Developing, choosing and using landscape evolution models to inform field-based landscape reconstruction studies

Citation for published version:

Digital Object Identifier (DOI):
10.1002/esp.4162

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Earth Surface Processes and Landforms

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Developing, choosing and using landscape evolution models to inform field-based landscape reconstruction studies

A.J.A.M. Temme\textsuperscript{1,2}, J. Armitage\textsuperscript{3}, M. Attal\textsuperscript{4}, W. van Gorp\textsuperscript{5}, T.J. Coulthard\textsuperscript{6}, J.M. Schoorl\textsuperscript{7}

1: Department of Geography, Kansas State University, Manhattan, Kansas, United States
2: Institute for Arctic and Alpine Research (INSTAAR), University of Colorado (Boulder), Colorado, United States
3: Dynamique des Fluides Géologiques, Univ Paris Diderot, Sorbonne Paris Cité, Institut de Physique du Globe de Paris, Paris, France
4: School of Geosciences, University of Edinburgh, Drummond Street, Edinburgh United Kingdom
5: Groningen Institute of Archaeology, University of Groningen, Groningen, the Netherlands
6: School of Environmental Sciences, University of Hull, Hull, United Kingdom
7: Soil Geography and Landscape, Wageningen University, the Netherlands

Abstract

Landscape evolution models (LEMs) are an increasingly popular resource for geomorphologists as they can operate as virtual laboratories where the implications of hypotheses about processes over human to geological timescales can be visualized at spatial scales from catchments to mountain ranges. Hypothetical studies for idealised landscapes have dominated, although model testing in real landscapes has also been undertaken. So far however, numerical landscape evolution models have rarely been used to aid field-based reconstructions of the geomorphic evolution of actual landscapes. To help make this use more common, we review numerical landscape evolution models from the point of view of model use in field reconstruction studies. We first give a broad overview of the main assumptions and choices made in many LEMs to help prospective users select models appropriate to their field situation. We then summarize for various timescales which data are typically available and which models are appropriate. Finally, we provide guidance on how to set up a model study as a function of available data and the type of research question.

Keywords: landscape evolution model, landscape reconstruction, calibration and validation, complexity

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/esp.4162

This article is protected by copyright. All rights reserved.
Introduction

Since the advent of computer modelling, landscape evolution models (LEMs) have been used to address a wide range of questions in geomorphology, often focusing on gaining explorative understanding of the processes that lead to the development of patterns in landscapes (e.g. Huggett, 1975; Kirkby, 1977; Ahnert, 1977; Willgoose et al., 1991; Tucker and Slingerland, 1994; Rodriguez-Iturbe and Rinaldo, 1997); see review by Tucker and Hancock (2010). Their main attraction is the possibility to illustrate and quantify the results of processes acting and interacting over large spatial and temporal scales (Coulthard, 2001) that are hard or impossible to replicate in physical laboratories. In that sense, LEMs have become the laboratories of choice (Tucker and Hancock, 2010) where landscape-scale hypotheses can be tested. Befitting this role, most LEM studies have been of exploratory nature, using hypothetical landscapes to test the effect of variations in boundary conditions, processes representations or parameter values on landscape morphology (e.g. Armstrong, 1982; Istanbulluoglu and Bras, 2005; Perron and Fagherazzi, 2012; Temme et al., 2013b). The discipline has benefited enormously from this work, perhaps mostly so where the effects of small-scale non-linear dynamics at the large-scale landscape level are concerned (Coulthard et al., 1998; Coulthard et al., 2000; Kirkby et al., 2000; Castellotort and Van Den Driessche, 2003; Wainwright, 2006; Van De Wiel and Coulthard, 2010; Jerolmack and Paola, 2010; Simpson and Castellotort, 2012).

As our understanding of geomorphic systems improved, a second use of landscape evolution models has increased as well: model testing against field data (e.g. Kirkby, 1984; Champel et al., 2002; van der Beek et al., 2002; Tomkin et al., 2003; van der Beek and Bishop, 2003; Braun and van der Beek, 2004; Loget et al., 2006; Attal et al., 2008b; Attal et al., 2011; Hancock et al., 2011; Forzoni et al., 2014; Hancock et al., 2015). Attention has been given to the selection of landscapes ("natural experiments") where testing is optimally possible (Tucker, 2009). Such landscapes ideally are subject to variation in only one (exogenous or endogenous) controlling factor, with other factors stable or uniform. However, as landscape evolution is a product of physical laws and the geographical, geological and environmental conditions of specific locations, it is possible that landscape evolution is by its own nature unique (Phillips, 2014). Such a hypothesis would make the goal of landscape evolution modelling particularly challenging.

In this study, however, we will particularly focus on a third category of LEM studies: those where models are used to simulate (i.e. post-dict or retro-dict) the partially known or observed evolution of landscapes, with the objective of supporting otherwise field-based landscape reconstruction efforts (Temme et al., 2013a). Landscape reconstruction is used here to denote attempts to explain how a
landscape has evolved to its present state. Landscape reconstruction lies at the interface of
gеомorphology and geology, and is sometimes called historical geomorphology (Rhoads, 2006). In
many cases, the landscapes of interest in these studies are not suitable for model testing because of
complicated tectonic, climatic or human histories, or large spatial variation in boundary conditions. In
fact, reconstruction is often of interest precisely because landscapes are not ideal natural experiments –
they are to some extent historically contingent (or badass, Phillips, 2014). The question remains what
LEMs can offer in these situations.

To answer this question, we suggest it is helpful to focus on landscape observables. For instance, it is
clear that stratigraphy in sedimentary basins contains information about landscape evolution, boundary
conditions (tectonics, climate) and their evolution through time (Bridgland and Westaway, 2012); yet,
sedimentologists may be relatively unaware of the corresponding postdictive capabilities of LEMs and
landscape evolution modellers may have limited knowledge of field observations of stratigraphy. To use
LEMs not only for insights in the basic mechanisms that control catchment form and evolution (Tucker
and Hancock, 2010), but also to explore possible pathways of evolution or to test hypotheses, we argue
that LEM studies should better connect to field reconstruction data. For fluvial systems, this involves
reconstruction of stratigraphical sequences, both at the broader scale of incision-aggradation phases,
but also at the scale of stratigraphic composition. The current state of science is that some models (see
below) can reconstruct these broader scale patterns, but tend to exclude measurable properties such as
grain size. Models that do include grain-size distribution tend to be limited to a 1-D profile, or data
requirements and model runtime limit their use for long-term studies. The capabilities, types, concepts
and possible uses of different LEMs have until now not been very well conveyed for potential model
users from outside the modelling community.

Therefore, our aim in this contribution is to assess how existing LEMs can support field-based landscape
reconstruction studies of landforms and stratigraphy, with a focus on fluvially shaped landscapes, and to
inform model choice and use by geomorphologists who think of themselves as field-oriented. To this
end, we first briefly review some main approaches and assumptions underlying LEMs. This is to help
users to select appropriate models based on knowledge of the models. Second, for timescales ranging
from orogenic to historical, we give an overview of data that are typically available and models that are
typically used. This aids users to select appropriate models based on their data situation and timescale.
Third, we provide guidance regarding the setup of model studies as a function of the kind of question
that needs to be answered, and the kind of data that are available.
Common properties of LEMs

LEMs are empirical

All LEMs are ultimately (partly) based on empirical measurements. This is because of practical and computational reasons: it is not feasible to both measure the properties, and calculate the actions and interactions of individual particles as a result of rainfall impact or flowing water at the spatial scale of landscapes (although such an approach can be used for small reaches of braided rivers - see Brasington and Richards, 2007). Unfortunately, we often do not understand the basic processes at larger spatial scale levels. In fact, it is likely that processes differ between larger scale levels as well (Reinhardt et al., 2010) – which requires different process formulations or parameter values for models operating at different scales. As an approximation, many models use proportionality parameters to empirically link a driving force to a result. For instance, models often link a proxy of the impact of water flow to the amount of eroded material using a proportionality parameter. This parameter then needs to be estimated. Estimates differ between case studies because the underlying, unresolved or unknown, processes cause different outcomes in different settings (Beven and Binley, 1992). Importantly, this empirical nature of LEMs requires calibration and sensitivity analysis in applications to real-world case study settings.

A diverse array of numerical models and model frameworks have been developed that approach the question of landscape evolution from different angles. Some models have been used in different versions adapted to different case study settings, such as the many versions of the Channel-Hillslope Integrated Landscape Development (CHILD) or LandscApe ProcesS modelling at mUlti-dimensions and Scales (LAPSUS) models. This means that potential users have to choose the optimal model and model version for their question, and typically that deliberation with model developers is necessary. This model choice depends strongly on the spatio-temporal scale of the research question – and hence on the type of observables that are available.

Digital landscapes

Most LEMs work with digital landscapes that are changed by geomorphic processes as simulated time progresses. These landscapes can broadly be divided into two categories (Figure 1): the well-known grid-based Digital Elevation Models (DEMs) and less well-known node-based Triangular Irregular Networks (TINs). The former are conceptually relatively simple, but only have a maximum of eight flow directions from each cell, while the latter can have any flow direction and can increase resolution in highly variable parts of a landscape. LEMs such as CAESAR (Coulthard et al., 1997), LAPSUS (Schoorl et al., 2000; Schoorl et al., 2002) and SIBERIA (Willgoose et al., 1991) use regular gridded DEMs as input,
whereas LEMs such as CASCADE (Braun and Sambridge, 1997) and CHILD (Tucker et al., 2001) use irregular grids as surface input. Conversion between DEMs and TINs is relatively simple, so that the availability of digital landscapes does not limit model options for prospective users.

Next to their altitude, nodes in a TIN or cells in a regular DEM typically hold additional information on properties that are affected by the simulated geomorphic processes such as stream width, the depth from surface to bedrock or grain size. For instance, a LEM that simulates the speed of flowing water based on Manning’s equation, may need to keep track of channel width in every cell that belongs to a drainage network (Attal et al., 2008a). Many models, capitalizing on increased computing power, go further and keep track of such types of information for a finite number of layers under the surface. For instance, a LEM may keep track of the median grain size ($D_{50}$) for several layers under the surface to simulate the formation, destruction and effects of bed armouring (Coulthard and Van De Wiel, 2007).

The growing class of soilscape-landscape evolution models, recently reviewed in (Minasny et al., 2015) keeps track of a range of soil properties for multiple soil layers to simulate profile development (Temme and Vanwalleghem, 2016). Selecting a model that keeps track of subsurface layers is key if connection to the spatial pattern of stratigraphic observables is desired.

**Process descriptions**

In this section, we review how processes are typically described in LEMs, focusing on the properties that are relevant to model choice (for an extensive review of processes in fluvial LEMs, see Tucker and Hancock, 2010). All LEMs begin with the conservation of mass. We begin with a simple idealized landscape composed of bedrock, of thickness $\eta$ (m), and a regolith, of thickness $h$ (m) that is not subdivided into internal layers (Figure 2). This landscape is forced externally through uplift or subsidence, $U$ (m a$^{-1}$). Bedrock is transferred into regolith at a rate $P$ (m a$^{-1}$), and sediment is transported across the system with a sediment flux, $q_s$ (m$^2$ a$^{-1}$). Assuming that the density of sediment produced and transported is equal to the density of bedrock, the rate of change in bedrock thickness is,

$$\partial_t \eta = U - P \tag{1}$$

and the rate of change in sediment thickness is,

$$\partial_t h = P - \partial_s q_s \tag{2}$$

It then follows that the rate of change in landscape elevation is the sum of the two rates of change,

$$\partial_t z = \partial_t \eta + \partial_t h \tag{3}$$
In getting to equation (3), we have assumed that the density of the sediment is equal to the density of the bedrock. This assumption is not true and many models allow relaxing it. When discussing various LEMs and their approach to solving equation (3), we will however try to place them into this simple framework so that their level of complexity can be assessed.

General equations for erosion, transport and deposition of sediment

In a LEM, denudation, the loss of mass from the surface, is due to the transport of mass down-slope. If the transport capacity for sediment is finite and it is assumed that there is always a supply of transportable material, then the system is defined as transport-limited, that is, the amount of erosion that can take place is limited by the transport capacity of the surface processes. Furthermore, ignoring changes in density, equation 3 becomes a form of the Exner equation (Smith and Bretherton, 1972; Paola et al., 1992a):

\[
\partial_t z = U - \partial_x q_s
\]  

(4)

Sediment flux is then usually modelled as a function of overland flow of water and local slope, i.e.

\[
\partial_t z = U - \partial_x (v(q_w) \partial_x z)
\]  

(5)

where \(v\) is a diffusion coefficient which can be a function of surface water flux, \(q_w\), and is not necessarily constant in space or time. The assumption that there is always a large supply of transportable material is more likely true when considering very long (Ma) timescales – but on smaller timescales, temporary lack of transportable material can occur, in particular in mixed bedrock-alluvial rivers. In that case, the system is defined as ‘detachment-limited’ as it is the detachment of mass from the bedrock that limits erosion (Howard and Kerby, 1983). Within the framework of equations 1 to 3, we can formalize this assumption by stating that \(\partial_t h \sim 0\), that is, the thickness of transportable material does not change with time, or that \(h \sim h\) everywhere, which is equivalent to exposed bedrock. This assumption means that there is no deposition, and is therefore only useful if deposition is not an important part of the landscape dynamics that need to be reconstructed. What remains is to find a relationship for the production of sediment, \(P\).

Based upon observed relationships between upstream area, water flux and the power of that water to detach rock of the base of an active channel, the popular stream power equation has been derived as an expression for \(P\) (Howard, 1994):

\[
\partial_t z = U - k \partial_x z^n A^m
\]  

(6)
where $k$ is a constant that is a function of lithology and climate. Analysis of real landscapes suggests the $m/n$ ratio (or concavity index) ranges between 0.3 and 0.7 (e.g. Sklar and Dietrich, 1998; Whipple and Tucker, 1999; Tucker and Whipple, 2002). No consensus exists at present as to which $m$ and $n$ exponents are best suited to simulate fluvial erosion (Lague, 2014). Hence, the values of these parameters are often adapted to the case study, usually through calibration (see below).

The stream power law (Eq. (6)), strictly applies only within the stream bed, and therefore landscape models typically add a linear diffusion term to the right-hand side of equation 6 to capture the provision of sediment from slopes outside of the river system, giving the following:

$$\partial_t z = U + k a^m (\partial_z^n + \kappa d^2 z)$$

where $\kappa$ is a diffusion constant. The diffusion equation simulates a group of processes that includes soil creep, due to various forms of bio- and cryoturbation. Below, in the section on hillslope processes, we will discuss how hillslope processes can be subdivided into component processes.

It is clear that both the transport of sediment and the detachment of bedrock can occur at the same time. To reflect this, various models treat the Earth’s surface as a two-layer system, with a layer of transportable sediment and erodible bedrock below. These hybrid models typically limit the rate of vertical erosion to the lesser of detachment capacity (equation (6)) or surplus transport capacity (equation (5), Tucker and Whipple, 2002; Whipple and Tucker, 2002). More complex hybrid models allow the amount of sediment in transport to be either more or less than transport capacity, depending on conditions (Davy and Lague, 2009; Schoorl et al., 2014) – or model specific fluvial processes in more detail (see below).

The stream power equation needs a positive value for downstream slope ($d_z$) to calculate erosion, and is therefore in essence unable to deal with lakes or depressions in the landscape (where slope is locally zero). Various models have found ways out of this problem (e.g. CAESAR (Coulthard et al., 1998; Coulthard et al., 2013) and LAPSUS (Temme et al., 2006)). This is of relevance to field studies where lakes significantly influence river incision and long term stream profile evolution, for instance due to landslide damming (e.g. Korup, 2012) or lava damming (Van Gorp et al., this issue).

Erosion is not only driven by the flow of water but also the flow of ice. In high mountain ranges or at high latitudes, abrasion and quarrying will act beneath the glacier. These two processes are linked to the sliding rate of the glacier, and so are typically formulated as a simple power law (e.g. MacGregor et al., 2000):

$$\partial_t z = U - c_1 u^2$$

This article is protected by copyright. All rights reserved.
where $u$ is the basal sliding velocity, and $c_1$ and $c_2$ are constants. The values for both constants are poorly constrained. Theoretical considerations for the effects of abrasion alone would suggest $c_2 \sim 2$ (Hallet, 1979), however various numerical studies have set $c_2 = 1$ (e.g. MacGregor et al., 2000; Herman and Braun, 2008; Egholm et al., 2009). The value of $c_1$ is typically used as a proportionality parameter to give an erosion rate that is in line with the available observations (e.g. Egholm et al., 2009). The sliding velocity is then assumed to be a function of the pressure exerted by the ice load above and the local slope, following from the shallow ice approximation (e.g. Braun et al., 1998; MacGregor et al., 2000). Equation (8) coupled with the solution to the flow of the ice layer under the shallow ice approximation then allows for the inclusion of glacial erosion within LEMs. This simple erosion model is however a first order approximation and ignores the effects of chemical dissolution and fluvial erosion due to melt water. Furthermore, recent work has suggested equation (8) is inadequate in capturing the full effects of quarrying beneath glaciers (Ugelvig et al., 2016).

Removing the spatial domain – Spatially lumped models

When modelling ancient sedimentary systems, the exact hydrological and spatial details of the catchment are not well known. Therefore, it is worth briefly describing a set of models that treat landscape in less than three dimensions. These models such as Hydrotrend (Kettner and Syvitski, 2008), PaCMod (Forzoni et al., 2013), LIMTER/TERRACE (e.g. Viveen et al., 2014a) or FLUVER2 (Tebbens et al., 2000; Veldkamp and Van Dijke, 2000) lump the study area into a finite element 1D river profile or into 1D-2D catchment properties such as catchment area and average slope. Erosion is then modelled as a slope dependent process (Eq. (6)), but with a single average slope for the whole catchment. Sediment flux is typically calculated using various forms of the BQART model (Syvitski and Milliman, 2007), where the flux of sediment is a simple function of catchment area and the spatially averaged water flux, rock type and temperature.

The mathematical advantage of this type of models is that the partial differential equation used to solve Equation (3) becomes an ordinary differential equation, as there is no spatial derivative. Therefore, these models can run very rapidly with temporal inputs of i) high resolution such as precipitation rate time series or ii) very long extent. Furthermore, where the properties of the catchment are unknown, it is arguable that such a simple model that requires no input DEM is more appropriate to explore past climate and tectonics. The disadvantage is that there is no information about the evolution of the 3D morphology as the landscape evolves and therefore model outcomes are less spatially explicit. In cases where the stratigraphy in a downstream basin gives detailed information about erosion history, but not much is known about the catchment that has produced this sediment, 1D or 2D models may be the best choice.
Processes on hillslopes

As mentioned above, a diffusion equation (Eq. (7)) is often used to lump hillslope processes. Depending on the timescale and objective of study, several studies have used equations describing the individual hillslope processes rather than lumping all the processes into one term: soil creep, solifluction, landsliding and soil tillage.

Soil creep and soil tillage are usually described with the linear diffusion equation (Eq. (7)). Solifluction, the movement of (nearly) saturated soil over a frozen or otherwise hard surface, may be specifically of interest in case study sites where slopes have experienced or are currently experiencing periglacial conditions. Following suggestions that solifluction is functionally equivalent to creep accelerated by wetness (Matsuoka, 2001), a suitable formulation may be, somewhat analogous to Temme and Veldkamp (2009):

\[ \frac{\partial z_{\text{solifluction}}}{\partial t} = k \frac{\partial^2 z}{\partial x^2} (\theta - \theta_{\text{ref}}) \quad \text{for} \quad \theta \geq \theta_{\text{ref}} \tag{9} \]

where diffusion is driven by soil wetness \( \theta \) [-] higher than a threshold wetness \( \theta_{\text{ref}} \) [-].

Landsliding is probably the process that is most commonly separated from the linear diffusion equation, reflecting the perceived important role of rapid failure of (undercut) steep slopes in landscape formation. Approaches to landsliding in landscape evolution, both for shallow and deep-seated types, include factor-of-safety approaches (Claessens et al., 2007) or the use of a non-linear diffusion equation including a critical slope (e.g. Martin and Church, 1997; Martin, 2000; Coulthard et al., 2000; Champel et al., 2002; Goren et al., 2014). This version of the diffusion equation is easily constrained with field data (Roering et al., 2001). The choice whether to select a model that includes landsliding as a process is best made after considering the steepness of the landscape that is being studied.

Sedimentary signals produced by rivers

Two potential advantages are to be gained by resolving fluvial processes in more detail than in the generalized stream power equation approach described above. The first is a better representation of water-flow dynamics, which can be important when rivers flood and affect wide areas, or more generally when topographic differences are small relative to differences in water flow height (see Tucker and Hancock, 2010). In fluvial landscape evolution models (e.g. CAESAR), temporal resolution is often dynamic and can be as low as seconds, to allow simulation of flow heights and flow speeds during floods. This better represents interactions between water, floodplain and banks, and potentially allows for simulation of lateral erosion and widening of both channels and valleys (Venditti et al., 2014; Langston et al., 2015) – something that traditional stream power equation-based models described above cannot
simulate. Reflecting the more detailed representation of hydrology, stream power equations used in such models are also more detailed.

The second advantage is more attention for downstream and overbank grain size and hence for grain size distributions of deposits. This can be a crucial advantage for field-based reconstructions where stratigraphy is the main type of observation. Grain size sorting down-system on a gross scale is described by the Sternberg relationship (Sternberg, 1875):

$$D = D_0 e^{-\alpha x}$$  \hspace{1cm} (10)

where \(D\) is a characteristic grain size at a downstream distance \(x\), \(D_0\) is the initial characteristic grain size and \(\alpha\) is the fining rate (typically expressed in km\(^{-1}\)). This empirical relationship has been found to be a good approximation at a basic level for many modern and ancient alluvial river systems. The fining rate \(\alpha\) is a function of many fluvial variables and can vary over orders of magnitude, from \(10^{-3}\) to \(10^0\) km\(^{-1}\) (e.g. Moussavi-Harami et al., 2004; Whittaker et al., 2011). The processes that control fining are (1) selective transport (e.g. Parker, 1991; Paola et al., 1992b; Hoey and Ferguson, 1994; Fedele and Paola, 2007), (2) abrasion (e.g. Parker, 1991; Attal and Lavé, 2009), and (3) spatial and temporal variation in the sediment supply (e.g. Pizzuto, 1995; Whittaker et al., 2010; Allen et al., 2015; Attal, 2015; Attal et al., 2015).

The hydraulics of grain segregation within an individual channel are highly complex and resolved only in few LEMs and for timescales less than \(10^5\) a (e.g. CAESAR; Coulthard et al., 1997 and CHILD; Gasparini et al, 2004). Such models may be the best choice when studying river systems that do not display downstream fining (e.g. Heller et al, 2001, Attal and Lavé, 2006, Sklar et al, 2006, Whittaker et al, 2010, Menting et al, 2015). For longer timescales, selective transport is typically approached through empirical or semi-empirical methods such as the Sternberg relationship (e.g. Snow and Slingerland, 1987; Parker, 1991; van Niekerk et al., 1992; Fedele and Paola, 2007). This empirical modelling approach to down-system sorting requires quantification of three controlling parameters.

First, the grain size distribution and volume of source sediment plays a major role on the down-system variation in grain size deposited, particularly for larger gravel clasts (Fedele and Paola, 2007). Changes in the input distribution cause changes in the fining rate (Sklar et al., 2006; Duller et al., 2010; Armitage et al., 2011; Allen et al., 2015). Second, the distribution of accommodation space along the river, and third, the hydraulics, topography and grain size interactions, can either be solved for as a coupled system where the available grain sizes inform the transport properties of the flow and deposition (e.g. Robinson and Slingerland, 1998; Granjeon, 1999; Clevis et al., 2004, Van de Wiel and Coulthard, 2007). While there is some theoretical justification for the basic equations that must be...
solved in the coupled system over geological timescales (>10^6 a) (e.g. Paola et al., 1992a; Marr et al., 2000; Smith, 2010), a drawback is that the multiple parameters and coefficients required as model input are not known for preserved stratigraphy and must be estimated to match our expectations of how the system evolves.

A second more simplistic approach is to solve for the sediment transport using a single equation (such as Eq. 5) and then to sort the grains using a second set of equations (e.g. Paola et al., 1992a; Armitage et al., 2011). Here the number of parameters for the sorting model are reduced to the input grain-size mean, variance and a small number of heuristic/empirical coefficients. However, these simple model suffer from the same basic problem as the Sternberg model, i.e., they only apply for transport down a 1-D line. Application of such a simple model therefore requires the sedimentary system to be well defined as a single sediment routing system (see Allen et al., 2013). Increased complexity in the observed system (strong convergence or divergence of channels) cannot be captured with such simple models, and would require a choice for a more complex numerical model.

Tectonics

Erosion and deposition are primarily controlled by the creation of elevated landscapes that can be eroded and intervening basins that can be filled. In most models, tectonics is treated as a displacement to the digital landscape’s surface. Models that have been developed to explore the response to change in catchment uplift over a period that is greater than a million years typically assume that uplift is a continuous function of time with a predefined functional form (see for example Tucker et al., 2001; Clevis et al., 2004; Pepin et al., 2010; Armitage et al., 2011). Some models account for the flexural response of the lithosphere to the loading and unloading resulting from deposition and erosion over large areas (e.g. Willett, 1999; Beaumont et al., 2001; Braun and van der Beek, 2004; Simpson, 2006). Another approach to model long-term landscape evolution is to couple a model of fault growth to that of surface erosion and deposition (e.g. Cowie et al., 2006a, Braun et al., 2013, Collignon et al., 2014).

As the temporal extent of interest decreases, uplift due to tectonics can sometimes not be simplified as a continuous function of time. If recognizable faults exist in the study area, in theory each individual slip event can be modelled. Over a timescale of days to hundreds of years, the cumulative effects of such individual slips on a fault relative to continuous uplift have been explored (see Coulthard and Van de Wiel, 2013). Yet these two approaches to tectonics are applied at opposite ends of the geological timescale of interest: millions of years and hundreds of years.
Complexity

The descriptions of the processes acting across landscapes given in the previous sections are all individually straightforward. However, when these are applied together and allowed to change over time and space (e.g. when applied to a DEM) their interaction becomes complex. At its most basic level, the feedback between erosive and depositional processes altering the surface of the model (its DEM or TIN), in turn altering the magnitude of erosion and deposition, leads to the development or emergence of complex landscape behaviours (Phillips, 1999; Phillips, 2003). For example, fluvial erosion and deposition can construct an alluvial fan where a channel joins a larger valley. But we cannot describe the formation of the alluvial fan from the erosive and depositional processes alone; instead it emerges from the interactions between these processes and the landscape. These complex process interactions can be expressed in terms of emergent landforms and features such as meander migration, river braiding, slope processes, as well as in the irregularity and complexity of fluxes or yields of sediment from landscapes. This is one of the most powerful and compelling aspects of landscape evolution modelling, that simple processes interacting over a grid or TIN can generate the patterns and complexity that we observe in reality.

However, the complexity and non-linearity of a landscape’s behaviour has revealed how hard it may be for us to use field stratigraphy and landscape reconstruction studies to determine what has happened in drainage basins in the past. The irregular and non-linear behaviour in landscapes is evident from field studies and laboratory studies. For example, non-linear changes in sediment delivery have been observed in natural channels (Cudden and Hoey, 2003; Croke et al, 2015) and laboratory experiments (Sapozhnikov and Foufoula-Georgiou, 1997; Gomez and Phillips, 1999). In all these examples, the changes in sediment delivery occurred with no external forcings, meaning that the internal feedbacks and processes operating within a channel alone were capable of generating changing sediment outputs. This internal or autogenic behaviour has important implications because observations in the field may only be due to the internal (autogenic) processes and re-organisation rather than due to any external forcing. Alternatively, the autogenic processes may be acting to modify, process or change any signals from external forcings.

Indeed, previous modelling research has shown how drainage basins can alter or buffer external inputs (Métivier, 1999; Castelltort and Van Den Driessche, 2003), and more recent research has suggested that (geologically) short term storage in floodplains may ‘shred’ any upstream signals of forcings (Jerolmack and Paola, 2010). Coulthard and Van de Wiel (2013) showed how tectonic signals within a basin were all but removed by a buffering section of floodplain, yet climate induced signals were visible. Other modelling work has found evidence of self-organised criticality in landscape evolution model.
simulations (Coulthard and Van De Wiel, 2007; Van De Wiel and Coulthard, 2010) whereby outputs from modelled basins show no relationship to any external drivers.

Therefore, modelling studies inform us that we have to be careful when inverting sedimentary records to imply past drivers (e.g. climate or land use) as there is a high likelihood that any signal within a basin will have been modified heavily by the autogenic processes or may represent no external forcing at all. In many cases, it will make sense to report on the (possibly wide) range of past drivers that may have resulted in the observed landscape or stratigraphy.

Models for various scales and observations

The explanation above about the various process formulations and the assumptions underlying them should help workers decide which kind of model is appropriate to their reconstruction setting. As such, it should clarify the demand side for models and modelling expertise. Below, we coarsely subdivide landscape reconstruction efforts by their timescale of interest, to frame a closer look at available observables and models (hence illuminating the supply side, Table 1). These timescales range from orogenic timescales (loosely defined as > $10^6$ a) to (pre-)history (typically $10^3$ a or less).

Orogenic timescales (> $10^6$ a)

When modelling over the timescales of the formation and destruction of mountain ranges, the initial topography is usually not known. Therefore, initial conditions for the numerical model are typically the steady state topography for a given set of boundary conditions (e.g. Attal et al., 2008b; Attal et al., 2011), the present day topography (e.g. Herman and Braun, 2006), or a reconstructed topography based on remnants in the landscapes (terraces, relict surfaces, e.g. van der Beek and Bishop, 2003; Braun and van der Beek, 2004). Models will typically forecast the evolution of the topography due to the interactions between tectonics and erosion; they will also calculate and record quantities that derive from these interactions, namely erosion rates and sediment fluxes across the study area. Models typically output a time evolution of erosion rates and patterns, topography and/or sediment fluxes. Some models also forecast the consequences of that sediment delivery in the depositional domain. Landscape evolution models therefore represent ideal tools to interrogate hypotheses for the formation of stratigraphic sequences, changes in grain size deposited, and/or the response of topography to tectonic and climatic change. However, especially at these large timescales with weakly constrained boundary conditions, it is crucial to assess whether there is a range of model inputs (such as uplift history) that produces statistically indistinguishable model outputs (such as an observed stratigraphy).
LEM-independent erosion rates across natural landscapes are obtained through a range of empirical methods, including thermochronology, cosmogenic nuclides and the dating of fluvial terraces; similarly, sediment fluxes are typically derived from integrating the erosion rates for the area of interest, over the timescale of interest (e.g., in rapidly eroding landscapes, cosmogenic \(^{10}\)Be and fission track thermochronology will typically provide rates that integrate thousands of years and millions of years, respectively). These quantities, their spatial distribution and their temporal evolution can be directly compared to model output. In the depositional realm, observed histories of sediment accumulation can be compared with predictions of sediment flux out of the eroding catchment or sediment routing system if the effect of hiatuses is appropriately considered. At a higher resolution, well logs and stratigraphic sections can be compared against the model output (e.g., models such as DIONISOS (Ganjeon, 1999), or the model published in Clevis et al. (2003) and Armitage et al. (2016)).

The topography of the modern landscape, as well as past landscapes (reconstructed through the dating of remnants such as fluvial terraces), can be compared to the topography forecast by the model at various stages of the experiment. It is important to note that even in reconstruction studies that focus on the history of a particular area (van der Beek and Bishop, 2003; Attal et al., 2008a; Attal et al., 2011), trying to replicate the exact morphology or depositional architecture is impossible. Instead, it is necessary to focus on metrics that can characterize the patterns in the landscape or stratigraphic record as a whole, such as valley spacing (e.g. Perron et al., 2009; Perron and Fagherazzi, 2012) or clustering of sedimentary bodies (e.g. Hajek et al., 2010; Flood and Hampson, 2014) (Figure 3); note that such metrics can also be used at greater spatial or temporal resolution, in studies that do not focus on a particular existing landscape. Other higher-level topographic metrics that can be used include relief, hypsometry, drainage density (Istanbulluoglu and Bras, 2005), and channel steepness via river profile analysis (van der Beek and Bishop, 2005; Attal et al., 2008b; Attal et al., 2011).

**Glacial-interglacial cycles (~10^4 – 10^6 a)**

At glacial-interglacial timescales, many of the observables that characterise landscape evolution over longer timescales are still useful, particularly erosion rates and volumes, and patterns and changes in profile gradient. Glacial-interglacial climatic variations can be reflected in phases of high and low geomorphic activity and fluvial terrace formation under a background uplift or sea-level fluctuation regime. At these timescales, models can demonstrate spatiotemporally different responses of regional fluvial systems to interacting climate and tectonic regimes (e.g. Tebbens et al., 2000; Veldkamp and Van Dijke, 2000; Simpson and Castelltort, 2012) and the growth and decay in glacier volume (e.g.
Pedersen and Egholm, 2013), and help test field-derived fluvial terrace reconstructions (Viveen et al., 2013; Viveen et al., 2014a). Furthermore, the response of small catchments to climatic regimes can be modelled, for instance by testing different base level regimes (van Gorp et al., 2015, this issue). Modelled outputs are often erosion and sediment output volumes, or the evolution of fluvial profile elevation and form. Stratigraphical output is typically not simulated at these timescales, but the models simulate possible timing and locations of when and where to find preserved records and landforms. Furthermore, in spatially explicit models, these timescales allow a closer connection to complex variability in sediment output. For instance, observed phases of catchment sediment yield can be either connected to external controls or to internal sediment storage and release phases (van Gorp et al., 2015). Landscape response at these timescales will more often lag changes in external drivers, which models can demonstrate (Moon et al., 2015).

As discussed above, spatially-lumped models can also be used at these timescales to explore how change in catchment forcing may be recorded in the sediment flux to a basin (e.g. Forzoni et al., 2013; Forzoni et al., 2014). While this type of model cannot simulate the geomorphology of the catchment, or terrace formation, the predictions of sediment flux can then be sorted down the depositional system using for example the fluvio-deltaic model DeltaSim (Forzoni et al., 2014). In this case the interplay between time-dependent processes within the eroding landscape and the depositional system can be explored. This may be a useful method for exploring the mechanisms behind the formation of stratal architecture where catchment properties are unknown.

Holocene (~10^4 a)

Including more processes explicitly becomes more important when looking at Holocene (~ 10^4 a) timescales. This is because over these relatively short timescales a greater amount of information about the catchment and its history typically remains preserved. In contemporary temperate climates, record of periglacial processes is often preserved in the earliest Holocene deposits, and natural processes can give way to human processes as the dominant factor determining landscape change in populated regions (Govers et al, 1996). After recognizing and describing important processes from the field situation, simulation of their interaction can explain complex landscape patterns, driven by high or low activity of different processes due to climatic change (Temme et al., 2009a). At this timescale, pollen records provide information about palaeo-vegetation which is sometimes simulated explicitly (Coulthard and Macklin, 2001; Collins et al., 2004; Istanbulluoglu, 2009; Saco and Moreno-De Las Heras, 2013). Pollen also can provide inferred precipitation and temperature estimates which can be used as model inputs. If knowledge of extreme events is available, these may become important observables that can constrain the range of suitable models. For instance, the annual temporal resolution of models such as LAPSUS
and SIBERIA makes it difficult to use them to study the (highly non-linear) effect of short, extreme
rainstorms (e.g. Baartman et al., 2013, Hancock et al., 2015). In those cases, models such as CAESAR
or CHILD may be better suited.

(Pre)history (<10\(^3\) a)

One main difference between studies at Holocene timescale and those focussing on reconstructing
landscapes at (pre)historical timescales, is the availability of instrumented and measured data to both
set up and to calibrate/validate models. For example, archaeological evidence can provide information
about the age of deposits, and about land use patterns which can be translated into maps (De Brue and
Verstraeten, 2014). For the historical period, maps play an important role in providing information about
changing land use and changing landscape infrastructure (such as roads, treelines and lynchets) on the
one hand, and landscape response on the other hand (particularly planform changes such as the
locations of channels). Other documentary evidence such as government records may give information
about timing and extent of extreme events such as landslides, floods and mudflows (e.g. Gruner, 2012;
Stedinger and Griffis, 2008). Models can also be validated using instrumented records of sediment
output and sequential aerial photography or satellite images to show channel pattern changes.

As these studies move into the timescale of humans and human influence, the importance of land use
changes as a driver of (on- or off-site) erosion and deposition patterns increases. The increase of human
use of the land has diachronously influenced sediment movement in different parts of the world since
Middle to Late Holocene times (Ellis et al., 2013). Thus the direct and indirect effects of e.g.
deforestation, burning, terracing, damming and tillage become important in LEM studies. Furthermore,
the shorter timescale of historical scale simulations suggests that higher resolution temporal data is
available to drive these simulations, such as precipitation data. Certainly, over shorter timescales, single
large events may have a greater relative impact and it could be argued that these should be better
represented. This may also be through better representation of certain processes such as river
hydrodynamics (Coulthard et al. 2013). Recent studies have shown how there may be inbuilt
sensitivities to the temporal and spatial resolution of precipitation data used to drive LEMs (Coulthard
and Skinner, 2016).

At very short timescales, landscape evolution modelling starts to overlap the realm of erosion modelling
(e.g. Merritt et al., 2003), in which hydrologic processes are sometimes better represented but
parametrisation of erosion models is even more case-study specific than for LEMs and has therefore
limited explanatory value on the long term. Some studies have attempted to bridge this gap while
maintaining a stream power equation approach (Debolini et al., 2015) while others used a more complete description of the fluvial process (Coulthard et al., 2012).

At historical timescales, LEMs start to have a more practical application for tackling environmental issues. For example, the SIBERIA and the CAESAR/CAESAR-Lisflood models have both been used to assess uranium mine stability in Northern Australia (Hancock et al. 2010; Hancock et al. 2011; Hancock et al. 2000; Hancock et al. 2002).

How to best use field data

One of the goals of the FACSIMILE community, from which ideas for this paper originated, is to find optimal ways in which field reconstruction and landscape evolution modelling can support each other. A first way to approach this question is by realizing that field reconstruction delivers a more or less detailed view of a landscape at one or more points in time, whereas modelling delivers process-based constrains on possible pathways between these points in time. Although field observations are always uncertain to some extent, more of them from different points in space and time can further constrain the number of simulated pathways (Figure 4).

There are many kinds of questions that can be answered using LEMs (e.g. Larsen et al., 2014). Below, we list those that are most relevant to reconstruction efforts, ordered by increasing data need:

- What would happen in or to a landscape under this set of assumptions? An example is: "what can happen to hillslopes when weathering and sediment transport interact with different strengths and different starting conditions?" (Strudley et al, 2006).

- Which observation will allow us to distinguish between two competing assumptions about landscape evolution? An example is: "can thermochronology help us distinguish between two hypotheses for the topographic evolution of the escarpment of south-eastern Australia?" (Braun and van der Beek, 2004).

- Which changes can this (property of, or position in the) landscape have experienced in between the points in time for which we have field observations? An example is: "is it possible that the first order streams in the top of our eroding catchment experienced deposition during short periods of downstream damming?" (Van Gorp et al, this issue).

- Which (sets of) processes and conditions may be responsible for the landscape changes that we observe? An example is: "are increased sediment flux from a glacier, increased sediment flux from hillslopes, increased rainfall intensity, or a combination of these, required to produce the observed downstream deposition during glacial intervals?" (Langston et al, 2015).
Can our hypothesis about timing, strength and/or interaction between processes over time in this landscape be correct? An example is: "are our models of combined tillage and water erosion, which were calibrated for one landscape, able to predict topographic evolution of a second landscape?" (Temme et al, 2011b).

The first two kinds of questions are the least strict and require the smallest amount of data (Figure 5). The second two kinds of questions explore what is possible and impossible in the study area, without evaluating what actually happened. Such evaluation happens in the fifth kind of question (hypothesis testing) and requires most information about the history of the study area.

The flow of information does not have to be only from model to field. Reconstruction efforts can also fruitfully indicate possible improvements to model assumptions or formulations – one of the topics of a companion paper (Briant et al, this issue). From this, it follows that model-fieldwork iterations are successful ways of increasing our knowledge of landscape evolution.

As a general guideline, interrogation of models is facilitated by greater constraints on input data and parameters that can be directly measured. It is important to note that models may not be able to provide a definite test or answer to questions, due to uncertainties in model output as a result of uncertainty in inputs and in model formulations and assumptions. A satisfying model result will show limited uncertainty about what happens in between points in time when landscape properties are known from observations (Figure 4).

Field data can be used in various ways to answer these model questions by driving, constraining and testing models. These uses fall into two main categories (Figure 5). The first category is using field data to run a model by providing model inputs or by setting model parameters. Providing inputs may mean the setting of the initial landscape and its properties, and the setting of time series of driving factors such as uplift and rainfall. In some model literature, starting conditions and time series are jointly referred to as boundary conditions. Using field data to set model parameters (parametrization in Figure 5) is only possible when model parameters are measurable (e.g. the bulk density of sediment, depth decay functions of weathering rates, or the m and n exponents of the stream power equation as discussed above).

The second category of data use is to learn about the model. This category contains calibration and validation (Figure 5). Calibration, which is also called inverse modelling, entails the repeated running of the model with various parameter settings to find out the parameter settings that result in outputs that best match field data. Validation then evaluates how well a model performs and is hence a one-way and
one-time process. The (portion of a) dataset that is used for validation should therefore not be used for calibration. A third category, pure prediction of future landforms, is rarely performed in LEM studies (but see Hancock et al. 2000; Temme et al., 2013a). This is possibly a symptom of the difference in philosophy between 'scientific' and 'engineering models' (e.g. Ben-Zion, 2017) where the aim of scientific modelling is to understand how the system works at the simplest level, whereas the aim of engineering modelling is to predict behaviour in actual cases. The use of landscape evolution models in field reconstruction studies may well mean a shift from 'scientific' modeling to 'engineering' modelling, and may hence increase the number of studies that have prediction as their objective.

Setting starting conditions

The most important starting condition is the initial landscape, the input landscape for modelling. Often, there are few if any remnants of palaeo-landscapes to support initial landscape creation. One strategy is to model backwards (inverse modelling), thus back-calculating palaeo relief by reversed process descriptions. However, this is conceptually difficult because the present landscape could have evolved from multiple palaeo-landscapes (geomorphic equifinality; Phillips, 1997) via different processes. Even if processes are known, numerical problems arise from calculating backwards, especially with multiple interacting processes (e.g. Temme et al., 2011b).

Alternatively, modellers have used either a steady state topography for a given set of boundary conditions (e.g. Attal et al., 2008a; Attal et al., 2011) or the present day topography (e.g. Herman and Braun, 2006) as input. This is justified when evidence suggests that the topography has been in equilibrium for the time considered or was in equilibrium at the time of the perturbation, when the study focuses on the landscape response to a known perturbation (e.g. Attal et al., 2008a; Attal et al., 2011, Hancock et al., 2016).

Field reconstructions can provide information on the palaeo-landscape based on geomorphic and stratigraphical evidence (Briant et al., this issue), or 1D models have to be used (see above). A popular technique in this respect is initial landscape reconstruction from interpolated terraces which are assumed to have remained stable (e.g. van der Beek and Bishop (2003), Van Gorp, 2015). In those cases, surface interpolation is rather simplified by either using interpolation from a triangular interpolated network, or GIS hydrological interpolation tools to reconstruct former floodplains. Differences between reconstructed and current landscapes then can yield estimates of volumes of material removed which can be used for model calibration/validation. These methods are simplified and it is therefore important
to stress that the choice of the initial topography can strongly influence the results of the model and the propagation of uncertainty of the initial surface altitude to the interpolated palaeo-landscape must be assessed, as illustrated in the review by Perron and Fagherazzi (2012) and in Figure 6. If field preservation of palaeo-surfaces is high, geostatistical methods are potentially valuable to address this uncertainty and predict the non-preserved part of the palaeo-landscape with a quantitatively informed model (Geach et al., 2014). Ideally, the sensitivity of the outcomes to the choice of initial topography is tested by producing a series of runs with different but equally plausible (i.e. equiprobable; Temme et al., 2009b) topographies, based on the existing constraints. This is an example of the well-known Monte-Carlo approach. Unfortunately, there is currently no dedicated toolset available that allows easy combination of the various types of information about a palaeo-landscape into a palaeo-DEM. If not enough information exists, then 1D models may be a better choice.

In addition to the palaeo-landscape, models often accept a map of other initial properties of the landscape, such as the thickness of the regolith $h$, or relative erodibility $K$ (Schoorl et al., 2002), as an input. If a map of such properties cannot be made, a spatially uniform estimate is often used. Here as well, the effect of these assumptions on model outputs can be tested with Monte-Carlo or sensitivity analysis approaches (Temme et al., 2011a).

Preparing other model input data

At a basic level there are two external factors that drive erosion within a landscape evolution model: i) base level change (uplift or sea-level driven), which changes topography and slope, and adds (or removes) mass above sea level by creating (or reducing) relief; and ii) surface run-off, driven by precipitation, which transports sediment and redistributes mass across the landscape.

Base level change typically takes the form of a spatial distribution of displacement of the surface that is applied at each numerical time step. This uplift field can mimic slip on a fault (e.g. Densmore et al., 2007; Armitage et al., 2011), be a simple block uplift of the entire numerical domain (e.g. Tebbens and Veldkamp, 2000; Pepin et al., 2010) or a part of it (Coulthard and Van de Wiel, 2013; Viveen et al., 2014b), or can be the output of a model of the deformation of the crust and lithosphere (e.g. Cowie et al., 2006b; Herman and Braun, 2006; Collignon et al., 2014). In case of smaller tributary catchment modelling, the base level change could be based on a field-reconstructed or known incision (or aggradation) rate of the main river, which may or may not be tectonically driven (e.g. van Gorp et al., this issue). The spatial pattern of uplift or base level change has obvious consequences on the routing of surface run-off and hence model response.
Change in the distribution of surface run-off will also directly impact erosion within the LEM. At a first order, sediment flux is directly related to the flow of water through the river network (e.g. Paola et al., 1992). In the absence of any sedimentary cover, the erosion by flowing water is likewise directly related to the water flux (e.g. Howard, 1994). In both cases, an increase in water flux would lead to an increase in the transport of material out of the eroding landscape. Yet, while a change in base level leads to a permanent change in the model system, a change in run-off creates a transient adjustment to the landscape, which then subsequently recovers to a pre-perturbed state (e.g. Bonnet et al., 2003). This adjustment comes with either a change in the active tectonics due to the redistribution of mass (e.g. Willett, 1999; Simpson, 2006), or an isostatic adjustment in the absence of active tectonics (e.g. Braun et al., 2013; Armitage et al., 2014).

The complexity of precipitation rate and uplift or base level input depends on the processes that are being modelled and the quantity of information about the ancient catchment that can be gained from field studies. Model selection and setup should be targeted towards what is knowable. To demonstrate this in an exploratory (rather than reconstruction) setting, we show typical model inputs for two very different modelling approaches, PacMod and Fastscape (see Table 1): (1) the spatially lumped model PaCMod includes the possible effects of vegetation cover and freeze thaw conditions and separates suspended and bed load sediment fluxes (Forzoni et al., 2013); (2) the spatially explicit LEM Fastscape (Braun and Willett, 2013) is based on the stream power equation.

If the catchment shape is not well known, yet there are high resolution paleo-climate proxies, a spatially-lumped model may be appropriate (Figure 7a-c; Forzoni et al., 2014). However, if the aim is to understand how landscape form responds to change in forcing conditions, then a 2D surface process model based on the stream power equation (Equation 7) may be more appropriate (Figure 7d-f and Figure 8). These two models are forced with the same change in uplift. In the spatially-lumped model PaCMod, uplift is essentially uniform, and therefore a spatially uniform uplift has also been applied to the stream power model Fastscape. Fastscape has a high spatial resolution, however relative to PaCMod it has a temporal resolution that is a few orders of magnitude less, with time steps typically of 10 ka compared to hundreds of years (or less) for PaCMod. Therefore PaCMod can be forced by a high resolution change in precipitation (Figure 7b) whilst Fastscape is forced by a simple precipitation rate that is averaged over 10 ka (Figure 7e). Despite the different temporal resolutions, there are similarities in the model responses to the change in forcing, namely a gradual increase in sediment flux in response to the increase in uplift rate (Figures 7c and f). With the greater resolution of the precipitation record in PaCMod, there is a greater depth of information on the sediment flux out of the catchment in terms of suspended and bed-load despite the loss of the spatial domain (compare Figure 7c and 7f). However,
the advantage of modelling the evolution of the landscape is the ability to interpret how the landscape responds, for example in terms of waves of incision (Figure 8). This highlights an important consideration: gaining spatial accuracy typically requires loss of temporal resolution in order to have a numerical model that can solve the basic set of equations within a reasonable computational time.

**Parametrization: provision of model parameters based on independent information**

Next to providing a model with the various input data, data from the field can also be used to derive model parameter values. This can be relatively simple, such as measurements of the bulk density of new sediment versus un-eroded soils that can inform density-sensitive versions of Eq. 1, or channel width measurements that can help constrain the hydraulic scaling relationships used to compute channel width in the model domain (e.g. Attal et al., 2008a, 2011). In some cases, the derivation of parameters is more complex, such as the derivation of the m and n parameters of the stream power equation (Eq. 7) from topography (e.g. Sklar and Dietrich, 1998).

**Calibration: obtaining parameter values by matching model outputs with field data.**

In some cases, model parameters cannot be measured in the field, or derived from independent datasets. In these cases, model calibration is useful. Calibration, sometimes called inverse modelling, is the process of repeatedly running a model with different parameter values, while matching the outputs with data from the field. This process requires a quantifiable measure of difference between the model output and field data (the objective function), which needs to be minimized. The parameter value (or set of parameter values) that leads to the minimal difference between field and model, is then selected.

Calibration can be done by manually varying parameter values (Martin and Church, 2004), or by automated procedures that can save time in complex settings but require time to set up (Heuvelink et al., 2006). Calibration typically distinguishes studies where models support field reconstruction from studies where models are used in a more exploratory manner to ask 'what-if' questions about landscape development (Larsen et al., 2014).

The main complication during calibration that model users should be aware of is equifinality: the notion that different starting conditions and different pathways can lead to statistically indistinguishable model outcome. In terms of calibration, this means that multiple parameters or parameter sets can result in equally correct model outputs. If this is the case, more independent data are needed to further constrain the model parameters (Figure 4). This is illustrated by a study of the South Australian margin by Braun and van der Beek (2004): the authors found that two different, synthetic initial topographies both led to a topography resembling the real one over the duration of the experiment. This meant that two evolutions of the study area were possible, viz. plateau degradation and escarpment retreat. To solve
this, the LEM was coupled with a thermal model of the lithosphere in order to use thermochronological data to discriminate between the two scenarios.

Model validation on independent data

Model validation is the independent testing of model performance by comparing model outputs to independent data, i.e. data that were neither used as inputs nor in inverse modelling, for instance where two adjacent catchments hardly differ, or only in a given factor, e.g. lithology or their (known) uplift or climate history. Calibrating a model with data from one catchment would then allow validating it for the other catchment (e.g. Temme et al., 2011b). However, LEMs used in landscape reconstruction typically face the difficulty that no similar second catchment or landscape is available. Some other options are available. If a substantial amount of information on the landscape (such as its topography) is available for a certain point in time somewhere between beginning and end of the period of interest, then models can be calibrated on the period between start and halfway point, and subsequently validated for the period between halfway point and end point. This strategy could work well for fluvial incision studies, where dated terraces can be used to constrain the landscape at a halfway point. Alternatively, the model can be calibrated on a subset of known landscape characteristics, and validated on another subset (e.g. Sapozhnikov et al., 1998). Finally, a model can be validated on the evolution of a downscaled experimental landscape (e.g. Hancock and Willgoose, 2002).

Conclusions

Landscape evolution models can be very useful tools for geomorphologists or geologists who wish to reconstruct or further understand the history of a particular landscape. However, the wide variety in approaches and assumptions can make it difficult to choose an appropriate model. In addition, designing the optimal connection between field stratigraphy and LEMs is often not straightforward. This inspired our work. We discussed the three most important steps when considering to use LEMs for landscape reconstruction. First, it is necessary to take note of the assumptions underlying various models and modelling approaches, and evaluate which set of assumptions best fits with the understanding of landscape evolution from the field. Second, it is necessary to consider the availability of data from the field, which will among others strongly differ between timescales of interest. This places additional constraints on model selection. Thirdly, the question that is to be asked during modelling should match the data availability and the capability of the model. For any given model, the answering of stricter questions will require more data, and more modelling steps to take. Ultimately, it appears that models are available for wide ranges of field processes, timescales and data availabilities. Disclosing and listing
strengths and limitations of different models such as is done in this paper will hopefully contribute to increasing model use in the future.

Acknowledgements

We acknowledge support from the members of the FACSIMILE network, and support for the FACSIMILE network meetings from the University of Twente in the Netherlands, and from the Volkswagen Foundation in Hannover, Germany. J.A. acknowledges a research fellowship from the Royal Astronomical Society. We also acknowledge the kind suggestions of three anonymous reviewers.
Bibliography


This article is protected by copyright. All rights reserved.


**Table 1**: Typical landscape observables and example landscape evolution models as a function of the timescale of interest.

<table>
<thead>
<tr>
<th>Timescale</th>
<th>Observables typical for timescale</th>
<th>Example models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Orogenic timescales</strong> (&lt;10⁶ a)</td>
<td>Erosion rates and patterns, sediment flux, depositional grain size, patterns of deposition (e.g., fan morphology), shoreline trajectory, landscape morphology (river long profiles, relief, hypsometry, valley spacing, drainage density, etc.), detrital sediment properties (grain provenance, distribution of fission track or cosmogenic nuclide ages), terraces and ages.</td>
<td>Transport-limited models: e.g. Dionisos, or publications such as Paola <em>et al.</em>, 1992, Armitage <em>et al.</em>, 2011. Detachment-limited models: Fastscape, DAC. Mixed models: CHILD, CASCADE, or publications such as Clevis <em>et al.</em>, 2003, Pepin <em>et al.</em>, 2010</td>
</tr>
<tr>
<td><strong>The Quaternary, glacial interglacial cycles</strong> (~ 10⁴ - 10⁶ a)</td>
<td>Fluvial terraces, fan morphology, landscape morphology, river profile development, raised beaches and other shorelines, buried soils and surfaces, strata, sediment accumulation, cosmogenic nuclide ages and denudation rates, luminescence dating</td>
<td>CAESAR, FLUVER2, CHILD, SIBERIA, PACMOD, LAPSUS</td>
</tr>
<tr>
<td><strong>Holocene</strong> (~ 10⁴ a)</td>
<td>Floodplain development, dated avulsions and buried terraces, strata, cosmogenic nuclide ages and denudation rates, sediment and peat accumulation, buried periglacial features, pollen records</td>
<td>CAESAR, LAPSUS, SIBERIA, LORICA, PACMOD</td>
</tr>
<tr>
<td><strong>(Pre-)history</strong> (~ 10³ a and less)</td>
<td>Buried and truncated (agricultural and other) soils, pollen records, buried archaeological remains, sediment volumes behind dams, denudation rates (¹³⁷Cs, ²¹⁰Pb, ²³⁸Pu inventories), parametric and non-parametric documentary evidence, particularly of extreme events and land use changes.</td>
<td>CAESAR, LAPSUS, WATEM, WATEM-SEDEM, HYDROTREND</td>
</tr>
</tbody>
</table>
Figure 1: Example of a regular gridded DEM (A) and a Triangulated Irregular Network (TIN, B) of the same fluvially dissected area of 800 by 800 m. The number of cells in the DEM is approximately the same as the number of triangles in the TIN. Note that the DEM and TIN are not equally detailed in all parts of the landscape.
Figure 2: A basic hillslope. $\eta$ and $h$ are thicknesses of bedrock and regolith, respectively. The landscape is uplifted at a rate $U$ and bedrock is turned into regolith (sediment) at a rate $P$. $q_s$ is sediment flux downslope.
**Figure 3**: example of general properties of landscape and stratigraphy that are quantifiable and comparable in models and in the field. TOP: modelled topographies with different valley spacing and drainage density; each image is 500 m wide and colour represent drainage area (blue low, red high) (adapted from Perron and Fagherazzi, 2012). BOTTOM: schematic across-valley cross sections showing different patterns of clustering of channel deposits (black) resulting from different depositional histories (adapted from Hajek et al., 2010). These can be compared to observed clustering patterns, rather than to the precise location of the clusters.
**Figure 4**: conceptual diagram illustrating the connection between field observations and modelling. Field reconstruction creates snapshots in time \((i, i+j, i+j+k)\), which lead to the reconstruction of fluvial networks and regimes. Modelling can connect these reconstructions in time, but intermediate states of the landscape remain uncertain (red lines). Additional field data can constrain credible pathways (orange area).
**Figure 5:** Upper: model pentagram providing an overview of various uses of (field-derived) data in a modelling context, with example types of data. Bottom left: pentagram for using models to provide new ideas and suggestions, where input data are needed and model parametrization adds value but is not required. Bottom middle: pentagram for using models to explore what is (im)possible in a given study area. Inputs and parametrisation are needed, and model calibration adds value but is not required. Bottom right: pentagram for using models to test hypotheses about landscape development. Inputs, parametrisation and calibration are needed. Validation is possible, and needed if calibration has been successful (and the hypothesis has not yet been rejected).
Figure 6: Sensitivity of modelled landscape evolution to initial topography using the CHILD model. The modelled domain is 70 x 150 km, with top and bottom boundaries open; node spacing is ~500 m and uplift is uniform; colours indicate elevation, ranging from 0 to 4000 m (blue low, red high). On the left, initial topography was flat with random noise with an amplitude of 1 m; on the right, initial topography was sloping towards the bottom boundary with a slope of only 1 m over 70 km (S = 0.000014), with the same random noise superimposed. The resulting topographies after 1 Ma are strikingly different, with the main drainage divide being more linear and shifted towards the top in the initially sloping run.
Figure 7: Comparison of model precipitation rate input for two different models of differing spatial and temporal resolution. (a) to (c) display the input uplift, precipitation rate and sediment flux out of the model domain for the spatially lumped model PaCMod (Forzoni et al., 2014). (d) to (f) display the uplift, precipitation rate and sediment flux out for the stream power model Fastscape (Braun and Willett, 2013), where $k=10^{-5}$, $m=0.5$ and $n=1$ in equation 7. In this case the model was run for a total duration of 5.7 Ma, giving a 5 Ma wind-up duration prior to the perturbation in uplift. Time steps are 10 ka, therefore the model cannot be forced by a high resolution precipitation rate based on marine cores.
Figure 8: Predicted topography from the stream power model Fastcape run in Figure 7. The consequence of a five-fold increase in uplift rate at 0.7 My BP (Figure 7d) can be observed within this model given the high spatial resolution. The spatially lumped model PaCMod (Figure 7a-c) can predict a high resolution sediment flux but cannot provide information about the catchment response. Note the change in colour bar scale between a and b, reflecting an increase in elevation.