The impact of different rainfall products on landscape modelling simulations

Christopher J. Skinner,[1]* [iD] Nadav Peleg,[2] Niall Quinn,[3] Tom J. Coulthard,[1] [iD] Peter Molnar[2] and Jim Freer[3,4]
[1] Energy and Environment Institute, University of Hull, Hull, UK
[2] ETH Zurich, Institute of Environmental Engineering, Zurich, Switzerland
[3] School of Geographical Sciences, University of Bristol, Bristol, UK
[4] Cabot Institute, University of Bristol, Bristol, UK

*Correspondence to: Christopher J. Skinner, Energy and Environment Institute, University of Hull, Hull HU6 7RX, UK. E-mail: c.skinner@hull.ac.uk

ESPL

Earth Surface Processes and Landforms

ABSTRACT: Rainfall products can contain significantly different spatiotemporal estimates, depending on their underlying data and final constructed resolution. Commonly used products, such as rain gauges, rain gauge networks, and weather radar, differ in their information content regarding intensities, spatial variability, and natural climatic variability, therefore producing different estimates. Landscape evolution models (LEMs) simulate the geomorphic changes in landscapes, and current models can simulate timeframes from event level to millions of years and some use rainfall inputs to drive them. However, the impact of different rainfall products on LEM outputs has never been considered. This study uses the STREAP rainfall generator, calibrated using commonly used rainfall observation products, to produce longer rainfall records than the observations to drive the CAESAR-Lisflood LEM to examine how differences in rainfall products affect simulated landscapes. The results show that the simulation of changes to basin geomorphology is sensitive to the differences between rainfall products, with these differences expressed linearly in discharges but non-linearly in sediment yields. Furthermore, when applied over a 1500-year period, large differences in the simulated long profiles were observed, with the simulations producing greater sediment yields showing erosion extending further downstream. This suggests that the choice of rainfall product to drive LEMs has a large impact on the final simulated landscapes. The combination of rainfall generator model and LEMs represents a potentially powerful method for assessing the impacts of rainfall product differences on landscapes and their short- and long-term evolution. © 2020 The Authors. Earth Surface Processes and Landforms published by John Wiley & Sons Ltd

KEYWORDS: landscape evolution; weather generator; numerical modelling; rainfall; uncertainty

Introduction

Landscape evolution models (LEMs; see review by Tucker and Hancock, 2010) are tools for understanding the role large-scale processes have in the long-term development of the Earth's surface (10^3–10^6 years). LEMs are most often operated in an exploratory manner, employing careful simplifications of physical laws to maximize computational efficiency (Tucker and Hancock, 2010). Recently, the increase in computational power available has allowed for increasing complexity and detail to be simulated (Coulthard *et al.*, 2013), yet understanding output uncertainties and sensitivities of these models to differences between input data has received far less treatment (Coulthard and Skinner, 2016; Hancock *et al.*, 2016; Skinner *et al.*, 2017; Chandra *et al.*, 2019).

Rainfall, both spatially and temporally, is considered a major source of input uncertainty in hydrological modelling (Keijsers *et al.*, 2011; McMillan *et al.*, 2011, 2012; Peleg *et al.*, 2017a). Measurements of rainfall are made using a variety of methods that produce estimations of the spatial and temporal distribution of rainfall, and these different techniques are referred to herein as rainfall products. Rainfall products can be formed from single-point observations (e.g. rain gauge), extrapolated using interpolation of many points in space (networks of rain gauges, disdrometers, etc.), or spatially estimated using remote sensing techniques (i.e. weather radar and satellite). For each of the products, differences in the estimates, spatially and temporally, are expected (e.g. Ciach, 2003; Villarini *et al.*, 2008; Villarini and Krajewski, 2010; Peleg *et al.*, 2013).

Rain gauges are able to provide an estimate of rainfall intensity at a single point, and when used in a network the spatial pattern of rainfall can be estimated using spatial interpolation methods (e.g. kriging) – these spatial datasets contain uncertainties originating from the point measurements, the interpolation (dependent on the density and design of the network), the terrain of the catchment, the spatial distribution of rainfall, and the interpolation method selected (e.g. Hofstra *et al.*, 2009; McMillan *et al.*, 2011). Weather radar is a useful technique for estimating the distribution of a rain field over a large area (Berne and Krajewski, 2013). Weather radar systems measure the backscattered signal and use it as a proxy for rainfall. Quantitative precipitation estimation from weather radar is considered to be highly uncertain, but the quality has improved in recent years, even allowing the study of extreme rainfall intensities

(e.g. Marra and Morin, 2015; Marra *et al.*, 2017). In combination, rain gauges and weather radar can provide a more detailed view of the properties of a rain field, with the added value of spatial coverage provided by the radar, especially between gauges.

Given the length of many LEM simulations, the impact of differences in the driving rainfall product used might seem trivial, however, research has shown that simulated geomorphic processes and sediment output from LEMs can be sensitive to the spatial and temporal resolution of rainfall inputs. Coulthard and Skinner (2016) showed that increased rainfall intensities, which may be hidden within averaged longer timesteps, or spatial averaging, led to greater first-order stream incision and second/third-order stream aggradation for the same rainfall totals. Therefore, Coulthard and Skinner (2016) suggested that the patterns of erosion and deposition predicted by LEMs are highly likely to be affected by uncertainties and variabilities in rainfall products. Coulthard *et al.* (2012b) showed that basins acted as a 'geomorphic multiplier', with a non-linear relationship between rainfall, discharge, and sediment transport acting to exponentially amplify the effects of increases in rainfall on sediment yield.

As landscape evolution operates over timescales spanning tens to millions of years, the choice of rainfall product with which to drive a LEM is often motivated by the requirement for an indicative record of rainfall to sample from. However, because rainfall observation records often lack the required temporal coverage, one approach has been to use as long as possible a record from a nearby rain gauge (even if located tens of kilometres away) and loop it to the required length (e.g. Coulthard *et al.*, 2000; Hancock, 2009; Hancock *et al.*, 2010, 2011, 2015; Hancock and Coulthard, 2012; Saynor *et al.*, 2012; Poeppl *et al.*, 2013; Hoober *et al.*, 2017). This clearly introduces several limitations as climate variability is reduced to the length of the record and introducing non-stationarity (e.g. climate change) can only be represented basically (e.g. by altering rainfall totals). A rainfall times series produced in this way will never contain higher extremes than those observed in the record, which is unrealistic. Rainfall generator models present an alternative way to derive long-term rainfall records (Wilks and Wilby, 1999; Smith *et al.*, 2014) and can be used to produce ensembles of synthetic rainfall calibrated using the climate variability and spatiotemporal characteristics from observed rainfall products. Rainfall generators have been used previously to produce time series of basin-averaged precipitation rates for LEMs (Tucker and Bras, 2000; Coulthard *et al.*, 2012a; Howard *et al.*, 2016; Coulthard and Van De Wiel, 2017; Hancock *et al.*, 2017), with no spatial distribution, and there has been no appraisal of how different methods of rainfall simulation, ergo rainfall products, can alter model outcomes.

This study aims to test how the different spatiotemporal estimates associated with different rainfall observation techniques impact on simulations of changes to basin geomorphology as they cascade through to LEM outputs. Three starting rainfall products will be used: (i) a single rain gauge product; (ii) an interpolated gridded rain field based on a rain gauge network; and (iii) weather radar observations. 50-Year records were produced by looping and bootstrapping (i), and using (ii) and (iii) to calibrate the STREAP weather generator (Paschalis *et al.*, 2013; Peleg *et al.*, 2017b) and to derive 30 ensemble members at a 1 h/1 km resolution. Whilst not a complete uncertainty evaluation of all plausible rainfall scenarios, we can start to quantify the impact and thus sensitivities to the LEM outputs driven by these different products.

Methods and Data

Study catchment

The study area was the Upper Swale catchment, UK (Figure 1), which has an area of 181 km^2 and an elevation range between 182 and 712 m. There is a strong orographic rainfall gradient across the catchment, with mean annual rainfalls of 1000 mm in the north-east and up to 1940 mm in the south-west. The catchment has been chosen as it has been widely used in previous studies testing the CAESAR-Lisflood model (Coulthard and Macklin, 2001; Coulthard *et al.*, 2012b; Coulthard and Van De Wiel, 2013; Coulthard and Skinner, 2016; Skinner *et al.*, 2017), and it has good coverage of the rainfall products.

In order to measure spatial patterns of simulated geomorphic change, the catchment was subdivided into areas based on stream orders derived from the proportion of the catchment drained, as per Skinner *et al.* (2017).

Rainfall products: Single gauge, gauge network, and weather radar

Three rainfall products were examined. The first (Tow) is based on a relatively long record from a single rain gauge located 2 km south-west of the study area, at a 1 h timestep. Even though this is the closest available gauge to the catchment, it is located in a topographically higher area with a greater mean annual rainfall than much of the catchment (see Figure A4 in the online Supporting Information A) – the record was not adjusted to account for this. It provided a 30-year record, however, years with more than 2000 missing hours (83 days) of observations were excluded from the analysis, reducing the record to 24 years (removing 1987–1993, 1995–2009, 2011, and 2014). This product reflects intensities sampled from a single point, therefore providing no representation of the spatial rainfall field, thus rainfall intensities were assumed to be uniformly distributed across the catchment.

The second product (TBR) aims to address the lack of spatial coverage in the single-gauge approach above by interpolating the point data from a network of rain gauges to produce a distributed and gridded rain field estimation (McMillan *et al.*, 2011). The gridded interpolated rainfall data was derived from a tipping bucket rain gauge (TBR) network at 1 h/1 km resolution, as described by Blenkinsop *et al.* (2017).

The third product (NIMROD) is derived from a rain radar network (UK NIMROD Composite product; Met Office, 2003), provided at a 5 min/2 km resolution, which was aggregated to a 1 h timestep.

Despite a much longer record being available for the TBR product (from 1990), both TBR and NIMROD were limited in length to a 6-year period where records for both the TBR and the NIMROD were available (2006–2011). This was so that it would be possible to directly compare their relative abilities in capturing the spatial rain field and point intensities over the same time period, using them in the calibration of STREAP, and subsequently propagating the impacts on changes to basin geomorphology.

An example of the spatial differences in rainfall intensity between the three products is shown in Figure 2, which illustrates the most extreme hourly rainfall intensities that were observed by the three products.

Further in-depth analysis of the differences between the observation products can be found in the online Supporting Information A, however, to summarize, the three products differed in their spatiotemporal representation of rainfall over the catchment. The Tow rain gauge record contained more information
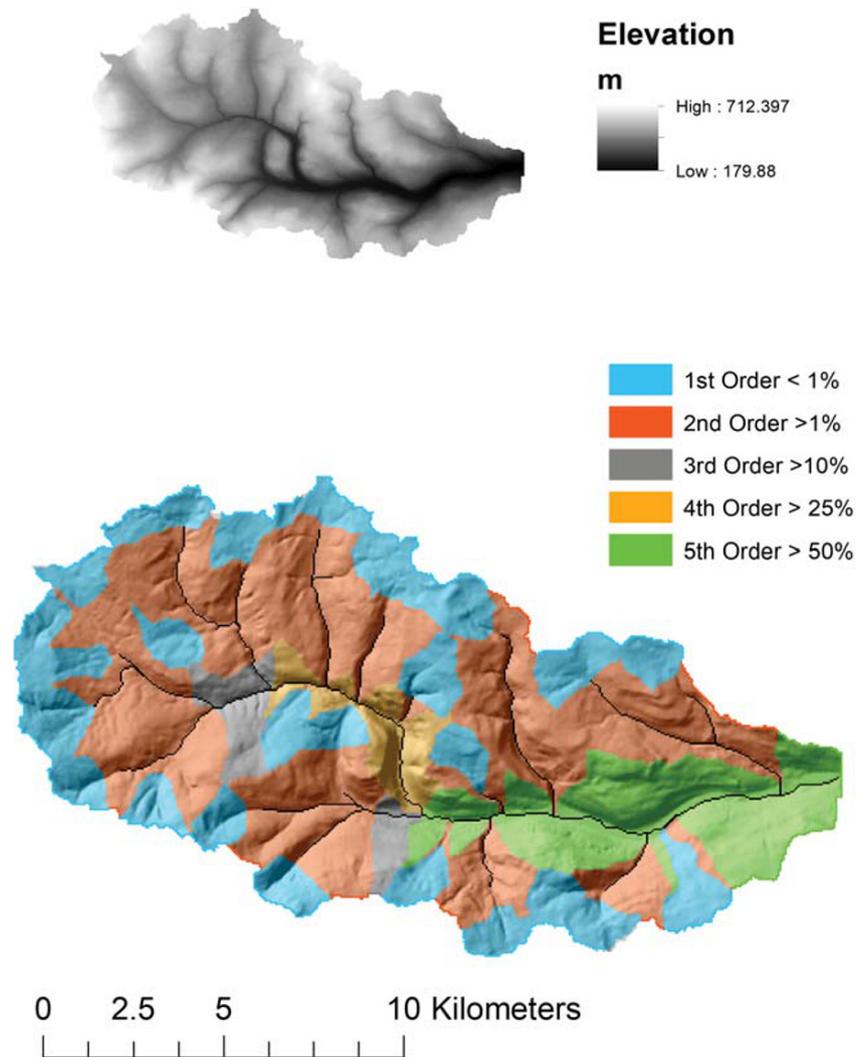
FIGURE 1. The elevations of the Upper Swale catchment, UK (top) and watershed orders (bottom). [Colour figure can be viewed at wileyonlinelibrary.com]
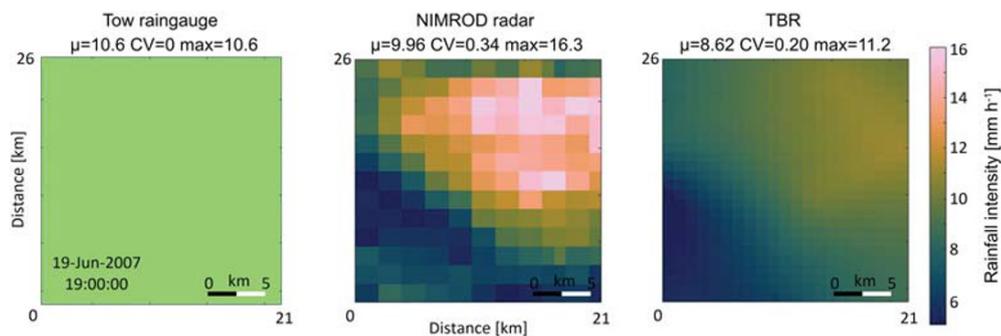

FIGURE 2. The largest hourly rainfall intensity (19/06/2007, 19:00) as recorded by the uniform Tow rain gauge (left), the distributed NIMROD weather radar (middle), and TBR (right) products. μ is areal rainfall intensity (i.e. the average rainfall over the field), CV is the spatial rainfall coefficient of variation (standard deviation of the rainfall intensities over the field divided by the areal rainfall intensity), and max is the maximum rainfall intensity observed in that rain field. [Colour figure can be viewed at wileyonlinelibrary.com]

on the natural variability of climate due to the length of the record, but only supplied this information for a single point – this point is just outside the catchment in the area of highest annual rainfall (1940 mm based off TBR). Point rainfall intensity estimates from the NIMROD weather radar are considered less accurate (not reported here) compared with rain gauges (Cecinati et al., 2017), but the weather radar supplies valuable information on the spatial structure of rainfall. Lastly, produced from interpolation of a network of ground rain gauges, rainfall intensity estimates from TBR were likely more reliable than the estimates

from the weather radar (Lewis et al., 2018), however, the spatial structure of the rainfall is smoothed compared to the radar (Jewell and Gaussiat, 2015), especially as the interpolation included no rain gauges from within the catchment itself.

STREAP rainfall generator

The STREAP (space–time realizations of areal precipitation) model is used to generate gridded rainfall at high spatial

(sub-kilometre) and temporal (minute) resolution. The model was originally developed by Paschalis *et al.* (2013) and was developed further by Peleg *et al.* (2017b). It has been applied for several hydrological applications (e.g. over a large rural catchment for flood investigations – Paschalis *et al.*, 2014, to simulate extreme rainfall intensity over small scales – Peleg *et al.*, 2018, and to study the impacts of spatial and climatological rainfall variability in urban drainage – Peleg *et al.*, 2017a), and was applied with the CAESAR-Lisflood model to study the landscape sensitivity to extreme rainfall events in the context of climate change (Peleg *et al.*, 2020b). STREAP reproduces the storm arrival process (i.e. the length of the storm and intra-storm periods), the temporal evolution of the storm (i.e. the areal intensity over the catchment and the fraction of wet areas), and the space–time structure of rainfall (i.e. intermittent rain fields). The model is ideal for LEM studies as it can be used to produce ensembles of synthetic rainfall reflecting the rainfall natural variability (assuming climate is stationary) from relatively short observation records, whilst reproducing the spatio-temporal characteristics of the rain fields.

CAESAR-Lisflood landscape evolution model

CAESAR-Lisflood is a LEM based on a regular grid that has been used extensively to simulate river basin morphodynamics over a wide range of time (1–10 000 years) and spatial (0.02–1000 km^2) scales (Skinner and Coulthard, 2017). The model can operate in a catchment mode where a spatially distributed rainfall input is converted into surface runoff using the TOPMODEL hydrological model (Beven and Kirkby, 1979), and this is routed across the grid using LISFLOOD-FP (Bates *et al.*, 2010), which generates flow depths and velocities. An active-layer system is used to simulate fluvial erosion and the model can handle up to nine grain sizes. The initial inception of the CAESAR model was to explore the impact large-scale processes have on the long-term evolution of landscapes (>1000 years), however, the incorporation of the LISFLOOD-FP hydraulic code (Coulthard *et al.*, 2013) enabled full hydrodynamics and validated inundation patterns and velocities to be simulated. Consequently, CAESAR-Lisflood is ideal to test the impact differences in high (temporal and spatial)-resolution rain data have on model outputs and on changes to basin geomorphology.

To assess the model's sensitivity to rainfall observation uncertainty, both within and between products, all model parameters were kept constant throughout the experiments. We used a pre-calibrated CAESAR-Lisflood model of the Swale catchment, UK, as used in several previous studies (e.g. Coulthard and Skinner, 2016) and used the same digital elevation model (DEM) derived from an air-borne LiDAR scan that had been resampled to 50-m grid cells. The DEM and grain size distributions were 'spun-up' before the experiments using a 10-year rainfall time series at 24 h/lumped (catchment-average) resolution based on the UK NIMROD 5 km Composite dataset (Met Office, 2003). It is important here to note that the model application is intended as a testbed for the impacts of rainfall products, and therefore does not include any representation of bedrock or vegetation that could alter the model sensitivity.

LEM ensemble simulations

The experimental procedure uses the different gridded rainfall products described to calibrate STREAP. Each calibration is used to produce a rainfall ensemble at a 1 h timestep and a 1 km spatial resolution. The generation of the ensembles follows a common setup when exploring climate impacts including stochastic uncertainties (e.g. Fatichi *et al.*, 2016), with each consisting of 30 individual realizations (to account for the natural climate variability) and covering a period of 50 years (the upper length boundary allowing us to assume climate stationarity). Each ensemble is designed to show how the different information provided by the TBR or NIMROD weather radar rainfall products spatially influences the generated rainfall records – any differences, effectively a source of representative uncertainty, will then be observed through the model cascade. Two records were generated using the Tow gauge: (i) a simple looped record, used as a proxy for a typically used rainfall product to allow comparison; and (ii) a bootstrapped ensemble (of 30 realizations), generated by sampling with replacement of the original records by block bootstrapping of the entire years, meaning that a specific year that was recorded can appear numerous times or never in each realization. It is possible to reduce the uncertainty in rainfall observations by merging multiple products, using the most useful information from each. Here, for example, a 'merged' rainfall record was produced by calibrating STREAP using the best information from the three original products: the temporal structure of rainfall follows that obtained from the Tow rain gauge, the spatial structure of rainfall follows that obtained from the radar, and rainfall intensities are adjusted to follow the TBR data (Table I). The generated rainfall ensembles will henceforth be referred to as products in their own right. The products used are shown in Table I, and an example of a single 50-year realization for each of the products is presented in Figure 3.

Whilst some difference in the discharge, sediment yields, and landscape change will likely be observed between simulations over the 50-year timeframe, longer simulation times will make landscape changes more prominent. To overcome this limitation, 1500-year records were produced – for TowLooped and TowBoot this was done by repeating the 50-year record, and for all other products all 30 ensemble realizations were combined into one long record.

Finally, to ascertain the impact of the extra variability afforded by the longer observation record with the Tow gauge, a further test was performed with normalized rainfall means and spatial representation. This was done using TowLoop, and the first members of the sNIM and sTBR ensembles. The two spatially distributed products were first lumped across the domain and then TowLoop and sNIM were normalized to the mean rainfall of sTBR. These three inputs were then used to drive a 50-year simulation. The peak rainfall intensities in these records were 13.7 mm h^{-1} for TowLoop (adjusted from 17.5 mm h^{-1}), 6.0 mm h^{-1} for sNIM (adjusted from 5.0 mm h^{-1}), and 6.9 mm h^{-1} for sTBR.

Evaluation of STREAP Products

STREAP's ability to reproduce the hourly areal rainfall was examined (Figure 4). The areal rainfall was well reproduced for the Tow rain gauge (Figure 4a). The simulated rainfall for the distributed rainfall products (gridded and radar, Figures 4b and c) is in general also well reproduced, but the high rainfall intensities (i.e. larger than the 95th percentiles) are underestimated. It is worth highlighting this underestimation, as it is these high rainfall intensities that can result in significant geomorphic activity, and subsequently sediment yields may be underestimated. The simulated spatial correlation of rainfall is a bit higher than observed, but the differences are not significant (see Figure A5 in the online Supporting Information B). More details from the evaluation, such as the comparison of the

Table I. Summary of the rainfall products used to drive the CAESAR-Lisflood model

Designation	Composition	Abbreviations
Tow	Tow gauge record (24 years) looped for 50 years and applied as a lumped estimate at 1 h timestep	TowLoop
Gauge	Tow gauge record (24 years) bootstrapped to produce an ensemble of 30 individual 50-year realizations, applied as lumped estimates at 1 h timestep	TowBoot
Gridded	STREAP calibrated using the gridded TBR record (6 years) to generate an ensemble of 30 individual 50-year realizations, applied at 1 h/1 km resolution	sTBR
Radar	STREAP calibrated using the NIMROD record (6 years) to generate an ensemble of 30 individual 50-year realizations, applied at 1 h/1 km resolution	sNIM
Merged	STREAP calibrated using intensity information from the gridded TBR record (6 years), spatial information from the NIMROD record (6 years), and climatic variability information from the Tow gauge (24 years) to generate an ensemble of 30 individual 50-year realizations, applied at 1 h/1 km resolution	sTowMer
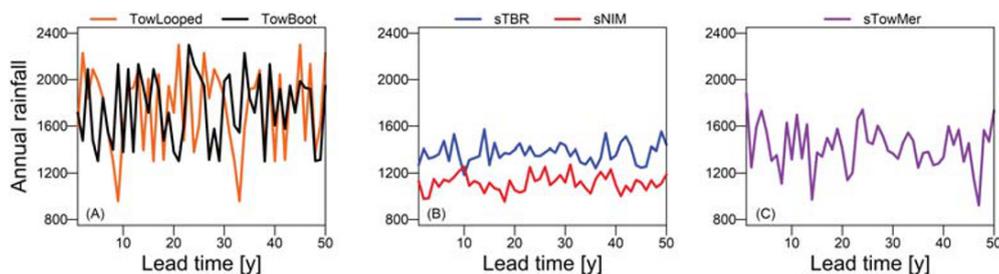
FIGURE 3. An example of a single 50-year realization, presented as a time series of the annual areal rainfall, derived from the Tow rain gauge (A), the gridded TBR and NIMROD weather radar (B), and the merged product (C). [Colour figure can be viewed at wileyonlinelibrary.com]

annual areal rainfall variability between observed and simulated products, can be found in the online Supporting Information B.

Results

LEM rainfall–discharge–sediment cascade

The cumulative rainfall, river discharge, and sediment volumes for all the products are shown in Figure 5. The cascade from the rainfall (Figure 5a) to the discharge (Figure 5b) was observed to be broadly linear, as observed previously (Coulthard et al., 2012b). Each product showed a small increase in volume in the conversion from rainfall to discharge that was caused by the setting of a low flow threshold in the CAESAR-Lisflood model (the model assumes that at discharges below the threshold, the erosion and transport of sediment is negligible and increases the model timestep to improve efficiency, assuming hydrology is running in a steady state). The spread of the products, expressed as the range of values as a percentage of the mean of the members, was similar for both rainfall and discharge, with the spread for sTowMer slightly reducing in the rainfall to discharge cascade (10.7 to 8.7%).

TowLoop, TowBoot, and sTowMer showed much greater rainfall and discharge volumes (Figures 5a and b), and as was expected the TowLoop and the mean of the gauge ensembles were similar (rainfall = 1.6×10^{10} m³; discharge = 2.0×10^{10} m³). sTBR (mean of ensemble = 1.2×10^{10} and 1.4×10^{10} m³), and sNIM (mean of ensemble = 9.6×10^{10} and 1.2×10^{10} m³) showed lower volumes, with sNIM the lowest. Despite being based on different observations of the same rainfall, the ensembles of the two products showed no overlap after the first few years of simulation. sTowMer (mean of ensemble = 1.4×10^{10} and 1.5×10^{10} m³) estimated greater volumes than all the other gridded products, but less than TowLoop and TowBoot.

The sediment yields produced using each rainfall product (Figure 5c) were more varied and complex than the hydrological response (Figure 5b). The same pattern is seen, however, with the Tow-based products having produced the most sediment yield (TowLoop = 1.7×10^6 m³), and sNIM the least (mean of ensemble = 1.5×10^5 m³). Again, there was no overlap between the products after the first few years. sTowMer had a mean cumulative sediment yield (7.3×10^5 m³), over four times greater than sTBR and SNIM yet lower than TowBoot – there was no overlap in range after the first few years.

In all instances the spread of ensembles, as a percentage of the ensemble mean, was greater for sediment yields than for discharges, with the greatest increase seen in sTowMer (discharge = 8.5%; sediment yield = 28.8%). The spread of ensembles was lower for products that did not use observations from the Tow rain gauge as part of the calibration.

When standardized for rainfall means and spatial distribution, each input produced similar hydrological outputs (TowLoop = 1.67×10^5 m³; sNIM = 1.63×10^5 m³; sTBR = 1.60×10^5 m³), yet diverging sediment yields (TowLoop = 9.87×10^5 m³; sNIM = 3.47×10^5 m³; sTBR = 2.65×10^5 m³).

Changes to basin geomorphology

Changes to basin geomorphology were assessed using the longer 1500-year simulations. Figure 6 shows the change in mean elevations for different stream order sub-basins when the model was driven by each of the rainfall products. The difference in the scale of changes between the products is evident, with TowLoop and TowBoot showing the greatest change. sTowMer also shows greater changes than sNim and sTBR. By normalizing the changes as a proportion of the total changes, the different spatial patterns between the products become clearer – TowLoop and TowBoot show a dominance of change (erosion) in the third-order region of the catchment, and also greater changes in the fifth-order region, with a relatively low proportion of change in the fourth-order region. This is in contrast to sNIM and sTBR, which both show a much greater proportion of change in the fourth-order region, a similar level to changes in the third-order region, and less relative change in the fifth-order region. sTowMer shows a mixture of the two
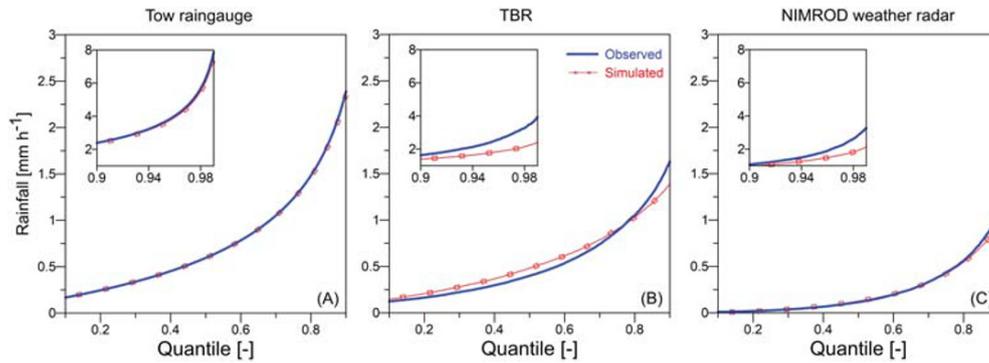
FIGURE 4. Comparison between the observed (blue) and simulated (red) inverse-cumulative distribution functions of the areal rainfall at hourly resolution. 30-Year observed and simulated (gauge) records from Tow rain gauge are presented in (A). In (B) and (C), 6-year observed and simulated (gridded and radar, respectively) records are shown. Internal boxes show differences at the upper quantiles. [Colour figure can be viewed at wileyonlinelibrary.com]
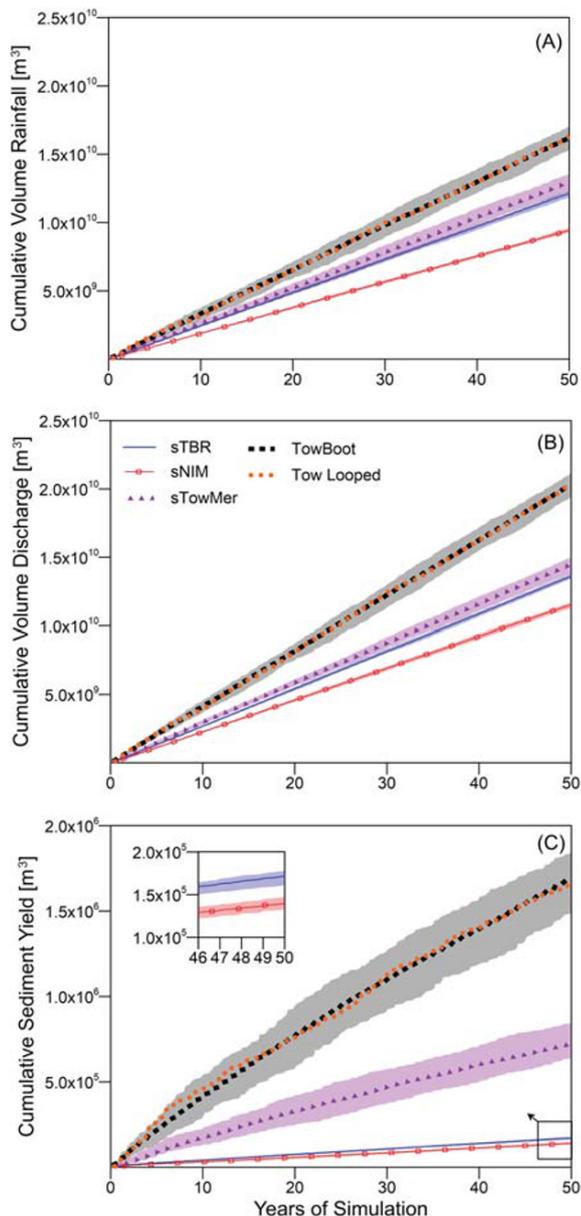

FIGURE 5. Cumulative volume rainfall (a), discharge (b), and sediment yields (c), for each rainfall product over 50 years. Shaded areas indicate the minimum and maximum bounds from the 30 ensemble members. Symbols/lines denote the mean of the 30 ensemble members (except for TowLoop). [Colour figure can be viewed at wileyonlinelibrary.com]

patterns, with third-order dominant, but more change in the fourth than the fifth.

Table II shows statistics drawn from elevation and volume changes at the individual cell level. The same differences observed in Figure 6 are also evident here, with all values greater for the products using the Tow rain gauge record. The mean elevation change across all cells appears low compared to the values seen in Figure 6, but this indicates that high erosion in the upper parts of the catchment is offset by deposition in the fifth-order region, which is a larger area. By looking at the proportion of the overall mean elevation change compared to the total erosion across the catchment, it is possible to see what proportion of eroded sediment is lost to the system at the catchment outlet – this is greatest for the two products containing the information from the Tow rain gauge.

Figure 7 shows the river profiles after 1500 years of simulation, and the changes from the initial river profile. The difference in the scale of the changes is evident, as with Figure 6 and Table II, however, there is a difference in the transition zone where the dominant change goes from erosion to deposition downstream. For sNIM and sTBR, the transition occurs around 7000 m downstream, after 7300 m for sTowMer, and 8000 m for TowLoop. The level of deposition tails off dramatically towards the catchment outlet, with some minor erosion at the edge of the model – this is possibly a result of a model parameter (which sets a fixed hydraulic slope at the edge of the domain).

Discussion

Differences between products

The sensitivity of the CAESAR-Lisflood model to different rainfall products was assessed using the characteristics of different rainfall observation products to calibrate a weather generator, then using ensembles of generated rainfall as an input for CAESAR-Lisflood. Differences in rainfall products occur as each product observes rainfall differently, with different abilities and spatial and temporal coverage. We have not assessed CAESAR-Lisflood's sensitivity to uncertainties associated with individual products, such as observation error in rain gauges or interpolation error in rain gauge networks, although many of the observations made here would apply.

Our simulations show that the variation of rainfall rate had the largest influence on landscape evolution in the tests – the Tow rain gauge had the greatest variability in rainfall of the products and the simulations that used rainfall derived from
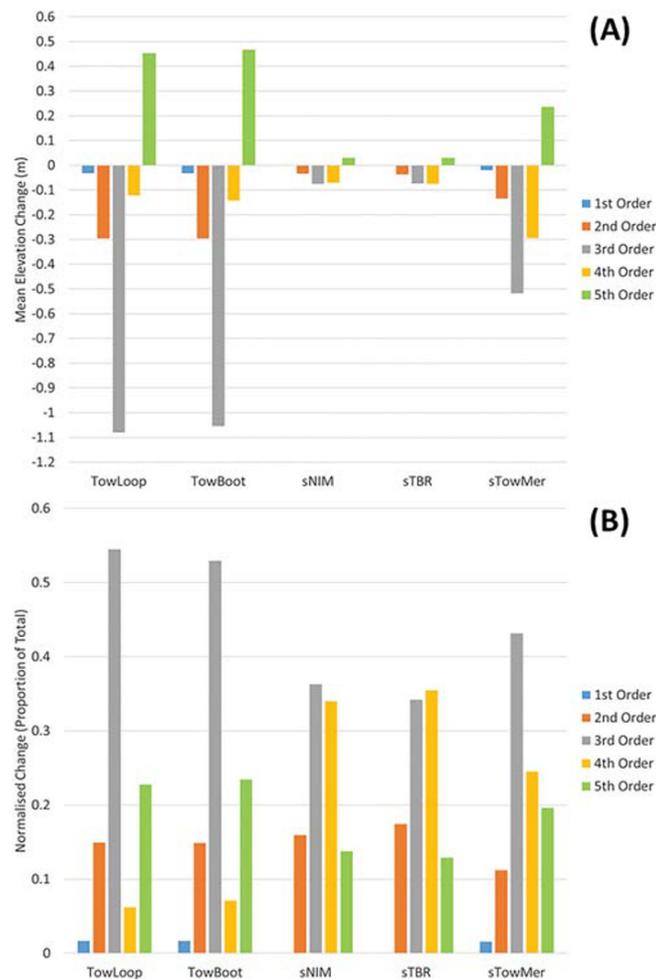
FIGURE 6. (A) Changes in the mean elevation across the different stream orders after 1500 years of simulation using each of the STREAP rainfall products. (B) The same changes normalized based on the proportion of total change within each of the stream orders. [Colour figure can be viewed at wileyonlinelibrary.com]

Table II. Summary of mean elevation changes across the catchment by 50-m DEM cells over 1500 years of simulation

Elevation change	sNIM	sTBR	TowLoop	sTowMer
Sum of all cells (10^6 m^3)	−3.2	−3.6	−23.7	−12.2
Mean of all cells (m)	−0.01	−0.01	−0.08	−0.04
Greatest depositional cells (m)	6.15	8.98	11.27	12.74
Greatest erosional cells (m)	−17.83	−18.31	−54.99	−46.02
Sum of depositional cells (10^6 m^3)	7.6	8.3	24.8	18.4
Sum of erosional cells (10^6 m^3)	−10.8	−11.9	−48.5	−30.6
Proportion of eroded material leaving catchment	0.3	0.3	0.49	0.4

the Tow rain gauge consequently showed the greatest sediment yields and changes in elevation. The difference in rainfall variability between TowLoop, sNIM, and sTBR was isolated as a factor through spatially averaging a single ensemble member from sNIM and sTBR to a catchment average, and standardizing the rainfall means of each – this produced cumulative discharges where TowLoop was 2% greater than sNIM and 4% greater than sTBR, yet cumulative sediment yields were 184% greater than sNIM and 273% greater than sTBR. Even when normalized, the Tow record contained higher rainfall intensities in the upper quantiles of events compared to the other two products, and it is these rare, high-intensity events that drive much geomorphic activity. The normalized TowLoop produced a sediment yield of 59% of the original. Rainfall variability, and by extension discharge variability, is important to erosion rates and changes to basin geomorphology (Deal *et al.*, 2017; Scherler *et al.*, 2017), and here it has been identified as a key

sensitivity and source of uncertainty in LEMs. The spatial representation of the rainfall is also important here, as the most extreme values of rainfall are often associated with convective storms that exist at a smaller spatial scale than the catchment. When using the Tow gauge, there is no spatial representation, so any observation of a storm is applied to the whole catchment, yet for the other products the peak of the storm is smoothed across the catchment. Typically, applying a spatially distributed rainfall record at a coarser resolution (spatially or temporally) results in a decrease in simulated sediment yields (Coulthard and Skinner, 2016; Battista *et al.*, 2019).

Ultimately, there is no reliable method of directly observing all characteristics of rainfall. No single product can be said to be correct, or even more correct than another. Rather, each product has its skill and deficiencies that must be balanced by the user, often through combining the best elements of each. This is what this study achieves with the calibration of the
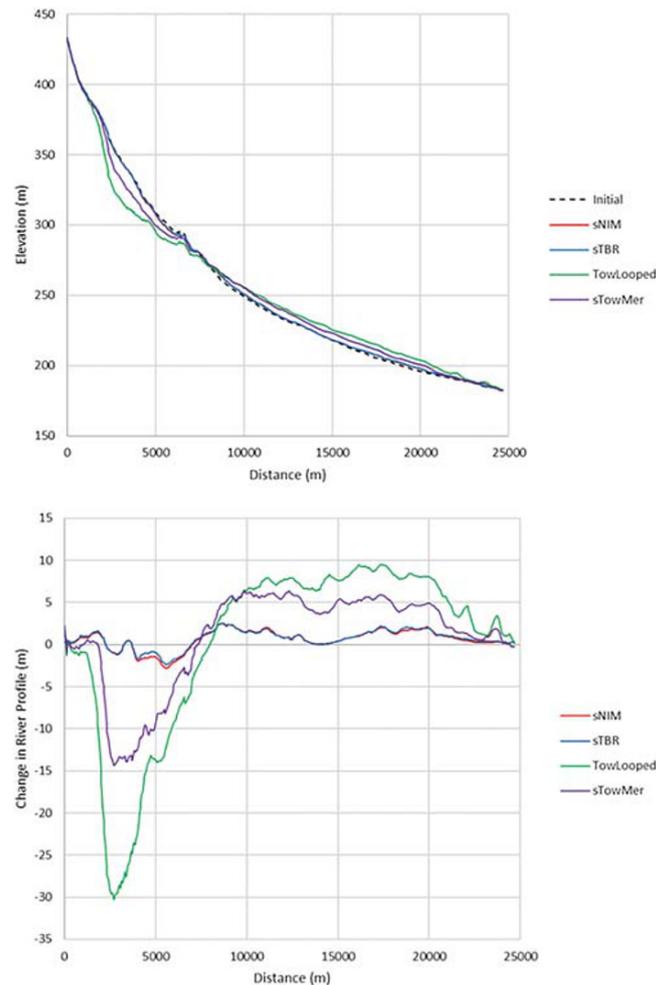
FIGURE 7. Stream profiles for the initial elevations and the final elevations after 1500 years of simulation (top), and changes from the initial elevation across the stream profile (bottom). [Colour figure can be viewed at wileyonlinelibrary.com]

sTowMer product, taking intensity from the rain gauge network, spatial and temporal coverage from the radar, and finally conditioning the distribution of events using the longer record afforded by the single gauge. This was by no means a novel approach for producing rainfall estimates, but it is for the application to a LEM, and this merged product produced outputs that were between those from the gauge ensemble and those of the other two products – and showed greater variability than sTBR and sNIM. Consequently, the additional variability from the Tow gauge allowed for higher-intensity upper-quantile events, increased discharges, and sediment yields, yet the better representation of the rainfall gradient across the catchment allowed by the information from the NIMROD and TBR network removed the spatial bias that is seen in the Tow only-based products (closer to the total produced by rainfall mean normalized TowLoop).

Geomorphic multiplier

The term 'geomorphic multiplier' was coined in Coulthard *et al.* (2012b) to describe the amplifying effect a catchment has on the signal moving from rainfall to discharge to sediment yield. For example, in Coulthard *et al.* (2012b) the predicted increase in rainfall by 1.28 times for a single extreme event resulted in estimated sediment yields over five times greater. Sediment response to flow variation is complex, non-linear, and difficult to predict (Gomez and Church, 1989; Cudden and Hoey, 2003; Coulthard *et al.*, 2007; Van De Wiel and

Coulthard, 2010). Field and laboratory observations have been used to develop sediment transport formulas in an attempt to quantify the relationship between flow velocities, shear stress, and transport initiation, and they often feature a cubic term linking changes in flow velocity to transport initiation. In the model itself, the geomorphic multiplier is expressed through the application of these physically derived equations (in this case, Wilcock and Crowe, 2003). The importance of this is clearly illustrated in our findings – where geomorphic responses to any differences between products are amplified. Effectively, as part of an uncertainty cascade, the geomorphic multiplier is magnifying any upstream uncertainty before uncertainty associated with the model (e.g. parameter uncertainty) is accounted for, increasing overall sensitivity.

This is evident in the results, where the differences in the rainfall transfer broadly linearly to the discharge yet resulting in much greater, non-linear, differences in the sediment yields (Figure 5). For example, both TowLoop and TowBoot produced cumulative discharge totals that were 150% greater than sTBR, yet cumulative sediment yields that were 1000% greater. A large difference was also observed in the scale of changes to the basin geomorphology (Figure 6).

Spatial divergence in geomorphic response and equifinality

Some authors have discussed the possibility of a 'geomorphic equifinality' within LEMs (e.g. Hancock *et al.*, 2016), suggesting

that landscape evolution is dominantly driven by the topography of the landscape itself, and thus LEMs are robust to input uncertainties over long-term simulations (e.g. Willgoose *et al.*, 1991a,b; Howard, 1994; Howard *et al.*, 1994; Hancock *et al.*, 2016). This concept would suggest that, regardless of differences in the rainfall products used here, given enough time, each would produce the same spatial pattern of geomorphic change, albeit to different extents. However, Figure 7 shows that there are differences in the patterns of erosion and deposition, especially between the Tow-based products and those of sNIM and sTBR, particularly in the third-order areas of the catchment where the TowLooped and sTowMer showed deep incision. One explanation would be that given enough time, the sNIM and sTBR would eventually match these levels of incision in third-order areas, however, for this to be the case it would be expected that the proportion of eroded material leaving the catchment would be the same, yet this is 0.30 for both sNIM and sTBR, 0.49 for TowLoop, and 0.40 for TowMer (Table II). The greater rainfall (volumes and intensities) provided by the climatic information in the Tow rain gauge produced greater stream powers within the catchment and consequently a greater proportion of the eroded sediment is removed from the system, causing the upstream area of erosion to extend further downstream. As sNIM and sTBR did not generate these greater stream powers, a greater proportion of the eroded sediment was deposited within the system, and this makes it unlikely that given enough time, the river profiles will resemble those for TowLoop and sTowMer. Therefore, our results should provoke concern – there are some clear and sometimes large differences (an order of magnitude) in the geomorphic response to different products, even over short timescales, which could potentially produce diverging landscapes over longer timescales than used in this study. Further work would be required to understand to what extent rainfall product differences influence the outputs of tests across different timeframes.

LEMs and climate change studies

The use of a rainfall generator model in combination with a LEM presents some intriguing benefits going forward that have not been explored here. The calibrations of a rainfall generator model are based on observations and assumed to represent the present climate. By altering the calibration parameters, including altering the spatial structure of rainfall and locations of convective events within catchments, possible future climate can be generated (Peleg *et al.*, 2019, 2020a). This has great potential for modelling future climate scenarios, where changes to the type of rainfall an area receives are likely to occur. For example, in the UK, climate predictions suggest that the frequency and intensity of extreme convective events are likely to increase disproportionately to changes in mean annual rainfall (Fowler and Ekström, 2009), and in Coulthard *et al.* (2012b) these changes were shown to have a non-linear impact on simulated sediment yields. However, as the version of CAESAR applied in that study could only utilize catchment average rainfall values, the impact is possibly underestimated, as representing convective events in this way smooths out local rainfall intensities and this has been shown to influence sediment yields (Coulthard and Skinner, 2016). This combination of a gridded high-resolution rainfall generator model and LEM could be used to investigate the impacts of convective events on catchments at high spatial and temporal resolution.

However, there must be caution in their applications for such. As in Coulthard *et al.* (2012b), we urge the geomorphic community to embrace probabilistic methods, even for exploratory use. Owing to the uncertain nature of rainfall estimation

and the propagation of this uncertainty into downstream applications, relevant fields, such as meteorology and hydrology, predominantly use probabilistic methods. Instead of representing rainfall as a single deterministic input, an ensemble of unique, yet equally probable, representations of rainfall are used instead. These are often combined with downstream models that also use probabilistic methods to represent model uncertainty (e.g. parameter uncertainty), providing a large ensemble output covering the range of probable outcomes. Both Coulthard *et al.* (2012b) and this study made use of rainfall ensembles but have not applied them to a probabilistic representation of the model for a full uncertainty cascade. The framework outlined by Pappenberger *et al.* (2005) could be adapted/expanded to do this.

Wider implications

For the potential application of LEMs in forecasting, in a similar way to hydrological models at present, several issues need to be addressed. First, as discussed above, they currently do not make use of probabilistic methods to assess uncertainty. Second, the models display far greater sensitivity to the rainfall input with regard to the choice of product (this study), resolution of the data (Coulthard and Skinner, 2016), and changes to the storm structure (Peleg *et al.*, 2020b). Third, geomorphic models retain a 'memory' of past events in a way hydrological models do not, where an erroneous change to the landscape is retained and will influence all future outputs. Fourth, there is a paucity in both metrics and observation data suitable for assessing the performance of the models (Tucker and Hancock, 2010). The developments of high-performance and cloud computing remove some of the barriers to probabilistic modelling using LEMs, yet it is likely that the barriers are challenging enough to impede the application of LEMs in the same way as hydrological models for decades to come. The development of the next generation of LEMs should seek to address these issues and also focus on determining what information can be extracted from the models that is of interest to stakeholders and decision-makers and is also reliable (i.e. useful). Some previous studies provide examples of what useful information can be extracted (e.g. Lane *et al.*, 2007; Coulthard *et al.*, 2012b).

Conclusions

Numerical models of geomorphic processes have been shown to be sensitive to numerous factors, with these sensitivities often being more acute than in models that simulate hydraulic processes alone. Here, the sensitivity of model outputs to rainfall inputs has been assessed using rainfall observations produced by different methods, showing that whilst the differences between these products transfer linearly to the hydraulic outputs from the model, the differences produced in the sediment yields were non-linear, magnifying the differences between the inputs. This poses a problem for the user as it is not possible to state that any single product is correct, or even the most correct, with each having a different skill for observing different characteristics of rainfall. The use of probabilistic methods to represent rainfall in LEM studies should be common practice in order to account for these uncertainties, and future development of LEMs should focus on better handling the non-linear uncertainty cascade resulting from the use of deterministic sediment transport equations. The use of LEM and a rainfall generator in combination, with full accounting for uncertainties, has the potential to provide useful information relating to the

impact climate change will have on the landscape, and consequently on society.

Data Availability Statement

The data used in this study can be made available on request.

The CAESAR-Lisflood model code can be found at https://sourceforge.net/projects/caesar-lisflood/

Conflict of Interest

The authors declare no conflict of interest.

References

Bates PD, Horritt MS, Fewtrell TJ. 2010. A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *Journal of Hydrology* **387**: 33–45. https://doi.org/10.1016/j.jhydrol.2010.03.027

Battista G, Molnar P, Burlando P. 2019. Modelling impacts of spatially variable erosion drivers on suspended sediment dynamics. *Earth Surface Dynamics Discussions*: Preprint. https://doi.org/10.5194/esurf-2019-59

Berne A, Krajewski WF. 2013. Radar for hydrology: unfulfilled promise or unrecognized potential? *Advances in Water Resources* **51**: 357–366. Available online at: https://www.sciencedirect.com/science/article/pii/S0309170812001157

Beven K, Kirkby M. 1979. A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal* **24**: 37–41. https://doi.org/10.1080/02626667909491834

Blenkinsop S, Lewis E, Chan SC, Fowler HJ. 2017. Quality-control of an hourly rainfall dataset and climatology of extremes for the UK. *International Journal of Climatology* **37**: 722–740. https://doi.org/10.1002/joc.4735

Cecinati F, Rico-Ramirez MA, Heuvelink GBM, Han D. 2017. Representing radar rainfall uncertainty with ensembles based on a time-variant geostatistical error modelling approach. *Journal of Hydrology* **548**: 391–405. https://doi.org/10.1016/J.JHYDROL.2017.02.053

Chandra R, Azam D, Müller RD, Salles T, Cripps S. 2019. Bayeslands: a Bayesian inference approach for parameter uncertainty quantification in Badlands. *Computers and Geosciences* **131**: 89–101. https://doi.org/10.1016/j.cageo.2019.06.012

Ciach GJ. 2003. Local random errors in tipping-bucket rain gauge measurements. *Journal of Atmospheric and Oceanic Technology* **20**: 752–759. https://doi.org/10.1175/1520-0426(2003)20<752:LREITB>2.0.CO;2

Coulthard T, Hicks D, Van De Wiel M. 2007. Cellular modelling of river catchments and reaches: advantages, limitations and prospects. *Geomorphology* **90**: 192–207. https://doi.org/10.1016/j.geomorph.2006.10.030

Coulthard T, Neal J, Bates P, Ramirez J, de Almeida G, Hancock G. 2013. Integrating the LISFLOOD-FP 2D hydrodynamic model with the CAESAR model: implications for modelling landscape evolution. *Earth Surface* **38**: 1897–1906. https://doi.org/10.1002/esp.3478

Coulthard TJ, Hancock GR, Lowry JBC. 2012a. Modelling soil erosion with a downscaled landscape evolution model. *Earth Surface Processes and Landforms* **37**: 1046–1055. https://doi.org/10.1002/esp.3226

Coulthard TJ, Kirkby MJ, Macklin MG. 2000. Modelling geomorphic response to environmental change in an upland catchment. *Hydrological Processes* **14**: 2031–2045. https://doi.org/10.1002/1099-1085(20000815/30)14:11/12<2031::AID-HYP53>3.0.CO;2-G

Coulthard TJ, Macklin MG. 2001. How sensitive are river systems to climate and land-use changes? A model-based evaluation. *Journal of Quaternary Science* **16**: 347–351. https://doi.org/10.1002/jqs.604

Coulthard TJ, Ramirez J, Fowler HJ, Glenis V. 2012b. Using the UKCP09 probabilistic scenarios to model the amplified impact of climate change on drainage basin sediment yield. *Hydrology and Earth System Sciences* **16**: 4401–4416. https://doi.org/10.5194/hess-16-4401-2012

Coulthard TJ, Skinner CJ. 2016. The sensitivity of landscape evolution models to spatial and temporal rainfall resolution. *Earth Surface Dynamics* **4**: 757–771. https://doi.org/10.5194/esurf-4-757-2016

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**: 13–27. https://doi.org/10.5194/esurf-1-13-2013

Coulthard TJ, Van De Wiel MJ. 2017. Modelling long term basin scale sediment connectivity, driven by spatial land use changes. *Geomorphology* **277**: 265–281. https://doi.org/10.1016/j.geomorph.2016.05.027

Cudden JR, Hoey TB. 2003. The causes of bedload pulses in a gravel channel: the implications of bedload grain-size distributions. *Earth Surface Processes and Landforms* **28**: 1411–1428. https://doi.org/10.1002/esp.521

Deal E, Favre A-C, Braun J. 2017. Rainfall variability in the Himalayan orogen and its relevance to erosion processes. *Water Resources Research* **53**: 4004–4021. https://doi.org/10.1002/2016WR020030

Fatichi S, Ivanov VY, Paschalis A, Peleg N, Molnar P, Rimkus S, Kim J, Burlando P, Caporali E. 2016. Uncertainty partition challenges the predictability of vital details of climate change. *Earth's Future* **4**: 240–251. https://doi.org/10.1002/2015EF000336

Fowler HJ, Ekström M. 2009. Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes. *International Journal of Climatology* **29**: 385–416. https://doi.org/10.1002/joc.1827

Gomez B, Church M. 1989. An assessment of bedload sediment transport formulae for gravel bed rivers. *Water Resources Research* **25**: 1161–1186.

Hancock GR. 2009. A catchment scale assessment of increased rainfall and storm intensity on erosion and sediment transport for Northern Australia. *Geoderma* **152**: 350–360. https://doi.org/10.1016/j.geoderma.2009.07.003

Hancock GR, Coulthard TJ. 2012. Channel movement and erosion response to rainfall variability in southeast Australia. *Hydrological Processes* **26**: 663–673. https://doi.org/10.1002/hyp.8166

Hancock GR, Coulthard TJ, Lowry JBC. 2016. Predicting uncertainty in sediment transport and landscape evolution – the influence of initial surface conditions. *Computers and Geosciences* **90**: 117–130. https://doi.org/10.1016/j.cageo.2015.08.014

Hancock GR, Coulthard TJ, Martinez C, Kalma JD. 2011. An evaluation of landscape evolution models to simulate decadal and centennial scale soil erosion in grassland catchments. *Journal of Hydrology* **398**: 171–183. https://doi.org/10.1016/j.jhydrol.2010.12.002

Hancock GR, Lowry JBC, Coulthard TJ. 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 GR, Lowry JBC, Coulthard TJ, Evans KG, Moliere DR. 2010. A catchment scale evaluation of the SIBERIA and CAESAR landscape evolution models. *Earth Surface Processes and Landforms* **35**: 863–875. https://doi.org/10.1002/esp.1863

Hancock GR, Verdon-Kidd D, Lowry JBC. 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

Hofstra N, Haylock M, New M, Jones PD. 2009. Testing E-OBS European high-resolution gridded data set of daily precipitation and

surface temperature. *Journal of Geophysical Research* **114**: D21. https://doi.org/10.1029/2009JD011799

Hoober D, Svoray T, Cohen S. 2017. Using a landform evolution model to study ephemeral gullying in agricultural fields: the effects of rainfall patterns on ephemeral gully dynamics. *Earth Surface Processes and Landforms* **42**: 1213–1226. https://doi.org/10.1002/esp.4090

Howard AD. 1994. A detachment-limited model of drainage basin evolution. *Water Resources Research* **30**: 2261–2285. https://doi.org/10.1029/94WR00757

Howard AD, Dietrich WE, Seidl MA. 1994. Modeling fluvial erosion on regional to continental scales. *Journal of Geophysical Research: Solid Earth* **99**: 13971–13986. https://doi.org/10.1029/94JB00744

Howard AJ, Knight D, Coulthard T, Hudson-Edwards K, Kossoff D, Malone S. 2016. Assessing riverine threats to heritage assets posed by future climate change through a geomorphological approach and predictive modelling in the Derwent Valley Mills WHS, UK. *Journal of Cultural Heritage* **19**: 387–394. https://doi.org/10.1016/j.culher.2015.11.007

Jewell SA, Gaussiat N. 2015. An assessment of kriging-based raingauge–radar merging techniques. *Quarterly Journal of the Royal Meteorological Society* **141**: 2300–2313. https://doi.org/10.1002/qj.2522

Keijsers JGS, Schoorl JM, Chang KT, Chiang SH, Claessens L, Veldkamp A. 2011. Calibration and resolution effects on model performance for predicting shallow landslide locations in Taiwan. *Geomorphology* **133**: 168–177. https://doi.org/10.1016/j.geomorph.2011.03.020

Lane SN, Tayefi V, Reid SC, Yu D, Hardy RJ. 2007. Interactions between sediment delivery, channel change and flood risk in a temperate upland catchment. *Earth Surface Processes and Landforms* **32**: 429–446. https://doi.org/10.1002/esp.1404

Lewis E, Quinn N, Blenkinsop S, Fowler HJ, Freer J, Tanguy M, Hitt O, Coxon G, Bates P, Woods R. 2018. A rule based quality control method for hourly rainfall data and a 1 km resolution gridded hourly rainfall dataset for Great Britain: CEH-GEAR1hr. *Journal of Hydrology* **564**: 930–943. https://doi.org/10.1016/J.JHYDROL.2018.07.034

Marra F, Destro E, Nikolopoulos EI, Zoccatelli D, Creutin JD, Guzzetti F, Borga M. 2017. Impact of rainfall spatial aggregation on the identification of debris flow occurrence thresholds. *Hydrology and Earth System Sciences* **21**: 4525–4532. https://doi.org/10.5194/hess-21-4525-2017

Marra F, Morin E. 2015. Use of radar QPE for the derivation of intensity–duration–frequency curves in a range of climatic regimes. *Journal of Hydrology* **531**: 427–440. https://doi.org/10.1016/j.jhydrol.2015.08.064

McMillan H, Jackson B, Clark M, Kavetski D, Woods R. 2011. Rainfall uncertainty in hydrological modelling: an evaluation of multiplicative error models. *Journal of Hydrology* **400**: 83–94. https://doi.org/10.1016/j.jhydrol.2011.01.026

McMillan H, Krueger T, Freer J. 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrological Processes* **26**: 4078–4111. https://doi.org/10.1002/hyp.9384

Met Office. 2003. 5km UK Composite Rainfall Data from the Met Office NIMROD System, NCAS Br. Atmos. Data Centre. Available at: http://catalogue.ceda.ac.uk/uuid/82adec1f896af6169112d09cc1174499

Pappenberger F, Beven K, Hunter N, Bates P, Gouweleeuw B, Thielen J, de Roo A. 2005. Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS). *Hydrology & Earth System Science* **9**: 381–393. Available at .http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=10275606&A-N=18850305&h=TofnL6Iz83pxgptLAQSCT0ou71W5bb7qdXUlZV-N0JEXhnwfvlsj2S5%2BNzikYeF1%2BVRHgU9zpgCNtX4K1jEbm1Q%3D%3D&crl=c

Paschalis A, Fatichi S, Molnar P, Rimkus S, Burlando P. 2014. On the effects of small scale space-time variability of rainfall on basin flood response. *Journal of Hydrology* **514**: 313–327. https://doi.org/10.1016/j.jhydrol.2014.04.014

Paschalis A, Molnar P, Fatichi S, Burlando P. 2013. A stochastic model for high-resolution space-time precipitation simulation. *Water*

Resources Research* **49**: 8400–8417. https://doi.org/10.1002/2013WR014437

Peleg N, Ben-Asher M, Morin E. 2013. Radar subpixel-scale rainfall variability and uncertainty: lessons learned from observations of a dense rain-gauge network. *Hydrology and Earth System Sciences* **17**: 2195–2208. https://doi.org/10.5194/hess-17-2195-2013

Peleg N, Blumensaat F, Molnar P, Fatichi S, Burlando P. 2017a. Partitioning the impacts of spatial and climatological rainfall variability in urban drainage modeling. *Hydrology and Earth System Sciences* **21**: 1559–1572. https://doi.org/10.5194/hess-21-1559-2017

Peleg N, Fatichi S, Paschalis A, Molnar P, Burlando P. 2017b. An advanced stochastic weather generator for simulating 2-D high-resolution climate variables. *Journal of Advances in Modeling Earth Systems* **9**: 1595–1627. https://doi.org/10.1002/2016MS000854

Peleg N, Marra F, Fatichi S, Paschalis A, Molnar P, Burlando P. 2018. Spatial variability of extreme rainfall at radar subpixel scale. *Journal of Hydrology* **556**: 922–933. https://doi.org/10.1016/j.jhydrol.2016.05.033

Peleg N, Molnar P, Burlando P, Fatichi S. 2019. Exploring stochastic climate uncertainty in space and time using a gridded hourly weather generator. *Journal of Hydrology* **571**: 627–641. https://doi.org/10.1016/j.jhydrol.2019.02.010

Peleg N, Sinclair S, Fatichi S, Burlando P. 2020a. Downscaling climate projections over large and data sparse regions: methodological application in the Zambezi River Basin. *International Journal of Climatology*. Early View. https://doi.org/10.1002/joc.6578

Peleg N, Skinner C, Fatichi S, Molnar P. 2020b. Temperature effects on the spatial structure of heavy rainfall modify catchment hydro-morphological response. *Earth Surface Dynamics* **8**: 17–36. https://doi.org/10.5194/esurf-8-17-2020

Poeppl R, Keesstra S, Keiler M, Coulthard T, Glade T. 2013. Impact of dams, dam removal and dam-related river engineering structures on sediment connectivity and channel morphology of the Fugnitz and the Kaja. *Research in Protected Areas* **5**: 607–614. Available from: http://www.landesmuseum.at/pdf_frei_remote/NP_Hohe_Tauern_Conference_5_0607-0614.pdf

Saynor MJ, Lowry J, Erskine WD, Coulthard T, Hancock G. 2012. Assessing erosion and run-off performance of a trial rehabilitated. *Proceedings of the Life of Mine Conference, July* **2012**: 10–12.

Scherler D, DiBiase RA, Fisher GB, Avouac JP. 2017. Testing monsoonal controls on bedrock river incision in the Himalaya and Eastern Tibet with a stochastic-threshold stream power model. *Journal of Geophysical Research: Earth Surface* **122**: 1389–1429. https://doi.org/10.1002/2016JF004011

Skinner C, Coulthard T. 2017. Caesar-Lisflood existing applications parameter listings – May 2017. https://doi.org/10.5281/ZENODO.800558

Skinner CJ, Coulthard TJ, Schwanghart W, Van De Wiel MJ, Hancock G. 2017. Global sensitivity analysis of parameter uncertainty in landscape evolution models. *Geoscientific Model Development Discussions*: 1–35. https://doi.org/10.5194/gmd-2017-236

Smith MW, Carrivick JL, Hooke J, Kirkby MJ. 2014. Reconstructing flash flood magnitudes using "Structure-from-Motion": a rapid assessment tool. *Journal of Hydrology* **519**: 1914–1927. https://doi.org/10.1016/j.jhydrol.2014.09.078

Tucker GE, Bras RL. 2000. A stochastic approach to modelling the role of rainfall variability in drainage basin evolution. *Water Resources Research* **36**: 1953–1964. https://doi.org/10.1029/2000WR900065

Tucker GE, Hancock GR. 2010. Modelling landscape evolution. *Earth Surface Processes and Landforms* **35**: 28–50. https://doi.org/10.1002/esp.1952

Villarini G, Krajewski WF. 2010. Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. *Surveys in Geophysics* **31**: 107–129. https://doi.org/10.1007/s10712-009-9079-x

Villarini G, Mandapaka PV, Krajewski WF, Moore RJ. 2008. Rainfall and sampling uncertainties: a rain gauge perspective. *Journal of Geophysical Research* **113**: D11. https://doi.org/10.1029/2007JD009214

Van De Wiel MJ, Coulthard TJ. 2010. Self-organized criticality in river basins: challenging sedimentary records of environmental change. *Geology* **38**: 87–90. https://doi.org/10.1130/G30490.1

Wilcock PR, Crowe JC. 2003. Surface-based transport model for mixed-size sediment. *Journal of Hydraulic Engineering* **129**: 120–128. https://doi.org/10.1061/(ASCE)0733-9429(2003)129:2(120)

Wilks DS, Wilby RL. 1999. The weather generation game: a review of stochastic weather models. *Progress in Physical Geography* **23**: 329–357. https://doi.org/10.1177/030913339902300302

Willgoose G, Bras RL, Rodriguez-Iturbe I. 1991a. A coupled channel network growth and hillslope evolution model: 1. Theory. *Water Resources Research* **27**: 1671–1684. https://doi.org/10.1029/91WR00935

Willgoose G, Bras RL, Rodriguez-Iturbe I. 1991b. Results from a new model of river basin evolution. *Earth Surface Processes and Landforms* **16**: 237–254. https://doi.org/10.1002/esp.3290160305

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Comparison of Rainfall Products
Data S2. Evaluation of STREAP Implementation