

# Blanket-peat restoration: numerical study of the underlying processes delivering Natural Flood Management benefits

Salim Goudarzi<sup>1</sup>, David G. Milledge<sup>1</sup>, Joseph Holden<sup>2</sup>, Martin Evans<sup>3</sup>, Tim Allott<sup>4</sup>, and Emma Shuttleworth<sup>3</sup>

<sup>1</sup>Newcastle University

<sup>2</sup>University of Leeds

<sup>3</sup>University of Manchester

<sup>4</sup>School of Environment

November 24, 2022

## Abstract

Headwater blanket-peat restoration activities, in particular revegetation and gully-blocking, are observed to deliver significant Natural Flood Management (NFM) benefits. A recent Before-After-Control-Intervention (BACI) experiment showed that these interventions reduce flood-peaks and increase lag-times, but the processes controlling these effects remain unclear. We seek to identify these processes at the same BACI sites by inverting the TOPMODEL rainfall-runoff model and linking the response-to-intervention in each catchment to model parameters through rigorous calibration. Through numerical experiments, we infer processes most likely to be driving the BACI observations. Our findings confirm the NFM benefits of these restoration-focused interventions. Independent of storm size/intervention, the increased lag is almost entirely due to surface roughness reducing the floodwave speed. We conceptualise this as a ‘mobile’ surface storage. In flood-relevant storms, at least 90% of the peak reduction in both interventions is delivered by mobile storage. The additional increase in the mobile storage due to gully-blocking is very significant and comparable to that of revegetation alone. The impact of interventions on ‘immobile’ storage (interception+ponding+evapotranspiration) becomes important for smaller storms, in which revegetation reduces peak discharge by increasing evapotranspiration but the not interception storage. Gully blocking however, increases ponding but reduces evaporation, such that there is no net gain in catchment immobile storage relative to revegetation alone. Although interventions always increase lag-times, they can be less effective in reducing peak magnitude in long duration frontal rainfalls. We propose two approaches to further increase catchment’s surface storage, while adhering to the restoration requirement to keep the water tables high.

# Blanket-peat restoration: numerical study of the underlying processes delivering Natural Flood Management benefits

Salim Goudarzi<sup>1</sup>, David G. Milledge<sup>1</sup>, Joseph Holden<sup>2</sup>, Martin G. Evans<sup>3</sup>, Timothy E. H. Allott<sup>3</sup>, Emma L. Shuttleworth<sup>3</sup>, Michael Pilkington<sup>4</sup>, and Jonathan Walker<sup>5</sup>

<sup>1</sup>Department of Civil Engineering, Newcastle University, NE1 7RU, UK

<sup>2</sup>Water@leeds, School of Geography, University of Leeds, LS2 9JT, UK

<sup>3</sup>Department of Geography, University of Manchester, M13 9PL, UK

<sup>4</sup>Moors for the Future Partnership, Fieldhead, S33 7ZA, UK

<sup>5</sup>Department of Biosciences, Swansea University, Swansea, SA2 8PU, UK

Submitted: 09/11/2020

## Abstract

Headwater blanket-peat restoration activities, in particular revegetation and gully-blocking, are observed to deliver significant Natural Flood Management (NFM) benefits. A recent Before-After-Control-Intervention (BACI) experiment showed that these interventions reduce flood-peaks and increase lag-times, but the processes controlling these effects remain unclear. We seek to identify these processes at the same BACI sites by inverting the TOPMODEL rainfall-runoff model and linking the response-to-intervention in each catchment to model parameters through rigorous calibration. Through numerical experiments, we infer processes most likely to be driving the BACI observations. Our findings confirm the NFM benefits of these restoration-focused interventions. Independent of storm size/intervention, the increased lag is almost entirely due to surface roughness reducing the floodwave speed. We conceptualise this as a 'mobile' surface storage. In flood-relevant storms, at least 90% of the peak reduction in both interventions is delivered by mobile storage. The additional increase in the mobile storage due to gully-blocking is very significant and comparable to that of revegetation alone. The impact of interventions on 'immobile' storage (interception+ponding+evapotranspiration) becomes important for smaller storms, in which revegetation reduces peak discharge by increasing evapotranspiration but the not interception storage. Gully blocking however, increases ponding but reduces evaporation, such that there is no net gain in catchment immobile storage relative to revegetation alone. Although interventions always increase lag-times, they can be less effective in reducing peak magnitude in long duration frontal rainfalls. We propose two approaches to further increase catchment's surface storage, while adhering to the restoration requirement to keep the water tables high.

## 1 Introduction

Peatlands cover only 2.84% of the land surface of the planet, yet they provide the human population with a wide range of services, including more than a third of soil carbon storage, river flow regulation, and biodiversity (Xu et al., 2018). These sensitive systems require high precipitation or impeded drainage to develop and remain stable. Industrialisation, mechanisation, land-use change (e.g., agriculture or forestry) and extraction of peat for horticultural and energy production have all led to degradation of peatlands.

Blanket peat, the dominant UK peat type, is unusual among peatlands in that it occurs on slopes up to 15° (Lindsay et al., 1988). When surface vegetation is damaged, their sloping nature means that erosion can rapidly occur (Evans et al., 2005, 2006). Rapid and extensive peatland degradation has been going on in the UK uplands for at least a century due to increased atmospheric pollution, e.g., Rothwell et al. (2007), drainage,

40 grazing and wildfire damage, e.g., [Evans & Warburton \(2011\)](#). Peatland degradation not only poses a severe  
41 threat to an important ecosystem, but in some circumstances may also enhance downstream flood risk ([Acreman  
42 & Holden, 2013](#)). For peat to develop and persist it must be largely saturated for much of the year, thus even  
43 intact peatlands are often able to store little additional water during rainfall events. Blanket peatlands typically  
44 exhibit a flashy response to rainfall, dominated by saturation-excess overland flow and/or near surface through-  
45 flow with minimal base-flow contribution ([Evans et al., 1999](#)). Peatland degradation through, for instance,  
46 vegetation loss and gully erosion can further increase the flashiness of the runoff response and increase flood  
47 risk downstream by raising peak discharge and reducing lag-times ([Grayson & Rose, 2010](#)).

48 In the UK, a growing recognition of the value of peatland ecosystem services has led to a proliferation of  
49 programmes to prevent further degradation and ultimately to restore their natural functions. In parallel, UK flood  
50 risk management policy and practice is increasingly interested in the potential for Natural Flood Management  
51 (NFM) - interventions to “reduce flood hazard, while also sustaining or enhancing other potentially significant  
52 co-benefits including enhanced eco-system services” ([Dadson et al., 2017](#)). Given that many river systems  
53 worldwide originate in upland catchments with a blanket peat cover, peatland restoration activities have the  
54 potential to offer significant NFM benefit. These activities often involve: drain blocking, gully blocking and  
55 re-profiling, bare peat stabilisation and revegetation ([Parry et al., 2014](#)). In the UK alone, over the past two  
56 decades, extensive peatland restoration programmes have been implemented ([Evans et al., 2005](#), [Wallage et  
57 al., 2006](#), [Armstrong et al., 2010](#), [Parry et al., 2014](#)). Although the impact of artificial drainage on peatland  
58 hydrology has been studied extensively ,e.g., [Holden et al. \(2004\)](#), studies of the effect of restoration efforts  
59 have primarily focused on sediment dynamics, e.g., [Shuttleworth et al. \(2015\)](#), carbon release, e.g., [Dixon et  
60 al. \(2014\)](#), [Holden \(2005\)](#), vegetation recovery, e.g., [Cole et al. \(2014\)](#). In fact there is very little experimental  
61 evidence available on the effects of restoration activities on hydrological functions of peatlands.

62 Most recently, in the first controlled catchment-scale experiment on the impact of revegetation and gully-  
63 blocking on hydrological functions, [Shuttleworth et al. \(2019\)](#) published the results of their Before After Control  
64 Intervention (BACI) study, showing that: lag-times increased by 106% and magnitude of flood peaks decreased  
65 by 27% after revegetation, and that lag-times increased by 200% and magnitude of flood peaks decreased by  
66 51% after a combined treatment of revegetation and gully blocking. Such BACI-type experimental data enable  
67 a change in discharge characteristics to be linked to an intervention in a particular catchment, while eliminating  
68 the noise due to ‘natural temporal variability’ resulting from inter-annual variation in synoptic hydrometeorol-  
69 ogy. These observations can later be assigned to a number of factors that are conceivable from a physical point  
70 of view (see [Shuttleworth et al. \(2019\)](#)).

71 However, solely based on variations in measured rainfall-discharge relationship, it is difficult to identify  
72 which hydrological processes are driving the observed changes. For example, it is not obvious whether the  
73 additional interception storage volume created by revegetating the bare peat areas can be sufficient to account  
74 for some of the observed peak magnitude reduction, and whether the time it takes to fill such additional storage  
75 accounts for some of the observed peak timing delay; similarly for the additional in-channel, pond-storage  
76 behind the blocks. It is also not clear how important is the reduction in the response-speed (celerity) of overland  
77 flow, that results from flow on a surface made rougher by more vegetation cover and/or stream blocking. In  
78 relation to evapotranspiration, it is not known how important a role is played by increased evapotranspiration  
79 rates resulting from a larger vegetated catchment area, or additional direct pond evaporation in the gully-blocked  
80 case. More importantly, perhaps, how do various processes interact (or counteract)? In reality all of the above  
81 processes, at least to some degree contribute to the observed changes to the hydrograph behaviour following  
82 interventions. Thus, studying both the ‘absolute’ and the ‘relative’ (to one another) magnitude of impact of  
83 these processes helps better understand the underlying mechanism(s) through which each intervention primarily  
84 delivers NFM benefit (i.e., reduce peak magnitudes and increase lag-times). This is desirable because it would:  
85 (1) condition our expectations for future interventions; (2) guide future decision making and implementation.

86 To this end, we invert the TOPMODEL rainfall-runoff model to relate the BACI observations to more spe-  
87 cific hydrological processes by attempting to link the response-to-intervention in each catchment to numerical  
88 model parameters through rigorous calibration. We then use the calibrated parameter-sets to perform a series  
89 of numerical experiments to see, through the lens of a numerical model, which underlying process(es) may  
90 be driving the observed changes. In doing so, we estimate and analyse the ‘relative’ and ‘absolute’ magni-  
91 tude of impact of each process. Finally, we discuss the implications of our findings regarding Natural Flood

92 Management at catchment scale.

## 93 2 Study sites

94 The BACI experiment of [Shuttleworth et al. \(2019\)](#) used rainfall and discharge data over a 5-year period for  
95 three micro catchments located on the Kinder Scout Plateau in the Southern Pennines, UK (see Fig. 1).

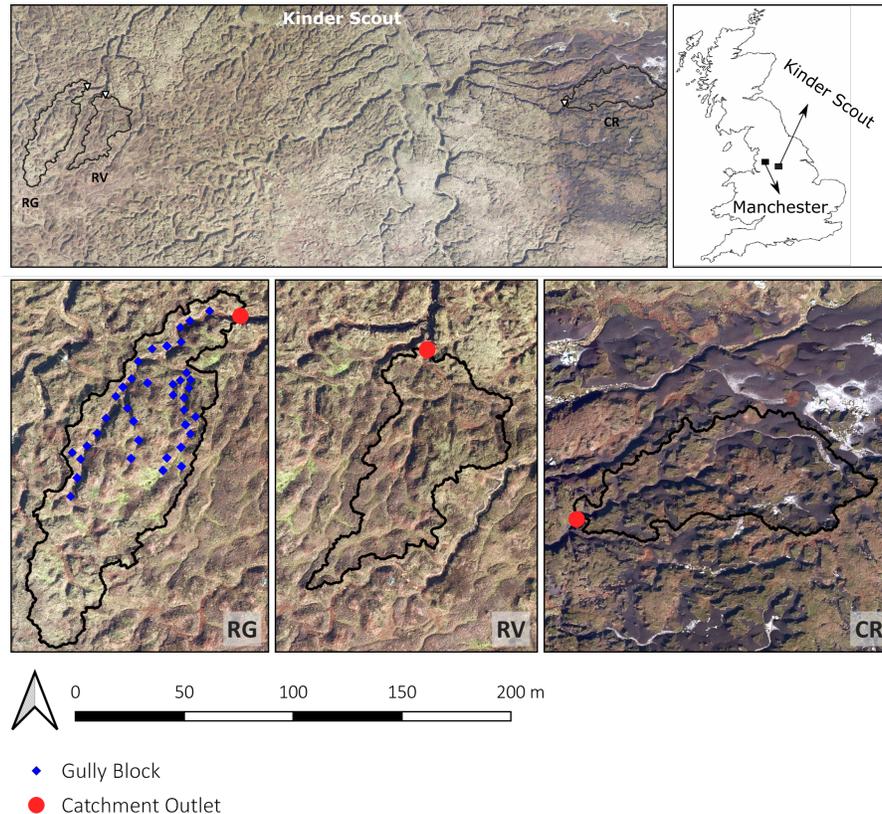


Figure 1: Study catchment outlines overlain on satellite images (obtained in 2020) to show: bare peat cover at the control site (CR), extensive revegetation at the intervention sites and the locations of gully blocks at the re-vegetated and blocked site (RG). Dark brown areas are bare peat, white areas are mineral bedrock exposed by total removal of the blanket peat

96 Kinder Scout is one of the most severely eroded areas of blanket peat in the UK, characterised (before  
97 intervention) by extensive areas of bare peat flats and networks of erosion gullies (Fig. 1). The three micro  
98 catchments were selected to be as close to each other as possible and have comparable geometries, erosion  
99 and gully characteristics prior to interventions (Table 1). Intensive monitoring began in June 2010 with a  
100 pre-intervention period of 15 months, followed immediately by more than 32-months of post-intervention mon-  
101 itoring. One catchment was retained as a control (CR) while interventions were implemented in the other two:  
102 revegetation alone (RV) and revegetation with gully blocking (RG). Between 2010 and 2014, revegetation led to  
103 76% reduction in bare peat cover, with no significant change at the control site. Gully blocking at RG involved:  
104 17 0.5 m high stone dams composed of loose piles of cobbles (75–200 mm diameter), across the width of the  
105 gully, spaced c. 6–7 m apart; and 20 timber dams in smaller tributary gullies of similar height but with a 38 mm  
106 deep ‘V notch’ cut into the top board. Timber dams generally retain a pond with its surface at, or close to the  
107 base of their v-notch. At the time of installation stone dams were permeable, but became progressively clogged

Table 1: BACI catchment characteristics (Shuttleworth et al., 2019). Values for gullied area and bare peat in non-gullied area, are in percentages of the total area.

Attribute (unit)	CR	RV	RG
Total area ( $m^2$ )	7008	4468	7096
Gullied area (%)	32.9	22.9	28.5
Bare peat in non-gullied area (%)	55	52	48
Max elevation ( $m$ )	618	617	619
Min elevation ( $m$ )	612	611	611
Mean hill-slope (degrees)	6.6	6.2	6.5

with eroded peat deposits. Thus, the permeability of the cobble stone dams over our study period is not clear, but we will discuss the implications of this uncertainty in the discussion section.

We chose a six-week study period starting from 22 August 2010 as the before-intervention period, and the same period in 2012 as the after-intervention period. These periods hereafter are referred to as ‘before’ and ‘after’ periods. Choosing the same period for each year reduces the possibility that seasonal variations play a role in the hydrograph behaviour. Other factors that informed our choice of study periods are: (1) large storms, with return periods within the largest 10% in the past 10 years, were present in both periods; (2) temperatures were always above zero; and (3) runoff coefficients for storms in the study period were never greater than 1 (a hydrograph mass balance check to identify gross errors in rainfall or discharge measurement).

### 3 Numerical Model

TOPMODEL (Beven & Kirkby, 1979) and its variants have been widely applied in rainfall runoff modelling of peat catchments in many parts of the world, e.g., Lane & Milledge (2013), Stocker et al. (2014), Gao et al. (2015, 2016, 2017). The full set of principles and mathematics underlying TOPMODEL have been discussed in detail by many authors including Beven & Kirkby (1979), Keeland et al. (1997) and Kirkby (1997), and will not be explained here. However, the core assumptions of the model are (Kirkby, 1997): (1) phreatic surface is always parallel to the surface, thus the hydraulic gradient in the saturated zone can be approximated using the local topographic slope ( $\tan\beta$ ); (2) transmissivity decays exponentially with water table depth (or local storage deficit); and (3) subsurface flow can be modelled as a series of steady-states. TOPMODEL’s assumptions fit well to the case of blanket peat catchments of interest, where runoff is dominated by surface or near surface flow and the rate of flow declines rapidly with depth in the top few centimetres of soil profile (Holden & Burt, 2002).

In its classic form, TOPMODEL does not distinguish between the connected and disconnected saturated areas, meaning that whether or not a saturated patch is connected to the drainage network, its overland flow is routed to the outlet. Lane et al. (2004) showed that this can lead to over-estimation of runoff generation and proposed a Network Index version of TOPMODEL that only routes the connected portions of saturated areas as overland flow. Later, Lane & Milledge (2013) coupled this Network Index version of TOPMODEL with spatially distributed unit hydrograph (SDUH) routing to account for the timing of delivery of water from different parts of the catchment to the outlet. They showed that even this simple spatially and temporally averaged treatment of overland flow was capable of considerably improving model performance for upland peat catchments. Therefore in this paper we use the Lane & Milledge (2013) model, which improves on the original TOPMODEL while keeping its major advantage of being very CPU-efficient for calibration purposes.

This version of TOPMODEL has nine uncertain parameters (Table 2). We set the parameter ranges based on a combination of literature review and previous experience. We chose ranges that were as wide as possible while remaining physically meaningful.

Overland flow celerity ( $c$ ) represents the speed with which rainfall-induced perturbations in the surface water travel overland and downstream. It is related to, but different from velocity which is the speed of travel of water particles as measured by tracers. Although both terms have often been used interchangeably, Beven

Table 2: Uncertain model parameters and their range.

parameter (unit)	range
$c$ (m/s), overland flow celerity	$1e^{-3} : 5e^{-1}$
$E_p$ (m/day), potential evapotranspiration rate	$1e^{-6} : 5e^{-2}$
$S_{rz_{max}}$ (m), maximum root-zone storage	$5e^{-3} : 2e^{-2}$
$m$ (m), exponential decay parameter	$5e^{-5} : 5e^{-3}$
$T_0$ (m <sup>2</sup> /s), maximum transmissivity (at saturation)	$1e^{-9} : 2e^{-5}$
$t_d$ (hr/m), unsaturated-zone time delay	$1e^0 : 3e^2$
$Q_0$ (m <sup>3</sup> /s), initial subsurface flux	$1e^{-8} : 1e^{-3}$
$S_{rz_0}$ (m), initial root-zone storage	$1e^{-8} : S_{rz_{max}}$
$n_c$ (-), number of topographic index classes (number of HRUs)	$5e^0 : 1e^2$

(2020) demonstrates their intrinsic differences and shows that celerity is the relevant property when hydrograph response is of interest (e.g., rainfall-runoff studies); he also shows that celerity in downstream direction is always  $\geq$  velocity. We will treat  $c$  as a spatially and temporally averaged celerity over the entire catchment area.

Note that the study catchments are small, such that the longest travel distance for water flowing overland to the outlet in all three catchments is 293 m. The observational rainfall-discharge time series was recorded at 10-min intervals, which dictates the time-step in the model. Thus, the combination of a maximum travel distance and a fixed time-step, sets an upper bound on  $c$  of 0.5 m/s. Average celerities exceeding 0.5 m/s result in all overland flow response reaching the outlet within one time step, thus the model becomes insensitive to values of  $c$  greater than 0.5 m/s. We do not expect this to pose any limitation on our study because: (1) surface flows will be shallow for the majority of the flowpaths; (2) measured velocities for such flows in peat catchments are less than 0.1 m/s (Holden et al., 2008); and (3) a 1-D diffusion (e.g. Manning) based estimate suggests that celerity is less than twice the velocity (McDonnell & Beven., 2014).

Potential evapotranspiration is defined as the maximum possible evaporation and transpiration when the catchment is fully saturated and root zone storage is full. We follow Metcalfe et al. (2015) and account for annual and diurnal variation by assuming a linear relationship between insolation and potential evapotranspiration, and sinusoidal variation in insolation both over a year and between sunrise and sunset. We take this approach given limited available data because it captures expected temporal dynamics in potential evapotranspiration at a cost of only one additional parameter ( $E_p$ , representing the annual average potential evapotranspiration rate in units of metres per day) and has been found to provide comparable performance to more complex formulations for a number of UK sites (Calder et al., 1983).

We found that modelled discharge became insensitive to number of the HRUs (or topographic index classes),  $n_c$ ,  $>20$ , therefore this value was used in all simulations. Further, we eliminated the initial subsurface flux ( $Q_0$ ) and the initial root zone storage ( $S_{rz_0}$ ) parameters by using a ‘spin-up’ period, where the model was run using measured rainfall data for a period of 7-days prior to the study period (i.e., 22 August 2010/2012). The spin-up duration was chosen to include a large storm so that the subsurface water conditions, in particular the initial root zone store and the initial subsurface flux, were balanced. The six remaining uncertain parameters were: exponential transmissivity decay parameter  $m$ , transmissivity at saturation  $T_0$ , overland flow celerity  $c$ , annual average potential evaporation rate  $E_p$ , unsaturated zone time delay,  $t_d$ , and maximum root zone storage,  $S_{rz_{max}}$ .

## 4 Framework

### 4.1 Model calibration using GLUE

Hydrological models in general, and TOPMODEL in particular, are well known to experience ‘equifinality’, where many different parameter-sets result in comparably good fits to observational data. Thus, we focus on parameter distributions and their changes following revegetation and gully-blocking rather than seeking a single

179 optimum parameter set that characterises each catchment under study. We apply a Generalised Likelihood Un-  
 180 certainty Estimation (GLUE) methodology (Beven & Binley, 1979), using uniform prior distributions for each  
 181 uncertain parameter, and running 160,000 TOPMODEL simulations with parameter-sets sampled randomly  
 182 from these distributions. We retain only 'behavioural' parameter-sets, for which modelled discharge is feasibly  
 183 within error (limits of acceptability) of the observed discharge for 99 % of the observation period.

184 Since the observed discharge record remains our best estimate of the true discharge (despite its associated  
 185 uncertainties) we then weight 'behavioural' sets based on their agreement with the observed record using three  
 186 metrics: Nash-Sutcliffe Efficiency (NSE), Root Mean Squared Error in magnitude of the 10 most distinct peaks  
 187 (RMQ) and Root Mean Squared Error in timing for the same peaks (RMT). By examining and comparing the  
 188 'weighted' distributions of each parameter in the 'before' and 'after' periods, we aim to identify the effects  
 189 of interventions, and use this information to infer catchment functions through numerical experimentation. To  
 190 calculate the likelihood weights assigned to each behavioural parameter-set, we evaluate a linear weight for  
 191 each metric and multiply them to obtain an overall weight. Thus, for each metric  $M$ , the normalised weight is  
 192 given by:

$$w_i^M = \frac{|M_i - M_0|}{\sum_{i=1}^{N_b} |M_i - M_0|} \quad ; \quad i = 1 : N_b \quad ; \quad M \in \{NSE, RMT, RMQ\} \quad (1)$$

193 where  $N_b$  is the number of behavioural parameter-sets and  $M_0$  is the value of the metric for the worst  
 194 performing set. The worst performing set is given the lowest weight and the best set the highest weight, while  
 195 sum of all weights equal to one. The overall likelihood weight, which takes into account the likelihoods from  
 196 each metric, is calculated by multiplying weights for all metrics and normalising by the sum of values:

$$W_i = \frac{\prod w_i^M}{\sum_{i=1}^{N_b} (\prod w_i^M)} \quad ; \quad M \in \{NSE, RMT, RMQ\} \quad (2)$$

197 To establish how large a sample size should be to ensure sufficiently dense sampling of the six-dimensional  
 198 parameter space, we: 1) generate three independent random sets (S1, S2, S3) each containing 5,000 randomly  
 199 sampled parameter sets; 2) run the model for each of S1, S2 and S3 and in each case calculate the weighted mean  
 200 of each parameter (Eq. 2); and 3) iteratively double the sampling size in each random-set until the weighted  
 201 means of the parameter values in all three random-sets converge. Convergence occurs at or before a sample  
 202 size of 160,000 for all parameters, study periods, and micro-catchments (CR, shown in Figure 2). In addition,  
 203 weighted mean parameter values remain stable even as the sample size is doubled from 80,000 to 160,000. Both  
 204 observations suggest that 160,000 samples provide sufficient sampling in our case.

205 We also performed a blind validation test, where the calibrated parameter-sets were used to predict runoff  
 206 responses for a period (5/10-20/10/2010 and 2012) adjacent but not overlapping the calibration period (22/8-  
 207 4/10/2010 and 2012). We found that none of the distributions of the objective functions were significantly  
 208 different with 95% confidence, providing more confidence that our behavioural sets are representative of the  
 209 sites being studied.

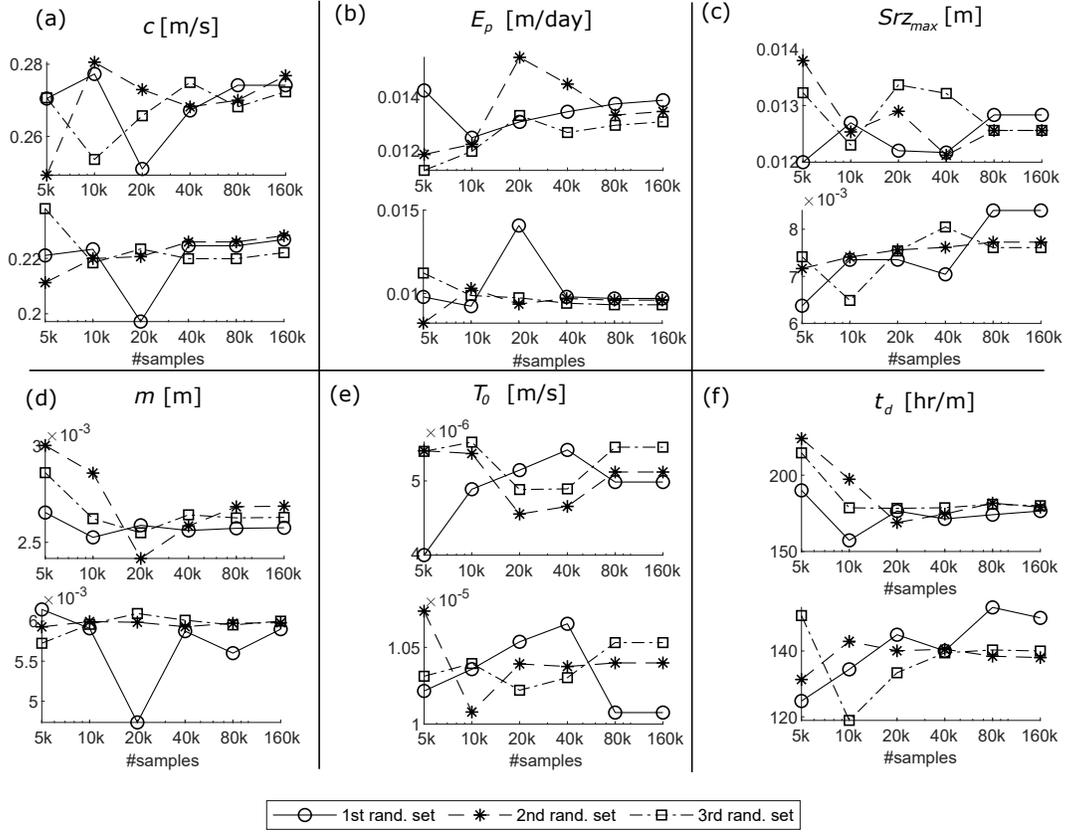


Figure 2: Results of the sampling convergence test for GLUE uncertainty estimation. In each panel top plot is the ‘before’ and bottom plot is the ‘after’ period.

## 4.2 Obtaining limits of acceptability

Limits of Acceptability (LOA) are upper and lower bound limits on the observational discharge within which the model predictions must lie to be deemed ‘behavioural’ and thus accepted. We seek to define LOA to minimise Type II errors, where a good model is rejected as a result of errors in validation or input data (Renard et al., 2010). These errors can arise both from measurement limitations/errors, and from commensurability and scaling issues where the time/space scale of measurements differ from those within the model. We estimate errors in: water level measurement, rating curve, and in the rainfall measurements that drive the model.

Discharge from each catchment was estimated at 10-minute resolution using a 90° triangular v-notch weir with stage behind the weir measured using pressure transducers. Stage was converted to discharge using a standard rating-curve relationship of the form  $\hat{q} = a\hat{h}^b$ , where  $\hat{q}$  is the observed discharge,  $\hat{h}$  is the observed stage, and  $a$  and  $b$  are empirical parameters related to geometry and discharge coefficient. Following Keeland et al. (1997), and making the conservative assumption that measurement accuracy was relatively low for such equipment, we use a stage measurement error standard deviation of  $\sigma = \pm 0.0075\text{m}$ . We then propagate this stage error through the rating curve using a standard frequentist inferential methodology (Petersen et al., 2005) to calculate an uncertainty ratio  $R$ , which is the span of the 95% confidence limit in discharge, divided by the observed discharge:

$$R = \frac{q_{95}}{\hat{q}} = \frac{3.92b\sigma}{\hat{h}} \quad (3)$$

where  $\sigma$  is the measurement error standard deviation. Therefore, the contribution of stage measurement errors in the LOA, calculated as upper and lower bounds on the discharge is taken to be  $(1 + R)\hat{q}$  and  $(1 - R)\hat{q}$ ,

228 respectively.

229 Stage was converted to discharge in the BACI experiment using theoretical rather than empirical rating-curve  
 230 relationships, thus we propagate parameter uncertainty in theoretical rating curves through the weir equation of  
 231 Shen (1981) relating observed discharge,  $\hat{q}[LT^{-1}]$  and stage,  $\hat{h}[L]$ :

$$\hat{q} = C_e \frac{8}{15} \sqrt{2g} \tan\left(\frac{\theta}{2}\right) \hat{h}^{\frac{5}{2}} \quad (4)$$

232 where  $C_e[-]$  is the non-dimensional coefficient of discharge,  $g[L^2T^{-1}]$  is the gravitational acceleration,  $\theta$   
 233 is the v-notch angle. For  $90^\circ$  weirs, the  $C_e[-]$  value varies between 0.5 and 0.7 (Shen, 1981), therefore these  
 234 values were used to obtain an upper and lower bound on discharge as maximum and minimum possible variation  
 235 in discharge as a result of rating curve uncertainty.

236 Since observed discharge is estimated from an instantaneous stage measurement every 10 minutes but mod-  
 237 elled discharge is the average of the volumetric water flux from the catchment over the 10-minute timestep,  
 238 there is a potential time-commensurability error between the two quantities. This error will be most severe  
 239 under rapidly changing discharge but it is difficult to conceive of scenarios in which instantaneous discharge  
 240 could differ from the 10-minute average by more than it differs from the instantaneous measurement 10 minutes  
 241 before or after. Thus the upper bound on discharge for timestep  $t$  is set to the maximum of the upper bounds  
 242 in timesteps  $t-1:t+1$ , unless  $t$  is a trough, when the upper bound is set to the minimum of  $t-1$  or  $t+1$ . The lower  
 243 bound for timestep  $t$  is set to the minimum in timesteps  $t-1:t+1$ , unless  $t$  is a peak, when the lower bound is set  
 244 to the maximum of  $t-1$  or  $t+1$ .

245 There is no commonly accepted method of evaluating rain gauge error, and particularly how it manifests  
 246 itself in the outlet discharge predictions when the erroneous rain gauge data is input to a numerical model to  
 247 predict a discharge (Vrugt & Beven, 2018). Rainfall measurement errors stem from errors in the measurements  
 248 themselves and from their inability to capture spatial and temporal variability (McMillan et al., 2012). Since  
 249 each catchment contains only a single raingauge recording at only 10-minute frequency we cannot estimate the  
 250 spatial and temporal variability of rainfall in our study catchments. However, the catchments are sufficiently  
 251 small (order of  $10^{-3}$  km<sup>2</sup>) relative to storm-cloud footprints (Sauvageot et al., 1999, Féral et al., 2006) and  
 252 the recording frequency sufficiently high relative to storm durations, that we expect errors due to spatial and  
 253 temporal variability to be small. Finally, point measurement errors can originate from different sources such  
 254 as the time-sampling effect caused by the discrete character of the tipping bucket measurements, water flow  
 255 instabilities in the gauge funnels, and turbulent airflow around the rain gauges caused by wind (Rodda, 1967,  
 256 Pollock et al., 2018). The BACI experiment has provided three rain gauge recordings, one for each catchment,  
 257 which are within 400 meters of each other ( $P_{CR}(t)$ ,  $P_{RV}(t)$  and  $P_{RG}(t)$ ). We use these to estimate raingauge  
 258 measurement errors by following Ciach (2003) in assuming that the average of rain gauge measurements is  
 259 likely to be closer to the true value than any of the individual recordings, due to smoothing of local random  
 260 errors. Thus, if  $\bar{P}(t)$  is the mean value of the precipitations, then, for the given timestep, the error is taken to be  
 261 the maximum deviation from the mean:

$$e_p(t) = \max(|\bar{P}(t) - P_i(t)|) \quad i \in \{CR, RV, RG\} \quad (5)$$

262 To include the contribution of rainfall measurement errors in the limits of acceptability, they must be trans-  
 263 lated into discharge error. However, non-linearity in rainfall-runoff models leads to accumulation of errors in  
 264 state variables, resulting in complex and non-traditional time series of residuals (Vrugt & Beven, 2018), and  
 265 making simple error propagation unfeasible. Instead, we choose to propagate the  $\bar{P}(t) \pm e_p(t)$  through the nu-  
 266 merical model and assess its magnitude of impact on discharge. Since, this must be performed prior to model  
 267 calibration, we run 160,000 simulations by sampling parameters from the prior distributions. We repeat this  
 268 exercise for the same parameter sets using three different rainfall inputs: (a)  $\bar{P} - e_p$ , (b)  $\bar{P}$  and (c)  $\bar{P} + e_p$ . The  
 269 resulting discharge prediction values for each rainfall case are then averaged across all the parameter values; the  
 270 distance between averaged discharge values of (a) and (c) compared to (b) is taken to be the upper and lower  
 271 bound as a consequence of rainfall measurement errors.

272 Finally, the limits of acceptability were calculated by adding the effects of all of the above mentioned errors,  
 273 in the form of upper and lower bounds on the observational discharge.

### 4.3 Obtaining parameter distributions

Figure 3 shows the step by step process that was undertaken in obtaining the parameter distributions which characterise each catchment-period under study. For the purpose of illustration, the potential evaporation parameter,  $E_p$ , for the CR catchment and in the ‘before’ period is shown here, but the procedure is the same for all parameters, catchments and periods.

A uniform random distribution of 160k samples is input to the model (Figure 3a), for each of which the model generates a discharge prediction. The parameter values associated with model predictions that sit outside of the LOA are discarded (Figure 3b). Note that since discharge predictions that remain within the LOA throughout the entire study period are extremely rare (Beven, 2006, Freer et al., 2003) we have allowed for 1% of predictions to be outside the LOA (similar to Blazkova & Beven (2009), Van Straten & Keesman (1991)). In Figure 3c, an extra condition is imposed on model predictions to reflect field observations in UK upland blanket peatlands that discharge response to rainfall is dominated by overland flow (Holden & Burt, 2002, 2003, Daniels et al., 2008, Holden, 2009). Thus the Overland Flow (OF) criteria requires that >70% of predicted outlet discharge should be composed of overland flow (a conservative, i.e. low, value based on field observations). Finally, after the model predictions are filtered by the LOA and OF criteria, the remaining (behavioural) parameter-sets are ranked according to model performance and linearly weighted (Eq. 2) such that better performers get higher weights. Using these weights, the weighted counts of the histogram bins are calculated, which are then normalised by dividing by the largest value of the weighted counts to obtain the Normalised Weighted Counts (NWC), which vary between 0 and 1. These normalised weighted distributions (Figure 3d) enable comparison between catchments’ distributions. All subsequent parameter distributions are normalised weighted distributions, their y-axis range is always 0-1, and thus is not shown.

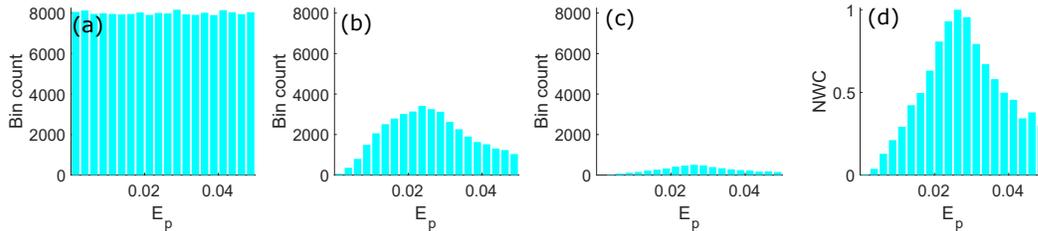


Figure 3: Step by step demonstration of how parameter distributions were obtained. Example for evapotranspiration parameter,  $E_p$ . NWC stands for Normalised Weighted Count of the histogram bins.

## 5 Results

### 5.1 Parameter distributions before and after interventions

Using the procedure described in Section 4.3, the normalised weighted distributions for all parameter-catchment-period combinations are generated and shown in Figure 4. The top and bottom rows of each panel contain the ‘before’ and ‘after’ distributions respectively. Note that even at the control site, where there has been no intervention, the parameter distributions differ between the ‘before’ and ‘after’ periods. Shuttleworth et al. (2019) also observed changes in hydrological behavior at the control site (emphasising the need for a BACI experiment) and referred to it as ‘natural temporal variability’. Thus distributions should be interpreted relative to the control site in each case.

These distributions highlight the parts of the parameter space that performed better while satisfying the LOA and OF criteria. In GLUE terms, these distributions can be interpreted as indicating the likelihood that particular parts of the parameter space represent the true values for the catchment of interest.

Examining the shifts in parameter distributions following interventions and starting from the top left panel of Figure 4, distributions of overland flow response (celerity,  $c$ ) in the ‘before’ period differ slightly among catchments, however, all three assign almost zero likelihood to  $c$  values smaller than 0.1 m/s. There is an

310 overall reduction in surface flow celerity in the ‘after’ period in all sites, but the reduction is larger in RV and  
 311 RG than in CR, and is largest in RG.

312 Root zone storage  $S[L]$ , in the model, accounts for overall interception storage of vegetation, within the  
 313 root zone or in surface ponds (whether natural or behind gully blocks). It has a maximum capacity  $S_{rz_{max}}$ ; is  
 314 emptied only by evapotranspiration at a rate  $E_a = E_p \frac{S}{S_{rz_{max}}}$ ; and is the only source of evapotranspirative loss in  
 315 the model. Unless this store is full rainfall cannot contribute to either vertical infiltration or surface flow. Before  
 316 the interventions, the distributions of potential evapotranspiration ( $E_p$ ) are very similar in CR and RV, while RG  
 317 is characterised by higher  $E_p$ . After the interventions, all sites shift towards higher  $E_p$  values, with RG larger  
 318 than CR and RV slightly larger than RG. Before the interventions, the distributions of  $S_{rz_{max}}$  are very similar in  
 319 all sites. After interventions, all sites shift towards higher  $S_{rz_{max}}$ , the increase is similar for CR and RV, while  
 320 RG shows a clear increase relative to CR.

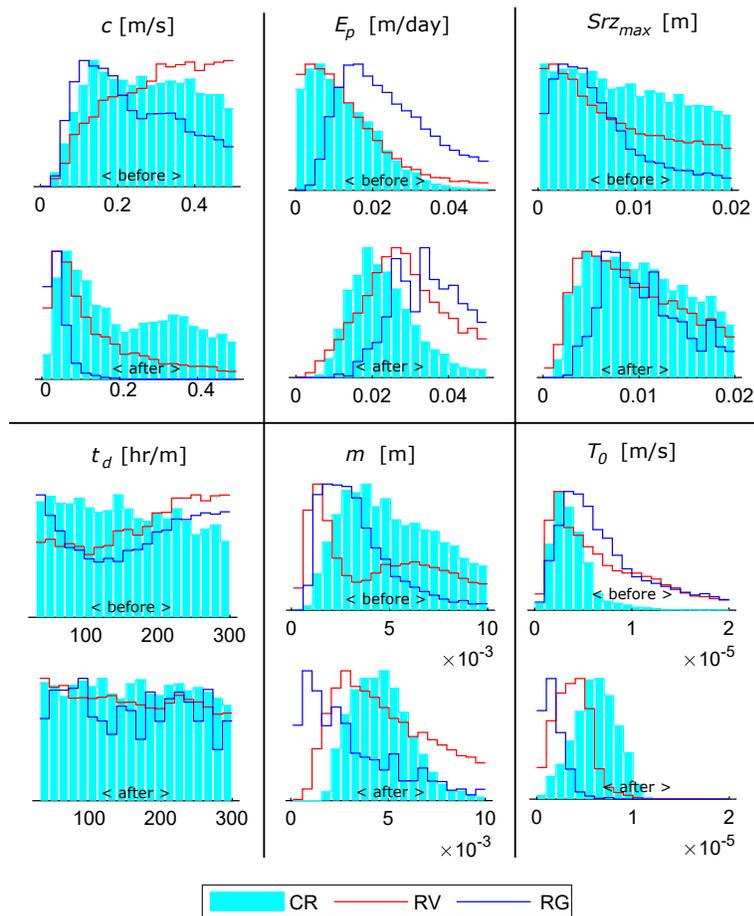


Figure 4: Before-After intervention parameter distributions for all three catchments under study. *CR* denotes the control, *RV* denotes the revegetated and *RG* denotes the revegetated & gully blocked case. Top row are the ‘before’ distributions and bottom are the ‘after’. The y-axis range is 0-1 for all distributions and thus is not shown. For description of parameters see Section 3.

321 The unsaturated zone time delay parameter ( $t_d$ ) controls the vertical flux of water from the peat surface to  
 322 the saturated zone below. The model is able to produce good fit to the observed discharge across the full range  
 323 of unsaturated zone time delays both before and after interventions and at all sites. This indicates that the model  
 324 is not sensitive to this parameter. This is perhaps due to the very shallow water table depth in peat catchments  
 325 and their propensity to form saturated areas during a storm, limiting both the vertical and lateral extent of the

326 unsaturated zone.

327 The parameters  $m$  and  $T_0$  control the catchment’s transmissivity-depth profile and thus the relationship be-  
328 tween subsurface flow and saturation deficit.  $T_0$  sets the maximum transmissivity when the peat is fully sat-  
329 urated, while  $m$  controls the rate of exponential decay in transmissivity as saturation deficit increases (higher  
330 values indicate slower decay). Before the interventions,  $T_0$  values are similar at all sites. After the interventions,  
331  $T_0$  shifts towards higher values for RV, but lower values for RG, though this difference may be offset by an  
332 increase in  $m$  for RV and decrease for RG after interventions.

333 Care is required in interpreting these parameter distributions. They show the magnitude and direction of  
334 individual parameter shifts in response to calibration to each catchment-period pair, but it is difficult to infer  
335 how significant these changes are when parameter ‘sets’ are used to predict runoff. This is because the impact of  
336 a parameter shift (e.g. in  $E_p$ ) is controlled by both its magnitude and the model’s sensitivity to that parameter.  
337 Since parameters interact this will also depend on the values of other parameters. This is clearly true for  $m$  and  
338  $T_0$ , which together control transmissivity but is also important for evapotranspiration, which is controlled by  
339 both  $E_p$  and  $S_{r,max}$ . For this reason, in the following sections we perform numerical experiments to investigate  
340 how the model’s runoff predictions change in response to varying sets of ‘parameter-sets’, each of which pertains  
341 to a catchment-period pair.

## 342 5.2 Numerical experiment #1: Digital Twins

343 The first numerical experiment uses the behavioural parameter-sets of Section 5.1 to examine the hydrograph  
344 response of each catchment in terms of peak magnitudes and timings. In order to isolate the impact of the inter-  
345 ventions we create catchments that are ‘digital twins’. These twins have the same topography, flow pathways,  
346 input rainfall, and antecedent conditions and differ only in the set of behavioral parameter-sets that represent  
347 each intervention.

348 First, we used RG’s topography, and input rainfall for the 2010 study period to drive the model. The  
349 model was then evaluated with four different sets of ‘behavioural’ parameter-sets: (1) pre-RV and (2) post-  
350 RV representing the pre- and post-intervention conditions at the re-vegetated site; (3) pre-RG and (4) post-RG  
351 representing pre- and post-intervention conditions at the revegetated and gully-blocked site. We then repeated  
352 this process but using the RG’s 2012 rainfall. Finally, we repeated the whole process but using RV’s topography,  
353 and RV’s rainfall from the 2010 and 2012 study periods (i.e., all combinations of the two rainfalls and two  
354 topographies).

355 The results are shown in Figure 5. Throughout this figure cooler colors represent pre-intervention conditions  
356 and hotter colors represent post-intervention. Due to the length of the records we only show the largest storms in  
357 the record (belonging to the 2010 ‘before’ period) in panels (a) and (b), and the most complex storm (belonging  
358 to the 2012 ‘after’ period) in panel (c). The middle line of each color band is the mean of all predictions and  
359 the width of the bands are indicative of the variation in the predictions across all behavioural parameter-sets.  
360 The numbers 1-8 are peak ID numbers and the smaller panels on the right of the figure are the associated  
361 Probability Density Function (PDF) of each peak-discharge and peak-timing prediction (left and right columns,  
362 respectively). In each of the small boxes, upper PDFs are for the RV case and lower (upside down) PDFs are  
363 for RG. These PDFs show how (in)significant the effects of interventions have been on each peak, such that the  
364 more overlap there is between the pre- and post-intervention PDFs, the less significant the effects have been for  
365 that particular peak, and vice versa.

366 Post-intervention hydrographs are noticeably smoother than pre-intervention (Figure 5). This reduced am-  
367 plitude and increased wavelength of the discharge timeseries is likely due to reduction in the speed of surface  
368 flow response (celerity) associated with increased surface roughness following revegetation and gully-blocking.  
369 Interventions almost always reduce peak discharge and delay the timing of the peaks. This attenuation is present  
370 in both large and small peaks (e.g. peaks 1 and 2) though it appears less marked in some peaks (e.g. peak 3)  
371 than others. Reduced attenuation of peak 3 is also clearly illustrated by the lack of separation between its pre-  
372 and post-intervention PDFs, particularly for peak magnitude and partially for timing (Figure 5).

373 To investigate whether there is a systematic relationship between storm size and peak-discharge/lag-time  
374 change, we examined the 40 most prominent peaks in the record (10 from each period-topography combina-  
375 tion, Figure 6a-b). For each peak, we used the weights from Eq. 2 to calculate the weighted-average of the

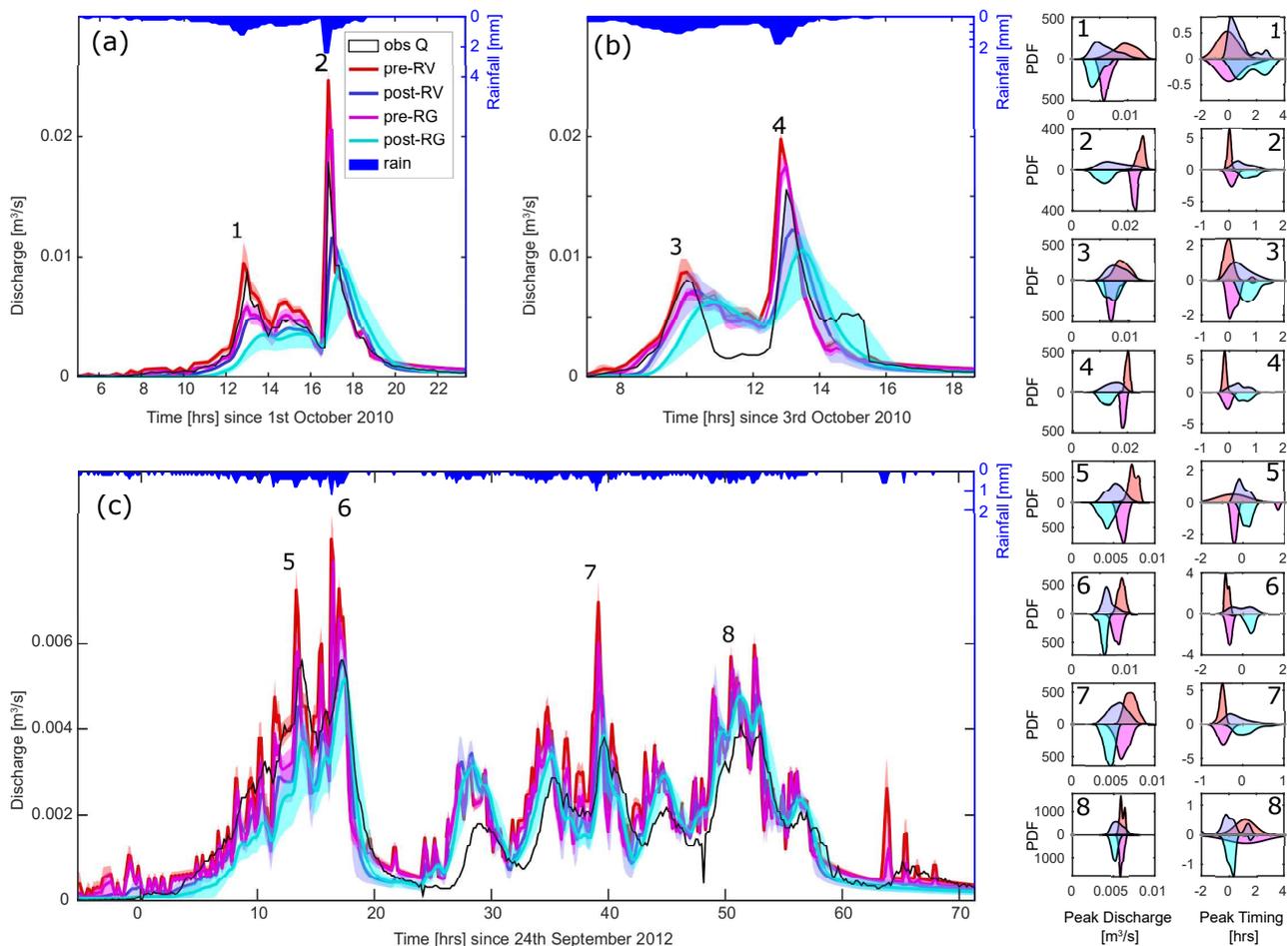


Figure 5: Digital Twins experiment: predicted discharge under four different model parameterisations (panels a to c). Numbers 1-8 are peak ID numbers. Right of the figure are Probability Density Functions (PDFs) of each peak.

376 behavioural model predictions pre- and post-intervention ( $Q_{pre}$  and  $Q_{post}$ , respectively). The post-intervention  
 377 change in peak discharge is then calculated as a percentage of the  $Q_{pre}$  using:  $\% = (Q_{post} - Q_{pre}) / Q_{pre} * 100$ ,  
 378 which is the y-axis of Figure 6a. For peak timings, the same procedure is followed but without normalisation.  
 379 Thus absolute changes in timing are shown on the y-axis of Figure 6b.

380 Note that in obtaining Figure 6 sometimes precise identification of the timing of the hydrograph peaks was  
 381 not possible. Delays and attenuations in the predicted hydrographs cause some peaks to interfere or merge  
 382 making it difficult or impossible to identify corresponding peaks in ‘pre’ and ‘post’ model evaluations. This  
 383 difficulty is demonstrated by the peak-timing PDF of peak 8 (Figure 5), where PDFs of RG show considerable  
 384 ‘decrease’ in lag-time following revegetation and gully-blocking which is inconsistent with the hydrographs  
 385 for peak 8 (Figure 5c). Peak magnitude change is not affected by this issue, because in this case it is only the  
 386 magnitude of the peaks that matter and thus peaks do not necessarily need to correspond exactly. However, this  
 387 introduces considerable uncertainty in estimates of timing change. To minimise these uncertainties we manually  
 388 identify ‘complex’ multi-modal peaks (such as peak 8), remove them from peak timing analysis (Figure 6b) and  
 389 flag them in our peak magnitude analysis (as circles rather than squares in Figure 6a).

390 Both interventions resulted in significant (95% CI) reduction in peak discharge for all storms independent  
 391 of storm size (Figure 6a). Even peaks 3 and 8, which show least reduction are significant at 95% confidence.  
 392 Mean discharge reduction was 48% for RV and 58% for RG suggesting that gully-blocking results in further

393 reductions in peak discharge (above those from revegetation alone). These mean values aggregate considerable  
 394 inter-storm variability with 90% reduction in peak discharge possible in some of the smaller storms but less  
 395 than 20% reduction in four storms for RV and two for RG. The 2010 study period includes the two largest  
 396 storms (peaks 2 and 4 in Figure 5). These storms have rainfall intensity-duration properties that suggest an  
 397 approximately annual return period (Jones et al., 2010) and peak discharges within the largest 10% over 10  
 398 years of observations at the control site. These two storms have modelled discharge reductions of 30-40% for  
 399 RV and 40-55% for RG.

400 For the non-complex storms for which timing analysis was possible, interventions always increased lag-  
 401 times (Figure 6b). The lag-time increase was always greater than 10 minutes for revegetation alone and greater  
 402 than 30 minutes for revegetation and gully blocking, with overall mean (across 40 peaks in 2 periods for 2  
 403 topographies) of 0.47 hr (28 min or 148%) in RV and 0.83 hr (50 min or 265%) in RG. For the two largest  
 404 storms, the lag increase in RV was roughly 0.5 hr (30 min or 159%) and in RG was around 0.9 hr (55 min or  
 405 280%).

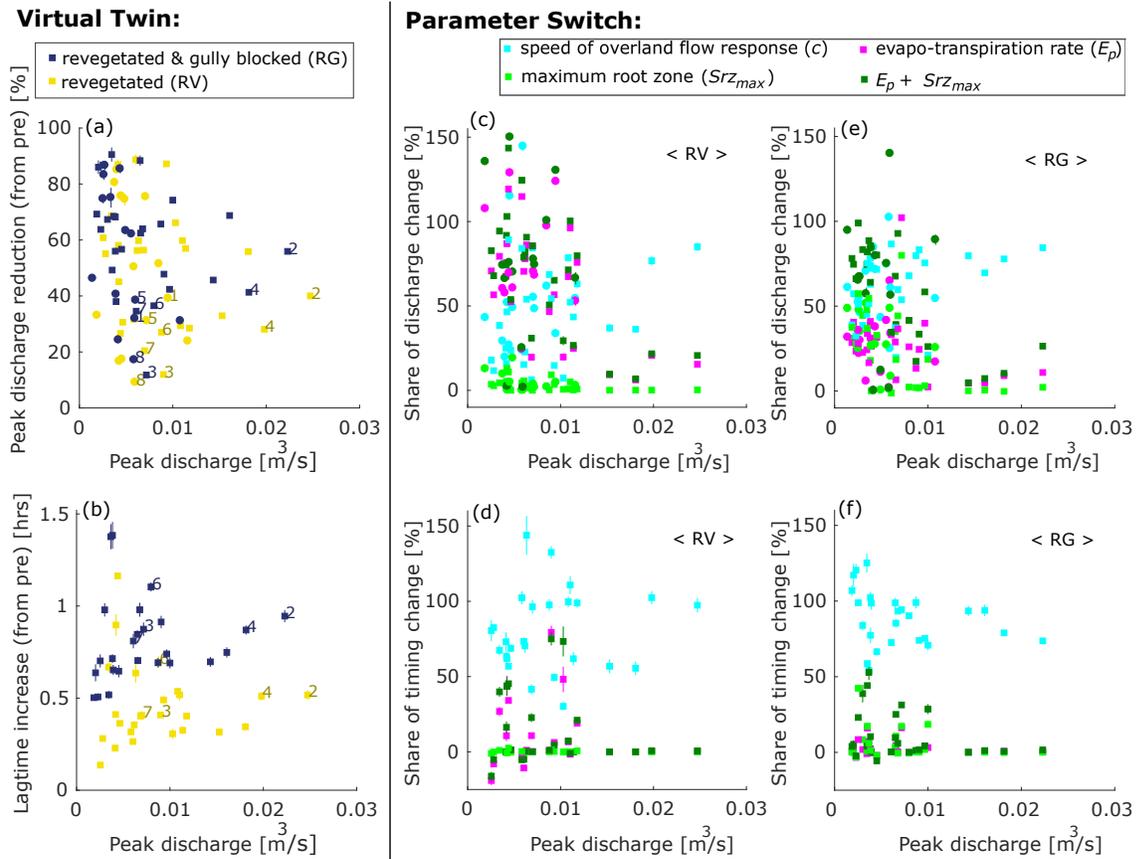
### 406 5.3 Numerical experiment #2: Parameter Switch

407 The Digital Twins experiment estimated the impact of interventions when all parameters are changed from pre-  
 408 to post-intervention values and Figure 4 indicates the magnitude and direction of parameter changes. However,  
 409 the relative importance of each parameter in driving the intervention impacts remains unclear. The second  
 410 numerical experiment aims to identify the parameters (and therefore processes) driving the observed effects of  
 411 a particular intervention.

412 The experimental design is similar to the ‘Digital Twins’ experiment, holding rainfall and topography con-  
 413 stant and varying only the input parameter-sets (the ‘behavioural’ parameter-sets for each intervention). How-  
 414 ever, here we perform a one-at-a-time parameter switch, where a single parameter (or pair of parameters) is  
 415 switched to its post-intervention values while all others are kept at their pre-intervention values. Changes in the  
 416 predicted peak magnitude and timing can then be assigned to the switched parameter. We calculate the asso-  
 417 ciated discharge change as a percentage of the post-intervention discharge where ‘all’ parameters are switched  
 418 to their ‘post’ values. To do this we calculate the weighted-average (Eq. 2) of all post-intervention discharge  
 419 predictions for the case where only one parameter  $p$  is switched and all other are kept at ‘pre’ values ( $Q_{post.p}$ ).  
 420 We then weighted-average the discharge predictions for the case where ‘all’ parameters are switched to their  
 421 ‘post’ values ( $Q_{post.all}$ ). The percentage change is then given by:  $\% = (Q_{post.p} - Q_{post.all}) / Q_{post.all} * 100$  (Figure  
 422 6 c&e). For peak timing a similar procedure is followed except that the results are not converted to percentage  
 423 but shown as absolute differences (Figure 6 d&f). Note that it is possible for the discharge/timing change due to  
 424 a single parameter switch to be larger than that when all parameters are at the post-intervention values, because  
 425 by switching parameters one-at-a-time we ignore parameter interactions, which might be counteracting one an-  
 426 other. Thus the y-axis in Figure 6c-f cannot be interpreted as the share of intervention impact in absolute terms,  
 427 rather as a relative indicator of that parameter’s importance.

428 Lateral hydraulic conductivity related parameters ( $m$  and  $T_0$ ) when switched in together (because they jointly  
 429 represent transmissivity) were found to have no significant share (with 95% CI) in delaying the flow or reducing  
 430 the peak discharge, indicating that transmissivity is not altered two years after interventions; similarly for the  
 431 vertical trasmissivity parameter ( $t_d$ ). Therefore, we assume the remaining parameters  $c$ ,  $S_{rz_{max}}$  and  $E_p$  are re-  
 432 sponsible for the total impact of each intervention and show only these parameters in Figure 6c-f. Additionally,  
 433 since  $S_{rz_{max}}$  and  $E_p$  are closely linked through their influence on  $E_a$  (see section 5.1), we study the overall effect  
 434 of rootzone storage using the paired parameters  $S_{rz_{max}} + E_p$ .

435 In both treatments, and for the larger ( $>0.01 \text{ m}^3/\text{s}$ ) peaks, discharge reduction is predominantly controlled  
 436 by celerity, although the impact of  $E_p + S_{rz_{max}}$  is still significant ( $>10\%$  of celerity’s impact). For smaller peaks  
 437 ( $\leq 0.01 \text{ m}^3/\text{s}$ ) discharge reduction is controlled primarily by  $E_p + S_{rz_{max}}$ . In terms of the relative role of  $E_p$  and  
 438  $S_{rz_{max}}$ , in RV,  $E_p$  dominates, while in RG, both are important. Unlike peak discharge plots, peak timings in  
 439 panels (d) and (f) are more consistently controlled by celerity as the primary delaying mechanism independent  
 440 of storm size or intervention. The effect of celerity is more pronounced in RG than RV (double on average).



**Estimates of interventions' impact in terms of NFM:**

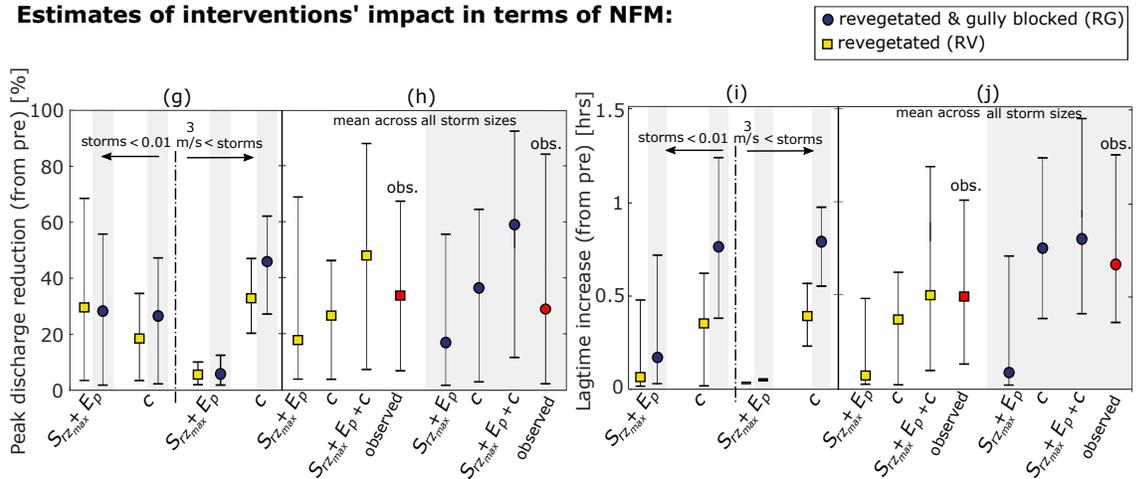


Figure 6: **Digital Twins experiment:** 'absolute' magnitude of intervention impacts for different storm sizes and in terms of: (a) peak magnitude reduction relative to pre, (b) peak timing delay relative to pre. **Parameter Switch experiment:** 'relative' (to one another) impact of restoration intervention for different parameters and storm sizes: (c) and (d) parameter shares in peak reduction and lag increase, respectively, in RV. (e) and (f) for RG. **Estimates of interventions' impacts:** (g) and (i) magnitude reduction and lag increase, respectively, for different storm categories. (h) and (j) mean impact across all storm sizes and comparison with observed means.

## 5.4 Estimating the NFM-potential of the restoration interventions

In their analysis BACI analysis [Shuttleworth et al. \(2019\)](#) noted significant variability in catchments' behaviour. Two important sources of uncertainty are: (1) differences between catchments in topography and material properties of surface and subsurface flow pathways; and (2) differences between rainfall characteristics in the pre- and post- intervention period. By assessing deviations from the control site, and examining a large sample of storms [Shuttleworth et al. \(2019\)](#) were able to detect the impact of restorations despite this variability. However, it is not clear how much of the observed effects were due to differences between catchments or in rainfall characteristics. Fixing topography and rainfall within the Digital Twins experiment will reduce, but not eliminate these uncertainties because the parameter-sets were calibrated to the observed rainfall-discharge time series.

Unlike the Digital Twins experiment, the magnitude of the changes predicted using Parameter Switch are not absolute. To estimate the absolute impact of each parameter ( $\approx$  process) on peak discharge and lag-times we combine the 'absolute' impact-magnitude of interventions predicted by the Digital Twins, with the 'relative' impact-magnitude of parameters predicted by the Parameter Switch. Since  $m$ ,  $T_0$  and  $t_d$  had negligible effects, we exclude them from this analysis. Due to the coupled nature of  $E_p$  and  $S_{rz_{max}}$  we treat them as a pair. Thus, we assume that the total absolute magnitude of interventions predicted by the Digital Twins, is shared between the speed of overland flow response  $c$ , and the parameter pair  $E_p+S_{rz_{max}}$ .

To obtain an estimate for the absolute impact-magnitude of each intervention that can be assigned to  $c$  and  $E_p+S_{rz_{max}}$ : (i) for each of the 40 peaks, and each parameter, we calculate a weight by dividing that parameter's share (from Parameter Switch) by sum of shares of both parameters, such that weights add to one; (ii) for each peak and each parameter, we multiply this weight by the absolute impact magnitude of the intervention (from the Digital Twins); (iii) for each parameter, we calculate mean values for peaks  $<0.01$  m<sup>3</sup>/s, peaks  $>0.01$  m<sup>3</sup>/s, and for the full sample of storm sizes. We then compare these to the case where all three parameters  $c$ ,  $E_p$ , and  $S_{rz_{max}}$  are switched to their post-intervention values (a full parameter-switch), and to the observed values (Figure 6g-j).

In Figure 6g-j, dots are the mean and the error bars represent the range (min-max). Panels g&i show that, celerity,  $c$ , alone is almost entirely responsible for lag-time increase, regardless of storm size or intervention. Peak reduction is dominated by  $c$  in large storms regardless of intervention. In RV and in smaller storms  $E_p+S_{rz_{max}}$  have slightly more impact in reducing the peak magnitude than does  $c$ . But in RG, and even for smaller storms,  $c$  reduction due to surface roughness is the dominant process reducing peak magnitudes.

Figure 6h&j show the mean of impact-magnitudes across all storm sizes, for  $c$  and  $E_p+S_{rz_{max}}$  individually, and for their combination. When comparing the impact of  $c+E_p+S_{rz_{max}}$  to the observed mean from the BACI experiment, there is considerable overlap in the predicted and observed range of peak discharge and lag-times. This agreement between the two alternative (mechanistic and statistical) analyses is encouraging. Note that, using only 2012 as the 'after' intervention period (modelled here), the observed reduction in peak discharge is smaller at RG than RV. This is unexpected given the additional gully blocking undertaken in RG and is not consistent with other years (2013 and 2014, see ([Shuttleworth et al., 2019](#))) where the reduction in peak discharge was significantly larger at RG than RV. However, the mean of modelled peak magnitude reductions in RV is 10% and in RG is 30% more than the observed mean, which is more consistent both with our expectations (that gully blocking should provide additional NFM benefit) and with results from the succeeding years. This suggests a possible underestimation of the impacts by the BACI experiment due to experimental limitations/noise.

## 6 Discussion

### 6.1 Process representation: immobile vs. mobile surface storage

[Shuttleworth et al. \(2019\)](#) showed that water tables at treatment sites rose following both revegetation and gully-blocking, which resulted in a reduction in catchment subsurface storage, leading to an enhanced production of saturation-excess surface flow. Despite the increase in surface flow production, lag-times increased and peak-discharges decreased, suggesting that: (1) the additional surface flow is stored in some form of 'surface' storage, (2) this increase in total surface storage more than offsets the effect of reduced subsurface storage due to water-table rise. Here we distinguish between two different types of surface storage, 'immobile' and 'mobile' surface

489 storage.

### 490 **6.1.1 Immobile surface storage**

491 We define 'immobile' surface storage as the portion of surface water that is intercepted (and immobilised) by  
492 vegetation and in-channel gully-blocks, and is only emptied through evapotranspiration. During a rainfall event,  
493 immobile storage can intercept some of the rainfall and reduce hydrograph peaks. It may also delay the peaks  
494 by a lag proportional to the time it takes to fill the store. This storage is represented by the root zone storage  
495 variable,  $S[L]$ , in the model, with a maximum capacity  $S_{rz_{max}}$  and emptying rate  $E_a = E_p \frac{S}{S_{rz_{max}}}$ . Together, the  
496 paired parameters  $S_{rz_{max}} + E_p$  capture the overall behaviour of immobile storage; individually they provide useful  
497 information on the interplay of storage capacity and evapotranspiration in the immobile storage unit. Note that  
498 due to its immobility, immobile storage also represents the portion of rainfall that does not contribute to runoff.  
499 Therefore, increasing immobile storage can reduce runoff coefficients (ratio of total catchment discharge to  
500 total storm rainfall). In connection to this, [Shuttleworth et al. \(2019\)](#) found no statistically significant change in  
501 runoff coefficients following either revegetation or gully-blocking, suggesting that the change in the immobile  
502 surface storage post-treatments was not the primary driver of observed peak reduction and lag increase in these  
503 sites.

### 504 **6.1.2 Mobile surface storage**

505 We define 'mobile' surface storage as the amount of surface water in motion, that is not yet delivered to the  
506 outlet, at any given point in time. Contrary to immobile storage, water in mobile storage will all eventually  
507 leave the catchment as discharge, thus will not alter runoff coefficients. In TOPMODEL, mobile surface storage  
508 is not explicitly represented by a parameter, but it is inversely proportional to the speed of surface flow response,  
509 i.e., flow celerity parameter ( $c$ ). Speed of response in this context is the time it takes for perturbations in  
510 the mobile surface storage unit, which travel overland, to arrive at the catchment outlet. Thus, during rainfall,  
511 rougher surfaces (for example due to re-vegetation or stream blocking) reduce celerities, retain water within the  
512 catchment for longer, and lead to accumulation of water in the mobile surface storage. Here, for a given rainfall  
513 event, the rougher the surface, the larger the mobile surface storage volume that the catchment can hold.

## 514 **6.2 Surface storage post-interventions: role of storm size**

### 515 **6.2.1 Immobile surface storage**

516 We first investigate the overall immobile surface storage effect, through the coupled parameters  $S_{rz_{max}} + E_p$ .  
517 Investigation of Figure 6g suggests that in both RV and RG, changes to immobile storage (i.e.  $S_{rz_{max}} + E_p$ ) are  
518 the primary cause of peak reduction in smaller storms ( $\leq 0.01 \text{ m}^3/\text{s}$ ), although closely followed by celerity  
519  $c$ . In larger storms ( $> 0.01 \text{ m}^3/\text{s}$ ) immobile storage is much less important for reducing the peaks, but is never  
520 negligible (it is at least 10% of the effect of celerity on average). Figure 6i shows that in both sites and in smaller  
521 storms, the role immobile storage in delaying the flow is small but not negligible. Also immobile storage in RG  
522 delays the flow more than RV in smaller storms. In larger storms however, regardless of treatment, the impact  
523 of immobile storage on peak timing delay is negligible.

524 Regarding the relative roles of  $S_{rz_{max}}$  and  $E_p$  following each treatment, their coupling in TOPMODEL (see  
525 section 5.3) means that: (1) if  $E_p$  is small relative to the input rainfall intensity, regardless of  $S_{rz_{max}}$ , evapotran-  
526 spirative losses are small 'during' a storm. If  $S_{rz_{max}}$  is also small, there are limited additional evapotranspirative  
527 losses between storms because there is little stored water to lose. But if  $S_{rz_{max}}$  is large, these losses will extend  
528 to the inter-storm period, making more rootzone storage available at the onset of the next storm. In this case  
529 the significance of this additional immobile storage depends on the storm sequencing pattern. (2) if  $E_p$  rate is  
530 comparable to the input rainfall intensity, regardless of  $S_{rz_{max}}$ , evapotranspirative losses 'during' the storm will  
531 also be considerable. If  $S_{rz_{max}}$  is small, again these losses primarily occur during the rainfall event but this time  
532 they are considerable. If  $S_{rz_{max}}$  is also large however, the evapotranspirative losses will extend to the inter-storm  
533 period and at higher rates, leading to even more rootzone storage available at the onset of the succeeding storm  
534 which must be filled before surface runoff can be generated (i.e., more peak reduction and delay).

535 In this context, Figure 6c shows that the increase in  $S_{rz_{max}}$  due to revegetation has only a small impact on peak  
 536 discharge independent of storm size. This suggests that revegetation changes immobile surface storage primarily  
 537 by raising  $E_p$  rates emptying the store more rapidly rather than by increasing interception storage capacity. This  
 538 in turn suggests that the mechanism through which revegetation increases the catchment immobile surface  
 539 storage is the evapotranspiration occurring 'during' the storm. Further comparing RV and RG in panel (g)  
 540 shows that independent of storm size, on average the net effect of  $S_{rz_{max}}+E_p$  in peak reduction is not significantly  
 541 different. On the other hand, Figure 6e shows that gully blocking increases not only  $E_p$  (as it did for RV) but  
 542 also  $S_{rz_{max}}$ . Together these two observations suggest that, gully-blocking has increased  $S_{rz_{max}}$ , but has reduced  
 543  $E_p$  rates such that blocks have no net effect on immobile surface storage when compared to the revegetation-  
 544 only case. This in turn suggests that the  $E_p$  rates of a unit vegetated surface area are higher than that of a pond  
 545 behind the blocks.

546 Higher  $E_p$  rates of a unit vegetated surface area than that of a pond predicted by our model is surprising, but  
 547 is consistent with findings from a range of other studies which examine the ratio of actual evapotranspiration  $E_a$   
 548 to open water evaporation  $E_w$  for wetland ecosystems in a range of climates (e.g., (Dolan et al., 1984, Campbell  
 549 & Williamson, 1997, Drexler et al., 2004)). These studies suggest significant variability in the ratio, both above  
 550 and below 1 that is strongly sensitive to biophysical factors linked to vegetation type and water supply. In  
 551 particular, when canopy resistance to evaporation is low ( $<100$  s/m) wetland vegetation increases  $E_a$  relative  
 552 to open water (Mohamed et al., 2012); similarly, evaporation from wet leaves may exceed that of open water  
 553 because the vegetation is able to withdraw more energy from the atmospheric boundary layer by reducing its  
 554 aerodynamic resistance to evaporation (Savenije, 2004).

555 In terms of peak timing, Figure 6i suggests that in large storms and independent of treatment, the delay  
 556 caused by the process of filling the immobile surface storage is negligible. In smaller storm this delay is more  
 557 significant and is on average 5 mins in RV and 10 mins in RG. These suggest that the additional immobile  
 558 surface storage added by revegetation can slightly delay the flow but only in smaller storms; it also suggest  
 559 that gully-blocking further increases this delay in smaller storms, but it is still not significant in large storms.  
 560 Additional inference can be made by comparing panels (d) and (f) in Figure 6, which suggest that the underlying  
 561 mechanisms through which these treatments delay the flow are different; in RV these lag-times are primarily  
 562 delivered through increased  $E_p$  rates whereas in RG primarily through increased  $S_{rz_{max}}$ . Again, this suggests  
 563 increased immobile storage capacity as a result of gully blocking but reduced  $E_p$  relative to revegetation.

## 564 6.2.2 Mobile surface storage

565 For both RV and RG, averaged across all storm sizes, celerity is the most important parameter contributing to  
 566 both peak discharge reduction and lag-time increase (Figure 6h&j). Similar to the findings of Shuttleworth et  
 567 al. (2019), the fraction of the hydrograph change that can be accomplished by changing celerity alone at RG is  
 568 almost double that at RV, both in peak-discharge reduction and lag increase. Note that being semi-distributed,  
 569 TOPMODEL only treats the hillslopes, channels and blocks in a bulk sense and therefore the celerity in the  
 570 model is also a bulk celerity averaged over the catchment area. Thus, although it will not represent the true  
 571 celerity of different parts of the hillslope, channel or individual blocks, it does represent the average response-  
 572 speed of surface flow necessary to reproduce the observed hydrograph for the studied rainfall-catchment com-  
 573 binations. Therefore, the significant reduction in celerity predicted by the model following these interventions  
 574 has three important implications. First, the additional mobile surface storage due to both interventions (RV and  
 575 RG) is consistently the dominant process for both peak reduction and lag increase. This is important because it  
 576 suggests that hydrograph attenuation (magnitude reduction and delay due to changes in hydrograph shape rather  
 577 than volume) resulting from interventions is not limited to smaller storms. Second, the mobile surface storage  
 578 associated with gully-blocks is much more important than their immobile storage. Third, in terms of mobile  
 579 surface storage the additional impact of gully-blocking is as large as that of revegetation (see Figure 6g-j).

580 Although it is intuitive that gully blocking should significantly reduce mean particle velocities of surface  
 581 flow (as measured by tracers), it is far less clear how gully blocking reduces celerity (see section 5.4). For  
 582 example, celerity estimates based on the St. Venant equation relate  $c$  to both flow depth and velocity ( $c =$   
 583  $v \pm \sqrt{gy}$ ),  $v[LT^{-1}]$ : velocity,  $g[LT^{-2}]$ : gravitational constant,  $y[L]$  flow depth) (McDonnell & Beven., 2014). It  
 584 is easier to imagine how celerity might be reduced in the case of permeable blocks due to their nonlinear storage

585 discharge relationship (Milledge et al., 2015, Metcalfe et al., 2017). For an impermeable block such as a timber  
 586 dam, the situation is more complex. Flow depth changes discontinuously along the channel length: due to gully  
 587 bed slope, flow depth gradually increases towards its maximum at the block, but then rapidly decreases to the  
 588 thickness of the sheet of water flowing over the top of the block. Therefore, compared to a non-blocked case,  
 589 mean particle velocity is much lower behind the block, while the  $\sqrt{(gy)}$  term in celerity is much higher due  
 590 to deeper flow. However, in the thin flow over the block the  $\sqrt{(gy)}$  term suddenly and significantly decreases,  
 591 while at the same time mass continuity forces a rapid increase in mean particle velocity (increasing the  $v$  term).  
 592 In reality water then flows into the subsequent blocks, with block interactions (block spacing, configuration,  
 593 etc.) introducing additional complexity.

594 In our study site (RG) gullies were blocked with 20 impermeable timber dams and 17 cobble dams that  
 595 were initially permeable but that tend to clog with sediments reducing their permeability over time. Our model  
 596 results suggest considerable celerity reduction due to gully blocking, but the mechanism for such a reduction  
 597 as well as the extent to which it can be attributed to permeable or impermeable blocks cannot be untangled in  
 598 this modelling framework. Explicit surface flow routing and block permeability representation could feasibly  
 599 address this uncertainty and would be valuable for informing future block design.

### 600 6.3 Role of storm duration

601 In the post-intervention period, despite significantly delaying the flow, peak 3 in Figure 5b exhibits very little  
 602 reduction in peak discharge when compared with other peaks, and particularly with peak 1, which is very similar  
 603 in magnitude and overall hyetograph shape. This can also be seen in the peak discharge PDFs of peak 3 where  
 604 there is no separation between ‘pre’ and ‘post’ PDFs for either treatment, meaning that the magnitude of change  
 605 in discharge was not significantly different (with 95% confidence) between these periods.

606 The main observed difference between rainfall-runoff behaviours of peaks 1 and 3 is in their duration of  
 607 peak rainfall intensity (see Figure 7). The storm responsible for Peak 3 has a hyetograph that features a wide  
 608 peak, such that the maximum intensity is almost constant for a span of 70 minutes. The hyetograph associated  
 609 with Peak 1 has a much narrower peak, spanning only around 10 minutes. On the other hand, the estimated  
 610 post-restoration mean lag-times for RV and RG were 28 and 50 minutes, respectively. Thus, the duration of the  
 611 peak rainfall intensity in peak 1 is shorter than the mean lag-times provided by both sites, whereas for peak 3  
 612 duration of peak intensity rainfall exceeds their lag-times.

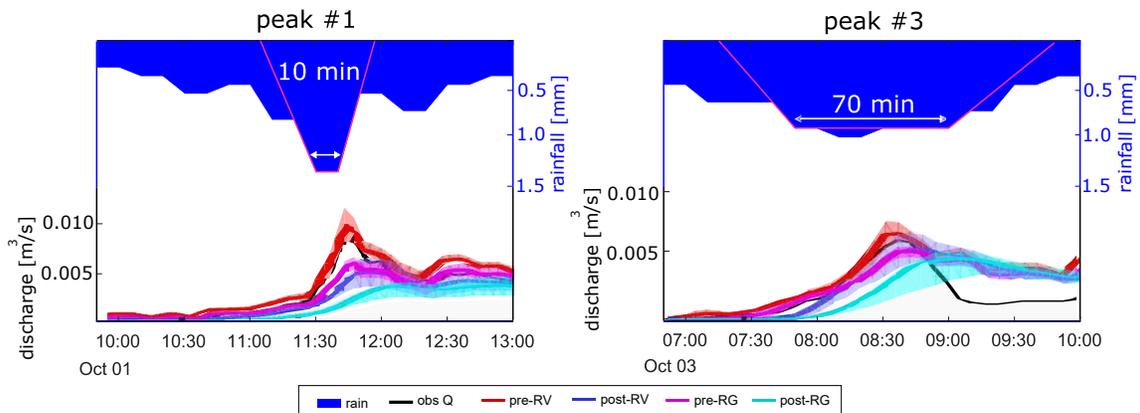


Figure 7: Comparing duration of peak rainfall intensity and its impact on peak magnitude reduction.

613 During peak 3 the duration of peak rainfall intensity exceeded the time to equilibrium for the catchment (the  
 614 time for the floodwave to propagate from the furthest point in the catchment). As a result the discharge peak  
 615 is delayed but not reduced in magnitude (with the catchment in an almost ‘steady-state’). This phenomenon is  
 616 independent of the magnitude of the rainfall and is driven by duration, which makes it possible for a prolonged  
 617 storm event to fill the entire catchment ‘immobile’ and ‘mobile’ surface storage, whilst experiencing no reduc-

tion in its magnitude, and thus removing the peak discharge reduction effect of NFM interventions (although not their delaying effect).

Note that peak 3 has a peak discharge less than half that of the largest storm in the study period, and is only used here to draw attention to the possible impact of such long-duration storms, irrespective of their probability of occurrence. In reality, it may be that the highest intensity storms rarely, or never assume such broad hyetograph shapes. Constraining this would require an Intensity-Duration-Frequency (IDF) study. Nonetheless, our findings and those of [Gao et al. \(2018\)](#), suggest that hyetograph characteristics can play an important role in modifying the effectiveness of NFM interventions.

## 6.4 NFM implications at catchment scale

Revegetation and gully blocking activities in our study catchments were motivated by, and designed for moorland restoration rather than for Natural Flood Management (NFM). It is unlikely that a flood management focus would have substantially changed the revegetation treatment; however, it likely would have changed the gully block designs/configurations, which were introduced to elevate water tables, trap sediment and prevent further gully erosion, rather than to attenuate flow. Given the difference in focus, it is perhaps surprising how successful the interventions have been in reducing and delaying peak discharge.

Further, our results show that the percentage reduction in peak discharge from a micro-catchment can vary widely (5%-90% in Figure 6a), and that this may be due to rainfall characteristics (Section 6.3). If so, the interventions studied here will be least effective in reducing micro-catchment peak discharge for long duration, 'frontal' rainfall, and more effective for shorter 'convective' rainfall, such as that responsible for the floods in nearby Glossop (August 2002) and Sheffield (June 2007).

It is also important to note that regardless of the duration of peak rainfall intensity, lag-times are always considerably increased following these interventions (minimum 10 minutes in RV and 30 minutes in RG). Consequently, even in the longest frontal rainstorms, these increased lag-times for upland headwater subcatchments can be utilised to attenuate the downstream flood wave, as the delayed delivery of water will allow flood peaks from other downstream subcatchments to pass first, before delivering water to catchments draining to communities at risk ([Pattison et al., 2014](#)). This, however, comes with the caveat that when alterations are primarily to flood wave timing rather than magnitude, it is possible that interventions synchronise sub-catchments rather than bringing them out-of-synch. To avoid such situations, catchment-scale models (such as [Metcalf et al. \(2018\)](#)) can be useful in testing different spatial patterns of NFM interventions and their impact on flood risk throughout a river network.

Nonetheless, possible susceptibility of these interventions to long duration, frontal rainfall suggests that: revegetation and gully-blocking of the design examined here, may not provide full protection against flooding in some cases. If so, this would be because the catchment's total surface storage (immobile and mobile) was not sufficiently increased by these interventions to reduce peak magnitude even in longest duration events. Note that the restoration interventions' main objective is to raise the water tables, which translates into reduced available subsurface storage. Therefore, to extend these interventions' effectiveness to long frontal rainstorms while remaining within the restoration objectives, the catchment's total 'surface' storage needs to be increased. We propose two approaches.

First, trade off immobile surface storage for more mobile surface storage, because in flood relevant storms, total immobile surface storage was insufficient to significantly reduce peak magnitude or increase lag-times. Instead, celerity reduction had a far greater effect. Thus making the blocks 'optimally' permeable could further increase surface roughness and reduce celerity. This would better utilise the mobile storage effect, which might not currently be maximally exploited. This can be done, for instance, by inserting pipes, or cutting holes into otherwise impermeable gully blocks (e.g. [Milledge et al. \(2015\)](#), [Metcalf et al. \(2017\)](#)), such that surface flow is partitioned into storage during times of peak flow. Here, for a given storm, optimal design would result in blocks that: fill to capacity, do so only at peak flow, and never overtop.

Second, increase the total immobile surface storage by creating more shallow open-water pools in other parts of the catchment as well as in channels and gullies, such that its volume is comparable to the large water volumes associated with flood relevant storms. Note that this may also increase the mobile surface storage by

667 further reducing surface flow celerities, but comes with the caveat that since total storm-water volumes scale  
668 with catchment area, full catchment-scale implementation may be costly/unfeasible.

## 669 **7 Summary and conclusions**

670 We have shown that a rainfall-runoff model can be inverted to corroborate findings and expand process under-  
671 standing associated with BACI observations. The extent to which such inversion provides useful information  
672 is critically dependent on uncertainties introduced by both observations (rainfall and discharge data) and the  
673 model parameters, as well as by model structure. We have sought to account for observational and parameter  
674 uncertainties within a limits of acceptability framework. However, model structural uncertainty remains and is  
675 difficult to quantify. For example, we have not accounted for the known depth-dependence of surface flow ve-  
676 locity (Holden et al., 2008) and celerity (McDonnell & Beven., 2014), which could result in under-estimation of  
677 the mobile surface storage in some parts of the catchment, and over-estimation in others; depending on whether  
678 the depth-averaged celerity is more or less than the true depth-dependent celerity in those parts. When calcu-  
679 lating the bulk mobile surface storage (as was the case in this study), it is conceivable that some of the errors  
680 associated with depth-averaging the celerity would cancel out due to smoothing that occurs by averaging over-  
681 and under-estimated values across different parts of the catchment. Additionally, our simple model structure  
682 cannot explicitly account for differences in gully block permeability between wooden and cobble blocks. Thus,  
683 the extent to which such model structural uncertainties could potentially have impacted our findings is not clear  
684 without further study.

685 However, confidence in our results comes from: (i) a large number of behavioural parameter-sets (>600)  
686 at all site-intervention combinations; (ii) good agreement between predicted and observed discharge for be-  
687 havioural models both in terms of overall fit (NSE) and peak characteristics (magnitude and timing); (iii) agree-  
688 ment between predicted and observed discharge that is retained with little degradation in performance metrics  
689 for a test period held back during calibration; and (iv) agreement with findings from the BACI experiment in  
690 terms of the distribution of changes in magnitude and timing of peak discharge.

691 In summary, we find that revegetation and gully blocking activities motivated by, and designed for moor-  
692 land restoration also provide significant additional catchment surface storage (immobile and mobile), which  
693 reduces peak discharge and increases lag-times, independent of storm size. Revegetation increases immobile  
694 surface storage through more evapotranspiration almost exclusively 'during' storms, and not through increased  
695 rainfall interception. Gully blocking extends the evapotranspirative losses to the inter-storm period through in-  
696 creasing the immobile storage capacity, but also reduces catchment average evapotranspiration rates. These two  
697 effects approximately offset one another, such that the overall impact of blocking on immobile surface storage  
698 is negligible when compared to revegetation alone. Importantly though, for flood relevant storms, the immobile  
699 storage plays only a secondary role to 'mobile' surface storage. In these storms, independent of intervention  
700 type, the key catchment change responsible for reducing peak discharge and delaying lag-times is increased  
701 surface roughness through its impact on surface flow celerity.

702 The dominant role of roughness on discharge attenuation means that hyetograph shape can strongly influence  
703 the interventions' effectiveness in reducing peak discharge. Long, broad, rainfall intensity peaks can lead to near  
704 steady-state discharge responses from these small catchments; in which case, reduced surface flow celerities still  
705 increase lag-times but they no longer reduce peak discharge. Although both are important, the latter has a more  
706 reliable impact on downstream discharge because lag-time increase in headwater catchments generally attenuate  
707 downstream discharge but can amplify it by synchronising subcatchment hydrographs. This highlights the  
708 importance of catchment-scale modelling to test the impact of intervention scenarios before implementation.  
709 Our pre- and post-intervention parameter-sets may provide useful constraints on efforts to parameterise such  
710 models.

711 Interventions focused more sharply on NFM could both reduce peak discharge and delay peak timing by  
712 increasing total catchment 'surface' storage. We propose two approaches: (1) trade-off 'immobile' surface  
713 storage for more 'mobile' surface storage by making the blocks 'optimally' permeable; (2) increase the total  
714 immobile surface storage by creating more shallow open-water pools in other parts of the catchment as well as  
715 in channels and gullies.

## 8 Acknowledgments

This research was conducted as part of the NERC PROTECT-NFM project (NE/R004587) with a project steering committee including representatives from: CH2M, Environment Agency, IUCN UK Peatland Programme, Moors for the Future Partnership, National Trust, Natural Resources Wales, Risk Management Solutions, SEPA, and United Utilities. The hydrological data were collected as part of the Making Space for Water projects funded by DEFRA and the UK Environment Agency. The work benefited from feedback from the wider PROTECT project team. The hydrological data used in this research are included in this paper: [Shuttleworth et al. \(2019\)](#).

## 9 Conflict of interest

No conflict of interest.

## References

- Acreman, M., & Holden, J. (2013). How wetlands affect floods. *Wetlands*, 33(5), 773-786. [1](#)
- Armstrong, A., Holden, J., Kay, P., Francis, B., Foulger, M., Gledhill, S. & Walker, A. (2010). The impact of peatland drain-blocking on dissolved organic carbon loss and discolouration of water; results from a national survey. *Journal of Hydrology*, 381(1-2), 112-120. [1](#)
- Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, 24(1), 43-69. [3](#)
- Beven, K., & Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes*, 6(3), 279-298. [4.1](#)
- Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1-2), 18-36. [4.3](#)
- Blazkova, S., & Beven, K. (2009). A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resources Research*, 45(12). [4.3](#)
- Beven, K. J. (2020). A history of the concept of time of concentration. *Hydrology and Earth System Sciences*, 24(5), 2655-2670. [3](#)
- Calder, I.R., Harding, R.J. & Rosier, P.T.W., 1983. An objective assessment of soil-moisture deficit models. *Journal of Hydrology*, 60(1-4), 329-355. [3](#)
- Campbell, D. I., & Williamson, J. L. (1997). Evaporation from a raised peat bog. *Journal of Hydrology*, 193(1-4), 142-160. [6.2.1](#)
- Ciach, G. J. (2003). Local random errors in tipping-bucket rain gauge measurements. *Journal of Atmospheric and Oceanic Technology*, 20(5), 752-759. [4.2](#)
- Cole, B., McMorrow, J., & Evans, M. (2014). Empirical modelling of vegetation abundance from airborne hyperspectral data for upland peatland restoration monitoring. *Remote Sensing*, 6(1), 716-739. [1](#)
- Dadson, S. J., Hall, J. W., Murgatroyd, A., Acreman, M., Bates, P., Beven, K., ... & O'Connell, E. (2017). A restatement of the natural science evidence concerning catchment-based 'natural' flood management in the UK. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2199), 20160706. [1](#)
- Daniels, S. M., Agnew, C. T., Allott, T. E. H. & Evans, M. G. (2008). Water table variability and runoff generation in an eroded peatland, South Pennines, UK. *Journal of Hydrology*, 361(1-2), 214-226. [4.3](#)

- 754 Dixon, S. D., Qassim, S. M., Rowson, J. G., Worrall, F., Evans, M. G., Boothroyd, I. M., & Bonn, A. (2014).  
755 Restoration effects on water table depths and CO<sub>2</sub> fluxes from climatically marginal blanket bog. *Biogeo-*  
756 *chemistry*, 118(1-3), 159-176. [1](#)
- 757 Dolan, T. J., Hermann, A. J., Bayley, S. E., & Zoltek Jr, J. (1984). Evapotranspiration of a Florida, USA,  
758 freshwater wetland. *Journal of Hydrology*, 74(3-4), 355-371. [6.2.1](#)
- 759 Drexler, J. Z., Snyder, R. L., Spano, D., & Paw U, K. T. (2004). A review of models and micrometeorological  
760 methods used to estimate wetland evapotranspiration. *Hydrological Processes*, 18(11), 2071-2101. [6.2.1](#)
- 761 Evans, M. G., Burt, T. P., Holden, J., & Adamson, J. K. (1999). Runoff generation and water table fluctuations  
762 in blanket peat: evidence from UK data spanning the dry summer of 1995. *Journal of Hydrology*, 221(3-4),  
763 141-160. [1](#)
- 764 Evans, M., Allott, T., Holden, J., Flitcroft, C., & Bonn, A. (2005). Understanding gully blocking in deep peat.  
765 *Derbyshire: Moors for the Future Report*, 4. [1](#)
- 766 Evans, M., Warburton, J., & Yang, J. (2006). Eroding blanket peat catchments: global and local implications of  
767 upland organic sediment budgets. *Geomorphology*, 79(1-2), 45-57. [1](#)
- 768 Evans, M., & Warburton, J. (2011). *Geomorphology of upland peat: erosion, form and landscape change*. John  
769 Wiley & Sons. [1](#)
- 770 Féral, L., Sauvageot, H., Castanet, L., Lemorton, J., Cornet, F. & Leconte, K., 2006. Large-scale modeling of  
771 rain fields from a rain cell deterministic model. *Radio Science*, 41(02), 1-21. [4.2](#)
- 772 Freer, J., Beven, K. J., & Peters, N. (2003). Multivariate seasonal period model rejection within the generalised  
773 likelihood uncertainty estimation procedure. *Calibration of watershed models*, 69, 88. [4.3](#)
- 774 Gao, J., Holden, J., & Kirkby, M. (2015). A distributed TOPMODEL for modelling impacts of land-cover  
775 change on river flow in upland peatland catchments. *Hydrological Processes*, 29(13), 2867-2879. [3](#)
- 776 Gao, J., Holden, J., & Kirkby, M. (2016). The impact of land-cover change on flood peaks in peatland basins.  
777 *Water Resources Research*, 52(5), 3477-3492. [3](#)
- 778 Gao, J., Holden, J., & Kirkby, M. (2017). Modelling impacts of agricultural practice on flood peaks in upland  
779 catchments: An application of the distributed TOPMODEL. *Hydrological Processes*, 31(23), 4206-4216. [3](#)
- 780 Gao, J., Kirkby, M., & Holden, J. (2018). The effect of interactions between rainfall patterns and land-cover  
781 change on flood peaks in upland peatlands. *Journal of Hydrology*, 567, 546-559. [6.3](#)
- 782 Grayson, R., Holden, J., & Rose, R. (2010). Long-term change in storm hydrographs in response to peatland  
783 vegetation change. *Journal of Hydrology*, 389(3-4), 336-343. [1](#)
- 784 Holden, J., & Burt, T. P. (2002). Infiltration, runoff and sediment production in blanket peat catchments: impli-  
785 cations of field rainfall simulation experiments. *Hydrological Processes*, 16(13), 2537-2557. [3](#), [4.3](#)
- 786 Holden, J., & Burt, T. P. (2003). Runoff production in blanket peat covered catchments. *Water Resources Re-*  
787 *search*, 39(7), 1191-1200. [4.3](#)
- 788 Holden, J., Chapman, P. J., & Labadz, J. C. (2004). Artificial drainage of peatlands: hydrological and hydro-  
789 chemical process and wetland restoration. *Progress in Physical Geography*, 28(1), 95-123. [1](#)
- 790 Holden, J. (2005). Peatland hydrology and carbon release: why small-scale process matters. *Philosophical*  
791 *Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 363(1837), 2891-  
792 2913. [1](#)
- 793 Holden, J., Kirkby, M. J., Lane, S. N., Milledge, D. G., Brookes, C. J., Holden, V., & McDonald, A. T. (2008).  
794 Overland flow velocity and roughness properties in peatlands. *Water Resources Research*, 44(6). [3](#), [7](#)

- 795 Holden, J. (2009). Peatland hydrology and carbon release. In Drivers of Environmental Change in Uplands;  
796 Bonn, A., Allott, T., Hubacek, K., Stewart, J. Routledge 2009 Jan 13. Eds, 113-134. 4.3
- 797 Jones, M., Fowler, H. J., Kilsby, C. G. & Blenkinsop, S. (2010). An updated regional frequency analysis of  
798 United Kingdom extreme rainfall 1961–2009. Proceedings of BHS Conference 2010, 110-118. 5.2
- 799 Keeland, B. D., Dowd, J. F., & Hardegree, W. S. (1997). Use of inexpensive pressure transducers for measuring  
800 water levels in wells. Wetlands Ecology and Management, 5(2), 121-129. 3, 4.2
- 801 Kirkby, M. J. (1997). TOPMODEL: A personal view. Hydrological processes, 11(9), 1087-1097. 3
- 802 Lane, S. N., Brookes, C. J., Kirkby, M. J., & Holden, J. (2004). A network-index-based version of TOPMODEL  
803 for use with high-resolution digital topographic data. Hydrological processes, 18(1), 191-201. 3
- 804 Lane, S. N., & Milledge, D. G. (2013). Impacts of upland open drains upon runoff generation: a numerical  
805 assessment of catchment-scale impacts. Hydrological Processes, 27(12), 1701-1726. 3
- 806 Lindsay, R., Charman, D. J., Everingham, F., O'reilly, R. M., Palmer, M. A., Rowell, T. A., & Stroud, D. A.  
807 (1988). The flow country: the peatlands of Caithness and Sutherland. Joint Nature Conservation Committee.  
808 1
- 809 McDonnell, J. J., & Beven, K. (2014). Debates—The future of hydrological sciences: A (common) path for-  
810 ward? A call to action aimed at understanding velocities, celerities and residence time distributions of the  
811 headwater hydrograph. Water Resources Research, 50(6), 5342-5350. 3, 6.2.2, 7
- 812 McMillan, H., Krueger, T., & Freer, J. (2012). Benchmarking observational uncertainties for hydrology: rainfall,  
813 river discharge and water quality. Hydrological Processes, 26(26), 4078-4111. 4.2
- 814 Metcalfe, P., Beven, K., & Freer, J. (2015). dynamic TOPMODEL: A new implementation in R and its sensitiv-  
815 ity to time and space steps. Environmental Modelling & Software, 72, 155-172. 3
- 816 Metcalfe, P., Beven, K., Hankin, B. & Lamb, R. (2017). A modelling framework for evaluation of the hydrologi-  
817 cal impacts of nature-based approaches to flood risk management, with application to in-channel interventions  
818 across a 29-km2 scale catchment in the United Kingdom. Hydrological Processes, 31(9), 1734-1748. 6.2.2,  
819 6.4
- 820 Metcalfe, P., Beven, K., Hankin, B., & Lamb, R. (2018). A new method, with application, for analysis of the  
821 impacts on flood risk of widely distributed enhanced hillslope storage. Hydrology and Earth System Sciences,  
822 22(4), 2589-2605. 6.4
- 823 Milledge, D., Odoni, N., Allott, T., Evans, M., Pilkington, M., & Walker, J. (2015). Annex 6. Flood risk  
824 modelling. In: Restoration of Blanket bogs; flood risk reduction and other ecosystem benefits; Pilkington,  
825 M., Walker, J., Maskill, R., Allott, T. and Evans, M. Moors for the Future Partnership, Edale. 6.2.2, 6.4
- 826 Mohamed, Y. A., Bastiaanssen, W. G., Savenije, H. H. G., Van den Hurk, B. J. J. M., & Finlayson, C. M.  
827 (2012). Wetland versus open water evaporation: An analysis and literature review. Physics and Chemistry of  
828 the Earth, Parts A/B/C, 47, 114-121. 6.2.1
- 829 Parry, L. E., Holden, J., & Chapman, P. J. (2014). Restoration of blanket peatlands. Journal of Environmental  
830 Management, 133, 193-205. 1
- 831 Pattison, I., Lane, S. N., Hardy, R. J., & Reaney, S. M. (2014). The role of tributary relative timing and sequenc-  
832 ing in controlling large floods. Water Resources Research, 50(7), 5444-5458. 6.4
- 833 Petersen-Øverleir, A., & Reitan, T. (2005). Uncertainty in flood discharges from urban and small rural catch-  
834 ments due to inaccurate head measurement. Hydrology Research, 36(3), 245-257. 4.2

- 835 Pollock, M. D., O'Donnell, G., Quinn, P., Dutton, M., Black, A., Wilkinson, M. E., & Kilsby, C. G. (2018).  
836 Quantifying and mitigating wind-induced undercatch in rainfall measurements. *Water Resources Research*,  
837 54(6), 3863-3875. 4.2
- 838 Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty  
839 in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*,  
840 46(5). 4.2
- 841 Rodda, J. C. (1967). The rainfall measurement problem. *Proc. IAHS Gen. Ass. Bern, IAHS Pub, (78)*, 215-231.  
842 4.2
- 843 Rothwell, J. J., Evans, M. G., & Allott, T. E. H. (2007). Lead contamination of fluvial sediments in an eroding  
844 blanket peat catchment. *Applied Geochemistry*, 22(2), 446-459. 1
- 845 Sauvageot, H., Mesnard, F. & Tenório, R. S. (1999). The relation between the area-average rain rate and the rain  
846 cell size distribution parameters. *Journal of Atmospheric Sciences*, 56(1), 57-70. 4.2
- 847 Savenije, H. H. (2004). The importance of interception and why we should delete the term evapotranspiration  
848 from our vocabulary. *Hydrological processes*, 18(8), 1507-1511. 6.2.1
- 849 Shen, J. (1981). Discharge characteristics of triangular-notch thin-plate weirs (No. 1617). United States Depart-  
850 ment of the Interior, Geological Survey. 4.2, 4.2
- 851 Shuttleworth, E. L., Evans, M. G., Hutchinson, S. M., & Rothwell, J. J. (2015). Peatland restoration: controls on  
852 sediment production and reductions in carbon and pollutant export. *Earth Surface Processes and Landforms*,  
853 40(4), 459-472. 1
- 854 Shuttleworth, E. L., Evans, M. G., Pilkington, M., Spencer, T., Walker, J., Milledge, D., & Allott, T. E. (2019).  
855 Restoration of blanket peat moorland delays stormflow from hillslopes and reduces peak discharge. *Journal*  
856 *of Hydrology X*, 2, 100006. 1, 2, 1, 5.1, 5.4, 6.1, 6.1.1, 6.2.2, 8
- 857 Stocker, B. D., Spahni, R., & Joos, F. (2014). DYPTOP: a cost-efficient TOPMODEL implementation to simu-  
858 late sub-grid spatio-temporal dynamics of global wetlands and peatlands. *Geoscientific Model Development*,  
859 7(6), 3089-3110. 3
- 860 Van Straten, G. T., & Keesman, K. J. (1991). Uncertainty propagation and speculation in projective forecasts of  
861 environmental change: A lake-eutrophication example. *Journal of Forecasting*, 10(1-2), 163-190. 4.3
- 862 Vrugt, J. A., & Beven, K. J. (2018). Embracing equifinality with efficiency: Limits of Acceptability sampling  
863 using the DREAM (LOA) algorithm. *Journal of Hydrology*, 559, 954-971. 4.2, 4.2
- 864 Wallage, Z. E., Holden, J., & McDonald, A. T. (2006). Drain blocking: an effective treatment for reducing  
865 dissolved organic carbon loss and water discolouration in a drained peatland. *Science of the total environment*,  
866 367(2-3), 811-821. 1
- 867 Xu, J., Morris, P. J., Liu, J., & Holden, J. (2018). PEATMAP: Refining estimates of global peatland distribution  
868 based on a meta-analysis. *Catena*, 160, 134-140. 1