Computers & Geosciences 90 (2016) 6–16

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Case study

Uncertainty quantification in modeling earth surface processes: more applicable for some types of models than for others

A. Brad Murray[a,*], Nicole M. Gasparini[b], Evan B. Goldstein[c], Mick van der Wegen[d,e]

[a] Division of Earth and Ocean Sciences, Nicholas School of the Environment; Center for Nonlinear and Complex Systems, Duke University, Box 90230, Durham, NC 27708, USA
[b] Department of Earth and Environmental Sciences, Tulane University, 101 Blessey Hall, New Orleans, LA 70118, USA
[c] Department of Geological Sciences, University of North Carolina at Chapel Hill, 104 South Rd, Mitchell Hall, Chapel Hill, NC 27515, USA
[d] UNESCO-IHE, PO Box 3015, 2601 DA Delft, Netherlands
[e] Deltares, PO Box 177, 2600 MH, Delft, Netherlands

ARTICLE INFO

Article history:
Received 11 March 2015
Received in revised form
2 February 2016
Accepted 10 February 2016
Available online 11 February 2016

Keywords:
Uncertainly quantification
Simulation models
Exploratory models

ABSTRACT

In Earth-surface science, numerical models are used for a range of purposes, from making quantitatively accurate predictions for practical or scientific purposes ('simulation' models) to testing hypotheses about the essential causes of poorly understood phenomena ('exploratory' models). We argue in this contribution that whereas established methods for uncertainty quantification (UQ) are appropriate (and crucial) for simulation models, their application to exploratory models are less straightforward, and in some contexts not relevant. Because most models fall between the end members of simulation and exploratory models, examining the model contexts under which UQ is most and least appropriate is needed. Challenges to applying state-of-the-art UQ to Earth-surface science models center on quantifying 'model-form' uncertainty—the uncertainty in model predictions related to model imperfections. These challenges include: 1) the difficulty in deterministically comparing model predictions to observations when positive feedbacks and associated autogenic dynamics (a.k.a. 'free' morphodynamics) determine system behavior over the timescales of interest (a difficulty which could be mitigated in a UQ approach involving statistical comparisons); 2) the lack of available data sets at sufficiently large space and/or time scales; 3) the inability to disentangle uncertainties arising from model parameter values and model form in some cases; and 4) the inappropriateness of model 'validation' in the UQ sense for models toward the exploratory end member of the modeling spectrum.

1. Introduction

Earth-surface science addresses a vast array of different processes, in different environments, and at different scales. Phenomena of interest include specific applications of fluid dynamics, in which the essential physics are relatively well understood. Models addressing fluid dynamics are typically applied over relatively small domains (meter to ten-kilometer scales) and relatively short periods (minutes to days). On the other end of the spectrum is landscape evolution, covering hundreds of square kilometers over time scales upwards of millions of years. Determining which interactions, over what spatial and temporal scales, are most critical in shaping a landscape and its associated ecosystem are often open questions. This range of scientific contexts in Earth-surface science gives rise to a range of different types of modeling endeavors, serving a range of purposes and employing a range of strategies.

In the end member of 'simulation models', modelers strive to explicitly represent all of the processes and interactions that quantitatively affect the way the system of interest behaves, and to represent those processes and interactions with maximal realism. In contrast, in the 'exploratory model' end member, modelers strive to leave out most of the processes and interactions occurring in the system of interest, in an effort to identify the essential causes of poorly understood phenomena. Simulation models are often used to make quantitative forecasts for practical purposes, or to provide quantitatively accurate information needed to address scientific questions (e.g., to provide input to other models). In either case, quantifying model uncertainty ('uncertainty quantification', or UQ) is essential. In contrast, when an exploratory model is used for hypothesis testing, questions about model uncertainties typically do not arise.

While the range of model types and how the appropriate model-evaluation approaches vary across the model spectrum

* Corresponding author.
E-mail address: abmurray@duke.edu (A.B. Murray).

http://dx.doi.org/10.1016/j.cageo.2016.02.008
0098-3004/Published by Elsevier Ltd.

have been discussed (e.g. Dietrich et al., 2003; Murray, 2002, 2007, 2013), an articulation of which modeling contexts UQ is appropriate for, and which contexts it is not, is currently missing. Even if the question of whether to employ UQ is clear for the end members of the simulation-exploratory model axis, most models fall between the ends of the spectrum, having some aspects of both simulation and exploratory modeling goals and strategies. Thus, a strong potential exists for confusion about when UQ makes sense, and when it does not. Here we analyze some modeling circumstances that make UQ more appropriate than might have been expected, and we analyze what limits the application of UQ in other modeling contexts, in the hopes of spurring further discussion and consideration of these issues.

Numerical models of all types involve uncertainty arising from multiple potential sources – model inputs, parameter values, and model form (e.g., Roy and Oberkampf, 2011). There are established methods for identifying and quantifying uncertainty, including model 'validation' to assess model-form uncertainty (e.g., Roy and Oberkampf, 2011; Oberkampf et al., 2002; Ferson et al., 2008). However, these methods are often only partly implemented in the Earth-surface-science modeling community. One reason for this may be that many modeling studies are used primarily for exploring emergent behavior of generic systems. The modeling is often motivated by specific field areas or observed phenomenon but the model outcomes are not meant to recreate the observations exactly (e.g., Attal et al., 2008; Cohen et al., 2015; Collins et al., 2004; Davy and Lague, 2009; Howard, 1997, 1999; Huang and Niemann, 2014; Pelletier et al., 2012; Roering, 2008; Saco and Moreno-de las Heras, 2013; Tucker and Bras, 1998; Yetemen et al., 2015). In many cases, data appropriate for strict validation does not even exist. In addition, strict model 'validation', as envisioned in the context of uncertainty quantification (UQ) may not be consistent with modeling goals in some cases.

This is not to say that Earth-surface-science modeling ignores the issue of uncertainty. For example, studies have sought to constrain parameter values for a given setting (e.g. Pelletier et al., 2011; van der Beek and Braun, 1998; Hancock et al., 2002) and have explored the impacts of variable model parameters, both absolute values and heterogeneity of values, (e.g. Moglen and Bras, 1995; Tucker and Whipple, 2002; van der Wegen and Jaffe, 2013) and model resolution (e.g., Pelletier, 2010; Schoorl et al., 2000), but these studies follow the structured framework of UQ (Roy and Oberkampf, 2011) only to a limited extent.

In this paper we first briefly overview an approach to UQ. We then describe the spectrum of models that are used in Earth-surface-science. We provide three examples that illustrate the challenges in applying UQ in some types of Earth-surface modeling. Finally, we discuss ways in which uncertainty can and should be addressed in Earth-surface modeling.

2. Uncertainty quantification approach

Here we briefly review the state-of-the-art approach to uncertainty quantification (UQ), as described in the review by Roy and Oberkampf (2011). In the following procedure, the situation envisioned is that a model is to be applied to make predictions about certain variables in a specific application. First, identify all uncertain model inputs and model parameters. Next, quantify the distributions of the possible values for those inputs and parameters, in the context of the specific application of interest, in ways appropriate for the nature of those uncertainties (unavoidable stochastic variations, "aleatory uncertainty", vs a lack of knowledge, "epistemic uncertainty", or a mixture). Then, propagate these distributions through the model, in an ensemble of model runs each involving different choices from the distributions (using a statistical method such as a Monte Carlo technique), to produce a distribution of results.

A separate but crucial set of steps quantifies the uncertainty in the model predictions arising from inevitable model imperfections, termed 'model form uncertainty.' Constituting a form of model 'validation', this step compares the distribution of values in an observational data set to the distribution of the modeled/predicted values (generated in an ensemble of runs using distributions of input and parameter values appropriate to the observed experiment or field situation). Model-form uncertainty is then based on the difference between observed values and the spread of predictions (e.g. Roy and Oberkampf, 2011).

3. A range of model types and modeling goals in earth-surface science

In the end-member limit, the ultimate goal of a simulation model is to mimic the natural or engineered system they are meant to represent—at least to produce the same output for a given set of inputs, and ideally to produce it for the same (usually physical) reasons. Model output can be used for practical purposes, providing an ability to make quantitative predictions or forecasts useful for planners, decision makers, and society more broadly. For example, forecasting the locations and depths of storm surge flooding from a range of possible storm scenarios provides crucial information for coastal managers and planners (e.g. Peng et al., 2004). In the simulation modeling end member context, when additional processes, additional detail, or additional resolution is believed to increase the quantitative reliability of model predictions, the additional processes, detail or resolution should be added (computing resource limitations aside). Because no model is perfect, and because inputs (and parameter values) are never known exactly, UQ is an essential part of the modeling endeavor. UQ provides the context for interpreting model predictions, and the model validation step in the UQ procedure outlined above, is completely appropriate.

In contrast, exploratory models are generally not used to make predictions for which quantitative accuracy is paramount. When designing an exploratory model, a modeler strives to include only processes hypothesized to be responsible for poorly understood phenomena and these processes are represented in simplified ways. Reducing the detail maximizes the clarity of the potential insights that can result. If such a highly simplified model produces, qualitatively, the phenomena of interest, the interactions in the model become a potential explanation. Of course, in an essential next step, the model behaviors need to be compared to observations in some way, to determine whether, or to what degree, the interactions in the model correspond to those in natural or anthropogenic systems (called 'natural' systems hereafter).

However, for models near the exploratory end member of the spectrum, simply comparing the value of model-output variables to values observed in a specific location and time may not be the most appropriate way of addressing the question of whether the interactions in the model captures the essence of the interactions in the natural system. First, such a model was constructed not to maximize quantitative accuracy, but rather to test hypotheses as clearly as possible. Directly comparing the magnitudes of modeled and observed variables does not necessarily help evaluate the model's success. Quantitative mismatches could be the result of either (intentional) model simplifications, or because the model interactions do not correspond to those in the prototype system. Conversely, if the model quantitatively reproduces the observations, the match could be coincidental—or could in many cases be achieved by tuning model parameters. For an exploratory end member, model success is directly evaluated by testing model

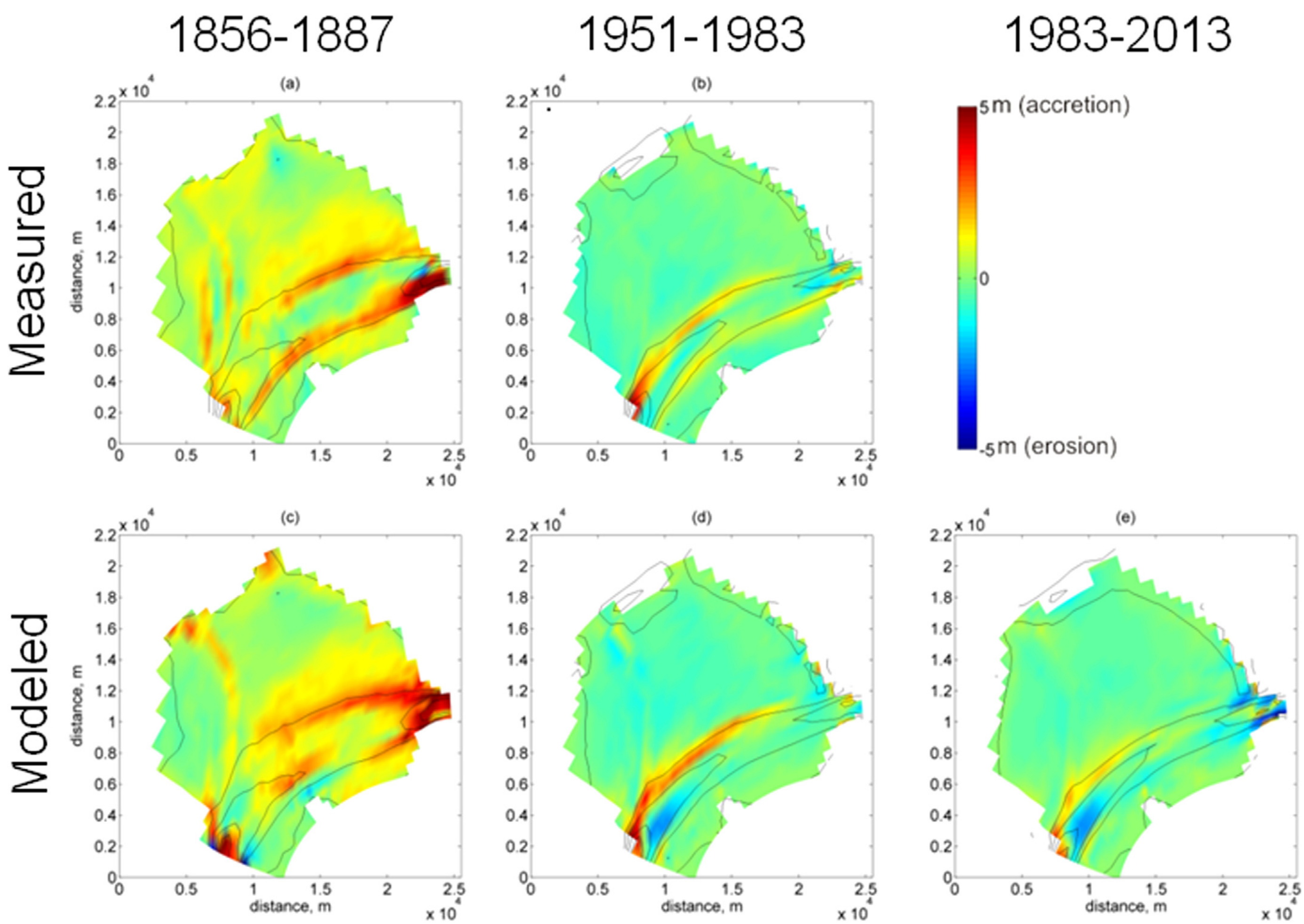

Fig. 1. Measured and predicted erosion and sedimentation patterns in San Pablo Bay, California, USA (from: van der Wegen and Jaffe (2013)).

predictions that are more robust (Murray, 2007; Tucker, 2009). For example, if a model always predicts that the value of output variable X increases as input variable Y increases, and the qualitative trend in nature does not match, then the interactions in the model must not correspond to the ones important in the natural system, and the model can be rejected and discarded.

Setting aside questions about the appropriateness of model 'validation' for models toward the exploratory end of the spectrum, other considerations can limit the ability to assess model-form uncertainty, as we discuss in the next two sections.

4. Free morphodynamics versus boundary conditions

Fluid dynamics can produce stunning examples of emergent structures and phenomena, such as turbulent eddies. However, many modeling contexts involve the mean flow, averaged over scales greater than those of the turbulence. In those cases, deterministic predictions are possible, and the accuracy of those predictions is limited not by the nature of the phenomena, but by modeling limitations (inputs, model approximations and parameterizations, solution procedures, and boundary conditions). When modeling changes in Earth-surface morphology, not only fluid flow but sediment transport is involved. When fluid flow and mobile sediment interact, especially if the sediment is transported as bed load, morphodynamic feedbacks typically lead to emergent structures and behaviors, such as bedforms of various sorts. Parameterizations for bulk sediment flux represent averages over the time and space scales of those structures—analogous to the way the effects of turbulence on mean flow are parameterized. Parameterizations representing the effects sub-grid-scale emergent morphodynamic feedbacks may be as quantitatively reliable as turbulence closure schemes. However, feedbacks between flow and sediment transport often also lead to autogenic behaviors at the scales of interest in Earth-surface science (e.g., Werner, 1999; Murray et al., 2008; Murray et al., 2014) that are not deterministically predictable.

For example, we can hope to predict, using a simulation model, how channels and shoals in an ebb tidal delta will change over short timescales (e.g. van Leeuwen et al., 2003; Coco et al., 2013). However channel and shoal location over longer timescales likely cannot be predicted deterministically (e.g. Cayocca, 2001; Elias and van der Spek, 2006; Coco et al., 2013)—just as the precise location of turbulent eddies cannot be deterministically predicted over timescales that are long compare to eddy lifetimes. Statistical forecasts are justified when such 'free' morphodynamic behaviors dominate the system's behavior. And when statistical predictions are desired, an appropriately modified approach to UQ would be required to judge the reliability of those predictions. However, the model validation stage should in that case not involve deterministic predictions. Mismatches between model predictions and observations could arise either from model imperfections or the effects of positive feedbacks—or a combination of the two that presents a challenge to disentangle.

In contrast to free and stochastic morphodynamic behavior the following example motivates the suggestion that boundary conditions can provide sufficiently strong constraints that prevailing patterns in the fluid dynamics largely dictate changes in the sediment bed. Ganju et al. (2009), van der Wegen et al. (2011) and van der Wegen and Jaffe (2013) report model efforts to hindcast measured decadal time scale morphodynamic developments observed in sub-embayments of San Francisco Estuary (using historical bathymetry from the beginning of the time periods as initial conditions). Remarkably (given the large amount of uncertain model input and process descriptions), model results had significant skill (quantified by the Briar Skill Score; van der Wegen and Jaffe, 2013) in reproducing the measured bathymetric developments over 30 year time frames (see also Fig. 1). In a classical UQ, van der Wegen and Jaffe (2013) showed that model results were mainly sensitive to boundary condition variations (i.e. sediment supply and wind wave forcing) and only to a limited extend to process-descriptions of the model itself (i.e. roughness, sediment transport characteristics). For example, for different 30 year time intervals, about 60–90% of the erosion and deposition volume was modeled with confidence given uncertainty levels in model input and process description. van der Wegen and Jaffe (2013) also explored a methodology to derive the model parameter leading to highest outcome uncertainty.

A probable explanation for the high skill scores was that the model included enough process description (the model was detailed enough) and that the estuaries' initial geometry (i.e. bathymetry and rocky plan form) played a governing role in erosion and deposition patterns. van der Wegen and Roelvink (2012) and Dam et al. (2013) also use this explanation in their skillful 110 year hindcast of morphodynamics in the Dutch Western Scheldt estuary. In other words, the interaction between the geometry and the modeled fluid dynamics and sediment transports captured a leading morphodynamic process. It would be questionable (or better: worth an upcoming modeling exercise) whether a similar type of modeling exercise could lead to comparable results in non-confined systems such as an ebb-tidal delta, where feedbacks and morphodynamic 'free' behavior steer system evolution.

This case study shows a model that is realistic enough to simulate changes at a particular field site, allowing for a skillful hindcast and prediction of morphodynamic developments. However, these modeling exercises also had exploratory elements. It was unknown beforehand whether or not the model would lead to acceptable results. The set of processes and input parameters acted as a hypothesis to explain observed behavior. In addition, the skill of the model is to some extent subjective, since its prediction is not perfect. It is very well possible that adding other processes would improve model results or that process formulations currently included capture the effects of more subtle processes that were not included. These are characteristics of exploratory models.

5. Scaling limitations on available parameterizations

Other challenges in applying UQ to some models of Earth surface processes and landscape change involve a lack of available parameterizations appropriate for the scales of interest, and for the range of processes involved. Often, the most well established and well calibrated parameterizations derive from studies on the scale of laboratory experiments or experimental plots in the field. As the examples below illustrate, when attempting to apply those parameterizations to model much larger scale phenomena, their interpretations are not straightforward. When the physical (or biological or ecological) meaning of the variables in a parameterization are not clear, finding the 'best' values for model parameters can become an exercise in model tuning, rather than an attempt to find the most physically (or biologically) realistic values. In these situations, the model 'validation' stage of the approach to UQ discussed above (Section 2) becomes entangled with the effort to identify uncertainties in model parameter values.

5.1. Sediment transport and evolution of continental shelves

Models representing sediment transport of mixed grain sizes on continental shelf systems highlight what we mean by model uncertainty in exploratory models. A series of models describing mixed grain size transport and bedform development on the inner continental shelf illuminates the uncertainty in process descriptions in heterogeneous non-idealized field environments. Murray and Thieler (2004) initially developed the first exploratory model of inner shelf 'sorted bedforms' using process-based rules. Coco et al. (2007a, 2007b) extended this model using accepted parameterizations of fluid mechanics and sediment transport processes that attempt to explicitly represent processes occurring on much smaller and faster scales than those of the phenomena of interest. Goldstein et al. (2014) further extended the model by replacing older empirical expressions with novel parameterizations built directly from data (using machine learning techniques).

Even with three generations of models over 10 years, significant gaps in knowledge remain when scaling up lab or theoretical fast- and small-scale fluid and sediment transport parameterizations to the field setting. For instance seabed ripples, whose size depends on grain size of the bed (Cummings et al., 2009; Goldstein et al., 2013) present roughness elements on the bed (1–10 s cm scale). Ripples control fluid flow, turbulence generation and, as a result, sediment suspension (e.g., Bolaños et al., 2012; Green et al., 2004; Green and Black, 1999). However it remains unclear how to parameterize the suspended sediment above large ripples in a model addressing spatial scales much larger than the ripples. Specifically, since the seabed elevation varies considerably in both space and time, it is unclear at what height reference concentration should be applied. Additionally the vertical suspended sediment concentration profile varies in space and time (it is different above the crest of a ripple vs. the trough of a ripple and different during different wave phases; Davies and Thorne, 2005; van der Werf et al., 2007; O'Hara Murray et al., 2011).

As another example, it remains unclear where the bottom current profile begins when roughness elements (ripples) with relief considerably greater than theoretical boundary layers are present and vary in size through the model domain (Coco et al., 2007a). Adjusting the location of the current profile relative to the suspended sediment profile (the location of the reference concentration) leads to dramatic changes in suspended sediment flux and bedform development (Coco et al., 2007a; Fig. 2).

These decisions highlight difficulties in accurately connecting the components of suspended sediment flux in models meant to capture large-scale pattern development where significant heterogeneity can occur at a sub-grid scale. One method to circumvent ambiguity in parameter values is to treat these parameters as tunable and adjusting reference concentration height, vertical concentration profile shape, and current profile height until model results match observations. (Note: this is only tenable when relevant observational datasets exist and they do not for inner shelf 'sorted bedforms'). This model calibration procedure may lead to the tuning of parameters to potentially nonphysical values, or values that are not prescribed by experimentation. These values do not necessarily compare favorably to observable values, and reflect a spatial average suggesting that reference concentration height, suspended sediment vertical shape, and current profile height are model parameters lacking a direct physical interpretation, and which might depend on model scale. After this calibration process, any adjustment in parameter values for UQ purposes ultimately

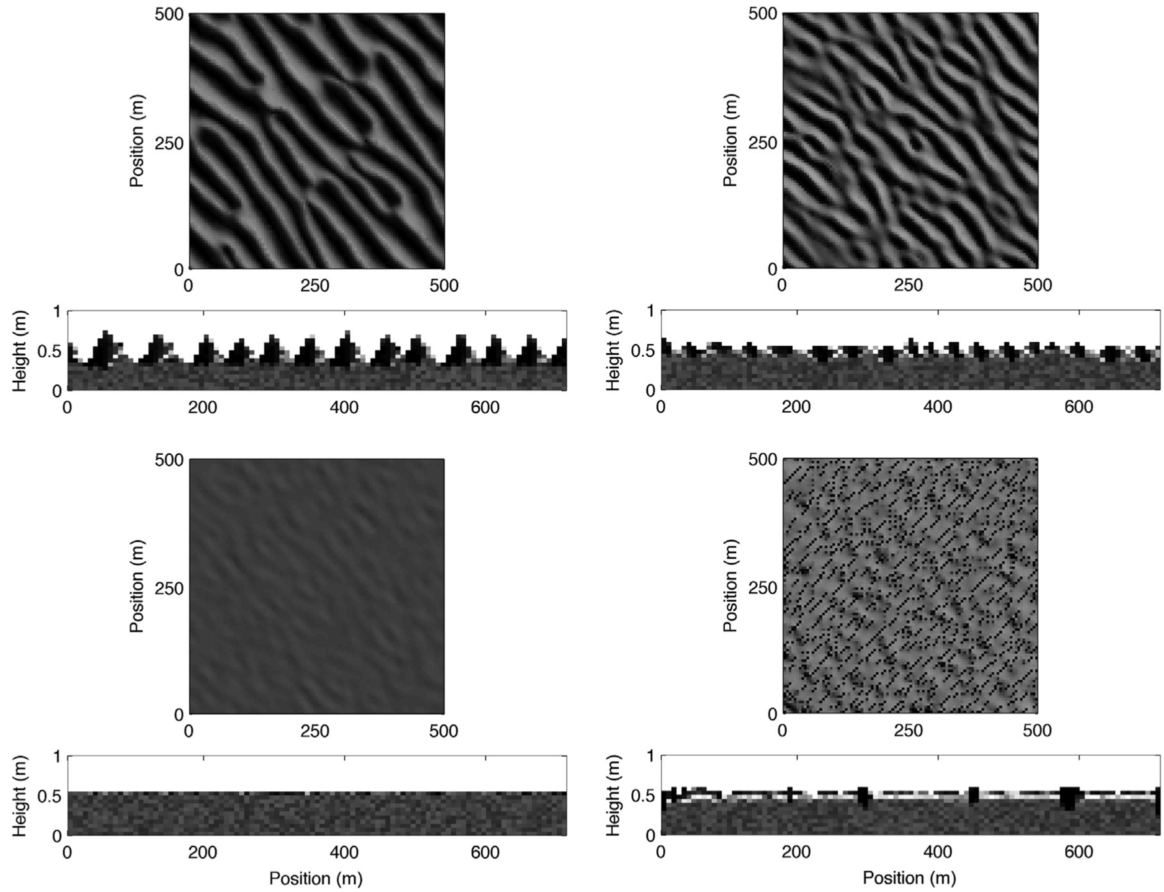

Fig. 2. Numerical experiments of 'sorted bedforms' using the model of Goldstein et al. (2014), showing changes in bedform pattern as the vertical location of the reference concentration is adjusted. Map-view and profile-view patterns after 200 model days are shown; black and white pixels represent fine and coarse sediment respectively (gray is well mixed). Profile view is taken along the diagonal from lower-left to top-right. The model domain has a vertical resolution of 0.05 m, a horizontal resolution of 5 m, and periodic boundary conditions in the horizontal direction. Initial conditions are a flat bed (with perturbations below 0.01 m) with well-mixed coarse (0.0005 m) and fine sand (0.0002 m) and 9 m of water depth. Forcing conditions are 2 m waves with 10 s period and a 0.2 m/s mean current along the diagonal (from bottom left to upper right) that reverses every 10 days. Each panel represents a different height at which the reference concentration is applied: Top left (ripple crest); Top right (half of the ripple crest); Bottom left (a quarter of the ripple crest); Bottom right (at the grain scale). Changes in sorted bedform pattern as a result of changes in reference concentration height is a robust observation, and has been seen in previous versions of the model (Murray et al., 2005; Coco et al., 2007a).

informs users about the model (i.e., what model behaviors are seen in specific regions of multidimensional parameter space), but does not necessarily yield insight into the range of variability in the natural system (i.e., variability in field examples).

5.2. Landscape evolution modeling of bedrock river systems

Landscape evolution models (LEMs) are generally applied as exploratory models. For example, landscape evolution studies are often focused on long time scales (e.g., implied or stated scales of thousands to millions of years) and large spatial scales (e.g., settings stated or suggested as mountain ranges or watersheds that are multiple to hundreds of square kilometers in area) (e.g., Anders et al., 2008; Attal et al., 2011, 2008; Braun et al., 2001; Colberg and Anders, 2014; Ellis et al., 1999; Gasparini and Whipple, 2014; Howard, 1994, 1997; Roe et al., 2003; Roering, 2008; Rosenbloom and Anderson, 1994; Tucker and Slingerland, 1997, 1996; Willgoose, 1994; Yanites and Ehlers, 2012). Over such scales, it is impossible to know exact initial conditions and the details of all the processes shaping the landscape, including the boundary conditions, such as climate and tectonics, which drive landscape evolution. However LEMs are a useful aid for quantitatively interpreting observed trends in real landscapes that are generally evolving at rates too slow to be observable over a human lifetime, leading to their increased use in the past two decades.

Geomorphic transport laws control topographic change in LEMs, and many of these laws fall into the category of "essential realism" as described by Dietrich et al. (2003) to mean laws, or equations, that can explain the "essential morphodynamic features of a landscape". These simplified laws are rooted in first principles and empirical relationships, yet they are highly simplified so that they can be applied over large spatial and temporal scales. LEMs run with such laws present a number of challenges for quantifying uncertainty, as will be illustrated with the stream power law or model (SPM).

A SPM is applied when the detachment rate of bedrock from a river channel controls the evolution of the channel, and the detachment rate can be described as a power law function of the fluvial shear stress, or similarly, the stream power per unit area of channel bed (e.g., Howard, 2004; Whipple and Tucker,1999). The details of the model will not be reiterated here, however for the unfamiliar, the SPM is based on relationships for conservation of mass and momentum and empirical relationships for downstream changes in channel width (hydraulic geometry) and basin hydrology. These relationships combine into a power law relationship between the river incision rate (I, L/T), drainage area (A, L^2) and channel slope (S, L/L):

$$I = KA^mS^n, \tag{1}$$

where K, m and n are all positive parameter values, and the units on K depend on the units of A and I and the value of m. The value of K varies with both rock strength (harder rocks are thought to have smaller K values) and climate (more erosive, or wetter climates are thought to have larger K values). The simplicity of Eq. (1) has led to its widespread use, despite the fact that it certainly does not directly include or parameterize all of the variables that affect fluvial incision into bedrock, such as rain storm variability through time (e.g., Tucker, 2004), the role that suspended and bed load sediment may have in increasing bedrock incision (e.g., Lamb et al., 2008), or the role that bed load sediment may play in inhibiting bedrock incision (e.g., Sklar and Dietrich, 2004). However, the effects of the latter may collapse to a power–law function of drainage area and slope (Gasparini and Brandon, 2011), and therefore can potentially be encapsulated in a form such as Eq. (1) with slightly different parameter values than one might predict from the shear stress or unit stream power formulation.

Uncertainty obviously arises in the appropriate values for the parameter values in the SPM when applied to a particular landscape. As an example of the difficulties in constraining the parameter values in the SPM, consider a bedrock river in which incision rates, rock type, and climate are uniform. In this case, the gradient of a log–log plot of slope versus drainage area should provide the value of m/n (illustrated by rearranging Eq. (1); e.g., Kirby and Whipple, 2012) (Fig. 3). However, if the goal is to model the evolution of a landscape through time, the values of K, m and n are all required. Even if the incision rate is known, K cannot be estimated from the slope–area plot because it co-varies with m and n, and therefore knowing only the value of m/n is not enough to estimate K (Fig. 3). If the incision rate is not known, then a channel with a steeper slope for a given drainage area could either be incising faster, or have harder rocks and/or a less erosive climate, or some combination of the two factors (assuming that m/n is fixed).

It would be ideal if K could be measured in the field, but no field calibration method for K currently exists. One challenge is

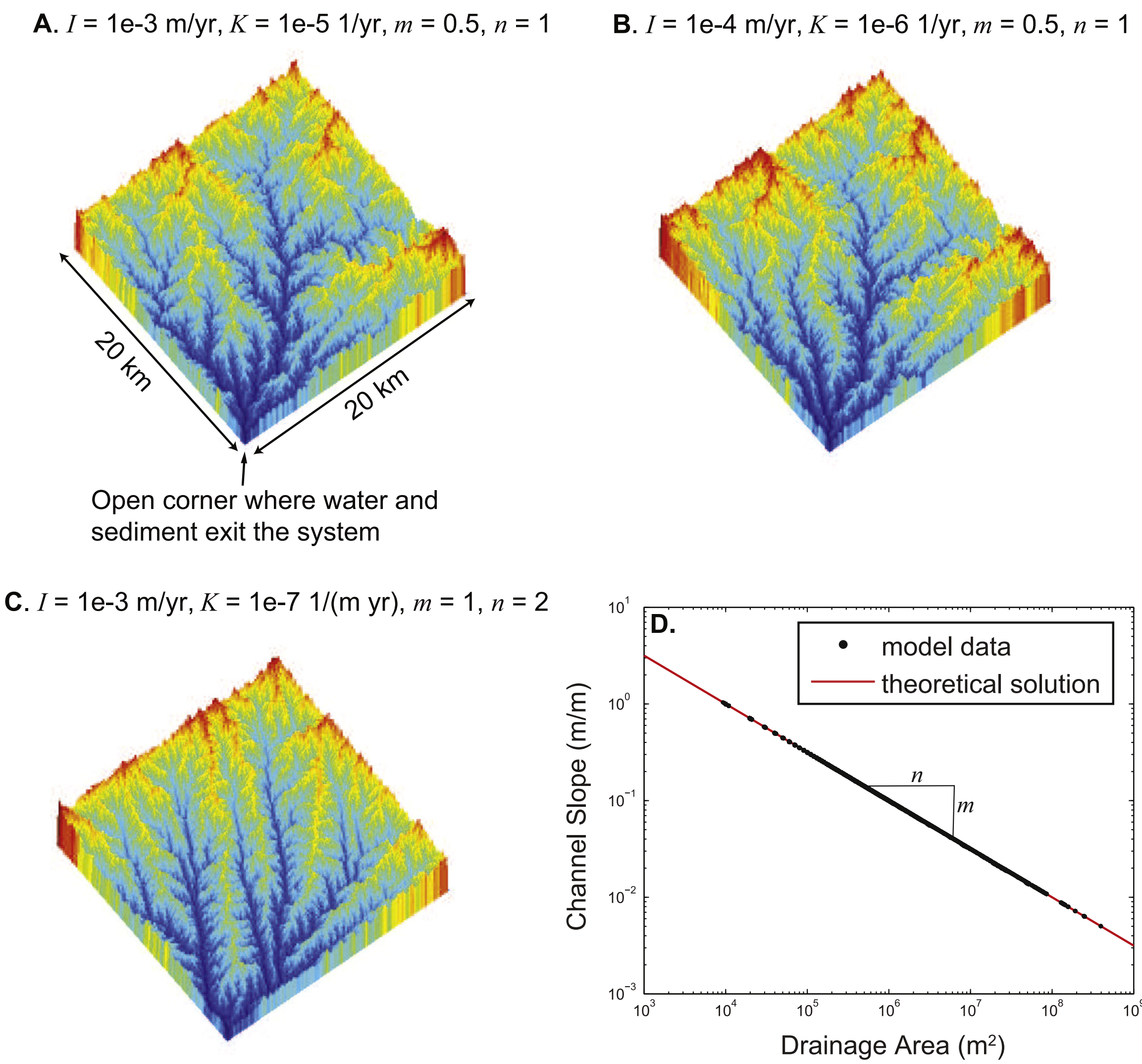

Fig. 3. A, B, and C illustrate example topographies (shaded by elevation, where blue is low and red is high) generated with the SPM model (using the CHILD LEM, (Tucker et al., 2001)), and D illustrates the slope–area data from the landscape shown in C. All of the landscapes have an ~100 m resolution (CHILD operates on a triangular irregular network, so node spacing is not perfectly regular), the same dimensions shown in A, the same initial condition (a flat surface of 1 m elevation with 0.5 m amplitude white noise, and the same exact noise in each) and the same boundary conditions (closed on all four sides with one open node in the lower corner that drains to a point of constant elevation of 0 m where water and sediment can exit the landscape, see A). Each landscape has uniform and steady rock uplift, and the landscapes shown have reached steady state, defined as the incision rate and uplift rate being equal at all locations on the landscape. Even though the incision rate, K, m, and n vary amongst the landscapes, they all produce the same exact slope–area relationship, the metric used for comparing K or incision rates among landscapes, and for estimating the value of m/n. In these examples, because all variables and parameters are uniform, the slope of the slope–area data in log–log space is equal to m/n. Note that only the slope–area data from landscape C are shown in D, but the slope–area data from A and B overlap those of C, making them indistinguishable. We also note that even though these landscapes all started from the same initial condition, the networks vary because the evolution to steady state varies depending on the parameters (see Tucker and Whipple (2002)). Note that in the labels, m refers to units of meters, and m refers to the exponent in Eq. (1).

that K parameterizes, at the very least, rock strength and climate. Even if climate could be parameterized in K as the mean upstream rainfall rate resulting in the effective discharge (as described by Whipple and Tucker, 1999), there is currently no way to take measurements of rock strength (such as from the Selby index (Selby, 1980) or Schmidt Hammer (Aydin and Basu, 2005) and translate them into a K value for use in Eq. (1).

The parameters in Eq. (1) can be calibrated, or tuned, to a given setting, and studies that have attempted to do so have met with varied results and successes. Stock and Montgomery (1999) fit K values through modeling of different landscapes (rivers in USA, Japan, and Australia) and found that K varied of over five orders of magnitude (although m and n were not fixed among these landscapes, adding to the variability in K). Similarly, Tomkin et al. (2003) calibrated a version of Eq. (1) to the Clearwater River in Washington State, USA. They arrived at the physically nonsensical value of m approximately equal to zero, which does not lead to the formation of a drainage network, an essential first step for the SPM to create realistic fluvial landforms. van der Beek and Bishop (2003) used four rivers in the Upper Lachlan watershed in southeast Australia to calibrate the SPM model. Although the parameter values they found were similar to those expected from standard derivations of the SPM model, the fit of the model to the observed channel form was not particularly good.

Despite these challenges, the SPM and the ideas behind it have been extremely useful in the geomorphic community. For example, the SPM has formed the basis for successfully interpreting many tectonically active landscapes to determine which parts of the landscapes are eroding at a relatively faster rate (e.g., Cyr et al., 2010; DiBiase et al., 2010; Harkins et al., 2007; Hilley and Arrowsmith, 2008; Kirby and Whipple, 2001; Ouimet et al., 2009; Safran et al., 2005; Snyder et al., 2000; Wobus et al., 2006), although it cannot be used to determine exact erosion rates without knowledge of K, m and n. The SPM has also provided a great deal of insight into how transient landscapes might evolve (e.g., Attal et al., 2011; Crosby et al., 2007; Ferrier et al., 2013; Gasparini and Whipple, 2014; Han et al., 2014; Hoke et al., 2007; Oskin and Burbank, 2007; Tucker and Whipple, 2002; Whipple and Tucker, 2002; Willenbring et al., 2013a,b).

6. Discussion

In summary, previous sections present examples that, although UQ is a valuable—even essential—component of an operational-modeling endeavor, it does not necessarily represent the most useful way to evaluate an exploratory model, or interpret predictions arising from it. Even models in Earth-surface science that are exploratory or address large-scale phenomena can make robust predictions, useful in applied and scientific contexts. Such predictions may be qualitative, of the form 'If forcing X increases, the landscape response will involve an increase (or decrease) in variable Y'. In such context the UQ approach, which is so essential for simulation/operational modeling contexts, is less relevant.

6.1. Limitations to model validation

The challenges in applying UQ to some models in Earth-surface science revolve around the steps in the approach related to model 'validation', which provide estimates of the model form uncertainty. When positive feedbacks and consequent autogenic dynamics and emergent events limit the time horizon over which deterministic prediction makes sense, a modified UQ approach could still be useful. Numerical models could be used to make statistical forecasts of relevant variables (van der Wegen and Jaffe, 2013), and the forecasts could be compared to observed statistical distributions in experimental or field observations, in an approach analogous to the deterministic model validation (e.g. Wang et al., 2011). Such an approach to quantifying model form uncertainty requires observational data sets that go beyond single snap shots in time—data sets that extend over spatial or temporal scales that are large compared to the characteristic timescales of the autogenic dynamics.

In cases where models address landscape dynamics over space and time scales that are large relative to the available observational data sets, such as the example in Section 5.2, the opportunity to assess model form uncertainty does not exist. In addition, when models address space and time scales that are large compared to the scales addressed by available parameterizations, as in the examples in Sections 5.1 and 5.2, parameterizations and parameter values can lose a direct correspondence to measurable quantities. In such cases, choosing parameter values typically involves tuning the values to produce reasonable model output. This empirical parameter tuning leads to underestimates of model-form uncertainty (e.g. Roy and Oberkampf, 2011).

In principle, uncertainty related to unknown model parameter values could be separated from model form uncertainty. The distribution of the possible values of poorly constrained parameters could be considered equally likely over some wide range (e.g. Roy and Oberkampf, 2011), and model form uncertainty could then be evaluated as a separate step. Perhaps the emphasis should shift in this direction in more modeling endeavors addressing large scale landscape phenomena. However, the UQ approach described in Section 2 does not apply directly to the way modeling of large-scale phenomena is often approached currently.

Exploratory modeling endeavors often arise when addressing phenomena on scales larger than those for which available parameterizations are appropriate (Murray, 2003, 2013; Werner, 1999). Testing whether the interactions represented by the parameterizations in such an exploratory modeling context correspond to those in the natural system is required. However, the most appropriate tests to address such questions do not necessarily correspond to the model validation procedure appropriate for simulation models. When quantitative assessments of model form uncertainty do not make sense—whether because of the exploratory goal of a modeling endeavor, or because of a lack of observations and/or parameterizations appropriate for the scales of interest—UQ is at best incomplete. When model form uncertainty is unknown but presumed to be large, assessing the uncertainty in model predictions arising from uncertainty in model inputs and parameter values provides information about model sensitivity, but not about how correct the model or its predictions might be.

The two examples in Section 5 make clear that applying a UQ analysis does not necessarily tell us much about how the model relates to the natural system, when model parameters are not easily physically interpretable, or when they are part of intentionally simplified representations of hypothesized processes and interactions. Quantitative predictions arising from such models need to be evaluated with care. However, both of the classes of models in these examples have been potentially useful in understanding the processes and interactions behind widespread, large-scale patterns on the Earth's surface. Evaluating to what degree the model interactions capture the essential features of interactions in the natural systems requires testing models in strategic ways (e.g., Murray, 2007; Tucker, 2009). However, UQ in the sense discussed here is not necessarily useful in such an evaluation of a model positioned toward the exploratory end of the modeling spectrum.

6.2. Uncertainty reduction by model inter comparisons

Exploratory model inter-comparisons, however, could be quite useful in evaluating the relationship between exploratory models and their prototypes. If different models addressing overlapping questions feature contrasting processes or interactions, or represent those processes or interactions in contrasting ways, testing which model makes the most accurate predictions can produce insights about how the natural system really works (e.g., Attal et al., 2011). The predictions involved may consist of quantitatively comparing the magnitude of model variables to measured values, or they may be robust predictions of qualitative trends.

On the other hand, if different models addressing overlapping questions that feature contrasting interactions or representations produce similar predictions (either of magnitudes or qualitative trends), we can also gain insights from model inter-comparisons (e.g., Kirwan et al., 2010). In such a case, the models may all involve fundamentally similar feedbacks that can then offer the most basic explanations for observed (or predicted) phenomena. Landscape evolution models of fluvial incision nicely illustrate this case. As discussed in Section 5.2, the SPM model results in a power–law relationship between channel slope and drainage area under steady, uniform conditions. Whipple and Tucker (2002) illustrated that this steady-state power–law relationship is also predicted by other, more complicated models of bedrock incision processes and sediment transport. As such, if steady-state conditions under uniform forcing apply, the details of more complicated processes may not be necessary to understand trends in the morphology of the channels.

6.3. Model refinement decreases epistemic uncertainty

Exploratory models often selectively leave out processes to focus on the feedbacks and processes of interest. In the parlance of UQ, exploratory models can be thought of as maximizing epistemic model uncertainty to gain insight (by isolating key feedbacks). Strategic testing or evaluation of exploratory models often yields an increased quantitative understanding of the processes and interactions responsible for poorly understood phenomena, or the importance of phenomena that are omitted. This can feed back into the modeling processes, allowing for the development of additional or refined parameterizations, and refinement of the exploratory model as a whole—in essence, a reduction of the epistemic uncertainty in the original model by making parameterizations increasingly empirically constrained (i.e. realistic). As this progression occurs, a model that was intentionally simplified can become useful for making quantitative predictions (e.g. Murray, 2007; Werner, 1999).

The progressive refinement of an originally exploratory model can either involve increasing the quantitative accuracy of parameterizations of interactions at relatively large scales (scales commensurate with the phenomena of interest), or it can involve the addition of explicit representations of processes at smaller scales. An example of this latter trend was presented in Section 5.1; The 'sorted bedform' model has been refined from its original exploratory model by replacing the original parameterization that lumped together the effects of multiple relatively small scale processes with a suite of the best available representations of those processes. However even with such refinement of the process descriptions, several significant issues hinder the conversion of the exploratory model into a simulation model. Considerable epistemic uncertainty still exists regarding: (1) how to parameterize sub-grid processes for a context in which available parameterizations cannot be interpreted literally; and (2) how to combine process descriptions with mismatched scales. Both of these issues are soluble based on data at larger scales that does not currently exist (an epistemic issue). As these epistemic uncertainties are reduced, UQ can become increasingly relevant.

Whereas strict UQ does not exist in landscape evolution modeling, refinement and/or comparison of process-based parameterizations does. LEMs are often used to explore how hypothesized processes may be evident in the landscape, or similarly, what hypothesized processes are likely dominating the evolution of a particular landscape. In the former, results from LEMs using contrasting process models guide exploration of real landscapes. For example, in contrast to their steady-state observations discussed above, Whipple and Tucker (2002) identified that non-steady-state channels, in which the process limiting the evolution of the bed is the transport of sediment, have a very different morphology from those in which bed evolution is controlled by detachment and removal of material according to the stream-power model. With this knowledge, an Earth scientist can explore the morphology of real channels in which a known-perturbation has occurred, and based on the modeling predictions, have a better understanding of the dominant process controlling long-term evolution of the channels. Further exploration of process models has led to predictions of the magnitude of perturbation necessary for different fluvial processes to be illustrated in the landscape (e.g., Crosby et al., 2007) or where to look in a landscape for the signature of particular fluvial processes (e.g., Gasparini et al., 2007).

Attal et al. (2011) offered an example of modeling specifically designed to explore the processes controlling a particular landscape. In their example, the real landscape that they were studying was subject to a well-constrained perturbation (constrained in both magnitude and timing), and they had, relative to many Earth-science studies, a large number of observations of the nature and morphology of the channels that they were modeling. They contrasted the behavior of two process models, and within each model they varied different parameters. However they did not explore the full distribution of parameters as would be called for under formal UQ. They quantitatively compared some aspects of the modeled system with the real system, although the modeled and real channel profiles (plots of elevation vs. distance) were compared qualitatively. The criterion for model success did not involve exactly recreating the currently observed morphology. Such a goal seems inappropriate when known variables that cannot be quantified through time are intentionally left out of the model system. Rather, Attal et al. (2011) sought to reproduce large-scale trends similar to those observed in the natural system without recreating it exactly. Despite the somewhat qualitative nature of the comparison between the modeled and real landscapes, in our opinion this study is extremely useful because it highlights which processes and variables likely play a role in shaping this landscape, and therefore suggest they may warrant further exploration.

The example of Attal et al. (2011) illustrates why strict UQ has probably not been done in many cases with LEMs. Limited data exist from real landscapes, and how to link the data measured at small scales in the field with the resolution of the model can be challenging. The expectation of exactly reproducing a landscape is unrealistic when, for example, each storm that drove landscape evolution, or the exact details of the uplift history, can never be known over the time-scales of landscape evolution. Further, many process models do not even parameterize variables such as individual storms. Although UQ has ways to deal with these challenges, because actually recreating the landscape is not the goal, this may be why UQ has not caught on in the LEM world. Yet, more methods for quantitatively comparing real and numerically evolved landscapes will be developed over time as more tools are developed from other parts of the Earth-science community that allow for quantification of many of the unknowns in landscape evolution.

For example, over the past few decades, methods were developed for estimating erosion rates using the concentration of cosmogenic radionuclides in quartz, both in situ and in detrital sediment (e.g., Lal, 1991; Granger et al., 1996; Von Blanckenburg, 2005). Erosion rate data from numerous locations are now available (greater than 1000 measurements) (Portenga and Bierman, 2011; Willenbring et al., 2013a,b). These data offer immense potential for more quantitative comparisons between landscape evolution models and real systems, but as of yet they have been virtually untapped by the landscape evolution modeling world (e.g. Schaller and Ehlers, 2006; Willenbring et al., 2013a,b).

6.4. Combining efforts

Exploratory and simulation models start from a different perspective. Exploratory models aim to capture dynamics by prescribing hypothesized governing interactions, whereas simulation models traditionally assume that inclusion of accepted essential physical processes lead to adequate predictions. Many models include aspects of both end member model types, striving towards a description that is simple enough to capture the essential dynamics, with a sufficient degree of quantitative accuracy, but not too complex to lose credibility. Given that computational resources are available, simulation models may play an important role to test the validity of exploratory models by systematic sensitivity analysis (in which processes are systematically included or excluded).

We mention the potentially large value of a comparison between results of an exploratory model and a simulation model for the same case. Similar results will provide a physically sound justification for exploratory models and point to the governing processes in the range of complicated process interactions of simulation models. In addition, uncertainty levels of the simulation model can thus be coupled to a more exploratory approach.

Combining simulation models and exploratory models is a recent promising development. This implies that parts of a simulation modeling domain become represented by exploratory models to limit computational effort while maintaining a certain amount of reasonable physics. An example is a coastal inlet model covering the foreshore and inlet at a 100 m grid size, but having an exploratory model attached to the inlet representing the tidal basin. Although combining these different types of models will have advantages it will definitely raise questions related to uncertainty levels.

6.5. A spectrum of models and uncertainty

Although this discussion has focused on the end members of exploratory and simulation models, many models in Earth surface science fall along a spectrum between these end members. Thus, the limitations in applying UQ procedures to exploratory models may inhibit UQ application to many models falling between the end members. The example in Section 5.1 involves a model with many aspects of a simulation model, including a representation of processes and interactions using state of the art parameterizations for the small-scale processes at play. However, applying a UQ procedure in that case would reveal more about the model than it would about what to expect for the natural system. The intent of UQ is to inform us about the range of outcomes to expect for some natural system. The examples described here suggest that applying the same UQ approaches that work well for simulation end members to models along the spectrum may or may not help us learn about what to expect in nature. Because of the wide range in model purposes and types, an associated spectrum of uncertainty evaluation approaches would be useful.

Acknowledgments

We thank Christopher Roy and three anonymous reviewers for extremely useful comments on earlier versions of this paper, and the Editor and Guest Editor for insightful suggestions.

References

Anders, A.M., Roe, G.H., Montgomery, D.R., Hallet, B., 2008. Influence of precipitation phase on the form of mountain ranges. Geology 36, 479–482. http://dx.doi.org/10.1130/g24821a.1.

Attal, M., Cowie, P., Whittaker, A., Hobley, D., Tucker, G., Roberts, G., 2011. Testing fluvial erosion models using the transient response of bedrock rivers to tectonic forcing in the Apennines, Italy. J. Geophys. Res.: Earth Surf., 116.

Attal, M., Tucker, G.E., Whittaker, A.C., Cowie, P.A., Roberts, G.P., 2008. Modeling fluvial incision and transient landscape evolution: Influence of dynamic channel adjustment. J. Geophys. Res.-Earth Surf., 113. http://dx.doi.org/10.1029/2007jf000893.

Aydin, A., Basu, A., 2005. The Schmidt hammer in rock material characterization. Eng. Geol. 81 (1), 1–14.

Bolaños, R., Thorne, P.D., Wolf, J., 2012. Comparison of measurements and models of bed stress, bedforms and suspended sediments under combined currents and waves. Coast. Eng. 62, 19–30.

Braun, J., Heimsath, A.M., Chappell, J., 2001. Sediment transport mechanisms on soil-mantled hillslopes. Geology 29, 683–686.

Cayocca, F., 2001. Long-term morphological modeling of a tidal inlet: the Arcachon Basin, France. Coast. Eng. 42, 115–142. http://dx.doi.org/10.1016/S0378 − 3839(00)00053 − 3.

Coco, G., Murray, A.B., Green, M.O., 2007a. Sorted bedforms as self-organized patterns: 1. Model development. J. Geophys. Res. 112, F03015. http://dx.doi.org/10.1029/2006JF000665.

Coco, G., Murray, A.B., Green, M.O., Hume, T., Thieler, R., 2007b. Sorted bedforms as self-organizing patterns. Part 2: complex simulations. J. Geophys. Res. Earth Surf. 112, F03016. http://dx.doi.org/10.1029/2006JF000666.

Coco, G., Zhou, Z., van Maanen, B., Olabarrieta, M., Tinoco, R., Townend, I., 2013. Morphodynamics of tidal networks: advances and challenges. Mar. Geol. 346, 1–16. http://dx.doi.org/10.1016/j.margeo.2013.08.005.

Cohen, S., Willgoose, G., Svoray, T., Hancock, G., Sela, S., 2015. The effects of sediment transport, weathering, and aeolian mechanisms on soil evolution. J. Geophys. Res. Earth Surf. 120, 260–274. http://dx.doi.org/10.1002/2014JF003186.

Colberg, J.S., Anders, A.M., 2014. Numerical modeling of spatially-variable precipitation and passive margin escarpment evolution. Geomorphology *207*, 203–212.

Collins, D.B.G., Bras, R.L., Tucker, G.E., 2004. Modeling the effects of vegetation–erosion coupling on landscape evolution. Journal. Geophys. Res.: Earth Surf. (2003–2012) 109 (F3).

Crosby, B.T., Whipple, K.X., Gasparini, N.M., Wobus, C.W., 2007. Formation of fluvial hanging valleys: theory and simulation. J. Geophys. Res.: Earth Surf. (2003–2012) 112 (F3).

Cummings, D.I., Dumas, S., Dalrymple, R.W., 2009. Fine-grained versus coarse-grained wave ripples generated experimentally under large-scale oscillatory flow. J. Sediment. Res. 79, 83–93.

Cyr, A.J., Granger, D.E., Olivetti, V., Molin, P., 2010. Quantifying rock uplift rates using channel steepness and cosmogenic nuclide–determined erosion rates: examples from northern and southern Italy. Lithosphere *2* (3), 188–198.

Davies, A.G., Thorne, P.D., 2005. Modeling and measurement of sed- iment transport by waves in the vortex ripple regime. J. Geophys. Res. 110, C05017. http://dx.doi.org/10.1029/2004JC002468.

Davy, P., Lague, D., 2009. Fluvial erosion/transport equation of landscape evolution models revisited. J. Geophys. Res.: Earth Surf. (2003–2012) 114 (F3).

DiBiase, R.A., Whipple, K.X., Heimsath, A.M., Ouimet, W.B., 2010. Landscape form and millennial erosion rates in the San Gabriel Mountains, CA. Earth Planet. Sci. Lett. 289 (1), 134–144.

Dietrich, W.E., Bellugi, D.G., Sklar, L.S., Stock, J.D., Heimsath, A.M., Roering, J.J., 2003. Geomorphic transport laws for predicting landscape form and dynamics. Predict. Geomorphol., 103–132.

Elias, E.P.L., van der Spek, A.J.F., 2006. Long-term morphodynamic evolution of the Texel Inlet and its ebb-tidal delta (the Netherlands). Mar. Geol. 225, 5–21. http://dx.doi.org/10.1016/j.margeo.2005.09.008.

Ellis, M.A., Densmore, A.L., Anderson, R.S., 1999. Development of mountainous topography in the Basin Ranges, USA. Basin Res. 11, 21–41. http://dx.doi.org/10.1046/j.1365 − 2117.1999.00087.x.

Ferrier, K.L., Huppert, K.L., Perron, J.T., 2013. Climatic control of bedrock river incision. Nature 496 (7444), 206–209.

Ferson, S., Oberkampf, W.L., Ginzburg, L., 2008. Model validation and predictive capability for the thermal challenge problem. Comput. Methods Appl. Mech. Eng. 197 (29), 2408–2430.

Ganju, N.K., Schoellhamer, D.H., Jaffe, B.E., 2009. Hindcasting of decadal-timescale estuarine bathymetric change with a tidal-timescale model. J. Geophys. Res. 114, F04019. http://dx.doi.org/10.1029/2008JF001191.

Gasparini, N.M., Whipple, K.X., Bras, R.L., 2007. Predictions of steady state and

transient landscape morphology using sediment-flux-dependent river incision models. J. Geophys. Res.: Earth Surf. (2003–2012) 112 (F3).
Gasparini, N.M., Brandon, M.T., 2011. A generalized power law approximation for fluvial incision of bedrock channels. J. Geophys. Res.: Earth Surf. (2003–2012) 116 (F2).
Gasparini, N.M., Whipple, K.X., 2014. Diagnosing climatic and tectonic controls on topography: eastern flank of the northern Bolivian Andes. Lithosphere 6, 230–250. http://dx.doi.org/10.1130/L322.1.
Granger, D.E., Kirchner, J.W., Finkel, R., 1996. Spatially averaged long-term erosion rates measured from in situ-produced cosmogenic nuclides in alluvial sediment. J. Geol., 249–257.
Green, M.O., Black, K.P., 1999. Suspended-sediment reference concentration under waves: field observations and critical analysis of two predictive models. Coast. Eng. 38, 115–141.
Green, M.O., Vincent, C.E., Trembanis, A.C., 2004. Suspension of coarse and fine sand on a wave-dominated shoreface, with implications for the development of rippled scour depressions. Cont. Shelf Res. 24, 317–335.
Goldstein, E.B., Coco, G., Murray, A.B., 2013. Prediction of wave ripple characteristics using genetic programming. Cont. Shelf Res. 71, 1–15.
Goldstein, E.B., Coco, G., Murray, A.B., Green, M.O., 2014. Data-driven components in a model of inner-shelf sorted bedforms: a new hybrid model. Earth Surf. Dynam. 2, 67–82. http://dx.doi.org/10.5194/esurf-2 – 67 – 2014.
Hancock, G.R., Willgoose, G.R., Evans., K.G., 2002. Testing of the SIBERIA landscape evolution model using the Tin Camp Creek, Northern Territory, Australia, field catchment. Earth Surf. Process. Landf. 27 (2), 125–143.
Han, J., Gasparini, N.M., Johnson, J.P., Murphy, B.P., 2014. Modeling the influence of rainfall gradients on discharge, bedrock erodibility, and river profile evolution, with application to the Big Island, Hawai'i. J. Geophys. Res.: Earth Surf. 119, 1418–1480.
Harkins, N., Kirby, E., Heimsath, A., Robinson, R., Reiser, U., 2007. Transient fluvial incision in the headwaters of the Yellow River, northeastern Tibet, China. J. Geophys. Res.: Earth Surf. (2003–2012) 112 (F3).
Hilley, G.E., Arrowsmith, J.R., 2008. Geomorphic response to uplift along the Dragon's Back pressure ridge, Carrizo Plain, California. Geology *36* (5), 367–370.
Hoke, G.D., Isacks, B.L., Jordan, T.E., Blanco, N., Tomlinson, A.J., Ramezani, J., 2007. Geomorphic evidence for post-10 Ma uplift of the western flank of the central Andes 18 30′–22 S. Tectonics 26, 5.
Howard, A.D., 1994. A detachment-limited model of drainage basin evolution. Water Resour. Res. 30 (7), 2261–2285.
Howard, A.D., 1997. Badland morphology and evolution: interpretation using a simulation model. Earth Surf. Process. Landf. 22, 211–227.
Howard, A.D., 1999. 11 Simulation of gully erosion and bistable landforms. In: Darby, S.E., Simon, A. (Eds.), Incised River Channels: Processes, Forms, Engineering and Management. Wiley, Chichester, ISBN 0 – 471 – 984469 – 9.
Huang, X., Niemann, J.D., 2014. Simulating the impacts of small convective storms and channel transmission losses on gully evolution. Rev. Eng. Geol. 22, 131–145.
Kirby, E., Whipple, K., 2001. Quantifying differential rock-uplift rates via stream profile analysis. Geology 29 (5), 415–418.
Kirby, E., Whipple, K.X., 2012. Expression of active tectonics in erosional landscapes. J. Struct. Geol. 44, 54–75.
Kirwan, M.L., Guntenspergen, G.R., D'Alpaos, A., Morris, J.T., Mudd, S.M., Temmerman, S., 2010. Limits on the adaptability of coastal marshes to rising sea level. Geophys. Res. Lett. 37, L23401. http://dx.doi.org/10.1029/2010GL045489.
Lal, D., 1991. Cosmic ray labeling of erosion surfaces: in situ nuclide production rates and erosion models. Earth Planet. Sci. Lett. 104 (2), 424–439.
Lamb, M.P., Dietrich, W.E., Sklar, L.S., 2008. A model for fluvial bedrock incision by impacting suspended and bed load sediment. J. Geophys. Res.: Earth Surf. (2003–2012) 113 (F3).
Moglen, G.E., Bras, R.L., 1995. The effect of spatial heterogeneities on geomorphic expression in a model of basin evolution. Water Resour. Res. 31 (10), 2613–2623.
Murray, A.B., 2003. Contrasting the goals, strategies and predictions associated with simplified numerical models and detailed simulations. In Prediction in Geomorphology. In: Wilcock, P.R., Iverson, R.M. (Eds.), Geophysical Monograph Series. American Geophysical Union, Washington, DC, pp. 151–165.
Murray, A.B., 2007. Reducing model complexity for explanation and prediction. Geomorphology 90 (3–4), 178–191.
Murray, A.B., Lazarus, E., Ashton, A., Baas, A., Coco, G., Coulthard, T., Fonstad, M., Haff, P.K., McNamara, D., Paola, C., Reinhardt, L., 2008. Geomorphology, complexity, and the emerging science of the Earth's surface. Gemorphology 103, 496–505.
Murray, A.B., 2013. Which models are good (enough), and when?. In: Shroder, J., Editor in Chief, Baas, A.C. (Eds.), Treatise on Geomorphology vol. 2. Academic Press, San Diego, pp. 50–58.
Murray, A.B., Coco, G., Goldstein, E.B., 2014. Cause and effect in geomorphic systems: complex-systems perspectives. Syst.: Complex-Syst. Perspect., Geomorphol. 214, 1–9. http://dx.doi.org/10.1016/j.geomorph.2014.03.001.
Murray, A.B., Thieler, E.R., 2004. A new hypothesis and exploratory model for the formation of large-scale inner-shelf sediment sorting and "rippled scour depressions". Cont. Shelf Res. 24, 295–315.
Murray, A.B., Coco, G., Green, M., Hume, T., Thieler, R., 2005. Different approaches to modeling inner-shelf 'sorted bedforms,' 4th International symposium on River, Coastal, and Estuarine Morphodynamics, Champaign-Urbana, Oct. 2005, 1009-1015.
Oberkampf, W.L., DeLand, S.M., Rutherford, B.M., Diegert, K.V., Alvin, K.F., 2002. Error and uncertainty in modeling and simulation. Reliab. Eng. Syst. Saf. *75* (3), 333–357.
O'Hara Murray, R.B., Thorne, P.D., Hodgson, D.M., 2011. Intrawave observations of sediment entrainment processes above sand ripples under irregular waves. J. Geophys. Res. 116 (C01001). http://dx.doi.org/10.1029/2010JC006216 2011.
Oskin, M.E., Burbank, D., 2007. Transient landscape evolution of basement-cored uplifts: example of the Kyrgyz Range, Tian Shan. J. Geophys. Res.: Earth Surf. (2003–2012) 112 (F3).
Ouimet, W.B., Whipple, K.X., Granger, D.E., 2009. Beyond threshold hillslopes: Channel adjustment to base-level fall in tectonically active mountain ranges. Geology 37 (7), 579–582.
Pelletier, J.D., 2010. Minimizing the grid-resolution dependence of flow-routing algorithms for geomorphic applications. Geomorphology 122 (1), 91–98.
Pelletier, J.D., McGuire, L.A., Ash, J.L., Engelder, T.M., Hill, L.E., Leroy, K.W., Chorover, J., 2011. Calibration and testing of upland hillslope evolution models in a dated landscape: Banco Bonito, New Mexico. Journal. Geophys. Res.: Earth Surf. 116 (F4).
Pelletier, J.D., DeLong, S.B., Orem, C.A., Becerra, P., Compton, K., Gressett, K., Spinler, J.C., 2012. How do vegetation bands form in dry lands? Insights from numerical modeling and field studies in southern Nevada, USA. J. Geophys. Res.: Earth Surf. 117 (F4).
Peng, M., Xie, L., Pietrafesa, L.J., 2004. A numerical study of storm surge and inundation in the Croatan-Albemarle-Pamlico Estuary System. Estuar. Coast. Shelf Sci. 59, 121–137. http://dx.doi.org/10.1016/j.ecss.2003.07.010.
Portenga, E.W., Bierman, P.R., 2011. Understanding Earth's eroding surface with 10 Be. GSA Today 21 (8), 4–10.
Roe, G.H., Montgomery, D.R., Hallet, B., 2003. Orographic precipitation and the relief of mountain ranges. J. Geophys. Res.-Solid Earth 108. http://dx.doi.org/10.1029/2001jb001521.
Roering, J., 2008. How well can hillslope evolution models "explain" topography? Simulating soil transport and production with high-resolution topographic data. Geol. Soc. Am. Bull. 120, 1248–1262. http://dx.doi.org/10.1130/b26283.1.
Rosenbloom, N.A., Anderson, R.S., 1994. Hillslope and channel evolution in a marine terraced landscape, Santa Cruz, California. J. Geophys. Res.-Solid Earth 99, 14013–14029.
Roy, C.J., Oberkampf, W.L., 2011. A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. Comput. Methods Appl. Mech. Engrg 200, 2131–2144.
Saco, P.M., Moreno-de las Heras, M., 2013. Ecogeomorphic coevolution of semiarid hillslopes: emergence of banded and striped vegetation patterns through interaction of biotic and abiotic processes. Water Resour. Res. *49* (1), 115–126.
Safran, E.B., Bierman, P.R., Aalto, R., Dunne, T., Whipple, K.X., Caffee, M., 2005. Erosion rates driven by channel network incision in the Bolivian Andes. Earth Surf. Process. Landf. 30 (8), 1007–1024.
Schaller, M., Ehlers, T.A., 2006. Limits to quantifying climate driven changes in denudation rates with cosmogenic radionuclides. Earth Planet. Sci. Lett. 248 (1), 153–167.
Schoorl, J.M., Sonneveld, M.P.W., Veldkamp, A., 2000. Three-dimensional landscape process modelling: the effect of DEM resolution. Earth Surf. Process. Landf. 25 (9), 1025–1034.
Selby, M.J., 1980. A rock mass strength classification for geomorphic purposes: with tests from Antarctica and New Zealand. Z. Geomorphol. 24 (1), 31–51.
Sklar, L.S., Dietrich, W.E., 2004. A mechanistic model for river incision into bedrock by saltating bed load. Water Resour. Res. 40, 6.
Snyder, N.P., Whipple, K.X., Tucker, G.E., Merritts, D.J., 2000. Landscape response to tectonic forcing: Digital elevation model analysis of stream profiles in the Mendocino triple junction region, northern California. Geol. Soc. Am. Bull. 112 (8), 1250–1263.
Stock, J.D., Montgomery, D.R., 1999. Geologic constraints on bedrock river incision using the stream power law. J. Geophys. Res.: Solid Earth (1978–2012) 104 (B3), 4983–4993.
Tomkin, J.H., Brandon, M.T., Pazzaglia, F.J., Barbour, J.R., Willett, S.D., 2003. Quantitative testing of bedrock incision models for the Clearwater River, NW Washington State. J. Geophys. Res.: Solid Earth (1978–2012) 108 (B6).
Tucker, G., Slingerland, R., 1997. Drainage basin responses to climate change. Water Resour. Res. 33, 2031–2047.
Tucker, G.E., Slingerland, R., 1996. Predicting sediment flux from fold and thrust belts. Basin Res. 8, 329–349.
Tucker, G.E., Bras, R.L., 1998. Hillslope processes, drainage density, and landscape morphology. Water Resour. Res. 34 (10), 2751–2764.
Tucker, G.E., Whipple, K.X., 2002. Topographic outcomes predicted by stream erosion models: Sensitivity analysis and intermodel comparison. J. Geophys. Res.: Solid Earth (1978–2012) 107 (B9), ETG-1.
Tucker, G.E., 2004. Drainage basin sensitivity to tectonic and climatic forcing: Implications of a stochastic model for the role of entrainment and erosion thresholds. Earth Surf. Process. Landf. 29 (2), 185–205.
Tucker, G.E., 2009. Natural experiments in landscape evolution. Earth Surf. Process. Landf. 34, 1450–1460. http://dx.doi.org/10.1002/esp.1833.
van der Beek, P., Bishop, P., 2003. Cenozoic river profile development in the Upper Lachlan catchment (SE Australia) as a test of quantitative fluvial incision models. J. Geophys. Res.: Solid Earth (1978–2012) 108 (B6).
van der Beek, P., Braun, Jean, 1998. Numerical modelling of landscape evolution on geological time-scales: a parameter analysis and comparison with the southeastern highlands of Australia. Basin Res. 10 (1), 49–68.
van der Wegen, M., Jaffe, B.E., Roelvink, J.A., 2011. Process-based, morphodynamic hindcast of decadal deposition patterns in San Pablo Bay, California, 1856–1887. J. Geophys. Res.: Earth Surf. (2003–2012) 116 (F2). http://dx.doi.org/10.1029/

2009JF001614.
van der Wegen, M., Roelvink, J.A., 2012. Reproduction of estuarine bathymetry by means of a process-based model: Western Scheldt case study, the Netherlands. Geomorphology 179, 152–167. http://dx.doi.org/10.1016/j.geomorph.2012.08.007.
van der Wegen, M., Jaffe, B.E., 2013. Towards a probabilistic assessment of process-based, morphodynamic models. Coast. Eng. 75, 52–63. http://dx.doi.org/10.1016/j.coastaleng.2013.01.009.
Dam, G., van der Wegen, M., Roelvink, D., 2013. Long-term performance of process-based models in estuaries. Proc. Coast. Dyn. 2013, 409–420.
van der Werf, J.J., Doucette, J.S., Donoghue, T., Ribberink, J.S., 2007. Detailed measurements of velocities and suspended sand concentrations over full-scale ripples in regular oscillatory flow. J. Geophys. Res. 112, F02012. http://dx.doi.org/10.1029/2006JF000614.
van Leeuwen, S.M., van der Vegt, M., de Swart, H.E., 2003. Morphodynamics of ebb-tidal deltas: a model approach. Estuar. Coast. Shelf Sci. 57, 899–907. http://dx.doi.org/10.1016/S0272−7714(02)00420−1.
Von Blanckenburg, F., 2005. The control mechanisms of erosion and weathering at basin scale from cosmogenic nuclides in river sediment. Earth Planet. Sci. Lett. 237 (3), 462–479.
Wang, Y., Straub, K.M., Hajek, E.A., 2011. Scale-dependent compensational stacking: an estimate of autogenic time scales in channelized sedimentary deposits. Geology 39 (9), 811–814.
Werner, B.T., 1999. Complexity in Natural Landform Patterns. Science 284, 102–104. http://dx.doi.org/10.1126/science.284.5411.102.
Whipple, K.X., Tucker, G.E., 1999. Dynamics of the stream-power river incision model: Implications for height limits of mountain ranges, landscape response timescales, and research needs. J. Geophys. Res.: Solid Earth (1978–2012) 104 (B8), 17661–17674.
Whipple, K.X., Tucker, G.E., 2002. Implications of sediment-flux-dependent river incision models for landscape evolution. J. Geophys. Res.: Solid Earth (1978–2012) 107 (B2), ETG-3.
Willenbring, J.K., Codilean, A.T., McElroy, B., 2013a. Earth is (mostly) flat: apportionment of the flux of continental sediment over millennial time scales. Geology 41 (3), 343–346.
Willenbring, J.K., Gasparini, N.M., Crosby, B.T., Brocard, G., 2013b. What does a mean mean? The temporal evolution of detrital cosmogenic denudation rates in a transient landscape. Geology 41 (12), 1215–1218.
Willgoose, G., 1994. A statistic for testing the elevation characteristics of landscape simulation models. J. Geophys. Res.-Solid Earth 99, 13987–13996.
Wobus, C., Whipple, K.X., Kirby, E., Snyder, N., Johnson, J., Spyropolou, K., Sheehan, D., 2006. Tectonics from topography: Procedures, promise, and pitfalls. Geol. Soc. Am. Spec. Pap. 398, 55–74.
Yanites, B.J., Ehlers, T.A., 2012. Global climate and tectonic controls on the denudation of glaciated mountains. Earth Planet. Sci. Lett. 325, 63–75. http://dx.doi.org/10.1016/j.epsl.2012.01.030.
Yetemen, O., Istanbulluoglu, E., Flores-Cervantes, J.H., Vivoni, E.R., Bras, R.L., 2015. Ecohydrologic role of solar radiation on landscape evolution. Water Resour. Res. 51, 1127–1157. http://dx.doi.org/10.1002/2014WR016169.