, 2021.

Schmidt, S., Alewell, C. and Meusburger, K.: Mapping spatio-temporal dynamics of the cover and management factor (C-factor) for grasslands in Switzerland, Remote Sens. Environ., 211(April), 89–104, doi:10.1016/j.rse.2018.04.008, 2018.