We are grateful for the time that volunteers have put in to review our paper. Reviewers have been thoughtful and thorough. Our paper is now stronger thanks to their comments. Our responses to their comments are shown in red below. In cases when the comments were long, we bolded what we thought was the essence of the comment for us to respond to.

Sincerely,
Nicole Gasparini, for all authors.

Reviewer 1 comments:

2 Overall Feedback

The structure of the manuscript is easy to follow. However, information is not always presented in a clear and straightforward way. The authors seem to try to include all possible different viewpoints and to show e.g. the widest spectrum of possibilities of application for some methods (e.g. L. 33– 47). The authors also focus a great deal on the process of how they arrived at certain details of their model set-up instead of simply stating their choice and reason (e.g. L. 141–151). **Considering that the manuscript is intended as short communication, there still is potential to shorten very detailed explanations to the quintessence of what is relevant to the topic.**

We find that most modelling studies do not provide enough information to recreate the experiments. We prefer to explain what we did and let reviewers decide to skip it than not make information available, so that reviewers cannot recreate what we did or understand why we did it.

Concerning the choice and "calibration" of methods, the **connection between the chosen metrics and thresholds used for the determination of the time to steady state in numerical simulations versus the analytical assumption does not become clear.**

We now have a motivation section that states how other papers describe methods used to reach steady state. We hope this will make it more clear why we have chosen the metrics we have chosen. The short answer is that essentially there are no established metrics.

The authors focus on an analytical topographic steady state definition that assumes steady state is achieved when knickpoints fully migrated through the system. It is not discussed how knickpoint migration affects their metrics and how this can be used as alternative method to refine their measurement of numerical time to steady state TE. The question of acceptable variability in the chosen metrics is raised, but not discussed with regards to the magnitude of fluctuations due to numerical artefacts or ongoing small-scale rearrangements of flow directions that complicate the automatic detection of response time in numerical simulations.
In that regard, **I think the usage of the terms "response time" and "steady state" should be used non-synonymously where the numerical model is concerned**. **The analytical solution TA for steady state is defined to be equal to the response time, but I think the central result that can be drawn from the evaluation of the numerical simulations is that the same does not apply there.** I think the logical conclusion is that two times need to be derived from the metrics, one for response time (also based on assummptions on knickpoint migration) and one for steady state in its stricter sense (h
t = 0).

We have added more discussion of the difference between topographic steady state and time to steady state, as predicted by the analytical equation.

I do not agree with the final conclusion that numerical simulations are basically unsuited to derive response times. It is certainly true that the most simple method of automatic detection does not work, but **that problem can be approached by separating between model behavior and numerical artefacts and finding a refined method for measurement that is less sensitiv to the latter.**

We are encouraged that the reviewer thinks the problem is tractable. We hope that the community will recognize the "nuances" of model behavior in the future, which is why we wrote this paper. We have not seen such a refined method as the reviewer describes used before, and we would appreciate a reference for the publication that describes such methods.

The other question is how much we need to depend on numerical simulations to determine response time under detachment-limited conditions, when an easy-to-use analytical solution does exist for that case. I appreciate the effort of the authors to investigate the applicability of landform evolution models (LEMs), but in many parts of the manuscript the question arises how much more important it would be to include different erosion laws into the comparison.

We do not argue against the use of the analytical solution. We only mean to illustrate that one needs to be aware that determining steady state, and time to steady state using a model, should be done with caution. Because many model results are presented on a "steady state" landscape, it is important for the community to reach agreement as to what this means and not over interpret numerical artifacts.

Using different erosion laws is beyond the scope of this study.

I think the manuscript needs at least a serious overhaul regarding the evaluation and discussion if it should be considered for publication. **The findings of the authors differ significantly from published literature regarding the reliability of the investigated models, but the causes are not identified or discussed properly**. I assume that most of the issues disappear if the evaluation of the metrics is revised and that **the results will mostly reflect what is already published on the behavior of these models**. Based on the introduction, I initially expected that the authors at least provide a recommendation how to deal with numerical artefacts, but **I was disappointed that the evaluation is based on what I perceive to be a mostly arbitrary threshold that has nothing to do with the chosen analytical estimation for response time**. I really regret to recommend rejection.

We are not aware of and could not find the published literature regarding the reliability of our study models. We wish the reviewer would have provided references. We are unsure what the reviewer means when they say "the results will mostly reflect what is already published on the behavior of these models". Without references, it is difficult for us to respond to this comment.

We removed the arbitrary threshold.

3 General Remarks

• I still miss the discussion of the applicability of detachment-limited erosion (and the derived theoretical steady state criterion presented in Eq. 9 and 10) where the interpretation of natural landscapes is concerned. **I raise this point (again) because the evaluation of natural landscapes is explicitly stated as main motivation in the introduction and because the authors insist on a very low and strict threshold for measuring time to steady state in their numerical models.** My main points are these:

Many studies use detachment limited models to understand natural landscapes. We have removed our "very low and strict threshold".

– Sediment transport and other secondary erosion processes (hillslope processes, aeolian processes, etc.) have a significant influence on response times (i.e. knickpoint velocity). The presented steady state criterion is based on the theoretical time a knickpoint needs to travel through the entire catchment (according to Whipple (2001)). **Doesn't this criterion only hold for purely detachment-limited conditions?**

We are unsure what the reviewer is saying. We use a detachment-limited model, and we compare it with the analytical solution for the detachment limited model. We have made changes that clarify that knickpoint propagation time and modeled time to steady state are not necessarily the same thing.

– **Is it not to be expected that variations due to different erosion laws are much higher in magnitude than the variations shown here between different numerical implementations of the same erosion law investigated here?** What is your estimation to that regard? I expect that the correct interpretation of natural landscapes depends much more on the choice of the correct erosion law than on differences purely due to numerical implementation.

Possibly variations from different erosion laws would be higher in magnitude. In our mind, that makes it even more compelling that the community come to agreement about what they are measuring.

– The argument that natural landscapes might behave different from what the model assumes is raised in the end **(L. 307f),** seemingly to deflect from the very high apparent discrepancy between theoretical response time and response/steady state time measured from the simulations. But the problem seems to lie in the method of evaluation and not model choice there, the analytical solution used for comparison is essentially derived from the basic model equation (with some integration and addition of Hack's empirical observation) and should be reproducable with the numerical model. The topic of initial model choice should be discussed right at the start.

The paragraph referred to above was deleted. We have added more discussion about the differences between the analytical solution and the model behavior.

I really only encourage to add a short clarification of the limitations where the importance of results of this study are concerned, not a detailed overview. Some assumptions for the evaluation of steady state/response times presented here probably cannot be transferred directly to other LEMs using different erosion laws.

We are not quite sure what the reviewer is asking for here. This may or may not address this comment, but we did add a Motivation section to the paper that discusses the lack of agreement or standard processes within the community. (At least there is a lack of agreed process in published studies that we are aware of.)

• In response to the reaction to my first review: I certainly did not wish to imply that errors in implementation and interpretation are not made by the user community. The point I tried to make is that the mathematical theory to correctly predict e.g. accuracy and stability of numerical schemes, i.e. the explicit scheme, already exists for several decades and it would have been very useful to apply this theory in the context presented in the last version of the manuscript, especially when a younger generation of model users without a rigorous mathematical background is addressed. It is my suspicion that model developers, especially when coming from a strong mathematical background, consider such knowledge basic on a textbook level and omit the details in model descriptions "because it is self-evident". The scope of the manuscript has changed, so **I expect no formal answer**, just felt the need to clarify why I am very insistent on some of what I consider to be fundamental points.

No answer is provided.

4 Major Remarks

4.1 Introduction

• L. 34–47 It seems your study only considers a one-time change in conditions and does not contribute to the scenarioes where conditions change continuously or faster than the response time. Maybe this part can be shortened accordingly?

We are not sure what the reviewer is asking of us.

4.2 Section 4 Experimental set-up

• L. 107–110 If you know that the landscape was static, it seems that you formally measured something.

Maybe something along the line "Initial conditions were static according to the criterion established in the discussion" is sufficient here?

We did not change anything.

• L. 116f Maybe explain the general setup and similarities between models before mentioning specific exceptions to simplify comprehension (from general to detail, so to say).

We did not change anything.

• L. 125 Is that simply a circumscription for Delauney triangulation? What is the average variation and what do you mean with "variation in a regular way" if the grid is actually irregular?

Changed to : The numerical experiments that use a Voronoi grid are created from a Delaunay triangulation of staggered rows of nodes spaced 100 m apart in the x direction.

• L. 135 correct within numerical errors, absolute accuracy with regard to the analytical solution usually still depends on dt (not only on machine precision/rounding errors).

Yes. We did not change anything.

• L. 141–149 The quintessence seems to be that the largest possible catchment area was deliberately overestimated to ensure a stable simulation, and rounded to have a nice dt value. This passage can be shortened.

Shortened.

4.3 5.1 Numerically modeled

• L. 196 Looks like Equation (8) is essentially the temporal derivative of equation (6)? Considering that Figures 1 and 3 are hard to decipher because of the many subplots and that the graphs of both metrics seem to deliver the same information regarding time to steady state, one of them could maybe be discarded to gain more space?

We have changed the figures to include two that are easier to read. We kept versions of Figures 1 and 3 for completion.

• L. 210 I think the graph should be interpreted by clearly dividing between the causes that are affecting the behaviour (e.g. what is the actual model behaviour and what are numerical artefacts). The first part of the graph (up to roughly TA) seems clearly affected by knickpoint migration and reflects where most of the adjustment of erosion rates towards a new steady state should be happening, and the second part seems to be dominated by numerical artefacts and some effects of (potentially related?) divide migration as suggested by the authors (possibly masking some knickpoint-related behaviour?).

We are unsure what the reviewer is asking of us.

Concerning the first, Whipple and Tucker (1999) and Whipple (2001) state that the equation for theoretical time to steady state is derived from the response time of catchments that is controlled by knickpoint migration (speed at which information can be transmitted through the system), and that steady state can at the earliest be achieved when knickpoints have migrated through the entire system. I propose that the metrics are quite clearly affected by knickpoint migration and that the change in behaviour of the graphs from relatively smooth trends to more chaotic swings occurs when knickpoints arrive at the divide in your modeled scenarios. I assume that the second phase is mainly influenced by what I will refer to as "numerical artefacts" for simplicity, because I strongly suspect that discretization of the grid is the driving factor that keeps local rearrangement of flow directions going on for such a long time. I stated before that I often observe a localized switching of flow directions. The reason seems to be due to the discretization of the grid that does not allow exact adjustment of stream lengths to the ideal equilibrium length (since increments

are limited by the gridspacing and there is also competition between catchments). At least anomalies in the steepness ks that appear as a group of too low and too high ks values seem to support that assumption. I would be interested if you observe the same in your simulations, or if you observe a different cause.
For the evaluation, I also think it should be considered that the variations occurring after the analytical response time are several magnitudes lower than the differences in the initial phase of your graphs. They are measurable in a computer simulation, but are they significant enough to be seriously regarded or should we find a way to filter them out of our results, especially if they might be numerical aretefacts and compared to all the much larger uncertainties that probably arise from parameter estimation or choice of erosion law?
In summary, I suggest separating between response time and time to steady state in numerical simulations and evaluate both times.

We have removed the strict interpretation of experimental time to steady state.

• L. 214f I cannot find a good explanation for the particular choice of these threshold values in the following. They remain arbitrary. Please state the criterion that you used. If you just used the lowest threshold that presented itself to you and used it for all simulations, how should a person choose a threshold if another software is used? Also, it seems that some metrics could use a higher threshold. This would partly significantly affect the results you presented in the following.

We no longer use this threshold value.

• L. 215 "Conservative" seems an understatement. Less than 0.1 nanometer vertical change per year does not seem to be a practical threshold for steady state, especially not where comparison to natural landscapes is concerned.

We no longer use this threshold value.

• L. 235 I understand that ka and h are not supposed to change much by divide migration. But how large is the variation between different catchments across the grid and does this introduce an error for the time to steady state calculation (especially where the detection of knickpoint migration in the simulations is concerned)? Did you perchance test the evaluation of (maybe only a few of the largest) individual catchments with the respective individual geometric parameters to test if this improves the detectability of response time/steady state? On that note, I believe that Whipple observed that vertical knickpoint velocity equals uplift rate, so could that not be used for a response time estimation that is less dependent on additional parameters that potentially add uncertainties to the evaluation?

We did not do more testing of the hack parameters. We did not do further tests of estimates of response time.

4.4 6 Time to steady state results
• L. 250 Is this a counterargument for using the knickpoint criterion? A systematic deviation in dependence of timestep is theoretically to be expected. Depending on the numerical scheme rates of change are systematically over or underestimated and the effect becomes usually worse with larger timesteps.

This was deleted.

• L. 251f Not clear if this is intended as a counterargument, or why you expect a smooth decrease for maximum elevation. The change in maximum elevation seems to behave as it theoretically should. Erosion rates are adjusted only below the knickpoint, mainly by adjustment of the slopes. Above the knickpoint all nodes including the highest peak are constantly eroded with the old erosion rate and at the same time elevated with the new uplift rate. Relative topography should not be affected by baselevel drop/baselevel increase until the knickpoint passes (so the maximum elevation can be expected to remain perfectly stationary at first). The resulting constant net elevation increase (40 m/1e5yr) of the peak seems to be reflected in your graph. A sudden change/drop in your metric is to be expected at the time when knickpoints arrive at the peak, which should also include the last knickpoint in the system since the peak should also be associated with the longest stream and longest knickpoint travel time.

This was clarified.

• Figure 1 is not readable at all. Beside reduction of the number of subplots, **I also suggest that a magnification of the initial phase (until the analytical time to steady state) is provided.** In the current state it is not possible to judge properly how the behaviour of the graphs changes at this important time because the interesting part is squished to the left boundary. Have you tried a double-logarithmic plot? There are several solid black lines in some of the diagrams where there should only be one according to the figure decription. Also, you include an additional analytical time to steady state based on an average river length, but do not explain or discuss it in the main text. I would be very interested how you calculated this, especially because this estimation of response time is a perfect fit to the empirical response time in Fig. 3 for max elevation and the knickpoint preserving algorithm (where I expect that empirical response time should be closest to TA based on the specifics of numerical implementation).

We magnified the initial time of two sets of experiments and made these figures 1 and 2. We kept the former figure 1 (it is now figure 3) because it seemed inappropriate to not illustrate all of the results. I find log time plots difficult to interpret, so I opted not to make that change. We removed the lines from these plots.

• L. 261 Can you include an explanation why the Voronoi grid suffers less from network rearrangement? The advantages/disadvantages of the Voronoi are not discussed before, although it seems from the model descriptions that it was deemed very important to include it. I recall that Voronoi grids are better suited for the representation of natural landscapes because they are more flexible, but need more effort to generate because triangles near the boundary tend to be disproportionally long.

We did not include an explanation of why the Voronoi grid has less rearrangement. We are not 100% sure why. It is not "very important" to include the Voronoi results, but we advocate for better understanding for the impact of different grid types on numerical modeling. Most of that aspect of our study is removed from this version of the paper because the reviewers were not supportive of the first version of this paper. But we kept in the Voronoi results because they help illustrate the impact that network rearrangement has on time to steady state. We can still remove these results though, if this version of the paper is invited for further revisions.

• L. 265 TE as measured by the threshold does probably not represent response time in the sense of TA. Also, would it not be better to just evaluate the largest catchment (that was also used to to estimate TA) if is to be expected that variation of ka, h and L add uncertainty to the results? Every metric except maximum elevation is probably affected because they would essentially middle between all the different knickpoint arrivel times.

We removed T_e calculations.

• L. 270 How are the numerical methods on a Voronoi grid different if they are supposed to be an implicit and explicit method respectively?

This was deleted, but good catch. We meant the flow routing algorithms are different.

• L. 273 If a threshold does not work reliably, have you considered another approach better suited to detect response times? It seems to me that using the temporal derivative of maximum elevation would work very good to find the time when knickpoints start to arrive at the peak.

We removed the threshold approach.

• L. 278 Do you assume that Braun and Willett (2013) are wrong or do you have a hypothesis why your results are so different? Even if you find that your results depend on initial flow directions, you seem not to be able to reproduce the relationship that you found in the literature?

No, of course we do not think that Braun and Willett are wrong!

• Figure 2 Please include the results of the time of completed knickpoint migration as ocurring in the simulations versus analytical response time (Eq. 9).

We did not calculate the time of completed knickpoint migration.

• 298 Your results seem to nicely show the effect of numerical diffusion/knickpoint smearing as well. TRT reduces knickpoint smearing and shows the nicest result for maximum elevation.

We are unsure of what the reviewer is asking of us.

• Figure 3 I am puzzled by what appears to be a solid double line in the first row. There also is a sentence fragment at the end of the figure description. Also, your derived results here appear to be particularly sensitiv to your choice of threshold, and I do not see why a higher threshold should not be used.

Threshold was eliminated. Lines from figure 3 were deleted.

• L. 303 I was under the impression that your main conclusion is that the metric combined with your single threshold is not suitable at all to detect response time.

That text was changed.

• L. 311

– Don't you state here that the assumption of detachment-limited erosion is inaccurate where the interpretation of natural landscapes is concerned, thereby devaluating your choice of numerical model (at least as far as the motivation for this study as stated in the introduction is concerned)?

That text was deleted.

– I think the more important question in this context is how to formulate the right criterion to reliably measure response times in this type of numerical model. Yes, the estimation for steady state presented in Eq. 9 might not be the ideally representative criterion where natural landscapes are concerned, but the numerical model is still required to reproduce predictions that are essentially derived from its model equations (if not, it is either an error in implementation or a numerical artefact). It seems quite clear that some unexpected fluctuations prevent perfect steady state (h
t = 0 during simulations
here, so model topographic steady state in the strictest sense obviously does not equal response time as predicted by knickpoint migration. I am surprised that you do not discuss refinement of the evaluation, e.g. by considering the potential effect of numerical artefacts and thresholding accordingly, trying to combine metrics or by changing metrics to make them better suited to the detection of knickpoints, maybe try to focus evaluation on single catchments to achieve a better precision.

I think some of our changes are along these lines.

• L. 327 At this point I think it becomes essential to discuss the possibility of numerical artefacts on the results of this study. It is tempting to see the counterpart of a natural process, but the cause of numerical artefacts are usually unintentionally an caused purely by limitations of numerical implementation and cannot be transferred to nature (nature is not limited by a discrete grid spacing, for example)

We now mention numerical artifacts.

• L. 334 **I think it is explicitly stated that TA is to be onsidered a minimum estimate in Whipple (2001).** The influence of divide migration and resulting error on TA can be estimated (or monitored during your simulations). Do you have any indication that the large discrepancy you present can realistically be explained by the effect of divide migration or network instability, as proposed in L. 316?

Yes, you are correct. We have clarified that.

5 Minor Remarks

• L. 63 "voronoi" is a proper noun I think?

Fixed.

• Table 1 Maybe change to "implicit (Fastscape)" in the last column for uniform style.

Done.

• L. 137 "v is the speed at which"?

Changed.

• L. 167 Waht do you mean by "empirical model"? Not the numerical model?

Fixed.

• Eq.5 i is the index of the node, I presume?

Fixed.

Reviewer 2

I am very interested in the finding that drainage rearrangement drives the longer-than-theoretical and inconsistent landscape response times. I comment on this a bit more below, but I have a couple of thoughts regarding possibliities. I am not sure if you will find them in-scope or out-of-scope, and leave this for the authors to decide.

First, **I thought of Kwang et al. (2021), who note that the lateral dynamics of rivers can be quite important in sustaining a dynamic, rather than static, equliibrium topography. Do you think that "noise" in erosion rates caused by this process could further delay the approach to equilibrium beyond the Whipple & Tucker (1999) analytical calculation?** This might further underscore your point towards the end of the article: even though the lateral-erosion component is not included in your tested LEMs, adding more expected realism might further separate from the theory rather than making the model closer. What do you think?
Note that I was Jeffrey Kwang's postdoc advisor and a graduate-student colleague with his co-author Abby Langston. Please do not feel any pressure to cite this work; there may be other papers out there, and I just know this one.

Yes, we do think these other processes would impact time to steady state. We have added this to our discussion.

Second, I considered how drianage basins form in the first place, as the Whipple & Tucker (1999) theoretical predction is based on the idea that the channel already exists. However, we see that the initial integration of drainage networks can produce lags in landscape evolution (cf. Lai & Anders, 2018: in this case, even when flow across internal depressions is allowed).

I then wondered about the more fundamental difference between propagation of a signal up an existing river and catchment (what Whipple & Tucker, 1999, were considering for "response time") and the time it takes for that order to co-develop alongside the signal propagation (what you are measuring in the models, albeit, with the additional consideration of what the models themselves are doing). This is all quite interesting, to me, in two specific ways. First, it lines up with some recent work demonstrating that cross-basin draingae integration may likely occur via overflow events (Hilgendorf et al., 2020) and that headward erosion is unlikely unless a channel already exists. Second, it underscores that we may have "noise" in how real landscapes respond to perturbations because of the nonlinear interactions of drainage sub-basins playing out atop whatever initial conditions existed, potentially giving both limit and geological guidance to geomorphic predictions.

We interpret this text as comments without anything for us to change or add.

[minor] The article is 342 lines of text+equations, with three large figures and one table. I checked the definition of "short communication" and confirmed my memory that this manuscript type should be "a few pages only". I would see it as a considerable overstep of my role here to discuss which kind of manuscript it *should* be, but I want to point this out to the editors and authors.
https://www.earth-surface-dynamics.net/about/manuscript_types.html

We leave it to the editors if this qualifies as a short communication or not.

LINE BY LINE

7
"The sensitivity of time to steady state to computational time step is not consistent among models or even within a single model."
The 2x "to" make the meaning ambiguous. I suggest:
"Time to steady state varies inconsistently with time-step length, both within a single model and among different models."
Is that right? I hope so. Otherwise I really didn't understand.

This sentence was changed as suggested.

95 (and other equations)
[minor] I think that ESurf has commas or other punctuation after equations, such that they fit into the surrounding sentence structure

I will wait for the copy editors on this suggestion, as I can't make that format correctly in Latex.

106
timesteps --> time-step lengths?

Changed.

132 (& hereafter)
[typographical]
1e-4 to 10^{-4} (imagine it being typeset) and so on

Changed.

133
Because you are discussing the workings of LEMs, might it be useful to consider the findings of Kwang & Parker (2017), that fluvial-only LEM outputs are scale invariant with m/n=0.5? I am wondering if this might be generally helpful when discussing the applicability of your results: despite your careful consideration of standardized initial conditions, you have also picked a parameter set that should be scale invariant.

I'm unsure what is being asked here.

136 (Eq. 3)
Could you pull the "v" outside of the differential operator?
Alternatively, you could change the "d"s to "\Delta"s, which would make this become a fraction instead of an operator.

Changed.

150-160
[self-narration; no response needed]
I am anxious about how the stability of your time steps might relate to your findings here, especially because only a subset of the models tested will be used for certain time steps. Because of how drainage-area changes can affect

knickpoint celerity, and drainage-capture events can occur instantaneously, I am likewise wondering if this could turn into the major time-step dependence. I'm continuing to read with this in mind.

152-153
Inline citation; consider "shown by"
Changed.

183 (& following)
Why "100,000"? And this is presumably years. It feels arbitrary, so a small note could help.

Explained below the equations.

190
Could this more precisely be called "maximum temporal change in elevation" (as opposed to "temporal change in maximum elevation")? Then the English would follow the math closely.

Changed.

195
A pair of nitpicks, one more real and one more convention.
The more "real" one is that dissolved, colloidal, etc. load are not considered to be "sediment". Therefore, perhaps you could consider noting the temporal change in total material removed from the landscape. This says what you are going for while removing the implicit assumption that eroded material becomes sediment.
Second, "flux" is mathematically and traditionally used as the transfer of a quantity across a plane, such that its units are [thing]/[area*time]. This is not how the sedimentologits & al. use it, or the climate scientists. So that's why this is a nitpick.

We changed this a bit, although we are using what we think is standard language in landscape evolution models, and we aren't sure what else we would call it.

214-218 (please ignore; including for narrative of my thinking during the review process)
This threshold seems arbitrary. Might it be better to consider a time scale that emerges naturally from the model outputs, e.g., finding an exponential decay time scale and defining some multiple of this?
On second thought, I scrolled down to your plots. I suppose they aren't all that exponential!

224 (Eq. 10)
Could you define "L"? (I've double-checked and hope I'm not just missing it!) I'm guessing that it is some kind of length scale -- perhaps half of the domain width?

Done.

Table 1 + Figure 1: This is a small thing, but I wonder if you could place the experiments within both of these in the same order. It would help in interpreting the results.

Changed.

Figure 3: Some dangling text has been left in the caption at the end.

Fixed

334-335 (also a bit of my own thinking, use/don't use at will)
I agree about this being the minimum response time, as (with your above note) the channels might not until later extend all the way up to the drainage divide with a steady-state basin area. I think that one other implication of this important point (though perhaps also beyond the scope of your article here) is the oft-invoked but seemingly mythical beast of channel headward propagation, used often in arguments of drainage integration. Knickpoints can propagate quickly upstream, but rearranging the channel head where drainage area --> 0 can be quite slow.

Reviewer 3

General comments
Dr. Gasparini and co-authors presented, as revealed by the title, issues with calculating time to steady state in landscape evolution models based on various time steps and landform metrics and compared that time between different LEMs. Compared to the original version, this revised manuscript is more centered on one specific theme (i.e. time to steady state) for model intercomparisons. I appreciate how the sections are now structured, with sufficient (but not too much) details of the theory, LEMs, and their methodology presented. Their results are also clearly explained and overall convincing. I especially agree with the implication of their work that "… it is critical to both report the metric being used and the threshold value for that metric to assess steady state (line 302)". I believe this work will encourage further investigations on other issues related with benchmarking LEMs, and this work provides a nice reference for those future studies. That will surely benefit the whole geomorphologic community. With all these thoughts, I recommend publishing this work after very minor revision.

We kept the line that this reviewer particularly liked.

Specific comments
You used different landform metrics and time steps and a threshold to calculate the time to steady state. I am curious about the values of those metrics (max Z, mean Z, local max Z, flux) at steady state and how different they are between models. Could you add those values in the subplots of Figure 1?

We did not add these values, simply because we were trying to stay focused and create a "short" communication.

Technical corrections
Figure 1: change the unit of dt to be yrs, making it more consistent with the main text. Same applies to Figure 3.

Because the values are so long we left them as My. This makes the plots easier to read, at least to us.

Line 315: cite Li et al 2018 in bracket.

That sentence was deleted.