EARTH SURFACE PROCESSES AND LANDFORMS
Earth Surf. Process. Landforms **43**, 2964–2980 (2018)

(wileyonlinelibrary.com) DOI: 10.1002/esp.4458

Applying Pattern Oriented Sampling in current fieldwork practice to enable more effective model evaluation in fluvial landscape evolution research

Rebecca M. Briant,[1]* Kim M. Cohen,[2] Stephane Cordier,[3] Alain J.A.G. Demoulin,[4] Mark G. Macklin,[5,6] Anne E. Mather,[7] Gilles Rixhon,[8] Tom Veldkamp,[9] John Wainwright,[10] Alex Whittaker[11] and Hella Wittmann[12]
[1] Department of Geography, Birkbeck, University of London, London, UK
[2] Department of Physical Geography, Utrecht University, Utrecht, The Netherlands
[3] Département de Géographie et UMR 8591 CNRS, Université Paris 1, Université Paris Est Créteil, Créteil Cedex, France
[4] Department of Physical Geography and Quaternary, University of Liège, Liège, Belgium
[5] School of Geography, University of Lincoln, Lincoln, UK
[6] Institute Agriculture and Environment, College of Sciences, Massey University, Palmerston North, New Zealand
[7] School of Geography, Earth and Environmental Sciences, University of Plymouth, Plymouth, UK
[8] Laboratoire Image, Ville, Environnement (LIVE), UMR 7362 - CNRS, University of Strasbourg-ENGEES, Strasbourg, France
[9] ITC, Faculty of Geo-Information Science and Earth Observation of the University of Twente, Enschede, The Netherlands
[10] Department of Geography, Science Laboratories, Durham University, Durham, UK
[11] Department of Earth Science and Engineering, Imperial College London, South Kensington Campus, London, UK
[12] Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany

Received 16 July 2015; Revised 11 June 2018; Accepted 11 June 2018

*Correspondence to: Rebecca M. Briant, Department of Geography, Birkbeck, University of London, Malet Street, London, WC1E 7HX, UK. E-mail: b.briant@bbk.ac.uk

---

ESPL
Earth Surface Processes and Landforms

ABSTRACT: Field geologists and geomorphologists are increasingly looking to numerical modelling to understand landscape change over time, particularly in river catchments. The application of landscape evolution models (LEMs) started with abstract research questions in synthetic landscapes. Now, however, studies using LEMs on real-world catchments are becoming increasingly common. This development has philosophical implications for model specification and evaluation using geological and geomorphological data, besides practical implications for fieldwork targets and strategy. The type of data produced to drive and constrain LEM simulations has very little in common with that used to calibrate and validate models operating over shorter timescales, making a new approach necessary. Here we argue that catchment fieldwork and LEM studies are best synchronized by complementing the Pattern Oriented Modelling (POM) approach of most fluvial LEMs with Pattern Oriented Sampling (POS) fieldwork approaches. POS can embrace a wide range of field data types, without overly increasing the burden of data collection. In our approach, both POM output and POS field data for a specific catchment are used to quantify key characteristics of a catchment. These are then compared to provide an evaluation of the performance of the model. Early identification of these key characteristics should be undertaken to drive focused POS data collection and POM model specification. Once models are evaluated using this POM/POS approach, conclusions drawn from LEM studies can be used with greater confidence to improve understanding of landscape change. © 2018 John Wiley & Sons, Ltd.

KEYWORDS: landscape evolution modelling; Pattern Oriented Sampling; catchments; fluvial systems; geological field data

Introduction

Traditionally landscape evolution models (LEMs) have been heuristic models based on elaborate fieldwork campaigns encompassing mapping and description of relevant landforms and deposits (e.g. Davis 1922). The interpretation of the collected data on topography, bedrock and sediments of hillslopes and valleys yielded chronological narratives centred around the available evidence (e.g. Maddy 1997; Gibbard and Lewin 2002). These narratives often used simple linear cause and effect reasoning tailored to specific locations and prone to disciplinary biases. A danger with such models is that they may then be applied as universal conceptual models in other locations where key processes differ. The growing awareness that Earth is a coupled system with many global dynamics caused researchers to incorporate known global oscillations such as in tectonics (e.g. Milliman and Syvitski 1992), climate (Bridgland and Westaway 2008; Vandenberghe 2008), base-level (Talling 1998) and glaciation (e.g. Cordier et al. 2017) into their heuristic models. However, since it has become more widely known that earth surfacse processes have non-linear complex dynamics it has also become clear that simple linear cause and effect stories do not accurately capture all real world behaviour. This non-linearity means that not all known global changes have left an imprint in all local records (e.g. Schumm 1973; Vandenberghe 1995; Blum and Törnqvist 2000; Jerolmack and Paola 2010).

Alongside this, the use of numerical LEMs has accelerated. Since the early 1990s (see review by Veldkamp et al. 2017) these have developed into tools used to undertake theoretical experiments about the complexity of earth surface processes, although under controlled and strongly simplified conditions. Because they were invented to explore theoretical questions about past forcings within landscapes, these LEMs are significantly different from other types of models that simulate and forecast processes operating at present. Not least, their relation to field data is only now being assessed in detail, since initial studies frequently used synthetic landscapes (e.g. Whipple and Tucker 1999; Wainwright 2006).

There are five main groups of numerical models that deal with the earth surface processes: climatological, hydrological, ecological, hydraulic-morphodynamic and LEMs. LEMs are distinctive because they combine elements of the other four, frequently enabling all domains to change during a model run rather than modelling one and specifying others as input parameters. In doing this, they focus on long-term geomorphology – both the form of the landscape and the processes operating within it (e.g. Temme et al. 2017). Whilst some geomorphological features form quickly and can be monitored and modelled in parallel to hydraulic measurement and modelling (e.g. Camporeale et al. 2007), evolution of a full geomorphological landscape takes several orders of magnitude longer than human monitoring. The record that remains is therefore scattered and incomplete. As such, the cases being modelled are inherently more intractable. This is not only because process observations, even 'long-term' ones, rarely scale to the geological timescales under study (parameters of the LEM can account partially for this, see Veldkamp et al. 2017), but even more so because the initial conditions required for the LEM cannot be specified simply from modern datasets, even though LEMs are notoriously sensitive to the specification of initial conditions. LEMs share these characteristics of underdetermination with geodynamic models (e.g. Garcia-Castellanos *et al.*, 2003), where key processes and features being modelled occur beneath the land surface and therefore very few initial conditions or processes can be directly measured. In addition, because more features of the landscape are allowed to change in a LEM than in the other types of earth surface models (Mulligan and Wainwright 2004), they require a different approach, analogous to the difference between modern climate and palaeoclimate modelling (Masson-Delmotte et al. 2013).

Many non-LEMs seek numerical prediction (e.g. Oreskes et al. 1994), or at least robust projection of potential scenarios into the future, based on detailed comparison to a short-time period of 'the past'. This is because many of these other types of model (climate, hydrology and ecology) are used as a basis for future policy planning. Thus such models seek to replicate 'reality' more and more closely, as can be seen in the explosion of complexity in General Circulation Models from the 1970s to the present day (e.g. Taylor et al. 2012). This replication of reality is seen in increased inclusion of processes, but also in calibration, where parameters are tuned to known field observations to produce outputs that are as close to measured reality as possible. Once these non-LEMs are validated using a different subset of past data, numerical prediction commences (Oreskes et al. 1994).

In contrast, landscape evolution modelling does not aim for exact replication of present day landscapes, although a measure of this is required to evaluate the usefulness of the model. Rather, the focus in most location-specific LEM studies is on narrowing down the range of processes likely to have been operating in a particular catchment in the geological past. For this reason calibration as defined earlier is rarely undertaken because numerical predictions are not required. This is not least because the difference between what is being modelled and what can be measured is greater than in (for example) hydrological models. For example in relation to temporal scale, the length of time being modelled means that the time steps necessarily used have little physical meaning (e.g. Codilean et al. 2006). Furthermore, some sets of parameter values that seem to fit the data well lack physical plausibility, questioning the value of applying calibration to LEMs, e.g. van der Beek and Bishop (2003). In addition, because of these longer timescales many properties are required to change in landscape evolution modelling that are frequently kept constant in hydrological models. These changing elements propagate impacts and uncertainties in space and time and the introduction of parameterization arguably increases these uncertainties by introducing an additional level of uncertainty (Mulligan and Wainwright 2004). Therefore, with LEMs, the aim is not for more and greater complexity over time, but to constrain uncertainties as much as possible. Because the research questions being addressed usually involve explanation, the goal is to generate a plausible narrative based on the (frequently sparse) data available – just as in a forensic investigation – and not to achieve a numerical outcome that is 'correct' although some measure of the accuracy of approximation of the landscape to the present day is of course required for evaluation. Key research questions are likely to be framed as (e.g. Larsen et al. 2014): which are the most likely modes of formation for the landscape observed? What types or scales of tectonic activity are most likely to produce the landforms observed? What characteristics of a catchment enable a climate signal to be successfully transferred into a sedimentary record? As noted by Temme et al. (2017), the more complete the data available, the more catchment-specific the questions that can be addressed. Often, however, complete landscape and process reconstruction is not possible. Providing evidence to choose between competing hypotheses is more common (e.g. Viveen et al. 2014).

In order to generate a plausible narrative of landscape change, complexity is often actively reduced (e.g. Wainwright and Mulligan 2005). Processes and parameters are only included in an LEM if there is evidence that they are likely to be relevant for explanation. This approach of 'insightful simplification' or 'reduced complexity modelling', does seek to explain what has happened in a specific place, as in the traditional heuristic model, but also to more broadly understand the known global driving factors within fluvial landscapes (Veldkamp and Tebbens 2001), and to create generalizable statements about the development of large-scale geomorphological features. A further advantage of seeking simplification with complex feedbacks is that it allows emergent behaviour. In this case, a relatively simple set of factors is modelled, but can lead to apparently complex behaviour (e.g. Schoorl et al. 2014).

The earlier listed differences in approach between LEMs and other groups of earth surface models, encompass both philosophical issues in modelling and the relationship between models and field observations. This paper, whilst exploring the philosophical issues, seeks mainly to address the issue of field-model data comparison to evaluate LEM output created using this insightful simplification approach. It is aimed predominantly at field scientists, enabling them to apply the multiplicity of papers discussing modelling approaches and philosophy to their specific setting of LEM output and geological field data. In this paper, we argue that field data collection strategies and LEM studies are best brought together by deploying Pattern Oriented Sampling (POS) approaches when collecting field data. In this way, key characteristics of a real-world catchment are identified (e.g. sediment distribution, thalweg gradient, floodplain width) in both past timeslices and in the end situation and used to compare with the same characteristics generated from LEM output. The POS approach that we advocate serves to collect field data that is more useful for comparison with model output. Improving our ability to evaluate model output will then allow us to use LEMs to narrow the range of plausible narratives that explain the field data observed. In this way, we will be able to generate more robust generalizations than either those based on location-specific heuristic/conceptual models (e.g. Bridgland and Westaway 2008) or those using synthetic landscapes (e.g. Whipple and Tucker 1999). Whilst there are philosophical difficulties with strict validation of models of inherently open natural systems (Oreskes et al. 1994), evaluation of such modelling work against relevant field datasets is still crucial to determine at least the empirical adequacy of each model (e.g. Coulthard et al. 2005; Van De Wiel et al. 2011; Veldkamp et al. 2016).

It is our contention that the nature and scarcity of much geological field data, which are typically not randomly generated, preserved or sampled, makes this a different and more intractable process for LEMs than for example hydrological modelling. Whilst it is true that all earth surface process models face problems of comparison with a limited set of field observations, this has mostly to do with bias and gaps in data collection. Because of the time scales involved, field data for comparison with LEM outputs have the additional problem that the geological and geomorphological records (deposits and erosional surfaces alike) are in large part removed and reworked by processes operating since they were first generated. Furthermore, most data are proxies for actual land surface characteristics that may or may not have analogues in the present day. Nonetheless, we argue that our POS can significantly improve the suitability of geological field data selected for model evaluation.

We focus on fluvial landscape evolution in this paper, but some of the general points raised are also relevant for modelling landscape evolution in other process domains. We will first discuss key philosophical considerations in applying field data to LEM evaluation. This is followed by advocating the use of a catchment wide POS approach to support fieldwork inventories, showing how such an approach might apply in different settings. This is a companion paper to Temme et al. (2017), which addresses a similar question from a numerical modelling perspective. Both papers arise from the newly created FACSIMILE (Field And Computer SIMulation In Landscape Evolution) network, which brings together European modellers and field-based geoscientists investigating landscape evolution at various scales with both tectonic and climatic drivers. This POS approach allows a more direct comparison with the Pattern Oriented Modelling (POM) approaches of numerical fluvial LEMs at multiple spatial and temporal scales.

Philosophical Considerations in Applying Field Data to LEM Evaluation

Calibration and parameterization

Parameterization is the inclusion of the most relevant processes for the questions being asked in a particular modelling study. Calibration is setting these parameters to meaningful values for the specific location being modelled. When LEMs are used for studies that fall within the historic time period, then field data is sometimes used for model calibration – i.e. to inform and empirically adjust the parameterization of the model (see for example Veldkamp et al. 2016). This process can also enable useful learning about model function (Temme et al. 2017). We would argue however that this full calibration is neither common nor useful for geological timescale LEM studies. This is despite the fact that LEMs contain multiple spatially-varying parameters that may have only a poor relation to field measurements (containing unmeasurable units such as erodibility) and would thus traditionally be targeted for significant calibration. This is because the aim of many LEMs is to explore process outcomes, rather than to closely mimic field results or provide numerical prediction. As stated by Temme et al. (2017, p. 2173) '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.' Whilst it could be argued that prediction could also be used as a term to refer to the interpolation of data spatially or temporally within the modelling process to estimate a value that has not been or cannot be measured this is not the definition of prediction that we are using here. We argue that such temporal interpolation is merely an extension of the process of exploring different pathways of landscape development. Because the models are not required for prediction, extensive calibration of parameters to a specific geomorphological setting is of less value, and indeed might 'tend to remove the physical basis of a model' (Mulligan and Wainwright 2004, p. 55), for example when parameters are given values that do not make physical sense. It is this physical basis that enables investigation of process outcomes and we would therefore argue needs to be retained.

This retention of basic physics is particularly important because rules drawn from short-term process observations do not scale up easily to longer timescales. One reason for this is that magnitude–frequency distributions of the parameterized events driving the process may have been different in the past, particularly when there is no suitable present day analogue. For example, whilst it is clear that periglacial processes have played an important role in fluvial activity and geomorphological change over Pleistocene timescales across Eurasia and North America (e.g. Vandenberghe 2008), and we understand the links between annual temperature cycle variations and periglacial processes in the modern circum-arctic very well, yet we have no understanding of how such annual freeze–thaw processes differ when occurring in mid-latitude rather than Arctic regions (e.g. Murton and Kolstrup 2003).

In the situation where one is forced to parameterize processes for settings lacking an analogue situation, which is very common when using LEMs, we argue that the researcher should avoid a full calibration of said parameters because it introduces greater certainty into the modelling than there is in the real world. Instead, a wider range of process pathways need to be explored in the LEM than possible using the subset of partial analogue settings for which calibration data would be available. Indeed, not calibrating parameters allows the investigation of process outcomes to also include experiments in which different values of these parameters are investigated, rather than a narrower range of experiments in which they have been 'optimized' in advance of the reported modelling study. For example, Attal et al. (2008) calibrated the model CHILD (Channel-Hillslope Integrated Landscape Development) to known tectonic settings, but other parameters in that LEM were varied in a series of experimental scenarios. Similarly, a restricted range of values can be set for a parameter on the basis of field data without specifying a single value through a traditional parameterization process (e.g. erosion rates estimated between two dated lava flow events – van Gorp *et al.*, 2015).

Validation versus evaluation

A second issue to be considered is that of validation. As Oreskes et al. (1994) state, this is intimately linked with the process of calibration, which we discuss earlier. Strict validation uses a separate dataset to that used for initial model specification and parameter calibration. However, over geological timescales, information relating to each parameter is often too sparse to afford the luxury of splitting a dataset into calibration and validation subsets. Indeed, it is usually the case that almost all the information available is used to specify initial conditions and narrow down the range of parameters used in model runs. Because of this, the only way in which a separate dataset can be generated for validation is by systematically leaving out part of the collected data and using only this data to compare with the key patterns emerging from model outputs in a form of quasi-validation (e.g. Veldkamp et al. 2016). Whilst not strictly independent, this type of quasi-validation is often sufficient to indicate if the LEM simulation is in the correct range of process rates and timing. As discussed in more detail later some quantification of the success of this evaluation/quasi-validation is useful if possible, even though the use of R^2 values to score performance is usually inappropriate.

Equifinality

Thirdly, equifinality is worth discussing because most LEMs of river catchments run forward from some initial situation and end in a simulation of 'the present'. The model output for the present is the simplest to both evaluate (comparing modelled and field data) and analyse (tracing development through time) for explanatory understanding of landscape evolution and the geological/geomorphological record preserved from it. This approach is of course sensitive for equifinality, considering that the generated end state in simulations can be reached in many ways starting from different initial conditions and physical assumptions, whereas in the real world it was just one path. Equifinality is well known to play an important role in fluvial records and their modelling by dedicated LEMs (Beven 1996; Nicholas and Quine 2010; Veldkamp et al. 2017). Such modelling is therefore often coupled with the use of multiple model runs to capture the range of statistical variability between different runs with either fixed or varying parameters. The narrative favoured for explanation is then adopted from the modelled scenario with the best fit to the present day (e.g. Bovy et al. 2016). Where only one scenario fits the geological data available for evaluation, equifinality is avoided. However, we argue here that whilst a single modelled scenario can sometimes be chosen, this is not always helpful in advancing understanding. Indeed, where more than one scenario fits well to the present day, we argue that this should be embraced as defining an envelope of possible explanations, narrowing down our understanding of the processes that could produce such a suite of features without suggesting an unrealistic level of certainty about which landscape history has taken place. If a single solution is still desired, a valuable way of dealing with equifinality in such settings is to gradually work through multiple competing hypotheses. This has traditionally been a common approach in geomorphology for assessing the plausibility of different conceptual models and has recently been adopted by some ecologists, e.g. Johnson and Omland (2004). It has been shown to be particularly useful in evolutionary biology, a field that bears remarkable similarity to landscape evolution modelling, given the long timescales involved, lack of data from many time periods other than the present, and the possibility of equifinality e.g. Lytle (2002). A more recent example of this in landscape evolution is the use of field data alone to determine the relative importance of seepage compared to runoff in canyon formation (Lamb et al. 2006). The two stage LEM strategy of Braun and van der Beek (2004) also demonstrates the gradual investigation of different hypotheses, with a second stage adding in modelling of the lithosphere to enable differentiation between two similar outputs based on different synthetic initial topographies.

Initial conditions

Fourthly, the influence of initial conditions should be considered. When the modelling exercise is carried out in a real-world (rather than synthetic) landscape, specifications of the initial digital elevation model (DEM – resolution, x, y and z accuracy) and surface characteristics (sediment thickness, grain size distribution and erodibility) are particularly important. Whilst all models that forward-simulate open systems require specification of initial conditions (e.g. snow cover or soil moisture in hydrological modelling), specifying initial conditions for geological timescales is particularly problematic because of the scale of difference from modern conditions. This is discussed earlier in relation to calibration and does not apply to other earth surface model types. This scale of difference is important because uncertainty propagation through the modelling process to output DEMs may be significant, and as discussed earlier equifinality can also play a role in such outcomes. For example, if starting topography 'contains the common processing artefact of steps near contour lines, these steps will tend to become areas of strong localized erosion and deposition that can obscure the larger patterns' (Tucker 2009, p. 1454). There are two approaches to specifying the initial DEM. The first is to use the modern land surface. This is only possible if change over time is minimal and topographic data are not used to evaluate model outputs. It has the advantage that the uncertainty relating to spatial resolution and associated interpolation is low [e.g. as investigated by Parsons et al. 1997 for hydrological modelling]. However, the longer the time period to be modelled, the greater the error associated with using such a surface, especially in models where sensitivity to initial conditions is a significant feature. For example, use of a modern DEM is not appropriate where sediments known to be deposited during the time period modelled are present below the modern land surface or when studying a tectonically triggered episode of deep valley incision (e.g. van de Wiel *et al.*, 2011).

Defining an alternative initial DEM or 'palaeoDEM' requires expert judgement based on field experience that is not easily harvested from the literature. For example, when incision over time is the main focus, it may be possible to determine surfaces within the landscape from which incision is likely to have started using modern land-surface DEMs as a starting point, such as relict long profiles (e.g. Beckers et al. 2015) or reliably reconstructed and dated palaeosurfaces (e.g. Fuchs et al. 2012).

A number of numerical approaches can be adopted here, as outlined by Demoulin et al. (2017). Expert judgement can also suggest palaeosurfaces based on sedimentological investigations. For example, erosional contacts may suggest initial surfaces lay higher prior to a period of erosion, but gradational contacts that initial surfaces were close to the base of the sequence. Such delineation is only worth doing however, if terraced depositional units have a thickness greater than the depth of a typical main channel and thus truly deviate from modern surface conditions (e.g. Boenigk and Frechen 2006). The disadvantage of using a reconstructed palaeosurface as an initial DEM is that they are 'typically of very coarse spatial resolution, smoothed and subject to considerable uncertainty' (van de Wiel *et al.*, 2011, p. 179). A useful recent development is the application of geospatial interpolation to refine field derived terrace data sets for palaeosurface reconstructions (Geach et al. 2014; van Gorp *et al.*, 2015). This approach can improve the resolution of the initial DEM and thus the quality of the end results but cannot resolve the fundamental problem of reconstructing the unknown.

The specification of an initial DEM is particularly important for LEMs because the scale of the difference between modern and past landscapes is likely to be large with different processes contributing to their formation (Temme and Veldkamp 2009). However, it should also be undertaken with caution because of this. We therefore propose that future studies should give more thought to initial land surfaces and their conditions whilst field investigation is being undertaken rather than at a later date. If field investigation suggests that the modern land surface is the most appropriate initial DEM to use then the field worker should liaise closely with the modeller to get the highest possible resolution data. This will be only over very short time periods of a century or less where the scale of change is sufficiently small that the additional error gained from using a non-modern initial DEM is no longer justifiable (van de Wiel *et al.*, 2011). If, as in most situations, investigation suggests that a palaeosurface/palaeoDEM should be constructed then additional information such as borehole and geophysical data should be collated to maximize the resolution of the surface created and appropriate geospatial interpolation should be applied (Geach et al. 2014; van Gorp *et al.*, 2015). Indeed, it might sometimes be wiser to turn the nature of the initial land surface into a research question comparing modern and palaeoDEMs in different model runs. In this way questions such as the scale of incision or of reworking of sediment within the landscape can be addressed. The multiple working hypotheses approach outlined earlier and advocated by Temme et al. (2017) can also be used to narrow down the most plausible initial DEM if possible.

Catchment choice

Finally it is important to consider which catchments are more suitable to study at this moment in time whilst we make the transition in landscape evolution modelling from synthetic to real landscapes. This is pivotal because not all catchments actually record the driving factor of interest (e.g. Fryirs et al. 2007). It has been argued that one should choose catchments that form a 'natural experiment' (Tucker 2009), where only one variable changes over the time period of interest – e.g. modelling channel incision in relation to differential rock uplift in the Mendocino Triple Junction region where other features of the catchments compared are broadly similar (Snyder et al. 2003; Tucker 2009). However such catchments are rare and we agree with Temme et al. (2017) that we are now at a stage where catchments exhibiting the 'badass geomorphology' of Phillips (2015) can be studied, although their complexity needs to be reflected in the research question. We must construct very tightly defined research questions for such catchments, by including or excluding specific external factors from experimental runs (e.g. Coulthard and van de Wiel 2013). Evidence for catchment response to climate change can be seen by comparing the coincidence of fossil or isotope based climatic reconstructions (e.g. Table I) with system response (e.g. Lewis et al. 2001; Schmitz and Pujalte 2007). This comparison shows whether the sediment flux signal coming out of the source region is buffered, or even 'shredded' with relation to the original signal (Métivier 1999; Castelltort and van den Driessche 2003; Wittmann et al. 2009; Jerolmack and Paola 2010; Armitage et al. 2013). We can also determine by how much and where it is delayed by intermittent sediment storage related to hill slope – channel (dis)connectivity (Michaelides and Wainwright 2002; Veldkamp et al. 2015). Evidence for tectonic response can be ascertained by geomorphologic markers distributed within the drainage network, such as slope break knickpoints resulting from the same regional uplift pulse (e.g. Table I, Beckers et al. 2015). Nonetheless, as noted by Blum et al. (2013), criteria for distinguishing between allogenic and autogenic control in catchments still remain to be tightly defined and it is recognized by Veldkamp et al. (2017) that there is an urgent need for research strategies that allow the separation of intrinsic and extrinsic record signals using combined fieldwork and modelling.

It is also worth discussing where the boundaries of the catchment should be drawn. In full source to sink modelling, all four of the following elements would be included: a record from the source, a record from the sink, a model for the source and a model for the sink. When catchments are small, downstream data can comprise field data from alluvial fans, floodplains and lakes containing deltaic and prodeltaic deposits. When a larger catchment is considered, the downstream regions are sedimentary basins with broad valleys and plains (e.g. megafans, distributive fluvial systems – e.g. Davidson et al. 2013; Nichols and Fisher 2007, Weissman *et al.*, 2015), lakes (e.g. Schillereff et al. 2015) and/or delta plains and coastal zones (e.g. basins that form part of continental shelves). Often, as discussed later, downstream data from the sink is not readily available and LEM studies simulate only the source area of the catchment, but this is likely to change as the application of LEMs becomes more widespread.

We therefore focus here on the small-medium catchment-scale (*c.* 10–1000 km long channels) over the later parts of the Quaternary where age control is more robust (*c.* 500 000 years to present) – there is only so much 'badass' behaviour that our LEMs can currently manage. We recognize that for now, this excludes ancient systems where preservation is fragmentary or dating absent or very limited. In such catchments, many originally deposited sediment sequences will have been modified by other depositional or erosional processes that may not be captured within the model specification. If numerical modelling is to be applied to such systems, we suggest that lower order research questions, i.e. a more speculative 'what if?' approach could be used to try to capture the main driving processes over longer timescales, and that detailed evaluation of model output in relation to field data is not yet possible.

POS of Field Data for Effective Evaluation of Model Outputs

We propose evaluation of model output using pattern-matching, because it is a practical solution to some of the

Table I. Comparison of areas with sedimentary records where the study focus is usually on climate and anthropogenic forcing of fluvial landscape dynamics, and the more erosional and morphological records which are often more focused on tectonic forcing

	Focus on climatic (+ anthropogenic) forcing	Focus on tectonic forcing, crustal movements and surface deformation
Characteristics of the fluvial response	Drainage system more likely to respond ubiquitously because when climate change imposes variations in hillslope sediment delivery or discharge, this affects a drainage network everywhere at once, if systems are buffered to this forcing	Drainage system response dependent on nature of forcing. May be highly localized (e.g. surface deformation from faulting) or more regional (e.g. regional uplift), with impacts propagating over time, for instance upstream of an active fault. Longer-term sediment flux histories as a result of tectonically-driven surface uplift or exhumation
Field data most commonly used for reconstructions	Aggradation and sedimentary data	Erosional and morphological data
Typical data set characteristics	Numerous and large data sets (number of observations, ages) available (e.g. Macklin et al. 2012a, 2012b)	Fewer data sets, often made of small numbers of data (e.g. Demoulin et al. 2015)
Scale of vertical evolution	Decimetre to metre	Decimetre
Typical time step in models	10^{-2} to 10^{2} years (days to centuries)	10^{0} to 10^{4} years (years to tens of thousands of years)
Models commonly used (non-exhaustive list)	CAESAR, LAPSUS, WATEM, CybErosion, PARALLEM (Coulthard et al. 2005; Wainwright 2007; Schoorl et al. 2014)	CHILD, CASCADE, SSTRIM (e.g. Berlin and Anderson 2007; Anthony and Granger 2007b; Tucker 2009)
External forcing data required for model input	Regional (or otherwise appropriate) climate time-series (temperature, precipitation). Time-series of human impacts (e.g. land clearance)	Regional uplift rates. Fault slip rates.
Means of evaluating/comparing the field and model data	Palaeoenvironmental data compilations + reconstructions of paleoclimatic variability and human-impact inferred from it (e.g. Lewis et al. 2001; Viveen et al. 2014; Benito et al. 2015). Downstream data	Structural geological data, palaeoseismicity investigations (fault history) Erosional morphological features (e.g. Beckers et al. 2015; Cohen et al. 2008; Whittaker and Boulton 2012) Downstream data

difficulties encountered in comparing it against geological data. This is an approach that has been used in ecological research for several decades (e.g. Grimm et al. 1996, 2005), and to some extent in fluvial geomorphology, e.g. Nicholas (2013). In this practical approach, adequate models should be able to (re-)create similar emergent properties to the field data, not only time-series.

Taking this approach requires that we are very specific in defining what these emergent properties or key characteristics are. For any one catchment these may be geomorphological features or sedimentary sequences. Different types of field data will therefore be available from each catchment, some of the most common of which are outlined in Tables I and II. Once identified, both field and model development can be focused on these catchment-specific properties (Figure 1). This will enable development of model outputs that can be most readily compared with field data in combined POM (Grimm and Railsback 2012) and POS approaches. These should be chosen to allow evaluation or quasi-validation, preferably using semi-quantitative measures, as discussed earlier. It is likely that some fieldwork will already have been undertaken at this stage, but we advocate that these discussions should not be left until after all field data has been collected. Identification of key characteristics to be used in a POM/POS approach should precede a further round of fieldwork and data gathering, this time focused purely on the key characteristics identified, rather than driven by opportunistic availability of sedimentary sequences (Figure 1). It is our contention that this approach will open up whole catchments and a wider range of field data to study. We do not therefore advocate more fieldwork, but more targeted collection of field data by considering comparison with model output at an earlier stage in the research process.

Figure 2 illustrates the type of records that could be sampled if occurring in the investigated research area. These proposed multi-scale records are both erosional landscape features and sedimentary records such as soil depth patterns, hillslope/colluvial records, local alluvial fan records, fluvial terrace records and delta records. The latter are particularly often overlooked in field studies and yet fundamental in providing an independent 'depositional' mirror record of the 'erosional' record in the catchment (e.g. Whittaker et al. 2010; Forzoni et al. 2014). Comparing the catchment and downstream data and partitioning the sediment budget to ensure that the budget 'closes' as effectively as possible (although see caveats in Parsons 2012) will improve the quality of model input data. Sediment budgeting also better quantifies the field data, enabling more precise evaluation of the match between modelled outputs and field observations. However, it is not always easy to include downstream data. Sometimes sediment budgets cannot be closed if small-scale sinks within the system store sediment over significant time periods (e.g. Blöthe and Korup 2013), or the downstream record is incomplete (e.g. Parsons 2012) or 'leaky' (i.e. sediment passes through to even more downstream areas such as the coast, sea or shelf). This 'leakiness' is hard to quantify from the geological record alone (e.g. Jerolmack and Paola 2010; Armitage et al. 2013; Godard et al. 2014). Non-linearities due to hillslope–channel (dis) connectivity and events such as river capture or glacial interventions would also cause a lack of a clear source to sink connectivity. In relation to other record types, an example is sub-catchment outlet beryllium-10 (^{10}Be) erosion rates which can be measured to get time aggregated erosion rates (e.g. Von Blanckenburg 2005) and combined with sediment budget estimates from source sink comparisons (item 8, Table II).

POS can also be applied not simply for evaluation but also for specifying initial conditions such as sediment thickness and composition for each grid cell, to avoid assuming a uniform cover across the catchment due to limited information.

Table II. Seven potential field data types that can be used to improve field-model pattern comparison

Item	Field data	Model data	References	Field data and model output requirements for effective comparison	Quantification of fit between data?
1	Estimated volume of observed sediments	Modelled volume of sediments	Viveen et al. (2014), Figure 3a; Coulthard et al. (2005), Figure 3b	• Spatial definition of area of key sediment body (e.g. fan) • Borehole records enabling extrapolation of sediment volumes • Modelled estimate of sediment volumes within specified area	Yes – compare volumes
2	Palaeo-hydrological discharge estimates	Modelled discharge estimates	Mather and Stokes (2016); Stokes et al. (2012); Van Balen *et al*. (2010); Westaway and Bridgland (2010, 2011); Busschers et al. (2011)	• Specification of location of discharge estimates within system • Measurements of key channel parameters • Multiple methods of estimating palaeohydrological discharges • Modern discharge data from catchment • Modelled discharges using a comparable methodology	Yes – compare discharge estimates
3	Observed borehole sequences	'Synthetic' boreholes	Newly proposed	• Borehole/section log data with age control to define time periods for comparability • Modelled time-series data for a spatial location corresponding to the observed sequences • Algorithm for aggregating erosion and deposition to generate 'synthetic' borehole record	Possibly – depends on details of synthetic borehole algorithm
4	Observed river aggradation and incision phases	Modelled river aggradation and incision phases	Coulthard et al. (2005), Figure 3	• Robustly defined and dated aggradation and incision phases using meta-analysis protocols • Aggradation/incision time series for a comparable area (catchment/region)	Possibly – depends on chronology and comparability of aggradation data
5	Planform characteristics of palaeochannels	Planform characteristics of river channels	Nicholas (2013)	• Measured wavelengths and widths of modern channels and palaeochannels • Measured wavelengths and widths of modelled channels	Yes – compare wavelengths
6	Measured long profile	Modelled long profile	Stange et al. (2016), Figure 4; Veldkamp et al. (2016)	• Quantitative measures of modern long profile shape/gradient • Reconstructed gradients of terrace bodies • Measures of convergence/divergence/parallelism • Modelled long profile shape/gradient	No
7	Knickpoint mapping/ TCN dating	Knickpoint propagation rates	Mather et al. (2002); Stokes et al. (2002); Yanites et al. (2009); Crosby and Whipple (2006)	• Knickpoint mapping • TCN dating of erosion waves across catchment • Modelled knickpoint propagation rates	Possibly but assumptions behind field data
8	Cosmogenic-based catchment denudation rates	Modelled denudation rates	Yanites et al. (2009); Willenbring et al. (2013), Figure 5	• Samples for cosmogenic-based denudation rates in appropriate and well-defined spatial locations • Directly comparable model output covering the same catchment area	Yes if directly comparable

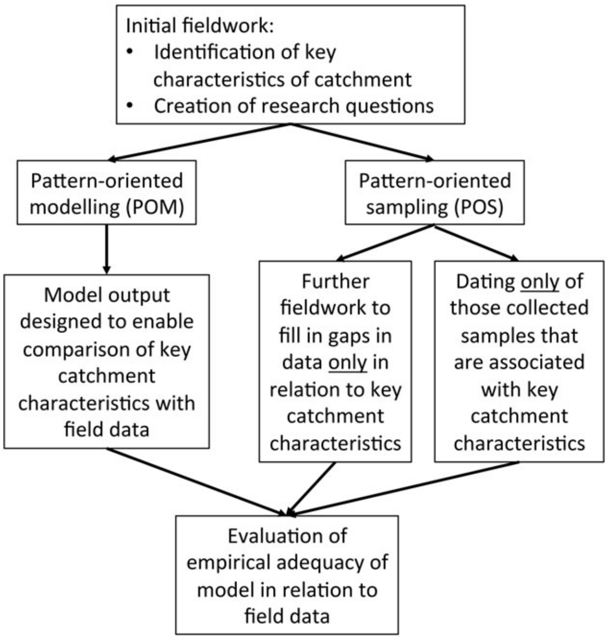

Figure 1. Flow chart for applying Pattern Oriented Modelling (POM) and Pattern Oriented Sampling (POS) within a joint field-model investigation of a specific catchment.

Whilst this may involve more fieldwork, it may rather involve creatively using existing datasets for this new purpose. Good pedological maps can be invaluable in achieving this aim (e.g. Bovy et al. 2016), as can use of geotechnical borehole data. These datasets can also be usefully used for making volumetric comparisons of various types, as noted in Table II. In parallel with developments in the automatic recognition of landforms (e.g. Jones et al. 2007) from DEMs, new technologies and data sources such as ground penetrating radar (GPR), other geophysical surveys, LiDAR (light detection and ranging) data (both airborne and scanning vertical faces) and the game changing use of structure-from motion (SfM) to generate high resolution digital surface models (DSMs) from aerial and unmanned aerial vehicle (UAV) imagery (e.g. Dabskia *et al.*, 2017) make the collection of geomorphological and spatially distributed sedimentary data much more feasible than was previously the case (Demoulin et al. 2007; Del Val et al. 2015). These data can be used iteratively with remotely sensed data both before and after field investigations. This spatially distributed dataset can provide information on erosional and depositional landforms as well as sedimentary units (Tables 1 and II).

Systematic collection of data from multiple landscape elements using a POS approach generates a better description and understanding of the catchment and thus allows for a more effective evaluation of model output than illustrated by Temme et al. (2017) in their figure 4.

The strength of POM is that it recognizes both the inherent (x, y,z,t) uncertainties in specification of initial conditions and the non-linearity of ecological and geomorphological processes and systems. Systematic POS will allow a more systematic characterization of the relevant landscape properties that can then be used for systematic sensitivity analysis of the developed LEM. It is for example equally relevant to know where sediments occur and where they do not. For LEMs, the inherent (x,y,z,t) uncertainties are primarily due to DEMs, sediment thickness/characteristics and dating technique uncertainties. Too often we have much data from particular locations while at the same time we have almost no data outside these unique locations (often boreholes and quarries). Non-linearity evaluation requires approaches such as Monte Carlo sensitivity ensembles to quantify the role of autogenic feedbacks in the model outcomes (Nicholas and Quine 2010). In order to do this in a meaningful way we have to quantify their spatial and

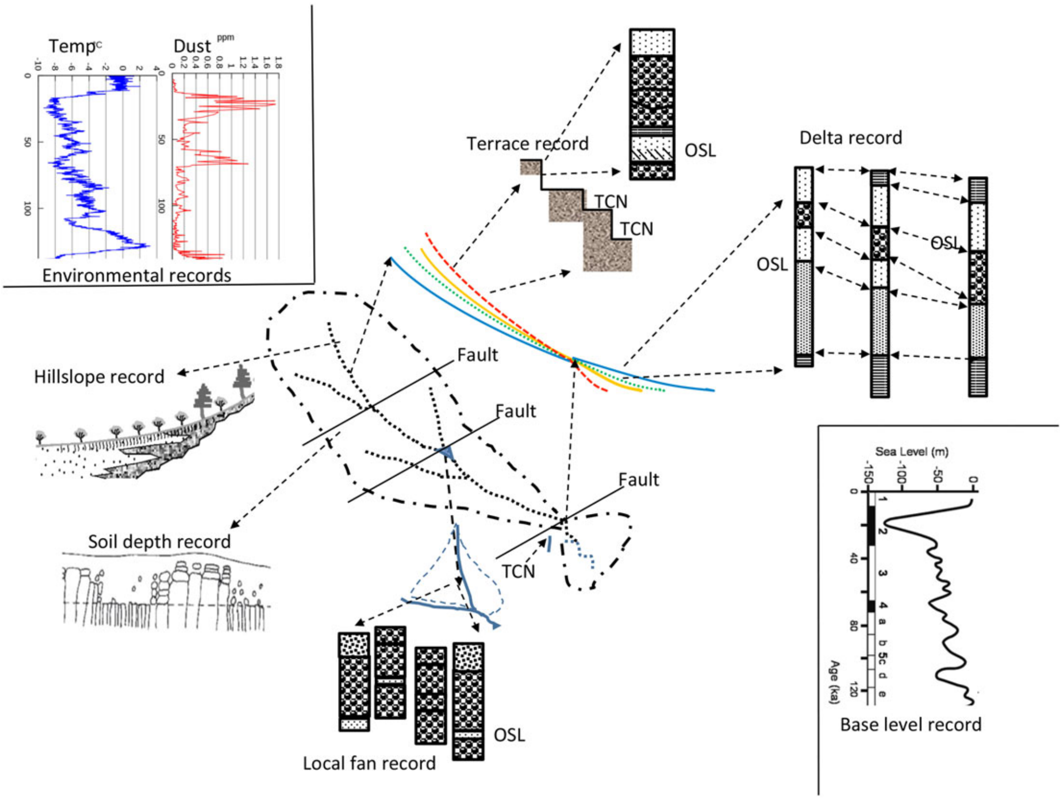

Figure 2. The type of records that should be sampled in a Pattern Oriented Sampling (POS) approach if occurring in the investigated research area. [Colour figure can be viewed at wileyonlinelibrary.com]

temporal distributions as well as possible. For example, Hajek et al. (2010) statistically define the degree of channel-belt clustering. By comparing the degree of spatial clustering between channel units observed in late Cretaceous-age rocks and a flume experiment, they conclude that the patterns observed could have formed as a result of self-organization within the system rather than due to external forcing (Humphrey and Heller 1995). A similar approach is taken with Quaternary age sequences by Bovy et al. (2016).

Similarly the strength of POS as illustrated in Figure 2 is that it recognizes the inherently stochastic nature of sediment preservation at the land surface compared with at-a-point comparisons. POS therefore widens the range of possible field data that can be used whilst simultaneously targeting only those data types that actually add information about the key characteristics identified. It is likely that this will include areas with no sedimentary records, running counter to much current geological fieldwork practice. It may also require the collection of field data for evaluation of model output across the whole catchment. As such it will require an intentional strategy and possibly some additional resources to observe and describe sedimentary successions and landforms even in hard to access locations. We propose here various new data types and patterns as useful for pattern-matching comparisons (Table II), many of which can be quantified and applied concurrently. As shown in Figure 1, identification of which of these can be used in model evaluation is crucial in guiding fieldwork strategy.

POS also aids in decision making when attempting to build a robust chronology because sample selection can be targeted to the key characteristics identified for the catchment as shown in Figure 1. For example, where depositional units are the focus, samples should be taken to enable robust comparison between sedimentary units. This means that whilst it is necessary only to undertake chronological analyses from suitable depositional settings (Table III), chronological data should be sampled both up and downstream (e.g. Chiverrell et al. 2011; Rixhon et al. 2011; Macklin et al. 2012a), combining vertical (successive terrace levels at a given location, e.g. Bahain et al. 2007) and longitudinal (same level at multiple places along the river profile, e.g. Cordier et al. 2014) sampling. This is especially important because many terraces and other fluvial sedimentary bodies are diachronous features (Veldkamp and Tebbens 2001; van Balen et al. 2010). Where stratigraphic relationships are well-known, Bayesian statistics can and should be used to increase age precision. We note, however, that Bayesian statistics are only helpful where units are in direct stratigraphic superposition (e.g. Toms 2013; Bayliss et al. 2015). Thus significant sediment bodies should be sampled more than once, with replication at each location of ideally up to five samples. In addition, as has been argued by many authors (e.g. Rixhon et al. 2017), multiple chronological methods (Table III) should be used where possible to improve robustness of the dating. Care should be taken to avoid both the use of techniques beyond their reliable limits and lack of clarity about the event being dated (e.g. Macklin et al. 2010).

In contrast, where erosional features are the key characteristic in a catchment, the determination of denudation rates using Terrestrial Cosmogenic Nuclide (TCN) data can provide values with which overall mean denudation rates of a catchment can be quantified (e.g. Schaller et al. 2001, 2002; Von Blanckenburg 2005; Wittmann et al. 2009). As discussed earlier, catchment averaged TCN data is a good target for model-data comparison because such long-term, spatially-averaged data are often produced by models (see for example Veldkamp et al. 2016). Low-temperature thermochronology is another source of (modelled) data complementary to TCN (Table III). It is used routinely for estimating (very) long-term denudation rates in active orogens (e.g. Willet et al. 2003) or in their adjacent basins. As an example, Valla et al. (2011) used thermochronology to demonstrate increased incision and relief production in the Alps since the Middle Pleistocene and King et al. (2016) show changes in the nature of uplift in the Himalayas.

Once appropriate data has been gathered, pattern-matching can and should be separated into the qualitative recognition of spatial patterns and the statistically quantified distribution of specific, quantifiable features (e.g. slopes, soil or sediment thickness or volume, Table II) within model output. Quantification of the goodness-of-fit should be applied wherever possible whilst bearing in mind the appropriate spatial scale. For example, statistical analysis has been used for comparing probability density functions of carbon-14 (^{14}C) dated Holocene flood units in New Zealand and the UK in order to demonstrate inter-hemispheric asynchrony of centennial- and multi-centennial-length episodes of river flooding related to short-term climate change (Macklin et al. 2012a). However, such meta-analyses sometimes aggregate data to too high a level, losing the spatial variability of the data and thus data that would be crucial for evaluating POM. Quantification of goodness-of-fit will not always be possible, but where it is, this is noted in Table II. It should be noted that there will always be an element of subjectivity/expert judgement about whether the fit is ‘good enough’. As discussed earlier, multiple uncertainties in LEMs over geological timescales negate the uncritical use of R^2 values as in a traditional validation process.

POS Applied to Specific Field Settings

Three main case study types can be distinguished where different types of field data are relevant to be used in comparisons with model output. These are (1) sedimentary records where the study focus is usually on climate and anthropogenic forcing of fluvial landscape dynamics (e.g. Viveen et al. 2014), (2) the more erosional and morphological records that are often more focused on tectonic forcing (e.g. Beckers et al. 2015; Demoulin et al. 2015) and (3) study of long-term denudation rates (e.g. Willenbring et al. 2013; Veldkamp et al. 2016). The two first categories are compared in Table I and discussed in more detail later in relation to POS. All case study types have still unresolved challenges related to the previously discussed issues of initial topography, equifinality and the separation of internal complex response from external forcing. Table I demonstrates the different data scale emphasis of the two first case study types. Table II gives seven potential field data types that can be used to improve field-model pattern comparison.

A detailed discussion of the data that will be most useful in evaluating model output is important because the data that is generated separately by the two endeavours (modelling and fieldwork) are by nature very different. For example, field data often comprises detailed study of only a very small part of the catchment (the best or ‘type’ example). Depending on the methods used to develop a chronology the reconstructed depositional history of a catchment may also lack significant temporal resolution, perhaps due to lack of dateable material or error bars that are too large. Indeed even the smallest error bars possible are frequently larger than the time intervals used in model runs. In contrast, model outputs have complete spatial coverage (e.g. mapped change in height/volume of sediment deposited) with high temporal resolution, but often lack local detail. Variables outputted by models are also different from those generated from field-based geological records – e.g. sediment and discharge variations which can only be inferred from sedimentary sequences, not directly measured. Whilst a combined

Table III. Table suggesting which chronological techniques are most appropriate for each timescale and event type

Feature to be investigated	What catchment activity does it constrain?	What timescale is being addressed?	What techniques can be used?	Example from literature	Modelled comparison feature
In-cave-deposited alluvium	Incision episode	>300 000 years BP	TCN burial dating (^{26}Al/^{10}Be): rapid and complete burial	Granger et al. (2001); Anthony and Granger (2007a); Wagner et al. (2010)	Incisional event
Fluvial deposit, e.g. accumulation terrace or fan unit	Aggradational event	>300 000 years BP	TCN burial dating (^{26}Al/^{10}Be): isochron method	Balco and Rovey (2008); Erlanger et al. (2012); Balco et al. (2013)	Sediment body
		>300 000 years BP	IRSL – pIRIR290	Cordier et al. (2012, 2014)	Sediment body
		>300 000 years BP	ESR dating in quartz	Chaussé et al. (2004); Bahain et al. (2007); Zhu et al. (2014)	Sediment body
Fluvial deposit, e.g. accumulation terrace or fan unit	Abandonment time of the landform; onset of incision	30 000–300 000 years BP	TCN surface exposure dating: concentration depth profile	Anderson et al. (1996); Brocard et al. (2003); Rixhon et al. (2011)	River incision - both temporal (e.g. chronological control on a terrace staircase) AND spatial (e.g. propagation of an upstream erosion wave - knickpoint diffusion)
	Aggradational event	30 000–300 000 years BP	OSL and IRSL – pIRIR290, SAR, sample dependent	Briant et al. (2006); Cordier et al. (2014); Kars et al. (2012)	Sediment body
	Period of stasis	30 000–300 000 years BP	U-Series dating of carbonate cements	Sharp et al. (2003); Adamson et al. (2014)	Surface
	Period of incision	30 000–300 000 years BP	U-Series dating of travertines	Veldkamp et al. (2004)	River incision timing and rate
Fluvial deposit, e.g. accumulation terrace or fan unit	Aggradational event	10 000–30 000 years BP	OSL - SAR	Choi et al. (2007); Kock et al. (2009)	Sediment body
Fluvial deposit, e.g. channel within terrace form or floodplain	Aggradational event incorporating material previously deposited on land surface	10 000–30 000 years BP – N.B. potential for contamination in samples at the limit of the technique	Radiocarbon – measurements should be on charcoal, bone or identified seeds and shells only, using appropriate pretreatments	Briant and Bateman (2009); Briant et al. (2018); Higham et al. (2006); Bird et al. (1999); Busschers et al. (2014)	Sediment body
Fluvial deposit, e.g. channel within terrace form or floodplain	Aggradational event incorporating material previously deposited on land surface	<10 000 years BP	Radiocarbon – constraining various different events and determining 'change after' dates for meta-analysis	Jones et al. (2015)	Sediment body
Fluvial deposit, e.g. channel within terrace form or floodplain	Aggradational event	<10 000 years BP	OSL - SAR	Schulte et al. (2008); Pierce et al. (2011)	Sediment body
Erosional landform, e.g. strath terrace	Abandonment time of the landform; onset of incision	Holocene/Late Pleistocene	TCN surface exposure dating	Burbank et al. (1996); Reusser et al. (2004); Seong et al. (2008)	Incisional event and knickpoint diffusion rate in actively uplifting orogens.
		1000–1 000 000 years			Denudation rates

(*Continues*)

Table 3. (Continued)

Feature to be investigated	What catchment activity does it constrain?	What timescale is being addressed?	What techniques can be used?	Example from literature	Modelled comparison feature
Modern river bedload, sediments of dated terraces, and recycling of sediments in floodplains	Catchment-scale (palaeo) denudation rate		TCN concentration (usually ^{10}Be and/or ^{26}Al)	Von Blanckenburg (2005); Wittmann et al. (2009)	
Valley walls in orogens or basin strata	Large-scale (orogen or basin-wide) denudation rate	Meso-Cenozoic, including Quaternary	Low-temperature thermochronology	Willett *et al.* (2003); Valla et al. (2011)	Denudation rates

POM/POS approach can aim to minimize these differences, it can never completely eliminate them.

Sedimentary records with a focus on climate and anthropogenic forcing

Comparison of sedimentary field data and modelled deposition will involve integration of borehole and three-dimensional (3D) surface data within a single system (Table II). For example Viveen *et al.* (2014, Figure 3a) used spatially constrained data on sediment thickness to compare with model output at multiple locations within a catchment, as do Geach et al. (2015). This is not as useful as volumetric data because it potentially masks the volumetric implications of variations in sediment thickness due to confluences, uneven floodplain bases and scour hollows. However, borehole data is not widely available from the regions in which these studies were based, so average sediment thickness had to be used instead. This limits the quality of the match between field and model data in these studies and means they are compared only qualitatively. It is also exemplified by the qualitative comparison of modelled and observed histograms of Holocene 500-year step sediment delivery for the Rhine and the Meuse delta sediments (Erkens et al. 2006; Erkens 2009) and catchment-data based quantifications. These studies could potentially be taken further by direct comparison of the modelled and observed volumes of key sediment bodies within a catchment, tightly, spatially constrained to ensure comparability (see item 1 in Table II). An alternative approach to understanding fluvial activity over time using estimates of palaeohydrology (item 2, Table II) over longer time periods shows that results are highly dependent on the approach used, highlighting a need to develop more standardized approaches for describing Quaternary river archives (both alluvial fans and terraces – e.g. Stokes et al. 2012; Mather and Stokes 2016; Mather et al. 2017).

Meta-analysis, a systematic approach to aggregating dated sedimentary units and landforms in catchments (e.g. Thorndycraft and Benito 2006; Macklin et al. 2013), can also be used in model evaluation at a catchment-scale. For example, it has been used for comparing periods of aggradation and quiescence found in the modelled and observed records in four adjacent upland catchments (e.g. Coulthard et al. 2005; item 4, Table II; Figure 3b). The use of consistent protocols for the aggregation of data is important in order to quantify reach-scale variability in the fluvial record (cf. Macklin et al. 2012b), enabling catchment-wide and regional patterns to be detected. What we advocate with the POS however is not only aggregation but also disaggregation of data to specific locations in the catchment to get a more comprehensive picture of the fluvial system pattern for model comparison. More work also needs to be undertaken on how to quantify the comparison of this data type because it is very dependent on the quality of the chronology (item 4, Table II).

Erosional and morphological records with a focus on tectonic forcing

Where the landscape is mostly erosional and the main landscape driver of interest is crustal uplift (Table I) high quality morphological data is relevant. Specific DEM-derived metrics (e.g. chi plots, hypsometric integrals, geophysical relief, R/SR – e.g. Cohen et al. 2008; Perron and Royden 2013; Demoulin et al. 2015) can be used to quantify field characteristics and integrated into a common geographic information system (GIS) software package, which will facilitate pattern-matching with model output in addition to greater understanding of the

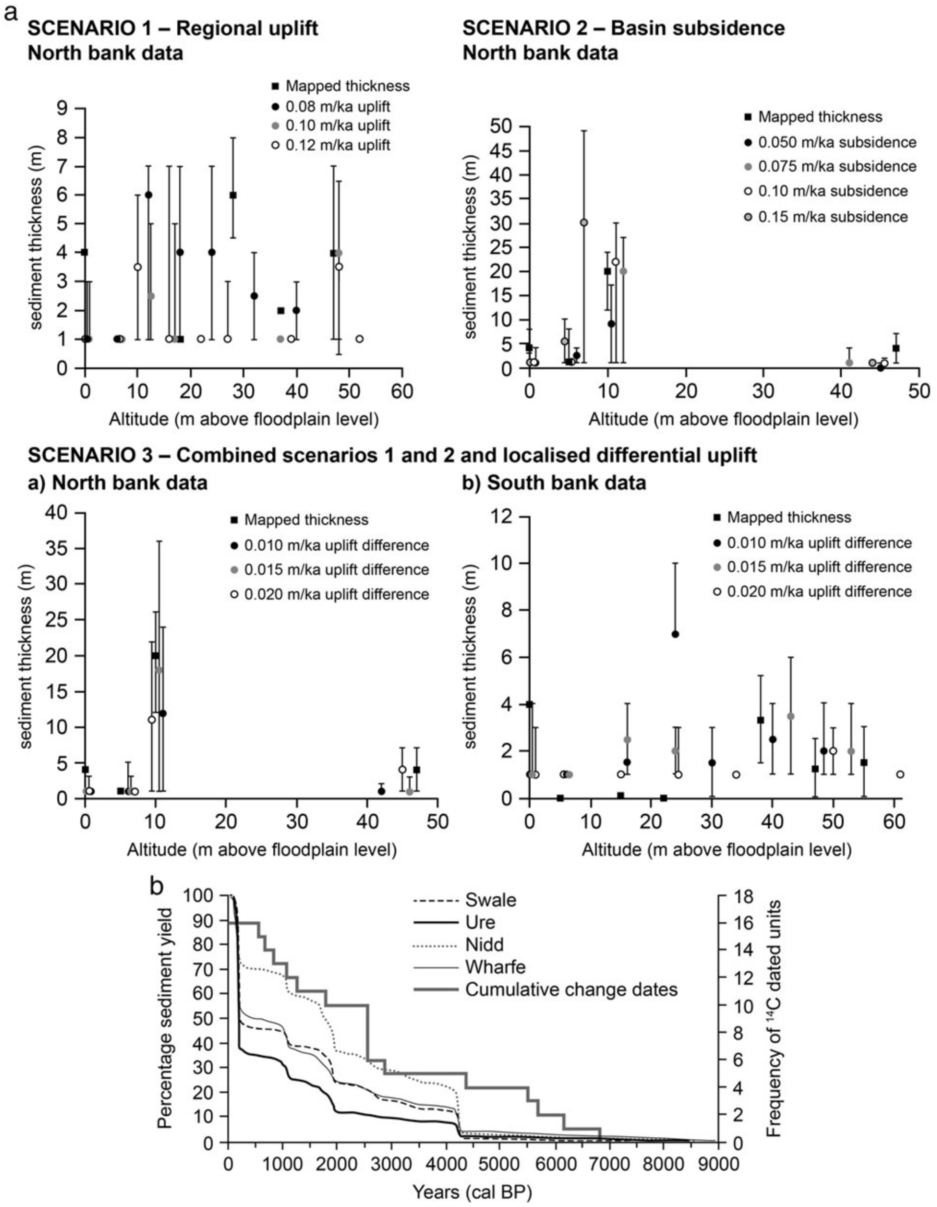

Figure 3. (a) Mapped and simulated sediment thicknesses for terrace levels from the northwest Iberian lower Miño River basin redrawn from Viveen et al. (2014, Figures 10–12). Median = dot, minima and maxima are given by the error bars. Three scenarios were modelled, as shown. The authors argue that model Scenario 3 matches the mapped sediment thicknesses most closely. (b) Comparison of modelled sediment yield from four upland catchments in northern England with sediment preservation as recorded by the frequency of radiocarbon dated units in these catchments. Redrawn from Figure 15 of Coulthard et al. (2005).

systems by comparison with other basins. Data such as non-lithologically controlled knickpoints or vertical spacing between fluvial terrace levels may additionally be useful for model output evaluation. As Stange et al. (2016) show, the spacing, timing and tilting (i.e. convergent, divergent or parallel) of exposure dated terrace forms can provide a powerful modelling test of competing hypotheses about the tectonic history of a region (item 6, Table II; Figure 4). Significantly more work is needed to quantify the match between field and modelled data in relation to long profiles however. At present this is possible only subjectively. Similarly, many studies support the usefulness of knickpoint mapping (item 7, Table II). They can be used to test the validity of river-incision models based on the stream power law (e.g. Berlin and Anderson 2007; Beckers et al. 2015) and evaluate the role of additional controls on incision (e.g. Whittaker et al. 2007, 2008; Whittaker and Boulton 2012). TCN dating of the progression of erosion waves across drainage systems also enables the two types of data to be compared (e.g. Anthony and Granger 2007b; Rixhon et al. 2011). However not all knickpoints are valid targets for model-data comparison. For example, a knickpoint in a highly erodible lithology or highly resistant lithology subject to structural discontinuities (e.g. Antón et al. 2015) is unlikely to be useful for evaluating landscape evolution modelling of longer timescales because climatic or tectonic controls on migration will be masked. In addition, other tectonic factors will influence fluvial systems, for example dislocation of river courses across laterally or vertically faulted landscapes, differential uplift or subsidence across substrate lithological boundaries or solution driven collapse.

Promising new techniques for quantifying denudation rates

In situ cosmogenic-based denudation rates, which are inherently spatially and temporally averaged (item 8, Table II)

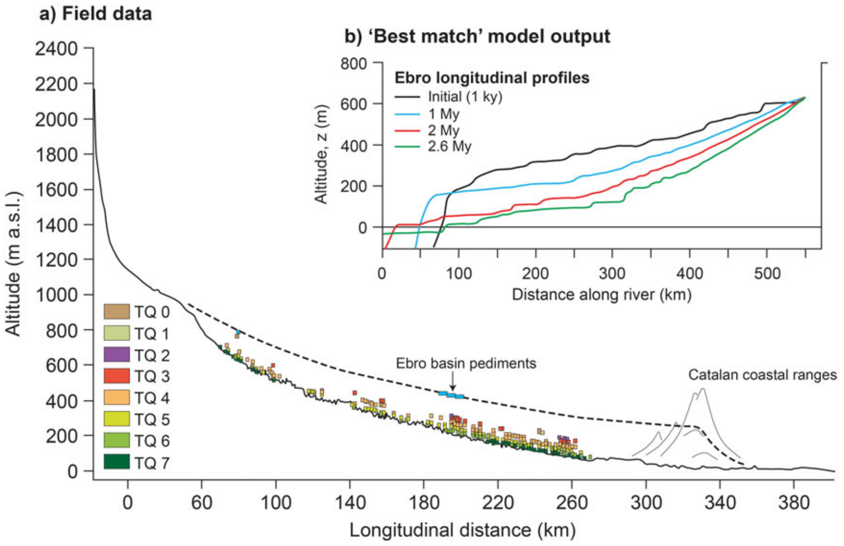

Figure 4. (a) Observed and mapped sub-parallel long profiles of both the present day Segre River (solid line), Pleistocene terrace remnants (coloured boxes) and Pliocene-Quaternary pedimentation surfaces (labelled). (b) 'Best-match' model output from the only one of four modelled scenarios which shows sub-parallel development of terraces. This scenario is continuous Quaternary uplift and climate variability. Redrawn from Stange *et al.* (2016, figures 4 and 12). [Colour figure can be viewed at wileyonlinelibrary.com]

provide an additional opportunity for a very powerful check on denudation rates produced from LEMs. They can only be used where the relevant assumptions hold (i.e. relatively steady rates of sediment production over time, well-mixed sediment). To date, most comparisons of numerical model output with cosmogenic denudation rates have been undertaken with the aim of better understanding the robustness of the TCN signal, for example in relation to different rates and styles of climate change (Schaller and Ehlers 2006), or in basins where sediment inputs to the system are dominated by landslides (Yanites et al. 2009). More recently, the ability of spatial analysis of such denudation rates to improve understanding of transient response to a tectonic perturbation has been effectively shown by Willenbring et al. (2013), with an acceptable match between independently modelled and cosmogenic-based basin-wide denudation rates (Figure 5). More recently Veldkamp et al. (2016) used fluvial terrace properties (thickness and timing of deposition/erosion for specific locations) to calibrate a longitudinal profile model. After an elaborate stepwise calibration and sensitivity analysis the derived temporal landscape erosion (sedimentary delivery) rates were compared with measured ^{10}Be catchment denudation rates (Schaller et al. 2002), proving to be comparable both in rate magnitudes and timing. We therefore propose that this approach has reached sufficient maturity that it should be used more widely in future studies, by using cosmogenic-based denudation rates as a means to evaluate landscape evolution modelling over timescales of 10^2 to 10^5 years.

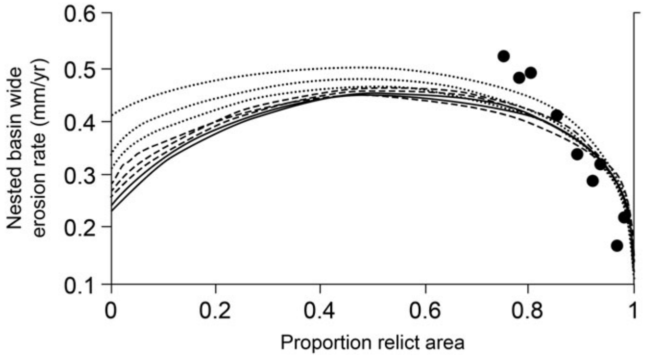

Figure 5. Plot comparing nested basin-wide erosion rates and relict proportion of nested watershed upstream of each sample site for nine sampling locations within the South Fork Eel River catchment. Solid lines show sampling locations 0–20 km upstream of the catchment outflow; dashed lines locations 20–60 km upstream and dotted lines 60–100 km upstream. Black dots show measured detrital ^{10}Be denudation rates and their close match with part of the modelled curves is given as evidence of the usefulness of the model across the full range of the curves. Redrawn from figure 1G of Willenbring et al. (2013).

With regard to intra-catchment pattern identification, TCN-based denudation rates can address this by sampling streams of different orders. Differences between catchments can highlight a specific intrinsic control such as lithology, steepness, a climatic gradient or different tectonic histories (Von Blanckenburg 2005), which are also key questions often addressed in landscape evolution modelling studies. TCN-based denudation rates help to constrain such controls across a wide range of spatial scales. However, one must bear in mind that a steady state assumption is intrinsic when deriving TCN-denudation rates, such that applications of this method to non-steady state settings should be exercised with care. Non-steady state settings are most common in catchments prone to mass-wasting processes, such as landsliding, where most of the sediments leaving the catchment may originate from a small area and there is therefore incomplete sediment mixing between hillslope and channels, as recorded in differing cosmogenic nuclide concentrations (Small et al. 1997; Norton et al. 2010; Binnie et al. 2006; Savi et al. 2014). Large contrasts in lithology within a catchment may also cause these assumptions to be violated (Von Blanckenburg 2005). In practice, although such situations should be avoided when they are obviously present, they are rare and in many cases TCN data have proven to record robust denudation rates over wide ranges of climatic and tectonic settings (Table II).

Recommendations

LEMs have moved away from purely theoretical research questions addressed in synthetic landscapes towards answering specific research questions in particular catchments. This brings into sharp relief the nature of the field data that enables effective evaluation of model outputs. We have argued earlier that the current practise of field data collection does not always allow for this. We believe that there are two key elements to be addressed.

Firstly, researchers need to be aware that LEMs are qualitatively different from other earth surface process models commonly used in the environmental sciences. They operate over longer geological time periods, with sparser datasets and a different purpose. Research questions usually seek explanation rather than numerical prediction, using an insightful simplification approach where minimum numbers of parameters are used. Instead of seeking an optimum set of parameters, different model runs often explore their relative importance and the effects of changing their amplitude. Whilst such forward modelling can result in equifinality, we argue that this should be embraced as narrowing down the plausible set of events that have occurred in the catchment, even if not converging on a single outcome. Indeed such convergence might suggest a greater level of certainty than is actually present and thus be misleading.

Secondly, we advocate the use of a quantitative pattern-matching approach for field-model evaluation such as that often used in ecological studies. Recognizing that fluvial landscape evolution modelling is also a POM approach (Grimm and Railsback 2012), generating geological field data that is comparable with model output will require adaptation of fieldwork strategies using POS. This sampling should focus only on data that provides information about identified key characteristics of the catchment (Figure 1). This will embrace a wider range of data types overall (Figure 2), but not increase the burden of data collection for study of a specific catchment. A number of suitable data targets for such an approach are outlined in Table II and exemplified in Figures 3–5 and related text.

We have shown that POS is starting to be applied in some cases. However, we believe that the community should more generally apply these principles in a structured way. Our aim as FACSIMILE is to facilitate a research approach that compares this wider range of field data with model output from a range of model types. Given that it is neither possible nor desirable to model all systems, we are in the process of working on a specific field catchment where initial pattern-matching model-data comparisons can be undertaken to determine further which approaches are most useful.

Acknowledgements—The authors are grateful to the Faculty of Geo-Information Science and Earth Observation of the University of Twente who provided support for two workshops during which these ideas were developed. The authors also thank three anonymous referees and the Associate Editor Professor Stuart Lane whose comments greatly clarified the thinking on these matters.

References

Adamson KR, Woodward JC, Hughes PD. 2014. Glaciers and rivers: Pleistocene uncoupling in a Mediterranean mountain karst. *Quaternary Science Reviews* **94**: 28–43.

Anderson RS, Repka JL, Dick GS. 1996. Explicit treatment of inheritance in dating depositional surface using in situ ^{10}Be and ^{26}Al. *Geology* **24**: 47–51.

Anthony DM, Granger DE. 2007a. A new chronology for the age of Appalachian erosional surfaces determined by cosmogenic nuclides in cave sediments. *Earth Surface Processes and Landforms* **32**: 874–887.

Anthony D, Granger D. 2007b. An empirical stream power formulation for knickpoint retreat in Appalachian Plateau fluviokarst. *Journal of Hydrology* **343**: 117–126.

Antón L, Mather AE, Stokes M, Munoz-Martin A, De Vicente G. 2015. Exceptional river gorge formation from unexceptional floods. *Nature Communications* **6**: 7963. https://doi.org/10.1038/ncomms8963.

Armitage JJ, Jones TD, Duller R, Whittaker AC, Allen PA. 2013. Temporal buffering of climate-driven sediment flux cycles by transient catchment response. *Earth and Planetary Science Letters* **369**: 200–210.

Attal M, Tucker GE, Whittaker AC, Cowie PA, Roberts GP. 2008. Modeling fluvial incision and transient landscape evolution: Influence of dynamic channel adjustment. *Journal of Geophysical Research: Earth Surface* **113**(F3).

Balco G, Rovey CW. 2008. An isochron method for cosmogenic-nuclide dating of buried soils and sediments. *American Journal of Science* **308**(10): 1083–1114.

Balco G, Soreghan GS, Sweet DE, Marra KR, Bierman PR. 2013. Cosmogenic-nuclide burial ages for Pleistocene sedimentary fill in Unaweep Canyon, Colorado, USA. *Quaternary Geochronology* **18**: 149–157.

Bahain J-J, Falguères C, Laurent M, Voinchet P, Dolo J-M, Antoine P, Tuffreau A. 2007. ESR chronology of the Somme River Terrace system and first human settlements in Northern France. *Quaternary Geochronology* **2**: 356–362.

Bayliss A, Brock F, Farid S, Hodder I, Southon J, Taylor RE. 2015. Getting to the bottom of it all: a Bayesian approach to dating the start of Çatalhöyük. *Journal of World Prehistory* **28**: 1–26.

Beckers A, Bovy B, Hallot E, Demoulin A. 2015. Controls on knickpoint migration in a drainage network of the moderately uplifted Ardennes Plateau, Western Europe. *Earth Surface Processes and Landforms* **40**(3): 357–374. https://doi.org/10.1002/esp.3638.

Benito G, Macklin MG, Zielhofer C, Jones AF, Machado MJ. 2015. Holocene flooding and climate change in the Mediterranean. *Catena* **130**: 13–33.

Berlin M, Anderson R. 2007. Modeling of knickpoint retreat on the Roan Plateau, western Colorado. *Journal of Geophysical Research* **112**: F03S06. https://doi.org/10.1029/2006JF000553.

Beven KJ. 1996. Equifinality and uncertainty in geomorphological modelling. In *The Scientific Nature of Geomorphology*, Rhoads BL, Thorn CE (eds). Wiley: Chichester; 289–313.

Bird MI, Ayliffe LK, Fifield LK, Turney CSM, Cresswell RG, Barrows TT, David B. 1999. Radiocarbon dating of 'old' charcoal using a wet oxidation-stepped combustion procedure. *Radiocarbon* **41**: 127–140.

Binnie SA, Phillips WM, Summerfield MA, Fifield LK. 2006. Sediment mixing and basin-wide cosmogenic nuclide analysis in rapidly eroding mountainous environments. *Quaternary Geochronology* **1**(1): 4–14.

Blöthe JH, Korup O. 2013. Millennial lag times in the Himalayan sediment routing system. *Earth and Planetary Science Letters* **382**: 38–46.

Blum M, Martin J, Milliken K, Garvin M. 2013. Paleovalley systems: insights from Quaternary analogs and experiments. *Earth-Science Reviews* **116**: 128–169.

Blum MD, Törnqvist TE. 2000. Fluvial responses to climate and sea-level change: a review and look forward. *Sedimentology* **47**: 2–48.

Boenigk W, Frechen M. 2006. The Pliocene and Quaternary fluvial archives of the Rhine system. *Quaternary Science Reviews* **25**(5): 550–574.

Bovy B, Braun J, Demoulin A. 2016. Soil production and hillslope transport in mid-latitudes during the last glacial–interglacial cycle: a combined data and modelling approach in northern Ardennes. *Earth Surface Processes and Landforms* **41**: 1758–1775.

Braun J, van der Beek P. 2004. Evolution of passive margin escarpments: what can we learn from low temperature thermochronometry? *Journal of Geophysical Research: Earth Surface* **109**(F4).

Briant RM, Bateman MD. 2009. Luminescence dating indicates radiocarbon age under-estimation in Late Pleistocene fluvial deposits from eastern England. *Journal of Quaternary Science* **24**: 916–927. https://doi.org/10.1002/jqs.1258.

Briant RM, Bates MR, Schwenninger J-L, Wenban-Smith FF. 2006. A long optically-stimulated luminescence dated Middle to Late Pleistocene fluvial sequence from the western Solent Basin, southern England. *Journal of Quaternary Science* **21**: 507–523.

Briant RM, Brock F, Langford HE, DeMarchi B, Penkman KEH, Schreve DC, Schwenninger J-L. 2018. Improving chronological control for environmental sequences from the last glacial period. *Quaternary Geochronology* **43**: 40–49.

Brocard GY, van der Beek PA, Bourlès DL, Siame LL, Mugnier JL. 2003. Long-term fluvial incision rates and postglacial river relaxation time in the French Western Alps from 10 Be dating of alluvial terraces with assessment of inheritance, soil development and wind ablation effects. *Earth and Planetary Science Letters* **209**(1): 197–214.

Bridgland D, Westaway R. 2008. Climatically controlled river terrace staircases: a worldwide Quaternary phenomenon. *Geomorphology* **98**: 285–315.

Burbank DW, Leland J, Fielding E, Anderson RS, Brozovic N, Reid MR, Duncan C. 1996. Bedrock incision, rock uplift and threshold hillslopes in the northwestern Himalayas. *Nature* **379**: 505–510.

Busschers FS, Cohen KM, Vandenberghe J, van Balen RT, Kasse C, Wallinga J, Weerts HJ. 2011. Comment on 'Causes, consequences and chronology of large-magnitude palaeoflows in Middle and Late Pleistocene river systems of northwest Europe' by Westaway and Bridgland (2010). *Earth Surface Processes and Landforms* **36**(13): 1836–1840.

Busschers FS, Wesselingh FP, Kars RH. 2014. Radiocarbon dating of Late Pleistocene marine shells from the southern North Sea. *Radiocarbon* **56**: 1151–1166.

Camporeale C, Perona P, Porporato A, Ridolfi L. 2007. Hierarchy of models for meandering rivers and related morphodynamic processes. *Reviews of Geophysics* **45**: RG1001. https://doi.org/10.1029/2005RG000185

Castelltort S, van den Driessche J. 2003. How plausible are high-frequency sediment supply-driven cycles in the stratigraphic record? *Sedimentary Geology* **157**(1): 3–13.

Chaussé C, Voinchet P, Bahain JJ, Connet N, Lhomme V, Limondin-Lozouët N. 2004. Middle and Upper Pleistocene evolution of the River Yonne valley (France). First results. *Quaternaire* **15**: 53–64.

Chiverrell RC, Thorndycraft VR, Hoffmann TO. 2011. Cumulative probability functions and their role in evaluating the chronology of geomorphological events during the Holocene. *Journal of Quaternary Science* **26**(1): 76–85.

Choi S-W, Preusser F, Radtke U. 2007. Dating of lower terrace sediments from the Middle Rhine area, Germany. *Quaternary Geochronology* **2**: 137–142.

Codilean AT, Bishop P, Hoey TB. 2006. Surface process models and the links between tectonics and topography. *Progress in Physical Geography* **30**: 307–333.

Cohen S, Willgoose GR, Hancock GR. 2008. A methodology for calculating the spatial distribution of the area-slope equation and the hypsometric integral within a catchment. *Journal of Geophysical Research: Earth Surface* **113**: F03027.

Cordier S, Adamson K, Delmas M, Calvet M, Harmand D. 2017. Of ice and water: Quaternary fluvial response to glacial forcing. *Quaternary Science Reviews* **166**: 57–73.

Cordier S, Frechen M, Harmand D. 2014. Dating fluvial erosion: fluvial response to climate change in the Moselle catchment (France, Germany) since the Late Saalian. *Boreas* **43**: 450–468.

Cordier S, Harmand D, Lauer T, Voinchet P, Bahain J-J, Frechen M. 2012. Geochronological reconstruction of the Pleistocene evolution of the Sarre valley (France and Germany) using OSL and ESR dating techniques. *Geomorphology* **165–166**: 91–106.

Coulthard TJ, Lewin J, Macklin MG. 2005. Modelling differential catchment response to environmental change. *Geomorphology* **69**: 222–241.

Coulthard TJ, van de Wiel MJ. 2013. Climate, tectonics or morphology: what signals can we see in drainage basin sediment yields? *Earth Surface Dynamics* **1**: 67–91.

Crosby BT, Whipple KX. 2006. Knickpoint initiation and distribution within fluvial networks: 236 waterfalls in the Waipaoa River, North Island, New Zealand. *Geomorphology* **82**(1): 16–38.

Davidson SK, Hartley AJ, Weissman GS, Nichols GJ, Scuderi LA. 2013. Geomorphic elements on modern distributive systems. *Geomorphology* **180–181**: 82–95.

Davis WM. 1922. Peneplains and the geographical cycle. *Geological Society of America Bulletin* **33**: 587–598.

Demoulin A, Beckers A, Hubert-Ferrari A. 2015. Patterns of Quaternary uplift of the Corinth rift southern border (N Peloponnese, Greece) revealed by fluvial landscape morphometry. *Geomorphology* **246**: 188–204.

Demoulin A, Bovy B, Rixhon G, Cornet Y. 2007. An automated method to extract fluvial terraces from digital elevation models: the Vesdre valley, a case study in eastern Belgium. *Geomorphology* **91**: 51–64.

Demoulin A, Mather A, Whittaker A. 2017. Fluvial archives, a valuable record of vertical crustal deformation. *Quaternary Science Reviews* **166**: 10–37.

Del Val M, Iriarte E, Arriolabengoa M, Aranburu A. 2015. An automated method to extract fluvial terraces from LIDAR based high resolution digital elevation models: the Oiartzun valley, a case study in the Cantabrian Margin. *Quaternary International* **364**: 35–43.

Erkens G. 2009. Sediment dynamics in the Rhine catchment – quantification of fluvial response to climate change and human impact. *Netherlands Geographical Studies* **388**: 278.

Erkens G, Cohen KM, Gouw MJP, Middelkoop H, Hoek WZ. 2006. Holocene sediment budgets of the Rhine Delta (The Netherlands): a record of changing sediment delivery. In *Sediment Dynamics and the Hydromorphology of Fluvial Systems*, Rowan JS, Duck RW, Werritty A (eds). IAHS Press: Wallingford; 406–415.

Erlanger ED, Granger DE, Gibbon RJ. 2012. Rock uplift rates in South Africa from isochron burial dating of fluvial and marine terraces. *Geology* **40**: 1019–1022.

Forzoni A, Storms JEA, Moreau J, Reimann T, de Jager G. 2014. Upstream versus downstream forcing: the response of the Golo River system, Corsica, to late quaternary climatic and eustatic oscillations. In *Proceedings of the Nederlands aardwetenschappelijk congres 12 (NAC 12), 8–9 April 2014, Veldhoven, The Netherlands*. Utrecht: Faculty Geosciences, University of Utrecht.

Fryirs KA, Brierley GJ, Preston NJ, Kasai M. 2007. Buffers, barriers and blankets: the (dis)connectivity of catchment-scale sediment cascades. *Catena* **70**: 49–67.

Fuchs M, Will M, Kunert E, Kreutzer S, Fischer M, Reverman R. 2012. The temporal and spatial quantification of Holocene sediment dynamics in a meso-scale catchment in northern Bavaria, Germany. *The Holocene* **21**: 1093–1104.

Garcia-Castellanos D, Vergés J, Gaspar-Escribano J, Cloetingh S. 2003. Interplay between tectonics, climate, and fluvial transport during the Cenozoic evolution of the Ebro Basin (NE Iberia). *Journal of Geophysical Research: Solid Earth* **108**(B7).

Geach MR, Stokes M, Telfer MW, Mather AE, Fyfe RM, Lewin S. 2014. The application of geospatial interpolation methods in the reconstruction of Quaternary landform records. *Geomorphology* **216**: 234–246.

Geach MR, Viveen W, Mather AE, Telfer MW, Fletcher WJ, Stokes M, Peyron O. 2015. An integrated field and numerical modelling study of controls on Late Quaternary fluvial landscape development (Tabernas, SE Spain). *Earth Surface Processes and Landforms* **40**(14): 1907–1926. https://doi.org/10.1002/esp.3768.

Gibbard PL, Lewin J. 2002. Climate and related controls on interglacial fluvial sedimentation in lowland Britain. *Sedimentary Geology* **151**: 187–210.

Godard V, Bourlès DL, Spinabella F, Burbank DW, Bookhagen B, Fisher GB, Léanni L. 2014. Dominance of tectonics over climate in Himalayan denudation. *Geology* **42**(3): 243–246.

Granger DE, Fabel D, Palmer A. 2001. Pliocene-Pleistocene incision of the Green River, Kentucky, determined from radioactive decay of cosmogenic ^{26}Al and ^{10}Be in Mammoth Cave sediments. *GSA Bulletin* **113**: 825–836.

Grimm V, Frank K, Jeltsch F, Brandl R, Uchmanski J, Wissel C. 1996. Pattern-oriented modelling in population ecology. *Science of the Total Environment* **183**: 151–166. https://doi.org/10.1016/0048-9697(95)04966-5.

Grimm V, Railsback SF. 2012. Pattern-oriented modelling: a multi-scope for predicitve systems ecology. *Philosophical Transactions of the Royal Society B* **367**: 298–310.

Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, Thulke H-H, Weiner J, Wiegand T, DeAngelis DL. 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *Science* **310**: 987–991. https://doi.org/10.1126/science.1116681.

Hajek EA, Heller PL, Sheets BA. 2010. Significance of channel-belt clustering in alluvial basins. *Geology* **38**: 535–538.

Higham TFG, Jacobi RM, Bronk Ramsey C. 2006. AMS radiocarbon dating of ancient bone using ultrafiltration. *Radiocarbon* **48**: 179–195.

Humphrey NF, Heller PL. 1995. Natural oscillations in coupled geomorphic systems: an alternative origin for cyclic sedimentation. *Geology* **23**: 499–502.

Jerolmack DJ, Paola C. 2010. Shredding of environmental signals by sediment transport. *Geophysical Research Letters* **37**(19): L19401. https://doi.org./10.1029/2010GL044638.

Johnson JB, Omland KS. 2004. Model selection in ecology and evolution. *Trends in Ecology & Evolution* **19**: 101–108.

Jones AF, Brewer PA, Johnstone E, Macklin MG. 2007. High-resolution interpretative geomorphological mapping of river valley environments using airborne LiDAR data. *Earth Surface Processes and Landforms* **32**: 1574–1592.

Jones AF, Macklin MG, Benito G. 2015. Meta-analysis of Holocene fluvial sedimentary archives: a methodological primer. *Catena* **130**: 3–12.

Kars RH, Busschers FS, Wallinga J. 2012. Validating post IR-IRSL dating on K-feldspars through comparison with quartz OSL ages. *Quaternary Geochronology* **12**: 74–86.

King GE, Herman F, Guralnik B. 2016. Northward migration of the eastern Himalayan syntaxis revealed by OSL thermochronometry. *Science* **353**: 800–804.

Kock S, Kramers JD, Preusser F, Wetzel A. 2009. Dating of Late Pleistocene terrace deposits of the River Rhine using Uranium series and luminescence methods: potential and limitations. *Quaternary Geochronology* **4**: 363–373.

Lamb MP, Howard AD, Johnson J, Whipple KX, Dietrich WE, Perron JT. 2006. Can springs cut canyons into rock? *Journal of Geophysical Research: Planets* **111**(E7): E07002, https://doi.org/10.1029/2005JE002663.

Larsen L, Thomas C, Eppinga M, Coulthard T. 2014. Exploratory modeling: extracting causality from complexity. *Eos, Transactions American Geophysical Union* **95**: 285–286.

Lewis SG, Maddy D, Scaife RG. 2001. The fluvial system response to abrupt climate change during the last cold stage: the Upper Pleistocene River Thames fluvial succession at Ashton Keynes, UK. *Global and Planetary Change* **28**: 341–359.

Lytle DA. 2002. Flash floods and aquatic insect life-history evolution: evaluation of multiple models. *Ecology* **83**: 370–385.

Macklin MG, Fuller IC, Jones AF, Bebbington M. 2012a. New Zealand and UK Holocene flooding demonstrates interhemispheric climate asynchrony. *Geology* **40**: 775–778.

Macklin MG, Jones AF, Lewin J. 2010. River response to rapid Holocene environmental change: evidence and explanation in British catchments. *Quaternary Science Reviews* **29**: 1555–1576.

Macklin MG, Lewin J, Jones AF. 2013. River entrenchment and terrace formation in the UK Holocene. *Quaternary Science Reviews* **76**: 194–206.

Macklin MG, Lewin J, Woodward JC. 2012b. The fluvial record of climate change. *Philosophical Transactions of the Royal Society A* **370**: 2143–2172.

Maddy D. 1997. Uplift-driven valley incision and river terrace formation in southern England. *Journal of Quaternary Science* **12**: 539–545.

Masson-Delmotte V, Schulz M, Abe-Ouchi A, Beer J, Ganopolski A, González Rouco JF, Jansen E, Lambeck K, Luterbacher J, Naish T, Osborn T, Otto-Bliesner B, Quinn T, Ramesh R, Rojas M, Shao X, Timmermann A. 2013. Information from Paleoclimate Archives. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). Cambridge University Press: Cambridge.

Mather AE, Stokes M. 2016. Extracting palaeoflood data from coarse-grained Pleistocene river terrace archives: an example from SE Spain. *Earth Surface Processes and Landforms* **41**: 1991–2004. https://doi.org/10.1002/esp.4001.

Mather AE, Stokes M, Griffiths JS. 2002. Quaternary landscape evolution: A framework for understanding contemporary erosion, southeast Spain. *Land Degradation and Development* **13**: 89–109.

Mather AE, Stokes M, Whitfield E. 2017. River terraces and alluvial fans: the case for an integrated Quaternary fluvial archive. *Quaternary Science Reviews* **166**: 74–90.

Métivier F. 1999. Diffusive like buffering and saturation of large rivers. *Physical Review E* **60**: 5827.

Michaelides K, Wainwright J. 2002. Modelling the effects of hillslope-channel coupling on catchment hydrological response. *Earth Surface Processes and Landforms* **27**: 1441–1457.

Milliman JD, Syvitski JP. 1992. Geomorphic/tectonic control of sediment discharge to the ocean: the importance of small mountainous rivers. *The Journal of Geology* **100**: 525–544.

Mulligan M, Wainwright J. 2004. Modelling and model building. In *Environmental Modelling: Finding Simplicity in Complexity*, Mulligan M, Wainwright J (eds). Wiley: Chichester; 7–73.

Murton JB, Kolstrup E. 2003. Ice-wedge casts as indicators of palaeotemperatures: precise proxy or wishful thinking? *Progress in Physical Geography* **27**: 155–170.

Nichols GJ, Fisher JA. 2007. Processes, facies and architecture of fluvial distributary system deposits. *Sedimentary Geology* **195**: 75–90.

Nicholas AP, Quine TA. 2010. Quantitative assessment of landform equifinality and palaeoenvironmental reconstruction using geomorphic models. *Geomorphology* **121**: 167–183.

Nicholas AP. 2013. Modelling the continuum of river channel patterns. *Earth Surface Processes and Landforms* **38**: 1187–1196.

Norton KP, Abbühl LM, Schlunegger F. 2010. Glacial conditioning as an erosional driving force in the Central Alps. *Geology* **38**: 655–658.

Oreskes N, Shrader-Frechette K, Belitz K. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* **263**: 641–646.

Parsons AJ. 2012. How useful are catchment sediment budgets? *Progress in Physical Geography* **36**: 60–71.

Parsons AJ, Wainwright J, Abrahams AD, Simanton JR. 1997. Distributed dynamic modelling of interrill overland flow. *Hydrological Processes* **11**: 1833–1859.

Perron JT, Royden L. 2013. An integral approach to bedrock river profile analysis. *Earth Surface Processes and Landforms* **38**(6): 570–576.

Phillips JT. 2015. Badass geomorphology. *Earth Surface Processes and Landforms* **40**: 22–33.

Pierce JL, Meyer GA, Rittenour T. 2011. The relation of Holocene fluvial terraces to changes in climate and sediment supply, South Fork Payette River, Idaho. *Quaternary Science Reviews* **30**: 628–645.

Reusser LJ, Bierman PR, Pavich MJ, Zen E, Larsen J, Finkel R. 2004. Rapid Late Pleistocene incision of Atlantic passive-margin river gorges. *Science* **305**: 499–502.

Rixhon G, Braucher R, Bourlès D, Siame L, Bovy B, Demoulin A. 2011. Quaternary river incision in NE Ardennes (Belgium) – insights from ^{10}Be/^{26}Al dating of river terraces. *Quaternary Geochronology* **6**: 273–284.

Rixhon G, Briant RM, Cordier S, Duval M, Jones A, Scholz D. 2017. Revealing the pace of river landscape evolution during the Quaternary: recent developments in numerical dating methods. *Quaternary Science Reviews* **166**: 91–113.

Savi S, Norton K, Picotti V, Brardinoni F, Akçar N, Kubik PW, Schlunegger F. 2014. Effects of sediment mixing on 10Be concentrations in the Zielbach catchment, central-eastern Italian Alps. *Quaternary Geochronology* **19**: 148–162.

Schaller M, Ehlers TA. 2006. Limits to quantifying climate driven changes in denudation rates with cosmogenic radionuclides. *Earth and Planetary Science Letters* **248**: 153–167.

Schaller M, von Blanckenburg F, Hovius N, Kubik PW. 2001. Large-scale erosion rates from *in situ*-produced cosmogenic nuclides in European river sediments. *Earth and Planetary Science Letters* **188**: 441–458.

Schaller M, von Blanckenburg F, Veldkamp A, Tebbens LA, Hovius N, Kubik PW. 2002. A 30000 yr record of erosion rates from cosmogenic 10Be in Middle European river terraces. *Earth and Planetary Science Letters* **204**: 307–320.

Schillereff DN, Chiverrell RC, Macdonald N, Hooke JM. 2015. Flood stratigraphies in lake sediments: a review. *Earth-Science Reviews* **135**: 17–37.

Schmitz B, Pujalte V. 2007. Abrupt increase in seasonal extreme precipitation at the Paleocene–Eocene boundary. *Geology* **35**: 215–218.

Schoorl JM, Temme AJAM, Veldkamp A. 2014. Modelling centennial sediment waves in an eroding landscape – catchment complexity. *Earth Surface Processes and Landforms* **39**: 1526–1537.

Schulte L, Julià R, Burjachs F, Hilgers A. 2008. Middle Pleistocene to Holocene geochronology of the River Aguas terrace sequence (Iberian Peninsula): fluvial response to Mediterranean environmental change. *Geomorphology* **98**: 13–33.

Schumm SA. 1973. Geomorphic thresholds and complex response of drainage systems. *Fluvial Geomorphology* **6**: 69–85.

Seong YB, Owen LA, Bishop MP, Bush A, Clendon P, Copland L, Finkel RC, Kamp U, Shroder JF, Jr. 2008. Rates of fluvial bedrock incision within an actively uplifting orogen: Central Karakoram Mountains, northern Pakistan. *Geomorphology* **97**: 274–286.

Sharp WD, Ludwig KR, Chadwick OA, Amundson R, Glaser LL. 2003. Dating fluvial terraces by ^{230}Th/U on pedogenic carbonate, Wind River Basin, Wyoming. *Quaternary Research* **59**: 139–150.

Small EE, Anderson RS, Repka JL, Finkel R. 1997. Erosion rates of alpine bedrock summit surfaces deduced from in situ 10Be and 26Al. *Earth and Planetary Science Letters* **150**: 413–425.

Snyder NP, Whipple KX, Tucker GE, Merritts DJ. 2003. Channel response to tectonic forcing: field analysis of stream morphology and hydrology in the Mendocino triple junction region, northern California. *Geomorphology* **53**: 97–127.

Stange KM, van Balen RT, Garcia-Castellanos D, Cloetingh S. 2016. Numerical modelling of Quaternary terrace staircase formation in the Ebro foreland basin, southern Pyrenees, NE Iberia. *Basin Research* **28**: 124–146.

Stokes M, Griffiths JS, Mather AE. 2012. Palaeoflood estimates of Pleistocene coarse grained river terrace landforms (Rio Almanzora, SE Spain). *Geomorphology* **149**: 11–26.

Stokes M, Mather AE, Harvey AM. 2002. Quantification of river capture induced base-level changes and landscape development, Sorbas Basin, SE Spain. In *Sediment Flux to Basins: Causes, Controls and Consequences*, Jones SJ, Frostick LE (eds) , Vol. **191**. The Geological Society: London; 23–35.London Special Publication

Talling PJ. 1998. How and where do incised valleys form if sea level remains above the shelf edge? *Geology* **26**: 87–90.

Taylor KE, Stouffer RJ, Meehl GA. 2012. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society* **93**: 485–498.

Temme AJAM, Armitage J, Attal M, van Gorp W, Coulthard T, Schoorl JM. 2017. Choosing and using landscape evolution models to inform field-stratigraphy and landscape reconstruction studies. *Earth Surface Processes and Landforms* **42**: 2167–2183.

Temme AJAM, Veldkamp A. 2009. Multi-process Late Quaternary landscape evolution modelling reveals lags in climate response over small spatial scales. *Earth Surface Processes and Landforms* **34**: 573–589.

Thorndycraft VR, Benito G. 2006. The Holocene fluvial chronology of Spain: evidence from a newly compiled radiocarbon database. *Quaternary Science Reviews* **25**: 223–234.

Toms PS. 2013. Optical dating of Broom. In *Quaternary History & Palaeolithic Archaeology in the Axe Valley at Broom*, Hosfield R, Green C, Chambers J (eds). Oxbow: Oxford.

Tucker GE. 2009. Natural experiments in landscape evolution. *Earth Surface Processes and Landforms* **34**: 1450–1459.

Valla P, Shuster D, van der Beek P. 2011. Significant increase in relief of the European Alps during mid-Pleistocene glaciations. *Nature Geoscience* **4**: 688–692.

van Balen RT, Busschers FS, Tucker GE. 2010. Modeling the response of the Rhine–Meuse fluvial system to Late Pleistocene climate change. *Geomorphology* **114**: 440–452.

Van De Wiel MJ, Coulthard TJ, Macklin MG, Lewin J. 2011. Modelling the response of river systems to environmental change: Progress, problems and prospects for palaeo-environmental reconstructions. *Earth-Science Reviews* **104**: 167–185.

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. *Journal of Geophysical Research: Solid Earth* **108**(B6): 2309. https://doi.org/10.1029/2002JB002125.

van Gorp W, Temme AJAM, Veldkamp A, Schoorl JM. 2015. Modelling long-term (300 ka) upland catchment response to multiple lava damming events. *Earth Surface Processes and Landforms* **40**: 888–900.

Vandenberghe J. 1995. Timescales, climate and river development. *Quaternary Science Reviews* **14**: 631–638.

Vandenberghe J. 2008. The fluvial cycle at cold–warm–cold transitions in lowland regions: a refinement of theory. *Geomorphology* **98**: 275–284.

Veldkamp A, Baartman JEM, Coulthard TJ, Maddy D, Schoorl JM, Storms JEA, Temme AJAM, van Balen RT, van De Wiel MJ, van Gorp W, Viveen W, Westaway R, Whittaker AC. 2017. Two decades of numerical modelling to understand long term fluvial archives: advances and future perspectives. *Quaternary Science Reviews* **166**: 177–187.

Veldkamp A, Candy I, Jongmans AG, Maddy D, Demir T, Schoorl JM, Schreve D, Stemerdink C, van der Schriek T. 2015. Reconstructing Early Pleistocene (1.3 Ma) terrestrial environmental change in western Anatolia: did it drive fluvial terrace formation? *Palaeogeography Palaeoclimatology Palaeoecology* **417**: 91–104.

Veldkamp A, Kroonenberg SB, Heijnis H, van den Berg van Saparoea R. 2004. The suitability of dated travertines as a record of fluvial incision: Allier (France) floodplain dynamics during the late Quaternary. *Quaternaire* **15**: 159–165.

Veldkamp A, Schoorl JM, Viveen W. 2016. Predicting location-specific fluvial archives with the FLUVER2 model for the late Quaternary Allier River (France). *Earth Surface Processes and Landforms* **41**: 2256–2268.

Veldkamp A, Tebbens LA. 2001. Registration of abrupt climate changes within fluvial systems: insights from numerical modelling experiments. *Global and Planetary Change* **28**: 129–144.

Viveen W, Schoorl JM, Veldkamp A, van Balen RT. 2014. Modelling the impact of regional uplift and local tectonics on fluvial terrace preservation. *Geomorphology* **210**: 119–135.

Von Blanckenburg F. 2005. The control mechanisms of erosion and weathering at basin scale from cosmogenic nuclides in river sediment. *Earth and Planetary Science Letters* **237**: 462–479.

Wagner T, Fabel D, Fiebig M, Häuselmann P, Sahy D, Xu S, Stüwe K. 2010. Young uplift in the non-glaciated parts of the Eastern Alps. *Earth and Planetary Science Letters* **295**: 159–169.

Wainwright J. 2006. Degrees of separation: hillslope-channel coupling and the limits of palaeohydrological reconstruction. *Catena* **66**: 93–106.

Wainwright J. 2007. Can modelling enable us to understand the rôle of humans in landscape evolution? *Geoforum* **39**: 659–674. https://doi.org/10.1016/j.geoforum.2006.09.011.

Wainwright J, Mulligan M (eds). 2005. *Environmental Modelling: Finding Simplicity in Complexity*. Wiley: Chichester.

Westaway R, Bridgland DR. 2010. Causes, consequences and chronology of large-magnitude palaeoflows in Middle and Late Pleistocene river systems of northwest Europe. *Earth Surface Processes and Landforms* **35**: 1071–1094.

Westaway R, Bridgland DR. 2011. Reply to comments by FS Busschers, KM Cohen, J Vandenberghe, RT Van Balen, C. Kasse, J. Wallinga, and HJT Weerts on 'Causes, consequences and chronology of large-magnitude palaeoflows in Middle and Late Pleistocene river systems of northwest Europe', by Rob Westaway and David R. Bridgland (2010). *Earth Surface Processes and Landforms* **36**: 1841–1846.

Whipple KX, Tucker GE. 1999. Dynamics of the stream-power river incision model: implications for height limits of mountain ranges, landscape response timescales, and research needs. *Journal of Geophysical Research: Solid Earth* **104**(B8): 17661–17674.

Whittaker AC, Attal M, Allen PA. 2010. Characterising the origin, nature and fate of sediment exported from catchments perturbed by active tectonics. *Basin Research* **22**: 809–828.

Whittaker A, Attal M, Cowie P, Tucker G, Roberts G. 2008. Decoding temporal and spatial patterns of fault uplift using transient river long profiles. *Geomorphology* **100**: 506–526.

Whittaker A, Boulton S. 2012. Tectonic and climatic controls on knickpoint retreat rates and landscape response times. *Journal of Geophysical Research* **117**: F02024. DOI: https://doi.org/10.1029/2011JF002157

Whittaker A, Cowie P, Attal M, Tucker G, Roberts G. 2007. Contrasting transient and steady-state rivers crossing active normal faults: new field observations from the central Apennines, Italy. *Basin Research* **19**: 529–556.

Willet SD, Fisher D, Fuller C, En-Cho Y, Chia-Yu L. 2003. Erosion rates and orogenic wedge kinematics in Taiwan inferred from fission-track thermochronology. *Geology* **31**: 945–948.

Willenbring JK, Gasparini NM, Crosby BT, Brocard G. 2013. What does a mean mean? The temporal evolution of detrital cosmogenic denudation rates in a transient landscape. *Geology* **41**: 1215–1218.

Wittmann H, Von Blanckenburg F, Guyot JL, Maurice L, Kubik PW. 2009. From source to sink: preserving the cosmogenic ^{10}Be-derived denudation rate signal of the Bolivian Andes in sediment of the Beni and Mamoré foreland basins. *Earth and Planetary Science Letters* **288**: 463–474.

Yanites BJ, Tucker GE, Anderson RS. 2009. Numerical and analytical models of cosmogenic radionuclide dynamics in landslide-dominated drainage basins. *Journal of Geophysical Research: Earth Surface* **114**(F1): F01007, https://doi.org/10.1029/2008JF001088.

Zhu S, Wu Z, Zhao X, Li J, Xiao K. 2014. Ages and genesis of terrace flights in the middle reaches of the Yarlung Zangbo River, Tibetan Plateau, China. *Boreas* **43**: 485–504.