Short Communication: ~~Motivation for standardizing and normalizing inter-model comparison of~~ The trouble with time to steady state calculated from computational landscape evolution models

Nicole M. Gasparini[1], Adam M. Forte[2], and Katherine R. Barnhart[3,4,5]

[1]Tulane University, Earth and Environmental Sciences Department, New Orleans, LA, USA
[2]Department of Geology & Geophysics, Louisiana State University, Baton Rouge, LA, USA
[3]University of Colorado at Boulder, Cooperative Institute for Research in Environmental Sciences, Boulder, CO, USA
[4]University of Colorado at Boulder, Department of Geological Sciences, Boulder, CO, USA
[5]Now at U.S. Geological Survey, Geologic Hazards Science Center, Golden, CO, USA

Correspondence: Nicole Gasparini (ngaspari@tulane.edu)

Abstract. ~~This manuscript is a call to the landscape evolution modeling community to develop benchmarks for model comparison. We illustrate the use of benchmarks in illuminating the strengths and weaknesses of different~~

Quantifying the timescales over which landscapes evolve is critical for understanding past and future environmental change. Computational landscape evolution models ~~(LEMs) that use the stream power process equation (SPPE) to evolve fluvial landscapes. Our examples compare three different modeling environments—CHILD, Landlab, and TTLEM—that apply three different numerical algorithms on three different grid types. We present different methods for comparing the morphology of steady-state and transient landscapes, as well as the~~ are one tool among many that have been used in this pursuit. We compare numerically modeled times to reach steady state for a landscape adjusting to an increase in rock uplift rate. We use three different numerical modeling libraries and explore the impact of time step, grid type, numerical method for solving the erosion equation, and metric for quantifying time to steady state. We ~~illustrate the impact of time step on modelbehavior. There are numerous scenarios and model variables that we do not explore, such as model sensitivity to grid resolution and boundary conditions, or processes beyond fluvial incision as modeled by the SPPE. Our examples are not meant to be exhaustive. Rather, they illustrate a subset of best practices and practices that should be avoided. We argue that all LEMs should be tested in systematic ways that illustrate the domain of applicability for each model. A community effort beyond this study is required to develop model scenarios and benchmarks across different types of LEMs.~~ find that modeled time to steady state is impacted by all of these variables. The sensitivity of time to steady state to computational time step is not consistent among models or even within a single model. In some cases, drainage rearrangement extends the time to reach steady state, but this is not consistent in all models or grid types. The two sets of experiments operating on voronoi grids have the most consistent times to steady state when comparing across time step and metrics. On a raster grid, if we force the drainage network to remain stable, time to steady state varies much less with computational time step. In all cases we find that modeled time to steady state is longer than that predicted by an analytical equation. Our results show that the predicted time to steady state from a numerical model is, in

many cases, more reflective of drainage rearrangement than the time for an uplift wave to propagate through a fixed drainage network.

1 Introduction

Thirty years ago there were only a handful of computational landscape evolution models (LEMs, Ahnert, 1976; Willgoose et al., 1991; Chas . Now there are multiple review papers on the theory behind, differences among, and uses of different computational landscape evolution models (Coulthard, 2001; Martin and Church, 2004; Tucker and Hancock, 2010; van der Beek, 2013; Temme et al., 2013; Chen . Yet, even as the number of computational landscape evolution models and modeling environments proliferate, models become more sophisticated, and modeling becomes a standard tool for testing hypotheses in the geomorphology community, the practice of inter-model comparison is not generally practiced within the LEM community.

All inter-model comparisons seek to compare models in some systematic way. One common type of inter-model comparison is called a 'benchmark' and typically involves a known answer. This known answer may derive from the analytical solution to the model governing equations, or anotherapproach, such as the method of manufactured solutions (Roache, 1998). The specification of benchmark problems are typically the result of community consensus. Model performance on benchmark problems indicates the model can reliably solve that type of problem. Benchmark problems typically have simple initial and boundary conditions such that multiple models can implement them.

Inter-model comparison is standard in the climate science community. Projects like the Coupled Model Intercomparison Project (CMIP) have been ongoing for decades (Meehl et al., 2000). (Here coupled model refers to global climate models (GCMs) that include energy fluxes among the ocean, atmosphere, cryosphere, and land.) The many iterations of CMIP have led to established scenarios for comparing model results and standardized formats of model output. Climate modeling has implications for policy and humanity. Thus the uncertainty and range in outcomes is highly scrutinized by the modeling community and the public.

However, inter-model comparison is not just a tool for the climate community. Other earth science modeling communities have embraced inter-model comparison. For example, Buiter et al. (2016) designed three numerical experiments for intercomparison of results among 11 different computational models of an accretionary wedge. If a researcher were to develop a new model, they could use the benchmark scenarios laid out in Buiter et al. (2016) to quantitatively test whether their model was providing reasonable solutions before applying it to new scenarios. Similarly, the community modeling dynamic earthquake rupture has produced a series of benchmark experiments and engaged in a process of inter-model verification of simulation results (e.g., Harris and Archuleta, 2004; Harris et al., 2009, 2011; Barall and Harris, 2015)The concepts of steady-state landscapes and characteristic timescales for landscapes to transition from one steady state to another, i.e., response times, have been widely used as framework for interpreting landforms and landscape evolution (e.g., Whipple, 2001; Whipple and Meade, 2004; Hilley et al., 2004; , and what signals may be preserved in the sedimentary record (e.g., Castelltort and Van Den Driessche, 2003; Simpson and Castelltort, 201 . If landscapes evolve to a predictable steady form that is a function of environmental drivers, presumably we could invert landscape form to infer these environmental drivers (e.g., Snyder et al., 2000; Densmore, 2004; Kirby and Whipple, 2012; Whittaker, 2012

. Further, steady-state morphology is a well-established way to compare modeled landscapes (e.g., Tucker and Bras, 1998; Tucker and Whip
.

Some model comparison studies have been performed in the LEM community, but these have primarily focused on how different processes impact model outcome. For example, there have been multiple studies that used the same modeling framework but different fluvial process equations (e.g., Whipple and Tucker, 2002; Crosby et al., 2007; Gasparini et al., 2007; Attal et al.,
or hillslope process equations (e.g., Tucker and Bras, 1998; Roering, 2008; Barnhart et al., 2019) or both (Barnhart et al., 2020b) to quantify the topographic signature of different processes. Further, both Campforts et al. (2017) and Dietterich et al. (2017) compared topographic outcomes produced using different numerical methods to solve the same process equation within the same model. Process model sensitivity to parameters has been explored (e.g., Tucker and Whipple, 2002; Barnhart et al., 2020b; Shobe et a
as well as sensitivity to parameters and grid resolution (Kwang and Parker, 2017). All of these studies are intra-model comparisons, in that they used the same modeling environment for all of the model experiments. However, there have been no inter-model comparisons or benchmark experiments of LEMs of which we are awareWhile steady state can be used to imply a variety of different conditions, here we specifically consider topographic steady state (e.g., Willett and Brandon, 2002). To determine whether a landscape has reached topographic steady state, we need to know how to measure it (either in the field or in a simulation). There are examples of two different models using the same algorithms and benchmarks for verification (de Almeida et al., 2012
, but these were done in series, not as a coordinated effort. It is reasonable to expect that different models that implement the same process equation, with the same parameters and boundary and initial conditions would lead to the same outcomes, at least statistically. However, the LEM community has never laid out how to set-up benchmark studies and what should be quantitatively compared in the model outcomes. many options for how to measure topographic steady state in a computational model, and accordingly, an initial question is which steady-state metric in a numerical model is most reliable. If we measure steady state using erosion rates or sediment fluxes, over what time scales should these fluxes be measured? Further, how much variability in sediment flux can we expect under steady conditions? If steady state is measured using landscape metrics, such as total relief, which metrics are used? Similarly, how much variability in these metrics is acceptable while still maintaining at steady state?

Here, we present results that motivate the establishment of a community effort to generate benchmark experiments of LEMs. We illustrate inter-model Once a criterion for steady state is established, a computational model can be used to measure the time it takes for a landscape to reach steady state following a perturbation. Given an understanding of time-to-steady-state, many different inferences may be drawn with implications for the interpretation of specific landscapes on Earth or other planets. If a particular configuration of initial and intra-model comparisons of landscapes modeled using the stream power process equation (SPPE). The SPPE is widely used in LEMs to describe fluvial incision as a power law function of drainage area and topographic slope (e.g., Howard, 1994; Whipple and Tucker, 1999). We perform the same experiments in three different modeling environments. In some cases, we compare the solution of the equation using different numerical algorithms in the same modeling environment. We also compare the solution of the equation using different grid types. Finally, we compare solutions using the same numerical algorithm and grid type but implemented in different modeling environments. When possible, we use the same initial condition for the different models, and in all cases we use the same boundary conditions

. We explore differences in model outcomes at steady state (here defined as bedrock incision rate equal to rock uplift rate) and following a perturbation (increase in rock uplift rate)after reaching boundary conditions results in landscapes that reach steady state relatively quickly (in comparison with variations in environmental forcings), then we can expect similarly configured real landscapes to reach steady state. The take-home message will not be surprising to anyone who has used a LEM; all numerical algorithms have limitations. However, the differences and similarities among the algorithms lead to insights on how to properly use these models and the need for established benchmarks for both current and future LEMs. However, if landscapes take so long to adjust that environmental changes happen more frequently, we might expect to rarely observe steady-state landscapes. Similarly, if we can establish that response times of simulated landscapes to perturbations behave systematically and predictably as a function of the environmental drivers, then simulated response times could theoretically be used to establish characteristic time scales for processes (e.g., Whipple et al., 2017; Lyons et al., 2020). Further, if time scales of landscape evolution and/or natural forcings are known in study landscapes, models could be used to test competing landscape evolution scenarios, interpret processes controlling landscape evolution, and establish the impact of competing timescales on landscape form (e.g., Densmore et al., 2007; Attal et al., 2008; Godard et al., 2013; Whittaker and Boulton, 2012; Mackey et al., 2014; Brocard et al., 201 . Establishing the reliability of time to steady state derived from landscape evolution models is thus a consequence of this metric's importance for interpreting landscape evolution models and connecting inferences drawn from them with datasets collected on Earth.

The goal of this paper is not to encourage users towards, or discourage users away from, the models we compare. It is also not to argue that the specific set of comparisons here are the best choice of benchmarks. Rather, it is an initial illustration of one vision of inter- and intra-LEM comparison intended to inspire community collaboration to establish benchmark problems for LEMs and support future inter- and intra-model comparison studies. This project was a side project and labor of love from three LEM community members. In contrast, other scientific model benchmarking studies were based on multiple funded meetings and input from a much larger swath of the community. We hope such an effort can grow from the seed of this paper.

2 Terminology

We define some terms that are often used with different meanings among studies. However, the specific definition that we use is important for understanding this study .

Boundary conditions: The rules that are applied around the perimeter of a modeling area so that numerical algorithms can be solved on the edges of the model domain. For example, whether water and sediment can flow into and out of nodes on model edges is a boundary condition.

Computer model: The application of a numerical algorithm or algorithms to solve process equations on a grid. In this case, the computer model is a landscape evolution model. We sometimes shorten this to model for readibility.

Geomorphic process: A physical process that is shaping the earth's surface, such as a landslide or abrasion of a bedrock riverbed by saltating sediment.

Modeling environment: The entire source code from which an individual computer model is created. For example, TTLEM is a modeling environment, but we implement different TTLEM computer In this short communication we show how time to steady state changes in otherwise identical computationally modeled landscapes. We quantify steady state in different models that use the different numerical algorithms contained within TTLEM. different numerical methods and grid types. In all of our model scenarios we evaluate four different metrics for quantifying steady state. We show that flow routing methodology is the biggest control on time to steady state in our modeling experiments, but that the impact of flow routing on drainage reorganization also varies with grid type. Similarly, computational time step and method for quantifying steady state also impact the time to steady state. In other words, we find time to steady state from numerical models to be inconsistent, and, in some cases, of minimal use when interpreting real landscapes, especially when considering the outcome of a single or very small set of landscape simulations.

Node: A location on a model grid. Variables that are required to solve a process equation are assigned to every node (e.g. elevation) or the connection between nodes (e.g. in some computer models slope is defined between nodes, in others it is defined at a node).

Numerical algorithm: A numerical algorithm uses what are commonly known as numerical methods to solve a differential equation so that it can be applied in a computer model. For example, the implicit finite difference method can be used as a numerical algorithm to calculate fluvial bedrock incision from the stream power process equation.

Process equation: An equation that describes a physical process and includes variables and parameters that can be measured, estimated, or calibrated. Examples of process equations in a LEM include sediment transport rate and bedrock incision rate equations. Process equations in geomorphology are often referred to as models because they are a simplification, or a model, of a natural process.

Simulation: A model run with a fully specified set of numerical algorithms, input parameters, and initial and boundary conditions.

We also include some scientific and computational terms with precise definitions, many of which were taken after Schlesinger et al. (1979 .

Accuracy: "the degree to which the result of a measurement, calculation, or specification conforms to the correct value or a standard." (Oxford English Dictionary accessed via Google dictionary).

Benchmark: A numerical simulation with specified initial and boundary conditions, anda known solution. Benchmarking a computer model is one part of verification.

Domain of applicability: The set of conditions for which a computer model has been verified and validated for use .

Numerically stable: An algorithm is considered numerically stable when it does not amplify errors, such as approximation or truncation errors. When an algorithm is unstable, the numerical solution diverges from the true solution, can oscillate between extreme positive and negative values, and can take on non-physical values.

Validation: Demonstration that a computer model agrees with reality within a stated level of accuracy. That is, is one using the correct model for the question at hand.

~~Verification: The demonstration that a computer model for a given system represents the stated conceptual model or set of governing equations to a stated level of accuracy. That is, solving the stated model equations correctly.~~

2 ~~Modeling environments and comparisons~~

2 Modeling environments

The three modeling environments used in this paper were chosen because the authors have experience using them. The choice of these models says nothing about the value of these ~~chosen~~ modeling environments or the variety of other modeling environments that exist. ~~Explicitly, our goal was not to develop a comprehensive inter- or intra-model comparison of these environments or others, but to show one possible way such comparisons could be done through broader community effort and to explore why such comparisons are warranted and needed~~We do not assume that these models represent the behavior of all modeling environments. That said, these three modeling environments allow us to explore the sensitivity of time to steady state to multiple grid types and numerical methods.

2.1 CHILD

The CHILD modelling environment was developed in the late 1990s (e.g., Tucker et al., 1999, 2001a, b). It operates on a triangular irregular network, forming a voronoi diagram, here referred to as a voronoi grid for consistency with the other grid types. ~~The CHILD experiments in this study all use~~ CHILD uses an explicit finite difference solution of the stream power process equation.

CHILD is a C++ code that is open source and available through GitHub (https://github.com/childmodel/child; accessed 08 September 2022). Although CHILD is no longer actively in development, it was widely used in the past 20 years and continues to be used at the time of writing. Because of its familiarity to the authors, its application on a voronoi grid, and its advanced stage of development, we ~~use~~ used it in this study.

2.2 Landlab

Landlab is Python library for modeling surface processes on regular and irregular grids (Hobley et al., 2017; Barnhart et al., 2020a). In this study we implement computer models that use raster, hexagonal, and voronoi grids. (Landlab also supports radial grids but these are not tested in this study.) Landlab is open source and available through GitHub (https://github.com/landlab/landlab; accessed October 14, 2022). This study used Landlab version 2.4.2.dev0. Landlab is in active development and is currently maintained through CSDMS (Tucker et al., 2022). The stream power process component used in all Landlab computer models in this study implements a version of the Fastscape implicit finite difference numerical algorithm (Braun and Willett, 2013). The Fastscape numerical algorithm was designed to be more stable than most finite difference methods (Braun and Willett, 2013)~~, the details of which are briefly discussed later~~.

2.3 TTLEM

TTLEM is part of the TopoToolbox Matlab library. Many readers may be familiar with the DEM (digital elevation model) analysis tools that are contained within TopoToolbox (e.g., Schwanghart and Scherler, 2014). TTLEM is an LEM that is distributed with these tools and uses the TopoToolbox library for many of the core functions within the LEM, e.g., flow routing (Campforts et al., 2017) (Available at https://github.com/wschwanghart/topotoolbox; accessed December 7, 2022).

TTLEM contains three numerical algorithms for solving the stream power equation on a raster grid. In this study we use TTLEM computer models that implement the Fastscape implicit numerical algorithm, an explicit finite difference algorithm, and the total variation diminishing finite volume method (TVD_FVM). TVD_FVM is designed to be highly accurate and limit numerical diffusion (Campforts and Govers, 2015).

2.4 LEMs and comparison rational

Using the three modeling environments described above, we created seven different LEMs which we ~~used for intra- and inter-model comparison~~use to calculate time to steady state. These different LEMs are described in Table 1. These three modeling environments allow us to compare how grid type and the numerical method for solving the stream power equation impact time to steady state.

LEM	Modeling Environment	Grid Type	Numerical ~~Algorithm~~ Method
CVE	CHILD	voronoi	explicit
LVI	Landlab	voronoi	fastscape (implicit)
LHI	Landlab	hexagonal	fastscape (implicit)
LRI	Landlab	raster	fastscape (implicit)
TRI	TTLEM	raster	fastscape (implicit)
TRT	TTLEM	raster	TVD_FVM
TRE	TTLEM	raster	explicit

Table 1. Table of different LEMs used in this study. We use the LEM abbreviated names in column 1 to refer to the different model scenarios. The first letter in the abbreviated name stands for the modeling environment: C for CHILD; L for Landlab; and T for TTLEM. The second letter stands for the grid type: V for voronoi; H for hexagonal; and R for raster. The third letter stands for the numerical ~~algorithm~~method used to solve the stream power equation: E for explicit; I for implicit; T for TVD_FVM.

~~These three modeling environments allowed for different types of comparisons (Fig. ??). Within the Landlab modeling environment we compared the outcomes of the same process equation and numerical algorithm implemented on different grid types (LRI, LVI, and LHI). Within the TTLEM modeling environment we compared the solution of the same process equation with different numerical algorithms (TRI, TRT, and TRE). Finally, among the three modeling environments we compared the~~

~~solution of the same process equation and numerical algorithm but on different grid types (CVE and TRE) and the same process equation and grid type, but different numerical algorithms (CVE and LVI).~~

~~Schematic of comparison types within the experiments.~~

3 Stream ~~Power Process Equation and Numerical Algorithms~~power equation

All of the model environments we ~~considered~~ consider use the stream power ~~process equation to model~~ equation to represent the evolution of fluvial profiles~~and use or include different numerical algorithms to solve this equation. Below, we first briefly review the stream power equation and then describe the different numerical algorithms employed to solve this equation within the modeling environments.~~

3.1 ~~Stream Power Process Equation~~

~~Within all three modeling environments, the~~ . We only consider fluvial erosion in our LEMs. Erosion is sustained throughout all simulations by uniform, steady rock uplift. The equation controlling the change in topographic elevation at each node is,

$$\frac{dz}{dt} = U - KA^m S^n \tag{1}$$

where z is node elevation; t is time, and dt is the ~~model~~ computational time step; U is the rock uplift rate; K is the erodibility parameter; A is the drainage area at a node; S is the topographic slope (negative of the spatial derivative in elevation, assuming directionality is in the downslope direction) at a node; and m and n are positive exponents. Note that the second set of terms on the right-hand-side of ~~Eq.~~ equation 1 is the widely used stream power ~~process equation (SPPE~~equation (SPE) that describes detachment limited fluvial incision,

$$E = KA^m S^n \tag{2}$$

where E is fluvial incision. Derivations, dynamics, and limitations of the ~~SPPE~~ SPE have been described in detail in numerous publications and we refer interested readers to such sources (e.g., Howard, 1994; Whipple and Tucker, 1999; Lague, 2014).

~~In the context of the model environments we used in our experiments, we define steady state as the condition when Eq. 1 is equal to zero, or, as said above, rock uplift rate U and fluvial incision rate E are equal at every node. Eq. 1 can be rearranged to solve for the analytical relationship between S and A at steady state,~~

$$S = \left(\frac{U}{K}\right)^{\frac{1}{n}} A^{\frac{-m}{n}}$$

~~Eq. ?? can be used to test whether each model experiment produces the analytical solution at steady state. As described below, this is one of the verification tests we explore in this study.~~

~~To consider the evolution of river profiles as predicted using SPPE, it is useful to define both the channel steepness k_s and transformed river coordinate χ. Empirically, the channel steepness was first described by Flint (1974), relating channel slope~~

~~S and drainage area A and considering the concavity index θ of a profile where,~~

$$S = k_s A^{-\theta}.$$

~~When erosion and~~

4 Experimental set-up

Each of our numerical experiments quantified the time it takes for a low rock uplift rate~~and erodibility are uniform, Eq. **??** can be recast in terms of Eqs. 2 and **??** such that,~~

$$k_s = \left(\frac{U}{K}\right)^{\frac{1}{n}}$$

~~and~~

$$\theta = \frac{m}{n}.$$

~~Thus, on a log-log plot of A and S and considering Eqs. **??**, **??**, and **??**, the~~ , steady-state ~~form of a channel should produce a straight line with a slope of $-\theta$ and a y-intercept of k_s. In analyses of real channel profiles, such slope-area data is often noisy so it is useful to consider an alternative form of the same data. Specifically, it has become common to transform river profile distances x using the χ integral transform proposed by Perron and Royden (2013),~~

$$\chi = \int_{x_b}^{x} \left(\frac{A_0}{A(x')}\right)^{\theta} dx'$$

~~where x_b is the x coordinate of base level of the profile, x is the x coordinate of the channel head, and A_0 is a reference drainage area. At steady state and if using an appropriate value for θ, a plot of z (y axis) versus χ (x axis) for a channel profile will be a straight line. If A_0 is set to 1, the slope of that line will be equivalent to k_s. In the context of the SPPE, θ is often replaced with $\frac{m}{n}$ in Eq. **??**, which holds when all parameters and variables are spatially uniform. χ-z relationships can thus be used in a similar manner as slope-area data to assess the degree to which a particular simulation reproduces the analytical expectation.~~

4.1 ~~Numerical Algorithms for Solving SPPE~~

~~We considered three different numerical algorithms to solve Eq. 1, specifically an explicit finite difference method (FDM), an implicit FDM~~landscape to fully adjust to an increase in rock uplift rate using a specific LEM scenario (Table 1), initial condition, and a ~~total variation diminishing finite volume method (TVD_FVM).~~

4.0.1 ~~Explicit FDM~~

The explicit FDM is widely used and easy to implement. The explicit FDM discretizes time t and space x, z such that,

$$x_i = x_0 + i * dx, i = 0, 1, ..., I$$

$$t_k = t_0 + k * dt, k = 0, 1, ..., K$$

Using a common upwind scheme, the solution to the differential Eq. 2 is computed as,

$$\frac{z(x_i)^{k+1} - z(x_i)^k}{dt} = KA(x_i)^m \left(\frac{z(x_{i+1})^k - z(x_i)^k}{dx} \right)^n$$

As described in more detail later, when using an explicit FDM, considerations must given to choices of both dx and dt to maintain numerical stability of the solution.

4.0.1 Implicit FDM

The common form of the implicit FDM for solving Eq. 2 was developed by Braun and Willett (2013), which is sometimes referred to as the *Fastscape* algorithm. This implicit FDM uses a similar discretization of both time and space as the explicit FDM, i.e., Eq. **??**. The primary difference between the implicit and explicit FDM is that the implicit solution is forward in both time and space. Specifically, the solution to Eq. 2 becomes,

$$\frac{z(x_i)^{k+1} - z(x_i)^k}{dt} = KA(x_i)^m \left(\frac{z(x_{i+1})^{k+1} - z(x_i)^{k+1}}{dx} \right)^n$$

Efficient solution of the implicit FDM relies on having a fixed (or known) base level for all modeled profiles at a given time step and a hierarchical list of *givers* and *receivers*, i.e., nodes that flow into each other, to overcome the challenge of having two unknowns, specifically, $z(x_{i+1})^{k+1}$ and $z(x_i)^{k+1}$. In detail, implementing the implicit FDM is different depending on the value of n. If $n = 1$, then Eq. **??** can be solved explicitly, but if $n \neq 1$, Eq. **??** must be solved iteratively, but the solution converges quickly. In detail, the two model environments that we test and which use the implicit FDM, i.e., TTLEM and Landlab, find the roots of this equation differently with the former using the Newton-Raphson method and the latter using Brent's method. As range of computational timesteps. We did not formally quantify that the initial low uplift landscapes were at steady state. Instead we generated initial conditions with simulations that ran for 100 million years. As will be apparent from our results, this is more than enough time for our experiments are limited to $n = 1$, these differences are not directly tested here. While more complicated than the explicit FDM, the potential benefit of the implicit FDM is that it is unconditionally numerically stable and thus in theory allows for longer time steps and shorter run times for equivalent model durations. We consider this point in the discussion of our results. For details of implementation of the implicit FDM and associated requirements, we refer readers to Braun and Willett (2013).

4.0.1 TVD_FVM

The TVD_FVM solution to Eq. 2 was developed by Campforts and Govers (2015) with the specific goal of minimizing numerical diffusion around knickpoints within modelled fluvial profiles. There are two different solutions depending on whether $n = 1$

or $n \neq 1$, we focused on the former since our experiments were all run with $n = 1$. The TVD_FVM solution for $n = 1$ recasts Eq. 2 as a scalar conservation law,

$$\frac{dz}{dt} + \frac{df(z)}{dx} = 0$$

$$f(z) = -KA(x_i)z$$

where the second equation represents a flux function for the conserved variable z. In this solution, Eq. **??** is assumed to evolve over a constant volume in both time and space where the spatial domain is discretized into finite volumes centered at x_i and with a cell size of dx,

$$dx = \left[x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}\right]$$

Eq. **??** is integrated and rearranged to derive,

$$z(x_i)^{k+1} = z(x_i)^k + \frac{dt}{dx}\left[f_{i-\frac{1}{2}} - f_{i+\frac{1}{2}}\right]$$

where $z(x_i)^{k+1}$ is the elevation at a discretized gridpoint representing the average value of z at the next time step and at grid position i and the quantity in brackets represents the numerical approximation of the flux from Eq. **??**. Other details of how the TVD_FVM algorithm works and additional steps to the solution are described in the supplementary materials of Campforts and Govers (2015), to which we refer interested readers. Of importance, the TVD_FVM is subject to similar stability criteria as the explicit FDM.

4.1 Other differences between model environments

While we focus on the differences in grid type and the numerical method to solve the SPPE among the model environments we compare, this is not meant to indicate that all other aspects of these models are directly comparable. Some of the other important differences between models stem from variability in the methods for (1) calculating drainage area, (2) flow routing, and (3) identifying and filling "pits" within fluvial profiles, among other details. All of these differences have the potential to change the resulting river network geometry and landscape structure, even when beginning from the same initial condition, with the same grid type, and the same numerical solution to the SPPE. In this effort, we do not consider these aspects in detail, but highlight that future inter-model comparisons should think more broadly about differences among modeling environments.

5 Experimental set-up

We ran two experiments, each of which required multiple simulations with each participating model. The first experiment used all the model environments to evolve landscapes to steady state and compared steady-state morphology among the LEMs.

~~The landscapes have the same uniform rock uplift rate, and we broadly consider that steady state is achieved when the fluvial incision rate is equal to the rock uplift rate everywhere on the landscape . For the steady-state experiments, we did not formally quantify steady state, but instead ran the models for an arbitrary long time (i.e.,~~ landscapes to reach steady state regardless of chosen metric or threshold for identifying steady state. In some cases we stopped the initial runs before 100 ~~Myr). The second experiment perturbed the previously-generated steady-state landscape by uniformly increasing the rock~~ million years because the landscape became perfectly static. This experimental design, in which an initially steady landscape is perturbed by changing the uplift rate, is similar to previous modeling studies (e.g., Rosenbloom and Anderson, 1994; Whipple and Tucker, 1999, 2002; Gasparini et al., 2007; Attal et al., 2011)~~. We compare the landscape morphology after 200, 000 years of evolution. Further, we more formally quantified steady state and compared the time to a new steady state after perturbation~~, which is why we chose it.

As described below, we kept as much constant among the simulations with different LEMs as possible. However, we did not do anything to change the models from "off the shelf". In other words, we used the modeling environments without changing any of the internal code that implements the numerical algorithms. Where appropriate, we did change ~~free~~ parameters within the different modelling environments to try to assure that their behavior was as comparable as possible. ~~There are some hard-wired differences among the codes that are not user-configurable and are not related to grid types and numerical algorithms. We found these differences led to some differences in the solutions, and this will be discussed later~~For example, all the raster models use D8 flow routing (Tarboton, 1997), and this was an option that we chose when using the LRI model to make it more similar to the TRI model. In contrast, there are multiple ways to calculate a topographic gradient at a grid cell in a computational model. Choosing the algorithm used to make this calculation is not exposed to the user for these models, and accordingly we can not ensure that each model calculates topographic gradient in the same way.

4.1 ~~Steady-state experiment~~

~~These simulations~~ The simulations that created the initial steady state landscape were started from a surface with elevation values randomly chosen between 0.0 and 1.0 ~~m~~meters. The random elevation initial surface was used because it creates more realistic looking drainage networks. The raster grids had ~~a grid of~~ 200 by 200 nodes, and the spacing between nodes in the x and y direction was 100 ~~m~~meters. All the simulations using a raster grid in different models used the same exact initial topographic surface (LRI, TRI, TRT, and TRE). The ~~numerical experiments~~ simulations that use a voronoi grid ~~have~~ had an average node spacing of ~~≈~~ 100 m, but the spacing varied (in a regular way) to define the grid. The initial condition was a similar noisy surface as used with the raster grids, and the same exact initial surface ~~is~~ was used in the CVE and LVI simulations. Regardless of the type of grid, all the grids were ≈ 20 km by ≈ 20 km with ≈ 100 m resolution.

The boundary conditions used in all simulations were the same. All of the nodes on the perimeter of grid were open boundaries where water can exit, but not enter, the grid. The elevation of the perimeter boundary nodes was fixed at zero meters and did not change during the ~~experiments~~simulations. In other words, the perimeter nodes were not uplifted but the rest of the grid was uplifted. U and K were spatially uniform and set at 1e-4 m/yr and 5e-6 yr^{-1}, respectively, in all the ~~steady-state~~

initial simulations. m and n were also spatially uniform and set at 0.5 and 1, respectively~~, in all the steady-state simulations, corresponding to a θ value of 0.5 given the uniformity of all variables and parameters~~.

We explored how the time step value, dt in ~~Eqs. 1, ??, ??, and ??, impacts whether the steady-state landscapes matches the analytical solution~~equation 1, impacts the time to steady state. Each model ~~experiment~~ should be stable, and produce the correct analytical solution, when the time step satisfies the ~~CourantâFriedrichsâ~~Courant?Friedrichs?Lewy condition:

$$C_{max} > \frac{vdt}{dx} \tag{3}$$

where C_{max} is the Courant number, ≈ 1.0 in stable conditions; v is the speed that an erosional wave will move through the network, and dx is the spacing between nodes. When using the stream power process equation with $n = 1$, v is approximated as

$$v \approx KA^m. \tag{4}$$

To calculate the maximum time step for a stable Courant condition (dt, ~~Eq.~~ equation 3), we needed to know the largest drainage area in the modeled network to estimate the fastest wave speed (~~Eq.~~ equation 4). Because we started with a noisy surface and drainages evolved out of the noise, the area of the largest watershed was unknown until the landscape reached steady state. However, we had to choose a stable dt before the landscape evolved. Therefore we estimated the size of the largest drainage. We assumed the maximum drainage area will be one-fourth of the total area of the grid, or 49 km^2. This value was intended as an overestimation, ensuring that we calculated a stable time step with Eqs. 3 and 4.

~~Based on the parameter values that we used in Eq. 1 and the effective version of this equation within the different numerical algorithms, we estimated that~~ We combined equations 4 and 3 to calculate a value of dt. For the values chosen for K and m, $C_{max} = 1$, and our estimated value for the maximum drainage area, we calculated a stable dt of 2857 years. We then chose dt =2,500 yrs ~~should satisfy Eq. 3 and~~ as the base-case model time step because it should result in stable ~~model experiments~~simulations. We explored how ~~the simulated landscape~~ landscape response time changed by changing dt. We ran four simulations with each model, reducing dt to 250 yrs and increasing dt to 25,000 and 100,000 yrs, or approximately 10 and 40 times greater than the stable condition, respectively.

4.1 ~~Transient experiment~~

~~In these simulations we modeled how a steady-state landscape evolves in response to a uniform increase in rock uplift. Here we illustrate the behavior of largest channel profile to represent the transient behavior. We explored the variation in response with time step dt size. All LEMs that operate on raster grids used the same initial topography. We used the topography produced using the Landlab modeling environment with dt=2,500 yrs from the steady-state model simulations. In the model simulations using a voronoi grid, we used the topography produced from the steady-state CHILD model simulation~~ The decision to use time steps longer than the estimated stable time step is motivated by the results shown in (Braun and Willett, 2013). They state that their numerical algorithm, which is the implicit numerical method used in TRI and all Landlab models, is accurate even when the time step is more than 100 times the stable condition. The stability of the algorithm is also discussed

in (Braun and Deal, 2023). However, simulations from models employing either the explicit finite difference or TVD_FVM algorithms did not produce accurate, or even sensible results, with time steps longer than the predicated stable value. Hence results with all times steps are only shown from models using the implicit solution. In one case, TRI, we also performed two extra simulations with dt = ~~2,500 yrs. Finally, for the Landlab LEMs using a hexagonal grid, we use the Landlab steady-state topography produced using the hexagonal grid and dt=2~~1,000 and 10,~~500 yrs.~~ 000 yrs. This was done based on preliminary results to fully explore sensitivity of time to steady state to dt.

~~Our transient model simulationsincreased~~ We use the steady initial conditions produced from each model simulation and time step combination as the initial condition for the corresponding transient simulations. We increase rock uplift rate uniformly across the grid by a factor of five, to U=5e-4 m/yr. ~~We did not change anything besides the rock uplift rate between our~~ As the landscape evolves in response to the new, higher uplift rate, we track the steady-state ~~and transient simulations. We used the same four time step values used in the steady-state simulations and ran four simulations for each participating LEM. We compared the profile of the river channel with the largest drainage area after 200, 000 years. We also compared the time to reach a new steady state following the perturbation. We explored how the~~ metrics to quantify differences in time to steady state~~differs depending on the metric used for calculating steady state and the time step dt~~.

5 ~~Comparison of steady-state results~~

4.1 ~~Raster LEM simulations~~

~~Here we illustrate the results of four LEMS (LRI, TRI, TRT, TRE) using raster grids in the Landlab and TTLEM modeling environments. All the simulations start with the same exact initial condition and used a 250 yr time step, the smallest time step used in this study. Therefore we expect none of the variation in observed model behavior to be attributable to numerical instability. The same parameters and boundary conditions are used. The three TTLEM LEMs use the explicit FDM, implicit FDM, and the TVD_FVM to solve the SPPE, and the Landlab LEM uses the implicit FDM.~~

~~Steady-state topography of the four computer models that use a raster grid.~~

~~Because all these landscapes evolved from the same initial random topography and use the same time step, one might expect the details of the network structure to be identical~~In all of our initial simulations we re-calculated flow directionality at every time step. However, ~~there are differences in the network organization among all the steady-state landscapes produced with these four raster LEMs, as illustrated by the path of the largest channel across the landscape (Fig. **??**). In all four cases the largest channels drain to the west of the range, but there are subtle differences in the main channel path and larger differences in the topology of the main drainage divides. These differences are highlighted in elevation difference plots, using TTLEM implicit as the base comparison case (Fig. **??**. Interestingly the TTLEM implicit and explicit models produce the most similar landscapes, when measured as the sum of absolute total difference in elevation values.~~

~~Steady-state topography produced using the TRE model (a) and the difference between (a) and the TRI steady-state topography (b), the TRT steady-state topography (c), and the LRI steady-state topography (d). A time step of 250 years was used to produce all the topographies used for this figure.~~

Even though the TTLEM and Landlab implicit LEMs (TRI and LRI) models use the same numerical algorithm, the topographic organization differs. This could be due the details of the algorithm implementation in each modeling environment. This could also be due to details in the flow routing algorithm and how sinks (local lows) are managed. The latter is generally more likely given the greater degree of similarity in gross drainage network and divides structure between the different TTLEM runs, which all share the same flow routing algorithm and methodology for dealing with sinks (Fig. **??**).

We compared the modeled slope and area data with the predicted analytical expression (Eq. **??**) in Fig. **??**. Despite differences in network organization, simulations from three of the four LEMs after running the simulations we observed that in some cases the drainage network rearranged, and we wanted to explore how this affects time to steady state. Thus, we performed two more numerical experiments using TTLEM with the three different numerical methods (TRI, TRE, LRI) created landscapes that exactly reproduced the slope-area relationship. The TRTmodel steady-state slope-area output have some scatter around the predicted analytical solution. This is not enough scatter that it is noticeable in the channel profile or chi-elevation plots (not shown here).

Steady-state slope-area data from the four computer models that use a raster grid. A time step of 250 years was used.

We suggest two takeaways from these simulations. First, be cautious when interpreting the details of planform morphology and flow organization of either rivers or divides produced from LEMs. Even in deterministic models with uniform parameters and forcings, the details of the network organization can vary solely based on the numerical algorithm and the details of how the algorithm is coded. Second, all the LEMs reproduced the predicted analytical relationship between slope and drainage area. However, some are more accurate than others. When interpreting model results, scatter that is a result of a numerical algorithm should not be misinterpreted as indicative of geomorphic processes. We suggest that the first step of any model verification study should be a test of whether the predicted analytical solution(s) is reproduced. Further, the community should agree on the accepted degree of model accuracy in reproducing the analytical solution.

4.1 Variable time step simulations

Here we illustrate the impact of changing the time step dt between multiple simulations using the same LEM. We used the CHILD model (CVE) because the impact of increasing the time step was very clear in the output. CHILD uses an explicit FDM solution, i.e., Eq. **??**. As such, it does not remain stable at time steps larger than the Courant condition (eq. 3), here estimated to be between 2TRT). First, 500 and 25,000 yrs.

The two landscapes produced using time steps less than the Courant condition illustrate stable behavior (Fig. **??**). The analytical slope-area relationship is exactly reproduced. There are some differences between the planform morphology in the steady-state landscapes produced with the two smallest time steps, despite starting from the same exact initial condition and using the same boundary conditions . However these differences are barely detectable from the topographic maps.

Topography of the entire model domain (a, b, c, d) and slope-area plot (e) of the largest channel from the CHILD steady-state simulations using different time steps.

The two landscapes produced using time steps greater than the Courant condition illustrate various symptoms of numerical instability. In the case of a dt we re-ran all of 25, 000 yr, the instability is not obvious from the elevation map alone.

However, the scatter in the slope-area relationship illustrates that the analytical relationship is not exactly produced in the largest drainage-area reaches of the main channel (Fig. **??**c). The steady-state solution is produced at smaller drainage-area reaches. This makes sense given Eq. 3, which predicts that the stable time step increases with drainage area. An even larger dt of 100the TTLEM empirical models but did not reroute flow at every time step during the transient simulations. In other words, we forced the network to remain static during the adjustment to a higher uplift rate. Second, we generated new steady-state initial conditions and used them for another set of uplift increase experiments. For these simulations, we started with a different initial white noise grid to create the low uplift steady state topography. Because the initial white noise determines the network details, 000 yr has obvious instability, even in the elevation map. Scatter in the slope-area relationship is evident at smaller drainage area reaches than in the 2these landscapes have a different network despite being run with the same model, 500 yr experiment.

The TTLEM raster explicit LEM (TRE) has the same sensitivity (results not shown). That is, for time steps of 25,000 parameters, and 100,000 years, there was mismatch between the TRE LEM slope-area relationship and the analytical solution. This mismatch occurs in smaller drainage area reaches with a larger time step. This behavior is indicative of the numerical algorithm, and not of the grid type, as supported by the comparison between TRE and CVE. We return to the consideration of time step in the transient experiments. boundary conditions.

The Landlab voronoi implicit (LVI) results do not show the same instability with time step as the CHILD model (CVE) does (Fig. **??**). This illustrates that the instability is not an artifact of a voronoi grid. Fig. **??** also illustrates the stability of the fastscape implicit algorithm. The steady-state analytical solution is produced even when the time step is greater than that predicted from the Courant condition. Finally, the LVI model results also illustrate the change in the detailsof flow organization with time step.

Topography of the entire model domain (a, b, c, d) and slope-area plot (e) of the largest channel from the Landlab voronoi (LVI) steady-state simulations using different time steps.

We have three take-aways from the variable time step steady-state numerical experiments. First, the details of flow organization can vary with model time step, even when the same LEM is used with the same flow routing and pit-filling algorithms. Second, modelinstability is not always obvious from a topographic map. If a LEM does not exactly reproduce an analytical solution, it may be an indicator of numerical instability. Third, the stability of the LEM depends on the time step and numerical algorithm. We did not find the stability of a LEM to be a function of the grid type.

5 Comparison of transient resultsTime to steady state

5.1 Morphology of transient landscapes

Here we illustrate the transient behavior of a landscape as it adjusts to an increase in rock uplift rate (5e-4 m/yr) from steady statewith a low rock uplift rate (1e-4 m/yr). We used three TTLEM LEMs (TRI, TRT, TRE) because they allowed us to illustrate the behavior of three different numerical algorithms and their sensitivity to time step during a perturbation while keeping other details (e.g., grid type, flow routing, etc.) the same.

Elevation, χ profile, and slope area data of main channel after 200,000 model years of evolution. All simulations used the TTLEM modeling environment.

The transient morphology of the largest channel after 200,000 years of evolution using the TRI, TRT, and TRE LEMs is shown in Figure **??**. The solution differed among the algorithms. None of the algorithms produced the exact solution above a time step of 2,500 yrs. We predicted that 250 and 2,500 yrs should both be stable time steps, and there are only slight differences in the solutions produced with these two time steps. The model inaccuracies produced with all of the numerical algorithms with a time step of 25,000 or 100,000 yrs are easily apparent in comparison with the solutions using smaller time steps. Below we discuss the details of the results produced with the different numerical solutions.

The implicit numerical solution (left column of figure **??**) smoothed the shape of the knickpoint when the time step did not satisfy the Courant condition (Eq. 3). In the simulations with a time step of 250 or 2,500 yrs, there was a sharp transition from the upstream, more slowly eroding part of the channel above 100 m elevation and the downstream part of the channel that has already responded to increase in rock uplift. The χ-z plots (middle row of figure **??**) show the divergence from the expected Here we describe the metrics we use to empirically quantify time to steady state. We also describe the analytical equation previously presented by Whipple and Tucker (1999) and Whipple (2001) that we use to predict time to steady state. In the results section we will compare the empirically derived steady-state solutions (dashed and dotted lines) in the two simulations with the largest time steps. In the channel profile (top row of figure **??**) this appears as a more rounded knickzone, as the transition from one erosion rate to another is spread above and below 100 m. The spreading of the knickzone with time step is visible in the slope-area plots (bottom row of figure **??**) as the experiment with the longest time step illustrates that points with smaller drainage area have started to adjust to the change in rock uplift. This smoothing of the knickpoint does not impact the total relief in the channel. times with the analytically predicted values.

In contrast to the implicit solution, when the time step is greater than 2,500 years both the TVD_FVM (middle column of figure **??**) and explicit FDM (right column of figure **??**) numerical algorithms generated behavior commonly associated with numerical instability. The channel profiles for both numerical algorithms had a stair-step appearance below the elevation of the knickpoint (100 m) with the 25,000 and 100,000 yr time steps. When the time step was 100,000 yrs, there were only two time steps modeled in these examples. Given the

5.1 Numerically modeled

Theoretically steady state is reached when equation 1 is equal to zero, or when rock uplift rate , each time step uplifts the landscape by 50 meters. The impact of these two large uplift *events* are clearly visible in the channel profiles and elevation-chi plots with both the TVD_FVM and explicit numerical algorithms. If you look very carefully, the impact of the 12.5 meter uplift *events* with the 25U and fluvial incision rate E are equal at every node. Although steady state is often used and referred to in numerical modeling studies, the criteria for reaching steady state, 000 yr time step is also visible in the lower portions of the profiles. Both algorithms smoothed the smaller profile steps (made with a dt of 25or determining when $E = U$ everywhere on the grid, 000 yrs) as they move up through the landscape.

~~The CHILD LEM (CVE) uses an explicit algorithm, and the results from the same numerical experiments with CVE match those produced with TRE (results not shown, only performed with time step $\leq$ 25,000 years). We note that CHILD and TTLEM have internal stability criteria which are meant to reduce numerical instabilities like those illustrated in figure ??. Usually these internal stability corrections in a model reduce the time step over which a process is calculated. However, these stability corrections only apply in the fluvial process calculations in both models and do not impact the time step over which rock uplift is applied. Thus, even with stability correction in both CHILD and TTLEM, a long time step creates a large cliff on the boundary of the model when rock uplift is added to the landscape.~~

5.2 ~~Time to steady state~~

~~Here we compare the time to reach a new steady statefollowing the perturbation used above. We illustrate the results using the three Landlab LEMs (LVI, LHI, LRI) and the three TTLEM LEMs (TRI, TRT, TRE) using different time steps. Thus we are able to document whether gridtype, numerical algorithm, and/or time step impact time to steady state.~~

~~We illustrate the time series of four different metrics whose value or temporal behavior could be indicative of steady state . The four metrics are the temporal change in maximum elevation, mean elevation, maximum location elevation, and sediment flux eroded from the entire landscape.~~ is not always explicitly described. Here we test four metrics for determining when steady state is reached. We refer to these collectively as the steady-state metrics. ~~In all cases the steady-state metrics are calculated over temporal difference 100,000 yrs, which was set by the longest time step of the simulations being compared. Therefore the first time that any metric was calculated is at 100,000 yrs into a transient simulation. This was done to ensure equivalent comparisons between the simulations as some degree of differences in time to steady state would be expected if we calculated these metrics over different time intervals between simulations.~~

The temporal change in maximum elevation, Δz^t_{max} was calculated as,

$$\Delta z^t_{max} = |max\left(z^t_i\right) - max\left(z_i^{t-100,000}\right)|, \tag{5}$$

where $max\left(z^t_i\right)$ is the maximum grid elevation within the domain at time t; and $max\left(z_i^{t-100,000}\right)$ is the maximum grid elevation within the domain at time $t - 100,000$.

The temporal change in mean elevation, Δz^t_{mean} was calculated as,

$$\Delta z^t_{mean} = |mean\left(z^t_i\right) - mean\left(z_i^{t-100,000}\right)|, \tag{6}$$

where $mean\left(z^t_i\right)$ is the mean grid elevation within the domain at time t; and $mean\left(z_i^{t-100,000}\right)$ is the mean grid elevation within the domain at time $t - 100,000$.

The temporal change in maximum local elevation, $\Delta z^t_{max(loc)}$ was calculated as,

$$\Delta z^t_{max(loc)} = max\left(|z^t_i - z_i^{t-100,000}|\right). \tag{7}$$

The difference between equations ~~7 and 5~~ 5 and 7 is that the former finds the difference between the maximum elevation of the entire domain at two different times, whereas the latter finds the difference in elevation at every node in the landscape between

two different time steps and uses the maximum of those differences. The units on all of the elevation change metrics are meters.

Finally, the temporal change in sediment flux, ΔQs^t, was calculated as,

$$\Delta Qs^t = \sum Qs_i^t - \sum Qs_i^{t-100{,}000}, \tag{8}$$

where $\sum Qs_i^t$ is the summation of the erosion rate at every node on the landscape at time t; and $\sum Qs_i^{t-100{,}000}$ is the summation of the erosion rate at every node on the landscape at time $t - 100{,}000$. The models do not track the sediment flux explicitly when using the ~~SPPE~~SPE, which is why we use the local erosion rate as a proxy for sediment flux. In all of the grids the cell size is nearly uniform, hence the summation of erosion rate is a good proxy for the sediment flux. The units of ΔQs are meters per year.

In all cases the steady-state metrics are calculated over a temporal difference 100,000 yrs, which was set by the longest time step of the simulations being compared. This was done to ensure equivalent comparisons between the simulations as some degree of differences in time to steady state would be expected if we calculated these metrics over different time intervals between simulations. The first time that the topographic metrics were calculated is at 100,000 yrs into the simulation. With the sediment flux metric, the first calculation is done at 200,000 yrs, because there is no sediment flux at time zero in the simulation.

As a landscape evolves towards steady state, all of the steady-state metrics should approach zero. When exactly a new steady state is reached could be defined as when the metric value passes below a predetermined threshold. The strictest definition would be when a metric reaches zero. We note that differences in the accuracy of calculations in different modeling environments could lead to some of the metrics never reaching zero. Also, the time at which a particular metric appears to reach zero will depend on the floating point precision of the programming environment in which the metric is calculated. One could also use the rate of change in the value of a steady-state metric as a criteria for reaching steady state. ~~As the intent of this paper was not to determine the best metrics or methods for LEM comparison, we do not settle the discussion. Rather we present the results for discussion among the community.~~

~~Time series of four different steady-state metrics that can be used for evaluating when a landscape has reached steady state. All metrics are calculated following an increase in rock uplift.~~ Here we define the time to steady state to be 10^{-5} meters for the elevation metrics or 10^{-6} meters per year for the flux metric. This may seem conservative. For example, this means the maximum elevation in a landscape changes by less than 10^{-5} meters over 100,000 years. We chose that threshold after seeing the results, as we did not know exactly what to expect in advance. However, the choice of the threshold value makes sense in light of the behavior of these metrics, as will be illustrated below.

5.2 Analytical equation

Following Whipple and Tucker (1999), Whipple (2001) showed that the predicted time to steady state following an increase in rock uplift in a landscape evolving according to equation 1 and with $n = 1$ is

$$T_A = \frac{\beta}{K} \tag{9}$$

where

$$\beta = k_a^{-\frac{m}{n}} \left(1 - \frac{hm}{n}\right)^{-1} \left(L^{1-\frac{hm}{n}} - x_c^{1-\frac{hm}{n}}\right). \tag{10}$$

In equation 9, T_A is the analytical time to steady state, K is the erodibility in equation 2. In equation 10, x_c is hillslope length, and m and n are the exponents in equation 2. Equation 10 also requires empirical parameters k_a and h from the equation developed by Hack (1957):

$$A = k_a x^h \tag{11}$$

where x is distance from the divide and A is drainage area. Note that equation 10 only holds when $\frac{hm}{n} \neq 1$ which is the case in all of our experiments.

We use data from the largest watershed in the modeled landscapes to calculate k_a and h. In simulations in which there is drainage rearrangement, and hence slight changes in these parameters while the simulation approaches a steady state, we use values at the end of the simulation time, which was 50 million years for the majority of our simulations. As none of our simulations formally included hillslopes, i.e., we do not consider diffusion and the diffusion constant was set to 0 in our simulations, we set x_c in equation 10 to 0.

6 Time to steady state results

Time series of all our metrics in all of our numerical experiments are illustrated in figure 1. Each column of ~~Figure~~ figure 1 shows the time series of one of the four steady-state metrics. The ~~top row of Figure 1 illustrates~~ different lines in each plot are the results using different computational time steps.

The top two rows of figure 1 illustrate the results for TRI and LRI LEMs, which have the same grid type and ~~numerical algorithm. Note that if one were to choose a threshold value for any of the elevation steady-state metrics , say 10^{-5} m~~use the same numerical method to solve equation 1. Using our threshold values (10^{-5} m for topographic metrics and 10^{-6} m/yr for the flux metric) to reach steady state, the estimated time to steady state ~~varied by as much as 10 my~~ varies by tens of millions of years for the same model environment and steady state metric, depending on the time step. ~~When comparing between~~ There is no consistent trend between time to steady state and time step. That is, time to steady state does not always increase or decrease with increasing time step. When comparing the steady-state elevation metrics among the TRI simulations, the time to steady state can increase by over 30 million years between the fastest and slowest modeled times to steady state. When comparing between the two model environments ~~and the different time step values, and using the same threshold, the~~ but with the same

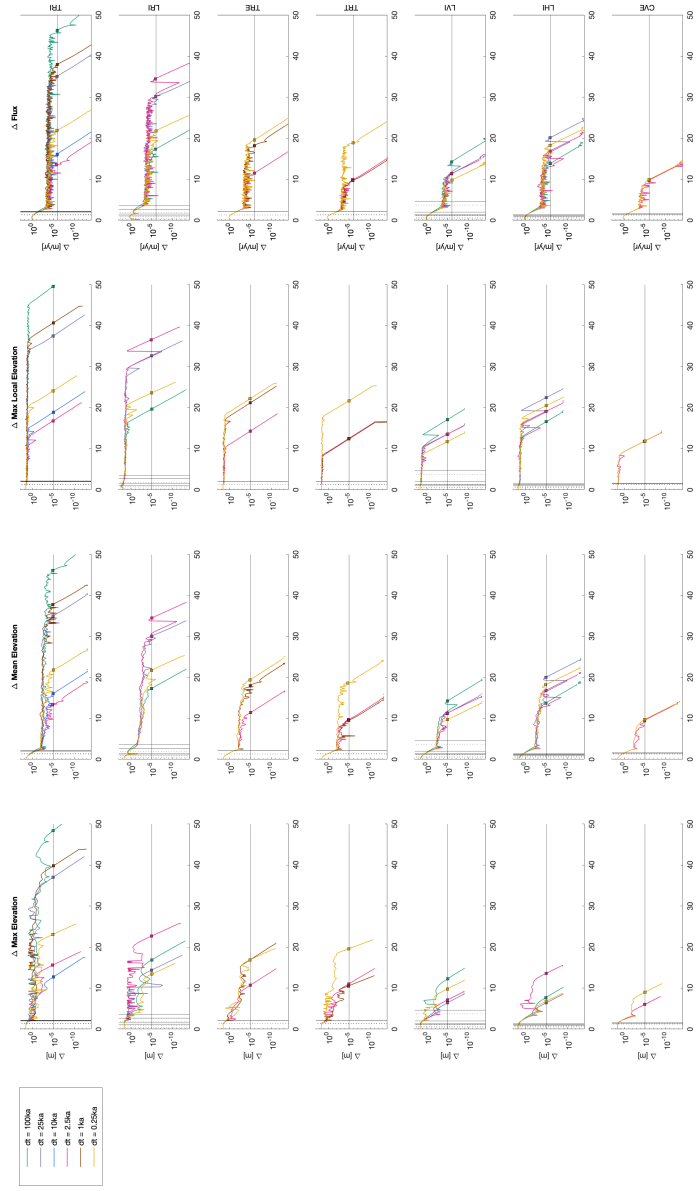

Figure 1. Time series of four different steady-state metrics, in columns, that can be used for evaluating when a landscape has reached steady state. Each row has the results from a different LEM (Table 1). The vertical lines are the analytical response times using the mean stream length (dashed) or longest stream length (solid). The x axis is time in millions of years in all the plots. Note that the extent of the x and y axes is the same in all the subplots.

time step and steady state metric, the predicted time to steady state ~~varied by almost~~ can vary by more than 20 ~~my. A similar conclusion would also be reached if one chose a threshold value less than 10^{-5} m/yr with the sediment flux steady-state metric.~~ million years.

In contrast, if one were to use the time at which a metric ~~asymptotes to a steady~~ levels off and oscillates around a value, the estimated times to steady state would be closer for a given metric. This is regardless of the time step or the LEM in the simulations with a raster grid. However, the value that a given metric ~~asymptotes to~~ levels off at differs with time step in some cases (e.g. ~~the sediment flux steady-state metric in Landlab simulations with different time steps).~~

~~The approachto steady state as gauged by the time series of the four different steady-state metrics in the transient Landlab LEMs using hexagonal and Voronoi grids (LHI, LVI) was similar to that of the TRI and LRI LEMS. In the hexagonal grid LEMs, when comparing among the time steps, some of the metrics could lead to ≈ 5 my estimated differences in time to steady state if using a threshold value. The time to steady state when calculated using a threshold value also varied among the Voronoi simulations with different time steps. However, there was less variation with the Voronoi simulations than there was with the hexagonal simulations~~change in maximum elevation for the LRI model). Also, it is not a smooth asymptotic approach, as one might expect from these simulations. Notably, variation in the change in maximum local elevation stays fairly steady at a value between 10 and 100 m for the entire simulation, until the landscape stabilizes and the value continually decreases. In other words, with the change in maximum local elevation, there is no initial decrease in the metric and then a stabilization, as occurs with the change in mean elevation and flux metrics.

The steady-state metrics from the ~~TTLEM~~ TRT and TRE ~~experiments~~ simulations (third and fourth rows of figure 1) showed similar behavior as discussed above ~~, but there are some caveats. First, these numerical algorithms~~ for the TRI and LRI simulations. The numerical methods used with TRT and TRE are not stable with larger time steps. Thus, the impact of time step on the steady-state metric values was not as large, given we only illustrate ~~the time stepsused satisfied Eq. 3. We are unsure why there were temporal oscillation in three of the steady-state metrics using the TRT model. Possibly this is related to results using three stable time steps. However, the shortest time step does result in~~ the ~~fact that the TRT model never converges to the exact analytical solution (Fig. **??**).~~ longest time to steady state for both TRT and TRE, regardless of the metric. In all cases TRT and TRE predict shorter times to steady state for a given time step and metric than the TRI models predict.

The ~~biggest take-away from this section is that time-to-steady-state metrics are likely not comparable among modeling studies. Here we used the same network for all of the raster transient simulations, so we eliminated network variability as a cause for differences in time to steady state~~ Landlab models using the hexagonal and voronoi grids (LVI and LHI, fifth and sixth rows of figure 1) all reach steady state at shorter times than the LRI simulations, for a given time step and steady state metric. Further, the LVI and LHI simulations have less of a range in predicted steady state time as a function of dt when comparing with the LRI simulations. ~~We showed that the metric and value for determining steady state matter, along with the time step, grid type, numerical model, and numerical algorithm.~~

7 ~~Summary, Implications, and Conclusions~~

6.1 ~~Stability, accuracy, verification, and the domain of applicability~~

~~A stable model avoids some of the tell-tale saw-tooth behavior of an unstable model (e.g. Fig. ??), but a stable model is not always accurate. We can see this in~~ These results are summarized in Figure 2 which compares the ratio of the modeled time to steady state using the metric threshold value T_E with the analytically predicted time to steady state (9), T_A. In theory, this should normalize the response times by expected differences related to either channel length or differences in length to drainage area scaling. One constant among all the simulations is that the ~~transient behavior of the models that use the implicit solution, i.e. Fastscape numerical algorithm with long time steps. The shape of the knickpoint is rounded when compared to the LEM thatuses the TVD_FVM numerical algorithm, which is designed to maintain the shape of a knickpoint. However, the same algorithm that is able to maintain the shape of the knickpoint has scatter in the slope-area relationship at steady state .~~ empirical time to steady state is greater than the analytically predicted value. In other words, $\frac{T_E}{T_A} > 1$ always. This holds true regardless of time step or steady-state metric. The models using a voronoi grid (LVI and CVE) consistently have the smallest values of $\frac{T_E}{T_A}$; these are the only two models that, regardless of the metric or time step, $\frac{T_E}{T_A} < 10$ always. Note that LVI and CVE use different numerical methods for solving the stream power equation.

~~Whether stability or accuracy or both are important in a modeling study depends on the scientific hypothesis being tested. What is always important is to not interpret instability or inaccuracy as being indicative of landscape processes. Model results that have clear instability have been presented in previous studies.~~ For some combinations of steady-state metrics and specific simulations, it would be possible to choose thresholds for the respective metric that would approximately equal the analytical time to steady state T_A, but not consistently. For example, ~~instability in fluvial processes near~~ the ~~outlet of a river, where drainage area is high and the boundary condition is imposed on a landscape . However, upstream of this instability, the model behaves accurately, and therefore, as long as only the stable and accurate part of the solution is used for hypothesis testing or scientific exploration, there is no problem.~~

~~We posit that the more important consideration is that the LEM community has not formally or informally defined the domain of applicability or engaged with model verification through benchmarking. Community driven definition of benchmark experiments and formal verification of models will clarify the circumstances under which a particular model is an appropriate choice for a scientific question. Furthermore, such benchmarks will demonstrate which aspects of model output are reliably interpretable from any given model (e.g., some studies may only want to interpret the relief of a landscape while others want to interpret the transient drainage reorganization). Enumeration and implementation of benchmark experiments associated with the most commonly interpreted LEM outputs would not elucidate the domain of applicability for each model, but these experiments would likely support beginner modelers in understanding which models are appropriate to use for different applications.~~ $\Delta z^t_{max(loc)}$ metric has limited change until well beyond T_A for most simulations and thus it would be challenging to choose a threshold that would result in a T_E value near T_A. For the other metrics, though, it would be generally possible to choose a threshold such that $T_E \approx T_A$. We emphasize that these threshold values would largely be different depending on the particular simulation, would not be known until after the experiments had been run, and that the continued variation in these

metrics beyond T_A imply that some amount of measurable landscape change persists beyond the predicted analytical response time.

~~Beginning LEM users must make numerous decisions. First, they must choose a modeling environment. Next they must make decisions on model initial conditions, boundary conditions, time step, and model resolution. They may also have to decide on grid type. Adding to the complexity of these decisions is the fact that different process models may apply on different spatial scales. As discussed by (Skinner and Coulthard, 2022), some decisions on model set-up are made arbitrarily based on model efficiency. By defining a domain of applicability for different LEMs, the community could remove the guesswork from these decisions.~~

6.1 ~~LEMs are not good at everything~~

Before doing the computational experiments, we had hypothesized that the empirical steady state would vary monotonically with time step, in part because this is implied by results of Braun and Willett (2013), e.g., their figure 4, exploring the dynamics of the implicit solution to the stream power equation. The lack of relationship between empirical steady state and time step is especially noticeable in the TRI simulations, which we ran with six time steps. We wanted to rule out that this lack of relationship was not because of an initial landscape and network that, just by chance, led to odd results. Thus, we reran all of the TTLEM simulations on different steady-state landscapes that were produced using the same process as the first set of simulations, just with a different initial random surface. In these ALT-TTLEM simulations, there is again no relationship between time to steady state and time step (second row of figure 2). Further, these results do not match the patterns in the first set of TTLEM simulations. For example, comparing the TRT simulations, in our first set of simulations, the smallest time step produced the longest times to steady state, regardless of metric. Contrasting that with the ALT-TRT simulations, the smallest time step produced the shortest times to steady state, regardless of the metric.

~~Our results highlight the danger in over-interpreting time~~ When comparing the evolving topographies, we observed that the CVE simulations have no drainage rearrangement, despite rerouting flow at every time step. The LVI simulations have some drainage rearrangement, but very little and only in the headwaters. In contrast, in some of the TTLEM simulations, we observed that the network across the entire landscape changed, though generally in subtle ways. Thus, we hypothesized that drainage rearrangement was playing a part in the variation in times to steady state~~from a LEM. When a model output is time step dependent or multiple models do not converge on the same answer, even when run with time steps below the Courant condition, it suggests that interpretation of that output may not be possible or must be done with caution. We illustrated that~~.

To test this hypothesis, we reran the initial set of TTLEM simulations, but this time we did not reroute the flow after every time step during the transient simulations, which is an available option within the TTLEM modeling environment. What we found is that there is much less variation in time to steady state ~~is not just time step dependent, but also algorithm, grid, and model environment dependent, as well as metric dependent.~~ with time step when comparing with the initial set of simulations (figure 3). The time series of all the steady-state metrics in the simulations without drainage rearrangement is very different from the TTLEM simulations with drainage rearrangement. With a fixed network, all of the metrics remain nearly steady in the initial response, and then they monotonically decline.

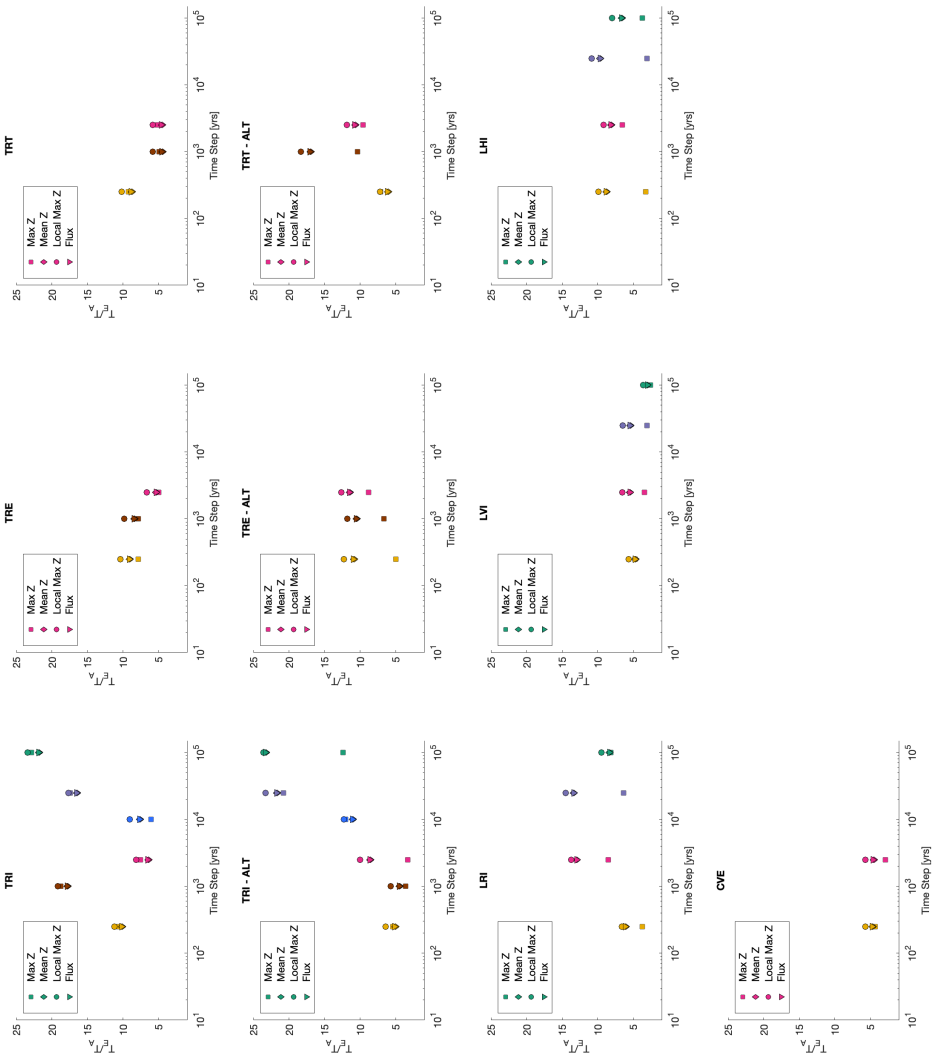

Figure 2. Ratio of the modeled to analytically predicted time to steady state as a function of time step for all of the modeling scenarios. The second row of results indicates the set of TTLEM simulations that used a different network for the initial conditions. Other than this change, everything was the same between TTLEM and TTLEM-ALT. Note that the extent of the x and y axes is the same in all of the subplots.

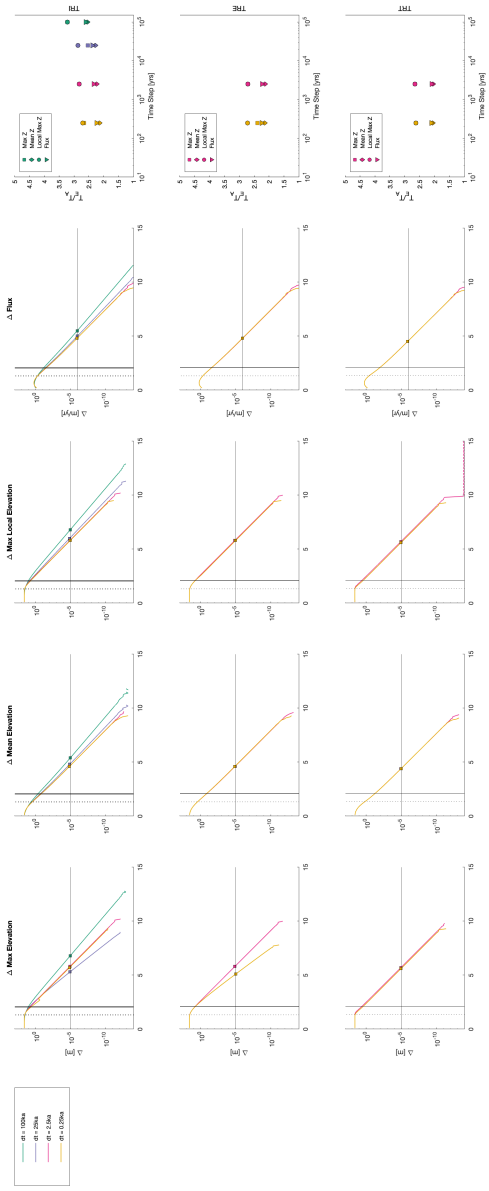

Figure 3. This figure is similar to Figures 1 and 2 except we only show TTLEM model results, and we did not reroute the flow after every time step. In other words, there was no drainage rearrangement in these simulations. Note the x axis in the first four columns is time in millions of years. The extent of the x and y axes in all of the time series is the same. The extent of the x and y axes

~~Our study was not extensive. A much deeper model comparison study could reveal other LEM strengths and weaknesses. This study only compared models using the stream power process equation. However, there are other processes shaping a landscape that may be included in different LEMs. There are also other fluvial process equations in LEMs. Most LEMs require flow routing and drainage area calculations in order to calculate fluvial transport or incision rates. Flow routing and drainage area algorithms can vary in and among modeling environments, and these details have been shown to matter (e.g., Shelef and Hilley, 2013; Adams et al., 2017; Lai and Anders, 2018). Explicitly stating what different LEMs can and cannot do is critical for appropriate model use.~~

7 Implications and Conclusions

7.1 ~~What (could be) next?~~

~~The surface processes community is pushing LEMs further and further. LEMs are being used for global prediction of sediment flux over 100 my (Salles et al., 2023). On the opposite timescale, Chen et al. (2021) used a LEM to explore how humans have impacted soil erosion over the Holocene. We are extremely excited about the proliferation of LEMs, and their use in exploring diverse surface processes questions. This paper is **not** suggesting that LEMs are only for an exclusive group of users. At present though, many of the limitations, best practices, and challenges related to~~ At a minimum, the results of our experiments highlight that when simulating landscape evolution scenarios and considering topographic steady state—and especially time to steady state—it is critical to both report the metric being used and the threshold value for that metric to assess steady state. Generally, the strictly topographic metrics, i.e., those only tracking mean and maximum elevations, behaved similarly in the sense that a single threshold value can be used to estimate time to steady state within a specific simulation (e.g., figures 1 & 2). The flux-based metric also behaved similarly among simulations, although it required a lower threshold value to yield a comparable estimate.

Given the variability in T_E with respect to T_A from our experiments, a deeper question, regardless of exactly how steady state is assessed, is whether response times as interpreted from 2D landscape evolution models are meaningful or reliable. At first glance, the ~~use of particular modeling environments are only known to small groups of users and are communicated informally. This informal sharing of knowledge makes using LEMs inherently exclusive despite the ease of access of the modelsthemselves. In response, we think that in order for LEMs to be used by everyone, our community deserves and requires a better, widely communicated, quantitative understanding of the strengths, limits, and pitfalls of different LEMs. Fortunately, a large number of scientific disciplines have already established practices for model benchmarking that the LEM development and application community can learn from~~combination of T_E being consistently greater than T_A but by inconsistent magnitudes for a given identical set of conditions (e.g., uplift rate change, erosional efficiency, algorithm, grid type, and dt), would understandably give one reason to question the utility of response times from 2D landscape evolution models. An important related question, though, is the extent to which we expect the conditions assumed in calculating analytical response times T_A to be met in real landscapes, i.e., is T_A a fair or reasonable benchmark? Implicit in the calculation of T_A is an assumption of static stream length and drainage area during response to a change in either K or U (e.g., Whipple, 2001; Goren, 2016). Further, T_A

reduces the network into 1-D, and the network structure provides a more complex transient response than predicted by 1-D models Li et al. (2018).

~~We suggest that the community of LEM developers and super users—those who have vast experience in using LEMs—should come together to design benchmark experiments that illustrate what particular LEMs can and cannot do, similar to other communities in~~ Comparing our experiments where flow routing is recomputed at every time step (figure 1) with those where the drainage network is fixed (figure 3) suggests that small perturbations in drainage network structure (e.g., single pixel to pixel changes in divide location or along the course of individual drainages) likely cause the longer response time compared to the predictions of T_A. Recent work highlighting that divide migration and drainage network instability may be more the norm (e.g. Willett et al., 2014; Whipple et al., 2017; Beeson et al., 2017; Forte and Whipple, 2018; Val et al., 2022), complicates the assumption of a static drainage network implied by calculations of T_A. Importantly, the ~~geosciences. These benchmarks would likely start with reproducing known analytical solutions and then expand into more difficult targets such as transient behavior and planview drainage patterns. Where analytical solutions are not known, established methods, such as the method of manufactured solutions (Roache, 1998) provide a way forward. A possible guide for identifying and prioritizing benchmark experiments is the type of model output most commonly interpreted.~~ majority of prior results arguing for pervasive drainage network reorganization consider scenarios with spatial gradients in either K or U, which our simple experiments do not include. It is possible that response times from 2D simulations driven by spatial gradients in environmental drivers and which generally induce greater amounts of network reorganization, like those considered for divide migration by Whipple et al. (2017) or Lyons et al. (2020), may be less sensitive to the initial conditions and small perturbations in the network than our simulations here. However, testing this is beyond the scope of this short communication.

~~After a benchmark experiment is designed, individual models can submit their performance. The collective development of these benchmarks and individual model performance on each benchmark would result in illuminating the domain of applicability for each model. That is~~ In summary, if we are primarily concerned with the legitimacy of using results of landscape simulations to establish the time scale of processes or the evolution of specific landscapes, it is not immediately apparent that T_E as estimated from 2D numerical simulations is any more or less grounded in reality than T_A, but the variability of T_E with respect to T_A is especially problematic. Based on our experiments and the above, we provide a set of recommendations for considering steady state and response times from 2D simulations. First and foremost, it is critical to remain mindful of the importance of the initial conditions (e.g., ~~what are the reasonable parameters, boundary and initial conditions , grid resolutions, and time steps for which a model will provide reliable results?~~ the random noise grid, or estimates of paleotopography) in dictating landscape evolution (e.g., Willgoose et al., 1991; Perron and Fagherazzi, 2012; Ferrier et al., 2013; Han et al., 2014; Ward and Ga and to treat properties related to response times as more stochastic. Thus, one potentially useful approach is to consider T_A as a minimum response time. Ranges of more reasonable response times, incorporating some stochastic degree of minor drainage reorganization, can be estimated from multiple 2D simulations with different initial topographic conditions (i.e., random noise) for a given fixed set of other environmental factors. Also, our results highlight that choice of grid is meaningful in terms of reliability of T_E, with either voronoi or hex grids more likely to produce T_E estimates that do not vary as much as a function of dt. Thus these grid types may be more suitable where reliability of response times is important or where simulations run with

different *dt* might be compared. The absolute most conservative approach to our results is to largely ignore response times as derived from 2D landscape evolution models and instead focus on landscape forms, either during the transient response or at quasi steady state, or as a function of non-dimensional time within a simulation.

~~Ultimately, the development of a vibrant culture of model benchmarking for LEMs will make our codes more reliable, our documentation better, and our science stronger. The possibilities grow rapidly and well beyond what was presented in this study. Yet even our highly contained exploration illustrated pitfalls that we , as experienced modelers, didn't know about. This raises the question of how some of what we presented might be interpreted by a novice modeler. No model developer wants results from their software incorrectly interpreted.~~

~~How such an effort might come to be remains an open question. Likely there would be multiple meetings and white papers written. Ideally the meeting participants would include scientists new to numerical modeling, as well as scientists who have designed and run models for decades. It would be important to include participants from diverse surface processes communities. For example, those who work on long and short time scales and those who work in diverse landscapes, e.g. glaciated, arid, and tropical. Scientists from different countries and different types of institutes should be part of the discussion. Only with broad community buy-in would such an effort have impact.~~

~~We also want to emphasize that model comparison is not solely about identifying when models do not work. It also highlights the strengths of different models and which models are best for testing different hypotheses. All modelers know the George Box saying "all models are wrong, but some models are useful". Maybe a more positive spin on the saying would be "all models are useful for something, but you have to know what that something is ". It's clunkier and won't take off like Box's saying, but the nuance is important. If we explicitly test for the strengths and weakness of a model, the science using LEMs will be stronger.~~

Code and data availability. CHILD is distributed from this repository: https://github.com/childmodel/child; accessed 1 May 2023. Landlab is distributed from this repository: https://github.com/landlab/landlab; accessed 1 May 2023. TTLEM is distributed from this repository: https://github.com/wschwanghart/topotoolbox; accessed 1 May 2023. The initial grids, input files, and code for creating the figures for this manuscript are available from this repository: https://github.com/nicgaspar/LEM_comparison.

Author contributions. NG initiated this study and ran all the CHILD simulations. KB ran all the Landlab simulations. AF ran all the TTLEM simulations and made figures for the manuscript. NG wrote the first draft of the manuscript. All authors contributed to the design and direction of the study, interpreting the results, and refining the manuscript.

Competing interests. The authors have no competing interests to declare.

Acknowledgements. We gratefully acknowledge funding from the Tulane Oliver Fund, NSF Awards OAC-1450338 (NMG), EAR-1349375 (NMG), and EAR-1725774 (KRB). We acknowledge contributions from Nathan Lyons in the initial stages of this project. Conversations with Kelin Whipple, Brian Yanites, and Leif Karlstrom encouraged us to pursue this work. CSDMS enabled this modeling study. We also acknowledge the impact of the global COVID-19 pandemic on this study. The first draft of this manuscript was written in January 2020, but the work was put aside for two years due to NMG being overwhelmed by the excess work and family burdens created during a pandemic.

References

Adams, B. A., Whipple, K. X., Forte, A. M., Heimsath, A. M., and Hodges, K. V.: Climate controls on erosion in tectonically active landscapes, Science Advances, 6, https://doi.org/10.1126/sciadv.aaz3166, 2020.

Adams, J. M., Gasparini, N. M., Hobley, D. E., Tucker, G. E., Hutton, E. W., Nudurupati, S. S., and Istanbulluoglu, E.: The Landlab v1. 0 OverlandFlow component: a Python tool for computing shallow-water flow across watersheds, Geoscientific Model Development, 10, 1645–1663, 2017.

Ahnert, F.: Brief description of a comprehensive three-dimensional process-response model for landform development, Zeitschrift fur Geomorphologie Supplement, 25, 29–49, 1976.

Anders, A. M., Roe, G. H., Montgomery, D. R., and Hallet, B.: Influence of precipitation phase on the form of mountain ranges, Geology, 36, 479, https://doi.org/10.1130/G24821A.1, 2008.

Armitage, J. J., Whittaker, A. C., Zakari, M., and Campforts, B.: Numerical modelling of landscape and sediment flux response to precipitation rate change, Earth Surface Dynamics, 6, 77–99, 2018.

Attal, M., Tucker, G., Whittaker, A. C., Cowie, P., and Roberts, G. P.: Modeling fluvial incision and transient landscape evolution: Influence of dynamic channel adjustment, Journal of Geophysical Research: Earth Surface, 113, 2008.

Attal, M., Cowie, P. A., Whittaker, A. C., Hobley, D., Tucker, G. E., and Roberts, G. P.: Testing fluvial erosion models using the transient response of bedrock rivers to tectonic forcing in the Apennines, Italy, Journal of Geophysical Research: Earth Surface, 116, https://doi.org/10.1029/2010JF001875, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2010JF001875, 2011.

Barall, M. and Harris, R. A.: Metrics for Comparing Dynamic Earthquake Rupture Simulations, Seismological Research Letters, 86, 223–235, https://doi.org/10.1785/0220140122, 2015.

Barnhart, K. R., Glade, R. C., Shobe, C. M., and Tucker, G. E.: Terrainbento 1.0: a Python package for multi-model analysis in long-term drainage basin evolution, Geoscientific Model Development, 12, 1267–1297, https://doi.org/10.5194/gmd-12-1267-2019, publisher: Copernicus GmbH, 2019.

Barnhart, K. R., Hutton, E. W. H., Tucker, G. E., Gasparini, N. M., Istanbulluoglu, E., Hobley, D. E. J., Lyons, N. J., Mouchene, M., Nudurupati, S. S., Adams, J. M., and Bandaragoda, C.: Short communication: Landlab v2.0: a software package for Earth surface dynamics, Earth Surface Dynamics, 8, 379–397, https://doi.org/10.5194/esurf-8-379-2020, 2020a.

Barnhart, K. R., Tucker, G. E., Doty, S. G., Shobe, C. M., Glade, R. C., Rossi, M. W., and Hill, M. C.: Inverting Topography for Landscape Evolution Model Process Representation: 1. Conceptualization and Sensitivity Analysis, Journal of Geophysical Research: Earth Surface, 125, e2018JF004 961, https://doi.org/10.1029/2018JF004961, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018JF004961, 2020b.

Beeson, H. W., McCoy, S. W., and Keen-Zebert, A.: Geometric disequilibrium of river basins produces long-lived transient landscapes, Earth and Planetary Science Letters, 475, 34–43, 2017.

Braun, J. and Deal, E.: Implicit algorithm for threshold Stream Power Incision Model, Journal of Geophysical Research: Earth Surface, p. e2023JF007140, 2023.

Braun, J. and Willett, S. D.: A very efficient 0(n), implicit and parallel method to solve the stream power equation governing fluvial incision and landscape evolution, Geomorphology, 180-181, 170–179, https://doi.org/10.1016/j.geomorph.2012.10.008, 2013.

Brocard, G. Y., Willenbring, J. K., Miller, T. E., and Scatena, F. N.: Relict landscape resistance to dissection by upstream migrating knickpoints, Journal of Geophysical Research: Earth Surface, 121, 1182–1203, 2016.

Buiter, S. J. H., Schreurs, G., Albertz, M., Gerya, T. V., Kaus, B., Landry, W., le Pourhiet, L., Mishin, Y., Egholm, D. L., Cooke, M., Maillot, B., Thieulot, C., Crook, T., May, D., Souloumiac, P., and Beaumont, C.: Benchmarking numerical models of brittle thrust wedges, Journal of Structural Geology, 92, 140–177, https://doi.org/10.1016/j.jsg.2016.03.003, 2016.

Campforts, B. and Govers, G.: Keeping the edge: A numerical method that avoids knickpoing smearing when solving the stream power law, Journal of Geophysical Research: Earth Surface, 120, 1189–1205, https://doi.org/10.1002/2014JF003376.Received, 2015.

Campforts, B., Schwanghart, W., and Govers, G.: Accurate simulation of transient landscape evolution by eliminating numerical diffusion: the TTLEM 1.0 model, Earth Surface Dynamics, 5, 47–66, https://doi.org/10.5194/esurf-5-47-2017, publisher: Copernicus GmbH, 2017.

Campforts, B., Shobe, C. M., Overeem, I., and Tucker, G. E.: The art of landslides: How stochastic mass wasting shapes topography and influences landscape dynamics, Journal of Geophysical Research: Earth Surface, 127, e2022JF006 745, 2022.

Castelltort, S. and Van Den Driessche, J.: How plausible are high-frequency sediment supply-driven cycles in the stratigraphic record?, Sedimentary geology, 157, 3–13, 2003.

Chase, C. G.: Fluvial landsculpting and the fractal dimension of topography, Geomorphology, 5, 39–57, https://doi.org/10.1016/0169-555X(92)90057-U, 1992.

Chen, A., Darbon, J., and Morel, J.-M.: Landscape evolution models: A review of their fundamental equations, Geomorphology, 219, 68–86, https://doi.org/10.1016/j.geomorph.2014.04.037, 2014.

Chen, H., Wang, X., Lu, H., and Van Balen, R.: Anthropogenic impacts on Holocene fluvial dynamics in the Chinese Loess Plateau, an evaluation based on landscape evolution modeling, Geomorphology, 392, 107 935, 2021.

Coulthard, T. J.: Landscape evolution models: a software review, Hydrological Processes, 15, 165–173, 2001.

Crosby, B. T., Whipple, K. X., Gasparini, N. M., and Wobus, C. W.: Formation of fluvial hanging valleys: Theory and simulation, Journal of Geophysical Research: Earth Surface, 112, https://doi.org/10.1029/2006JF000566, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2006JF000566, 2007.

de Almeida, G. A., Bates, P., Freer, J. E., and Souvignet, M.: Improving the stability of a simple formulation of the shallow water equations for 2-D flood modeling, Water Resources Research, 48, 2012.

Densmore, A. L.: Footwall topographic development during continental extension, Journal of Geophysical Research, 109, F03 001, https://doi.org/10.1029/2003JF000115, 2004.

Densmore, A. L., Allen, P. A., and Simpson, G.: Development and response of a coupled catchment fan system under changing tectonic and climatic forcing, Journal of Geophysical Research: Earth Surface, 112, 2007.

Dietterich, H. R., Lev, E., Chen, J., Richardson, J. A., and Cashman, K. V.: Benchmarking computational fluid dynamics models of lava flow simulation for hazard assessment, forecasting, and risk management, Journal of Applied Volcanology, 6, 9, https://doi.org/10.1186/s13617-017-0061-x, 2017.

Ferrier, K. L., Huppert, K. L., and Perron, J. T.: Climatic control of bedrock river incision, Nature, 496, 206–209, 2013.

Flint, J. J.: Stream gradient as a function of order, magnitude, and discharge, Water Resources Research, 10, 969–973, 1974.

Forte, A. M. and Whipple, K. X.: Criteria and tools for determining drainage divide stability, Earth and Planetary Science Letters, 493, 102–117, https://doi.org/10.1016/j.epsl.2018.04.026, 2018.

Forzoni, A., Storms, J. E., Whittaker, A. C., and de Jager, G.: Delayed delivery from the sediment factory: Modeling the impact of catchment response time to tectonics on sediment flux and fluvio-deltaic stratigraphy, Earth Surface Processes and Landforms, 39, 689–704, 2014.

Gasparini, N. M., Tucker, G. E., and Bras, R. L.: Network-scale dynamics of grain-size sorting: Implications for downstream fining, stream profile concavity, and drainage basin morphology, Earth Surface Processes and Landforms, 29, 401–421, 2004.

Gasparini, N. M., Whipple, K. X., and Bras, R. L.: Predictions of steady state and transient landscape morphology using sediment-flux-dependent river incision models, Journal of Geophysical Research: Earth Surface, 112, https://doi.org/10.1029/2006JF000567, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2006JF000567, 2007.

Gerya, T. V.: Cambridge University Press, 2010.

Godard, V., Tucker, G. E., Burch Fisher, G., Burbank, D. W., and Bookhagen, B.: Frequency-dependent landscape response to climatic forcing, Geophysical Research Letters, 40, 859–863, 2013.

Goren, L.: A theoretical model for fluvial channel response time during time-dependent climatic and tectonic forcing and its inverse applications, Geophysical Research Letters, 43, https://doi.org/10.1002/2016GL070451, 2016.

Hack, J. T.: Studies of longitudinal stream profiles in Virginia and Maryland, vol. 294, US Government Printing Office, 1957.

Han, J., Gasparini, N. M., Johnson, J. P., and Murphy, B. P.: Modeling the influence of rainfall gradients on discharge, bedrock erodibility, and river profile evolution, with application to the Big Island, Hawai'i, Journal of Geophysical Research: Earth Surface, 119, 1418–1440, 2014.

Han, J., Gasparini, N. M., and Johnson, J. P. L.: Measuring the imprint of orographic rainfall gradients on the morphology of steady-state numerical fluvial landscapes: OROGRAPHIC RAINFALL AND STEADY-STATE FLUVIAL LANDSCAPES, Earth Surface Processes and Landforms, 40, 1334–1350, https://doi.org/10.1002/esp.3723, 2015.

Harris, R. A. and Archuleta, R. J.: Earthquake rupture dynamics: Comparing the numerical simulation methods, Eos, Transactions American Geophysical Union, 85, 321, https://doi.org/10.1029/2004EO340003, 2004.

Harris, R. A., Barall, M., Archuleta, R., Dunham, E., Aagaard, B., Ampuero, J. P., Bhat, H., Cruz-Atienza, V., Dalguer, L., Dawson, P., Day, S., Duan, B., Ely, G., Kaneko, Y., Kase, Y., Lapusta, N., Liu, Y., Ma, S., Oglesby, D., Olsen, K., Pitarka, A., Song, S., and Templeton, E.: The SCEC/USGS Dynamic Earthquake Rupture Code Verification Exercise, Seismological Research Letters, 80, 119–126, https://doi.org/10.1785/gssrl.80.1.119, 2009.

Harris, R. A., Barall, M., Andrews, D. J., Duan, B., Ma, S., Dunham, E. M., Gabriel, A.-A., Kaneko, Y., Kase, Y., Aagaard, B. T., Oglesby, D. D., Ampuero, J.-P., Hanks, T. C., and Abrahamson, N.: Verifying a Computational Method for Predicting Extreme Ground Motion, Seismological Research Letters, 82, 638–644, https://doi.org/10.1785/gssrl.82.5.638, 2011.

Hilley, G., Strecker, M. R., and Ramos, V.: Growth and erosion of fold-and-thrust belts with an application to the Aconcagua fold-and-thrust belt, Argentina, Journal of Geophysical Research: Solid Earth, 109, 2004.

Hobley, D. E. J., Adams, J. M., Nudurupati, S. S., Hutton, E. W. H., Gasparini, N. M., Istanbulluoglu, E., and Tucker, G. E.: Creative computing with Landlab: an open-source toolkit for building, coupling, and exploring two-dimensional numerical models of Earth-surface dynamics, Earth Surface Dynamics, 5, 21–46, https://doi.org/10.5194/esurf-5-21-2017, 2017.

Howard, A. D.: A detachment-limited model of drainage basin evolution, Water Resources Research, 30, 2261–2285, https://doi.org/10.1029/94WR00757, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/94WR00757, 1994.

Hurst, M. D., Grieve, S. W., Clubb, F. J., and Mudd, S. M.: Detection of channel-hillslope coupling along a tectonic gradient, Earth and Planetary Science Letters, 522, 30–39, https://doi.org/10.1016/j.epsl.2019.06.018, 2019.

Kirby, E. and Whipple, K. X.: Expression of active tectonics in erosional landscapes, Journal of Structural Geology, 44, 54–75, 2012.

Kwang, J. S. and Parker, G.: Landscape evolution models using the stream power incision model show unrealistic behavior when m/ n equals 0.5, Earth Surface Dynamics, 5, 807–820, 2017.

Lague, D.: The stream power river incision model: evidence, theory and beyond, Earth Surface Processes and Landforms, 39, 38–61, https://doi.org/10.1002/esp.3462, 2014.

Lai, J. and Anders, A. M.: Modeled postglacial landscape evolution at the southern margin of the Laurentide ice sheet: Hydrological connection of uplands controls the pace and style of fluvial network expansion, Journal of Geophysical Research: Earth Surface, 123, 967–984, 2018.

Li, Q., Gasparini, N. M., and Straub, K. M.: Some signals are not the same as they appear: How do erosional landscapes transform tectonic history into sediment flux records?, Geology, 46, 407–410, 2018.

Lyons, N. J., Val, P., Albert, J. S., Willenbring, J. K., and Gasparini, N. M.: Topographic controls on divide migration, stream capture, and diversification in riverine life, Earth Surface Dynamics, 8, 893–912, https://doi.org/10.5194/esurf-8-893-2020, 2020.

Mackey, B. H., Scheingross, J. S., Lamb, M. P., and Farley, K. A.: Knickpoint formation, rapid propagation, and landscape response following coastal cliff retreat at the last interglacial sea-level highstand: Kaua ‘i, Hawai ‘i, Bulletin, 126, 925–942, 2014.

Martin, Y. and Church, M.: Numerical modelling of landscape evolution: geomorphological perspectives, Progress in Physical Geography, 28, 317–339, 2004.

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J.: The Coupled Model Intercomparison Project (CMIP), Bulletin of the American Meteorological Society, 81, 313–318, https://www.jstor.org/stable/26215108, publisher: American Meteorological Society, 2000.

Perron, J. T. and Fagherazzi, S.: The legacy of initial conditions in landscape evolution, Earth Surface Processes and Landforms, 37, 52–63, 2012.

Perron, J. T. and Royden, L. H.: An integral approach to bedrock river profile analysis, Earth Surface Processes and Landforms, 38, 570–576, https://doi.org/10.1002/esp.3302, 2013.

Roache, P. J.: Verification of codes and calculations, AIAA Journal, 36, 696–702, https://doi.org/10.2514/3.13882, 1998.

Roe, G. H., Whipple, K. X., and Fletcher, J. K.: Feedbacks among climate, erosion, and tectonics in a critical wedge orogen, American Journal of Science, 308, 815–842, 2008.

Roering, J. J.: How well can hillslope evolution models “explain” topography? Simulating soil transport and production with high-resolution topographic data, GSA Bulletin, 120, 1248–1262, https://doi.org/10.1130/B26283.1, 2008.

Romans, B. W., Castelltort, S., Covault, J. A., Fildani, A., and Walsh, J.: Environmental signal propagation in sedimentary systems across timescales, Earth-Science Reviews, 153, 7–29, 2016.

Rosenbloom, N. and Anderson, R. S.: Hillslope and channel evolution in a marine terraced landscape, Santa Cruz, California, Journal of Geophysical Research, 99, 14 013–14 029, 1994.

Salles, T., Husson, L., Rey, P., Mallard, C., Zahirovic, S., Boggiani, B. H., Coltice, N., and Arnould, M.: Hundred million years of landscape dynamics from catchment to global scale, Science, 379, 918–923, https://doi.org/10.1126/science.add2541, publisher: American Association for the Advancement of Science, 2023.

Schlesinger, S., Crosbie, R. E., Gagné, R. E., Innis, G. S., Lalwani, C. S., Loch, J., Sylvester, R. J., Wright, R. D., Kheir, N., and Bartos, D.: Terminology for model credibility, Simulation, 32, 103–104, https://doi.org/10.1177/003754977903200304, 1979.

Schwanghart, W. and Scherler, D.: Short Communication: TopoToolbox 2 - MATLAB based software for topographic analysis and modeling in Earth surface sciences, Earth Surface Dynamics, 2, 1–7, https://doi.org/10.5194/esurf-2-1-2014, 2014.

Shelef, E. and Hilley, G. E.: Impact of flow routing on catchment area calculations, slope estimates, and numerical simulations of landscape development, Journal of Geophysical Research: Earth Surface, 118, 2105–2123, 2013.

Shobe, C. M., Tucker, G. E., and Barnhart, K. R.: The SPACE 1.0 model: a Landlab component for 2-D calculation<? xmltex\break?> of sediment transport, bedrock erosion, and landscape evolution, Geoscientific Model Development, 10, 4577–4604, 2017.

Shobe, C. M., Tucker, G. E., and Rossi, M. W.: Variable-Threshold Behavior in Rivers Arising From Hillslope-Derived Blocks, Journal of Geophysical Research: Earth Surface, 123, 1931–1957, https://doi.org/10.1029/2017JF004575, 2018.

Simpson, G. and Castelltort, S.: Model shows that rivers transmit high-frequency climate cycles to the sedimentary record, Geology, 40, 1131–1134, 2012.

Skinner, C. J. and Coulthard, T. J.: The sensitivity of Landscape Evolution Models to DEM grid cell size, Earth Surface Dynamics Discussions, pp. 1–26, 2022.

Snyder, N. P., Whipple, K. X., Tucker, G. E., and Merrits, D. J.: Landscape response to tectonic forcing: Digital elevation model analysis of stream profiles in the Mendocino triple junction region, northern California, Geological Society of America Bulletin, 112, 1250–1263, 2000.

Stolar, D. B., Willett, S. D., and Roe, G. H.: Climatic and tectonic forcing of a critical orogen, in: Tectonics, Climate, and Landscape Evolution, Geological Society of America, https://doi.org/10.1130/2006.2398(14), 2006.

Straub, K. M., Duller, R. A., Foreman, B. Z., and Hajek, E. A.: Buffered, incomplete, and shredded: The challenges of reading an imperfect stratigraphic record, Journal of Geophysical Research: Earth Surface, 125, e2019JF005 079, 2020.

Tarboton, D. G.: A new method for the determination of flow directions and upslope areas in grid digital elevation models, Water Resources Research, 33, 309–319, https://doi.org/10.1029/96wr03137, 1997.

Temme, A. J. A. M., Schoorl, J. M., Claessens, L., and Veldkamp, A.: Quantitative Modeling of Landscape Evolution, vol. 2, pp. 180–200, Elsevier, London, UK, http://oar.icrisat.org/7445/, 2013.

Tofelde, S., Bernhardt, A., Guerit, L., and Romans, B. W.: Times associated with source-to-sink propagation of environmental signals during landscape transience, Frontiers in Earth Science, 9, 628 315, 2021.

Tucker, G. and Whipple, K.: Topographic outcomes predicted by stream erosion models: Sensitivity analysis and intermodel comparison, Journal of Geophysical Research: Solid Earth, 107, ETG–1, 2002.

Tucker, G. E. and Bras, R. L.: Hillslope processes, drainage density, and landscape morphology, Water Resources Research, 34, 2751–2764, https://doi.org/10.1029/98WR01474, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/98WR01474, 1998.

Tucker, G. E. and Hancock, G. R.: Modelling landscape evolution, Earth Surface Processes and Landforms, 35, 28–50, https://doi.org/10.1002/esp.1952, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/esp.1952, 2010.

Tucker, G. E. and Slingerland, R. L.: Erosional dynamics, flexural isostasy, and long-lived escarpments: A numerical modeling study, Journal of Geophysical Research: Solid Earth, 99, 12 229–12 243, https://doi.org/10.1029/94JB00320, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/94JB00320, 1994.

Tucker, G. E., Gasparini, N. M., Bras, R. L., and Lancaster, S. L.: A 3D Computer Simulation Model of Drainage Basin and Floodplain Evolution: Theory and Applications, Technical report prepared for U.S. Army Corps of Engineers Construction Engineering Research Laboratory, 1999.

Tucker, G. E., Lancaster, S. T., Gasparini, N. M., and Bras, R. L.: The channel-hillslope intergrated landscape development model (CHILD), in: Landscape erosion and evolution modeling, edited by Harmon, R. S. and Doe, W. W., pp. 349–388, Springer, New York, 2001a.

Tucker, G. E., Lancaster, S. T., Gasparini, N. M., Bras, R. L., and Rybarczyk, S. M.: An object-oriened framework for distributed hydrologic and geomorphic modeling using triangulated irregular networks, Computers & Geosciences, 27, 959–973, 2001b.

Tucker, G. E., Hutton, E. W. H., Piper, M. D., Campforts, B., Gan, T., Barnhart, K. R., Kettner, A. J., Overeem, I., Peckham, S. D., McCready, L., and Syvitski, J.: CSDMS: a community platform for numerical modeling of Earth surface processes, Geoscientific Model Development, 15, 1413–1439, https://doi.org/10.5194/gmd-15-1413-2022, 2022.

Val, P., Lyons, N. J., Gasparini, N., Willenbring, J. K., and Albert, J. S.: Landscape evolution as a diversification driver in freshwater fishes, Frontiers in Ecology and Evolution, 9, 788 328, 2022.

van der Beek, P.: Modelling Landscape Evolution, in: Environmental Modelling, pp. 309–331, John Wiley & Sons, Ltd, https://doi.org/10.1002/9781118351475.ch19, section: 19 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118351475.ch19, 2013.

Ward, D. J. and Galewsky, J.: Exploring landscape sensitivity to the Pacific Trade Wind Inversion on the subsiding island of Hawaii, Journal of Geophysical Research: Earth Surface, 119, 2048–2069, 2014.

Whipple, K. X.: Fluvial landscape response time: How plausible is steady-state denudation?, American Journal of Science, 301, 313–325, 2001.

Whipple, K. X. and Meade, B.: Controls on the strength of coupling among climate, erosion, and deformation in two-sided, frictional orogenic wedges at steady state, Journal of Geophysical Research, 109, F01 011–F01 011, https://doi.org/10.1029/2003JF000019, 2004.

Whipple, K. X. and Meade, B.: Orogen response to changes in climatic and tectonic forcing, Earth and Planetary Science Letters, 243, 218–228, 2006.

Whipple, K. X. and Tucker, G. E.: Dynamics of the stream-power river incision model: Implications for height limits of mountain ranges, landscape response timescales, and research needs, Journal of Geophysical Research: Solid Earth, 104, 17 661–17 674, https://doi.org/10.1029/1999JB900120, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/1999JB900120, 1999.

Whipple, K. X. and Tucker, G. E.: Implications of sediment-flux-dependent river incision models for landscape evolution, Journal of Geophysical Research: Solid Earth, 107, ETG 3–1–ETG 3–20, https://doi.org/10.1029/2000JB000044, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2000JB000044, 2002.

Whipple, K. X., Forte, A. M., DiBiase, R. A., Gasparini, N. M., and Ouimet, W. B.: Timescales of landscape response to divide migration and drainage capture: Implications for the role of divide mobility in landscape evolution, Journal of Geophysical Research: Earth Surface, https://doi.org/10.1002/2016JF003973, 2017.

Whittaker, A. C.: How do landscapes record tectonics and climate, Lithosphere, 4, 160–164, 2012.

Whittaker, A. C. and Boulton, S. J.: Tectonic and climatic controls on knickpoint retreat rates and landscape response times, Journal of Geophysical Research: Earth Surface, 117, 2012.

Willett, S. D. and Brandon, M. T.: On steady states in mountain belts, Geology, 30, 175–178, 2002.

Willett, S. D., McCoy, S. W., Perron, J. T., Goren, L., and Chen, C.-Y.: Dynamic reorganization of river basins, Science, 343, 1248 765–1248 765, https://doi.org/10.1126/science.1248765, 2014.

Willgoose, G., Bras, R. L., and Rodriguez-Iturbe, I.: A coupled channel network growth and hillslope evolution model: 1. Theory, Water Resources Research, 27, 1671–1684, https://doi.org/10.1029/91WR00935, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/91WR00935, 1991.