JGR Earth Surface

RESEARCH ARTICLE
10.1029/2020JF005795

Projections of Landscape Evolution on a 10,000 Year Timescale With Assessment and Partitioning of Uncertainty Sources

Katherine R. Barnhart[1,2,3], Gregory E. Tucker[1,2], Sandra G. Doty[4], Rachel C. Glade[2,5,6], Charles M. Shobe[1,2,7], Matthew W. Rossi[1,8], and Mary C. Hill[9]

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA, [2]Department of Geological Sciences, University of Colorado, Boulder, CO, USA, [3]Now at U.S. Geological Survey, Landslide Hazards Program, Golden, CO, USA, [4]Denver, CO, USA, [5]Institute for Arctic and Alpine Research, University of Colorado, Boulder, CO, USA, [6]Los Alamos National Laboratory, Los Alamos, NM, USA, [7]Helmholz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany, [8]Earth Lab, University of Colorado, Boulder, CO, USA, [9]Department of Geology, University of Kansas, Lawrence, KS, USA

Abstract Long-term erosion can threaten infrastructure and buried waste, with consequences for management of natural systems. We develop erosion projections over 10 ky for a 5 km^2 watershed in New York, USA. Because there is no single landscape evolution model appropriate for the study site, we assess uncertainty in projections associated with *model structure* by considering a set of alternative models, each with a slightly different governing equation. In addition to model structure uncertainty, we consider the following uncertainty sources: selection of a final model set; each model's parameter values estimated through calibration; simulation boundary conditions such as the future incision of downstream rivers and future climate; and initial conditions (e.g., site topography which may undergo near-term anthropogenic modification). We use an analysis-of-variance approach to assess and partition uncertainty in projected erosion into the variance attributable to each source. Our results suggest one sixth of the watershed will experience erosion exceeding 5 m in the next 10 ky. Uncertainty in projected erosion increases with time, and the projection uncertainty attributable to each source manifests in a distinct spatial pattern. Model structure uncertainty is relatively low, which reflects our ability to constrain parameter values and reduce the model set through calibration to the recent geologic past. Beyond site-specific findings, our work demonstrates what information prediction-under-uncertainty studies can provide about geomorphic systems. Our results represent the first application of a comprehensive multi-model uncertainty analysis for long-term erosion forecasting.

Plain Language Summary Erosion of ground material is a hazard to buildings and other infrastructure, and can pose an environmental risk when it occurs in areas such as radioactive waste repositories and post-industrial sites. We make projections for erosion over the next 10,000 years and assess uncertainty sources at a 5 km^2 watershed in New York state. Natural systems, like the study site, are not as well understood as engineered systems. However, information from studies like this one can provide useful insights to guide management decisions. Parts of the watershed experienced up to 50 m of erosion over the past 13,000 years. The type of model used to simulate land surface evolution over thousands of years is a Landscape Evolution Model (LEM). We use a set of alternative LEMs identified by prior work as successful in simulating this watershed from 13,000 years ago to the present. We consider uncertainty in future climate, incision of downstream rivers, the models used and their parameters, and small changes to the initial topography. We find that just under 20% of the study site will experience erosion exceeding 5 m at +10,000 years. The most important uncertainty source varies across the watershed and increases with time.

Key Points:

- Projections of erosion to 10 ky and uncertainty assessment demonstrate that the dominant uncertainty sources vary in space and time
- Two alternative approaches to model selection result in similar projections and uncertainty at 10 ky
- Variation across alternative models is subtle and initial conditions influence the location of gullies that drain low-relief areas

Supporting Information:

- Supporting Information S1

Correspondence to:
K. R. Barnhart,
krbarnhart@usgs.gov

Citation:
Barnhart, K. R., Tucker, G. E., Doty, S. G., Glade, R. C., Shobe, C. M., Rossi, M. W., & Hill, M. C. (2020). Projections of landscape evolution on a 10,000 year timescale with assessment and partitioning of uncertainty sources. *Journal of Geophysical Research: Earth Surface*, *125*, e2020JF005795. https://doi.org/10.1029/2020JF005795

Received 21 JUL 2020
Accepted 19 NOV 2020

1. Introduction

Erosion poses a threat to roads, buildings, and other infrastructure, and to containment of long-lived hazardous materials produced by modern society. For the latter, radioactive waste repositories, mine waste deposits, and landfills are the most concerning. Performance assessment for such sites often requires

long-term (centuries to millennia) estimates of contaminant release by erosion. Making such long-term projections for natural sites is a bold endeavor, and it is important to acknowledge and quantify as many uncertainty sources as possible. Uncertainty sources include, for example, future environmental conditions such as climate and vegetation, human activity and management (or lack thereof), parameters that describe material properties and process rates, and initial conditions such as site topography. In addition, because mathematical and computational models imperfectly approximate natural systems, there is uncertainty in any model's representation of the operative geomorphic processes. Efforts to minimize projection uncertainty and identify data needs require: (1) identifying uncertainty sources, and (2) quantitatively constraining their relative contributions to the uncertainty in projections.

Long-term erosion forecasting is a relatively new field, and systematic approaches to uncertainty quantification have not yet been widely formulated or adopted. For example, prior attempts to project erosion over time spans of a century or more have been performed for uranium mine tailings (Evans et al., 2000; Hancock et al., 2008, 2015; Hancock, Coulthard, & Lowry, 2016; Hancock, Lowry, & Coulthard, 2016; Hancock, Lowry, & Saynor, 2017; Lowry et al., 2013; Willgoose & Riley, 1998), other mining-related sites (Hancock et al., 2019; Sharmeen & Willgoose, 2007; Slingerland et al., 2018), landfills (Neuhold & Nachtnebel, 2011), and radioactive waste repositories (Atchley et al., 2019; French et al., 2008; C. J. Wilson et al., 2005). However, efforts to quantify uncertainty associated with long-term erosion projections have been relatively limited. Where uncertainty estimates have been attempted, they are often based on sensitivity analysis (El-Ghonemy et al., 2005; Hancock, Verdon-Kidd, & Lowry, 2017; Hancock, Coulthard, & Lowry, 2016; Sharmeen & Willgoose, 2007), on a small number of alternative scenarios (Atchley et al., 2019; Hancock, Lowry, & Coulthard, 2016; C. J. Wilson et al., 2005), or on comparisons between two alternative numerical model algorithms (Hancock, 2006; Hancock et al., 2015; Hancock, Lowry, & Saynor, 2017). There remains a need for projections that include quantitative, probabilistic estimates of uncertainty and its spatial and temporal variation at a site of interest. Comprehensive uncertainty quantification should encompass future environmental conditions, process models and their assumptions, and model parameters (El-Ghonemy et al., 2005). While the characterization of uncertainty sources is most pressing for geohazards like the case study here, the landscape evolution community as a whole would strongly benefit from formalized methodologies that identify limitations to the long-term projections we collectively produce.

To address these needs, we develop and apply a method for long-term erosion projections with quantitative, probabilistic, and spatially distributed uncertainty estimates. The method includes multi-model analysis to address uncertainty arising from incomplete scientific knowledge (Foglia et al., 2013). As a case study, we focus on the West Valley Demonstration Project site in New York state, USA (Figure 1). At this site, which lies within a ~5 km^2 watershed, concerns about contaminant release from buried waste in the distant future due to progressive erosion creates a need for long-term (10 ky) projections of erosion together with quantitative uncertainty estimates. Because potential contaminants are located at varying depths and geographic positions, it is especially important to estimate how the total depth of projected erosion varies in space and time.

The site's recent geological history is relatively well known. Advances of the Laurentide ice sheet during the last glacial cycle left behind a low-relief, valley-filling surface that caps a sequence of glaciogenic sediments (M. Wilson & Young, 2018; Young et al., 2020). After the final deglaciation of the area at ~13 ka, the stream network incised this surface, creating a network of canyons, ravines, and gullies (Figure 1). These include Buttermilk Creek, the major stream that serves as the outlet to the study watershed, which has incised up to ~50 m into young glacial deposits (Fakundiny, 1985; M. Wilson & Young, 2018; Young et al., 2020). While most of Buttermilk Creek's recent incision has been into the glaciogenic sediments, sandstone bedrock is now exposed in parts of the channel (M. Wilson & Young, 2018).

Barnhart, Tucker, et al. (2020b, 2020c, 2020d) used this geomorphic natural experiment to conceptualize and test a set of 37 alternative Landscape Evolution Models (LEMs), each of which represents a plausible representation of the geomorphic processes for the late glacial to Holocene. LEMs were calibrated by running forward in time from a reconstruction of the 13 ka paleotopography to the present day. Calibrated models were validation-tested on a nearby watershed of similar size, geology, and geomorphic history. Here we use the best-performing subset of these models to create projections of erosion depth from the present to +10 ky.

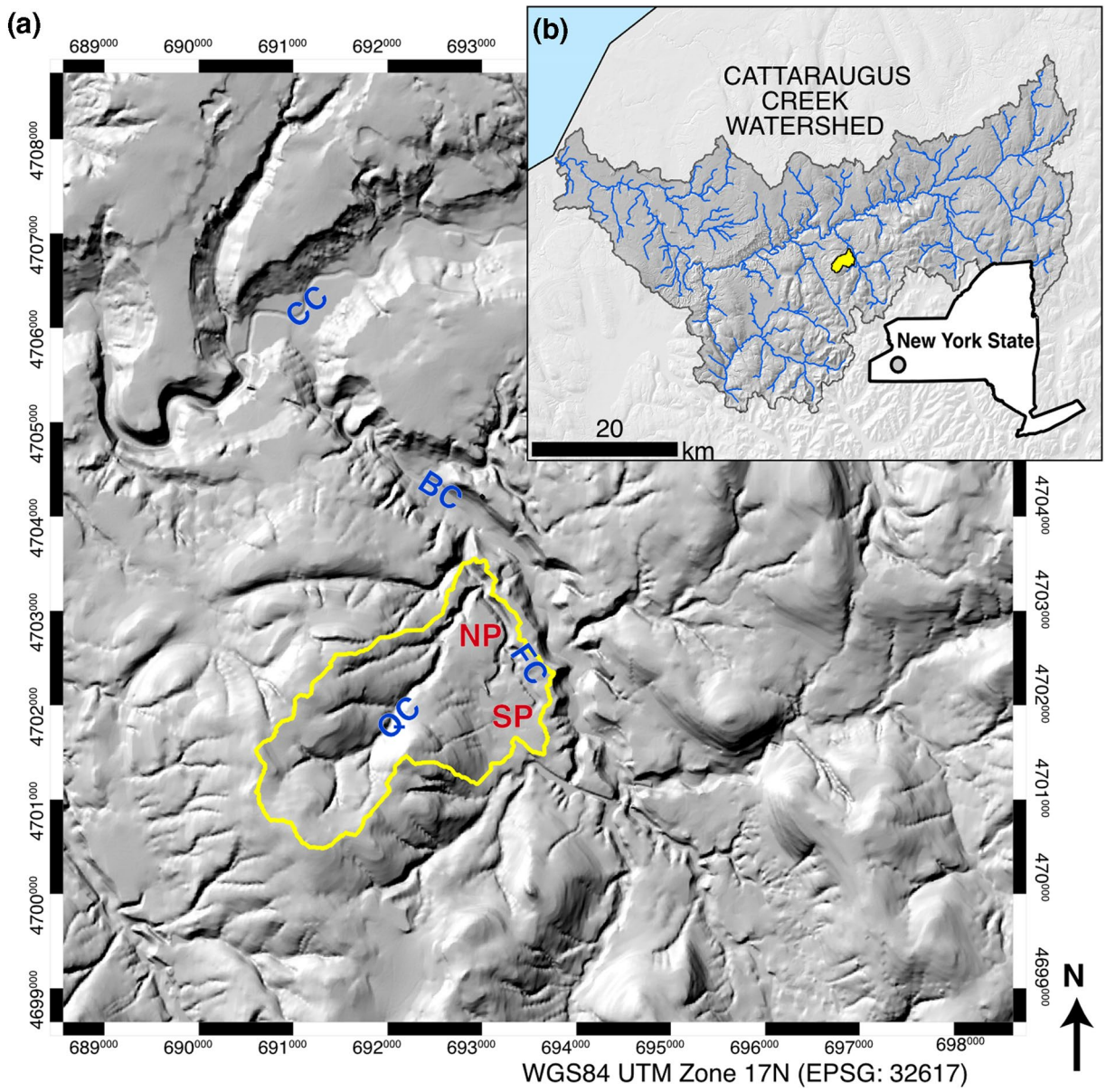

Figure 1. Study watershed (yellow polygon) in the context of the surrounding terrain (a) and within New York State and the Cattaraugus Creek watershed (b). Major drainages (blue): CC, Cattaraugus Creek; BC, Buttermilk Creek; FC, Franks Creek; QC, Quarry Creek. Remnant plateau surfaces (red): NP, North plateau; SP, South plateau.

Projections are accompanied by quantitative, spatially distributed estimates of uncertainty arising from five input sources: model structure, parameter calibration, initial topography, future climate, and future incision on Buttermilk Creek. Although the above list does not cover all possible uncertainty sources—it omits, for example, low-probability, high-impact events such as a single massive landslide or a major stream capture from an adjacent basin—it spans those sources that can be readily identified and quantified. To our knowledge, this is the first application of a comprehensive, spatially distributed, and multi-model uncertainty analysis to long-term erosion forecasting.

2. Model Set

Because there is no single established LEM formulation, but rather a variety of alternative approaches, the model equations are an uncertainty source, reflecting limitations of present scientific understanding (Willgoose, 2018). Barnhart, Tucker, et al. (2020b) constructed a set of 37 plausible LEMs for the study watershed,

Table 1
Summary of Eight Considered Models

Model name	N_p	Description[a]	Calibration[b]				Validation			
			AIC_c	Δ_i	p_i	Rank[c]	AIC_c	Δ_i	p_i	Rank
BasicRt	3	Differentiation between erodibility coefficients for shale bedrock and glacial till (Rt)	160.6	36.1	0.000	5	509.5	83.0	0.00	5
BasicRtTh	5	Rt and fluvial erodibility threshold (Th)	138.2	13.7	0.001	2	435.7	9.2	0.01	2
BasicRtSs	3	Rt and shear stress exponents ($m = 1/3$, $n = 2/3$) in lieu of stream power exponents (Ss)	160.5	36.0	0.000	4	519.9	93.4	0.00	8
BasicDdRt	5	Rt and depth dependent fluvial erodibility threshold (Dd)	157.8	33.4	0.000	3	426.6	0.0	0.99	1
BasicHyRt	4	Rt and a hybrid erosion and deposition form of fluvial erosion (Hy)	162.9	38.4	0.000	6	484.9	58.3	0.00	4
BasicChRt	4	Rt and nonlinear hillslope sediment transport (Ch)	166.2	41.8	0.000	8	514.5	87.9	0.00	6
BasicRtVs	4	Rt and variable source area hydrology (Vs)	165.7	41.3	0.000	7	519.4	92.8	0.00	7
BasicChRtTh	6	Ch, Rt, and Th	124.4	0.0	0.999	1	468.2	41.6	0.00	3

[a]Model name reflects a combination of two-letter keys which are provided in parens when first introduced in the description. [b]AIC_c: Akaike information criterion; Δ_i: AIC_c difference (Equation 4); p_i: posterior model probability (Equation 3). [c]Note that models with high probability have the best (lowest) ranks.

representing variations in *model structure* (Table S1). The simplest of these models, called "Basic," uses a common formulation: uniform lithology, water erosion based on stream power, hillslope soil transport by linear diffusion, and deterministic and uniform hydrology (e.g., Duvall & Tucker, 2015; Miller et al., 2007; Miller & Slingerland, 2006; Pelletier, 2010; Perron et al., 2009). The Basic model has the following governing equation:

$$\frac{\partial \eta}{\partial t} = -KA^{1/2}S + D\nabla^2\eta, \tag{1}$$

where η is the surface elevation, K and D are the rate constants, S is the positive-downward slope gradient in the direction of steepest descent, and A is the upstream drainage area. The first term on the right represents water erosion by channelized flow, and the second represents gravitational soil transport by diffusive processes.

The numerical models were built using the Landlab Toolkit (Barnhart, Hutton, et al., 2020; Hobley et al., 2017) and distributed via the terrainbento Python package (Barnhart, Glade, et al., 2019). Each simulation used a global timestep of 10 years, and the model grid was discretized with a horizontal cell spacing of 7.3 m (24 ft). Each model used between three and six input parameters (Table 1 and Table S2).

Each model was independently calibrated using Dakota (Adams et al., 2017a, 2017b; Dennis Jr. et al., 1981; Jones et al., 1998) with a simulation period from 13 ka to the present (Barnhart, Tucker, et al., 2020c). Such a calibration requires a formal statement of goodness of fit, which we established in Barnhart, Tucker, et al. (2020b, 2020c). Model fit was assessed by a direct topographic difference between modern and end-of-simulation topography. We reduced the number of model-data comparison metrics from the total number of grid cells ($\sim 10^5$) to 20 by grouping grid cells into domains with similar processes. Fit metrics were calculated using the umami Python package (Barnhart, Hutton, & Tucker, 2019).

3. Model Selection and Probabilities

The considered models (Table 1) represent a set of alternative hypotheses for system representation (e.g., Clark et al., 2015; Poeter & Anderson, 2005). Once calibrated, each model—along with estimated parameter values—represents the best possible version of that model, conditional on the choice of model fit metric, for representing the considered system. Given such a set of calibrated models and a desire to make projections with them, the modeler is faced with the challenge of how much confidence to assign to each model (model weighting) and which models to use for subsequent analysis (model selection). It is intuitive that calibrated

models that perform better should be given more weight (higher probability) than those that perform more poorly, all else equal. We thus seek a method that converts model performance (as assessed through calibration) into a statement of model probability (the *posterior model probability*).

Following Burnham and Anderson (2003) and Poeter and Hill (2007) we use the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951), approximated by the corrected Akaike information criterion (AIC_c, Akaike, 1973, 1974; Sugiura, 1978), as a basis for model selection and weighting. The KL divergence represents the information lost when a model is used to approximate the full truth. The AIC_c is a function of the goodness of model fit, represented by the maximum likelihood function F'_{obj} (Hill & Tiedeman, 2007, their Appendix A), the number of calibrated parameters N_p, the number of observations N_d, and the number of prior information values N_{pr}:

$$AIC_c = F'_{obj} + 2N_p + \frac{2N_p\left(N_p + 1\right)}{N_d + N_{pr} - N_p - 1}. \tag{2}$$

A low value of F'_{obj} represents good model fit, and a low value of AIC_c represents a better model (e.g., less information has been lost). The AIC_c thus rewards models that fit the data well, and penalizes models with a large number of input parameters. In our application the model performance is assessed with a maximum likelihood function based on comparing observed and simulated topography.

There are many alternatives to the to AIC_c to approximate the KL divergence (e.g., Bayesian information criterion [BIC] or Kashyap's information criterion [KIC] as discussed by Hill & Tiedeman, 2007, p. 189). While the AIC_c, BIC, and KIC can perform similarly, and all penalize models as the number of parameters increases, each model fit statistic differs in its theoretical development. For example, the development of the BIC assumes that the "true" model exists in the set of candidate models (Burnham & Anderson, 2003, pp. 284–288, 289; 2004, as discussed by Poeter & Hill, 2007). This manifests in a dependence on the number of observations, such that as N_d increases, the probability of the "true" model based on the BIC converges to 1.0. In contrast, the development of the AIC_c does not assume that the "true" model is in the considered model set, and permits the most probable model to change as more data is collected (Poeter & Hill, 2007, pp. 15–19). For our application, we use the AIC_c because the complexity of the study system means that we have no *a priori* reason to believe that a "true" model is present in the considered model set.

The set of AIC_c values from calibration presented in Table 1 represents a quantitative statement of relative model performance across our set of alternative models. We use the approach of Burnham and Anderson (2003, pp. 70–75) to calculate posterior model probabilities based on the AIC_c. The posterior probability p_i of model i in a set of R candidate models is given as

$$p_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{r=1}^{R}\exp\left(-\frac{1}{2}\Delta_r\right)}, \tag{3}$$

where Δ_i, the AIC_c difference for model i, is defined as

$$\Delta_i = AIC_{c,i} - AIC_{c,\min} \tag{4}$$

and $AIC_{c,min}$ is the minimum value of AIC_c among the R considered models.

Based on calibration results alone, model BasicChRtTh outperforms all other models and receives nearly all the probability weighting (Table 1; see Barnhart, Tucker, et al., 2020c, their Table S3 for all successfully calibrated models). However, when validation is considered, the model probabilities and rankings change (Table 1; Figure 2).

Equation 3 considers only calibration results, and neglects validation results, in constructing posterior probabilities. Because combining calibration and validation results into a single set of posterior model probabili-

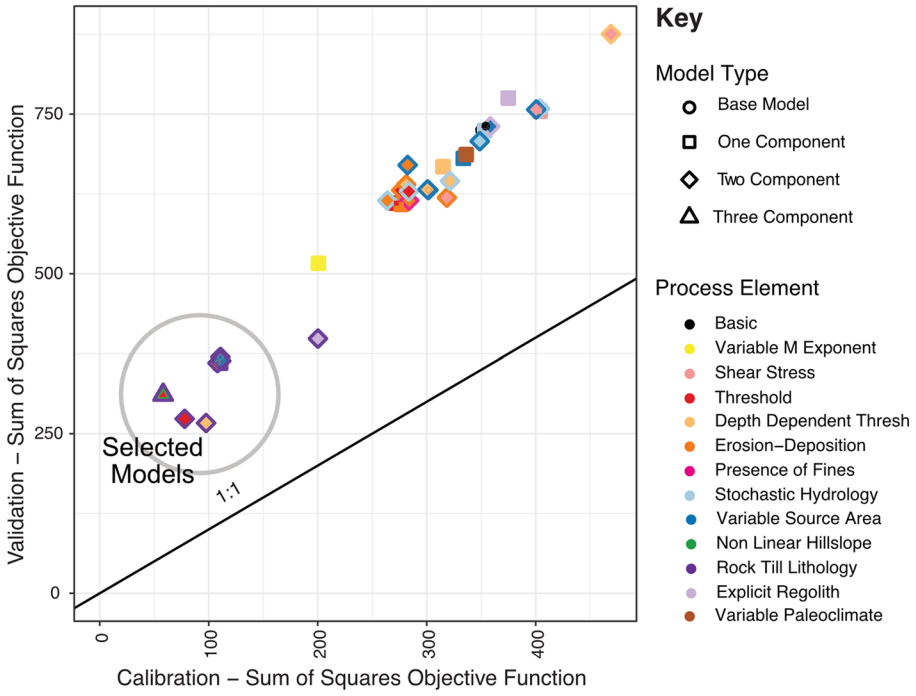

Figure 2. Annotated reproduction of Figure 10 from Barnhart, Tucker, et al. (2020c) showing a comparison of the validation and calibration results for the 34 models that succeeded in calibration. Annotation indicates the group of eight selected models used for projection here.

ties requires subjective choices regarding the bias-variance tradeoff, we evaluate two alternative approaches to model selection to assess how this choice influences the results.

First, we consider only the top-ranked model in calibration (BasicChRtTh), with a probability of unity. Second, we consider the top eight models with equal weighting. We chose these eight models out of the original set of 37 because they showed a clear performance break in calibration, and they had the top performance scores in validation testing (Figure 2). All models share the common element of distinguishing between shale-dominated bedrock and glacial till.

4. Numerical Experimental Design

We designed two numerical experiments to quantify how the five uncertainty sources propagate into projected erosion (Table 2). Expected erosion was estimated by averaging across all model simulations, and the standard deviation due to each source was calculated using analysis-of-variance (ANOVA; Hawkins & Sutton, 2009; Madden, 1976; Yip et al., 2011, and Text S1).

The first experiment quantified projection uncertainty from model structure, both external forcings, and initial condition. Here variation in the initial condition reflects uncertainty in the topographic data as well as contemporary human modification to the terrain (e.g., Lazarus & Goldstein, 2019).

Uncertainty from parameter calibration is excluded from Experiment 1 for computational reasons. It is addressed in Experiment 2, which also considers model structure uncertainty but neglects external forcings and initial conditions. We obtain an estimate of the total uncertainty by adding all variances.

The results of Barnhart, Tucker, et al. (2020c) indicated that initial estimates of posterior parameter distributions from calibration were non-Gaussian, so the posteriors were refined using Bayesian calibration (Markov Chain Monte-Carlo using Delayed Rejection Adaptive Metropolis implemented by QUESO; Haario et al., 2006; Prudencio & Schulz, 2012). Estimates of the refined posteriors fall within reasonable bounds identified by Barnhart, Tucker, et al. (2020d, Figures S1–S8, Tables S3–S10).

Table 2
Summary of Experimental Design

UNCERTAINTY SOURCE:[a]	Experiment 1		Experiment 2	
Model structure[b]	8 Rt Variants OR BasicChRtTh		8 Rt Variants OR BasicChRtTh	
Parameter Calibration	Fixed as calibrated		1000 samples from posterior	
Outlet incision[c]	3 scenarios		1 scenario (slow)	
Climate[d]	3 scenarios		1 scenario (no change)	
Initial Condition[e]	100 samples		No perturbation	
NUMBER OF RUNS	(1 or 8) × 3 × 3 × 100 = 7,200		(1 or 8) × 1,000 = 8,000	
ANALYSIS METHOD	ANOVA		ANOVA	
OUTPUTS	8 Rt variants[f]	BasicChRtTh[f]	8 Rt variants[f]	BasicChRtTh[f]
Expected Erosion	a, p	i, x	Comparable to Exp 1	
Variance from:				
Model Structure (M)	b, q	-	Comparable to Exp 1	
Parameter Calibration (P)[g]	-	-	c, r	j, y
Outlet Incision (O)	d, s	k, z	-	-
Climate (C)	e, t	l, aa	-	-
Initial Condition (I)	f, u	m, bb	-	-
Categorical interactions[h]:				
C-M	g, v	-	-	-
O-M	g, v	-	-	-
C-O	g, v	n, cc		
Total Uncertainty = Model Structure + Parameter Calibration + Outlet Incision + All Interactions.				

[a]*Italic* indicates stochastic input, **bold** indicates categorical input. [b]Model selection evaluated by comparing these two options. [c]Three outlet incision scenarios: slow, average, fast. [d]Three climate scenarios: No change, RCP4.5, RCP8.5. [e]Perturbations added to topography distributed as $\mathcal{N}(0, 5\ \mathrm{m})$. [f]Letters indicate subpanel in Figure 5 where results at 5 ky and 10 ky are presented. - indicates not evaluated in the experiment. [g]Experiment 2 permits assessment of correlation between model structure and parameter calculation (Buckland et al., 1997), which reduced the uncertainty by 20% or less (Figure S9, Text S1.1.4). For the sake of simplicity, here we ignore the correlation and consider the two variances as independent and therefore additive. [h]Categorical interaction row in Figure 5 is the combination of all evaluated interaction terms.

5. Future External Forcing

We consider two types of external forcing: future variations in climate and the elevation of the watershed outlet. Because information about each of these on a 10 ky time scale is limited, we take a simple approach to uncertainty quantification and define three scenarios for each.

5.1. Climate

We represent changes in climate as a change in the effective erodibility coefficient (K in Equation 1). Climate models are usually only run to the year ~2100 and in the absence of longer-term climate model projections, we simply assume that erodibility increases linearly for the first 100 years of simulation time and then stabilizes (Solomon et al., 2009). This choice captures the basic behavior of rapid change, but does not unrealistically increase in perpetuity.

Three scenarios are considered, with K increasing by 0%, +14%, and +25% (Barnhart, Tucker, et al., 2020d, their Section 6.3). Representing changes in climate as changes in K reflects propagating changes in local precipitation distributions into effective erodibility changes using a simple hydrologic model (Barnhart, Glade, et al., 2019; Rossi et al., 2016). This choice stems from the observation that while changes in global precipitation are strongly linked to global temperature (Held & Soden, 2006; O'Gorman & Schneider, 2009;

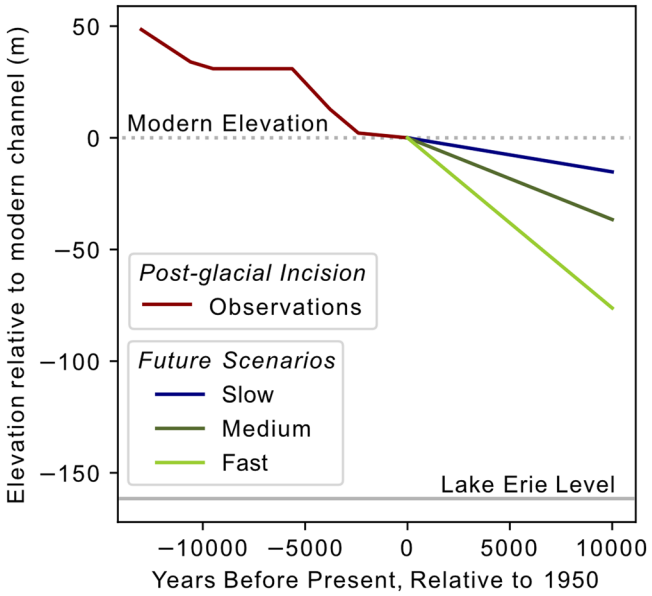

Figure 3. Postglacial incision history and future watershed outlet lowering scenarios.

Trenberth, 2011), assessing regional differences requires local projections (Abatzoglou & Brown, 2012). The two scenarios with increases in K are based on standard Representative Concentration Pathways (RCPs) that extend out to 2100 (RCP4.5 and RCP8.5).

To connect future climate with changes in geomorphic processes and rates, one would ideally want both knowledge of future climate on a 10 ky time scale, and a detailed mechanistic understanding of exactly how climate influences erosion rates and processes. However, projections of the climate state on this time scale are unavailable, apart from the likelihood that global warming will persist to some extent for thousands of years (Solomon et al., 2009). Our understanding of climate impacts on erosion remain rudimentary—though this is an active area of research (e.g., Deal et al., 2018). Because both necessary pieces are poorly constrained we use a simple approach for relating changes in the precipitation distribution to changes in the erodibility coefficient (Barnhart, Glade, et al., 2019).

5.2. Watershed Outlet Lowering

Scenarios for the incision of Buttermilk Creek are based on the 13 ka to present incision history, and local geomorphic and chronologic constraints on paleotopography (M. Wilson & Young, 2018; Young et al., 2020). Buttermilk Creek experienced alternating periods of fast (~0.007 m yr^{-1}) and slow (~0.002 m yr^{-1}) incision, with an average incision rate over the past 13 ka of ~0.004 m yr^{-1} (Figure 3). The incision history likely reflects erosion first through glaciogenic sediments and more recently bedrock, based on exposures of sandstone in its lower reaches (M. Wilson & Young, 2018).

We assume that the geologic substrate will control Buttermilk Creek incision rates over the next 10 ky. The local stratigraphic section is primarily composed of shale, with interbedded siltstones and sandstones (Bergeron, 1985; Rickard, 1975; Smith & Jacobi, 2001; Tesmer, 1963). Nearby subsurface observations are sparse, and thus we consider that Buttermilk creek will erode either shale or sandstone over the simulation duration.

Given the dominance of shale in the local stratigraphic section, it is plausible that Buttermilk Creek's incision will be primarily into shale. Lacking other constraints, we assume that the fastest plausible incision rate through shale can be estimated by the fastest rate observed over the last 13 ka (0.007 m.yr^{-1}). We interpret the recent (2.5 ka to present) slow incision rate as incision through more resistant siltstones and sandstones.

Next, we assess the plausibility of the slow incision rate persisting for the entire 10 ky simulation duration. Assuming the current rate is set by incision into sandstone or similarly resistant bedrock, this would require an additional 20 m of resistant bedrock in the local subsurface. Regional geologic mapping identifies two approximately 50 m thick interbedded siltstone units, the lower of which may be at an elevation close to the current confluence of Buttermilk Creek and Franks Creek (Bergeron, 1985; Smith & Jacobi, 2001; Tesmer, 1963, and personal communication between Doty & Fakindiny, 2017).

To reflect the uncertainty in the future incision of Buttermilk Creek, our three scenarios for future watershed outlet incision span the range of rates observed in past 13 ka. The three scenarios all use a constant incision rate for the 10 ky simulation period. The rates are 0.002, 0.004, and 0.007 m yr^{-1}, reflecting the slow, average, and fast rates from the postglacial period, and an expectation of either incision into sandstone, shale, or a mixture of the two.

6. Results

Projection results are depicted as maps of the expected value of erosion and each uncertainty source (Figure 5 and Figures S10–S19). Most erosion is concentrated in the lower portions of the study watershed, on the steep valley side walls and the adjacent portions of the low-relief remnant plateau surfaces (NP and SP

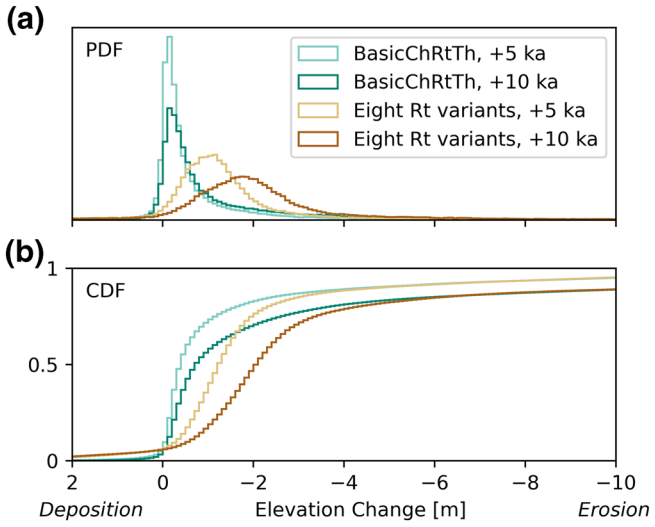

Figure 4. Probability distribution function (a) and cumulative distribution function (b) of expected erosion and deposition for +5 ka and +10 ka and both model set options. The mean and median are very different across model set and time. However, the proportion of the domain that sees erosion exceeding 5 m is consistent across model sets and increases through time.

in Figure 1). Maximum erosion depth on the plateau edges can reach 25 m. The highest forecast erosion depths occur along the main valley near the watershed outlet, where the channel is strongly influenced by imposed outlet incision. The spatial patterns of total expected erosion and total uncertainty are remarkably similar between the single ("best") model and eight-model ensemble.

The two model sets differ in overall median erosion depth (0.7 m for BasicChRtTh model; 2 m for the eight-model ensemble) because the ensemble includes extensive but shallow erosion in the upper watershed (Figure 4). However, the two model sets are consistent in the portion of the total domain that experiences significant erosion through time. Both indicate that ~17% of the domain will erode in excess of 5 m at +10 ky.

The magnitudes of each component of uncertainty grow with time, with each uncertainty source showing a characteristic spatial pattern. In the plateaus, the erosion-depth uncertainty is dominated by the initial topography. Comparing the drainage patterns from two model runs that differ only in the initial condition indicates that small variations in topography influence the planform shapes and positions of gullies that incise the plateaus (Figures 6a–6c). This is consistent with prior work demonstrating that drainage patterns on low-relief surfaces are sensitive to small variations in starting topography (Hancock, 2006; Hancock, Coulthard, & Lowry, 2016; Ijjász-Vásquez et al., 1992; Kwang & Parker, 2019; Willgoose et al., 2003).

Uncertainty in calibration and climate forcing manifests most substantially along the downstream valley side walls of the study watershed. These areas have some of the steepest slopes in the domain and are impacted by variation in the value of K. The uncertainty signature of the outlet incision scenarios is focused on valley bottoms and hillslopes near the watershed outlet.

Model structure uncertainty is highest along the bottom of main channels near the gullies that cut into the remnant plateau surfaces. This spatial pattern represents differences between models that do and do not include a fluvial erosion threshold (Figures 6d–6f). Models with a threshold (BasicRdTh, BasicDdRt, and BasicChRtTh) have initially faster incision in channel bottoms compared to the no-threshold models. Incision then slows as the channel slope approaches a threshold slope for incision (Figure 6g). This result is counter to the intuition that adding a threshold will slow incision rates, all else equal. However, in this case, all else is not equal because the values for both erodibility coefficients and thresholds were calibrated by Barnhart, Tucker, et al. (2020c). Models without an erosion threshold had calibrated erodibility values that were smaller than models with thresholds.

The threshold models also forecast little to no erosion in the upper hillslope portions of the watershed. Despite these differences, a comparison of representative longitudinal profiles shows patterns of incision that are remarkably similar across models (Figure 6). Our experimental design permits us to estimate model structure uncertainty in both experiments. We find similar model structure uncertainty in the two experiments (Figure S20).

Uncertainty derived from the interaction between categorical inputs (model structure, climate, and outlet incision) is relatively unimportant. At +10 ky, the largest contributor to this uncertainty source is the interaction between model structure and outlet incision (Figure S21), reflecting the variable impact of outlet incision on models with and without a fluvial erosion threshold.

7. Discussion

The ANOVA-based approach to uncertainty partitioning with LEMs facilitates comparison of the spatial and temporal evolution of uncertainty in projections due to multiple sources. We expect that this approach is suitable for other well-constrained sites. The most important uncertainty sources will vary in other appli-

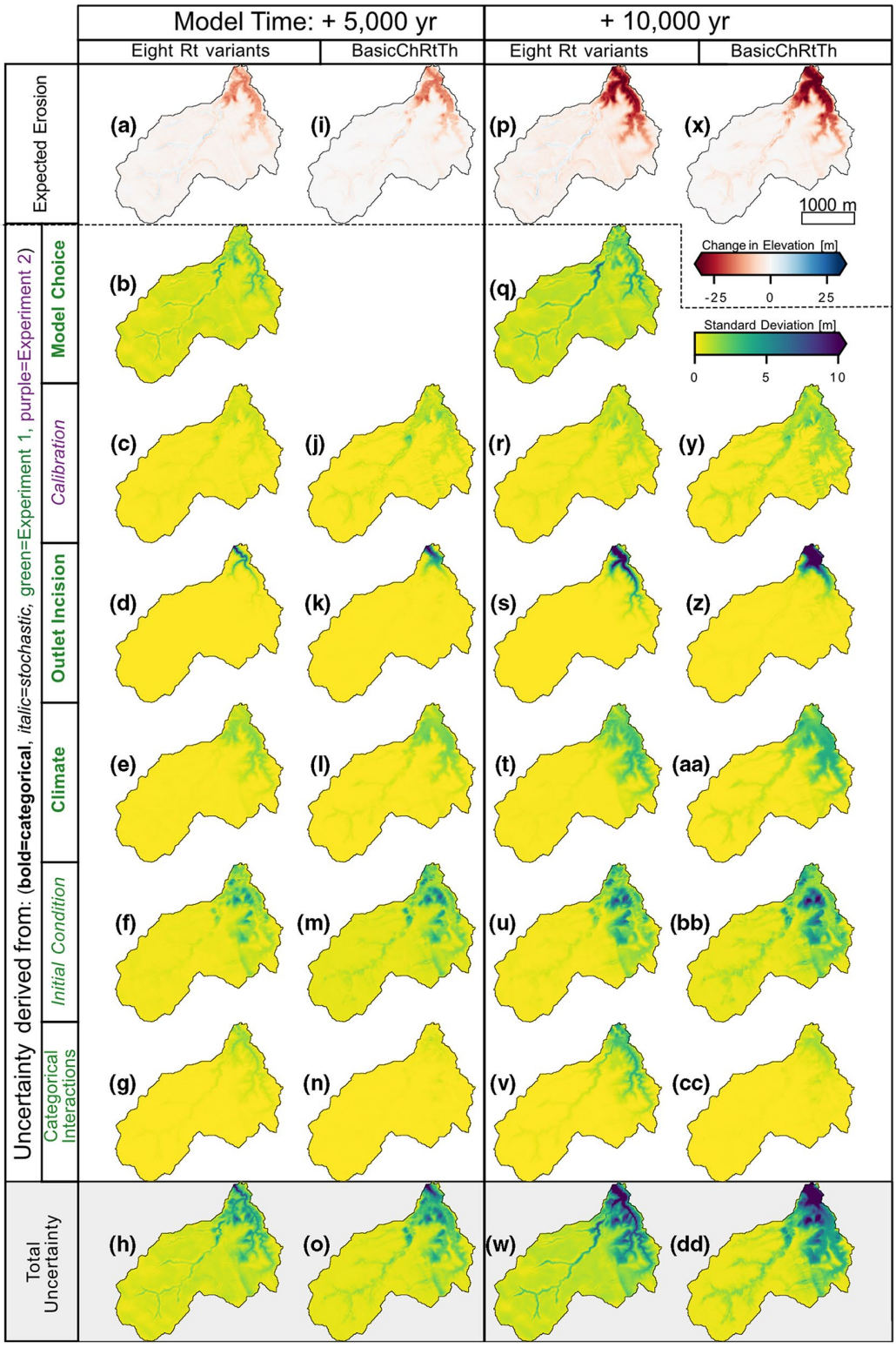

Figure 5. Maps of expected erosion (a, i, p, x) and uncertainty sources at +5,000 years (left) and +10,000 years (right). Results are shown for eight rock-till model variants and for model BasicChRtTh only. Uncertainty due to model structure is only constrained for the case of all eight Rt variants (b, q). Total uncertainty (h, o, w, dd) is the additive variance from model structure, parameter calibration (c, j, f, y), outlet incision (d, k, s, z), climate (e, l, t, aa), initial condition (f, m, u, bb), and all interactions between categorical inputs (g, n, v, cc). Elevation change scale is truncated at 30.5 m and the standard deviation scale is truncated at 10 m.

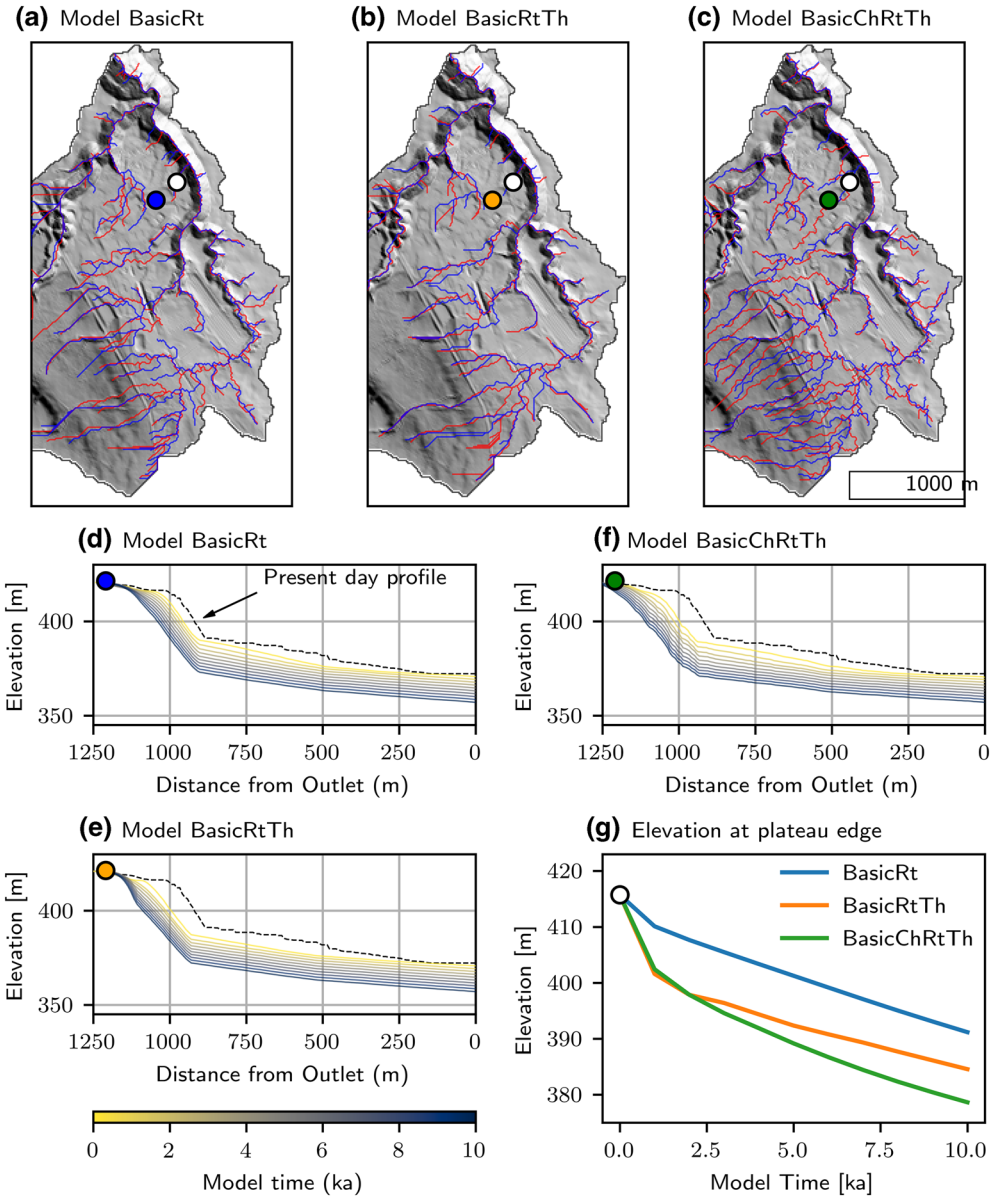

Figure 6. Impact of model selection and initial condition on erosion through time. Upper row shows channel network (drainage area threshold = 5.35*E*3 m^2) for two initial condition samples (#0 = red, #1 = blue) for models BasicRt (a) BasicRtTh (b), and BasicChRtTh (c). Panels (d), (e), and (f) show the channel profile through time at 1000 years increments starting from the blue, orange, and green dots in (a), (b), and (c), respectively. Panel (g) shows the elevation at the plateau edge (white dot) through time in each of the three models.

cations depending on watershed characteristics and the projected quantities of interest. This contribution represents both an assessment of projection uncertainty for the study site, and an exploration of what this approach can illuminate about our current understanding of process geomorphology in general.

The calibration and validation results of Barnhart, Tucker, et al. (2020c) provided the starting point from which we created our model sets. However, model selection was not straightforward because calibration and validation pointed toward incompatible posterior model probabilities. By selecting two end-member model sets, and comparing the results, we demonstrate which aspects of the results are insensitive to the method of model selection. For example, while the spatial patterns of erosion are relatively insensitive

to this choice (Figures 5b and 5q), the mean erosion depth (and thus mean sediment export) is sensitive (Figure 4).

A natural next step is to identify where uncertainty can be further reduced. The relatively high uncertainty that derives from initial conditions implies that the detailed configuration of gully networks that form on the two plateau surfaces are sensitive to minor variations in the starting (contemporary) topography. The considered models do not formally treat elements of the built environment (e.g., pavement and culverts), which we expect to strongly influence the evolving gully networks. With regard to reducing model structure uncertainty, the identification and monitoring of locations that show large variance due to model structure would be a reasonable approach. There are clearly benefits to ongoing monitoring of terrain features such as gully longitudinal profiles, but the similarity among modeled profiles suggests that this alone may not be sufficient to reduce model structure uncertainty (Figures 6d–6f). We also suspect that supplementing topography with additional state variables, such as long-term sediment yield, may help reduce model uncertainty (e.g., Furbish, 2003).

If we had found a large component of projection uncertainty from model structure, we could point to refinement of our governing equations as an important step in reducing uncertainty. This is not the case. Instead, we show that a set of models which vary in their governing equation and represent distinct implementations of fluvial erosion (threshold/no threshold) and hillslope transport (linear/nonlinear) produce quite similar projections. We suspect that model structure uncertainty is relatively low in this study because we were able to incorporate the results of calibration and model selection based on a well-constrained natural experiment. This result may also reflect the similarity in duration between the calibration and projection periods (13 ky and 10 ky, respectively).

Future calibration and prediction-under-uncertainty efforts should assess model structure uncertainty (and thus robustness of the model set) in order to determine under what circumstances model structure uncertainty is low. Is this a common outcome, or is it unique to unusually well-constrained sites where calibration and projection timeframes are similar? Only through future calibration and projection efforts leveraging other well-constrained natural experiments will we be able to assess the extent to which these results can be generalized.

8. Conclusions

In this work, we show how to better quantify uncertainty sources for LEMs using formal model analysis. We make projections of erosion for a 5 km^2 watershed in New York, USA, and show that one sixth of the watershed can expect to see erosion in excess of 5 m over the next 10 ky. While this result is insensitive to the choice of model, other simulation results (e.g., mean erosion rate and thus sediment outflux) are not.

We consider multiple sources of uncertainty (model structure, selection, parameter estimation, boundary conditions, and initial conditions) and use an ANOVA to identify where in the watershed each of these sources dominates. By partitioning uncertainty in this way, we place each source on a level field. This approach allows us to conclude that, for this site, model structure uncertainty is low, whereas uncertainty related to potential near-term anthropogenic landscape modification translates into uncertainty in the scale and location of gullies that incise the till plateau over the next 10 ky. Overall, our approach can be considered a framework for distinguishing when the main uncertainty lies in the details of geomorphic process representation, and when it arises instead from external factors such as lithology, topography, or climate.

Data Availability Statement

Data were not created for this research. The creation and analysis of models presented here was fully scripted. Instructions for reproducing the results, input files, model and analysis code, and the model output files are available through a GlobusConnect endpoint (endpoint name: Barnhart_WVDP_EWG_STUDY3, endpoint identifier UUID 89df0600-bd11-11e8-8c12-0a1d4c5c824a). In addition, the input files and code are housed on GitHub (https://github.com/kbarnhart/inverting_topography_postglacial) and archived with Zenodo (Barnhart, Tucker, et al., 2020a). This work used Landlab version v1.3.1 + 41.g37a1e60.

Acknowledgments
The authors thank Editor Noah Finnegan, Stuart Grieve, and two anonymous reviewers for constructive and thoughtful comments that improved the quality and clarity of the manuscript. Support for this work was provided by a contract with Enviro Compliance Solutions, Inc. (Contract Number DE-EM0002446/0920/13/DE-DT0005364/001), NSF Award 1450409 to Tucker, an NSF EAR Postdoctoral Fellowship to Barnhart (NSF 1725774), and a National Defense Science and Engineering Graduate Fellowship and a University of Colorado Chancellor's Fellowship to Shobe. Landlab is supported by by NSF ACI-1450409 and by the Community Surface Dynamics Modeling System (CSDMS; NSF 1226297, 1831623). This work utilized the RMACC Summit supercomputer, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The Summit supercomputer is a joint effort of the University of Colorado Boulder and Colorado State University. We acknowledge computing time on the CU-CSDMS High-Performance Computing Cluster. Data storage supported by the University of Colorado Boulder "PetaLibrary." Discussion with Christopher Miller regarding statistical inference and semantics substantially improved this manuscript.

References

Abatzoglou, J. T., & Brown, T. J. (2012). A comparison of statistical downscaling methods suited for wildfire applications. *International Journal of Climatology*, *32*(5), 772–780. https://doi.org/10.1002/joc.2312

Adams, B., Bauman, L., Bohnhoff, W., Dalbey, K., Ebeida, M., Eddy, J., et al. (2017a). *Dakota, A multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 6.6 theory manual* (Sandia Technical Report SAND2014-4253). Retrieved from https://dakota.sandia.gov/sites/default/files/docs/6.6/Theory-6.6.0.pdf

Adams, B., Bauman, L., Bohnhoff, W., Dalbey, K., Ebeida, M., Eddy, J., et al. (2017b). *Dakota, A multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 6.6 user manual* (Sandia Technical Report SAND2014-4633). Retrieved from https://dakota.sandia.gov/sites/default/files/docs/6.6/Users-6.6.0.pdf

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Atchley, A. L., Birdsell, K. H., Crowell, K., Middleton, R. S., & Stauffer, P. H. (2019). Simulating 10,000 years of erosion to assess nuclear waste repository performance. *Geosciences*, *9*(3), 120. https://doi.org/10.3390/geosciences9030120

Barnhart, K. R., Glade, R. C., Shobe, C. M., & Tucker, G. E. (2019). Terrainbento 1.0: A Python package for multi-model analysis in long-term drainage basin evolution. *Geoscientific Model Development*, *12*(4), 1267–1297. https://doi.org/10.5194/gmd-12-1267-2019

Barnhart, K. R., Hutton, E. W. H., & Tucker, G. E. (2019). umami: A Python package for Earth surface dynamics objective function construction. *Journal of Open Source Software*, *4*(42), 1776. https://doi.org/10.21105/joss.01776

Barnhart, K. R., Hutton, E. W. H., Tucker, G. E., Gasparini, N. M., Istanbulluoglu, E., Hobley, D. E. J., & Bandaragoda, C. (2020). Short communication: Landlab v2.0: A software package for Earth surface dynamics. *Earth Surface Dynamics*, *8*(2), 379–397. https://doi.org/10.5194/esurf-8-379-2020

Barnhart, K. R., Tucker, G. E., Doty, S., Shobe, C. M., Glade, R. C., Rossi, M. W., & Hill, M. C. (2020a). Calculation Package: Inverting topography for landscape evolution model process representation. *Zenodo*. https://doi.org/10.5281/zenodo.2799489

Barnhart, K. R., Tucker, G. E., Doty, S., Shobe, C. M., Glade, R. C., Rossi, M. W., & Hill, M. C. (2020b). Inverting topography for landscape evolution model process representation: Part 1. Conceptualization and sensitivity analysis. *Journal of Geophysical Research: Earth Surface*, *125*, e2018JF004961. https://doi.org/10.1029/2018JF004961

Barnhart, K. R., Tucker, G. E., Doty, S., Shobe, C. M., Glade, R. C., Rossi, M. W., & Hill, M. C. (2020c). Inverting topography for landscape evolution model process representation: Part 2. Calibration and validation. *Journal of Geophysical Research: Earth Surface*, *125*, e2018JF004963. https://doi.org/10.1029/2018JF004963

Barnhart, K. R., Tucker, G. E., Doty, S., Shobe, C. M., Glade, R. C., Rossi, M. W., & Hill, M. C. (2020d). Inverting topography for landscape evolution model process representation: Part 3. Determining parameter ranges for select mature geomorphic transport laws and connecting changes in fluvial erodibility to changes in climate. *Journal of Geophysical Research: Earth Surface*, *125*, e2019JF005287. https://doi.org/10.1029/2019JF005287

Bergeron, M. (1985). *Records of wells, test borings, and some measured geologic sections near the Western New York Nuclear Service Center* (U.S. Geological Survey Open-File Report(83-682), p. 95) Cattaraugus County, New York. https://doi.org/10.3133/ofr83682

Box, G. E., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for experimenters: Design, innovation, and discovery* (2nd ed.). Hoboken, NJ, USA: Wiley-Interscience New York.

Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics*, 603–618. https://doi.org/10.2307/2533961

Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multimodel inference: A practical information-theoretic approach*. New York, NY: Springer-Verlag.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., & Rasmussen, R. M. (2015). A unified approach for process-based hydrologic modeling: 1 Modeling concept. *Water Resources Research*, *51*(4), 2498–2514. https://doi.org/10.1002/2015WR017198

de González, A. B., Cox, D. R. (2007). Interpretation of interaction: A review. *The Annals of Applied Statistics*, *1*, (2), 371–385. https://doi.org/10.1214/07-aoas124

Deal, E., Braun, J., & Botter, G. (2018). Understanding the role of rainfall and hydrology in determining fluvial erosion efficiency. *Journal of Geophysical Research: Earth Surface*, *123*(4), 744–778. https://doi.org/10.1002/2017JF004393

Dennis Jr., J. E., Gay, D. M., Walsh, R. E. (1981). An Adaptive Nonlinear Least-Squares Algorithm. *ACM Transactions on Mathematical Software*, *7*(3), 348–368. https://doi.org/10.1145/355958.355965

Duvall, A. R., & Tucker, G. E. (2015). Dynamic ridges and valleys in a strike-slip environment. *Journal of Geophysical Research: Earth Surface*, *120*(10), 2016–2026. https://doi.org/10.1002/2015JF003618

El-Ghonemy, H., Watts, L., & Fowler, L. (2005). Treatment of uncertainty and developing conceptual models for environmental risk assessments and radioactive waste disposal safety cases. *Environment International*, *31*(1), 89–97. https://doi.org/10.1016/j.envint.2004.07.002

Evans, K.G., Saynor, M.J., Willgoose, G.R., Riley, S.J. (2000). Post-mining landform evolution modelling: 1. Derivation of sediment transport model and rainfall-runoff model parameters. *Earth Surface Processes and Landforms*, *25*(7), 743–763. https://doi.org/10.1002/1096-9837(200007)25:7<743::aid-esp95>3.0.co;2-0

Fakundiny, R. (1985). Practical applications of geological methods at the West Valley low-level radioactive waste burial ground, western New York. *Northeastern Environmental Science*, *4*(3–4), 116–148.

Foglia, L., Mehl, S., Hill, M., & Burlando, P. (2013). Evaluating model structure adequacy: The case of the Maggia Valley groundwater system, southern Switzerland. *Water Resources Research*, *49*(1), 260–282. https://doi.org/10.1029/2011WR011779

Foreman-Mackey, D. (2016). corner.py: Scatterplot matrices in python. *Journal of Open Source Software*, *1*(2), 24. https://doi.org/10.21105/joss.00024

French, S., Schuman, R., Cole, G. L., Crowell, K. J., Gable, C. W., Gard, M. O., & Stauffer, P. H. (2008). *Performance assessment and composite analysis for Los Alamos National Laboratory Technical area 54, area G, Revision 4*(Los Alamos National Laboratory Report LA-UR-08-06764). Retrieved from https://www.energy.gov/sites/prod/files/2015/10/f27/PACA_2008.pdf

Furbish, D. J. (2003). Using the dynamically coupled behavior of land-surface geometry and soil thickness in developing and testing hillslope evolution models. In P. R. Wilcock & R. M. Iverson (Eds.), *Prediction in geomorphology* (pp. 169–181). Washington, DC: American Geophysical Union (AGU). https://doi.org/10.1029/135GM12

Haario, H., Laine, M., Mira, A., & Saksman, E. (2006). DRAM: Efficient adaptive MCMC. *Statistics and Computing*, *16*(4), 339–354. https://doi.org/10.1007/s11222-006-9438-0

Hancock, G. R. (2006). The impact of different gridding methods on catchment geomorphology and soil erosion over long timescales using a landscape evolution model. *Earth Surface Processes and Landforms*, *31*(8), 1035–1050. https://doi.org/10.1002/esp.1306

Hancock, G. R., Coulthard, T. J., & Lowry, J. B. C. (2016). Predicting uncertainty in sediment transport and landscape evolution—The influence of initial surface conditions. *Computers and Geosciences*, *90*(Part B), 117–130. https://doi.org/10.1016/j.cageo.2015.08.014

Hancock, G. R., Duque, J. F. M., & Willgoose, G. R. (2019). Geomorphic design and modeling at catchment scale for best mine rehabilitation—The Drayton mine example (New South Wales, Australia). *Environmental Modelling & Software*, *114*, 140–151. https://doi.org/10.1016/j.envsoft.2018.12.003

Hancock, G. R., Lowry, J. B. C., & Coulthard, T. J. (2015). Catchment reconstruction—Erosional stability at millennial time scales using landscape evolution models. *Geomorphology*, *231*, 15–27. https://doi.org/10.1016/j.geomorph.2014.10.034

Hancock, G. R., Lowry, J. B. C., & Coulthard, T. J. (2016). Long-term landscape trajectory—Can we make predictions about landscape form and function for post-mining landforms? *Geomorphology*, *266*, 121–132. https://doi.org/10.1016/j.geomorph.2016.05.014

Hancock, G. R., Lowry, J., Moliere, D., & Evans, K. (2008). An evaluation of an enhanced soil erosion and landscape evolution model: A case study assessment of the former Nabarlek uranium mine, Northern Territory, Australia. *Earth Surface Processes and Landforms*, *33*(13), 2045–2063. https://doi.org/10.1002/esp.1653

Hancock, G. R., Lowry, J. B. C., & Saynor, M. (2017). Surface armour and erosion—Impacts on long-term landscape evolution. *Land Degradation & Development*, *28*(7), 2121–2136. https://doi.org/10.1002/ldr.2738

Hancock, G. R., Verdon-Kidd, D., & Lowry, J. B. C. (2017). Sediment output from a post-mining catchment—centennial impacts using stochastically generated rainfall. *Journal of Hydrology*, *544*, 180–194. https://doi.org/10.1016/j.jhydrol.2016.11.027

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1107. https://doi.org/10.1175/2009BAMS2607.1

Held, I. M., & Soden, B. J. (2006). Robust responses of the hydrological cycle to global warming. *Journal of Climate*, *19*(21), 5686–5699. https://doi.org/10.1175/JCLI3990.1

Hill, M. C., & Tiedeman, C. R. (2007). *Effective groundwater model calibration: With analysis of data, sensitivities, predictions, and uncertainty*. Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/0470041080

Hobley, D. E. J., Adams, J. M., Nudurupati, S. S., Hutton, E. W. H., Gasparini, N. M., Istanbulluoglu, E., & Tucker, G. E. (2017). Creative computing with Landlab: An open-source toolkit for building, coupling, and exploring two-dimensional numerical models of Earth-surface dynamics. *Earth Surface Dynamics*, *5*(1), 21–46. https://doi.org/10.5194/esurf-5-21-2017

Ijjasz-Vasquez, E. J., Bras, R. L., & Moglen, G. E. (1992). Sensitivity of a basin evolution model to the nature of runoff production and to initial conditions. *Water Resources Research*, *28*(10), 2733–2741. https://doi.org/10.1029/92WR01561

Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, *13*(4), 455–492. https://doi.org/10.1023/A:1008306431147

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694

Kwang, J. S., & Parker, G. (2019). Extreme memory of initial conditions in numerical landscape evolution models. *Geophysical Research Letters*, *46*(12), 6563–6573. https://doi.org/10.1029/2019GL083305

Lazarus, E. D., & Goldstein, E. B. (2019). Is there a bulldozer in your model? *Journal of Geophysical Research: Earth Surface*, *124*(3), 696–699. https://doi.org/10.1029/2018JF004957

Lowry, J., Coulthard, T., & Hancock, G. (2013). Assessing the long-term geo-morphic stability of a rehabilitated landform using the CAESAR-Lisflood landscape evolution model. In *Proceedings of the Eighth International Seminar on mine closure* (pp. 611–624). https://doi.org/10.36487/ACG_rep/1352_51_Lowry

Madden, R. A. (1976). Estimates of the Natural Variability of Time-Averaged Sea-Level Pressure. *Monthly Weather Review*, *104*(7), 942–952. https://doi.org/10.1175/1520-0493(1976)104<0942:eotnvo>2.0.co;2

Miller, S. R., & Slingerland, R. L. (2006). Topographic advection on fault-bend folds: Inheritance of valley positions and the formation of wind gaps. *Geology*, *34*(9), 769–772. https://doi.org/10.1130/G22658.1

Miller, S. R., Slingerland, R. L., & Kirby, E. (2007). Characteristics of steady state fluvial topography above fault-bend folds. *Journal of Geophysical Research*, *112*, F04004. https://doi.org/10.1029/2007JF000772

Neuhold, C., & Nachtnebel, H. (2011). Assessing flood risk associated with waste disposals: Methodology, application and uncertainties. *Natural Hazards*, *56*(1), 359–370. https://doi.org/10.1007/s11069-010-9575-9

O'Gorman, P. A., & Schneider, T. (2009). The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proceedings of the National Academy of Sciences*, *106*(35), 14773–14777. https://doi.org/10.1073/pnas.0907610106

Pelletier, J. D. (2010). How do pediments form?: A numerical modeling investigation with comparison to pediments in southern Arizona, USA. *The Geological Society of America Bulletin*, *122*(11–12), 1815–1829. https://doi.org/10.1130/B30128.1

Perron, J. T., Kirchner, J. W., & Dietrich, W. E. (2009). Formation of evenly spaced ridges and valleys. *Nature*, *460*(7254), 502–505. https://doi.org/10.1038/nature08174

Poeter, E. P., & Anderson, D. (2005). Multimodel ranking and inference in ground water modeling. *Groundwater*, *43*(4), 597–605. https://doi.org/10.1111/j.1745-6584.2005.0061.x

Poeter, E. P., & Hill, M. C. (2007). MMA, A computer code for multi-model analysis. *United States Geological Survey Techniques and Methods*, *6-E3*, 1–113. https://doi.org/10.3133/tm6E3

Prudencio, E., & Schulz, K. W. (2012). The parallel C++ statistical library 'QUESO': Quantification of uncertainty for estimation, simulation and optimization. In Euro-Par 2011: Parallel Processing Workshops (Vol. *7155*, pp. 398–407). Berlin; Heidelberg: Springer. https://doi.org/10.1007/978-3-642-29737-3_44

Rickard, L. (1975). *Correlation of the Silurian and Devonian rocks in New York State, Map and Chart Series No. 24*, Albany, NY: . New York State Museum and Science Service. Retrieved from https://ngmdb.usgs.gov/Prodesc/proddesc_90884.htm

Rossi, M. W., Whipple, K. X., & Vivoni, E. R. (2016). Precipitation and evapotranspiration controls on daily runoff variability in the contiguous United States and Puerto Rico. *Journal of Geophysical Research: Earth Surface*, *121*(1), 128–145. https://doi.org/10.1002/2015JF003446

Sharmeen, S., & Willgoose, G. R. (2007). A one-dimensional model for simulating armouring and erosion on hillslopes: 2. Long term erosion and armouring predictions for two contrasting mine spoils. *Earth Surface Processes and Landforms*, *32*(10), 1437–1453. https://doi.org/10.1002/esp.1482

Slingerland, N., Isidoro, A., Fernandez, S., & Beier, N. A. (2018). Geomorphic analysis for tailings dam design in consideration of a 1000-year closure design life. In *Proceedings of the 2nd International Congress on Planning for Closure of Mining Operations* (pp. 1–9). https://doi.org/10.36487/ACG_rep/1915_120_Slingerland

Smith, G., & Jacobi, R. (2001). Tectonic and eustatic signals in the sequence stratigraphy of the Upper Devonian Canadaway Group, New York state. *AAPG Bulletin*, *85*(2), 325–357. https://doi.org/10.1306/092601860696

Solomon, S., Plattner, G.-K., Knutti, R., & Friedlingstein, P. (2009). Irreversible climate change due to carbon dioxide emissions. *Proceedings of the National Academy of Sciences*, *106*(6), 1704–1709. https://doi.org/10.1073/pnas.0812721106

Storch, H. V., & Zwiers, F. W. (1999). *Statistical analysis in climate research*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511612336

Sugiura, N. (1978). Further analysts of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics: Theory and Methods*, *7*(1), 13–26. https://doi.org/10.1080/03610927808827599

Tesmer, I. (1963). *Geology of Chautauqua county, New York, Part 1 stratigraphy and paleontology (upper Devonian)* (Technical Report No. 391).

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, *47*(1/2), 123–138. https://doi.org/10.3354/cr00953

Willgoose, G. R. (2018). *Principles of soilscape and landscape evolution*. Cambridge University Press. https://doi.org/10.1017/9781139029339

Willgoose, G. R., Hancock, G. R., & Kuczera, G. (2003). A framework for the quantitative testing of landform evolution models. In P. R. Wilcock, & R. M. Iverson (Eds.), *Prediction in geomorphology* (pp. 195–216). Washington, DC: American Geophysical Union (AGU). https://doi.org/10.1029/135GM14

Willgoose, Garry, Riley, Steven (1998). The long-term stability of engineered landforms of the Ranger Uranium Mine, Northern Territory, Australia: application of a catchment evolution model. *Earth Surface Processes and Landforms*, *23*, (3), 237–259. https://doi.org/10.1002/(sici)1096-9837(199803)23:3<237::aid-esp846>3.0.co;2-x

Wilson, C. J., Crowell, K. J., & Lane, L. J. (2005). *Surface erosion modeling for the repository waste cover at Los Alamos National Laboratory Technical area 54, material disposal area G* (Los Alamos National Laboratory Report LA- UR-05-7771). Retrieved from https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-05-7771

Wilson, M., & Young, R. A. (2018). Phase 1 erosion studies: Study 1—Terrain analysis. West Valley Erosion Working Group Retrieved from http://www.westvalleyphaseonestudies.org/Documents/EWG%20Final%20Study%201%20Report_Vol%201_2.21.18.pdf

Yip, S., Ferro, C. A., Stephenson, D. B., & Hawkins, E. (2011). A simple, coherent framework for partitioning uncertainty in climate predictions. *Journal of Climate*, *24*(17), 4634–4643. https://doi.org/10.1175/2011JCLI4085.1

Young, R. A., Gordon, L. M., Owen, L. A., Huot, S., Zerfas, T. D. (2020). Evidence for a late glacial advance near the beginning of the Younger Dryas in western New York State: An event postdating the record for local Laurentide ice sheet recession. *Geosphere*, http://dx.doi.org/10.1130/ges02257.1