AGU ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research

RESEARCH ARTICLE
10.1029/2024WR037320

Special Collection:
Quantifying Nature-based Climate Solutions

Natural Flood Management Through Peatland Restoration: Catchment-Scale Modeling of Past and Future Scenarios in Glossop, UK

Salim Goudarzi[1], David Milledge[2], Joseph Holden[3], Martin Evans[4], Tim Allott[4], Adam Johnston[4], Emma Shuttleworth[4], Martin Kay[4], David Brown[5], Joe Rees[4], Donald Edokpa[4], and Tom Spencer[6]

[1]School of Geosciences, University of Aberdeen, Aberdeen, UK, [2]Department of Civil Engineering, Newcastle University, Newcastle Upon Tyne, UK, [3]Water@leeds, School of Geography, University of Leeds, Leeds, UK, [4]Department of Geography, University of Manchester, Manchester, UK, [5]Greater Manchester, Merseyside and Cheshire Environment Agency, Manchester, UK, [6]Moors for the Future Partnership, Manchester, UK

Key Points:

- At catchment scale, it is not necessary (nor feasible) to delay the flood-wave to meaningfully attenuate it at the outlet
- Deferring only a portion of the flood volume to the receding limb can be sufficient for meaningful flood risk mitigation at catchment scale
- Through Sphagnum re-establishment, significant Natural Flood Management + ecosystem benefits can be delivered at scales/storms relevant to communities at risk

Correspondence to:
S. Goudarzi,
salim_goodarzi@yahoo.com

Citation:
Goudarzi, S., Milledge, D., Holden, J., Evans, M., Allott, T., Johnston, A., et al. (2024). Natural flood management through peatland restoration: Catchment-scale modeling of past and future scenarios in Glossop, UK. *Water Resources Research*, *60*, e2024WR037320. https://doi.org/10.1029/2024WR037320

Received 10 FEB 2024
Accepted 17 JUL 2024

Author Contributions:
Conceptualization: Salim Goudarzi, David Milledge, Joseph Holden, Martin Evans, Tim Allott
Data curation: Martin Evans, Tim Allott, Adam Johnston, Emma Shuttleworth, Martin Kay, Joe Rees, Donald Edokpa
Formal analysis: Salim Goudarzi, Joseph Holden, Martin Evans, Tim Allott, David Brown
Funding acquisition: David Milledge, Joseph Holden, Martin Evans, Tim Allott, Tom Spencer
Investigation: Salim Goudarzi, David Milledge, Adam Johnston, Emma Shuttleworth, David Brown

Abstract Field-scale experiments have shown the Natural Flood Management (NFM) potential of peatland restoration. The likelihoods of effectiveness are yet unknown at scales and storms large enough to impact human lives. Using GMD-TOPMODEL, we upscale a rare Before-After-Control-Intervention empirical data set to a 25 km^2 catchment with >600 properties at flood-risk, and test storms of up to a 1,000-year return period (RP). Under these scales/storms, we find that it is not necessary (nor feasible) to delay the outlet flow-peak to meaningfully attenuate it. Enhancing catchment "kinematic" storage, for example, through restoration, can be sufficient to reduce flow magnitudes without detectable changes to peak-flow timing. NFM benefit increases exponentially with restoration area size under smaller storms, but linearly under larger storms. At RP ≤ 100 years, longer-lasting frontal-type storms are more challenging to defend against via NFM, but at RP > 100 years shorter-duration convectional-type events become more challenging. In the order of 1,000–10 years storms: (a) revegetating the bare-peat areas in 15% of the catchment is 31%–61% likely to reduce peak-flows by >5%; (b) revegetating & damming the erosion gullies in ∼20% of the catchment is 42%–71% likely to reduce peak-flows by >5%; (c) Growth of Sphagnum in the dammed gullies of ∼20% and ∼40% of the catchment increase the likelihoods of >5% peak reductions to 65%–86% and 90%–98%, respectively. The numerical evidence of significant NFM benefit due to Sphagnum re-establishment is an important finding, because it shows that meaningful flood-risk mitigation in headwater catchments under scales/storms relevant to communities at risk can be delivered alongside other ecosystem benefits of Sphagnum re-establishment.

1. Introduction

Natural Flood Management (NFM) is part of the Nature Based Solutions (NBS) initiative (Cohen-Shacham et al., 2016). NBS involves restoring disrupted ecosystems and the services they provide to tackle societal challenges, thus simultaneously improving human well-being and biodiversity. NFM is the branch of NBS that is concerned with flooding (Dadson et al., 2017); it focuses particularly on those ecosystems that if restored, can act as a natural buffer against flooding for their downstream communities (Nilsson et al., 2018).

Peatlands are good examples of such damaged systems, and have been suggested as ideal targets for NFM (Bonn et al., 2009). Although they cover only 2.84% of the global land area, they provide a wide range of important services including: more than a third of soil carbon storage, river flow regulation, drinking water supply, and biodiversity (Bonn et al., 2016; Holden, 2005; Holden et al., 2009; Milledge et al., 2015; Xu et al., 2018). These sensitive systems require high precipitation or impeded drainage to develop and remain stable (Acreman & Holden, 2013; M. Evans & Warburton, 2011; M. J. Evans et al., 1999; Holden & Burt, 2002, 2003). Factors contributing to their erosion include: industrialization (and associated atmospheric pollution), overgrazing, mechanization, land-use change (e.g., agriculture or forestry) and extraction of peat for horticultural and energy production (M. Evans & Warburton, 2011; Parry et al., 2014; Rothwell et al., 2007; Shuttleworth et al., 2015).

Blanket peatlands are the dominant UK peatland type (and the focus of this study), but the majority of UK blanket peats have been degraded by drainage, erosion and/or the loss of keystone species such as Sphagnum mosses (M. Evans & Warburton, 2011). Blanket peatlands typically occur on sloping terrain sometimes up to 15° (Lindsay et al., 1988) and their runoff is dominated by saturation-excess overland flow (Holden & Burt, 2003).

Methodology: Salim Goudarzi, David Milledge, Joseph Holden
Project administration: David Milledge, Joseph Holden, Martin Evans, Tim Allott
Resources: Salim Goudarzi, David Milledge, Martin Evans, Adam Johnston, Emma Shuttleworth, Martin Kay, Tom Spencer
Software: Salim Goudarzi
Supervision: David Milledge, Joseph Holden, Martin Evans, Tim Allott
Validation: Salim Goudarzi, Adam Johnston
Visualization: Salim Goudarzi
Writing – original draft: Salim Goudarzi
Writing – review & editing: Salim Goudarzi, David Milledge, Joseph Holden, Martin Evans, Tim Allott, David Brown

Consequently, once surface vegetation cover is damaged, fast overland flow can lead to more rapid erosion (M. Evans et al., 2005, 2006). Over time, erosion often leaves behind large areas of bare peat and networks of deep erosion gullies (Pilkington et al., 2015; Shuttleworth et al., 2019), which can severely disrupt ecosystem service provision (M. Evans & Warburton, 2011). To stop and reverse the erosion effects, blanket peat restoration techniques commonly applied today include: hillslope re-vegetation using a nurse crop (e.g., amenity grasses), and blocking of erosion gullies using natural materials such as peat, timber, or cobble stone dams (Howson et al., 2023; Milledge et al., 2020; Parry et al., 2014; Pilkington et al., 2015; Shuttleworth et al., 2019). In the last decade substantial restoration efforts have also been directed to the re-introduction of key native species, particularly Sphagnum through widespread plug planting and other techniques (Wittram et al., 2015).

In addition to improved ecosystem functions, field-scale experiments show that blanket peat restoration can also attenuate storm peak runoff significantly, reducing the magnitude and delaying the timing of peak discharge (McIntyre et al., 2012; Parry et al., 2014; Pilkington et al., 2015; Shuttleworth et al., 2019). The focus of most peatland restoration schemes has traditionally been on habitat recovery, protecting stored carbon, and reducing wildfire risk, though a broader range of benefits have been more recently considered (Bonn et al., 2016; Chimner et al., 2017; Loisel & Gallego-Sala, 2022). In the UK, because a substantial number of headwater systems are dominated by blanket peat, there is interest in whether flood mitigation benefits can be derived from peatland restoration (Acreman & Holden, 2013). However, implementation is currently limited due to an evidence-gap (Bark et al., 2021; Klaar et al., 2020; Wilkinson et al., 2019; Wingfield et al., 2019): is NFM capable of mitigating flooding at scales large enough to impact communities (and thus justify its cost)?

A logical step prior to large-scale implementation of NFM is numerical modeling. Previous attempts have been successful in modeling NFM interventions and estimating their impact at scale (e.g., non-blanket peat examples Addy & Wilkinson, 2019; Dixon et al., 2016; Ferguson & Fenner, 2020; Hankin et al., 2019, 2021; Marshall et al., 2009; Metcalfe et al., 2017, 2018; O'Connell et al., 2007; Odoni & Lane, 2010; O'Donnell et al., 2011; Reaney, 2022, and blanket peat examples, Gao et al., 2015, 2016). The advantage of our modeling work is that it is coupled with a Before-After-Control-Intervention (BACI) experimental data set, which informs the choice of values for our (otherwise uncertain) model input parameters; not only to represent conditions "after" the interventions, but also "before" (typically only one of these periods is available). We consider this to amount to a reduction in uncertainty (and thus an increase in confidence) in our numerical results. But we also recognize that there are other sources of uncertainty that a BACI data set does not alleviate. We account for some of these uncertainties by adopting a Limits of Acceptability (LOA) approach (e.g., Liu et al., 2009). The remaining uncertainties and their implications are also discussed.

In short, this article aims to contribute to two main questions: (a) Within an uncertainty estimation framework, how can we upscale field observations of NFM to catchment scale? (b) Are the current blanket peat restoration strategies likely to continue to mitigate flooding at scales and storms large enough to improve human lives?

2. Background: Related Previous Studies

Our study substantially builds upon three preceding studies: (a) the BACI experimental study of Shuttleworth et al. (2019), (b) numerical modeling study of Goudarzi et al. (2021), and (c) the numerical model of Goudarzi et al. (2023), Generalized Multistep Dynamic (GMD) TOPMODEL; we summarize each one here, but for more details readers are referred to the respective journal articles.

(a) Shuttleworth et al. (2019) published the first set of findings from their (ongoing) BACI experiment, involving three severely eroded blanket-peat micro-catchments (~0.4–0.7 ha each) located on the Kinder Plateau (within a 3 km radius of Glossop catchment). In the "before" (the interventions) period (2010), all three micro-catchments were characterized by large areas of bare peat and deeply incised erosion gullies. During the "after" period (2012–2015), one catchment was kept unchanged as control; the second received revegetation treatment in 2012, where bare peat areas were revegetated with nurse crop (mainly amenity grasses); the third site underwent a more intense restoration involving revegetation and gully-blocking (i.e., damming the erosion gullies) in 2012, and high density planting of plugs of Sphagnum moss, in 2015. This latter treatment was designed to rapidly establish Sphagnum cover in the micro-catchment. Rainfall and discharge data were recorded separately for each site. The controlled nature of the experiment allowed for eliminating noise due to natural temporal variability resulting from inter-annual variation in synoptic hydrometeorology. The longitudinal nature of the study allows monitoring the interventions' effects across time, that is, as vegetation grows, gullies fill with trapped sediments, and

Table 1
Different Stages of Upland Peat Restoration in the Before-After-Control-Intervention Experiment

Intervention	Characteristic
Severely eroded	Presence of large areas of bare peat and deep networks of erosion gullies
Early-stage revegetation	Previously bare peat areas are now covered with young (≤2 years old) vegetation cover while gullies are still present
Early-stage reveg. + gully blocks	Previously bare peat areas are now covered with young vegetation cover while gullies are blocked with cobble stone, timber and/or peat dams
Late-stage revegetation	Previously bare peat areas are now covered with mature (8–10 years old) vegetation cover while gullies are still present
Sphagnum cover + gully blocks	Sediment has accumulated to a depth of circa 20 cm (Howson et al., 2023), equivalent to >42% of gully volume, and overall, Sphagnum covers 25% of the micro-catchment area but >85% of gully floors/walls

Note. See also Figure 1.

Sphagnum grows in the flow-lines. While data collection at these micro-catchments is still ongoing, the results of Shuttleworth et al. (2019) only included the intervention effects up until (but not including) the expansion of Sphagnum cover following growth from plug plants and natural spread from wetter areas. They found that, relative to the control, at the micro-catchment scale, and for the top 10% largest storms within the 2005–2015 period: (a) revegetation resulted in 106% increase in flood-peak delay, and 27% decrease in magnitude; (b) revegetation & gully-blocking resulted in 196% increase in delay, and 51% decrease in magnitude. Table 1 describes the different stages of restoration, which we will refer to throughout this paper.

(b) Goudarzi et al. (2021) applied the TOPMODEL hydrological model (K. J. Beven & Kirkby, 1979; Lane & Milledge, 2013) to the BACI data in order to better understand the underlying processes that drive the observed NFM effects of blanket peat restoration. Most of the family of TOPMODEL models are referred to as "semi" spatially distributed, because in order to save computation time, they lump together the parts of a catchment that tend to generate the same runoff response to the same rainfall. The lumped parts are referred to as Hydrologically Similar Units (HSU), where the "similarity" is gauged based on a catchment-specific Topographic Index (TI) value calculated for individual Digital Elevation Model (DEM) cells. The most common form of TI is: upslope contributing area per unit contour length divided by the tangent of the slope. Due to the semi-distributed nature of TOPMODEL, it is not generally possible to model spatially distributed features such as land cover change. In the case of Goudarzi et al. (2021) however, because the land cover changes were applied across the entire area of the micro-catchments, the choice of a semi-distributed model was justified. For calibration they applied the Generalized Likelihood Uncertainty Estimation (GLUE) methodology (K. Beven & Binley, 1992) to the three experimental micro-catchments, once at a "before" period (2010) and then at an "after" period (2012), for a total of six cases (i.e., six different model parameter sets/distributions). To further control for the variations in rainfall and topography, which is not possible in an experimental setup, they numerically repeated the BACI experiment but kept rainfall and topography fixed to the control site. Their study also confirmed the NFM benefits of blanket peat restoration at the micro-catchment scale. Further, through a number of parameter-switch numerical experiments they identified kinematic storage (or temporary/flow/mobile storage, see Goudarzi et al., 2021) as the primary driver of the observed NFM effects at that scale (static/immobile storage, i.e., interception/pond storage and evapotranspiration were only important in smaller storms). In this paper we use the same calibration approach to obtain model parameters pertaining to each of the cases listed in Table 1.

(c) To investigate whether the findings in (a) and (b) also hold at catchment scale, a numerical model is needed that allows modeling spatially distributed features (because interventions are unlikely to be possible/feasible everywhere at that scale). GMD-TOPMODEL (Goudarzi et al., 2023) is such a model that is built upon TOPMODEL concepts but with a number additional features/improvements. A feature particularly relevant to this study is the application of an "iso-basin" discretization layer to constrain spatially distributed information. Isobasins are smaller subcatchments within the larger catchment of interest, utilized (in this case) to enable representing different land covers in different locations within the catchment. Note that we have used a modified version of GMD TOPMODEL in which one model parameter is eliminated (namely maximum subsurface storage) by implementing a different transmissivity-depth relationship to what was originally used by Goudarzi

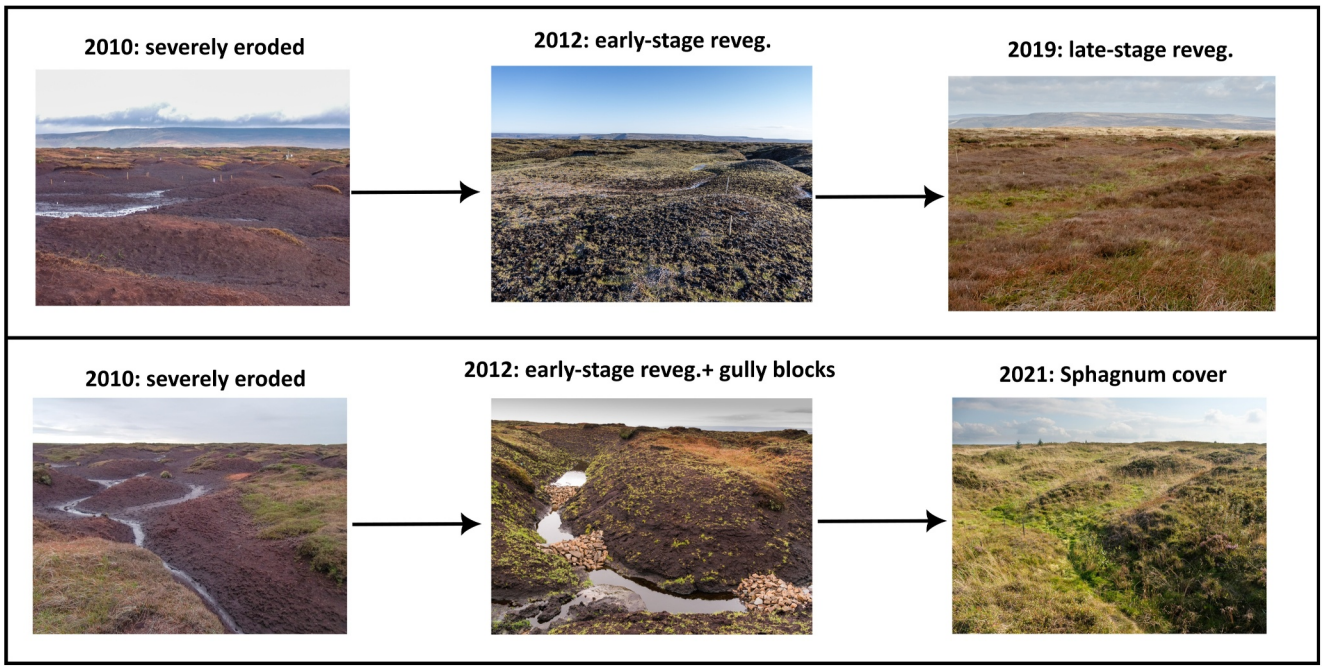

Figure 1. The different stages of blanket peat restoration at Kinder Plateau. See also Table 1.

et al. (2023). This was because we found it difficult to constrain all three parameters representing the subsurface. For reference, originally the three parameters were: exponent of transmissivity decay with depth $\hat{d}[-]$, maximum transmissivity $\hat{T}_0$ (m/s) and, maximum subsurface storage $\hat{H}_m[m]$, with the following transmissivity-depth relationship:

$$T = \hat{T}_0 \left(\frac{S_w}{\hat{H}_m} \right)^{\hat{d}} \tag{1}$$

where S_w $[m]$ is subsurface storage. The new transmissivity-depth relationship used in this study is:

$$T = \hat{T}_0 \left(\frac{1}{1 + \hat{d} \times D_{wt}} \right)^{\hat{d}} \tag{2}$$

where D_{wt} (m) is water-table depth. Note that similar to the original version, the modified version of GMD TOPMODEL is also publicly available from Goudarzi (2022). Thus, in short, in this study we use the BACI experimental data in (a) as the raw materials for representing peatland restoration by model parameters; we then apply the same calibration approach used in (b) to convert the BACI data into model parameters; we then use GMD-TOPMODEL in (c) to run a series of past and future scenarios at catchment scale.

3. Study Site: Glossop Catchment

The Glossop catchment (Figure 2) is a ~25 km^2 headwater system located on the western edge of the south Pennines, ranging from 135 to 620 m in altitude with a mean of 1,600 mm of precipitation per year. It contains more than 600 properties at flood risk (Allott et al., 2019). Floods were recorded in 1930, 1944, 1976, 1986, and 2002 at significant economic cost. The town has a flood alleviation scheme based primarily on increasing the channel conveyance, but this is limited by the infrastructural constraints within the town. Risks are expected to increase under the latest climate projections. As a result, future mitigation options will likely need to also consider interventions in the headwater catchments draining to Glososp.

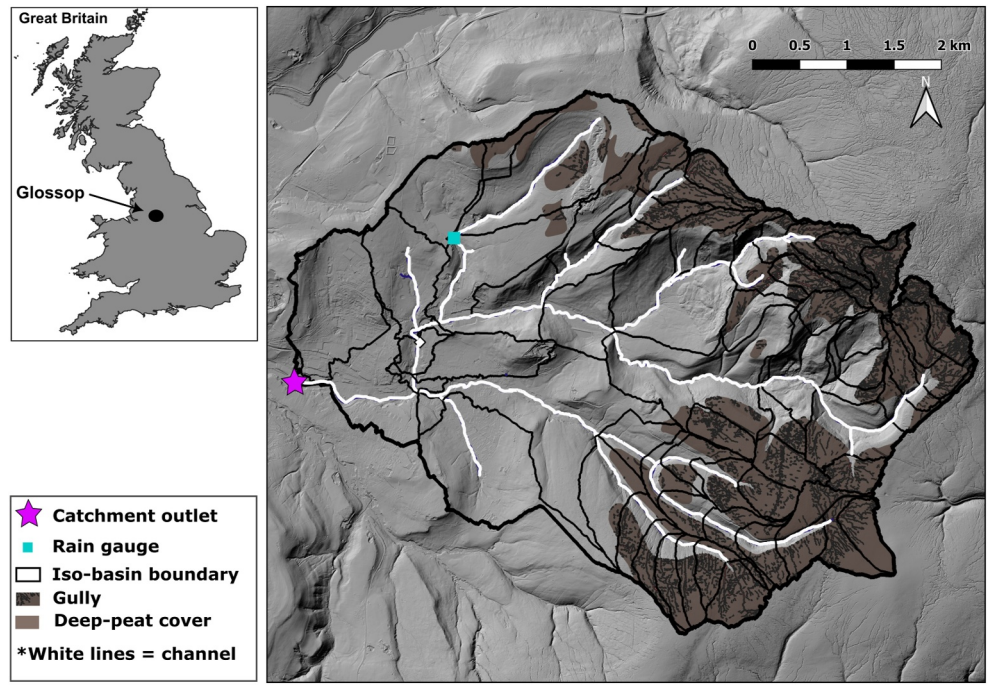

Figure 2. Glossop catchment in the western edge of the south Pennines. Total plan area = 25 km^2; upland deep peat cover = 38%.

The catchment is represented in the model through a raster DEM of 2 m resolution. First, the TI values for individual DEM cells are calculated and cells are grouped into five TI bins (five TI classes). Then the catchment is discretized into iso-basins as can be seen in Figure 2. We use smaller iso-basins in the upland deep peat (DP) areas (where peatland restoration typically occurs), and larger iso-basins in the lower parts, to allow a higher resolution where the interventions will be represented within the model. HSUs are then formed by lumping together those DEM cells that have the same TI class, and are also within the same iso-basin. To derive the channel network, we used a channel initiation threshold catchment area of 100,000 m^2. We further break up the channel network into river reaches. The break points are defined by channel confluences but also at the iso-basin boundaries, such that no river reach spans across more than one iso-basin. This resulted in 80 reaches in total. Each reach is assigned a separate HSU ID to its neighboring hillslope cells to allow independent parameter values for river and hillslope cells. Note that the catchment outlet is also defined as a separate HSU, from which the modeled hydrograph is extracted. This setup led to a total of 373 HSUs. Flow routing is based on the well-known M8 multiple flow directional algorithm which leads to a weighting matrix that assigns fractional flow from each DEM cell to its eight neighboring cells. This matrix is used to aggregate the appropriate weighting for each HSU. Given that clusters of cells form an HSU, this leads to the majority of the flow being "recycled" within the same HSU (representing the time a flow parcel spends within a given HSU), while the remainder of the flow is given to the neighboring HSUs. Being self-contained, surface flow does not occur across the iso-basin boundaries, but subsurface flow can occur across iso-basin boundaries (because subsurface flow does not necessarily experience the same topographic focusing as does surface flow; see Goudarzi et al. (2023) for more details).

A 2-month calibration period was chosen (Figure 5d), starting from 23 September 2019, and ending 23 November 2019, with a validation period (for blind-validation testing) starting 2 months prior to the start of the calibration period and up to the start of the calibration period. The choice of the calibration year (2019) was informed by a number of factors: (a) presence of large storms; (b) above zero temperatures to avoid snow-melt effects; and (c) runoff coefficients ≤ 1 (which is the portion of rainfall that leaves the system as river discharge, and thus values >1 indicate a violation the mass continuity principle). Note that the first 15 days of the record are taken to be the "spin up" period, during which we disregard the model's prediction or performance. The choice of 15 days for spin-up period was based on the time it took for catchment internal fluxes to re-balance after arbitrary assumptions regarding the antecedent conditions at $t = 0$ (i.e., assuming subsurface is full initially, see end of spin-up period in

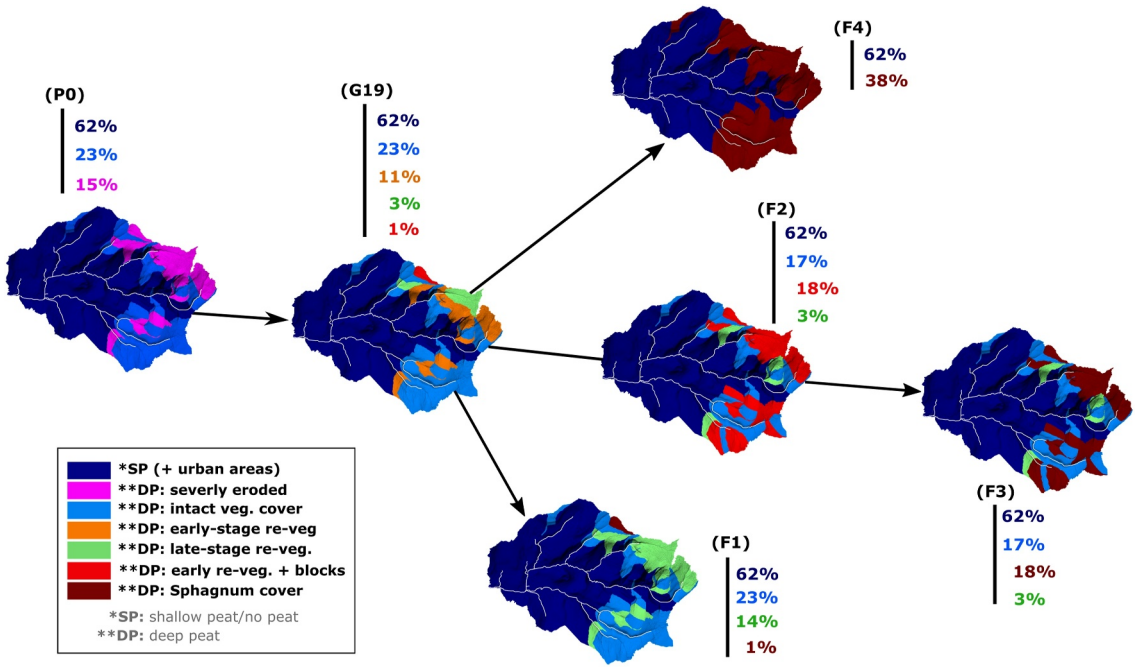

Figure 3. The scenarios considered in this study. Percentages are of the total catchment area (25 km^2).

Figure 5d showing convergence between the observation and predictions). This is to allow internal fluxes to re-balance and thus minimize the effects of unknown antecedent moisture conditions. Finally, rainfall data are from a tipping bucket rain-gauge located in the northwest of the catchment (see Figure 2), recorded every 15 min. Simultaneous water-level (stage) measurements, taken at the catchment outlet, were converted to discharge through an empirical rating curve equation; this observed discharge is the benchmark for model performance evaluations during the calibration period.

4. Scenarios

4.1. Design

We test six different scenarios as shown in Figure 3. Each scenario represents Glossop at a different time. They are: a reference scenario, G19, which represents Glossop during the calibration period in 2019; a past scenario, P0, which represents Glossop before it received any restoration treatment (i.e., pre-2000 as the first large-scale restoration initiatives date from 2003); and four future scenarios, F1–4.

Scenarios differ from one another through their iso-basin characteristics in the upland DP areas, which is why they are highlighted with different colors in different scenarios. Note that each color is also an item in Table 1, except "NDP" and "intact veg. cover." "intact veg. cover" (the light blue areas), represents areas that have not had bare peat cover in the past (and therefore have not received any restoration treatment). Vegetation surveys suggest fully grown vegetation cover but negligible Sphagnum, therefore we take "intact" areas to be the same as "late-stage re-veg." (note that similar to "late-stage re-veg.," "intact" areas can still have gullies, but none with bare peat floors).

Non-Deep-Peat (NDP) areas include a range of soil types including peaty podzols on valley slopes, gley soils in valley bottoms, and a range of land uses including improved grazing land, golf course and urban areas, none of which are targets for peatland restoration (though they may be subject to considerable changes due to grassland management or seasonality (Bond et al., 2020) and urban development, but they are outside our scope). As such, we treat NDP areas as one large zone (that still contains iso-basins of different sizes, see Figure 2). This zone (i.e., its HSUs and their parameter values), is kept as the common denominator among the six scenarios. This way any change in the catchment's runoff behavior detected by the model will be solely due to the treatments applied in the upland DP areas.

Table 2
GMD-TOPMODEL's Input (Calibration) Parameters

Parameter	Unit	Description
$\hat{d}$	–	Exponent of decay of transmissivity with depth
$\hat{T}_0$	m/s	Maximum transmissivity, that is, at the surface
$\hat{e}_p$	mm/day	Annual average daily potential evapotranspiration rate
$\hat{S}_m$	m	Maximum rootzone storage
$\hat{n}_{hs}$	s/m$^{1/3}$	Hillslope Manning's roughness coefficient
$\hat{n}_{ch}$	s/m$^{1/3}$	Channel Manning's roughness coefficient

4.2. Implementation

Each scenario is represented in the model through the parameter values assigned to each iso-basin. The modified GMD-TOPMODEL has a set of six input parameters which are listed in Table 2. To obtain reasonable values for the "sets" of parameters pertaining to each iso-basin color in Figure 3, we calibrate our model using the GLUE method (K. Beven & Binley, 1992) with LOA (Liu et al., 2009) calculated using the rating curve voting point method of McMillan and Westerberg (2015) which accounts for both aleatory and epistemic uncertainties (for more details regarding our LOA estimation refer to Goudarzi et al. (2023)). A three-step process was involved in obtaining model parameters for the six scenarios, which are as follows. Note that both steps (1) and (2) are applications of the GLUE method but to different parts of the catchment.

Step (1): aims to obtain parameter values for the micro-catchment cases listed in Table 1. To do this, following Goudarzi et al. (2021), we randomly and uniformly sample, using the Latin Hypercube method, 50,000 (50k) values for each of the parameters within the set of six (making 50k parameter-sets); then for each item in Table 1, we run GMD-TOPMODEL 50k times, we then filter model predictions through the LOAs of each item and discard the predictions that sit outside their respective LOA for more than 1% of the calibration time period. The remaining sets of parameters are considered behavioral (acceptable). Out of the 50k runs we were left with 5, 15, 7, 6, and 55 sets for the cases listed in Table 1, respectively. The ranges and distributions of these parameters, together with example hydrographs are shown in Figure 4. Note that due to complex parameter interactions, differences in distributions of a given parameter under different interventions do not necessarily reflect differences in physical processes. Therefore, attempts to interpret these shifts could lead to misleading conclusions as is shown by Goudarzi et al. (2021). Parameter distributions are shown here for reproducibility purposes only, that is, to provide ranges and locations of "good" models within the parameter space. Further, hydrographs are generated using the same "parameter-switch" experiment as in Goudarzi et al. (2021) to control for rainfall/topographic variation. For consistency across scales (i.e., hectare-scale micro-catchments to km-scale Glossop) we treat micro-catchments as hillslopes with no channels, because the channel initiation threshold of 100,000 m^2 that was used for Glossop, would yield no channel if applied to the micro-catchments. This is why there are only five parameters in Figure 4 (i.e., $\hat{n}_{ch}$ is missing).

Step (2): aims to obtain parameter values for the NDP areas, which is the common denominator in all six scenarios. To do this we first construct and then calibrate G19 to the observed discharge record in 2019 and perform the same LOA filtering as in step (1) but this time using the LOA for Glossop's outlet gauge (see Figure 5d). Thus, we discretize Glossop according to G19 in Figure 3, assigning each isobasin a class based on its intervention status; then for each of the DP iso-basins, we uniformly and randomly (using Latin Hypercube) sample 50k parameter "sets" from the values in the behavioral pools obtained in step (1) (and according to the G19 colors in Figure 3). Note that since there were less than 55 values for each parameter in the G19 pool, we take multiple samples of the same parameter values. But because different iso-basins across the catchment sit in different intervention classes, the full (catchment-wide) parameter sets will contain very few (if any) duplicates. For the NDP areas, we randomly and uniformly sample 50k values for each of the six parameters from the ranges shown by parameter distributions in Figures 5e1–5e6. Once all DP and the NDP iso-basins are allocated with their respective parameter sets, we aggregate them to make 50k G19 parameter-sets, where each set contains 30 parameters, that is, 6 for each zone in G19. We then run GMD-TOPMODEL for the 50k G19 sets and perform the LOA filtering. In this case we were left with 12 acceptable models. These acceptable models provide 12 sets of parameters for the NDP areas. Given that the NDP areas are shared between all scenarios, and that only the relative (to P0) behavior of the catchment is of interest (i.e., NDP cancels out), in all remaining simulations we represent the NDP areas with the single best performing parameter-set (with respect to the G19 calibration case).

Step (3): Once the pools of acceptable DP parameters and the single-best NDP parameter-set are known, the remaining scenarios in Figure 3 can be formed using the same sampling approach in step (2); but here we only

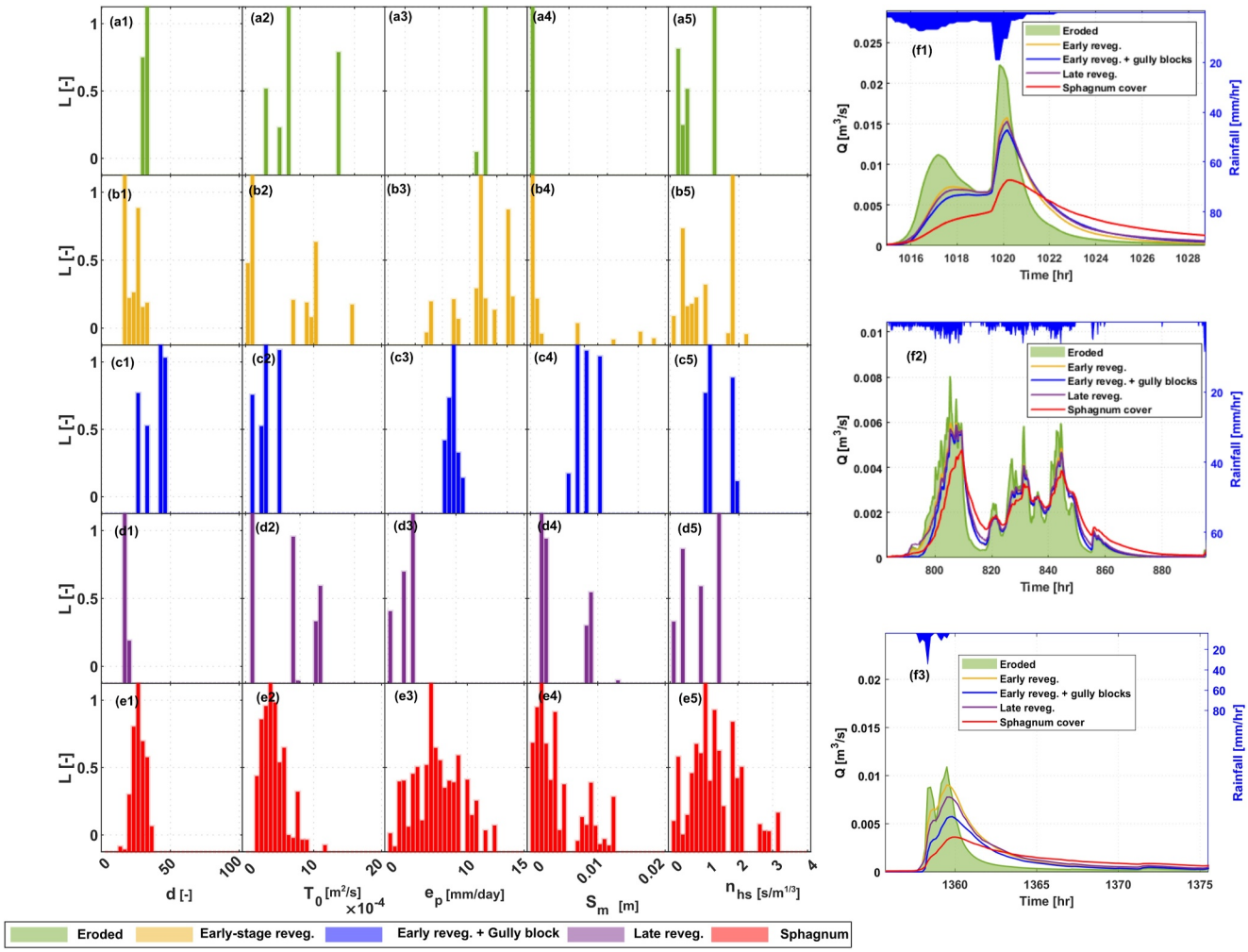

Figure 4. (a1–e5) Range and normalized likelihood (L [–]) distribution of Generalized Multistep Dynamic (GMD) TOPMODEL's parameters for each of the items in Table 1. (f1–f3) Example hydrograph comparing medians of behavioral predictions in each case.

sampled 250 parameters, since there is no further calibration involved. The aim is then to quantify the NFM impact of each scenario relative to P0.

4.3. Design Storms

To examine the different scenarios in terms of NFM impact, it is important to consider the role of storm size, because the storms in our calibration period are 2-year events (in rainfall, rather than discharge, return period terms) which are not thought to pose a flood risk. Thus, in the sections that will follow we assess the NFM impact of different scenarios for 2, 10, 100, 500, and 1,000 years rainfall events. These events were generated using Glossop's rainfall Intensity-Duration-Frequency (IDF) curve obtained from the Flood Estimation Handbook (Institute of Hydrology, 1999). We selected two storms: a long "frontal" rainfall event (Figure 5c, with continuous rainfall spanning over 19 hr), and a shorter "convectional" rainfall event (Figure 5a spanning 6.5 hr), because they generate the two largest responses in our record and represent very different hyetograph shapes. We then multiplied the total rainfall volumes across the respective durations by relevant coefficients calculated from Glossop's IDF curve. For the frontal event the rainfall volumes across a 19 hr window were: 48, 67, 105, 163, and 196 mm, which formed our 2, 10, 100, 500, and 1,000 years return periods (RPs) events, respectively. For the convectional event, rainfall volumes across a 6.5 hr window (containing two rainfall peaks) were: 22, 36, 67, 116, and 143 mm, for the same RPs.

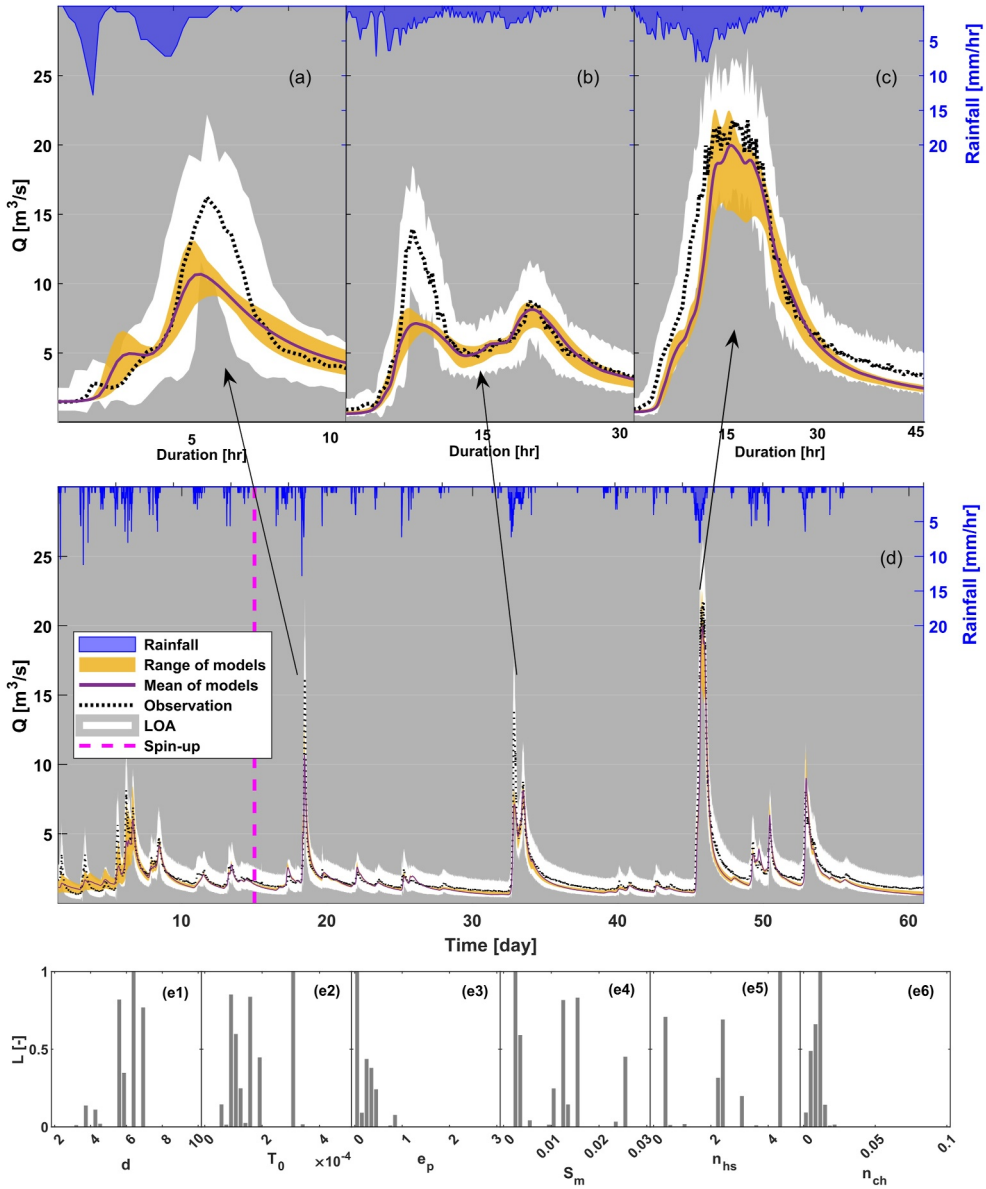

Figure 5. (d) Discharge and rainfall for the calibration period with the largest storms highlighted in (a–c). (e1–e6) Parameter ranges and distributions for the Non-Deep-Peat areas (see Table 2 for parameter descriptions/units). Number of behavioral models/parameter-sets = 12.

5. Results and Discussion

5.1. Model Calibration

Figure 5d shows both the observed and the predicted discharge time-series for the calibration period described in Section 3. Orange shaded bands represent the range of behavioral models (i.e., 12 models) and solid purple line shows the mean of behavioral models. (a–c) Highlight the three largest events in the record. Although the model remains within the LOA for >99% of the record, it provides a better fit to the largest storm in panel (c) than it does to the event in (a). Also, the first peak in (b) is underestimated consistently by all 12 behavioral models. This poor "fit" could be due to rainfall/discharge measurement errors, or the inability of the model to capture some underlying physical processes that would generate a larger first peak in (b) (i.e., model structural errors, K. Beven (2016)). It is generally difficult to identify the root cause, especially since behavioral predictions are, by definition, within the observational uncertainty (LOA). Though, the model produces very good overall fits to the

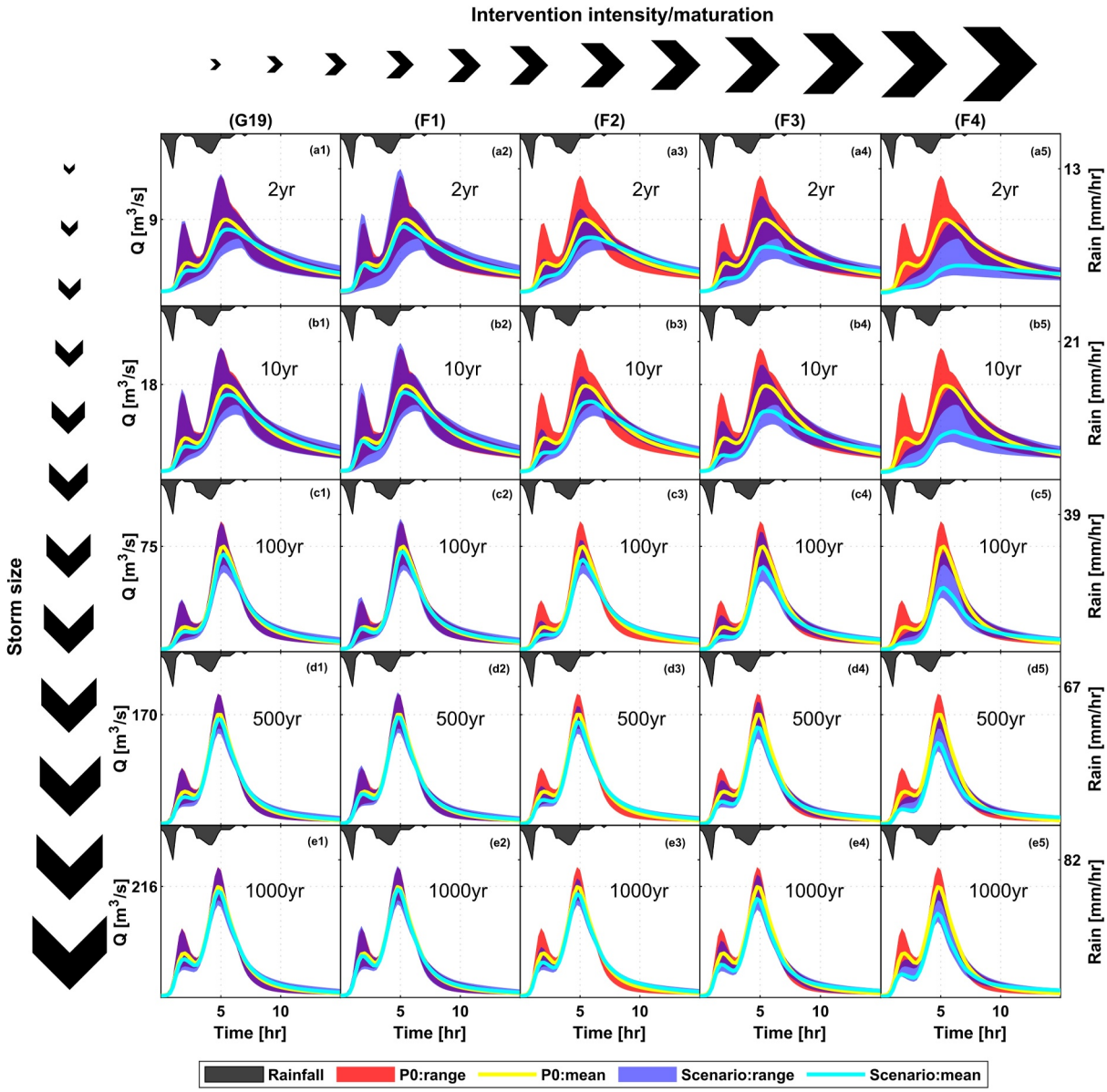

Figure 6. Hydrograph comparison between different scenarios and P0, for different return periods, and for the convectional rainfall event in Figure 5a. Note the different y-axis scales both for discharge and rainfall on different rows.

observed record according to the Kling-Gupta efficiency values of the 12 behavioral predictions which ranged from 73% to 86% in the calibration period, and 67%–82% for the validation period (not shown here).

5.2. Model Prediction

5.2.1. Hydrographs

Figure 6 shows the hydrographs extracted at the catchment outlet for different scenarios, at different rainfall return periods (RP) and for the convectional rainfall event in Figure 5a. Blue shaded areas represent the range of the 250 models in each scenario. The solid cyan line is the median value for that scenario. Shaded red bands are the range of predictions for P0, and solid yellow lines are median of P0 predictions.

Down the columns, hydrographs tend to preserve their double-peak shape as events become larger, with the exception of F4 for RP < 500 years where the smaller first peak is not present. As events become larger,

hydrograph peaks occur earlier in the record, which is expected because depth-dependant flow velocities implemented in the model will produce faster flows with increasing flow thickness, which is proportional to storm size.

Across the rows, as interventions intensify/mature, peak reductions and delays tend to increase, though less clearly for delay which also has some exceptions (i.e., delay decreasing with increasing intervention intensity/maturation). For example, F1's peak-flow for RP = 10 years clearly comes before that of P0, while this is not the case for G19 for RP = 10 years. Aside from such exceptions which may indicate an interference in flow timing of the different parts of the catchment (i.e., stream re-/de-synchronization as discussed in Section 5.2.3), the general pattern of increased peak reduction/delay with intervention intensity/maturation is expected because the same behavior has been observed at the micro-catchment scale, both experimentally (Shuttleworth et al., 2019) and numerically (see Figures 4f1–4f3, also Goudarzi et al. (2021)).

Figure 7 is similar to Figure 6 but for the frontal rainfall event in Figure 5c. The same general observations for the convectional rainfall case can also be made for the case of frontal rainfall: across the rows, as interventions intensify/mature both peak reductions and delays tend to increase. Down the columns, hydrographs tend to have broader peaks at shorter return periods (RP) but as events become larger, hydrograph shapes tend to become more similar which is in line with the findings of Mindham et al. (2023), who observed a convergence in the characteristics of runoff response of UK upland catchments during larger storms.

5.2.2. Likelihood of Meaningful NFM Benefit at 25 km^2

The boxplots in Figures 8a1–8a5 show the interquartile ranges of "all" possible differences between peak magnitudes of the 250 P0 (worst case) models and the 250 predictions for the rest of the scenarios (separately for each scenario), and for the convectional rainfall event in Figure 5a. Thus, each boxplot contains 62,500 (250 × 250) datapoints. In each panel, x-axis is return period and y-axis is peak magnitude change at the catchment outlet relative to P0. In (a1), which compares G19 (the calibration period, representing the "current" state of Glossop) with P0 (the "past" scenario representing Glossop at its most eroded state), error bars tend to shrink with increasing return period (RP) values. This indicates that the same P0 and G19 models that deviated much more from one another at shorter RPs, tend to deviate much less at longer RPs. The same behavior can be observed for other scenarios, a2–a5. This is consistent with our expectation and may be attributed to the threshold behavior of UK upland catchments reported by Edokpa et al. (2022). They found that catchment storage, connectivity and antecedent conditions control small discharge peaks but become increasingly irrelevant for larger storms, during which surface flow attenuation dominates. Similarly, through comparing hydrographs across 17 UK upland micro-basins over a 4-year period, Mindham et al. (2023) found that at smaller rainfall events there was a higher variance in hydrograph shape quantifiers and that these variances decrease as storms became larger.

Thus, the reduction in variability with increasing storm size is, at least partly, due to the reduction in the number of physical processes that remain relevant (i.e., can impact the runoff response) as rainfall intensifies. The same analysis may also explain the diminishing peak reduction effect with increasing RP values reflected in the median values (dots within each boxplot), which is true in all cases (a1–a5). In other words, peak-flow attenuation diminishes, because (a) the (perceived) fixed potency of the "active" process-drivers of attenuation naturally decreases with increasing rainfall volumes; (b) some of those process-drivers simply turn "passive" due to a threshold behavior (Edokpa et al., 2022). For example, during a storm, once catchment static storage (interception + ponding + evapotranspiration) is overwhelmed, kinematic storage (or flow storage Goudarzi et al., 2021) becomes the only major contributor to its response to the incoming precipitation volumes.

Comparing different interventions, in (a1–a5), median peak reductions tend to be higher for more mature/intense interventions, which is expected. Figures 8b1–8b5 shows the likelihood contours of different peak reduction levels for each scenario. The y-axes of these panels show the percentage of the 62,500 datapoints that sit below a specific ΔQ level, as denoted by the legend in (b5). According to (b1), for RP = 2 years, it is 62% likely that G19 will deliver peak reductions greater than 5%, relative to P0. Similarly, the likelihood of >5% peak reduction, $L(\Delta Q > 5\%)$, for G19 is 60%, 58%, 44%, and 42%, for RP = 10, 100, 500, and 1,000 years, respectively. Table 3 lists the same likelihoods for different cases but only for flood-relevant events (RP ≥ 10 years). In all cases in Figures 8b1–8b5, overall, likelihoods of meaningful NFM impact (here taken to be at least 5% peak reduction) goes down with event size as expected. Also, likelihoods of meaningful impact tend to increase with intervention intensity/maturation, with the exception of F1 which exhibits a lower

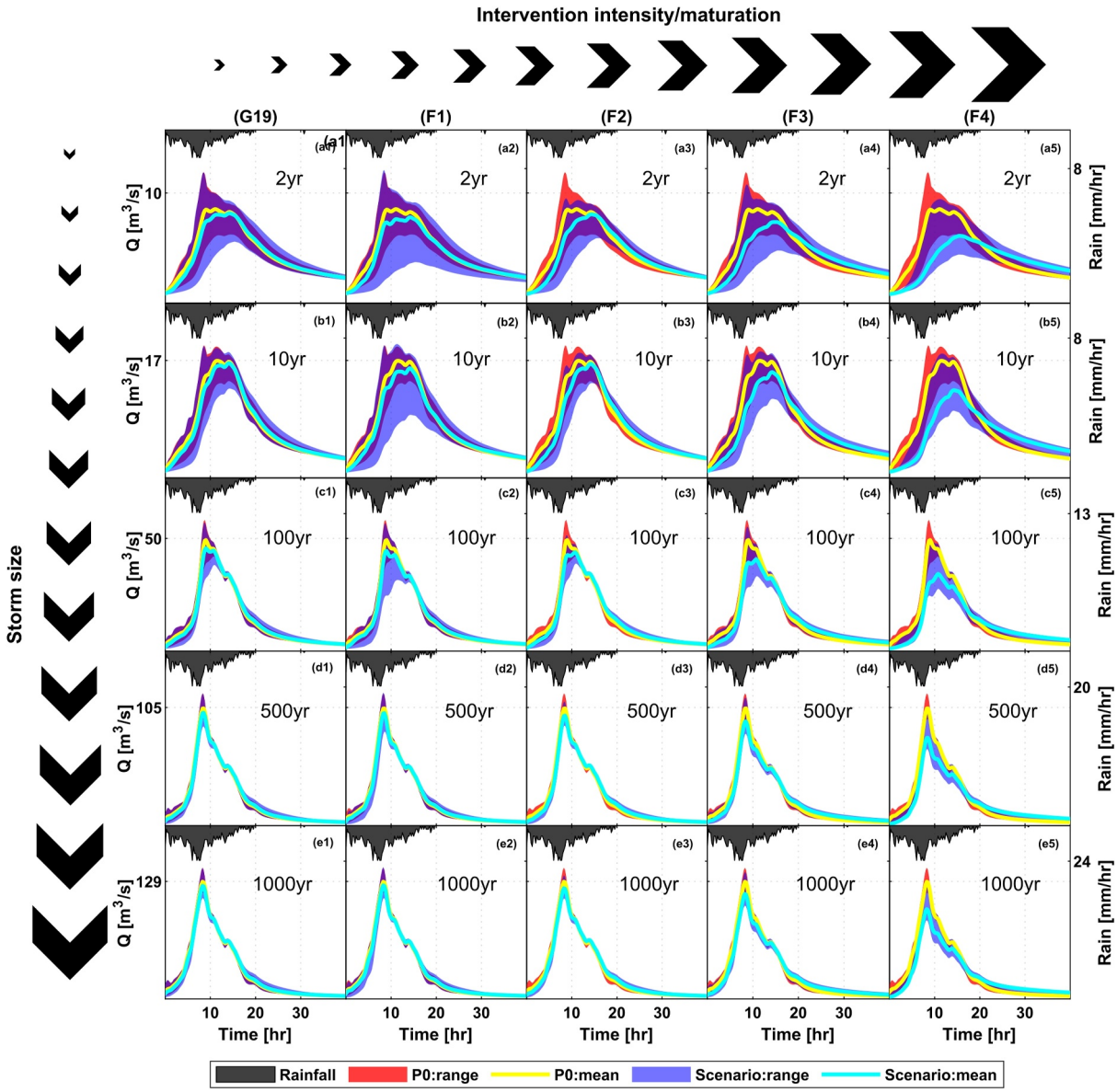

Figure 7. Hydrograph comparison between different scenarios and P0, for different return periods. This figure is similar to Figure 6 but for the frontal rainfall event in Figure 5c.

likelihood of meaningful impact compared to G19 (see also Table 3). Given that at micro-catchment scale (see Figure 4) F1 ("Late-stage reveg.") reduced peak discharge slightly more than G19 (which is mostly "Early-stage reveg."), it is perhaps surprising that at catchment scale F1 results in lower likelihood of meaningful NFM impact. This suggests that the intervention in F1, while likely locally attenuating the flow more, it is also re-synchronizing the stream network. Equally it is possible that G19 may be de-synchronizing the stream network leading to the anomalous behavior (where G19 is more effective than F1 against expectation). The possibility of re-/de-synchronization will be further examined in Section 5.2.3.

Panels (c1–c5) are the same (a1–a5) but their y-axes show lagtimes at the catchment outlet relative to P0. Also, similar to (b1–b5), (d1–d5) show the likelihood contours but for different lagtime levels as denoted by the legend in (d5). Similar to peak magnitudes in (a1–a5), lagtimes in (c1–c5) exhibit the same patterns: (a) lagtimes tend to decrease with storm size, (b) variability in lagtimes tends to diminish with storm size, and (b) lagtimes tend to be longer for more intense/mature interventions. A couple of noteworthy observations can be made from the likelihood plots in (b1–b5) and (d1–d5). First, the likelihoods of positive lag at the outlet tends to be much lower than

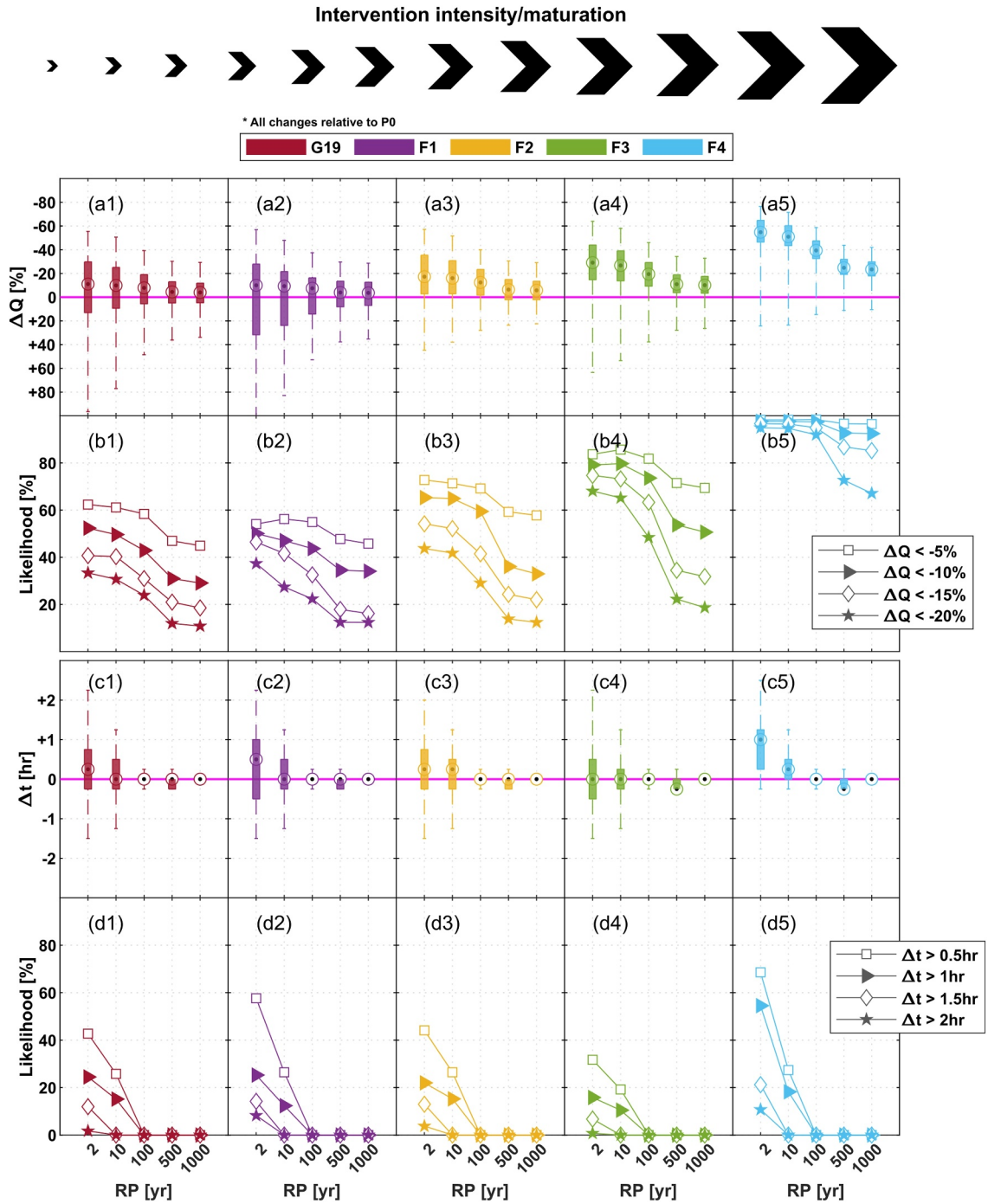

Figure 8. Each boxplot marks the interquartile ranges of "all" possible differences between the 250 P0 models and the 250 predictions for scenarios, that is, each boxplot contains 62,500 datapoints. (a1–a5) Relative (to P0) change in peak-flow magnitude for different scenarios and for the convectional rainfall event in Figure 5a. (b1–b5) The likelihood contours representing the percentage of datapoints that sit below specific contour levels in terms of ΔQ. (c1–c5) Similar to (a1)–(a5) but for change in peak-flow timing. (d1–d5) Similar to (b1)–(b5), but for peak-flow timing.

the likelihood of positive peak reduction (or negative ΔQ in (a1–a5)) for the same scenario/RP pairs (compare with b1–b5 with d1–d5). For example, in G19 for RP = 2 years, $L(\Delta Q > 5\%)$ is 67% while $L(\Delta t > 0.5$ hr) is around 42%; for RP = 500 years, $L(\Delta Q > 5\%)$ is 40% while $L(\Delta t > 0.5$ hr) is zero (similar pattern can be seen for the rest of the scenarios). Second, once events are larger than RP = 10 years, the likelihood of positive lag at the outlet, and for the convectional event, seems to be zero irrespective of the scenario. However, when considering that

Table 3
Likelihoods of Meaningful Natural Flood Management Impact of Different Blanket Peat Restoration Scenarios at the 25 km² Scale and for Flood-Relevant Events

	Convectional rainfall				Frontal rainfall			
	10 years	100 years	500 years	1,000 years	10 years	100 years	500 years	1,000 years
	G19 (Int. area = 15%)				G19 (Int. area = 15%)			
L(ΔQ > 5%)	61%	50%	40%	31%	40%	52%	45%	43%
L(ΔQ > 10%)	58%	43%	31%	24%	16%	37%	28%	26%
L(ΔQ > 15%)	47%	31%	21%	12%	16%	28%	17%	14%
L(ΔQ > 20%)	45%	31%	18%	11%	16%	20%	7%	4%
	F1 (Int. area = 15%)				F1 (Int. area = 15%)			
L(ΔQ > 5%)	56%	47%	42%	27%	34%	47%	49%	49%
L(ΔQ > 10%)	55%	44%	32%	22%	20%	38%	36%	26%
L(ΔQ > 15%)	48%	35%	18%	12%	20%	37%	23%	19%
L(ΔQ > 20%)	46%	34%	16%	11%	20%	31%	13%	8%
	F2 (Int. area = 21%)				F2 (Int. area = 21%)			
L(ΔQ > 5%)	71%	65%	52%	42%	54%	63%	57%	52%
L(ΔQ > 10%)	69%	59%	41%	29%	29%	50%	31%	30%
L(ΔQ > 15%)	59%	36%	24%	14%	16%	35%	32%	16%
L(ΔQ > 20%)	58%	33%	22%	12%	15%	22%	9%	5%
	F3 (Int. area = 21%)				F3 (Int. area = 21%)			
L(ΔQ > 5%)	86%	80%	73%	65%	73%	81%	73%	71%
L(ΔQ > 10%)	82%	74%	63%	48%	62%	69%	55%	51%
L(ΔQ > 15%)	72%	54%	34%	22%	46%	56%	34%	29%
L(ΔQ > 20%)	69%	51%	32%	19%	27%	43%	19%	15%
	F4 (Int. area = 38%)				F4 (Int. area = 38%)			
L(ΔQ > 5%)	98%	98%	96%	95%	90%	95%	95%	95%
L(ΔQ > 10%)	98%	97%	95%	92%	85%	92%	92%	91%
L(ΔQ > 15%)	97%	93%	87%	73%	77%	87%	85%	83%
L(ΔQ > 20%)	96%	92%	85%	67%	71%	82%	74%	67%

G19	Early-stage reveg.
F1	Late-stage reveg.
F2	Early-reveg. + gully blocks
F3	Sphagnum + gully blocks
F4	Sphagnum + gully blocks

Note. L($\Delta Q > X\%$) = the likelihood that an intervention reduces peak magnitude at catchment outlet by X% relative to P0 (the "Eroded" state of Glossop), and for different rainfall return periods and types. "Int. area" = percentage of catchment area intervened.

significant peak reduction at the outlet still occurs at RPs > 10 years (see a1–a5) despite the lack of delay, these results suggest that: in order to reduce peak magnitudes at the outlet, it is not necessary to delay the timing of peak-flow, instead it can be sufficient to defer only a portion of the flow to the falling limb of the hydrograph (see the stretched falling limbs in Figure 6), even if that portion is much smaller than what would be needed to delay the hydrograph peak itself.

Figure 9 is the same as Figure 8, except that it is for the frontal rainfall event in Figure 5c. The general observations that were true for the convectional rainfall are also true for the frontal rainfall: (a) variation decreases with storm size in all cases; (b) peak reduction and delay tend to increase with intervention intensity/maturation; (c) all scenarios tend to produce zero-delay at the outlet once RP ≥ 500 years (note that this threshold was RP ≥ 100 years for the convectional event which is expected given the higher rainfall intensity of convectional

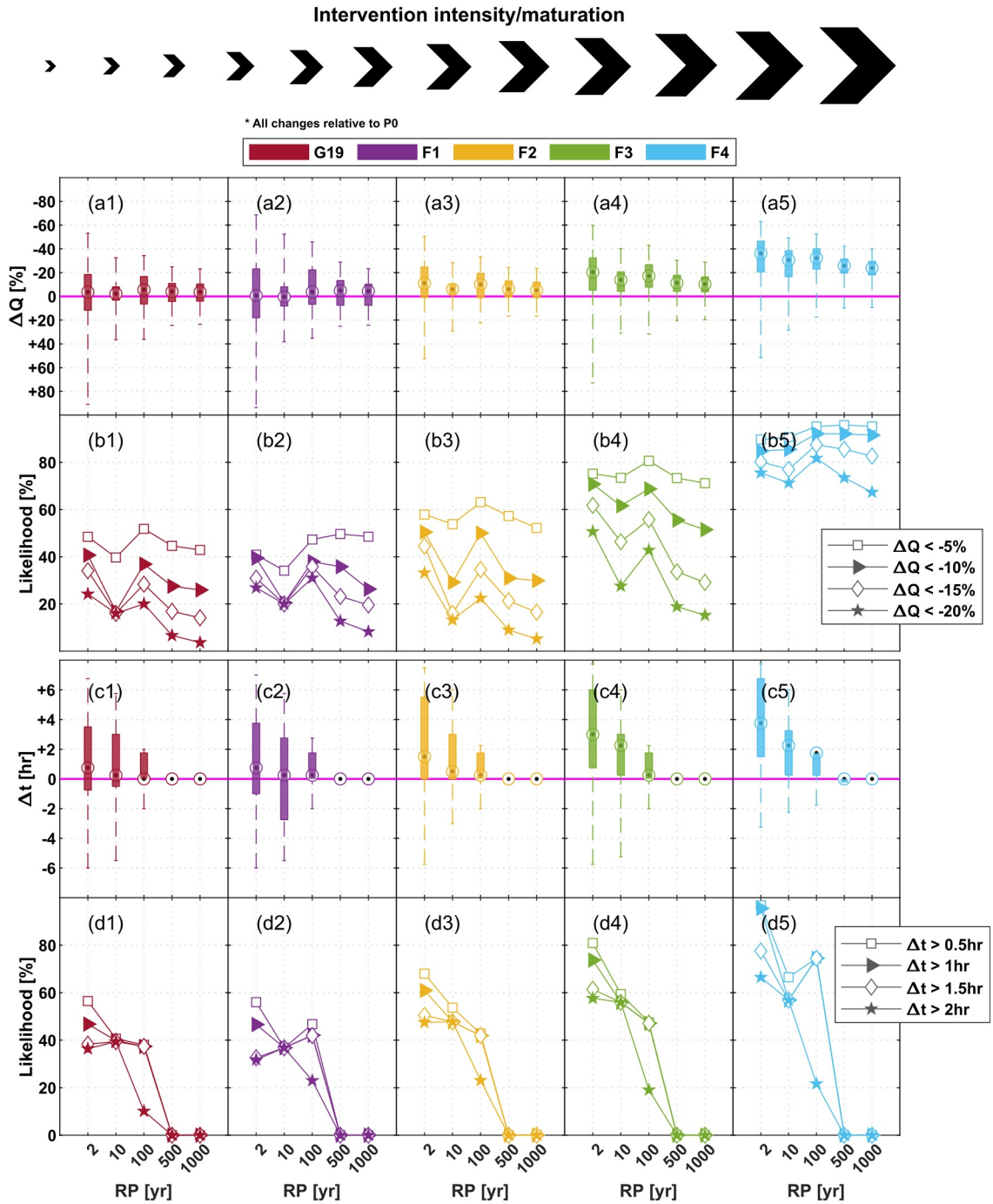

Figure 9. Each boxplot marks the interquartile ranges of "all" possible differences between the 250 P0 models and the 250 predictions for scenarios, that is, each boxplot contains 62,500 datapoints. (a1–a5) Relative (to P0) change (negative = attenuation, and vice versa) in peak-flow magnitude for different scenarios and for the frontal rainfall event in Figure 5c. (b1–b5) The likelihood contours representing the percentage of datapoints that sit below specific contour levels in terms of ΔQ. (c1–c5) Similar to (a1)–(a5) but for change in peak-flow timing. (d1–d5) Similar to (b1)–(b5), but for peak-flow timing.

events); (d) peak reduction and delay tend to decrease with storm size, with one exception. This exception occurs in all scenarios (see b1–b5) going from RP = 10 years to RP = 100 years, where the likelihoods of peak reduction (as well as median peak reductions in a1–a5) increase with storm size, which is contrary to our expectation. Note that similar to the convectional storm, G19 seems to perform better than F1 also during the frontal storm, albeit not for all RP values (see b1–b2, d1–d2, and Table 3).

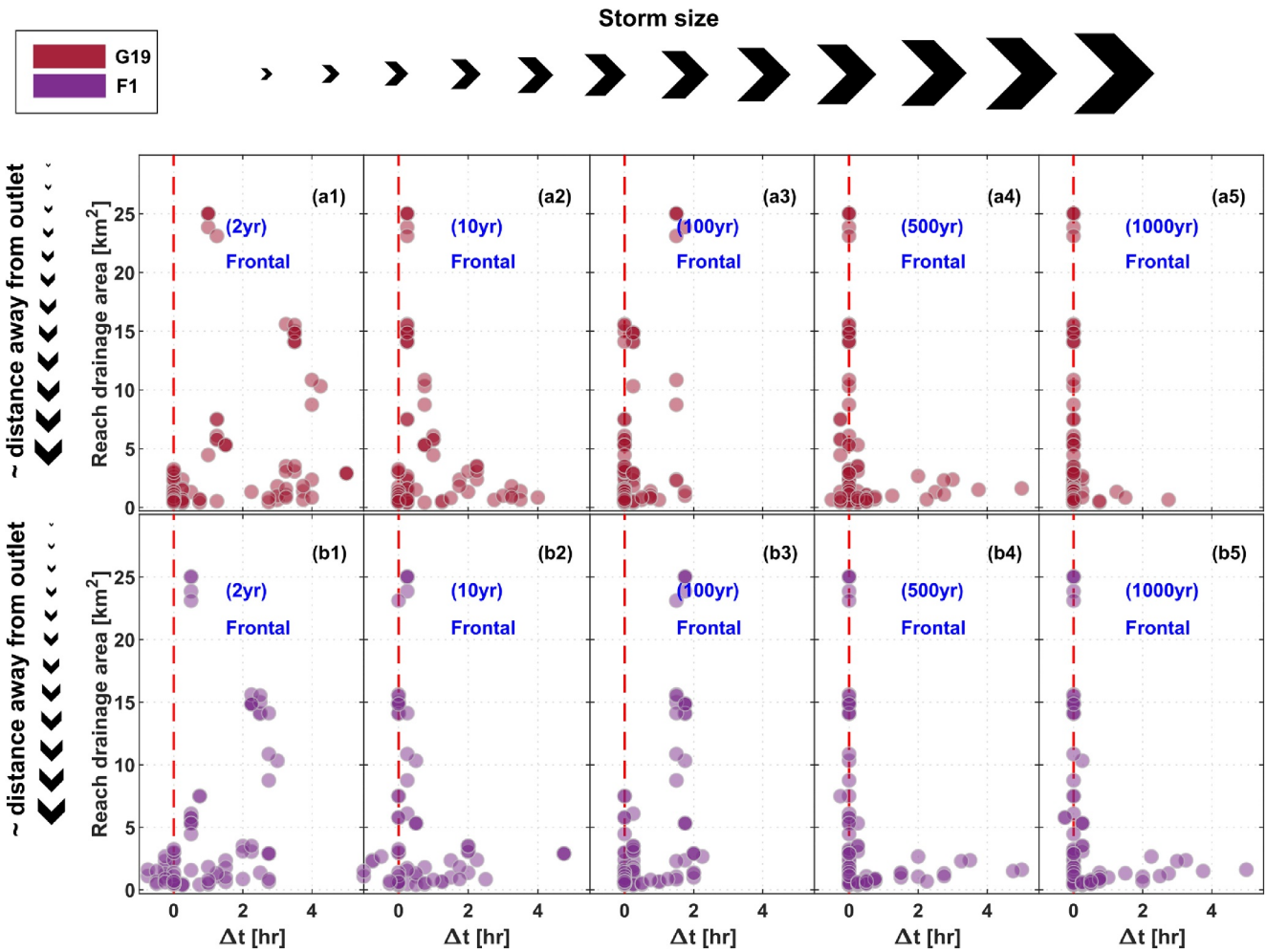

Figure 10. (a1–a5) G19's reach-by-reach lagtimes (relative to P0) (x-axis) for different return periods and for the frontal rainfall event in Figure 5c, versus reach drainage area (y-axis). Note that these Δt values are extracted from one (out of 250) of G19 models for demonstration, that is, different models/parameter-sets will likely show different timing distributions along the river network. (b1–b5) The same as (a1)–(a5) but for F1 scenario.

Thus our results show two instances of unexpected behavior: (a) one that seems to be storm-dependent, that is, the "jump" in performance of all scenarios going from RP = 10 years to RP = 100 years in the frontal event (see Figures 9b1–9b5), which is absent during the convectional event (see Figures 8b1–8b5); (b) one that seems to be scenario-dependent, that is, G19 outperforming F1. Note that the outperformance of G19 over F1 seems to vanish with storm size, and more so in the case of the frontal event (which can be seen more clearly in Table 3). Therefore (b) is neither entirely independent of storm-size, nor on storm type, but the fact that it only occurs between G19 and F1 suggests that it is predominantly due to the characteristics of one or both of those scenarios. We will examine both (a) and (b) in Section 5.2.3.

5.2.3. Stream Re-/De-Synchronization

If stream re-/de-synchronization is responsible for the two unexpected behaviors noted in the previous section, its signal is to be found in flow timing along the stream network. In Figure 10, the x-axes shows reach-by-reach lagtimes for G19 (a1–a5) and for F1 (b1–b5), both for the frontal rainfall case. The y-axes show the size of the drainage area of individual reaches, which corresponds (though not strictly) to distance from outlet. Note that in generating this figure we have selected one model (out of the 250 models) for demonstration purposes. But to ensure that these plots do not represent the extremes in our range of model behaviors, for each scenario we have picked the model with the closest outlet discharge to the median of all 250 models (i.e., dots in Figures 9a1–9a5). Also note that in these plots, longer delay for a particular reach, in and of itself is neither good nor bad with

regards to NFM performance at the outlet; what is more important is the timing sequence amongst the adjacent reaches. Further, it is important to note that while we take drainage area as a proxy for distance from the outlet, the dots sitting within the same size class (i.e., closer together in the y-axis) do not necessarily drain to one another, which makes it difficult to accurately assess the timing sequence of a point-cluster within a size class. It is generally difficult to properly visualize flow timings of the different catchment parts, but we hope that the overall pattern of points presented here can help provide at least some indication regarding the occurrence (or lack thereof) of re-/de-synchronization.

To examine the unexpected jump in performance of all scenarios going from RP = 10–100 years under the frontal rainfall (Figures 9b1–9b5), that is, stream de-synchronization, first consider panels (a2 and a3) in Figure 10. While subcatchments furthest from the outlet tend to have a shorter delay under the 100 years (a3) event than under the 10 years event (a2) (i.e., as expected), larger subcatchments further downstream (in particular two out of the three subcatchments around the 10 km^2 mark) experience a longer lagtime under the 100 years event than under the 10 years event. In turn, this seems to have slightly de-synchronized the larger subcatchments around the 15 km^2 mark during the 100 years event. Continuing downstream, the rather small segregation at the 15 km^2 mark seems to have translated into a large delay near the outlet under the 100 years event, compared to the 10 years event. Thus, while it is difficult to fully make sense of the timing signal at the upper end of the stream network (where there are many smaller subcatchments), the interactions there seem to have resulted in de-synchronization further downstream, and we suspect that this is the main cause for the unexpected boost in performance going from a 10 years event to a 100 years event. A similar explanation can be applied to F1 (compare b2 and b3).

To examine the unexpected outperformance of G19 over F1, that is, re-synchronization, under frontal rainfall and for RP ≤ 10 years (note that the threshold for re-synchronization was RP ≤ 100 years under convectional rainfall which is not demonstrated here), we compare Figure 10a1 with Figure 10b1. It can be seen that, under F1 (panel b1), many of the smaller subcatchments peak even before P0 (hence the negative lag), while this is not the case under G19. The same is true at RP = 10 years (see a2 and b2). Thus, it seems that earlier arrival of peak flows in the smaller subcatchments further from the outlet under F1 could be responsible for its slightly worse performance relative to G19.

Thus, according to our limited numerical evidence, unintended stream re-synchronization due to changes to the landcover via restoration seems possible. Applying our modeling approach to a much larger number of storm shapes, sequencing patterns (and directions in the case of convectional rainfall) as well as catchments would be needed to quantify the true likelihood of such phenomenon. Though, in any case this is unlikely to be of concern in large events during which this effect seems to vanish (see Table 3).

5.2.4. Role of Storm Characteristics

Figure 11 presents multiple data dimensions which are as follows. In (a1), the y-axis shows the G19 lagtimes relative to P0, which were extracted from the mean of all 250 G19 hydrographs, and the mean of the 250 P0 hydrographs (i.e., not "all" possible differences as in Figure 8), and for the convectional rainfall event. Similarly, (a2–a5) are for F1–F4 scenarios. Each dot represents an individual river reach within the model (of which there are 80 in total including the outlet, see the white lines in Figure 2). The x-axes of all panels are percentage of the drainage basin of each reach that is altered through an intervention. The size of each dot is proportional to the drainage area of that reach, while colors (green to purple) represent rainfall return periods (RP), 2 to 1,000 years. Solid lines are the linear regression lines for different RP levels (distinguished by line colors); the associated linear correlation coefficients (R) are shown in the top left corner of each panel. (b1–b5) Are the same as (a1–a5) but for the frontal rainfall case. Finally, Figure 12 shows the same multi-dimensional data set but for the case of peak-reduction (on the y-axes).

According to Figures 11a1–11b5 under both storm types, overall, lagtimes tend to decrease with catchment area (dot sizes) and storm size (colors green to purple) but increase with intervention intensity/maturation (across the rows) as well as with percentage intervention area (x-axes). Except under smaller (RP ≤ 100 years) frontal events and for interventions not involving Sphagnum, (i.e., G19 to F2, b1–b3) where smaller intervention areas seem to experience a longer lag (indicated also by the slopes of the linear regressions). This unexpected behavior seems to be storm-related since it is not observed under convectional rainfall (a1–a3). The dependence of these

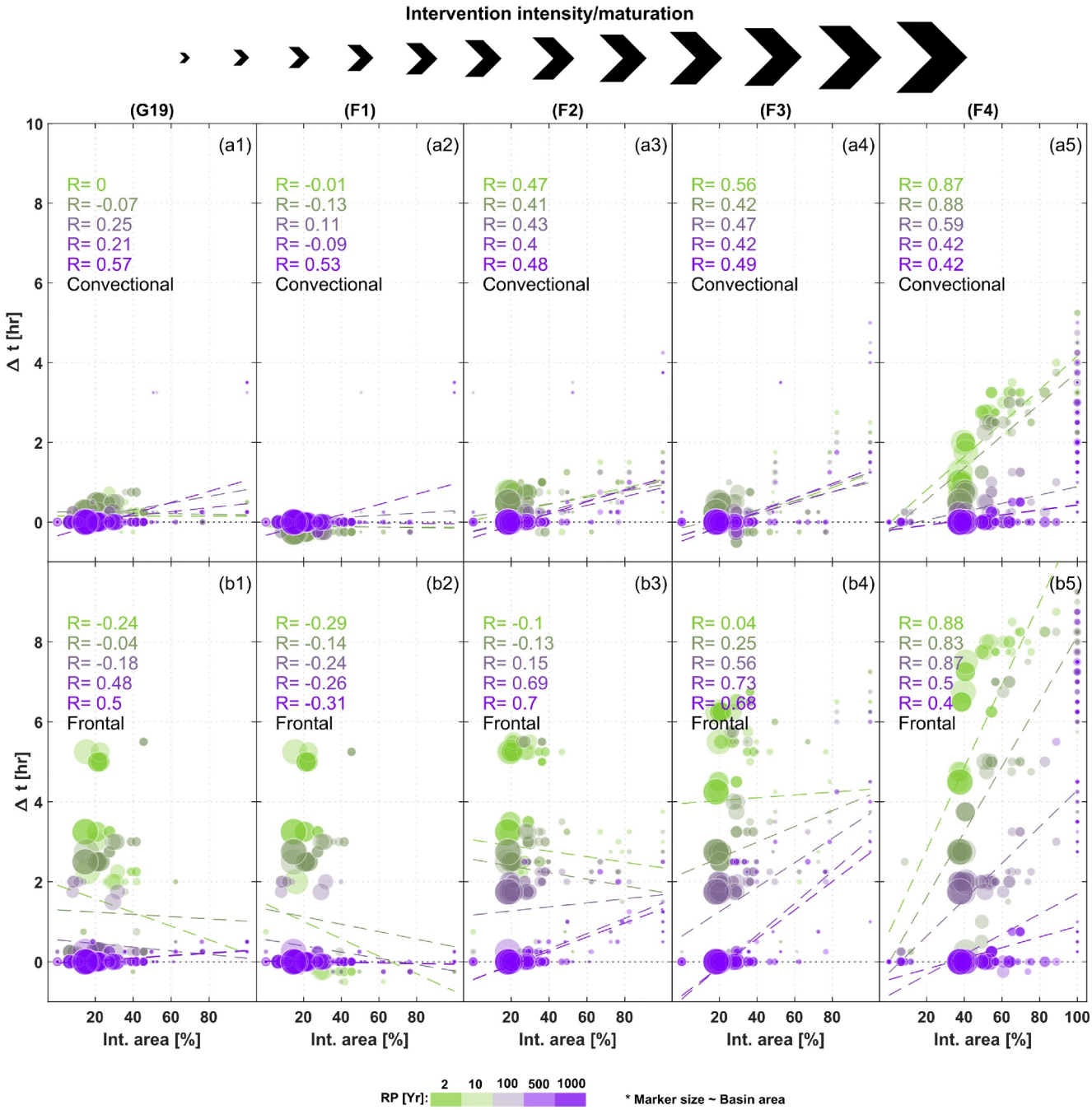

Figure 11. (a1–a5) Reach-by-reach relative (to P0) change in peak-flow timing for different scenarios and return periods (y-axis) for the convectional rainfall event in Figure 5a, versus percentage intervention area (x-axis). Each dot represents a subcatchments that drains to a river reach, of which there are 80 in the model. Size of the dots approximates subcatchment size, for example, largest dot = the entire Glossop catchment, or 25 km^2, and smallest dot represents a 40 ha subcatchment. (b1–b5) Are similar to (a1)–(a5), but for the case of frontal rainfall event in Figure 5c.

interventions' behavior on storm characteristics can also be viewed from a different angle. Catchments of all sizes, but especially larger catchments tend to experience longer lagtimes delays under frontal rainfall than under convectional rainfall (a1–a5), though only for RP ≤ 100 years, beyond which lagtimes delays of larger catchments tend to zero under both storm types. Surprisingly, however, considering the associated peak reductions in Figure 12, interventions tend to reduce peak-flow magnitudes more under convectional rainfall than under frontal rainfall (i.e., the opposite of the delay behavior). Also, at RPs > 100 years, despite significant peak reductions,

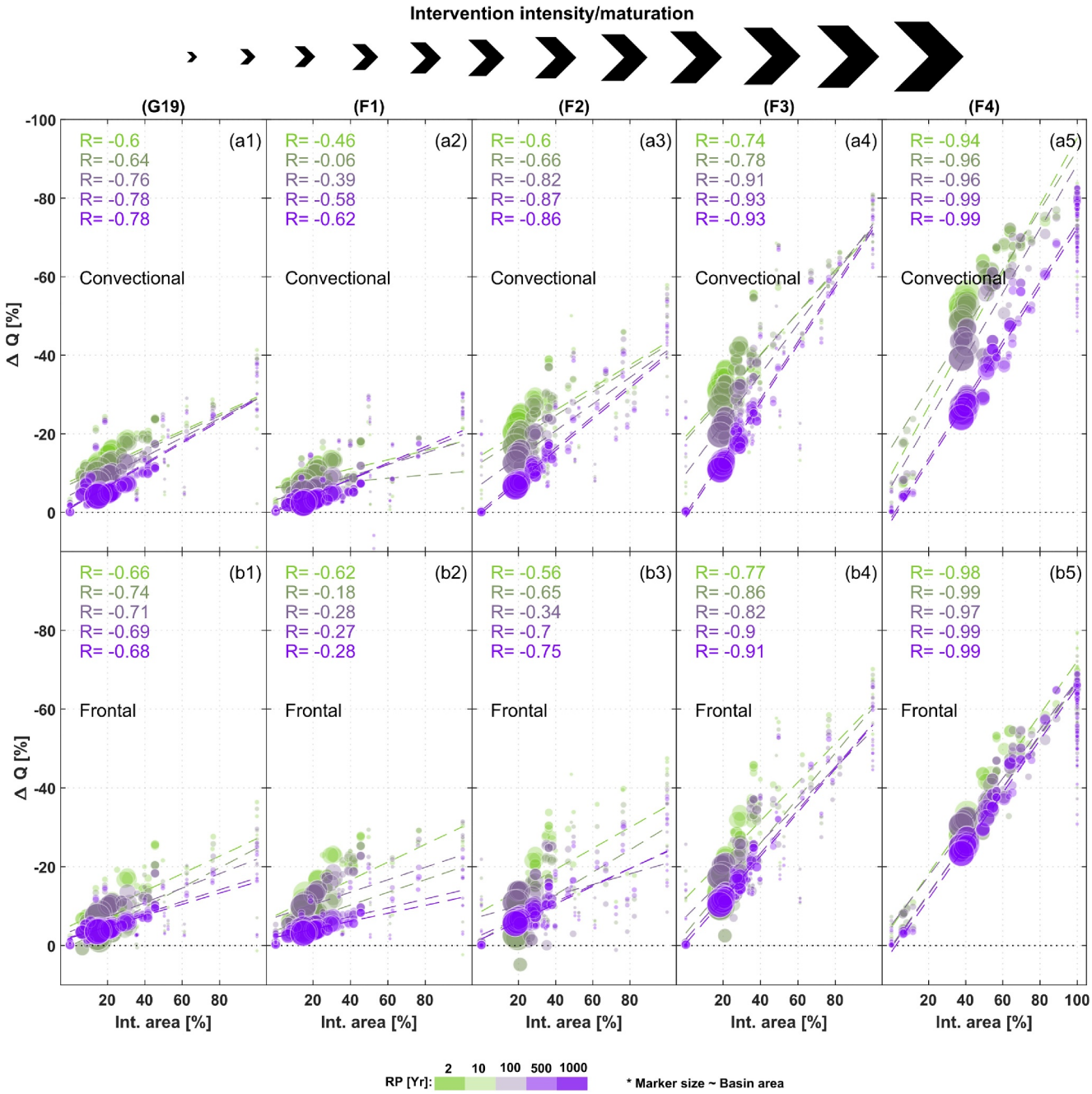

Figure 12. (a1–a5) Reach-by reach relative (to P0) change in peak-flow magnitude for different scenarios and return periods (y-axis) for the convectional rainfall event in Figure 5a, versus percentage intervention area (x-axis). Each dot represents a subcatchments that drains to a river reach, of which there are 80 in the model. Size of the dots approximates subcatchment size, for example, largest dot = the entire Glossop catchment, of 25 km^2, and the smallest dot represents a 40 ha subcatchment. (b1–b5) Are similar to (a1)–(a5), but for the case of the frontal rainfall event in Figure 5c.

lagtime delays are predominantly zero, except at (or near) the 100% intervention level (which tends to represent endmember isobasins located at the tips of the river network in Figure 2).

The above observations may seem to suggest that lagtime (or kinematic storage, which attenuates storm flow, both reducing peak discharge and increasing lagtime (Goudarzi et al., 2021)) is not a major contributor to peak reduction at catchment-scale since the two do not seem to correlate for larger dot sizes. But kinematic storage with its associated increased lagtimes has already been shown to be the primary mechanism of peak reduction at smaller scales and under flood-relevant events (Goudarzi et al., 2021; Shuttleworth et al., 2019) and we expect this

to remain the case at larger scales. Note that the endmember isobasins (those near 100% intervention area) are by far more representative of the conditions in the hectare-scale experimental study of Shuttleworth et al. (2019), and the subsequent numerical study of Goudarzi et al. (2021). It is therefore not surprising that interventions in the endmember isobasins behave similarly (i.e., maintain a positive lag even at larger events) to those studied in isolated micro-catchments nearby. It is notable that this is the case despite the significant difference in size (the smallest dot in these plots represents a 40 ha catchment, which is at least 50 times larger than the size of the experimental micro-catchments mentioned earlier).

We therefore suggest that even though it may not be as obvious, the same mechanism (of peak-reduction via delay) is likely to dominate also in larger catchments (which typically have a less than 100% intervention area); the difference being that interference from the un-intervened parts most likely dilutes the timing signal in a way that peak reduction at the outlet appears to not have a delay component or basis (i.e., despite significant peak reductions, lagtime delays are predominantly zero at intervention areas $\leq$40% and at RP > 100 years). This points to the same conclusion as in Section 5.2.2, that at large-enough scales/events at which flooding can impact communities, it is not necessary, nor is it feasible given the flow volumes (see Figures 6 and 7), to delay the flood wave in order to meaningfully attenuate it at the outlet. Instead, it seems sufficient to defer only a portion of the flood volume to the receding limb, that is, store water in catchment's kinematic storage unit (Goudarzi et al., 2021).

On the basis of the central role of lagtime in peak reduction in large events and catchments, the aforementioned unexpected behavior of the interventions under different storm types (i.e., more delay under frontal rainfall, but more peak reduction under convectional rain for RP $\leq$ 100 years) may be explained as follows. Precipitation volumes being the same, a higher-intensity event (i.e., convectional) will render a more "pointy" hydrograph shape (see Figure 6), whereas a longer-lasting event (i.e., frontal) will give rise to a more "blunted" shape (see Figure 7). We therefore suggest that it is perhaps more useful to assess the significance of a given lagtime relative to some measure of peak flow duration (e.g., the duration of discharge/stage >X% of the maximum/minimum). This is because it seems sensible for the effectiveness of a given lagtime delay magnitude to be inversely proportional to the duration of peak flow that the delay is trying to attenuate, that is, by holding water back in the kinematic storage long enough until peak flow has passed.

In this context, the lagtimes produced by these interventions under frontal-type events, while longer in absolute terms, they would likely be shorter if adjusted for peak flow duration. Regardless, such difference in behavior under the different storm types tend to vanish for RPs > 100 years, where even a frontal event renders a "pointy" hydrograph shape, just as a convectional event would (compare Figures 6 and 7, peak flow durations are no longer very different at large events). This would explain why for RPs $\leq$ 100 years, the likelihoods of meaningful peak reduction at Glossop's outlet are higher under convectional rainfall than under frontal events for smaller sizes, but it is not for RPs > 100 years, where in fact the reverse tends to be the case (see Table 3).

These results therefore suggest that longer lasting frontal-type events are likely to be more challenging to defend against via NFM for medium to large storms (RP10–100 years), which is consistent with the findings of Gao et al. (2018) who showed that events with late rainfall-peaks tend to produce the largest discharge peaks at the outlet. Our results however also suggest (see Table 3) that for large to very large storms (RP = 100–1,000 years) convectional-type storms are likely to be more difficult to defend against. Though, importantly, even for RP = 1,000 years the likelihoods of meaningful NFM impact is quite high for scenarios involving Sphagnum (F3 and F4), and certainly considerable for other scenarios (see Table 3).

Finally, note that there is a caveat regarding the modeled differences between the two storm types; relative to catchment orientation, the direction that a convectional rain cell moves may play a role in stream re-/de-synchronization, whereas this is less likely to be a factor for a frontal type event, at least at the scales considered here (i.e., 25 km^2). Nevertheless, in this study we have assumed that rainfall has been uniform across the catchment under both storm types. Thus, while this makes our findings independent of storm direction, it might also make them less generalizable when it comes to the differences between the two storm type (but not with respect to storm sizes that we consider).

5.2.5. Interplay Between Catchment Size and Percentage Restoration Area

As expected, and as can be seen in Figure 12, NFM benefit of all scenarios tends to increase with increasing restoration area (as a percentage of total catchment area). Note that dots should not be compared like-for-like because it is not possible from these plots to distinguish between the different reaches that may have similar percentage restoration/drainage areas. We have used linear regression to show the trends, but note that most of the regression lines do not pass through the origin, as if zero intervention would not result in zero ΔQ. Thus, while there are clear trends with increasing restoration area, the trends are not necessarily linear. Although not shown here, we also fitted the data with exponential functions of the form $\Delta Q = a(Int.\ area)^b$. We found that the relationships were more nonlinear for smaller events, but they were almost completely linear for larger events (this can also be gauged by how far the y-intercept of the lines deviate from zero for the smaller events but not for larger events). This is consistent with the findings of Edokpa et al. (2022) and Mindham et al. (2023).

According to Figure 12, if 100% restoration were to be feasible, depending on the restoration strategy anywhere between 15% and 80% peak reduction could be possible. Though, at the 100% intervention level, the sub-catchments are not very large (smaller than 240 ha), meaning that it is not clear from these plots whether similar NFM benefits would have resulted if the catchments at the 100% level were larger. Note that Shuttleworth et al. (2019) reported a 27% median peak reduction following early-stage revegetation (i.e., G19) in a 0.7 ha experimental catchment, and a 51% median peak reduction following additional gully-blocking (i.e., F2) in an adjacent catchment of roughly the same size. Considering our model predictions for these two interventions and at the 100% level (see a1/b1 and a3/b3), it is interesting that the observed median values sit well within our range of predictions. Further, the dots occupying the 100% level for these two interventions range between 40 and 85 ha in catchment area, that is, anywhere between 40 and 120 times the size of the experimental micro-catchments.

Together, the last two points suggest that catchment size may not be a deciding factor on NFM performance as long as these interventions are applied everywhere across the catchment. Once again considering the central role of kinematic storage (or lagtime delay), such scale-intensiveness at the 100% level could make sense because if flows are delayed everywhere, lagtimes at the outlet will simply increase with catchment area in such a way as to offset some/most of the increased area effect. In fact, maximum relative (to the eroded case) lagtime delay reported by Shuttleworth et al. (2019) at the experimental sites was 80 min for G19 and 110 min for F2. Our results in Figures 11a1, 11b1 and 11a3, 11b3 show lagtimes of up to ca. 4 hr for both interventions (i.e., up to 2–3 times longer lag at scales 40 to 120 times larger).

In practice, percentage restoration area tends to decrease with catchment size, because the larger the catchment, (a) the smaller the fraction of its area that is likely to be covered by DP and thus suitable for restoration; and (b) the higher the financial cost of achieving the same percentage restoration. Thus, while interesting in terms of process understanding, the results at >40% intervention level are unlikely to be relevant for catchment-scale NFM. Further, a caveat of the reach-by-reach plots in both Figures 11 and 12 is that they show ΔQ and Δt values that were extracted from the mean of the 250 models (per scenario). So while they likely provide sensible "mean" predictions, they are not associated with a likelihood, as in Figures 8 and 9, nor do they reflect the full ranges of possible ΔQ and Δt (e.g., the full range will most likely show maximum delays greater than 4 hr). Avoiding these limitations would have required generating these plots with "all possible" differences between P0 and individual scenarios and on a reach-by-reach basis. Here we deemed it sufficient to use the mean of all models to examine the trends and patterns.

5.2.6. Role of Sphagnum

So far we have compared our scenarios (Figure 3) to P0, that is, the worst/eroded case. For G19/F1, this is essentially the same as estimating the NFM benefit of revegetating bare peat (see Table 3) but for F2 this indicates the combined effect of revegetation and gully-blocking. Similarly, for F3 it calculates the combined effect of gully-blocking and Sphagnum. From a decision making perspective it would be valuable to be able to isolate the effect of each restoration technique. For example, comparing F2 to G19 (rather than P0) would have estimated the impact purely due to gully-blocking, but restored areas are not the same (15% in G19 whereas 21% in F2, see Figure 3) which would confound such comparison. This, however, is not the case between F2 and F3 where the restored areas are identical and a comparison between the two would indicate

the Sphagnum effect (but relative to gully-blocking + revegetation rather than gully-blocking alone). In any case gully-blocking is unlikely to be implemented on bare peat without revegetation thus separating their impacts is perhaps less important, whereas it has often been implemented without Sphagnum planting in the first instance, making it more important from a decision making perspective to be able to isolate Sphagnum's contribution to NFM. Further, Holden et al. (2008) showed from plot studies that Sphagnum cover was very effective at slowing overland flow velocity on blanket peat, so understanding how this effect upscales is important also from a process understanding perspective.

Based on the median values (dots) in Figure 8a3 (for F2), at the 25 km^2 scale, revegetating bare-peat areas while also blocking the erosion gullies using timber and/or cobble stone dams in ∼18% of the catchment (+ and extra 3% revegetation without gully-blocking, see Figure 3) leads to median peak reductions of 16%, 12.4%, 6.4% and 5.8% under the convectional events of RP = 10, 100, 500 and 1,000 years, respectively; and from Figure 9a3, 6%, 10.1%, 6.1% and 5.2% under the frontal event of the same RPs. On the other hand, additional Sphagnum planting in the same areas (i.e., F3) leads to median peak reductions of 26.8%, 19.5%, 11%, and 10.2% under the convectional event, and 13.9%, 17.2%, 11.4%, and 10.3% under the frontal event. Accordingly, in peak-reduction terms, the increase in gully-blocking's effectiveness due to Sphagnum can be anywhere between 57% and 132% (by comparing median values like-for-like). Performing a similar exercise but for the likelihood values in Table 3 suggests that, additional Sphagnum planting can increase the likelihood of >5% peak reduction by 15%–23%; >10% by 13%–33%; >15% by 8%–30%; and >20% by 7%–21%, all of which are significant improvements particularly if taken to be as indications of the likelihood of flood mitigation/prevention.

It may be worth mentioning that gully-blocking activities considered here have been motivated by, and designed for, moorland restoration rather than for NFM. Though yet to be demonstrated, it is conceivable that gully-block permeabilities may be altered to convert some of their static storage to kinematic storage during low flows, to allow static storage to be utilized more effectively at high flows (Goudarzi et al., 2021). Tighter control on block permeability could be engineered by, for example, inserting a pipe of a specific diameter, and above a certain height from the gully floor. Thus, under better designs, the NFM performance of the gully-block/Sphagnum system can potentially be improved even further.

Regardless, the gully-block designs in our study are typical for current restoration activities in blanket peat, and a 57%–132% boost in their ability to reduce peak flows, as well as a 7%–33% increase in the likelihood that they might successfully mitigate/prevent flooding, just by adding Sphagnum to the current designs is notable. In our view Sphagnum planting should therefore be widely adopted as an important component of peatland restoration which delivers significant NFM benefit in addition to its role as a keystone species providing a number of other important ecosystem benefits (Rochefort, 2000; Verhoeven & Liefveld, 1997) in bog ecosystems (although there may be locations where other species could be favored, perhaps in places where there is no palaeo evidence of Sphagnum presence during the peatland's development). This is particularly the case given that the proportion of upland peatlands that are characterized by bare peat is relatively small (circa 1%), reducing the opportunity for NFM through revegetation. Reduced Sphagnum (or other keystone species) coverage is however a common condition across much of the uplands of UK due to historic industrial pollution and land management practices for example, Blundell and Holden (2015), Lee et al. (1987), and Mackay and Tallis (1996). There is therefore a significant opportunity to bring NFM benefits to headwater communities through Sphagnum re-establishment in headwater catchments.

6. Uncertainties, Limitations, and Future Work

(a) The degree to which our results might be model-dependent is unknown, because we did not test our scenarios using different types of rainfall-runoff models, although the assumptions underlying GMD-TOPMODEL (see Goudarzi et al., 2023) fit the case of blanket peat catchments well. Importantly also, our impact estimations are of a similar magnitude to those of Gao et al. (2016), who modeled the impact of riparian Sphagnum planting on flood peaks in a 11.4 km^2 UK blanket peat catchment. For a 10 years event, and with 20% (of catchment's area) Sphagnum cover in the headwaters, they estimated a 13% flood peak reduction at the catchment outlet. In our case, for a 10 years event and with a total of 21% upland intervention area (but 18% Sphagnum cover, see Figure 3), we find 26.7% median peak reduction during a convectional-type event (Figure 8a4) and 13.9% median peak reduction under a frontal-type storm (Figure 9a4). This is significant because the estimates provided by Gao et al. (2018) are for similar percentage intervention but at a

different catchment, using different storms and spatial resolution, and using a fundamentally different (i.e., fully distributed with a stochastic surface flow routing) hydrological model as well as modeling approach. This provides some confidence that our results are likely to not be severely affected by our choice of model. (b) It is typically not possible to account for all observational uncertainties (e.g., we have not accounted for possible under-catch and v-notch blockage), nor is it possible to account for them with absolute accuracy (i.e., there is paradoxically uncertainty in the way we account for uncertainties). We have sought to account for this fact by retaining models even when their predictions fall outside our LOA for up to 1% of the time. (c) To obtain model parameters for different restoration strategies, we calibrated our model to experimental blanket peat micro-catchments near Glossop (3 km away, on the Kinder plateau). In doing so we have assumed that porting model parameters between the sites (of different topographies/flow pathways) is acceptable. The basis of such assumption is more the current lack of a better alternative, rather than rigorous testing. Although both sites being part of the same stretch of blanket peat cover provides some justification for the above assumption, nevertheless this is an unquantified source of uncertainty. (d) Finally, the degree to which our model's hydrograph-fitting capability (and therefore calibration) is sensitive to the choice of calibration period/size is unclear. Although the model's good performance in split-sample testing provides encouragement that such uncertainties may be fairly constrained, different storm shapes, sequencing patterns, and travel directions could potentially yield very different results. Further, spatial variability in rainfall intensity-duration characteristics could also play an important role, particularly in larger catchments. These are important topics that future research should explore.

7. Conclusions

We have shown that by introducing iso-basins into a semi-distributed hydrological model it is possible to constrain spatially distributed information such as landcover. Our approach to upscale hectare-scale experimental data sets to kilometer-scale using a hydrological model and, importantly, within an uncertainty estimation framework is novel. We considered the 25 km^2 Glossop catchment with 600 properties at flood-risk, and under flood-relevant storms of 10–1,000 years return period (RP).

Our results demonstrate that NFM through moorland restoration can be an effective strategy in mitigating flooding even at scales and storms large enough to impact communities. Our results show that blanket peat restoration activities attenuate peak discharge primarily through increasing the catchment's kinematic storage. Importantly, at catchment scale and during large storms, it seems not necessary (nor is it feasible) to delay the flood wave in order to meaningfully attenuate it at the outlet. Instead, it is likely sufficient to defer only a portion of the flood volume to the receding limb and into the kinematic storage unit.

Our limited numerical evidence suggests that unintended stream re-synchronization due to changes to landcover may be possible though not for large storms of concern to NFM. The NFM benefit of the restoration strategies considered here seems to increase exponentially with intervention area under smaller storms, but this relationship is quite linear for larger storms. For medium to large storms (RP10–100 years), longer lasting frontal-type events are likely to be more challenging to defend against via NFM, but there seems to be a threshold behavior which shifts that trend at large to very large storms (RP = 100–1,000 years) where shorter convectional-type events become more challenging for NFM.

We find that at the 25 km^2 scale and under storms of 10–1,000 years RP, revegetating bare-peat in ∼15% of the catchment leads to a median peak reduction of 2.2%–9.9% (depending on storm size), with a 31%–61% likelihood of >5% peak reduction (again depending on storm size). Revegetating bare peat areas while also blocking the erosion gullies using timber and/or cobble stone dams in ∼20% of the catchment results in a median peak reduction of 5.8%–16% with a 42%–71% likelihood of >5% peak reduction. Additional Sphagnum planting in the same areas (i.e., ∼20% of the catchment) increases the median peak reductions to 10.2%–26.8% while it increases the likelihood of >5% peak reduction to 65%–86%. Additional Sphagnum planting in 40% of the catchment increases the median peak reductions to 23.5%–51% while it increases the likelihood of >5% peak reduction to 90%–98%. Accordingly, adding Sphagnum to existing gully-blocking activities can improve their NFM performance anywhere between 57% and 132% while increasing the likelihood that they might prevent flooding by 7%–33%.

The clear numerical evidence of significant NFM benefits associated with the re-establishment of Sphagnum in headwater peatlands is an important finding with implications for peatland restoration. In the UK less than 1% of

peatland area remains un-vegetated but limited or absent coverage of Sphagnum (or other keystone species) is a common condition in degraded bogs. When also considering its wider ecosystem benefits, our findings mean that meaningful flood risk mitigation in headwater catchments at scales relevant to communities at risk can be delivered alongside these co-benefits. Sphagnum re-establishment is therefore likely to be a "no regrets" restoration strategy.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The hydrological data were collected as part of the Making Space for Water funded by DEFRA and the UK Environment Agency, and PROTECT NFM project funded by NERC, which are included in this paper: Shuttleworth et al. (2019). The numerical model (coded in MATLAB) is available from Goudarzi (2022).

Acknowledgments
This research was conducted as part of the NERC PROTECT-NFM project (NE/R004587) with a project steering committee including representatives from: Environment Agency, IUCN UK Peatland Programme, Moors for the Future Partnership, National Trust, Natural Resources Wales and SEPA. In addition to UK NERC funding, Joseph Holden has also received funding for peatland research in the last 5 years from Defra, Yorkshire Water, The National Trust, Forest Research, Research England, Peak District National Park Authority, North York Moors National Park Authority, the Environment Agency and the EU's Horizon 2020 programme.

References

Acreman, M., & Holden, J. (2013). How wetlands affect floods. *Wetlands*, *33*(5), 773–786. https://doi.org/10.1007/s13157-013-0473-2
Addy, S., & Wilkinson, M. E. (2019). Representing natural and artificial in-channel large wood in numerical hydraulic and hydrological models. *Wiley Interdisciplinary Reviews: Water*, *6*(6), e1389. https://doi.org/10.1002/wat2.1389
Allott, T., Auñón, J., Dunn, C., Evans, M., Labadz, J., Lunt, P., et al. (2019). Peatland catchments and natural flood management.
Bark, R. H., Martin-Ortega, J., & Waylen, K. A. (2021). Stakeholders' views on natural flood management: Implications for the nature-based solutions paradigm shift? *Environmental Science & Policy*, *115*, 91–98. https://doi.org/10.1016/j.envsci.2020.10.018
Beven, K. (2016). Facets of uncertainty: Epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal*, *61*(9), 1652–1665. https://doi.org/10.1080/02626667.2015.1031761
Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305
Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, *24*(1), 43–69. https://doi.org/10.1080/02626667909491834
Blundell, A., & Holden, J. (2015). Using palaeoecology to support blanket peatland management. *Ecological Indicators*, *49*, 110–120. https://doi.org/10.1016/j.ecolind.2014.10.006
Bond, S., Kirkby, M. J., Johnston, J., Crowle, A., & Holden, J. (2020). Seasonal vegetation and management influence overland flow velocity and roughness in upland grasslands. *Hydrological Processes*, *34*(18), 3777–3791. https://doi.org/10.1002/hyp.13842
Bonn, A., Allott, T., Evans, M., Joosten, H., & Stoneman, R. (Eds.) (2016). *Peatland restoration and ecosystem services: Science, policy and practice*. Cambridge University Press.
Bonn, A., Allott, T., Hubacek, K., & Stewart, J. (2009). *Drivers of environmental change in uplands*. Routledge.
Chimner, R. A., Cooper, D. J., Wurster, F. C., & Rochefort, L. (2017). An overview of peatland restoration in North America: Where are we after 25 years? *Restoration Ecology*, *25*(2), 283–292. https://doi.org/10.1111/rec.12434
Cohen-Shacham, E., Walters, G., Janzen, C., & Maginnis, S. (2016). *Nature-based solutions to address global societal challenges* (p. 97). IUCN. 2016-2036.
Dadson, S. J., Hall, J. W., Murgatroyd, A., Acreman, M., Bates, P., Beven, K., et al. (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. https://doi.org/10.1098/rspa.2016.0706
Dixon, S. J., Sear, D. A., Odoni, N. A., Sykes, T., & Lane, S. N. (2016). The effects of river restoration on catchment scale flood risk and flood hydrology. *Earth Surface Processes and Landforms*, *41*(7), 997–1008. https://doi.org/10.1002/esp.3919
Edokpa, D., Milledge, D., Allott, T., Holden, J., Shuttleworth, E., Kay, M., et al. (2022). Rainfall intensity and catchment size control storm runoff in a gullied blanket peatland. *Journal of Hydrology*, *609*, 127688. https://doi.org/10.1016/j.jhydrol.2022.127688
Evans, M., Allott, T., Holden, J., Flitcroft, C., & Bonn, A. (2005). Understanding gully blocking in deep peat. Moors for the Future Report, 4.
Evans, M., & Warburton, J. (2011). *Geomorphology of upland peat: Erosion, form and landscape change*. John Wiley & Sons.
Evans, M., Warburton, J., & Yang, J. (2006). Eroding blanket peat catchments: Global and local implications of upland organic sediment budgets. *Geomorphology*, *79*(1–2), 45–57. https://doi.org/10.1016/j.geomorph.2005.09.015
Evans, M. G., Burt, T. P., Holden, J., & Adamson, J. K. (1999). Runoff generation and water table fluctuations in blanket peat: Evidence from UK data spanning the dry summer of 1995. *Journal of Hydrology*, *221*(3–4), 141–160. https://doi.org/10.1016/s0022-1694(99)00085-2
Ferguson, C. R., & Fenner, R. A. (2020). The potential for natural flood management to maintain free discharge at urban drainage outfalls. *Journal of Flood Risk Management*, *13*(3), e12617. https://doi.org/10.1111/jfr3.12617
Gao, J., Holden, J., & Kirkby, M. (2015). A distributed TOPMODEL for modelling impacts of land-cover change on river flow in upland peatland catchments. *Hydrological Processes*, *29*(13), 2867–2879. https://doi.org/10.1002/hyp.10408
Gao, J., Holden, J., & Kirkby, M. (2016). The impact of land-cover change on flood peaks in peatland basins. *Water Resources Research*, *52*(5), 3477–3492. https://doi.org/10.1002/2015wr017667
Gao, J., Kirkby, M., & Holden, J. (2018). The effect of interactions between rainfall patterns and land-cover change on flood peaks in upland peatlands. *Journal of Hydrology*, *567*, 546–559. https://doi.org/10.1016/j.jhydrol.2018.10.039
Goudarzi, S. (2022). Generalised multistep dynamic TOPMODEL in MATLAB [Software and Data]. Retrieved from https://zenodo.org/badge/latestdoi/514527559
Goudarzi, S., Milledge, D., & Holden, J. (2023). A generalized multistep dynamic (GMD) topmodel. *Water Resources Research*, *59*(1), e2022WR032198. https://doi.org/10.1029/2022wr032198

Goudarzi, S., Milledge, D. G., Holden, J., Evans, M. G., Allott, T. E., Shuttleworth, E. L., et al. (2021). Blanket peat restoration: Numerical study of the underlying processes delivering natural flood management benefits. *Water Resources Research*, *57*(4), e2020WR029209. https://doi.org/10.1029/2020wr029209

Hankin, B., Metcalfe, P., Beven, K., & Chappell, N. A. (2019). Integration of hillslope hydrology and 2D hydraulic modelling for natural flood management. *Hydrology Research*, *50*(6), 1535–1548. https://doi.org/10.2166/nh.2019.150

Hankin, B., Page, T. J., Chappell, N. A., Beven, K. J., Smith, P. J., Kretzschmar, A., & Lamb, R. (2021). Using micro-catchment experiments for multi-local scale modelling of nature-based solutions. *Hydrological Processes*, *35*(11), e14418. https://doi.org/10.1002/hyp.14418

Holden, J. (2005). Peatland hydrology and carbon release: Why small-scale process matters. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *363*(1837), 2891–2913. https://doi.org/10.1098/rsta.2005.1671

Holden, J. (2009). Peatland hydrology and carbon release. In A. Bonn, T. Allott, K. Hubacek, & J. Stewart (Eds.), *Drivers of environmental change in uplands* (pp. 113–134). Routledge.

Holden, J., & Burt, T. P. (2002). Infiltration, runoff and sediment production in blanket peat catchments: Implications of field rainfall simulation experiments. *Hydrological Processes*, *16*(13), 2537–2557. https://doi.org/10.1002/hyp.1014

Holden, J., & Burt, T. P. (2003). Runoff production in blanket peat covered catchments. *Water Resources Research*, *39*(7), 1191–1200. https://doi.org/10.1029/2002wr001956

Holden, J., Kirkby, M. J., Lane, S. N., Milledge, D. G., Brookes, C. J., Holden, V., & McDonald, A. T. (2008). Overland flow velocity and roughness properties in Peatlands. *Water Resources Research*, *44*(6), W06415. https://doi.org/10.1029/2007wr006052

Howson, T., Evans, M., Allott, T., Shuttleworth, E., Johnston, A., Rees, J., et al. (2023). Peatland gully restoration with stone and timber dams (Kinder Plateau, UK). *Ecological Engineering*, *195*, 107066. https://doi.org/10.1016/j.ecoleng.2023.107066

Institute of Hydrology. (1999). *Flood estimation handbook*. Centre for Ecology Hydrology.

Klaar, M. J., Carver, S., & Kay, P. (2020). Land management in a post-Brexit UK: An opportunity for integrated catchment management to deliver multiple benefits? *Wiley Interdisciplinary Reviews: Water*, *7*(5), e1479. https://doi.org/10.1002/wat2.1479

Lane, S. N., & Milledge, D. G. (2013). Impacts of upland open drains upon runoff generation: A numerical assessment of catchment-scale impacts. *Hydrological Processes*, *27*(12), 1701–1726. https://doi.org/10.1002/hyp.9285

Lee, J. A., Press, M. C., Woodin, S., & Ferguson, P. (1987). Responses to acidic deposition in ombrotrophic mires in the UK. In *Effects of atmospheric pollutants on forests, wetlands and agricultural ecosystems* (pp. 549–560). Springer Berlin Heidelberg.

Lindsay, R., Charman, D. J., Everingham, F., O'reilly, R. M., Palmer, M. A., Rowell, T. A., & Stroud, D. A. (1988). *The flow country: The peatlands of Caithness and Sutherland*. Joint Nature Conservation Committee.

Liu, Y., Freer, J., Beven, K., & Matgen, P. (2009). Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error. *Journal of Hydrology*, *367*(1–2), 93–103. https://doi.org/10.1016/j.jhydrol.2009.01.016

Loisel, J., & Gallego-Sala, A. (2022). Ecological resilience of restored peatlands to climate change. *Communications Earth Environment*, *3*(1), 1–8. https://doi.org/10.1038/s43247-022-00547-x

Mackay, A. W., & Tallis, J. H. (1996). Summit-type blanket mire erosion in the Forest of Bowland, Lancashire, UK: Predisposing factors and implications for conservation. *Biological Conservation*, *76*(1), 31–44. https://doi.org/10.1016/0006-3207(95)00087-9

Marshall, M. R., Francis, O. J., Frogbrook, Z. L., Jackson, B. M., McIntyre, N., Reynolds, B., et al. (2009). The impact of upland land management on flooding: Results from an improved pasture hillslope. *Hydrological Processes: An International Journal*, *23*(3), 464–475. https://doi.org/10.1002/hyp.7157

McIntyre, N., Ballard, C., Bulygina, N., Frogbrook, Z., Cluckie, I., Dangerfield, S., et al. (2012). The potential for reducing flood risk through changes to rural land management: Outcomes from the Flood Risk Management Research Consortium. In *BHS Eleventh National Symposium, Hydrology for a changing world*. British Hydrological Society.

McMillan, H. K., & Westerberg, I. K. (2015). Rating curve estimation under epistemic uncertainty. *Hydrological Processes*, *29*(7), 1873–1882. https://doi.org/10.1002/hyp.10419

Metcalfe, P., Beven, K., Hankin, B., & 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 29-km^2 scale catchment in the United Kingdom. *Hydrological Processes*, *31*(9), 1734–1748. https://doi.org/10.1002/hyp.11140

Metcalfe, P., Beven, K., Hankin, B., & Lamb, R. (2018). A new method, with application, for analysis of the impacts on flood risk of widely distributed enhanced hillslope storage. *Hydrology and Earth System Sciences*, *22*(4), 2589–2605. https://doi.org/10.5194/hess-22-2589-2018

Milledge, D., Goudarzi, S., & Dixon, S. (2020). Water environment grant-building blocks: Hydrological analysis to prioritise Gully block locations.

Milledge, D., Odoni, N., Allott, T., Evans, M., Pilkington, M., & Walker, J. (2015). Annex 6. Flood risk modelling. In M. Pilkington, J. Walker, R. Maskill, T. Allott, & M. Evans (Eds.), *Restoration of Blanket bogs; flood risk reduction and other ecosystem benefits*. Moors for the Future Partnership.

Mindham, D., Beven, K., & Chappell, N. (2023). Rainfall–streamflow response times for diverse upland UK micro-basins: Quantifying hydrographs to identify the nonlinearity of storm response. *Hydrology Research*, *54*(2), 233–244. https://doi.org/10.2166/nh.2023.115

Nilsson, C., Riis, T., Sarneel, J. M., & Svavarsdóttir, K. (2018). Ecological restoration as a means of managing inland flood hazards. *BioScience*, *68*(2), 89–99. https://doi.org/10.1093/biosci/bix148

O'Connell, P. E., Ewen, J., O'Donnell, G., & Quinn, P. (2007). Is there a link between agricultural land-use management and flooding? *Hydrology and Earth System Sciences*, *11*(1), 96–107. https://doi.org/10.5194/hess-11-96-2007

Odoni, N. A., & Lane, S. N. (2010). Assessment of the impact of upstream land management measures on flood flows in Pickering Beck using overflow. Project RMP55455: Slowing the flow at Pickering.

O'Donnell, G., Ewen, J., & O'Connell, P. E. (2011). Sensitivity maps for impacts of land management on an extreme flood in the Hodder catchment, UK. *Physics and Chemistry of the Earth, Parts A/B/C*, *36*(13), 630–637. https://doi.org/10.1016/j.pce.2011.06.005

Parry, L. E., Holden, J., & Chapman, P. J. (2014). Restoration of blanket peatlands. *Journal of Environmental Management*, *133*, 193–205. https://doi.org/10.1016/j.jenvman.2013.11.033

Pilkington, M., Walker, J., Maskill, R., Allott, T., & Evans, M. (2015). *Restoration of Blanket bogs; flood risk reduction and other ecosystem benefits. Final report of the Making Space for Water project*. Moors for the Future Partnership.

Reaney, S. M. (2022). Spatial targeting of nature-based solutions for flood risk management within river catchments. *Journal of Flood Risk Management*, *15*(3), e12803. https://doi.org/10.1111/jfr3.12803

Rochefort, L. (2000). Sphagnum—A keystone genus in habitat restoration. *The Bryologist*, *103*(3), 503–508. https://doi.org/10.1639/0007-2745(2000)103[0503:sakgih]2.0.co;2

Rothwell, J. J., Evans, M. G., & Allott, T. E. H. (2007). Lead contamination of fluvial sediments in an eroding blanket peat catchment. *Applied Geochemistry*, *22*(2), 446–459. https://doi.org/10.1016/j.apgeochem.2006.11.002

Shuttleworth, E. L., Evans, M. G., Hutchinson, S. M., & Rothwell, J. J. (2015). Peatland restoration: Controls on sediment production and reductions in carbon and pollutant export. *Earth Surface Processes and Landforms*, *40*(4), 459–472. https://doi.org/10.1002/esp.3645

Shuttleworth, E. L., Evans, M. G., Pilkington, M., Spencer, T., Walker, J., Milledge, D., & Allott, T. E. (2019). Restoration of blanket peat moorland delays stormflow from hillslopes and reduces peak discharge. *Journal of Hydrology X*, *2*, 100006. https://doi.org/10.1016/j.hydroa.2018.100006

Verhoeven, J. T. A., & Liefveld, W. M. (1997). The ecological significance of organochemical compounds in Sphagnum. *Acta Botanica Neerlandica*, *46*(2), 117–130. https://doi.org/10.1111/plb.1997.46.2.117

Wilkinson, M. E., Addy, S., Quinn, P. F., & Stutter, M. (2019). Natural flood management: Small-scale progress and larger-scale challenges. *Scottish Geographical Journal*, *135*(1–2), 23–32. https://doi.org/10.1080/14702541.2019.1610571

Wingfield, T., Macdonald, N., Peters, K., Spees, J., & Potter, K. (2019). Natural flood management: Beyond the evidence debate. *Area*, *51*(4), 743–751. https://doi.org/10.1111/area.12535

Wittram, B. W., Roberts, G., Buckler, M., King, L., & Walker, J. S. (2015). *A practitioners guide to Sphagnum reintroduction*. Moors for the Future Partnership.

Xu, J., Morris, P. J., Liu, J., & Holden, J. (2018). PEATMAP: Refining estimates of global peatland distribution based on a meta-analysis. *Catena*, *160*, 134–140. https://doi.org/10.1016/j.catena.2017.09.010