Peer-Review of "Short Communication: The trouble with time to steady state calculated from computational landscape evolution models"

Anonymous

January 24, 2024

1 Summary of the Content

The authors present a study that investigates the quality of time to steady state evaluation in numerical simulations of landscape evolution based on a detchment-limited erosion law.

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.

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. 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 T_E. 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 T_A 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 ($\frac{\Delta h}{\Delta t} = 0$).

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.

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.

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.

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:
  - 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?
  - 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.
  - 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.

  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.

- 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.

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?

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?

- **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).

- **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?

- **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).

- **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.

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?

- **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 T_A) 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?).

  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.

- **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.

- **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.

- **L. 235** I understand that k_a 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?

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.

- **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.

- **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 T_A based on the specifics of numerical implementation).

- **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.

- **L. 265** T_E as measured by the threshold does probably not represent response time in the sense of T_A. Also, would it not be better to just evaluate the largest catchment (that was also used to to estimate T_A) if is to be expected that variation of k_a, 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.

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

- **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.

- **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?

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

- **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.

- **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.

- **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.

- **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)?
  - 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 ($\frac{\Delta h}{\Delta t} = 0$ 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.

- **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)
- **L. 334** I think it is explicitly stated that T_A is to be onsidered a minimum estimate in Whipple (2001). The influence of divide migration and resulting error on T_A 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?

5 Minor Remarks

- **L. 63** "voronoi" is a proper noun I think?
- **Table 1** Maybe change to "implicit (Fastscape)" in the last column for uniform style.
- **L. 137** "v is the speed *at which*"?
- **L. 167** Waht do you mean by "empirical model"? Not the numerical model?
- **Eq.5** i is the index of the node, I presume?