A new method, with application, for analysis of the impacts on flood risk of widely-distributed enhanced hillslope storage

Peter Metcalfe[1*], Keith Beven[1,2], Barry Hankin[3] and Rob Lamb[4,1]

Correspondence to p.metcalfe@lancs.ac.uk

[1] Lancaster Environment Centre, Lancaster University, Lancaster, LA14YQ, UK
[2] Department of Earth Sciences, Uppsala University, Uppsala 75263, Sweden
[3] JBA Consulting, Sankey Street, Warrington, Cheshire WA1 1NN, UK
[4] JBA Trust, South Barn, Broughton Hall, Skipton, North Yorkshire, BD23 3AE, UK

Abstract. Enhanced hillslope storage is utilised in "natural" flood management in order to retain overland storm runoff and to reduce connectivity between fast surface flow pathways and the channel. Examples include excavated ponds, deepened or bunded accumulation areas, and gullies and ephemeral channels blocked with wooden barriers or debris dams.

The performance of large, distributed networks of such measures is poorly understood. Extensive schemes can potentially retain large quantities of runoff, but there are indications that much of their effectiveness can be attributed to desynchronisation of subcatchment flood waves. Inappropriately-sited measures may therefore increase, rather than mitigate, flood risk. Fully-distributed hydrodynamic models have been applied in limited studies but introduce significant computational complexity. The longer run-times of such models also restrict their use for uncertainty estimation or evaluation of the many potential configurations and storm sequences that may influence the timings and magnitudes of flood waves.

Here a simplified overland flow routing module and semi-distributed representation of enhanced hillslope storage is developed. It is applied to the headwaters of a large rural catchment in Cumbria, UK, where the use of an extensive network of storage features is proposed as a flood mitigation strategy. The models were run within a Monte Carlo framework against data for a two-month period of extreme flood events that caused significant damage in areas downstream. Acceptable realisations and likelihood weightings were identified using the GLUE uncertainty estimation framework. Behavioural realisations were rerun against the catchment model modified with the addition of the hillslope storage. Three different drainage rate parameters were applied across the network of hillslope storage.

The study demonstrates that schemes comprising widely-distributed hillslope storage can be modelled effectively within such a reduced complexity framework. It shows the importance of drainage rates from storage features while operating through a sequence of events. We discuss limitations in the simplified representation of overland flow routing and representation and storage, and how this could be improved using experimental evidence. We suggest ways in which features could be grouped more strategically and thus improve the performance of such schemes.

1 Introduction

A catchment-based approach to flood risk management is becoming widely adopted (Werritty, 2006; Dadson et al., 2017). Its principle is that storm runoff can be managed most effectively with a combination of catchment-scale measures and downstream flood defences (Lane, 2017). An example of this approach, often referred to as Natural Flood Management (NFM) or Working with Natural Processes (WWNP: EA, 2014, Hankin et al., 2017) utilises soft-engineered structures and interventions that both utilise and enhance the natural processes within the catchment in order to provide flood resilience (Calder and Alywood, 2006; SEPA, 2016; Lane, 2017). It is argued that NFM is a low-cost, scalable, approach that, in addition to improved flood resilience, can yield considerable benefits in terms of stakeholder engagement (Lane et al., 2011) and improved ecosystem services (Iacob, et al., 2014).

A wide range of interventions are employed in NFM. These are reviewed by, amongst others, Quinn et al. (2013), EA (2014), SEPA (2016), Dadson et al., (2017) and Lane (2017). Many of these are intended to increase the capacity of hillslopes to retain fast overland runoff and release it slowly, thus reducing its contribution to the rising limb of the storm hydrograph. Measures to achieve this encompass excavated ponds, deepened or bunded existing hollows, or wooden barriers or debris dams in ephemeral channels.

The term Runoff Attenuation Feature (RAF) has been used to refer to any structure to retain storm runoff, whether in channel (online) or on the hillslope (offline) (e.g. Barber and Quinn, 2012; Nicholson et al., 2012; Quinn et al., 2013). Here the term "enhanced hillslope storage" (EHS) is used to disambiguate online and offline approaches, and will relate to any measure intended to increase the capacity of the hillslopes to retain storm runoff. EHS was implemented in the Belford Burn catchment as wooden structures and earth bunds placed across overland flow pathways identified from topographic analysis (Nicholson et al., 2012; Wilkinson et al., 2010a). In Pickering Vale, North Yorkshire (Nisbet et al., 2011), Pontbren (Wheater et al., 2008; Nisbet and Page, 2016) and Holnicote, Devon (National Trust, 2015) blocking of moorland gullies added storage to upper parts of the catchments.

Despite the increasing use of EHS, there have been few relatively attempts to model its effects on the storm response at a catchment scale, and to quantify the requirements of a scheme to meet a certain level of flood risk mitigation. The expectation is currently that NFM will have an impact on small to moderate scale events. However, to be a practical strategy, it will be required to operate effectively across extreme events, but there is as yet little evidence of what will be required to have an effective impact in these conditions.

Ghimire et al. (2014) applied a fully-distributed hydrodynamic model (TUFLOW) to simulate a potential single hillslope pond of capacity 27000 m³ within the 74 km² Tarland Burn catchment in Aberdeenshire. This showed that the intervention could reduce peak flows by 9 % for an event of the scale of the median annual flood (known as QMED). Using the fully-distributed JFLOW model (Lamb et al., 2009, Environment Agency, 2013), Hankin et al. (2016, 2017) simulated the effects of multiple hillslope ponds across three catchments in Cumbria This showed the relative impacts on the hydrograph, in terms of downstream benefits or damages avoided. The model was driven by multiple rainfall event sets incorporating spatial joint

probabilities observed in the extremes from time series of observed rainfalls around the catchment (Lamb et al., 2010; Keef et al., 2013). Nicholson et al. (2012) and Wilkinson et al. (2010b) used a lumped representation in the Belford Burn catchment, whereby the storage requirement of the scheme was estimated from the reduction required to prevent damage in the hydrograph of the smallest storm that caused flooding. This indicated that 20000 m³ of additional storage would have

5   sufficed to prevent damage. However, the interactions of storage and the complex routing of flood waves may mean that the theoretical static storage of a scheme could be substantially underutilised, even during large events (Metcalfe at al., 2017).

The capacity of individual storage features is in the UK constrained by legislation that limits their size to 10,000 m³, above which significant legal responsibilities are imposed (Wilkinson et al., 2010b; Ghimire et al., 2014). Ghimire (2014) suggested that a network of ponds just under the critical capacity could be used to replace the single large pond first

10   modelled. Analysis of the effects of just one of these 9500 m³ ponds indicated that it could reduce the peak of the QMED event by 2.5 %. A pond of 5200 m³ would reduce the peak of the same storm by 1 %. The storage required for a significant mitigating effect on storm flows is much greater, however. Metcalfe et al. (2017) showed that within a 29 km² catchment 168000 m³ of hydrodynamic storage provided by a single large in-channel feature in the middle reaches would just have prevent flooding in a 1 in 75 year event. A scheme of a realistic scale could therefore involve installation of extremely large

15   numbers of features on the hillslope, although the quantity required might be reduced by applying other measures such as tree-planting and online interventions.

Representing such extensive schemes within computer models will be challenging. Physically-based models in general employ a gridded digital terrain model (DTM) to represent the surface. Features can be introduced into the landscape representation by raising cells on their boundary to represent wooden walls or bunds. For example, Hankin et al. (2016,

20   2017) simulated hillslope ponds by deepening the appropriate cells in a 2 m DEM by 1 m. This was also the approach of Ghimire et al. (2014), this time using a 5 m DEM. Hydraulic models of individual structures could be achieved by applying analogies to engineered interventions whose characteristics are understood. Metcalfe et al. (2017) modelled the effects of the wooden channel barriers by analogy with underflow sluices employed in in irrigation schemes. Chow (1959) describes analytical storage-discharge relationships for such structures that utilise empirically-determined parameters. Dams

25   constructed of spaced timber members could be modelled by the Kirschmer-Mosonyi formula for flow through trash screens (Mosonyi, 1966), such as used in the intakes to power plants and waste-water treatment works. Overflow of these structures that run out of capacity during the course of a storm event could be modelled analytically as though across a weir, another well-studied structure.

Drainage from structures to retain runoff can be by infiltration, direct evaporation from the water surface, or through

30   permeable or "leaky" walls. It will, however, be difficult in the models mentioned to account dynamically for the drainage pathways out of large numbers of hillslope storage features. Areas will be unable to drain down during the course of a simulation, and thus the modelling approach will have limitations when applied to multiple storm events. In these cases, recovery of storage capacity during recession periods could have a significant impact on the effectiveness of a scheme through multiple events (Metcalfe et al., 2017).

Much of the effectiveness of natural and distributed flood management schemes has been attributed to their ability to desynchronise subcatchment flood waves (Thomas and Nisbet, 2007; Blanc et al., 2012). Pattison et al. (2014) examined the interacting effects of the Eden subcatchment flood waves using data for a large flood event in Carlisle in 2005, and concluded that their timing and magnitude predicted the majority of the variance in modelled downstream flood peaks. It will therefore be necessary to design such installations with some care. Slowing runoff that would have contributed to the hydrograph *before* the peak could have the effect of increasing the peak magnitude. This is called the synchronicity problem. Peak timings vary, however, with the pattern and timing of rainfalls and antecedent wetness in the catchment, and also in the way in which the hydrographs from different subcatchments interact. Thus, it will be necessary to test the sensitivity of a design before implementation using an appropriate model of runoff generation and mitigation measures. This might, in itself, require many model runs that reflect different event characteristics and patterns, and storage and drainage characteristics of features. Given the immense number of combinations of potential configurations and varieties of storm events, a pragmatic approach to evaluating the impacts of NFM would be "experimental" modelling whereby many possible realisations of the catchment model and event sets are generated (Hankin et al., 2016, 2017). A computationally efficient runoff model and representation of hillslope storage would allow such an approach in reasonable timescales. This will necessarily involve some simplification of the hydraulic behaviour, as a fully hydrodynamic treatment will introduce complexity, additional parameters, and much increased run-times.

There is significant uncertainty in predictions of the spatial distribution and quantities of runoff (Beven, 2006, 2012). Attempts to assess the effectiveness of NFM will compare predictions for unaltered and modified catchments and introduce even further uncertainty. Uncertainty estimation and sensitivity analysis can provide a more realistic assessment of the reliability of predictions of the impacts of NFM (Hankin et al., 2017). These techniques will, however, also require the generation of thousands or even millions of model realisations in which parameters are sampled from prior distributions of realistic values. With continued growth in computing power it is now feasible to run the fully-distributed JFLOW model over 750 km² with a 2 m resolution grid (e.g. 175 million cells) in approximately real-time (Hankin et al., 2017). This will still require a significant commitment in time and computing power to produce more than a few scores of realisations. Thus, while high-performance computing may be fast enough for assessment for a selection of features and hillslope properties, it may still be inadequate for uncertainty analysis across larger catchments and wide-scale NFM schemes comprising the large number of features required to meet the storage requirements to significantly attenuate real flood events.

In 2016 the Life IP Natural Course project was undertaken by Lancaster University and JBA Consulting for the UK Rivers Trust. Its aim was provide better understanding of how NFM could be applied strategically to the headwaters of three catchments in Cumbria, UK (Hankin et al., 2016, 2017). Measures considered included large-scale tree-planting in order to increase evaporation losses and improve soil structure, and restoration of peat and heath to increase surface roughness. Measures to provide enhanced hillslope storage were also considered. The interventions were modelled separately, allowing the effects of each to be discerned.

Given the uncertainties in both predictions of storm runoff and the impacts of NFM, a key objective of this project was to develop an approach that could allow the modelling results to be presented to the client alongside realistic estimates of their uncertainty. This required a model that could simulate the storm runoff with sufficient computational efficiency to generate the thousands of runs required for uncertainty analysis.

Previous work has shown the importance of the drainage characteristics of features that retain runoff whilst operating across a series of closely-spaced events (Metcalfe at al., 2017), as a degree of "leakiness" is required to allow the scheme to recover capacity between events. Another objective was therefore to develop an approach that could reflect drainage from "leaky" EHS and could be applied to the thousands of features required, but again without excessive computational cost. A period of three significant storms in November and December 2015, in which the catchments were badly affected, was chosen as the basis for the study. This paper documents the modelling approach that was developed and describes how it was applied in the largest of these catchments, the 223 km^2 Upper Eden.

2   Material and methods

2.1   Runoff modelling

A new implementation of the semi-distributed Dynamic TOPMODEL (Metcalfe et al., 2015) was employed to predict storm runoff in both unaltered catchments models and those with the addition of hillslope storage. The original version (Beven and Freer, 2001a) has been applied in many studies (e.g. Liu et al., 2009; Page et al., 2007). In conjunction with a spatially explicit hydraulic channel routing module, the later implementation was used to evaluate the effect on flood risk of an NFM-style scheme in an agricultural catchment in North Yorkshire, UK (Metcalfe et al., 2017).

The model extends TOPMODEL (Beven and Kirkby, 1979). The basic approach in both models is the aggregation of "similar" landscape areas into Hydrological Response Units (HRUs), that are treated during the course of a simulation as having identical hydrological responses. The units may be of arbitrary size and not necessarily spatially contiguous, although they, along with their internal states, can be mapped back into space. This "discretisation" approach significantly reduces the complexity of the landscape model whilst retaining hydrological connectivity of the hillslope. The model can therefore be run extremely quickly, and this approach therefore helped meet the objective of the LifeIP project to evaluate NFM impacts within an uncertainty framework.

In TOPMODEL the catchment classification is strictly according to the Topographic Wetness Index (TWI), a measure of the tendency of a point on the hillslope to become saturated. An improved subsurface routing algorithm introduced in Dynamic TOPMODEL allows a more flexible approach to aggregation of catchment areas, and any characteristic or combination of characteristics may be used to identify a HRU. A routing matrix is developed from the surface topography, taken as an approximation to the hydraulic gradient, and applied to route subsurface runoff downslope through the units, with a kinematic wave relationship between specific discharge and storage deficit.

Model parameters are shown in Table 1.

Once a HRU discretisation has been defined the model can be run against rainfall and potential evapotranspiration (E_p) data for a specified time period. For that period, it produces a time series of simulated discharges at the catchment outlet and of the internal states of the HRUs. These include storage deficit, saturation excess (surface) and root zone storage. In order to reflect spatial variability, distinct rainfall and / or E_p records may be applied to the HRUs.

2.2   Overland flow routing

Hunter et al. (2007) suggest that in some situations simplified, but physically-based, surface flow models can perform as well as fully hydrodynamic formulations. In TOPMODEL and the first version of Dynamic TOPMODEL (Beven and Freer, 2001a) a network width approach was taken to routing surface flow (see Beven, 2012). In the implementation described by Metcalfe et al. (2015) semi-distributed, storage based surface flow routing was introduced. This uses a routing scheme similar to that applied to the subsurface. Saturation excess from upslope HRUs is routed to downslope units by a surface flow distribution matrix $\boldsymbol{W_{of}}$, again derived from the surface topography:

$$\boldsymbol{W_{of}} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ r_1 & p_{11} & \cdots & p_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ r_n & p_{n1} & \cdots & p_{nn} \end{pmatrix} \qquad r_i + \sum_{j=1}^{n} p_{ij} = 1 \qquad\qquad (1)$$

Each row in this matrix gives the proportions of the corresponding unit's flow that is directed to other units. For example, p_{ij} is the proportion of unit i's flow that is directed to unit j, and p_{ii} is the proportion that remains within unit i. The vector $\boldsymbol{r}$ represents a lumped "river" unit such that r_j is the proportion of downslope flux entering the channel network from unit number j. With an extended matrix a multi-reach river unit can also be defined. The matrix approximates transfer of flux between the different landscape units, and thence into the channel unit(s), by averaging the intercell slopes of the elevation raster between cells falling into each of the HRU categories.

An assumption of a linear storage-discharge relationship is now made, whereby the downslope discharge overland through a unit contour within a unit is proportional to depth of flow d (m). This implies a uniform velocity profile, so that the specific discharge is $q_{out} = v_{of}d$, with v_{of} (m h^{-1}) the mean overland wave velocity within that unit. It can be shown (Metcalfe et al., 2015) that this leads to a coupled series of ordinary differential equations for the vector of discharges $\boldsymbol{q_{out}}$ from all of the HRUs:

$$\boldsymbol{q_{out}} = \boldsymbol{Bd} \qquad\qquad (2)$$

where $\boldsymbol{A} = diag\left(\frac{1}{a}\right)\boldsymbol{W_{of}^T}diag(\boldsymbol{a})$, $\boldsymbol{B} = \boldsymbol{v_{of}}^T(\boldsymbol{A} - \boldsymbol{I})$, $\boldsymbol{d}$ the vector of average water depths and $\boldsymbol{a}$ the vector of areas of each of the HRUs. This system can be solved analytically by the so-called Eigenvalue method (Dummit, 2012). During a simulation storage distributed downslope is calculated to the end of the time interval, and any directed to the channel is routed to the outlet using the network width approach. Surface excess storages thus redistributed are then added to the

rainfall input for those units in the next time step. This approach allows for possible re-infiltration as "run on" if there is a soil moisture deficit in downslope units.

2.3   Modelling of enhanced hillslope storage

In order to assess the effects on the storm runoff of additional hillslope storage, simulations with identical parameters and inputs, but using representations of unaltered and modified hillslope, are undertaken and the respective results compared.

The more flexible discretisation approach introduced in Dynamic TOPMODEL allows a highly efficient way of modelling spatially-distributed hillslope storage as areas identified with interventions can be lumped into one or more HRUs. In model realisations reflecting unaltered catchments these units behave identically to surrounding landscape areas. To then simulate the effect of applying EHS, the surface routing through these units can be altered to reflect their reduced connectivity with the hillslope. Intervention areas may be grouped into multiple units according to their characteristics. These could include, for example, height and permeability, position on the hillslope, upslope area, slope and proximity to the channel. Even if many such aspects are identified and result in multiple groupings, this approach will substantially reduce computational overheads against a fully-spatially distributed representation.

Drainage is an important aspect of hillslope storage. If operational the ground making likely to be saturated and little storage will infiltrate to the subsurface. There will be evapotranspiration losses, however, which will occur at the maximum potential rate from the open surface. Direct drainage from a bunded area or dam could be applied via a pipe which, ignoring friction, gives a dependence of discharge on the square of the hydrostatic head. Leaky wooden structures or permeable bunds may instead drain as though through a porous medium. In this case it be more realistic to relate output discharge linearly with the head (Beven, 2012). A structure could also be impermeable so that storage not removed by evapotranspiration is lost downslope only when the features starts overflowing.

In deepened hollows and ground scrapes any overflow in excess of the storage capacity will leave the feature in a similar direction and velocity across the unmodified hillslope. In both leaky and impermeable structures, assuming they are designed to maximise interception by following the local contours, the directions are again likely to be similar. The flow rates will be more dependent on its material and construction and can, for example, be simulated by a non-contracted broad-crested Weir equation (Chow, 1959; Brater and King, 1976). The semi-distributed surface routing algorithm outlined in Section 2.2 is now applied in order to model catchments where hillslope storage has been added.by any of the methods described. The units representing storage are modified so that smaller proportions of their downslope surface flow are directed to other units than for the unaltered landscape representation, and thus features fill when there is a net input of runoff. Subsurface routing beneath the structure is unaltered.

This change in the rate of storage draining out areas due to emplacement of features with permeable walls can be emulated by reducing the effective overland flow velocity in the corresponding units. A leakiness" factor Λ is defined as the proportional reduction in effective downslope flow velocities from an aggregated storage unit with respect to the unaltered

hillslope. Each unit in effect behaves for the duration of a time step as a linear store with mean residence time $T_{res}(\Lambda, v_{of,i}) = \frac{1}{\Lambda v_{of,i}}$ where $v_{of,i}$ is the overland flow velocity for the unaltered hillslope representation.

A modified surface distribution matrix $\boldsymbol{W_{of}}$ can also be used to reflect changes in the directions of downslope flow due to the presence of the storage features. For example, a downslope pond drained by a pipe the drainage direction is likely to be

5   towards the nearest channel. In this case, for unit i draining to reach j, the element r_{ij} of $\boldsymbol{W_{of}}$ is set to unity and all the other units set to zero.

For a roughly rectangular area, storage could be considered as proportional to water depth, but other storage – depth relationships could be specified for more realistic morphologies, for example in blocked gullies. A parameter ex_{max} [L] can be introduced to control the maximum storage; if $ex > ex_{max}$ the structure starts to overflow. An additional overflow matrix

10 $\boldsymbol{W_{over}}$ can be defined to direct excess water out of the lumped features. This matrix is likely to be identical to the unaltered surface distribution matrix, $\boldsymbol{W_{surf}}$ but could be altered, for example to emulate an overflow channel.

2.4   Uncertainty estimation framework

It has been observed that different model parameterisations can lead to similar results which meet some type of "acceptability" criteria when evaluated against observational data. This has been termed equifinality (Beven, 2006). The

15 approach taken by the project is a form of the Generalised Likelihood Uncertainty Estimation methodology (GLUE: Beven and Binley 1992; Beven and Binley, 2014). It accepts the possibility of many model realisations that can fit the observed behaviour. GLUE has been employed in many studies (e.g. Beven and Freer, 2001b; Blazkova and Beven, 2004; Liu et al., 2009).

Many realisations are produced using Monte Carlo approach, whereby a large number of model parameters are sampled

20 randomly from one or more prior distributions and are then applied in turn to the system model. For each a simulation is performed across the desired period. Each output is then scored with a likelihood or weighting function calculated from performance metrics against observational data and other criteria. This is then normalised to produce a weighting factor for the corresponding realisation.

In general only those simulations achieving an acceptability threshold are retained, yielding a "behavioural" parameter set,

25 giving a posterior distribution of likelihoods. A triangular weighting function is common (Beven and Freer, 2001b; Beven, 2006; Blazkova and Beven, 2009; Liu et al., 2009). whereby the value is unity at the likeliest value of a predefined acceptability interval, zero at either limit, and linearly interpolated between these points. A trapezoidal form (e.g. Blazkova and Beven, 2009) could also be applied, where the value is unity between two threshold values and linearly interpolated outside these to the acceptability criteria. All of the predicted values of the observables must lie within acceptable ranges or

30 the realisation is rejected, even if the overall sum of weights is non-zero.

An approach to incorporate this uncertainty framework into evaluation of NFM interventions is shown in Figure 1. Weighted behavioural model realisations for the unaltered catchment model are obtained through the procedure described. One or more

modifications due to the applications of interventions are applied to the catchment model, and each of the base realisations rerun against this altered model. For each altered realisation the weighting score is carried through from the baseline case. If there is available information on the likelihoods of the effects of the interventions, a score for the modified realisation can be calculated by combining it with the weighting of the associated behavioural model.

5 **[INSERT FIGURE 1]**

3 Study catchment and storm events

The Eden headwaters in Cumbria, UK (Figure 2). have a drainage area of 223 km² above Great Musgrave Bridge (2.363234 W, 54.5126 N). The catchment is 55.4 % acid, improved or rough grassland and 36.0 % bog, scrub or heath. Bedrock geology is Permian and Triassic sandstones lain on Carboniferous limestones and there is some influence of groundwater

10 pathways (Ockenden and Chappell, 2010). Tree cover is minimal, comprising just 2.5 % of its area, although there has been significant recent tree planting that is not yet thought to have had any major effect on flood peaks. Overall annual rainfall is in the region of 1200 mm, but there is a strong synoptic and orographic influence (EA, 2009).

There are two UK Environment Agency flow gauges in the catchments (see Figure 2). Gauge number 76806 is located at Great Musgrave Bridge and was installed in July 2000. It is of a velocity- area type and of the form of a shallow V weir 30 m

15 in width. Gauge number 76014 is located near Kirkby Stephen (-2.351097 W, 54.482588 N) and drains an area of 69.4 km². It dates from 1971 and is a compound broad-crested weir, also using a velocity-area method area for flow estimation.

The early autumn of 2015 was unusually dry and soil water deficits were in October more than 10 mm greater than the long term average (Marsh et al., 2016). In November a south-westerly "atmospheric river" became established that brought warm moisture-laden air from the subtropics. A period of exceptional rainfall events followed that included Storms Abigail (15th-

20 16th November), Barney (18th November) and Desmond (5th-6th December). This final extra-tropical cyclone caused significant damage and over 2000 homes were flooded in Carlisle near the catchment outlet. A record 1680 m³ s^{-1} discharge was recorded at 9am on December 6th at Sheepmount Weir (54.905332 N, 2.952091 W) in Carlisle. The town of Appleby (54.578719 N, 2.488839 W), was also badly affected by this event; a peak discharge of 372 m³ s^{-1} was recorded at 18:00 on the 5th of December at the Great Musgrave Bridge gauge a few kilometres upstream. Matthews et al. (2018) provide a fuller

25 description of Storm Desmond and its hydrological extremes.

 [INSERT FIGURE 2]

The study period runs from close to the end of September 2015, when the soil moisture deficit was at its peak, to the recession period of Storm Desmond in early December. Processed 15 minute time series of rated discharges were obtained from the EA for gauge 76806. Tipping-bucket recorder (TBR) rainfall data at 15 minute intervals were also provided by the

30 EA and a set of these gauges lying within 10 km of the catchment was identified. Given the extreme rainfall, some gauges went off-line during the storms and were removed from the set, leaving four gauges with complete records over the period. A

rainfall record was interpolated from this and met the water balance to within 5 %. This was then applied to the entire catchment area.

Although significant events, Storms Abigail and Barney did not cause damaging flooding in this or downstream areas of the catchment. This suggests that a reduction of the 7.0 mm h^{-1} peak of Storm Desmond to that of Abigail, at 2.4 mm h^{-1}, the larger of these storms, would be a successful outcome of any NFM intervention. This corresponds to a reduction of around 4.6 mm h^{-1} or 65 %. It should be borne in mind the potential for large degree of uncertainty in the rating of these discharges, particularly at the extreme levels seen during Storm Desmond.

3.1   Model setup

An initial screening stage was undertaken to identify potential sites for enhancement of hillslope storage. Rainfall from a designed event of 30 year return period was routed to the catchment outlet with JFLOW. Areas that accumulated significant water depths, such as natural depressions, flow pathways or small channels, were tagged as suitable candidates for enhanced storage (see Figure 3). Their areas were then constrained in size to between 100 and 5000 m² and those within 2 m of roads and buildings excluded. This yielded 4500 distinct sites of average area 506 m², occupying 4.0 % of the catchment, leading to a potential static storage of just over 8 million m³.

[INSERT FIGURE 3]

A Dynamic TOPMODEL "discretisation" was now undertaken. The catchment was divided into 8 response units according to the topographic wetness index, and hillslope areas identified as potential sites for EHS were associated with an additional single response unit within the catchment. This HRU classification overrode any underlying classification determined from the TWI. This unit was thenceforth treated as though a single aggregated feature bunded or dammed by a "leaky" barrier, 1 m in height with upslope sides open to receive surface runoff.

Overflow was directed over the top of the barrier in the same direction as the original distributions given by the surface flow weighting matrix. The specific overflow per unit length along the top side of the feature was calculated as though for a broad-crested weir of width 50cm. Discharge coefficients are taken from the appropriate entries in the tables provided by Brater and King (1976). The unit took the same hydrological parameters as the surrounding regions. The unaltered overland flow velocity was taken as 100 m h^{-1} for all units. Three scenarios were considered, labelled RT1, RT10 and RT100, corresponding to residence times of 1 hr, 10 hr and 100 h and Λ factors of 0.01, 0.1 and 1, respectively.

3.2   Monte Carlo exercise

An ensemble of 5000 parameter set were sampled randomly from uniform prior distributions with ranges given in Table 1. The observables used to calculate likelihood weighting score were: A_c, the maximum saturated contributing area, or area proportion of the catchment that generates overland flow, the Nash-Sutcliffe statistic (NSE, Nash and Sutcliffe, 1970), and q_{max}, the maximum simulated discharge relative to the observed rated value. The acceptability criteria for A_c were derived from considerations of physically-feasible values. These could be approximated from observations across actual events, for

example, by remotely-sensed imaging (EA, 2013; Luscombe et al., 2015). The criteria observables are given in Table 2, alongside their limits of acceptability. The likelihood score for the NSE was the actual value calculated from the simulated discharges versus the observed, rated values. For the others, the likelihood score was triangular in the corresponding acceptability intervals, with value of unity at the midpoint of the range. The overall weighting score for a realisation is then
5   calculated by taking the mean of the individual scores. The behavioural sets identified by applying the limits shown in Table 2 are then used as the basis for investigation of the effects of applying a number of NFM scenarios.

[INSERT TABLE 2]

4   Results

Of the 5000 realisations undertaken, 384 were identified with outputs that met the acceptability criteria given in Table 2.
10   Figure 4 shows the discharges simulated by these behavioural cases, alongside the observed rated discharges at the EA gauge at Great Musgrave bridge.

The "dotty" plots in Figure 5 show overall GLUE weightings against the three metrics used to select acceptable cases. The maximum saturated area, A_c, shown in the leftmost plot, has a discontinuous shape. This is because when a response unit reaches saturation it contributes its entire area at once. Within the behavioural realisations only a limited number of HRUs,
15   along with the aggregated EHS unit, ever contribute saturated surface flow, leading to just seven distinct values for A_c. Points corresponding to likelihood weightings for realisations producing these distinct values of A_c are shown in the leftmost plot applying a distinct colour for each. The same colours are applied to points corresponding to the same realisations in the other plots. The stratified appearance of the NSE plot is a by-product of its correlation with the contributing area; the correlation with the maximum predicted discharge is less clear. A value of A_c of around 65 % is associated with the best NSE fits, but is
20   spread throughout the maximum discharges. The bias towards higher values suggests that realisations producing greater quantities of fast overland flow better reflect the storm response in this period. This would be consistent with the extreme nature of the storms and observational evidence (e.g. Marsh et al., 2016).

[INSERT Figure 5]

Parameters for each of the 384 accepted cases were applied to catchment models modified to reflect insertion of hillslope
25   storage networks with each of the residence times considered, and the simulations rerun. Reduction in peak Δq for an intervention case is defined as the difference between a base line case and intervention simulation based on this case. For each pair of simulations the value of Δq was multiplied by the base case normalised weighting to reflect the likelihood of this reduction effect. The likelihood-weighted statistics for each of the events and intervention levels are given in Table 3, and the impacts on the arrival time of the main peaks of each in Table 4. The Kolmogorov-Smirnov test included is a non-
30   parametric approach to comparing empirical distribution, equal to the maximum vertical separation of the CDF of the discharges. This allows an evaluation of the relative effectiveness of each intervention.

[INSERT TABLE 3]

Figure 6 shows, as rainfall-equivalent specific surface storage within the aggregated hillslope storage unit across the entire simulation period. When this is filled, and overflow discharge predicted with a Weir equation, the crest height will lead to a slight excess over the specified maximum static storage of 1 m.

[INSERT FIGURE 6]

5 In the RT1 case the areas fill and drain quickly but appear to never fill completely. The available storage is therefore not utilised effectively: utilisation is highest at the peak of Storm Desmond, at 29 % of the theoretical hydrostatic capacity. This is equivalent to approximately 2.45 million m³ volume of storage retained across the catchment. The corresponding effect on the hydrograph is small, with a likelihood-weighted median reduction in the peak of Storm Desmond of 1.7 %. For RT10 the features appear under-utilised in the earlier storms, peaking at about 50 % capacity, and recover almost all their capacity in

10 the recession period after Storm Barney. Retained hillslope runoff across the catchment peaks at just over 10 million m³. The impact of the additional storage is significant in the final storm, and reduces the peak by a likelihood-weighted median of 5.8 % and maximum of 17.3%, the greatest impact of any of the interventions. This intervention would not by itself have prevented flooding, however combined with more conventional FRM measures could have had a significant mitigating effect even through this extreme event.

15 In the RT100 intervention the features fill completely across the course of Storm Abigail near the start of the period, and the drain-down is insufficient to allow complete recovery of capacity before the final event. They are overflowing during much of the event, with the excess following the fast pathways to the channel blocked by the hillslope. The impact is therefore much reduced compared with the 10 hr case, with a median reduction of 1.9 % in the peak, although some cases do show a reduction almost as effective as RT10.

20 Ensemble hydrographs through each of the named storms are shown in the following figures. These present the discharges simulated for the unmodified catchment model using the parameters for each of the acceptable cases against those produced by applying the corresponding parameters to catchment models with the addition of enhanced hillslope storage.

[INSERT Figure 7]

In the initial storm, Abigail, the first peak is attenuated as much by the RT100 case as the RT10 case, but in the second peak

25 RT10 again provides much greater reduction. The RT1 case has the lowest impact on all storm peaks. The arrival of the main peak is retarded most by the RT100 case, with a median delay of 30 minutes. The later peak appears to be brought forward marginally by the RT100 cases, however. This may indicate that the hillslope storage unit has run out of storage and is delivering flow downslope by overflow. In all cases the recession curve is extended by the intervention, indicating that storage is draining for some time after the storm peak.

30 **[INSERT Figure 8]**

In the second event the RT100 case appears to have most impact than any of the other cases across the initial peak but then becomes less effective than RT10 through the main peak. There is less delay to the arrival of the main peak than for Abigail, with a median delay of only 15 minutes for the RT10 and RT100 cases. In the RT10 case the median brings the peak forward

by 15 minutes. This might indicate that their effect has been to slow down fast-responding catchments so that their flood waves contribute more to the rising limb of the overall storm hydrograph.

[INSERT Figure 9]

In this largest event the RT10 case significantly outperforms the others, suggesting that it has recovered much more capacity. There is virtually no attenuation from the other cases and the peak is hardly delayed for RT1 and RT100**.** Figure 9 shows clearly that the RT100 units fill quickly during Abigail and remain full for the duration of Barney and Desmond. There is only a week-long period in later November when they recover a little capacity through drainage during the recession period of Barney.

The RT10 case has the greatest impact on the flood peak for all storms, with its advantage over the other cases increasing through the period. In Abigail the K-S distance of the RT10 case is 150 % that of the RT100 case, whereas across Desmond this has increased to 350 %.

5   Discussion

The surface routing algorithm is computationally very efficient, particularly as it can be solved analytically. The semi-distributed representation is also straightforward and quick, but and allows examination of relevant characteristics such as the drainage times and multiple model runs that apply different network configurations. More sophisticated physically-based modelling of surface flow will introduce computational overheads without accounting for large uncertainties in the input, and may not add greater insight into the catchment response.

The RT10 case was significantly more effective across the largest event than the other two cases, as features had recovered capacity during the previous recession and drained sufficiently to have an impact on the flood peak in Storm Desmond. The mitigation was less across the earlier events but, given they did not cause significant flooding, this was not a consideration.

These results provide further evidence that the theoretical, static storage provided by an NFM scheme can be significantly underutilised, assuming it is configured so that it can recover capacity effectively between storms events. It has already been suggested that the effectiveness in reducing flood risk of a NFM intervention is not a simple function of the additional storage it provides, but also of its contribution to desynchronising a subcatchment's flood wave with those downstream (Thomas and Nisbet, 2007; Blanc et al., 2012).

A more sophisticated approach to implementing NFM is probably required. For instance, the locations on the hillslopes of interventions are likely to be a significant factor on a scheme's performance as they will influence flood wave timing. In this study the network width approach to routing channel flow was applied across the entire headwater catchment, and flood wave size and timings were not available for the individual subcatchments. It was thus unclear from the results what proportion of attenuation of storm hydrographs were due to desynchronisation of flood waves, as opposed to simple retention of surface run-off. A regression analysis similar to that undertaken by Pattison et al. (2013), including as predictor variable the hillslope storage location and quantities could provide insight on the relative contributions of additional hillslope storage

and flood wave desynchronisation on downstream flood mitigation. This can be done in a further analysis, but does raise the question of how to calibrate meaningfully flood wave velocities for the individual subcatchments.

Metcalfe et al. (2015) showed that the runoff predictions simulated by Dynamic TOPMODEL stabilised at around eight to twelve subdivisions of the catchment. In this study eight HRUs were used: despite being at the lower limit of the region of stability identified by Metcalfe et al. (2015), the model gave good fits to the observed hydrographs, with many realisations exceeded efficiencies of 0.9 (although the uncertainties introduced through rating of discharges at extreme storm levels should be noted). The values of the maximum saturated area, A_c were highly discontinuous, however, as when the response units start to produce saturated flow they contribute their entire area to the metric. A more fine-grained discretisation would produce a more continuous distribution which might be advantageous in terms of better identification of saturated contributing areas, in particular how the positioning of hillslope storage features in proposed schemes may influence runoff from these areas.

The features appeared to have an effect through the course of even the largest storm, albeit that the slowest-draining cases were full and overflowing for much of the simulation period. There is, however, a possibility that features such as woody debris dams will be unable to withstand hydraulic loading across such extreme events, or across a series of storms. Complete or partial failure could lead to debris being introduced to the flood waters and potential damage or blockage of downstream infrastructure which will increase risk of flood damage. Ideally, these scenarios should be incorporated in a risk reduction – cost matrix. There is, as yet, little information about the potential for failures that can be used to estimate residual risk., although initial findings suggest that failure sequences can behave non-linearly.

Enhanced hillslope storage was treated as a single unit in the catchment discretisation. In practise different types of interventions, e.g. hillslope ponds, bunded accumulation areas and blocked gullies, will have distinct responses under loading and their locations may also have a significant impact on the runoff. A more realistic approach would recognise the variety of interventions and group features and categorise them by their construction, geometries, and their position on the hillslopes. The number of possibilities will increase rapidly according to the number of varieties and configurations defined, but these are likely to be constrained by considerations of realistic design and implementation issues. In addition, the simplified representation used in this study allows for relatively rapid investigation of many different configurations. This approach can be incorporated into a sensitivity analysis to determine which characteristics are most important to the impact of enhanced hillslope storage features on the storm response. A Monte Carlo approach selecting from multiple design event sets, for example generated by the method of Keef et al. (2013), could help assess the robustness of the conclusions drawn from modelling and help to site features more strategically. It could also help identify situations where flood waves become synchronised by emplacement of features at different catchment scales.

6 Conclusions

This study has analysed the performance across a series of extreme events of a NFM scheme made up of many "leaky" features providing additional storage for hillslope runoff. The model incorporated a simplified overland routing module and achieved a high level of efficiency in simulating the observed discharges. The best base line simulations were applied to the

5 catchment model with the NFM features incorporated. Every variety of enhanced hillslope storage were simulated using a single simple aggregated linear store model. It showed that the combined impact could have significantly attenuated the flood peak, even during the largest storm. It also showed that their effectiveness was contingent on how their drainage characteristics would allow them to recover capacity between events.

The study demonstrates that, given a computationally efficient modelling strategy, uncertainty estimation can be applied to

10 evaluation NFM. A well-established uncertainty analysis framework was used, whereby multiple realisation of a hillslope runoff model applied to the events were enacted and scored with a likelihood function combining output criteria. Acceptable realisations were selected according to limits of acceptability of those criteria and were likelihood-weighted according to their score. This allowed results to presented to project and catchment stakeholders alongside meaningful estimates of their uncertainty.

15 The representation is efficient and allows investigation of many different configurations whilst retaining the important aspects of their behaviour and impacts, namely storage addition and residence times It contains many assumptions and simplifications, however, and a key aim of further work will be to determine whether other significant characteristics are adequately simulated in this representation, and which will require refinement. For example, in actual applications there is likely to be a more complex storage-water depth relationship than the straightforward equivalence used here. Hydrodynamic

20 analysis of simulated individual structures sited in realistic topographic representations may provide insights into the applicability of the simplification across a range of loading scenarios.

The storage considered in this study were lumped into a single HRU to maximise computational efficiency. A more sophisticated approach could utilise many more classifications that reflect their behaviour and impact on the runoff. This could include capacity, position on the hillslope, distance from access tracks and the channel and sources of construction

25 material, type of location (e.g. within ephemeral channels, shallow hillslope accumulation areas or on the floodplain). More work will be needed to determine which of these characteristics will be most significant and how well classifications reflect actual implementations by NFM practitioners.

It may be that the most beneficial effect of additional hillslope storage is likely to be seen on a small scale in reaches immediately downstream of a feature or sets of features. It could be, even given the overall mitigation effect, that some

30 asynchronous flood peaks in the unmodified catchment model became synchronised when the storage was introduced and thus reduced the overall effectiveness of the interventions. Much further work is required to better understand the synchronisation problem, particularly as catchment scale increases.

Obtaining observational evidence to support modelling predictions will be difficult in the field as, by their nature, the extreme events that load features are rare and gradual processes such as sediment deposition difficult to measure or simulate everywhere. An innovative experimental approach is needed to address these questions. Detailed hydrometric data are required, collected by instruments such as stage and flow gauges upslope, downslope and within features. However only 6 %

5   of NFM schemes in the UK currently have any type of monitoring (JBA Trust, 2016). The effects of the additional storage on the overall flood hydrograph will be increasingly difficult to discern as the catchment size increases, including the potential for synchronicity effects. The methodology used in this paper, however, extended to incorporate better-supported representations of small scale impacts due to feature emplacement, may provide a productive way to take forward research into natural flood management and its effectiveness.

10   **Acknowledgements**

This work was funded by the JBA Trust (project number W12-1866), the European Regional Development Fund, and the UK National Environmental Research Council (NERC).

References

Barber, N., and Quinn, P., 2012. Mitigating diffuse water pollution from agriculture using soft engineered runoff attenuation features. Area, 44(4), 454-462.

Beven, K., 2006. A manifesto for the equifinality thesis. Journal of hydrology, 320(1), 18-36.

Beven, K, 2012. Rainfall-Runoff Modelling: The Primer (2nd Edition), Wiley-Blackwell: Chichester, UK. ISBN: 978-0-

Beven, K. and Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes, 6 (3), 279-298.

Beven, K. and Binley, A., 2014. GLUE: 20 years on. Hydrological Processes, 28(24), 5897-5918.

Beven, K. and Freer, J., 2001a. A Dynamic TOPMODEL. Hydrological Processes, 15(10), 1993-2011.

Beven, K., and Freer, J., 2001b. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249(1), 11-29.

Beven, K. J. and Kirkby, M. J., 1979. A physically based variable contributing area model of catchment hydrology, Hydrol. Sci. Bull., 24, 43-69.

Beven, K., & Wood, E. F. (1983). Catchment geomorphology and the dynamics of runoff contributing areas. Journal of Hydrology, 65(1–3), 139-158. doi: http://dx.doi.org/10.1016/0022-1694(83)90214-7

Blanc, J., Wright, G. and Arthur, S. 2012. Natural Flood Management knowledge system: Part 2 – The effect of NFM features on the desynchronising of flood peaks at a catchment scale. CREW report.

Blazkova, S. and Beven, K., 2004. Flood frequency estimation by continuous simulation of subcatchment rainfalls and discharges with the aim of improving dam safety assessment in a large basin in the Czech Republic. Journal of Hydrology, 292(1), 153-172.

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).

Brater, E. F., & King, H. W., 1976. Handbook of hydraulics for the solution of hydraulic engineering problem, 6th Ed. McGraw-Hill. New York.

Calder, I. R. and Aylward, B., 2006. Forest and floods: Moving to an evidence-based approach to watershed and integrated flood management. Water International, 31(1), 87-99.

Chow, V.T., 1959. Open channel hydraulics. McGraw-Hill Book Company, Inc; New York.

Dadson, S. J., Hall, J. W., Murgatroyd, A., Acreman, M., Bates, P., Beven, K., Heathwaite, L., Holden, J., Holman I.P., Lane, S.N., O'Connell, E., Penning-Rowsell, E., Reynard, N., Sear, D., Thorne, C., & Wilby, R. (2017). A restatement of the natural science evidence concerning catchment-based "natural" flood management in the United Kingdom. UKProc. R. Soc. A 473: 20160706. http://dx.doi.org/10.1098/rspa.(2016).0706

Dummit, E (2012). Differential Equations (part 3): Systems of First-Order Differential Equations. Downloaded 16/01/2015 from http://www.math.wisc.edu/~dummit/docs/diffeq_3_systems_of_linear_diffeq.pdf

Environment Agency, 2008. River Eden, Cumbria, Catchment Flood Management Plan. Environment Agency, Bristol, UK.

Environment Agency, 2009. Using airborne remote sensing for flood risk management with the Environment Agency. Downloaded 12/12/2015 from http://www.ggy.bris.ac.uk/hydrology/bhsmeeting/AlastairDuncan.pdf

Environment Agency, 2012. Greater working with natural processes in flood and coastal erosion risk management – a response to Pitt Review Recommendation 27. Environment Agency led working group. Environment Agency, Bristol, UK.

Environment Agency, 2013, Benchmarking the latest generation of 2D hydraulic modelling packages, Report SC120002, ISBN: 978-1-84911-306-9, Environment Agency, Bristol, UK.

Environment Agency, 2014. Working with natural processes to reduce flood risk. R and D framework. Report SC130004/R2. Environment Agency, Bristol, UK.

Ghimire, S., Wilkinson, M. and Donaldson-Selby, G., 2014. Application Of 1D And 2D Numerical Models For Assessing And Visualizing Effectiveness Of Natural Flood Management (NFM) Measures.

Hankin, B., Lamb, R., Craigen, I., Page, T., Chappell, N., Metcalfe, P., 2016. The Rivers Trust Life-IP Natural Course Project: Strategic Investigation of Natural Flood Management in Cumbria. JBA Consulting report number 2016s4667. Available at http://naturalcourse.co.uk/uploads/2017/04/2016s4667-Rivers-Trust-Life-IP-NFM-Opportunities-Technical-Report-v8.0.pdf

Hankin, B., Metcalfe, P., Johnson, D., Chappell, N., Page, T., Craigen, I., Lamb, R., Beven, K., 2017. Strategies for Testing the Impact of Natural Flood Risk Management Measures. In Hromadka, T. & Rao, P. (eds). Flood Risk Management. InTech, Czech Republic. ISBN 978-953-51-5526-3.

Hunter, N.M., Bates, P.D., Horritt, M.S. and Wilson, M.D., 2007. Simple spatially-distributed models for predicting flood inundation: a review. Geomorphology, 90(3), pp.208-225.

Iacob, O., Rowan, J. S., Brown, I. & Ellis, C. (2014). Evaluating wider benefits of natural flood management strategies: an ecosystem-based adaptation perspective. Hydrology Research, 45(6), 774-787.

Keef, C., Tawn, J.A. and Lamb, R., 2013. Estimating the probability of widespread flood events. Environmetrics, vol 24, no. 1, pp. 13-21., 10.1002/env.2190

Lamb, R., Crossley, A., Waller, S., 2009. A fast 2D floodplain inundation model. Proceedings of the Institution of Civil Engineers: Water Management, Volume 162.

Lamb, R., Keef, C., Tawn, J., Laeger, S., Meadowcroft, I., Surendran, S., Dunning, P. and Batstone, C., 2010. A new method to assess the risk of local and widespread flooding on rivers and coasts, Journal of Flood Risk Management. 3, 4, p. 323-336 14 p, DOI: 10.1111/j.1753-318X.2010.01081.x

Lane, S. N., 2017. Natural flood management. Wiley Interdisciplinary Reviews: Water, 4(3).

Lane, S.N., Odoni, N., Landström, C., Whatmore, S.J., Ward, N. and Bradley, S., 2011. Doing flood risk science differently: an experiment in radical scientific method. Transactions of the Institute of British Geographers, 36(1), pp.15-36.

Lavers, T. and Charlesworth, S., 2016. Natural Flood Risk Management and its Role in Working with Natural Processes. In Charlesworth, S. and Booth, C. (eds.). Sustainable Surface Water Management: A Handbook for SUDS, p156-172. John Wiley and Sons.

Liu, Y., Freer, J., Beven, K. and Matgen, P., 2009. Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error. Journal of Hydrology, 367(1), pp.93-103.

Luscombe, D.J., Anderson, K., Gatis, N., Grand-Clement, E. and Brazier, R.E., 2015. Using airborne thermal imaging data to measure near-surface hydrology in upland ecosystems. *Hydrological processes*, *29*(6), pp.1656-1668.

Marsh, T.J., Kirby, C., Muchan, K., Barker, L., Henderson, E. & Hannaford, J., 2016. The winter floods of 2015/2016 in the UK - a review. Centre for Ecology & Hydrology, Wallingford, UK

Metcalfe, P., Beven, K. and Freer, J., 2015. Dynamic TOPMODEL: A new implementation in R and its sensitivity to time and space steps. Environmental Modelling and Software, 72, 155-172.

Metcalfe P., Beven, K. and Freer, J., 2016. dynatopmodel: Implementation of the Dynamic TOPMODEL Hydrological Model. R package version 1.1. https://CRAN.R-project.org/package=dynatopmodel

Metcalfe, P., Beven, K., Hankin, B. and Lamb, R., 2017. A modelling framework for evaluation of the hydrological impacts of nature-based approaches to flood risk management, with application to in-channel interventions across a 29km² scale catchment in the United Kingdom. Hydrological Processes. DOI 10.1002/hyp.11140

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - A discussion of principles. Journal of hydrology, 10(3), pp.282-290.

Matthews, T., Murphy, C., McCarthy, G., Broderick, C., and Wilby, R. L., 2018. Super Storm Desmond: a process-based assessment. Environmental Research Letters, 13(1), 014024.

Mosonyi, E., 1966. Wassrekraftwerke, Band I, Niederdruckanlagen. Düsseldorf.

National Trust, 2015. From source to sea: natural flood management – the Holnicote experience. Final Report to Defra.

Nicholson, A., Wilkinson, M., O'Donnell, G. M. and Quinn, P., 2012. Runoff attenuation features: a sustainable flood mitigation strategy in the Belford catchment, UK. Area, 44(4), 463-469.

Nisbet, T.R., Marrington, S. Thomas, H., Broadmeadow, S. and Valatin, G. , 2011. Slowing the flow at Pickering. Final Report to Defra, 28pp.

Nisbet, T. & Page, T., 2016. Case study 16. Pontbren catchment land use change study. In Hankin, B., Burgess-Gamble, L., Bentley, S., Rose, S. (Eds.). How to model and map catchment processes when flood risk management planning. Science report SC120015/R1, Environment Agency, Bristol, UK.. Available at http://evidence.environment-agency.gov.uk/FCERM/Libraries/FCERM_Project_Documents/SC120015_case_study_16.sflb.ashx.

Ockenden, M.C. and Chappell, N.A., 2011. Identification of the dominant runoff pathways from data-based mechanistic modelling of nested catchments in temperate UK. Journal of hydrology, 402(1), pp.71-79.

Page, T., Beven, K. J., Freer, J., & Neal, C., 2007 . Modelling the chloride signal at Plynlimon, Wales, using a modified dynamic TOPMODEL incorporating conservative chemical mixing (with uncertainty). Hydrological Processes, 21(3), 292-307.

Pattison I, Lane, S.N., Hardy, R.J. and Reaney , .M., 2014. The role of tributary relative timing and sequencing in controlling large floods. Water Resources Research 50(7): 5444-5458.

Quinn, P., O'Donnell, G., Nicholson, A., Wilkinson, M., Owen, G., Jonczyk, J., Barber, N., Hardwick, M., and Davies, G., 2013. Potential use of Runoff Attenuation Features in small rural catchments for flood mitigation. England: Newcastle University, Environment Agency, Royal Haskoning DHV.

Scottish Environment Protection Agency (SEPA), 2016. The Natural Flood Management Handbook. Available at: http://www.sepa.org.uk/media/163560/sepa-natural-flood-management-handbook1.pdf

Thomas, H. and Nisbet, T., 2007. An assessment of the impact of floodplain woodland on flood flows. Water and Environment Journal 21(2): 114-126.

Werritty, A., 2006. Sustainable flood management: oxymoron or new paradigm?. Area, 38(1), 16-23.

Wheater, H., Reynolds, B., McIntyre, N., Marshall, M., Jackson, B., Frogbrook, Z. and Chell, J., 2008. Impacts of upland land management on flood risk: Multi-scale modelling methodology and results from the Pontbren Experiment.

Wilkinson, M., Quinn, P., Benson, I., and Welton. P., 2010. "Runoff management: mitigation measures for disconnecting flow pathways in the Belford Burn catchment to reduce flood risk." In British Hydrological Society International Symposium. British Hydrological Society, Newcastle upon Tyne,.

Wilkinson, M., Quinn, P., and Welton, P., 2010. Runoff management during the September 2008 floods in the Belford catchment, Northumberland. Journal of Flood Risk Management, 3(4): 285-295.

Tables

Table 1: Runoff model parameter and ranges applied in calibration and uncertainty analysis.

Parameter	Description	Units	Lower	Upper
m	Form of exponential decline in conductivity	m	0.0011	0.033
srz_{max}	Max root zone storage	m	0.1	0.3
srz_0	Initial root zone storage	%	20	100
v_{chan}	Channel routing velocity	m h^{-1}	500	5000
$ln(T_0)$	Lateral saturated transmissivity	m² h^{-1}	3	12
sd_{max}	Maximum effective deficit of saturated zone	m	0.5	-
ex_{min}	Minimum surface storage	m	0	-
ex_{max}	Maximum surface storage	m	1	1

Table 2. Observables collected for each model realization and acceptability criteria applied

Metric	Description	Units	Criteria applied
NSE	Nash Sutcliffe statistic	-	>= 0.85
A_c	Maximum saturated contributing area	%	[10, 95]
q_{max}	Specific discharge at largest storm peak	mm h^{-1}	[5.2, 8.2]

Table 3. Statistics for the impacts of interventions relative to base cases on the peaks of the named storms within the simulation period. Each difference is weighted by the likelihood score for the base case simulation. Δq_{max} (mm h^{-1}) is the maximum and Med the median percentage likelihood weighted reduction in peak. K.S. is Kolmogorov-Smirnov Statistic, a non-parametric statistic equal to the maximum separation of the CDF of the simulated base case peaks versus the peaks calculated for the intervention case. This shows that the RT10 intervention consistently has the greatest impact across the storms.

	Abigail			Barney			Desmond		
Name	Δq_{max}	Med	K-S	Δq_{qmax}	Med	K.S.	Δq_{max}	Med	K.S.
RT1	8.4	2.6	0.013	18.7	3.5	0.016	7.9	1.7	0.033
RT10	36.7	12	0.103	34	10.4	0.141	17.3	5.8	0.129
RT100	22.2	6.6	0.05	33.5	7.7	0.063	17	1.9	0.037

Table 4. Mean and median delays (h) to simulated peaks of each of the named storms for the intervention cases compared to the corresponding unmodified cases.

Name	Abigail		Barney		Desmond	
	Mean	Median	Mean	Median	Mean	Median
RT1	0.3	0.3	0.4	0.3	1.6	0
RT10	-0.1	0	-0.2	-0.3	6.2	6.5
RT100	0.5	0.5	0.3	0	1.6	0

Figures
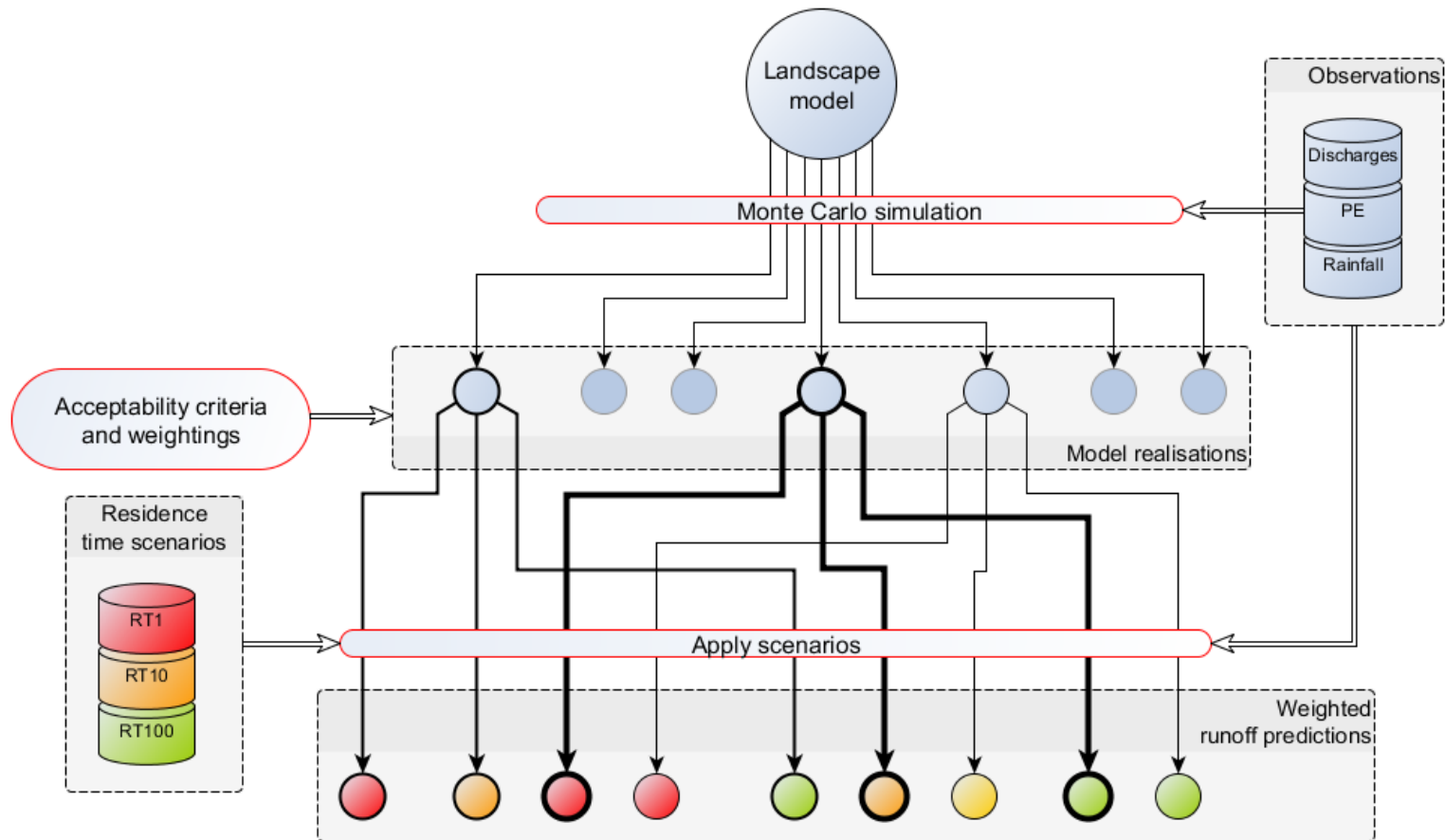
Figure 1. Work flow diagram for Monte Carlo simulation of storm runoff, and selection and weighting of behavioural realisations and application of NFM scenarios for forward prediction of change. The weight of lines leading from acceptable simulations reflects the weighting likelihood score in the validity of that realisation.
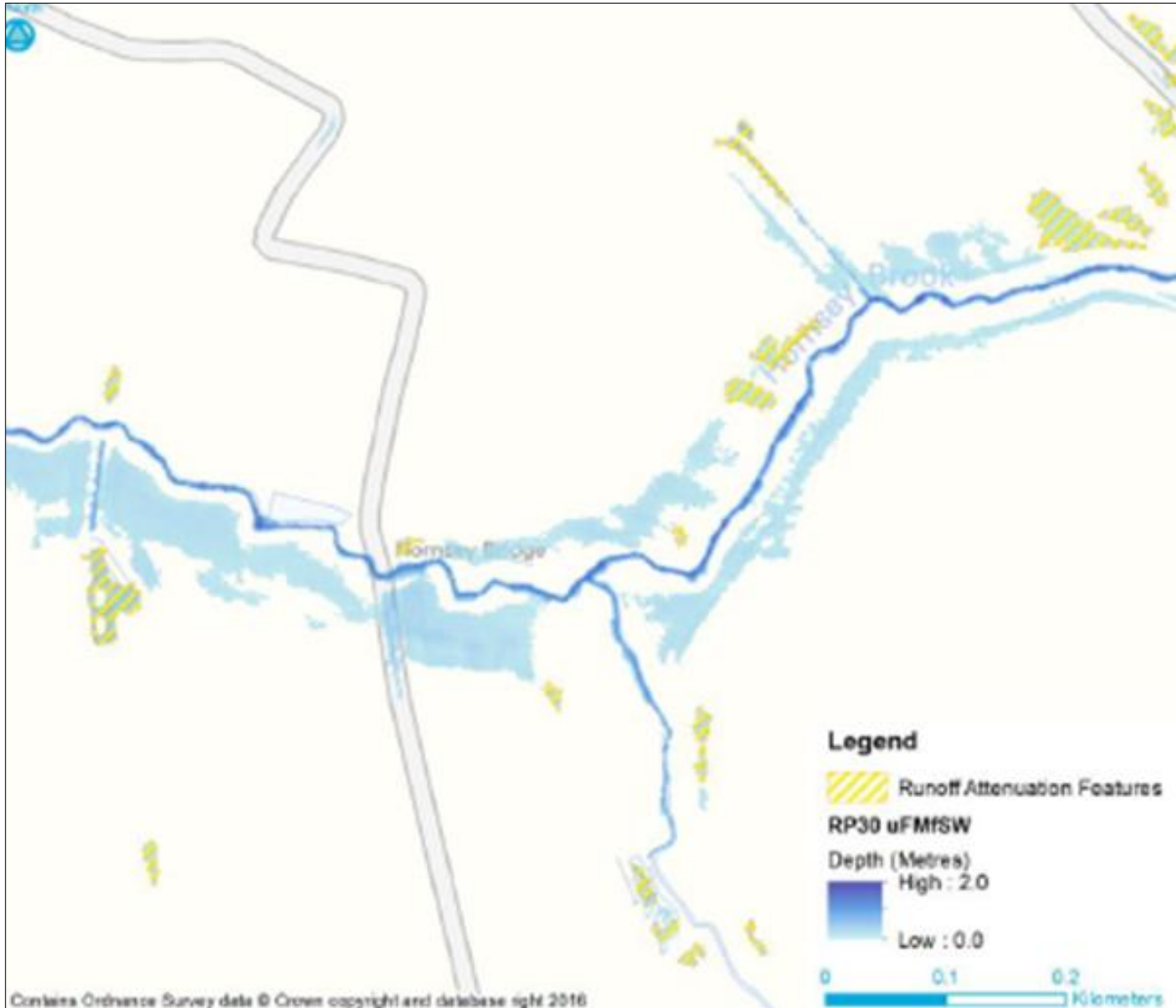
Figure 2. Hydrodynamic accumulation areas within Eden, identified by JFLOW analysis for a designed storm of return period of 30 years (Hankin et al., 2017). Maximum water depths are indicated, and areas that exceed the threshold depth and other criteria (minimum area, slope angle and proximity to roads and buildings) are highlighted as potential sites for EHS.
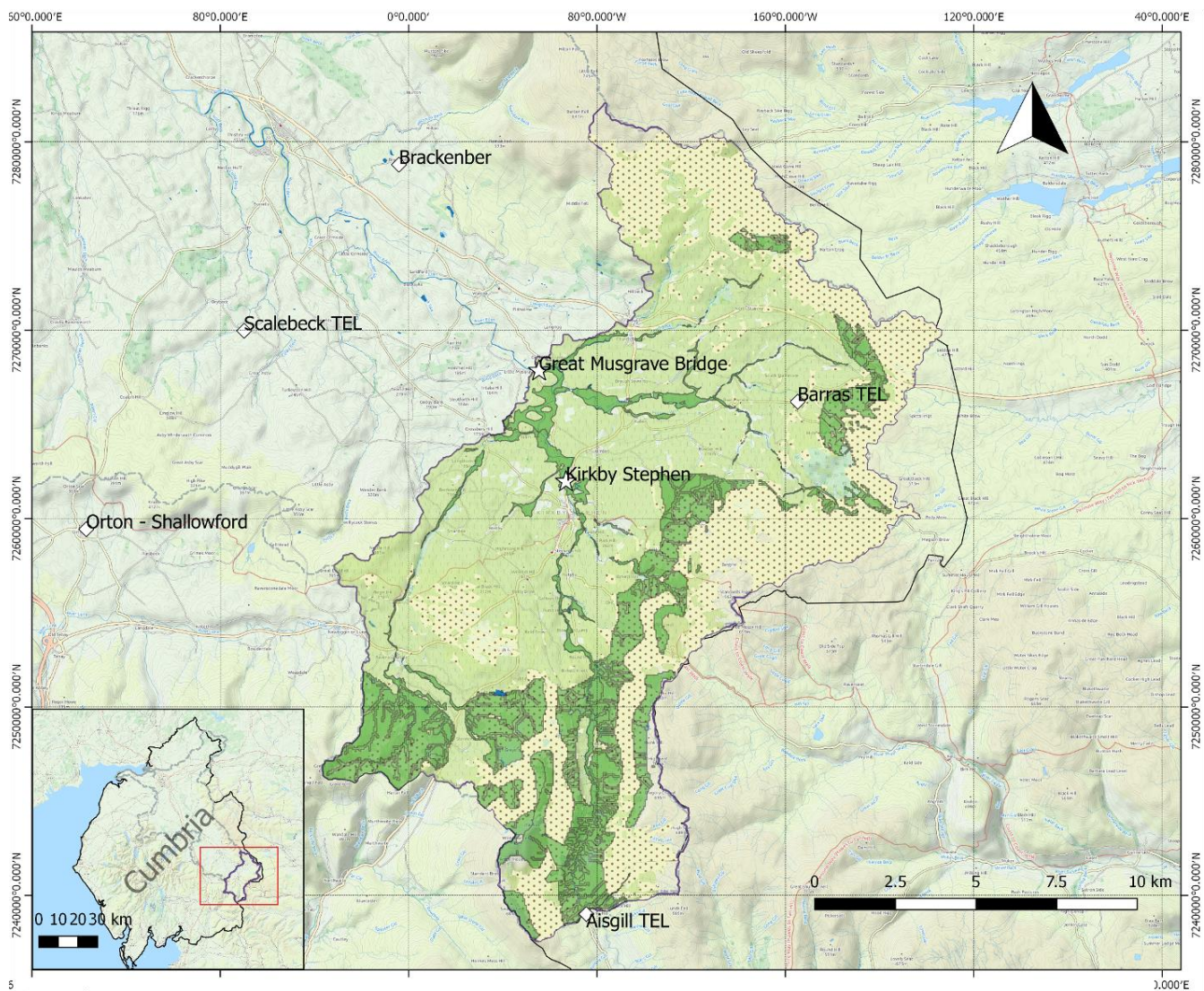
Figure 3. Study catchment, the Eden headwaters to Great Musgrave Bridge (223 km²), showing context within Cumbria, UK, predominant land cover and location of TBR rain gauges and gauging stations. Woodlands for Water tree planting opportunity areas are shown. These were applied in another application of the NFM modelling framework developed for the project described, which are not discussed in detail here.
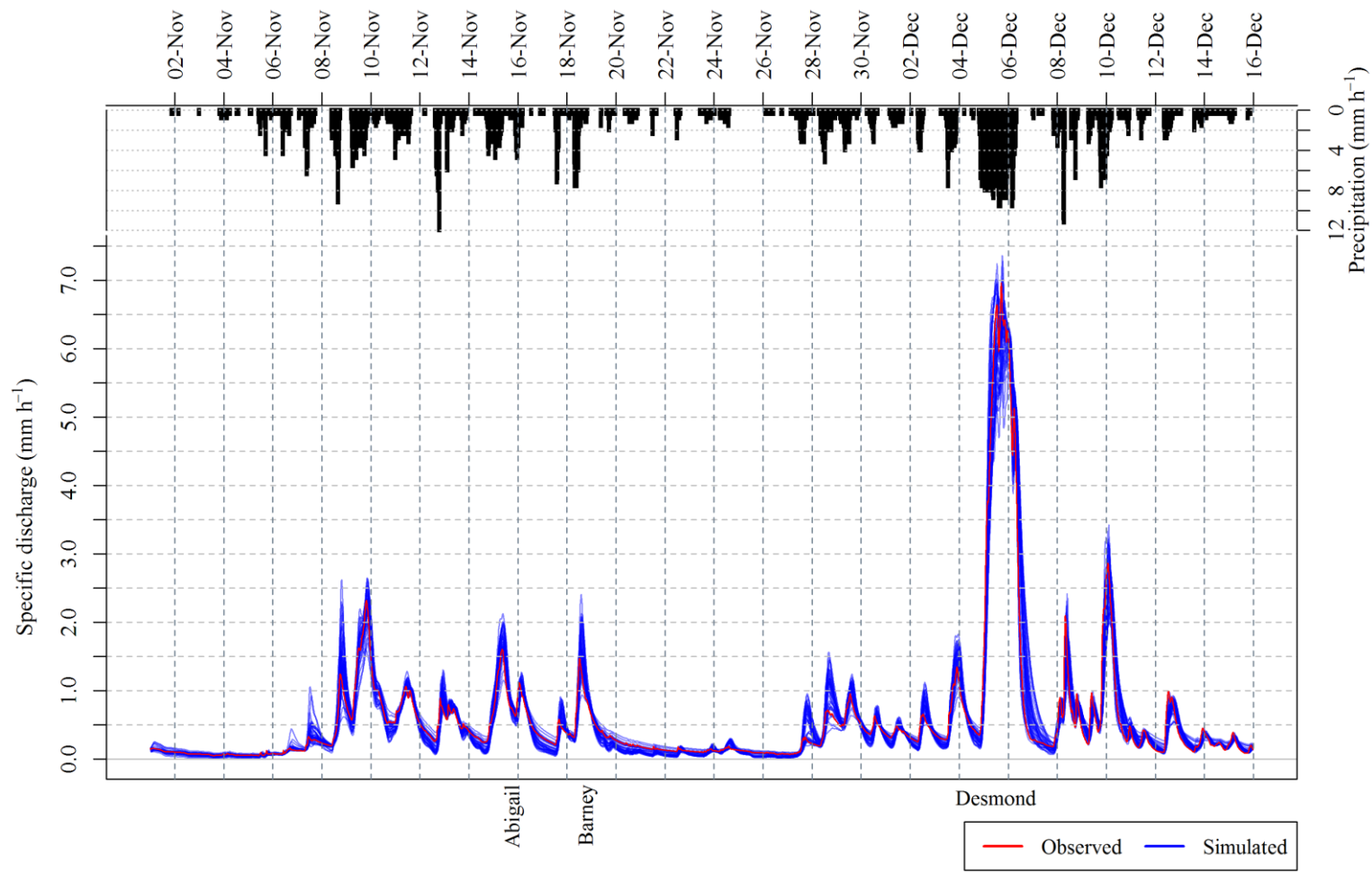
Figure 4. Simulated discharges across the storm period alongside rated observed discharges at Great Musgrave Bridge. The three named storms are indicated. Rainfall is interpolated between the gauges shown in Figure 3.
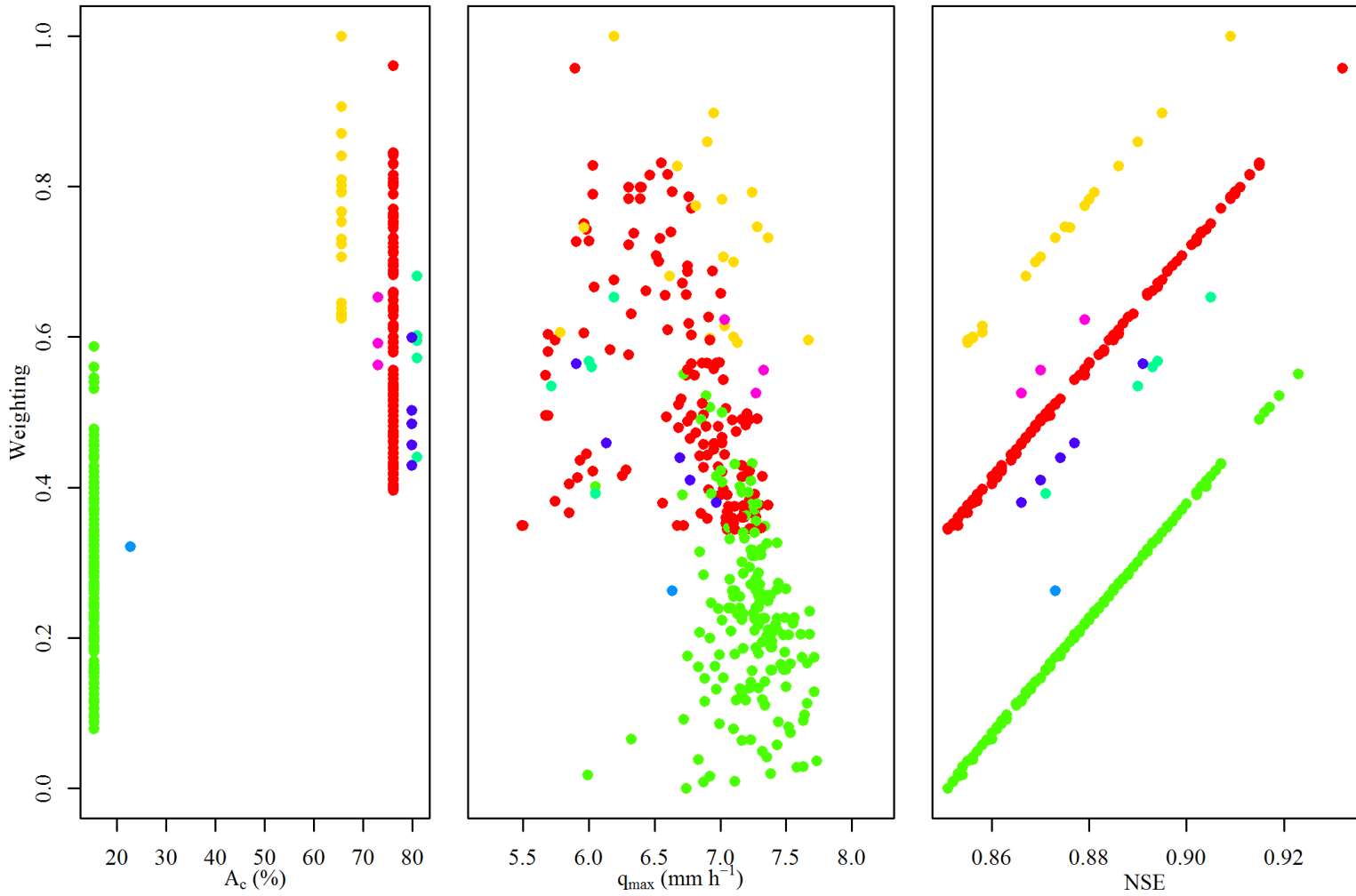
Figure 5. GLUE "dotty" plots showing overall likelihood (weighting) scores calculated for the 348 behavioural runoff simulations against the three model outputs described in the text. The discontinuous appearance of the maximum saturated contributing area A_c is due to the relatively coarse discretisation applied such that once a HRU begins to produce any saturated overland flow, its entire area is added. Each unique A_c value takes a separate colour that is carried through to corresponding points in the other plots.
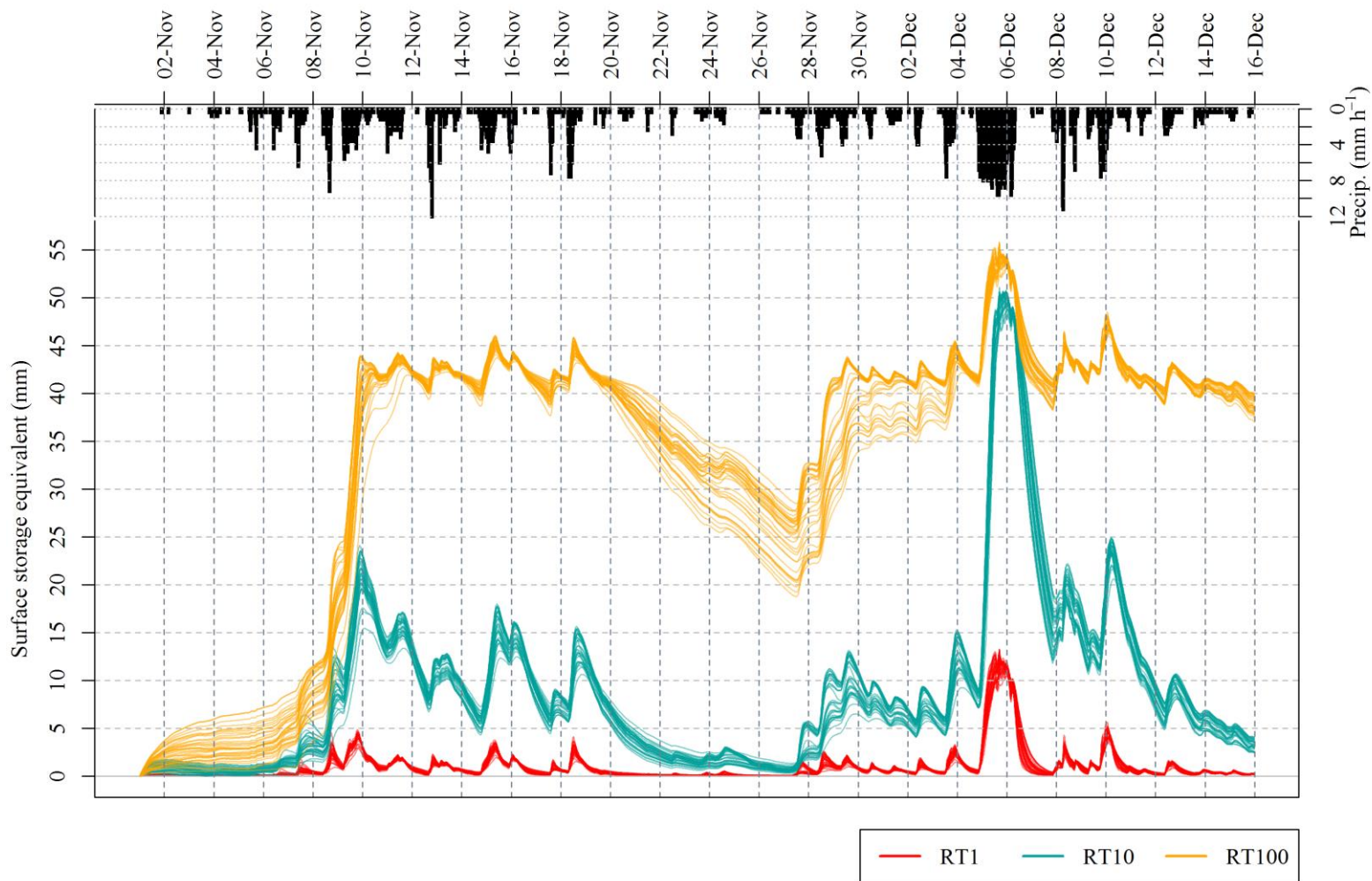
Figure 6. Surface excess storages, expressed as specific rainfall equivalent, across one of the lumped RAF units with maximum storage 1 m through a single intervention cases and for the three mean residences times considered. The slight excess at the peaks of the storm reflects the weir crest height of the overflow function applied.
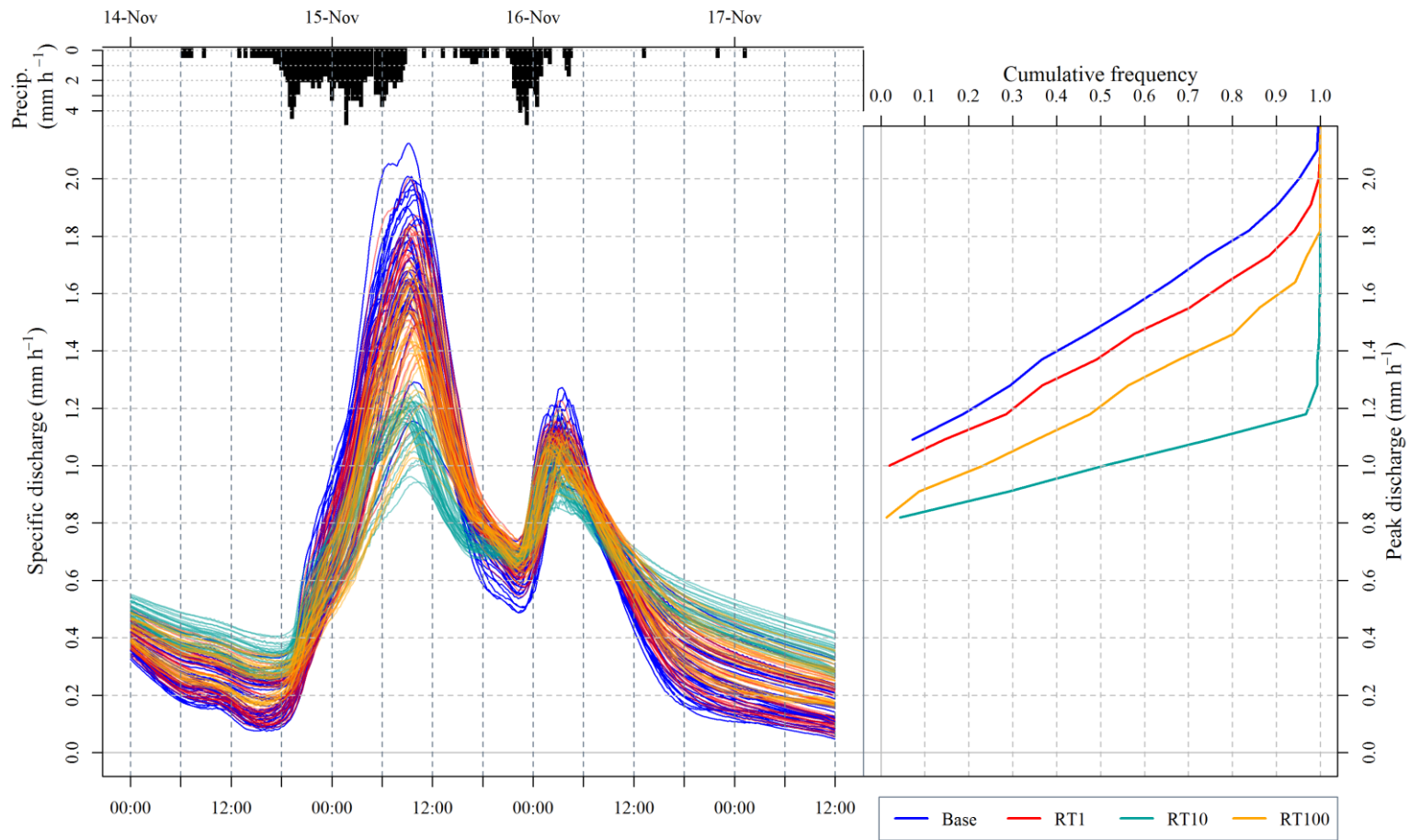
Figure 7. (L) 90th percentile likelihood scored baseline and corresponding intervention cases through Storm Abigail. (R) Likelihood-weighted cumulative frequency plot of peak discharges for base and intervention cases. Note that, in order to share the same vertical axis, the Cumulative frequency plot is transposed relative to convention.
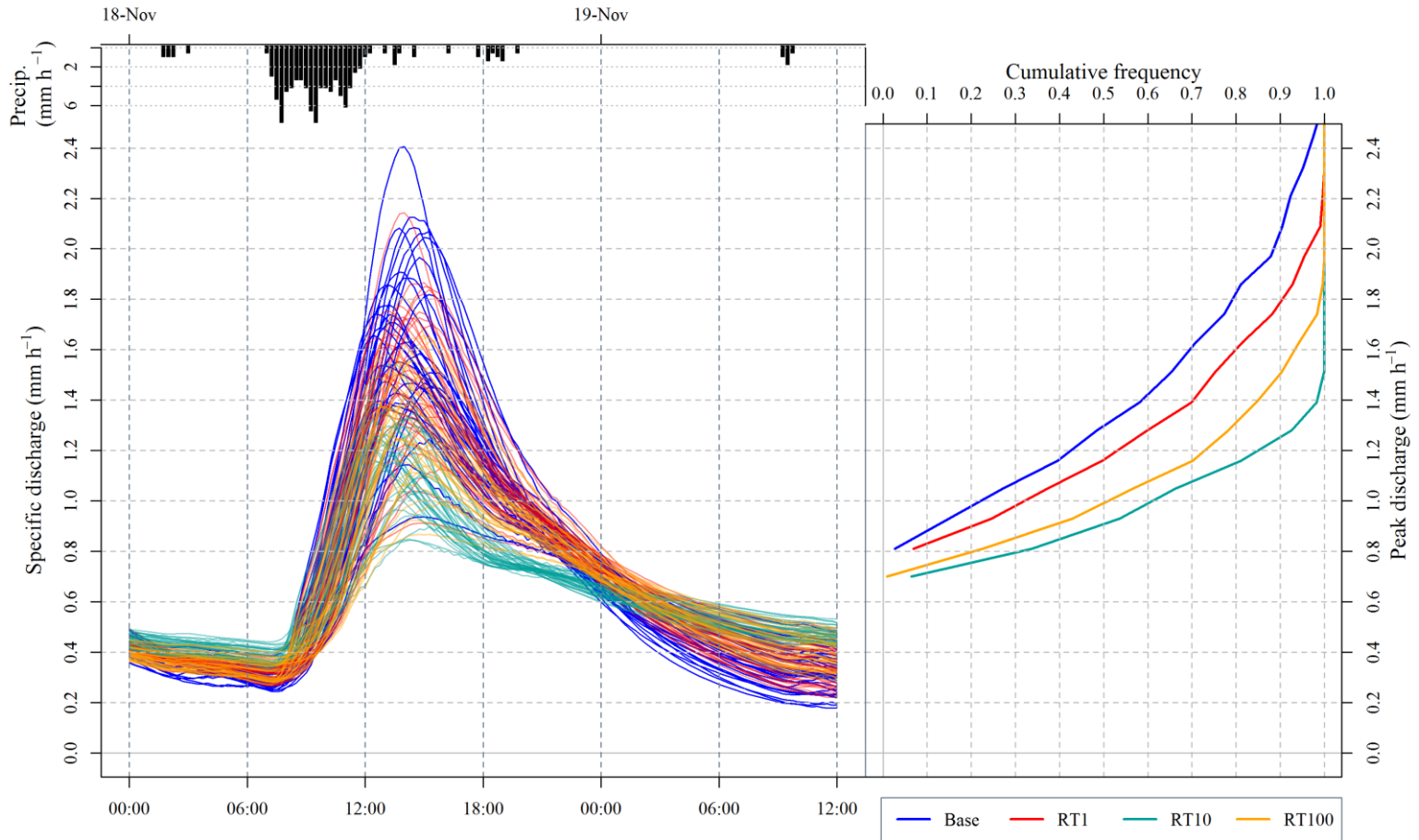
Figure 8. (L) 90th percentile scored base line and corresponding intervention cases through Storm Barney. (R) Cumulative frequency plot of peak discharges for base and intervention cases.
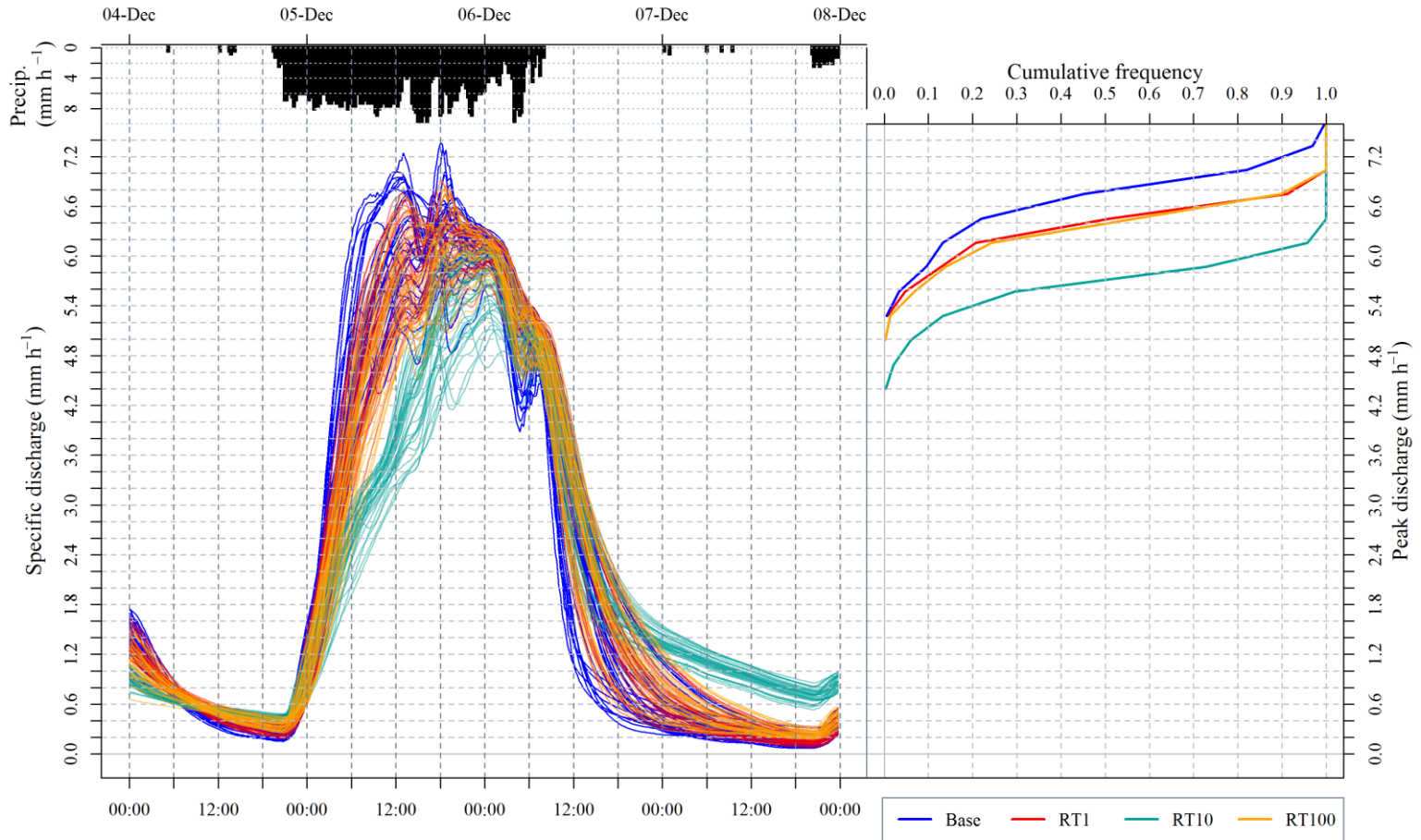
Figure 9. (L) 90th percentile scored base line and corresponding intervention cases through Storm Desmond. (R) Cumulative frequency plot of peak discharges for base and intervention cases.