Available online at www.sciencedirect.com
ScienceDirect
Geomorphology 90 (2007) 263–282
GEOMORPHOLOGY
www.elsevier.com/locate/geomorph

Does hydrological connectivity improve modelling of coarse sediment delivery in upland environments?

Simon C. Reid [a], Stuart N. Lane [b,*], David R. Montgomery [c], Christopher J. Brookes [d]

[a] School of Geography, University of Leeds, Leeds, LS2 9JT, UK
[b] Department of Geography, University of Durham, Durham, DH1 3LE, UK
[c] Department of Earth and Space Sciences/Quaternary Research Center, University of Washington, Seattle, WA98195, USA
[d] School of Earth Sciences, Victoria University of Wellington, PO Box 600, Wellington, New Zealand

Received 30 October 2005; received in revised form 6 May 2006; accepted 17 October 2006
Available online 21 April 2007

Abstract

Modelling the delivery of landslide-generated sediment to channel networks is challenging due to uncertainty in the magnitude–frequency distribution of failures connected to the channel network. Here, we investigate a simplified treatment of hydrological connectivity as a means for improving identification of coarse sediment delivery to upland rivers. Sediment generation from hillslopes and channel banks and its delivery to the channel network are modelled based on a modified form of SHALSTAB coupled to a network index version of TOPMODEL. The network index treatment has two important hydrological effects: (a) it only allows saturated areas to connect to the hydrological network when there is full saturation along the associated flow path; and (2) overland flow associated with unconnected but saturated zones is assumed to remain within the catchment and to contribute to a reduction in the catchment-averaged saturation deficit. We use this hydrological treatment to restrict sediment delivery to situations where there is surface hydrological connection (i.e. saturation) along the complete flow path that connects failure areas to the drainage network. This represents an extreme restriction on the possibility of connected failure as it does not allow for failed material to connect if failures are associated with partial saturation or where delivery involves runout across areas where hydrological connection is not maintained. The impact of this restriction is assessed by comparing model predictions with field mapping of connected failures and data from continuously recording coarse sediment sensors, for two storm events. The hydrological connection requirement restricted connected failures to zones closer to the drainage network and resulted in a better level of agreement with the field mapped failures. Simulations suggested that in the study catchment the majority of sediment inputs occur from hydrologically-connected areas close to the channel network during moderate sized rainstorms that occur relatively frequently.

Keywords: TOPMODEL; SHALSTAB; SEDMAP; Reduced complexity modelling; Coarse sediment delivery; Hydrological connectivity; Network index; Topographic index

1. Introduction

The identification of sediment source areas and the way they connect to the channel network are fundamental to environmental management in upland environments, especially where delivery rates to the drainage

* Corresponding author.
E-mail address: s.n.lane@durham.ac.uk (S.N. Lane).

0169-555X/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.geomorph.2006.10.023

network exceed the capacity of the network to transport that sediment. Sediment accumulation in the drainage network affects stage–discharge relationships and increases water levels (and hence flood risk) associated with given flows (e.g. Hey and Winterbottom, 1990; James, 1999; Stover and Montgomery, 2001). Sediment accumulation also needs to be understood as a result of its potential habitat impacts (Rimmer et al., 1983; Cunjak, 1988; Heggenes, 1996; Rice et al., 2001), which may differ according to the size distribution of delivered material.

Hillslope sources of coarse sediment commonly involve failure thicknesses of between 0.5 and 3.0 m as rain infiltrating into the soil increases soil pore-water pressure and reduces the shear strength of the soil (e.g., Hearn and Griffiths, 2001), simultaneously increasing the downslope component of the weight of the soil mass. Whilst the general mechanisms of slope failure are well-established, the actual process is complex and depends upon: local slope; soil depth; the degree of differentiation between the soil horizons; soil saturation; and antecedent rainfall conditions (Brooks et al., 1993, 2004). Thus, hillslope failures are spatially complex and commonly scale-dependent. The generation of fully-saturated conditions is uncommon on very steep, highly conductive soils, but can be significant in lower gradient unchannelled valleys (e.g. Dunne and Black, 1970; Wilson and Dietrich, 1987; Dietrich et al., 1995; Montgomery and Dietrich, 1995), where the soils are shallow, or where the vertical structure of the soil impedes drainage (e.g. Brooks et al., 1993). Although bedrock has considerable cohesion, as well as frictional strength, and soil is generally more conductive, the underlying bedrock may also be highly fractured and may conduct large amounts of storm flow (Wilson and Dietrich, 1987; Johnson and Sitar, 1990; Montgomery et al., 1997, 2002). Thus, there may be broad geological controls of the propensity to higher levels of saturation, and therefore slope stability and sediment delivery rates.

The last 20 years has seen significant progress in the application of slope stability models at the catchment-scale due to growing availability of high-resolution topographic data, improvements in computational aspects of data handling, and the associated development in modelling capabilities. Much of this progress has been based upon catchment-scale application of infinite-plane slope models, often in a Geographical Information System (GIS) framework, to determine the conditions at which failure would occur. At the same time, there has been a move away from analyses that predict absolute hazard towards a more probabilistic approach that emphasises the likelihood of failure (e.g. Montgomery and Dietrich, 1994; Montgomery et al., 1998; Dietrich et al., 2001) in order to provide a different means of considering failure as compared to the traditional stable/unstable output of slope stability models.

In some situations (e.g., assessing the impacts of future climate changes or alternative land management options upon sediment delivery rates) it is necessary to determine sediment delivery rates, over time, in response to the series of rainfall events impacting upon a catchment. The latter causes the failure likelihood to evolve spatially and temporally. Under the assumption that the hydraulic gradient of the saturated zone can be approximated by the local surface topographic slope (Beven, 2000), estimation of soil saturation dynamics is possible using topographic indices (Beven and Kirkby, 1979; O'Loughlin, 1981; Moore et al., 1993; Wolock, 1993). If the dynamics of the water table can be approximated by uniform subsurface runoff production per unit area over the area draining through a point (Beven, 2000), and with an exponential decline of hydrological transmissivity with depth, then the local wetness is given by $\ln(a/\tan\theta)$, where a is the area drained per unit contour or the specific area, and θ is the slope in degrees (Moore et al., 1993; Wolock, 1993). It is then possible to account for moisture at the level of sub-catchments, as in TOPMODEL (Beven and Kirkby, 1979). As the sub-catchment-averaged saturation deficit is updated in response to rainfall inputs and catchment outflow, comparison back to the local topographic index allows identification of the spatio-temporal development of changing levels of saturation and therefore the proportion of the local soil thickness that is saturated. Thus, the TOPMODEL approach provides a computationally efficient mechanism for driving the dynamic saturation component of failure models.

In this paper, we develop the SHALSTAB approach to test, explicitly, the role played by surface hydrological connection in delivering coarse sediment sources to the river channel system. The way in which hillslopes couple to the channel system has long been recognised as a key control upon coarse sediment delivery to the system and subsequent river channel and floodplain response (e.g. Harvey, 1991; Gerrard, 1999; Harvey, 2001, 2002). Concurrently (e.g. Bracken and Croke, 2007), we now recognise that the ease of connection of sources of surface overland flow to the drainage system is as important a control on flood hydrology as the generation of rapid runoff due to local hydrological dynamics (e.g. water table rise). Connectivity has been shown to be a central component of understanding fine

sediment and soil erosion processes (Wainwright et al., 2002) but despite field evidence that it matters in relation to coarse sediment delivery, most coarse sediment delivery models fail to include a connectivity component. This raises an interesting question: what are the implications of hydrological disconnection for the delivery of failed sediment to the drainage network? Traditionally, landslide runout distance (L) has been described as a function of height loss (H) under the assumption that greater height loss leads to longer runout distances. However, H/L ratios are commonly much lower than expected from simple friction analyses (Legros, 2002) which may mean that fluids exert a significant control upon landslide runout distance. Bagnold (1954) showed that addition of an interstitial fluid could reduce the effective coefficient of friction of a granular material by increasing interstitial pore-water pressures, thereby reducing effective normal stresses. Legros (2002) showed: (1) that if the fluid pressure gradient is hydrostatic, the effective coefficient of friction may become low if the density of the interstitial fluid (water and fine particles) is high relative to that of the larger particles; and (2) that if the fluid pressure gradient is in excess of hydrostatic, the whole load of solid material may be supported by the fluid, and viscous dissipation becomes dominant, proportional to velocity gradient, and with energy dissipation potentially higher on steeper slopes. As landslide runout velocities appear to be low, Legros argues the associated high velocities at failure must be rapidly dissipated, with the landslide flowing (as a fluid) rather than sliding, inertial terms of reduced importance, and runout distance dependent upon local slope and the thickness of the moving material (as with debris flows). Water may be particularly important in this process as it is dense, incompressible and viscous, such that large volumes of water may not be required to achieve fluidization as compared with air (Legros, 2002). The zone of fluidization may not need to extend throughout the landslide; the presence of a saturated layer of material at the base of a landslide may extend landslide runout (Legros, 2002). Whilst failed material is likely to contain water that can contribute to this process, there is the possibility of entraining additional water if the landslide takes a flow path that is fully saturated, whether from overland flow or saturated sediment over which the landslide passes. Thus, the presence of hydrological connection, along a flow path from failed material source to the drainage network, may be a condition that discriminates between failures that deliver sediment to the drainage network and those that do not. There are three conditions where this is likely to be of importance: (1) where the failure volumes are small, and where higher pore-water pressures may counter the reduction in landslide depth as it flows; (2) where zones of lower gradient topography occur along a flow path, and where the reduction in normal stress helps to sustain transport at lower gradients; and (3) where there is a mix of much finer (i.e. silt and sand) and much coarser material, with the finer material assisting with fluidization.

Lane et al. (2004) describe a simple modification of the topographic index, the network index, to represent surface hydrological connectivity. The network index describes the level of catchment-averaged saturation required for a saturated zone to be hydrologically-connected to the drainage network along the complete hydrological flow path that connects it to the drainage network. The network index is defined for each location within each sub-catchment. We assume that the topographic index represents the propensity to generate saturation overland flow at a location. If we consider all the locations along a hydrological flow path from a possible source of overland flow through to the drainage network then the lowest value of the topographic index along that flow path becomes the limiting location for full surface hydrological connection. Thus, each location has two index values: (1) a topographic index, which describes the local propensity to generate saturation overland flow; and (2) a network index, which describes the propensity for that location to be connected by surface hydrological flow to the drainage network. Determination of a network index allows a basic representation of the dynamics of overland flow generation within a TOPMODEL formulation, by only allowing the delivery of overland flow to the drainage network if there is complete saturation along a given flow path, and then with an appropriate delay for travel time, at each point in time (see Lane et al., 2004). As the catchment-averaged saturation deficit falls as the catchment wets up, so the network index required for connection falls. Hence, saturated areas become more readily connected to the drainage network. This allows dynamic simulation of both the development of connected overland flow and how significant areas of saturation may remain unconnected during storm events (Lane et al., 2004). The question that we address in this paper is, given that landslide runout distances may be enhanced through saturation along a given flow path, can estimates of the delivery of failed sediment to the drainage network be improved by requiring failure to occur within a hydrologically-connected surface flow path. Thus, we extend the SHALSTAB approach (Montgomery and Dietrich, 1994) to include a

simplified but dynamic representation of hydrological connectivity, and assess the impact of this on sediment delivery to the drainage network.

2. Model development

The model we develop in this paper seeks to identify both hillslope and channel bank sources in situations where high quality, high-resolution topographic data are available. We refer to the model as SEDMAP. Slope stability is derived from a modified form of SHALSTAB (Montgomery and Dietrich, 1994; Montgomery et al., 1998; Dietrich et al., 2001), leading to determination of relative failure likelihood (F) through the ratio of the steady state rainfall intensity to transmissivity (Q/T) needed to generate the degree of soil saturation (h/z) necessary to cause slope failure. The failure likelihood equations are given by:

$$F_{\mathrm{co}} = \left[\frac{C_{\mathrm{s}} + C_{\mathrm{r}}}{\rho_{\mathrm{w}} g z \cos^2\theta \tan\phi} + \frac{\rho_{\mathrm{s}}}{\rho_{\mathrm{w}}}\left(1 - \frac{\tan\theta}{\tan\phi}\right)\right] \quad (2a)$$

which for the case of cohesionless soils with no root cohesion reduces to:

$$F_{\mathrm{no}} = \left[\frac{\rho_{\mathrm{s}}}{\rho_{\mathrm{w}}}\left(1 - \frac{\tan\theta}{\tan\phi}\right)\right] \quad (2b)$$

where F is the failure likelihood as expressed by the critical value of Q/T, C_{s} is soil cohesion (KPa), C_{r} is root cohesion (KPa), ρ_{w} is water density (kg m^{-3}), ρ_{s} is the soil bulk density (kg m^{-3}), g is the acceleration due to gravity (m s^{-2}), z is soil depth (m), θ is the slope angle (°) and ϕ is the soil internal friction angle (°). SHALSTAB does not predict failure. Rather, it seeks to classify the landscape into zones of equal probability of failure. In order to predict failure, we specify the maximum Q/T found in a given event, and use this to determine the locations and magnitude of failure during that event.

We restrict our analysis to an end member case defined for situations where runout distances, in the absence of fluidization within or throughout the moving material, are insufficient to connect failures to the drainage network. In this end member, we only define a failure as connected if it occurs within a flow path that is hydrologically-connected within the same event. Hence, as the drainage network expands and contracts during a storm event, network extension into zones of failure becomes the mechanism by which failures connect to the drainage network. As a result of the way that the network index is defined, full saturation of the soil column is also required for a connected failure, whereby providing a doubly-restricted criterion for failure (point saturation and hydrological connection). Whilst hydrological connection is required throughout a flow path, this does not lead to failure throughout a flow path as the failure condition is not only dependent upon saturation. This end member state is likely to be associated with: (1) smaller failure volumes given the positive correlation observed between volume of failure and runout distance (Legros, 2002); and/or (2) complex topography where low or reverse gradients along a flow path allow for the probability of deposition before connection with the drainage network. Shorter return period rainfall events are likely to result in more restricted patterns of surface saturation as well as smaller extents and durations of hydrological connectivity. This will result in more spatially restricted zones of failure and hence smaller failure volumes. These are likely to be more dependent upon fluidization in order to connect them to the drainage network. Following Legros (2002), larger failure volumes are more likely to connect as they result in longer runout distances. If the onset failure is driven by increasing pore-water pressures (i.e., increasing levels of saturation), larger volumes will result from longer return period rainfall events, when levels of landscape saturation are greater and hydrological connectivity is more likely to be present throughout a flow path length and so not become a limiting factor. Thus, the use of our end member is most likely to impact smaller magnitude and relatively more frequent coarse sediment delivery events, associated with shorter return period rainfall events.

The channel network herein is defined from the upslope contributing area, calculated using a multiple flow routing algorithm and 5 m resolution topographic data. Continuous hydrological simulation is used to identify a set of discrete storm events when catchment-averaged saturation deficits will lead to saturation at locations within the catchment. Each of these saturation deficits (herein called the Topographic Failure Threshold, TFT) is expressed as a topographic index in order to identify where in the catchment saturation has occurred and the failure model can be applied. Following the TOPMODEL framework (Beven and Kirkby, 1979), the saturation deficit (S) is defined as the amount of water required from rainfall or upslope drainage to bring that part of the landscape to full saturation (allowing it to become a potential coarse sediment source):

$$S = m\ln\left(\frac{R}{T_{\max}}\right) - m\ln\left(\frac{a}{\tan\theta}\right) \quad (3)$$

where, m is the 'soil parameter' equal to the soil porosity divided by the scaling parameter f which governs the

rate of decrease of the hydraulic conductivity with soil depth, R is the recharge rate of the soil and T_{max} is the maximum transmissivity of the soil. In order to operationalise (3), the equation is integrated over the catchment and divided by the total catchment area, to yield the catchment-averaged saturation deficit:

$$\bar{S} = -m\ln\left(\frac{R}{T_{max}}\right) - m\lambda \quad (4)$$

where λ is the mean topographic index for the catchment. The catchment-averaged storage deficit is used to drive the saturated zone flux (or baseflow). Rainfall (through an unsaturated zone flux) reduces the average storage deficit and baseflow increases it. At any one time, the average storage deficit can be mapped back onto (3) in order to identify which areas of the landscape are saturated and therefore generating overland flow.

Application of (4) in its default form assumes that all generated overland flow leaves the catchment. Lane et al. (2004), with higher resolution topographic data, questioned this assumption, arguing that a significant proportion of generated overland flow may reinfiltrate before reaching the channel network. Thus, they used a network index approach with a single flow routing algorithm to identify those cells generating overland flow that are connected to the channel network by saturated cells along an entire flow path. This is essentially a representation of network extension and contraction and has now been extended to the case of multiple flow routing. In this paper, we use the network index approach suggested by Lane et al. (2004) but modified to include the multiple flow routing algorithm of Tarboton (1997).

In the next sections of this paper, we test the effects of this restricted definition of connected failure as compared with the original version of SHALSTAB from which the approach has been developed.

3. Model application and testing

3.1. Study catchment

The model is tested for the Buckden Beck subcatchment of the River Wharfe, Yorkshire Dales National Park, U.K. (Fig. 1). The area is 2.16 km^2 with an elevation range from 218 m above Ordnance Datum (OD) (sea level) where the stream joins the main channel, to 701 m on the divide close to Buckden Pike. The catchment land cover comprises rough grass and moorland on the upper and middle slopes with low intensity pasture on the lower slopes and higher intensity pasture on the floodplain. The stream passes through the village of Buckden, although this does not alter the flow regime of the channel or severely limit the channel's capacity for sediment transfer. Modelled river sediment discharges suggest that, for the Wharfe in general, the Holocene period was characterised by short (10 to 100 year) periods of high sediment discharges (Coulthard and Macklin, 2003). These peaks correspond closely to a wetter climate indicating that climate is the main driver of changes in sediment yield for this catchment over the long term, whilst land-use changes can have large, yet short-term effects. Tree clearance in the last 2000 years has led to significantly larger sediment discharge than prior to deforestation, with the climate change trend removed (Coulthard and Macklin, 2003). The catchment receives high rainfall in the region of 1750–2000 mm per year from the prevailing westerly air streams. The long-term average for Buckden is 1710 mm, with just 433 mm of annual evapotranspiration (Heritage and Newson, 1997). The area experiences both localised convective summer thunderstorms and winter cyclones, which tend to produce high-intensity rainfall events (Merrett and Macklin, 1999).

Catchment geology comprises two types of limestone (Great Scar Limestone and Yoredale Series) and a series of interbedded sandstones and shales (Millstone Grit) deposited during the early part of the Upper Carboniferous (White, 2002). Distinctive horizontal lithologically-controlled terraces have been formed by a combination of erosion and glacial processes. The headwaters of the stream are, in many places, deeply incised into Millstone Grit and/or glacial till. Due to glacial activity and steep slopes, very thin soils overlie the bedrock/till. The soils are generally thin, commonly comprising a shallow root zone, a thin organic A horizon and a deeper B horizon, the latter comprising silty/sandy/gravely material. The main channel is bedrock floored, with negligible sediment storage. This was important as it allowed us to link predictions from the model to measurements of instream coarse sediment transport (see below).

3.2. Data sources

Two sources of Digital Elevation Model (DEM) data were available. Light-induced Detection and Ranging (LIDAR) data were available at a resolution of 2 m. These data were processed (James, 2004) in order to filter out non-topographic features (e.g. buildings) that could lead to spurious estimates of slope instability. The data were tested by comparison with real time kinematic

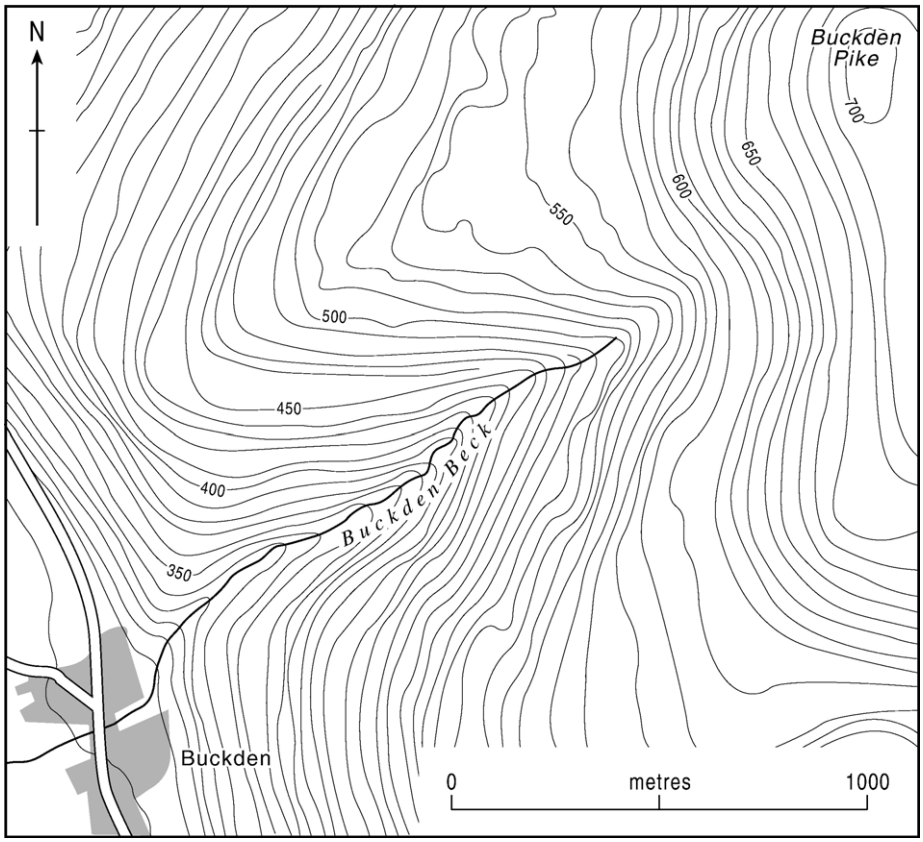

Fig. 1. The study catchment.

Global Positioning System data, with a precision better than ±0.02 m. These tests suggested that the LiDAR data were precise to ±0.22 m for this sub-catchment, which corresponds to a local slope uncertainty of ±8.8°. Second, we used INSAR data, supplied post-processed to a 5 m resolution. GPS tests suggested that the INSAR data were precise to ±0.90 m, which corresponded to a local slope uncertainty of ±14.2°.

For the hydrological component of the model, rainfall data were collected from a Campbell ARG100 Tipping Bucket rain gauge with an automated data logger (Hobo event logger). The rain gauge was located at co-ordinates 395542; 477910 (BNG) on a North West facing slope at 570 m elevation, close to the summit of the catchment. Data were available from March 19th, 2003 to March 16th, 2004 (Fig. 2). Discharge data for the catchment outlet were determined at the site of a V-notch weir at the catchment outlet using an Eijkelkamp/Van Essen pressure transducer and automated data logger. Rainfall totals (in millimetres) and mean discharges (in $m^3\ s^{-1}$) were calculated on a 15 minute basis for the 363 days and these data were applied to the network index version of TOPMODEL (Lane et al., 2004).

Fig. 3 shows the upslope contributing area used to determine the topographic index and compares the single and multiple flow routing algorithms. The single flow algorithm is dominated by vertical, diagonal and horizontal patterns, emphasising the importance of using multiple flow routing to get effective hydrological representation.

For the application of SEDMAP, the soil depth was set as 1.0 m and the bulk soil density as 1600 kg m^{-3}. Coulthard (1999) suggested that soils in neighbouring Cam Gill Beck were often thinner than 1.0 m on the steeper slopes and greater on shallower slopes, so 1.0 m was used as a first approximation. The root cohesion map (Fig. 4) required for (2a) was generated from a mosaiced series of geometrically corrected and ortho-rectified aerial photographs, with ±0.003 m root mean square error and 0.635 m ground resolution. The mosaiced photograph was subject to a histogram equalisation and then classified for land use using supervised classification, supported by field observations and an overall accuracy of over 98%. A root cohesion value was assigned to each land-use classification using representative values from Sidle (1991) and Wu and Sidle (1995) (Fig. 4).

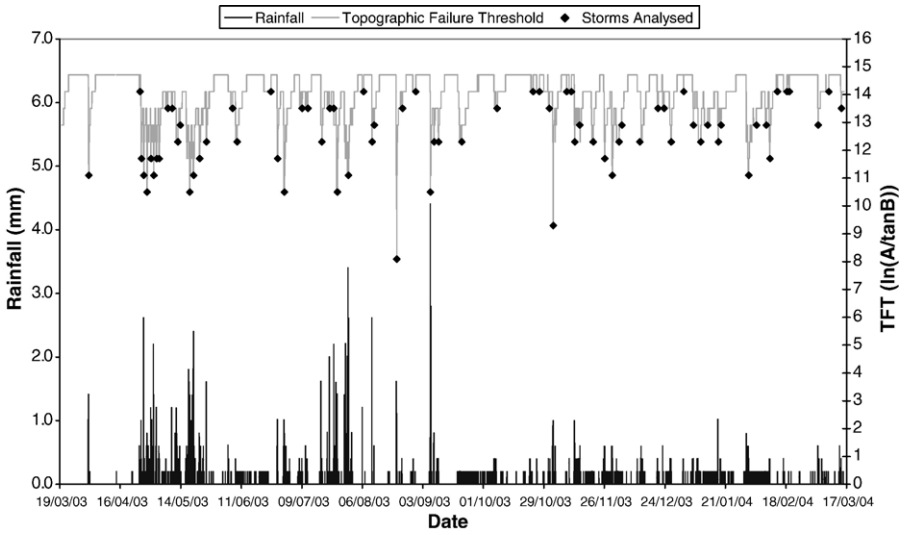

Fig. 2. Rainfall and topographic failure threshold values for Buckden Beck from March 19th, 2003 to March 16th, 2004 with the effective storms that caused sediment transport.

3.3. Parameter values

TOPMODEL discharge, the catchment-averaged saturation deficit and the topographic failure threshold (TFT) have been shown to be sensitive to two main parameters: m (the soil parameter); and T_0 (the local downslope transmissivity at soil saturation) in $m^2\ h^{-1}$. Using incorrect values of the TFT in this analysis could wrongly identify possible failures. Therefore, parameter perturbation analysis was undertaken to assist in the identification of optimal values for m and T_0, in relation to measured flows at the V-notch weir. The Nash–Sutcliffe Efficiency (Nash and Sutcliffe, 1970) criterion (NSE) was used to test the fit of the network index version of TOPMODEL's predicted catchment discharge to the observed discharge for the yearly data period at a 15 minute interval. Values of m between 0.0001 and 0.1 and T_0 between −5.0 and 2.5 were tested.

The simulation suggested that the optimal values were $m = 0.0007$ (Fig. 5a) and $T_0 = 1.5$ (Fig. 5b) which are reasonable values for small and responsive catchments in upland environments. The soil parameter (m) was optimised first, followed by T_0, since m has been shown in the literature to be the most sensitive of the TOPMODEL parameters. For the yearly period, the NSE value using the combination of optimal parameters was 0.755, which suggested a relatively close fit (Fig. 6). The optimised fit had a mean error of 0.007 $m^3\ s^{-1}$ and a standard deviation of error of 0.032 $m^3\ s^{-1}$.

3.4. Model validation: failure mapping

Field mapping of erosional scars that formed during the 12 month period was undertaken using a Leica Geosystems SR530 RTK 24 channel dual-frequency real time Global Positioning System (GPS) receiver with a mean point precision of better than ±0.005 m ± 0.5 ppm. Two main surveys were planned: late November 2003 and late March 2004. However, the focus here is on events recorded during a survey from the 16th to the 18th November 2003, as the continuous hydrological simulation suggested that the minimum value of the TFT occurred before this date. The maximum rainfall intensity for the yearly period was 8.0 $mm\ h^{-1}$ on September 6th, but this was of such short duration that the TFT values were not significantly reduced to cause many failures, and no coarse sediment output from the catchment was recorded during this event (see below). Under the model assumptions, any predicted failures in this event would have been found within the same predicted failure areas estimated for when the minimum averaged saturation deficit was estimated.

The largest event occurred on August 21st and was of relatively long duration (11 h 45 min) and high rainfall intensity with a total rainfall of 33 mm and a maximum hourly rainfall of 5.4 mm (Fig. 7a). Although this is only a moderate sized event for this sub-catchment when viewed over a decadal timescale, it became a suitable validation event as field visits indicated that it caused slope failures and levels of revegetation allowed these

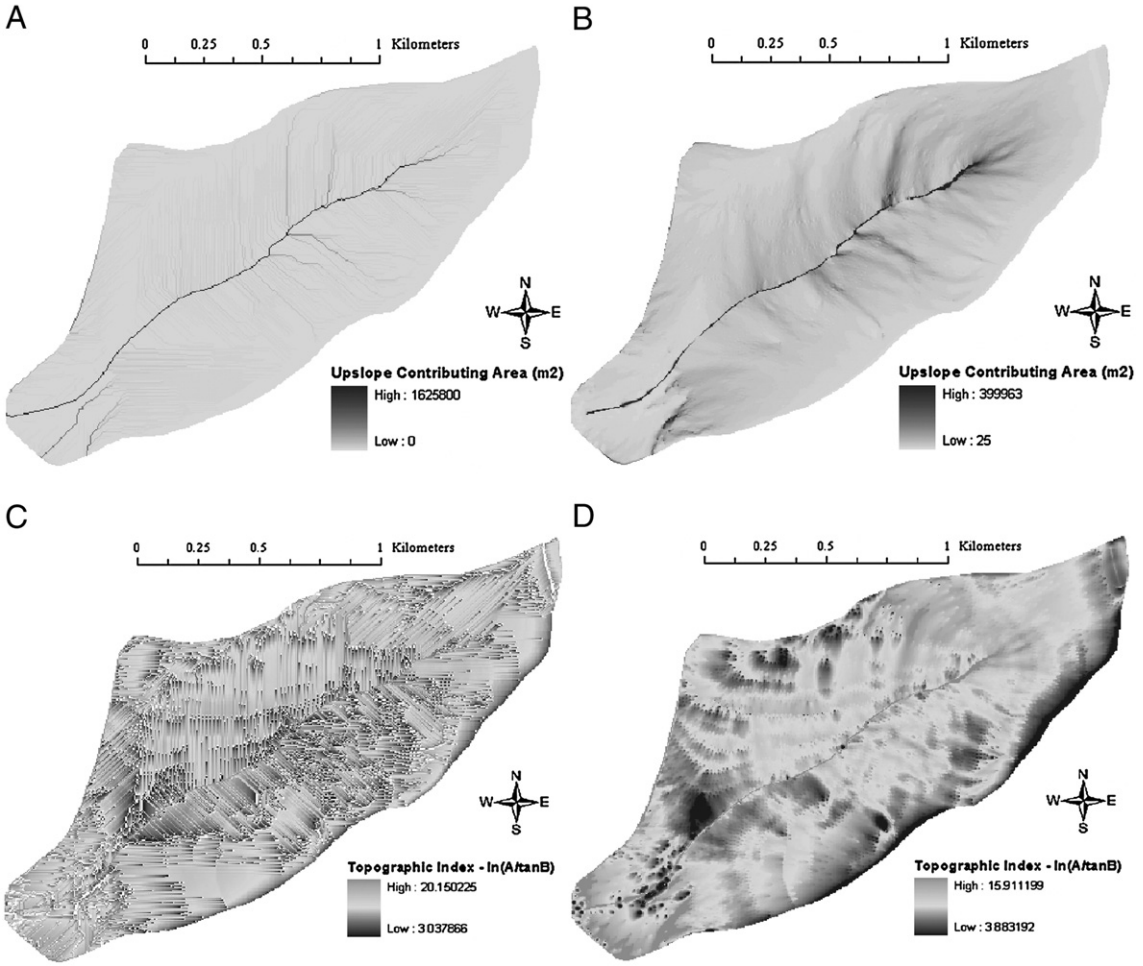

Fig. 3. (A) The upslope contributing area calculated using the ESRI single flow direction algorithm; (B) using the SEDMAP multiple flow direction algorithm (after Tarboton, 1997); (C) the topographic index calculated using the ESRI single flow direction algorithm; and (D) using the SEDMAP multiple flow direction algorithm (after Tarboton, 1997).

failures to be distinguished from subsequent scars that occurred in November. A second event also chosen for comparison resulted in much smaller scars. The storm event of November 2nd was prolonged but of low intensity with maximum rainfall of 3.2 mm h^{-1} and a total rainfall of 11.2 mm (Fig. 7b). Fresh failures observed after this event could not be dated to the previous largest storm event on September 6th, as this was observed to produce little coarse sediment.

We evaluated model performance based upon formal accuracy assessment. The model predicts two conditions; (i) connected failure, and (ii) either no failure or failure that was disconnected. The same conditions were recorded in the field. This gives a 2×2 matrix that needs to be quantified. We compute four accuracy measures. The overall accuracy is computed as the sum of the

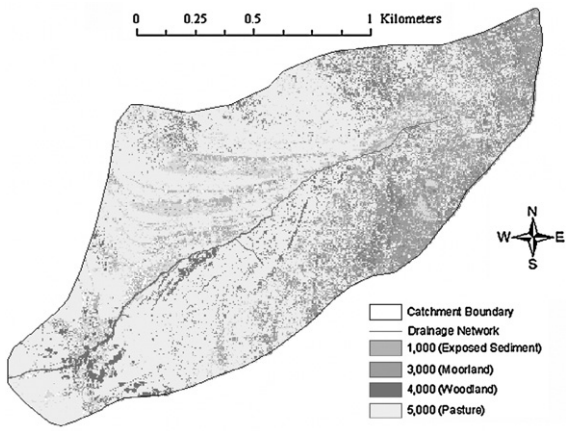

Fig. 4. The mosaiced aerial photograph classified for land use and soil cohesion (values in Pascals). The banding is not a contour artefact but the topographic expression of interbedded limestones.

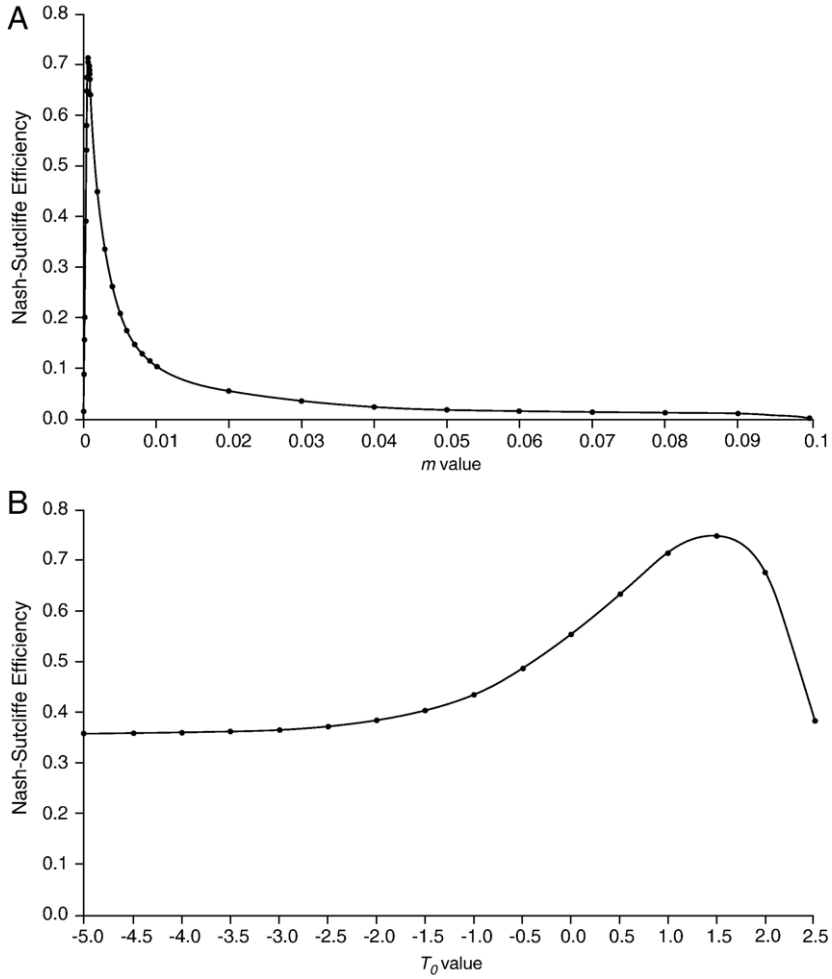

Fig. 5. Nash–Sutcliffe Efficiency values for the tested values of: (a) m (T_0=1.0); and (b) T_0 (m=0.0007).

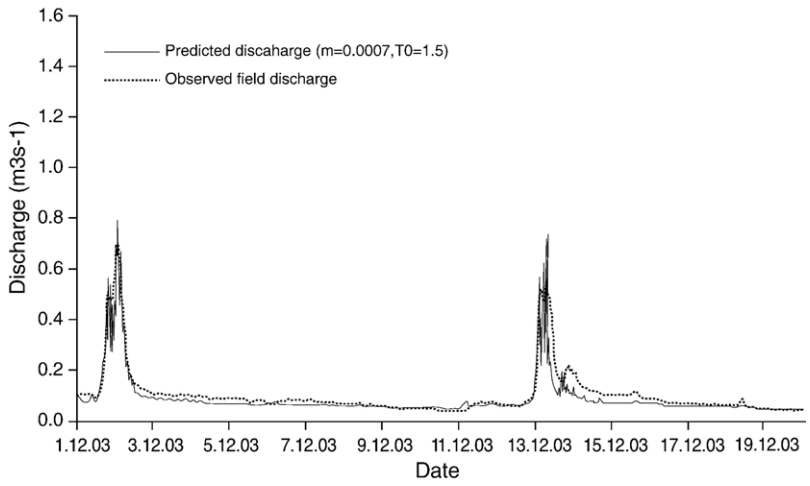

Fig. 6. The observed and predicted discharges (using the optimal parameter values) for a sample period.

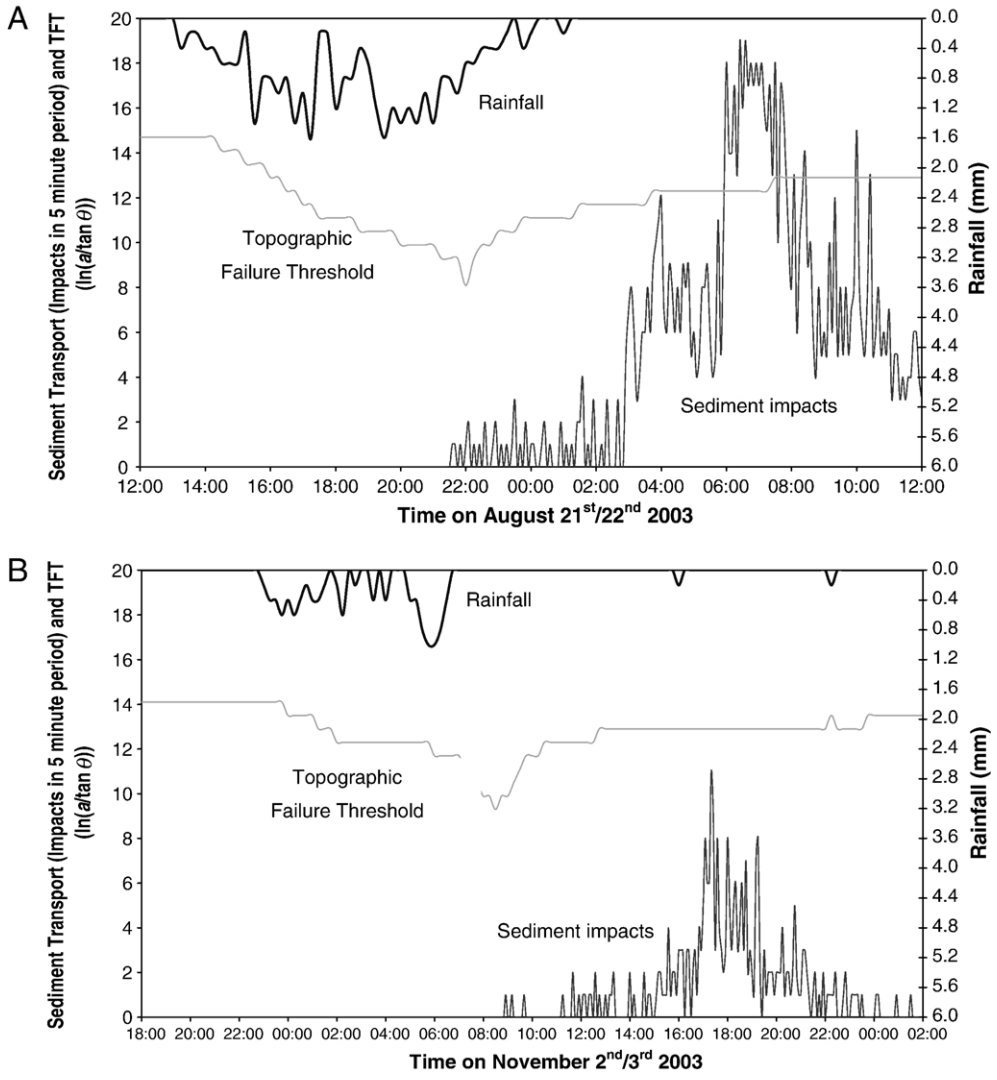

Fig. 7. Measured hourly rainfall, measured sediment impacts and modelled Topographic Failure Threshold values for Buckden Beck for: (a) August 21st; and (b) November 2nd.

correct predicted cells divided by the total number of cells:

$$\text{Overall accuracy} = \frac{\sum_{i=1}^{2} n_{ii}}{n} \tag{5}$$

where n is the number of cells where failure ($i=1$) was both observed and predicted or not observed and not predicted ($i=2$). The main disadvantage with this statistic is it is strongly dependent upon catchment area: the statistic may appear to do very well if the number of 'no failure' cells is large. It also takes no account of the expected levels of agreement due to chance, given the sub sample sizes in use. Thus, we calculate a Kappa (or Kappa–Hat) statistic, which is often cited as a more reliable measure of accuracy or agreement than overall accuracy (Cohen, 1960):

$$\hat{K} = \frac{n\sum_{i=1}^{2} n_{ii} - \sum_{i=1}^{2} n_{i+}n_{+i}}{n^2 - \sum_{i=1}^{2} n_{i+}n_{+i}} \tag{6}$$

where:

$$n_{i+} = \sum_{j=1}^{2} n_{ij} \tag{7}$$

$$n_{+i} = \sum_{i=1}^{2} n_{ij} \tag{8}$$

(6) is essentially expressing the ratio of the observed excess over chance agreement to the maximum possible excess over chance, with Kappa = 1.0 at perfect agreement and Kappa = 0.0 when observed agreement equals chance agreement (Everitt, 1998). Although this statistic still retains a dependence upon the number of no failure cells in the analysis, it is well suited to a 2 × 2 matrix, the non-diagonal terms are given greater weight and it corrects for random agreement. However, the Kappa statistic does not deal completely with the bias that occurs in the results due to the effects of increasing the number of no failure cells. Two main alternatives emerge. First, in most cases, we are interested in how the model-predicted areas of connected failure agree with the validation dataset. Thus, model performance can be assessed using a measure of fit (M) that compares the model-predicted failure extent with the validation dataset. This has been used in flood inundation modelling, where there is a similar wet–dry predicted-observed issue (e.g. Horritt and Bates, 2001, 2002; Yu and Lane, 2006).

Second, a natural extension of this, and one that is important for statistical testing, is the conditional Kappa statistic. This is based upon the maximum likelihood estimate of the Kappa coefficient for the conditional agreement of the ith category. The expected number of cells that would fail (n_{ii}) under a random simulation is ($n_{i+}n_{+i}/n$) and the maximum number of cells that are predicted as failing (n_{i+}) is the maximum number of cells that could be classified correctly as failed. Thus, denoting failure as f,

$$\hat{K}_f = \frac{n_{ff} - \left(\frac{n_{f+}n_{+f}}{n}\right)}{n_{f+} - \left(\frac{n_{f+}n_{+f}}{n}\right)} = \frac{nn_{ff} - n_{f+}n_{+f}}{nn_{f+} - n_{f+}n_{+f}}. \quad (10)$$

Use of Eq. (10) eliminates the effects on the numerator of a large number of cells that are always stable in both the model and the classification data, although both the numerator and the denominator are affected by the total number of cells used. We use Eq. (10) in this study.

Although this kind of accuracy assessment provides a means of quantifying levels of agreement, the associated statistics still need to be interpreted in terms of what is an acceptable level of agreement. In a formal sense, the only inference that can be made from Kappa values is where they are greater than zero. In this case, the model is delivering more information than would be derived if the model were replaced by labelling a random sample of locations from the sub-catchment. Thus, we adopt the subjective classification of Kappa values recommended by Altman (1991) (Table 1) in order to assist with reporting model performance.

3.5. Model assessment: sediment movement data

Model assessment also involved use of a continuous coarse sediment transport monitor located close to the catchment outlet on the bed of the channel. Reid et al. (in press) describe the monitoring system in full. It is based upon a continuously logging impact sensor that detects the acceleration of a steel plate fixed to a rock channel bed or large boulder, after the plate is struck by a clast (Richardson et al., 2003). The instrument is installed flush with the channel bed or gravel surface to ensure that it only detects true bedload motion. Comparison with the literature (e.g. Drake et al., 1988) suggested that the sensor plate (150 × 130 × 6 mm) is of the correct size to record most rolling particles once, yet large enough to catch those saltating. Bedload is often routed along well defined and narrow pathways within upland rivers (Richardson et al., 2003) and installation location was optimised through analysis of the first autumn floods in September 2002 through identification of clean gravel 'tracks' which identify the most active areas of the bed (Richardson et al., 2003). The impact sensor was attached to a fixed continuous logger (Tinytag Plus TGPR1201 Count Input Data Logger) and the sensor was installed next to the V-notch weir described above. Data were recorded as the number of impacts within the preceding 15 min so that the relationship between rainfall, Topographic Failure Threshold, and sediment movement could be seen (e.g. Fig. 7). The sensor was adjusted to only detect grains greater than 20 mm in diameter, which was taken as the definition of coarse sediment in this study.

4. Results

4.1. Model calibration on mapped, connected failure areas

The model was run using the INSAR DEM at 5 m resolution and an internal friction angle of 30°, with the

Table 1
Values of conditional Kappa and their strength of agreement (from Altman, 1991)

Value of conditional Kappa	Strength of agreement
<0.20	Poor
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Good
0.81–1.00	Very Good

optimised TOPMODEL parameter values described previously, and the requirement that a failed area must also be hydrologically-connected for sediment delivery to occur. For the TFT value of 8.1 (August 21st, 2003), the no-cohesion routine predicted 92 sources totalling 1.2% of the catchment area (25,250 m^2). The with-cohesion routine predicted just 62 sources covering 0.3% (6250 m^2). For the TFT value of 9.3 (November 2nd, 2003), the no-cohesion routine predicted 44 sources accounting for 0.2% of the catchment area (3200 m^2). The with-cohesion routine predicted just 32 sources representing 0.1% of the catchment area (1830 m^2).

The effect of cohesion must first be analysed to determine which routine provides a more reliable estimate of sediment sources. Field observations mapped as many scars within the catchment as possible. However, estimates of failure areas may be conservative since some smaller scars may have been missed. Mapping suggested 26 fresh failures (November 2nd) accounting for 1430 m^2 and 58 larger revegetating scars of 5030 m^2 (August 21st). The preliminary conditional Kappa values (Table 2) suggested that the with-cohesion routine (if root cohesion data are available for a site) provided a more reliable estimate, largely because the no-cohesion routine predicted many more failures than were observed. This is partly because not all sites with equal predicted failure probability would fail in any given storm, reflecting the disadvantage of moving away from the probability-based analysis implicit in the original SHALSTAB approach. The spatial distribution of the estimate of failure likelihoods for the with-cohesion and no-cohesion routines (Fig. 8) shows that in areas of high failure likelihood (lower stability value), the no-cohesion routine classifies a greater area as likely to fail. Thus, our subsequent analysis focuses on the with-cohesion case.

Even with the cohesion routine and the restricted requirement for hydrological connection, there remained a residual over-prediction of failed areas (Table 3). This is commonly the case in models of sediment failure. The reasons for the continued residual over-prediction are one or more of: (i) the model is over-predicting areas acting as connected sources; (ii) not all sources were mapped due to the steep slopes and catchment size; and (iii), following from the point made above, not all sites with equal mapped/predicted failure probability would fail in any given storm. A combination of all three is likely. However, the angle of internal friction (ϕ) will also affect the stability of the slope, with lower values increasing the failure likelihood for a given pixel. Thus, the model was run using values of ϕ between 29° and 35° and the Kappa analysis was used to test the predicted sources against those observed during the 2 storm events (Table 4).

This analysis suggests that an internal friction angle of 32° best calibrates the model through the range of storm events when using the parameter set of $m=0.0007$ and $T_0=1.5$ m^2 h^{-1}. Using this angle of internal friction slightly under-estimates both the number and coverage of failure areas. However, it is also likely that some failures may have been missed during the survey

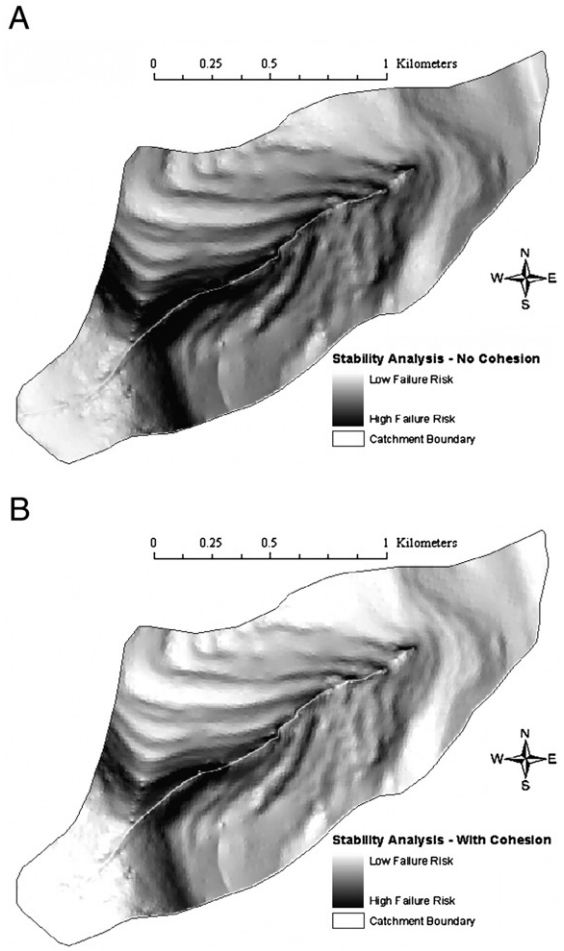

Fig. 8. Likelihood of failure estimate for the Buckden Beck subcatchment for: (A) the no-cohesion routine; and (B) the with-cohesion routine, both using the INSAR DEM at 5 m resolution.

Table 2
Conditional Kappa values for the with- and no-cohesion routines and the two storm events

Event	Value of conditional Kappa	Strength of agreement
August 21st — with-cohesion	0.502	Moderate
August 21st — no-cohesion	0.170	Poor
November 2nd — with-cohesion	0.534	Moderate
November 2nd — no-cohesion	0.319	Fair

Table 3
Comparison of observed and predicted failures for the two validation storm events

Data type	Number of sources	Source area (m^2)
August 21st — observed failures	58	5030
August 21st — predicted failures	62	6250
November 2nd — observed failures	26	1430
November 2nd — predicted failures	32	1830

periods. The with-cohesion routine and an internal friction angle of 32° are used in all further analysis within this paper. Fig. 9 shows the good correspondence between modelled (A) and mapped (B) sediment sources in terms of failure location.

4.2. Model assessment using observations of sediment transport

Data from the impact sensor (Fig. 10) provides an independent assessment of the calibrated model. The largest predicted sediment generation event of the study year occurred on August 21st, 2003 (TFT=8.1). This coincided with the largest sediment transport event observed at the site during the study period (Fig. 10). The event saw 1032 impacts for the 0.13 m width of the sensor within the channel. The lag time from peak rainfall to peak transport was 10 h 55 min. The second largest predicted generation event of the study year occurred on November 2nd, 2003 (TFT=9.3). This coincided with the second largest transport volume during the study (Fig. 10). This event saw 296 impacts and a lag time of 11 h 35 min.

4.3. Comparison of model application without consideration of connection during failure

SHALSTAB was applied to the catchment for comparison with the SEDMAP approach reported above (the network index modification to TOPMODEL is retained — it is only the sediment connectivity that is removed). The INSAR DEM at 5 m resolution was used. This considered both: (i) no-cohesion; and (ii) soil and root cohesion, as the effect of reinforcement of the slope by roots that penetrate the basal failure surface (Montgomery et al., 1998). This was undertaken for both of the storm events for which validation data were available. The category of SHALSTAB output used in the comparison was the log $Q/T \leq -3.1$ category since this has the highest potential for instability. The 'chronic instability' category applies to areas where slope failure is predicted to occur even when the soil is dry, and these areas are therefore interpreted to represent areas such as bedrock outcrops or cliffs and hence these areas are unconditionally unstable. Therefore a comparison is not made between these areas and the mapped failures. Without a connectivity constraint, the no-cohesion routine classified 1,068,000 m^2 in the highest relative likelihood of failure category, accounting for 49.5% of the catchment area. The with-cohesion routine classified 801,000 m^2 areas totalling 37.1%. These areas relate to the places that have the highest relative failure hazard based on topography, since SHALSTAB does not purport to predict sites of failure during individual storm events. Comparison was made between the areas classified as the log $Q/T \leq -3.1$ category and the mapped failures. The comparison suggested (Table 5) that the significant majority of observed failures occurred within the highest relative failure hazard (i.e. Q/T band) of the default version of the model. This was significant, both with and without the cohesion component, although the with-cohesion routine once again provided a better fit to the observed data, for both validation events. DRM

However, only a very small percentage (Table 5) of the highest relative failure category (i.e. log $Q/T \leq -3.1$) predicted by SHALSTAB actually failed during the

Table 4
Agreement between predicted and observed failures for different values of the internal friction angle, using the August 21st and November 2nd events, INSAR DEM at 5 m and the with-cohesion routine

ϕ in degrees	August 21st, 2004 event			November 2nd, 2004 event		
	Conditional Kappa	Strength of agreement	Predicted failures and area	Conditional Kappa	Strength of agreement	Predicted failures and area
29	0.393	Fair	73, 7530 m^2	0.517	Moderate	38, 2180 m^2
30	0.502	Moderate	62, 6250 m^2	0.534	Moderate	32, 1830 m^2
31	0.536	Moderate	61, 5280 m^2	0.588	Moderate	27, 1680 m^2
32	0.611	Good	55, 4850 m^2	0.637	Good	24, 1280 m^2
33	0.549	Moderate	46, 3700 m^2	0.505	Moderate	21, 1100 m^2
34	0.495	Moderate	38, 2650 m^2	0.396	Fair	17, 1050 m^2
35	0.384	Fair	34, 2180 m^2	0.370	Fair	12, 780 m^2

The observed data for August 21st were 58 failures covering 5025 m^2 and for November 2nd, 26 failures covering 1425 m^2.

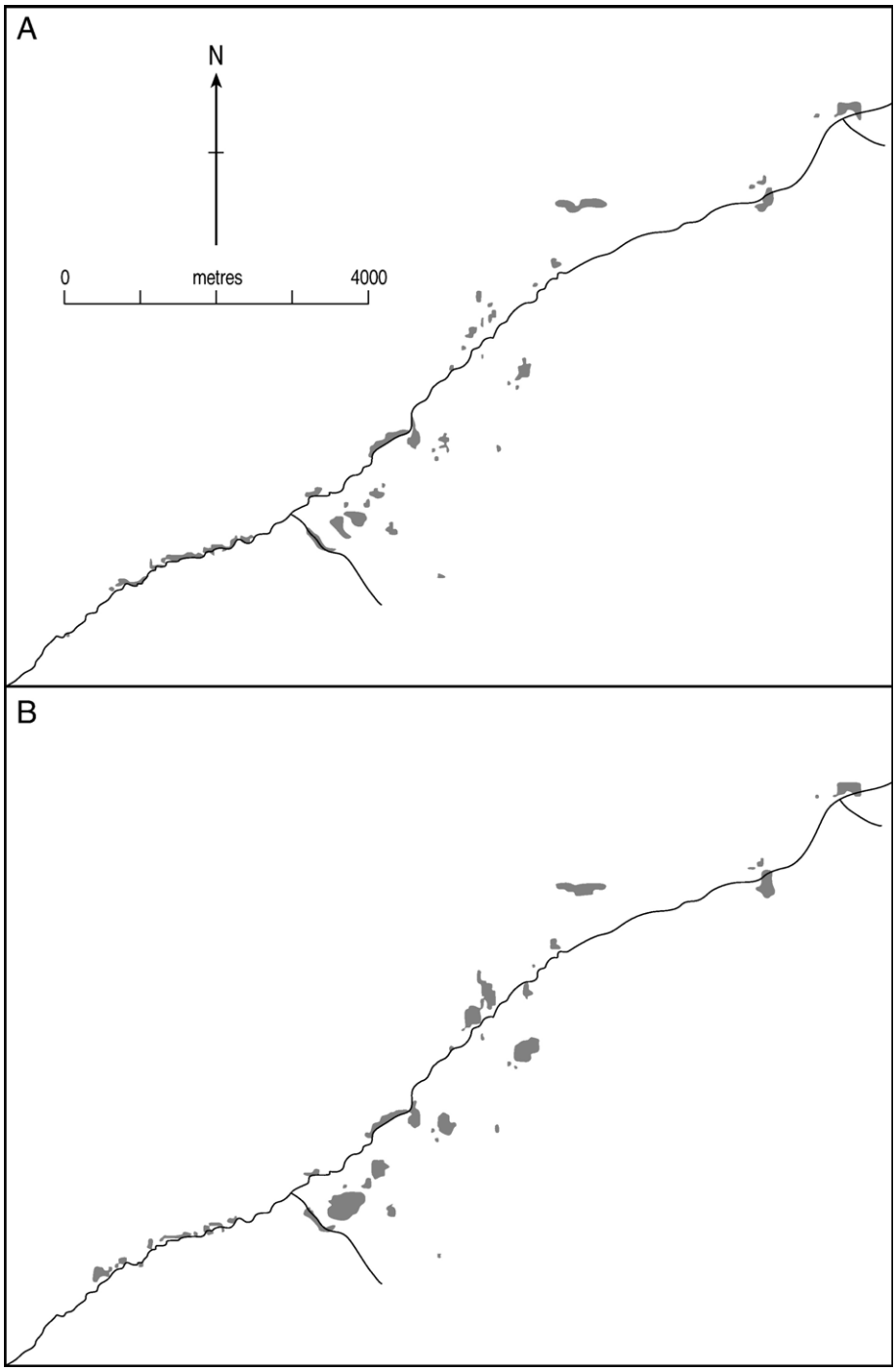

Fig. 9. Areas of connected sediment sources modelled for the August 21st event (A) and observed during field survey (B).

August and November events. Thus, the modification for the effects of hydrological connectivity is important in terms of constraining observed failure areas, whilst extension of the drainage network into the failure zone is important for creating the local conditions that allow for failures to connect to the drainage network.

4.4. Application of the SEDMAP model to a longer time period

The data period (March 19th, 2003 to March 16th, 2004) allowed the model to be run over a longer timescale to see the effect upon source areas of storms of

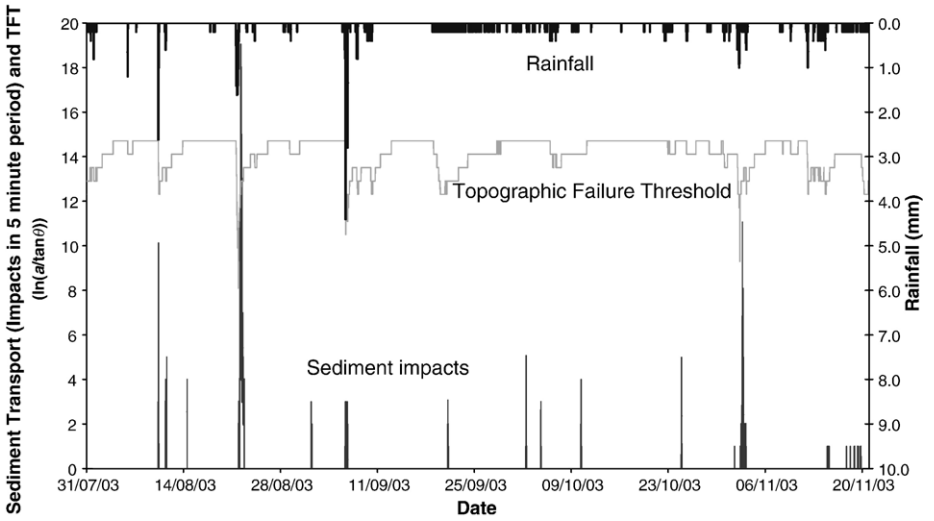

Fig. 10. The relationship between rainfall, Topographic Failure Threshold and sediment transport during autumn 2003.

different magnitudes and durations and with different antecedent conditions. During this period, 73 individual storm events were recorded (Fig. 2) where the value of the TFT permitted coarse sediment generation to occur (TFT < 14.7). An individual storm is classed as where the TFT value increased to or above 14.7 during post rainfall-recession. This TFT value is the highest (i.e. least saturated) at which connected sources are still seen. Application of the model suggested that the total volume of material failed during this period was 28,180 m^3, although this is only an indicative value as a result of assumptions made about failure depth and the assumption that all material failed is coarse sediment (rather than soil or vegetation). The assumption of a 1.0 m soil depth is likely to over-estimate delivery volumes as soils are probably thinner than 1.0 m on the steeper slopes where the majority of failure occurred.

Table 5
The relationship between the two SHALSTAB predicted high failure hazard areas and the observed failure areas for the two validation storm events

Event	% of observed failures found within the log $Q/T \leq -3.1$ category	% of highest relative failure category which failed during this event
August 21st — with-cohesion	98.5	0.62
August 21st — no-cohesion	90.1	0.42
November 2nd — with-cohesion	96.5	0.17
November 2nd — no-cohesion	82.5	0.11

5. Discussion

5.1. Does a basic representation of hydrological connection improve estimates of sediment delivery?

The key finding from the above results is that the very simple representation of hydrological connectivity based upon the network index appears to be effective in constraining modelled failure locations to give a better agreement with observations. This follows from the theoretical analyses and review of Legros (2002) who argued that landslide runout distances may be significantly enhanced in the presence of water due to reduction in the effective coefficient of friction. In this case, the topography within which the failures occur and the moderate size of the events considered means that the extension of runout distances through hydrological connectivity was important as compared with situations where either: (1) the relief is such that short runout distances allow ready connection; or (2) larger events result in larger failure volumes and hence longer runout distances. Reaching the conclusion that the hydrological treatment is important using the mapped data needs caution, as optimisation of conditional Kappa values using angle of internal friction (Table 4) was required to obtain a level of agreement described as 'good' by Altman (1991) and reflected in the agreement between observations and predictions shown in Fig. 9. However, the very large reduction in failed areas to levels found in the measured data (Table 3) and moderate levels of agreement before calibration (Table 2), both suggest that this constraint has beneficial effects upon model

predictions. Independent comparison with the measurements of sediment transport was particularly encouraging, especially for larger events. For smaller events, predictions of sediment delivery were unrelated to records of coarse sediment movement within the channel, and we believe that this reflects the residual storage of material within the tributary. After calibration using the angle of internal friction, a continued difference between predicted and observed failures remained. The model defines connected failures, but the nature of the slope stability component still retains a probabilistic aspect, such that there will still be a difference between model-predicted failures and observed failures. One of the underlying philosophies of SHALSTAB was a refocusing of analysis upon failure probabilities. Adopting the restrictive assumption of hydrological connection as a requirement for sediment delivery allows us to restrict failures to smaller areas of the landscape and these appear to be well-related to the storm events observed during the study period. Continuing the probabilistic interpretation, as the highest failure category was shown (by the calibration, Table 5) to have a 1:200 to 1:1000 chance of failing in a given storm is a key finding (and problem) for landslide hazard management, emphasising that interpretation of the kinds of models reported in is likely to have to remain probabilistic.

5.2. Geomorphological implications

The results suggested that only the largest events (August 21st, 2003 and November 2nd, 2003) have both the storm intensity and duration to: (i) generate coarse sediment from the channel banks, bed or hillslopes and deliver this to the channel network; and (ii) transfer this sediment along the network to the catchment outlet. During the study year, larger storm events occurred less frequently but, when they did, larger areas failed due to greater areas of connected saturation (Fig. 11A). Smaller storm events, which permitted coarse sediment generation (TFT < 14.7), were not much more common than moderately sized events. These results also support other studies which suggest that a small area of a catchment (in this case <1%) may account for a large proportion of sediment production (Kelsey, 1980). The most frequent events were 15 storms with a TFT of 12.3, leading to a total volume of 4500 m^3 during the study period. The second most common events were 13 storms of value 13.5 and 14.1 giving a total volume of 2275 m^3 and 975 m^3 respectively. More significant events were 8 storms of 11.701 (3400 m^3), 7 storms of 11.100 (4025 m^3) and 5 storms at 10.499, where 4625 m^3 was generated. Two sizes of event only occurred once: (i) the largest storm (TFT = 8.1) on August 21st which produced 4850 m^3; and (ii) the second largest (TFT = 9.3) on November 2nd which saw 1275 m^3 of connected failure. Even though larger, less frequent events (e.g. August 21st) can produce more sediment over a yearly timescale than the combined total of many smaller events, over the longer timescale, moderately sized events may produce much more sediment than larger events (with are less frequent), or smaller events (which have much less connected failure). Extension of the model to include events greater than the largest considered here would be valuable to test this hypothesis.

During the moderate events (such as November 2nd), all identified connected sources were within 12 m of the channel network, which suggests that during these flow conditions, the majority of generated sediment was formed through mass failures of the channel banks, or failure initiated by basal cutting as the channel eroded into

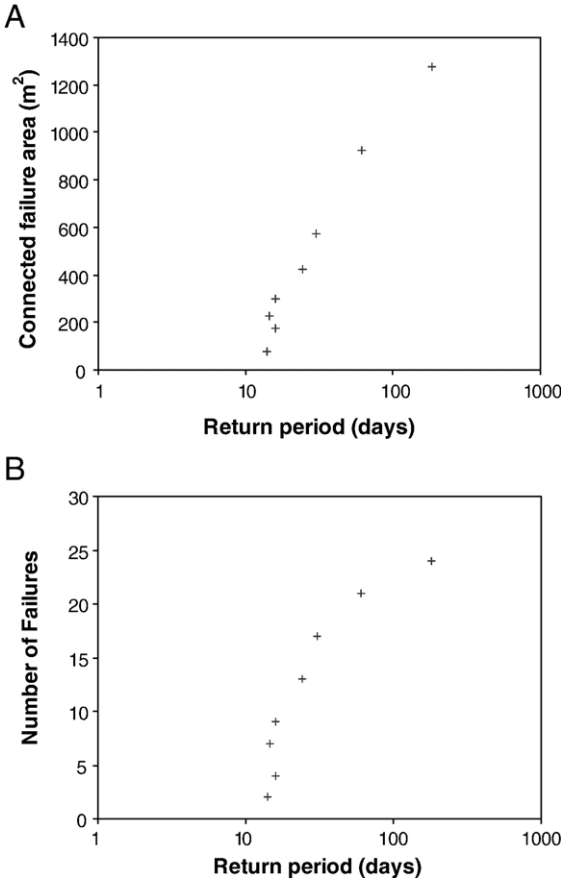

Fig. 11. Plots of: (A) connected failure area; and (B) number of connected source areas, both against storm return period.

a hillslope. The effectiveness of the model in reproducing this may be because it constrains failure to saturated footslopes, which also happen to be the areas of bank undercutting, leading to the 'right results' for the 'wrong reasons'. Regardless of whether or not this is the case, understanding failure for smaller events requires an ability to include small-scale bank failure processes in the model. In turn, this requires topographic data of sufficient precision and resolution and that the models physical basis is correct for this type of behaviour. For the larger event that occurred on August 21st, all connected failures were within 150 m of the main tributary or a tributary branch, and it is clear from Fig. 9A that the larger event, the greater the relative importance of hillslope as opposed to channel adjacent failure sites.

Assuming that process representation and data are sufficient, this evidence suggests that on the longer timescale (where moderate events generate more sediment), the major sources are located within, or very close to, the channel banks. Only during large events do the hillslopes act as connected sources (see Harvey, 2001). A magnitude–frequency analysis of the 73 storm events shows that events where hillslope sources are seen (TFT ≤ 8.1) accounted for 17.2% of sources during the data period. Channel banks accounted for 82.8% of sources over the longer timescale. These values are similar to the work by Keinholz et al. (1991) who found values of 17% for hillslopes and 83% for channel banks. Schmidt (1994) found values of 24% and 76% and Johnson and Warburton (2002) values of 25% and 75% respectively. The Buckden Beck catchment does not have significant channel bed sources. The bed is exposed Great Scar Limestone bedrock and material in the channel is only temporarily stored, with few areas of longer-term accumulation. However, the channel in the headwaters is deeply incised into the Millstone Grit and the majority of modelled bank sources are found here. Finally, although the model has only been run for a 12 month period, it is possible to begin to explore the relationship between event return period and the number and area of failures (Fig. 11). If we ignore the largest recorded event (for which the estimation of return period is likely to be particularly unreliable) and recognising the dependence of the results obtained upon the meteorological characteristics of the study period, it appears that there is a semi-logarithmic relationship between return period and the total area of connected sediment sources (Fig. 11A) and a possibly semi-logarithmic relationship between return period and the number of connected sediment sources (Fig. 11B). The possibility of a scaling relationship between failure area and return period is of particular interest as it is thought to be a general property of shallow landsliding, as noted by Fuyii (1969), Turcotte (1992), Sugai et al. (1994), Hovius et al. (1997; 2000) and Densmore and Hovius (2000).

5.3. Reflections on reduced complexity modelling

The model as presented represents a special case of reduced complexity modelling that in turn has implications for how SEDMAP is subsequently developed and applied. We developed this model in response to an observed river management problem: high rates of coarse sediment delivery to an upland river system were leading to channel instabilities and flood risk problems. In order to move away from instream gravel management, with its associated problems (e.g. the impacts of gravel removal upon instream biota), and towards managing the sediment at source, we needed a model that was capable of capturing the spatial detail of the sediment delivery process. Central to model development was the field observation of extensive sediment scars distributed across the landscape, but much of that sediment was accumulating in topographic hollows and at other similar reductions in slope, such that it was not always actively delivered to the system. We also observed in the field, during a major event in February 2002, that a key criterion in the delivery of large amounts of gravel to the river was the extension of the drainage network (i.e. surface hydrological connection) into zones of sediment failure. Thus, our approach to model development was grounded in a set of largely informal field observations that, by chance, were coincident with a major sediment delivery event. In turn, this caused us to conceptualise the river sedimentation problem as having diffuse causes: a number of distributed locations where both failure and connection resulted in coarse sediment delivery. SEDMAP captures this process using a simplified representation of surface hydrological connection, and a series of associated assumptions (e.g. the end member condition described above). In conceptual terms, the model is clearly incomplete. We know that certain aspects of the physics have been overlooked. Some of this is necessary (or perhaps convenient) as whilst the physics may be known, it is not easily resolved as a result of uncertainties over boundary conditions, geometry and the challenges of numerical solution.

As with other reduced complexity models (e.g. Murray and Paola, 1994), achieving an adequate representation of a geophysical process does not necessarily require complex process representation at small spatial and temporal scales, although it does depend on a sufficient discretisation of space and time.

In our case, the discretisation must capture to a first approximation the topographic controls upon the patterns of surface saturation, important for both failure and connection. In other words, representation of the physics is minimised, the resolution in space and time is maximised, and the focus is upon predicting emergent properties rather than system detail. The need for this kind of approach is reflected in other models of shallow landsliding (e.g. Burton and Bathurst, 1998) and relates to the relatively small spatial scale at which shallow landslides happen coupled to the need to understand the landsliding process over potentially large spatial extents. If we accept that a stronger physical basis is necessary, then this only has meaning if we retain the spatial scale necessary to evaluate correctly the gradient terms that inevitably result in the associated equations and have the capability to determine necessary boundary conditions at that spatial scale. The ways in which we have reduced complexity in SEDMAP mean that it is implicitly contextualised to certain types of shallow landslide events and probably certain types of upland environments. Models are often advocated as embodying a certain level of generality that makes them widely applicable when contextualised by information acquired by particular places. Rarely do we acknowledge the ways in which the models themselves, as we decide how to reduce their complexity, become 'calibrated' on the 'informal' observations we make of the systems that we study. In this case, we take a reduced complexity model and calibrate it upon a particular place in order to improve it. This calibration is not just about forcing empirical adequacy (after Oreskes et al., 1994) through fitting a given model to a set of observations or perturbing model parameters in order to determine model uncertainty, leaving the underlying model structure intact. Rather, it uses wider empirical understanding to reflect upon the complexity in the system being modelled and hence the necessary underpinning perceptual model. In this case, the observation that surface hydrological connection is important provides an additional conceptual constraint that yields significantly better predictions of those locations that are likely to fail in storms. Whilst this appears to have meaning for this location, the reduction in complexity that we have adopted may not necessarily hold for other locations as a result of the implicit contextualisation of the model that limits its spatial relevance.

6. Conclusion

This paper has shown that a highly simplified treatment of hydrological connection can be used to constrain predictions of failure area to give a better level of agreement with both observed, connected failures and records of coarse sediment transport from an impact sensor. This simplified treatment is a restrictive end member of the failure case, especially given observations that failures do not necessarily require point saturation throughout the soil depth for failure to occur. However, this restriction is useful as it allows distributed estimation of sediment generation and delivery and a distributed assessment of failure prediction locations. Application of the model to a full year of data suggested that there was a semi-logarithmic relationship between total failure area and event return period. This merits further consideration using a longer data record. Model results indicate that short duration, high intensity storms do not have the same effect upon sediment generation and delivery as long duration but low intensity storms as sources are limited to hydrologically well-connected parts of the landscape. However, over the longer timescale and due to their relatively high frequency, it is likely to be moderately sized events which generate most sediment. This sediment has been shown to come primarily from channel banks, especially areas of incision into glacial material but during high magnitude events, the hillslopes become effective and significant source areas.

Acknowledgements

This research is supported by NERC Connect Grant NER/D/S/2000/01269 awarded to SNL, Mike J. Kirkby and Adrian T. McDonald, by Environment Agency R and D award E1-108 awarded to Adrian T. McDonald and SNL, and by the National Trust. Simon C. Reid was funded by a University of Leeds Ph.D. studentship.

References

Altman, D.G., 1991. Practical Statistics for Medical Research. Chapman and Hall, London.

Bagnold, R.A., 1954. Experiments on a gravity-free dispersion of large solid spheres in a Newtonian fluid under shear. Proceedings of the Royal Society of London 225, 49–63.

Beven, K.J., 2000. Uniqueness of place and process representations in hydrological modelling. Hydrology and Earth System Sciences 4, 203–213.

Beven, K.J., Kirkby, M.J., 1979. A physically based variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24, 43–69.

Bracken, L.J., Croke, J.M., 2007. The Concept of Hydrological Connectivity and its Contribution to Understanding Runoff Dominated Geomorphic Systems. Hydrological Processes 21. doi:10.1002/hyp.6313.

Brooks, S.M., Richards, K.S., Anderson, M.G., 1993. Approaches to the study of hillslope development due to mass movement. Progress in Physical Geography 17, 32–49.

Brooks, S.M., Crozier, M.J., Glade, T.W., Anderson, M.G., 2004. Towards establishing climatic thresholds for slope instability: use of a physically-based combined soil hydrology-slope stability model. Pure and Applied Geophysics 161, 881–905.

Burton, A., Bathurst, J.C., 1998. Physically based modelling of shallow landslide sediment yield at a catchment scale. Environmental Geology 35, 89–99.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.

Coulthard, T.J., 1999. Modelling upland catchment response to Holocene environmental change. Unpublished Ph.D. dissertation, University of Leeds, U.K.

Coulthard, T.J., Macklin, M.G., 2003. Long-term and large scale high resolution catchment modelling: innovations and challenges arising from the NERC Land Ocean Interaction Study (LOIS). In: Lang, A., Hennrich, K., Dikau, R. (Eds.), Long Term Hillslope and Fluvial System Modelling: Concepts and Case Studies from the Rhine River Catchment. . Lecture Notes in Earth Sciences. Springer, Berlin, pp. 123–136.

Cunjak, R.A., 1988. Behaviour and microhabitat of young atlantic salmon (*Salmo salar*) during winter. Canadian Journal of Fisheries and Aquatic Science 45, 2156–2160.

Densmore, A.L., Hovius, N., 2000. Topographic fingerprints of bedrock landslides. Geology 28, 371–374.

Dietrich, W.E., Reiss, R., Hsu, M.L., Montgomery, D.R., 1995. A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. Hydrological Processes 24, 383–400.

Dietrich, W.E., Bellugi, D., De Asua, R., 2001. Validation of the Shallow Landslide Model, SHALSTAB, for Forest Management. In: Wigmosta, M.S., Burges, S.J. (Eds.), Land Use and Watersheds: Human Influence on Hydrology and Geomorphology in Urban and Forest Areas, Water Science and Application, vol. 2. American Geophysical Union, pp. 195–227.

Drake, T.G., Shreve, R.L., Dietrich, W.E., Whiting, P.J., Leopold, L.B., 1988. Bedload transport of fine gravel observed by motion picture photography. Journal of Fluid Mechanics 192, 193–217.

Dunne, T., Black, R.D., 1970. Partial area contributions to storm runoff in a small New England watershed. Water Resources Research 6, 1296–1311.

Everitt, B.S., 1998. The Cambridge Dictionary of Statistics, second ed. Cambridge University Press, Cambridge. 420 pp.

Fuyii, Y., 1969. Frequency distribution of the magnitude of landslides caused by heavy rainfall. Seismological Society of Japan Journal 22, 244–247.

Gerrard, A.J., 1999. Landsliding in the Likhu Khola drainage basin, Middle Hills of Nepal. Physical Geography 20, 240–255.

Harvey, A.M., 1991. The influence of sediment supply on the channel morphology of upland streams: the Howgill Fells, North–West England. Earth Surface Processes and Landforms 16, 675–684.

Harvey, A.M., 2001. Coupling between hillslopes and channels in upland fluvial systems: implications for landscape sensitivity, illustrated from the Howgill Fells, northwest England. Catena 42, 225–250.

Harvey, A.M., 2002. Effective timescales of coupling within fluvial systems. Geomorphology 44, 175–201.

Hearn, G.J., Griffiths, J.S., 2001. Landslide hazard mapping and risk assessment. In: Griffiths, J.S. (Ed.), Land Surface Evaluation for Engineering Practice. Geological Society Engineering Geology Special Publication, vol. 18, pp. 43–52.

Heggenes, J., 1996. Habitat selection by brown trout (*Salmo trutta*) and young atlantic salmon (*S. salar*) in streams: static and dynamic hydraulic modeling. Regulated Rivers: Research and Management 12, 155–169.

Heritage, G.L., Newson, M.D., 1997. Geomorphological Audit of the Upper Wharfe. University of Newcastle upon Tyne.

Hey, R.D., Winterbottom, A.N., 1990. River engineering in national parks: the case of the River Wharfe, U.K. Regulated Rivers: Research and Management 5, 35–44.

Horritt, M.S., Bates, P.D., 2001. Effects of spatial resolution on a raster based model of flood flow. Journal of Hydrology 253, 239–249.

Horritt, M.S., Bates, P.D., 2002. Evaluation of 1D and 2D numerical models for predicting river flood inundation. Journal of Hydrology 268, 87–99.

Hovius, N., Stark, C.P., Allen, P.A., 1997. Sediment flux from a mountain belt derived by landslide mapping. Geology 25, 231–234.

Hovius, N., Stark, C.P., Hao-Tsu, C., Jiun-Chuan, L., 2000. Supply and removal of sediment in a landslide-dominated mountain belt: Central Range, Taiwan. Journal of Geology 108, 73–89.

James, A., 1999. Time and the persistence of alluvium: river engineering, fluvial geomorphology, and mining sediment in California. Geomorphology 31, 265–290.

James, T.D., 2004. Error identification and correction methods for automatically derived digital elevation models. Unpublished Ph.D. dissertation, University of Leeds, U.K.

Johnson, K.A., Sitar, N., 1990. Hydrologic conditions leading to debris flow initiation. Canadian Geotechnical Journal 27, 789–801.

Johnson, R.M., Warburton, J., 2002. Annual sediment budget of a UK mountain torrent. Geografiska Annaler 84, 73–88.

Keinholz, H., Lehmann, C., Guggisberg, C., Loat, R., Hegg, C., 1991. Bedload transport in Swiss Mountain torrents with respect to the disaster in 1987. Zeitschrift für Geomorphologie 983, 53–62.

Kelsey, H.M., 1980. A sediment budget and an analysis of geomorphic processes in the Van Duzen River Basin, North Coastal California 1941–1975: summary. Geologic Society of America Bulletin 91, 190–195.

Lane, S.N., Brookes, C.J., Kirkby, M.J., Holden, J., 2004. A network index based version of TOPMODEL for use with high resolution digital topographic data. Hydrological Processes 18, 191–201.

Legros, F., 2002. The mobility of long-runout landslides. Engineering Geology 63, 301–331.

Merrett, S.P., Macklin, M.G., 1999. Historic river response to extreme flooding in the Yorkshire Dales, Northern England. In: Brown, A.G., Quine, T.A. (Eds.), Fluvial Processes and Environmental Change. Wiley, Chichester, pp. 345–360.

Montgomery, D.R., Dietrich, W.E., 1994. A physically based model for the topographic control on shallow landsliding. Water Resources Research 30, 1153–1171.

Montgomery, D.R., Dietrich, W.E., 1995. Hydrologic processes in a low-gradient source area. Water Resources Research 31, 1–10.

Montgomery, D.R., Dietrich, W.E., Torres, R., Anderson, S.P., Heffner, J.T., Loague, K., 1997. Hydrologic response of a steep unchanneled valley to natural and applied rainfall. Water Resources Research 33, 91–109.

Montgomery, D.R., Sullivan, K., Greenberg, H., 1998. Regional test of a model for shallow landsliding. Hydrological Processes 12, 943–955.

Montgomery, D.R., Dietrich, W.E., Heffner, J.T., 2002. Piezometric response in shallow bedrock at CB1: implications for runoff generation and landsliding. Water Resources Research 38, 1274–1292.

Moore, I.D., Gallant, J.C., Guerra, L., Kalma, J.D., 1993. Modelling the spatial variability of hydrological processes using Geographical Information Systems. In: Kovar, K., Nachtnebel, H.P. (Eds.),

HydroGIS 93: Application of Geographical Information Systems in Hydrology and Water Resources. IAHS Publication, vol. 211, pp. 161–169.

Murray, A.B., Paola, C., 1994. A cellular model of braided rivers. Nature 371, 54–57.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models, part 1, a discussion of principles. Journal of Hydrology 10, 282–290.

O'Loughlin, E.M., 1981. Saturation regions in catchments and their relation to soil and topographic properties. Journal of Hydrology 53, 229–246.

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

Reid, S.C., Lane, S.N., Berney, J.M., Holden, J., in press. The timing and magnitude of coarse sediment transport events within an upland gravel-bed river. Geomorphology.

Rice, S.P., Greenwood, M.T., Joyce, C.B., 2001. Macroinvertebrate community changes at coarse sediment recruitment points along two gravel bed rivers. Water Resources Research 37, 2793–2803.

Richardson, K., Benson, I., Carling, P.A., 2003. An instrument to record sediment movement in bedrock channels. IAHS Publication 283, 228–236.

Rimmer, D.M., Paim, U., Saunders, R.L., 1983. Changes in the selection of microhabitat by juvenile atlantic salmon (*Salmo salar*) at the summer–autumn transition in a small river. Canadian Journal of Fisheries and Aquatic Science 41, 469–475.

Schmidt, K.H., 1994. River channel adjustment and sediment budget in response to a catastrophic flood event (Lainbach catchment, Southern Bavaria). In: Ergenzinger, P., Schmidt, K.H. (Eds.), Dynamics and Geomorphology of Mountain Rivers. Springer-Verlag, London, pp. 109–127.

Sidle, R.C., 1991. A conceptual model of changes in root cohesion in response to vegetation management. Journal of Environmental Quality 20, 43–52.

Stover, S.C., Montgomery, D.R., 2001. Channel change and flooding, Skokomish River, Washington. Journal of Hydrology 243, 272–286.

Sugai, T., Ohmori, H., Hirano, M., 1994. Rock control on the magnitude–frequency distribution of landslides. Transactions of the Japanese Geomorphological Union 15, 233–251.

Tarboton, D.G., 1997. A new method for the determination of flow directions and contributing areas in grid digital elevation models. Water Resources Research 33, 309–319.

Turcotte, D.L., 1992. Fractals and Chaos in Geology and Geophysics. Cambridge University Press, Cambridge. 412 pp.

Wainwright, J., Calvo Cases, A., Puigdefábregas, J., Michaelides, K., 2002. Editorial: linking sediment delivery from hillslope to catchment scales. Earth Surface Processes and Landforms 27, 1363–1364.

White, R., 2002. The Yorkshire Dales: A Landscape through Time, 2nd edition. Great Northern Books, Ilkley. 128 pp.

Wilson, C.J., Dietrich, W.E., 1987. The contribution of bedrock groundwater flow to storm runoff and high pore pressure development in hollows. In: Beschta, R.L., Blinn, T., Grant, G.E., Ice, G.G., Swanson, F.J. (Eds.), Erosion and Sedimentation in the Pacific Rim. International Association of. Hydrological Sciences Publication, vol. 165, pp. 49–60.

Wolock, D.M., 1993. Simulating the variable-source-area concept of streamflow generation with the watershed model TOPMODEL. USGS Water Resources Investigations Report, pp. 93–4124.

Wu, W., Sidle, R.C., 1995. A distributed slope stability model for steep forested hillslopes. Water Resources Research 31, 2097–2110.

Yu, D., Lane, S.N., 2006. Urban fluvial flood modelling using a two-dimensional diffusion wave treatment: 1. mesh resolution effects. Hydrological Processes 20, 1541–1565.