ORIGINAL ARTICLE

CIWEM Chartered Institution of Water and Environmental Management — Journal of Flood Risk Management — WILEY

Can hydrological models assess the impact of natural flood management in groundwater-dominated catchments?

Hèou Maléki Badjana[1,2] | Hannah L. Cloke[1,3,4,5] | Anne Verhoef[1] |
Stefan Julich[6] | Carla Camargos[7] | Sarah Collins[8] |
David M. J. Macdonald[9] | Patrick C. McGuire[3,10] | Joanna Clark[1]

[1]Department of Geography and Environmental Science, University of Reading, Reading, UK

[2]Laboratory of Botany and Plant Ecology, Faculty of Sciences, University of Lome, Lome, Togo

[3]Department of Meteorology, University of Reading, Reading, UK

[4]Department of Earth Sciences, Uppsala University, Uppsala, Sweden

[5]Centre of Natural Hazards and Disaster Science, CNDS, Uppsala, Sweden

[6]Institute of Soil Science and Site Ecology, Technische Universität Dresden, Tharandt, Germany

[7]Institute for Landscape Ecology and Resources Management (ILR), Research Centre for BioSystems, Land Use and Nutrition (iFZ), Justus Liebig University Giessen, Giessen, Germany

[8]The Lyell Centre, British Geological Survey, Edinburgh, UK

[9]British Geological Survey, Wallingford, Oxfordshire, UK

[10]National Centre for Atmospheric Science (NCAS), University of Reading, Reading, UK

Correspondence
Hèou Maléki Badjana, Department of Geography and Environmental Science, University of Reading, Reading, RG6 6DW, UK.
Email: m.badjana@reading.ac.uk; maleki.badj@yahoo.fr

Funding information
Natural Environment Research Council, Grant/Award Number: NE/R004668/1

Abstract

Natural flood management (NFM) is widely promoted for managing flood risks but the effectiveness of different types of NFM schemes at medium (100–1000 km^2) and large scales (>1000 km^2) remains widely unknown. This study demonstrates the importance of fully understanding the impact of model structure, calibration and uncertainty techniques on the results before the NFM assessment is undertaken. Land-based NFM assessment is undertaken in two medium-scale lowland catchments within the Thames River basin (UK) with a modelling approach that uses the Soil and Water Assessment Tool (SWAT) model within an uncertainty framework. The model performed poorly in groundwater-dominated areas (P-factor <0.5 and R-factor >0.6). The model performed better in areas dominated by surface and interflow processes (P-factor >0.5 and R-factor <0.6) and here hypothetical experiments converting land to broadleaf woodland and cropland showed that the model offers good potential for the assessment of NFM effectiveness. However, the reduction of large flood flows greater than 4% in medium-sized catchments would require afforestation of more than 75% of the area. Whilst hydrological models, and specifically SWAT, can be useful tools in assessing the effectiveness of NFM, these results demonstrate that they cannot be applied in all settings.

1 | INTRODUCTION

Natural flood management (NFM) has gained much interest in recent years, driven by increased recognition of its potential for reducing flood risk, and enhancing and sustaining ecosystem services and functions (Dadson et al., 2017; McLean et al., 2013). Catchment-based NFM is the implementation of land and soil management measures to protect, restore, or enhance the natural features, characteristics, and functions of catchments, to reduce peak flood flows and/or delay their timing, and, in turn, to reduce flood impact (Connelly et al., 2020; SEPA, 2016). Although NFM is recognised as an effective natural approach, its effects vary across the scales and cannot always be translated between sites/catchments (Wingfield et al., 2019). Investing in NFM therefore requires robust evidence on its effectiveness where it is to be implemented.

Despite the wide adoption of NFM across the UK, most evidence for its effectiveness has only been established at small scales (catchment areas of <100 km^2). There remains a distinct and concerning lack of evidence at larger scales (Connelly et al., 2020; Dadson et al., 2017; Kay et al., 2019; Waylen et al., 2018; Wilkinson et al., 2019; Wingfield et al., 2019), and limited investigations of the detailed processes governing the catchment hydrological system and associated uncertainties, and how this can be affected by NFM. Few studies on the effectiveness of NFM have focused on permeable lowland catchments with complex hydrology and where integrated modelling of surface water-groundwater interactions remains challenging (Lane et al., 2019; Wagener et al., 2021; Wheater et al., 2007); yet, these catchments are subject to frequent flooding from groundwater and intensive rainfall (Ascott et al., 2017; Hughes et al., 2011; Macdonald et al., 2012). There is a clear need to test NFM effectiveness in large lowland catchments, to provide the necessary evidence required to take appropriate actions.

Due to the complexity of large-scale and lowland groundwater-dominated catchments systems, NFM assessment in these requires a consistent modelling framework that allows realistic simulation of the surface and groundwater interactions while considering the spatial variability of landscape features, acknowledging uncertainties, and assessing the changes in the processes and parameters due to NFM (Ellis et al., 2021). This would enable the assessment of the catchment scale impacts of NFM, the interactions between individual NFM interventions and sub-catchments, and catchment process changes after implementation of NFM measures (Hankin et al., 2017; Metcalfe et al., 2017). Within such a framework, the catchment model itself should be able to represent the spatial complexity of the catchment features and the NFM-related processes, whilst still being practical and computationally realistic to employ (Metcalfe et al., 2017).

The Soil and Water Assessment Tool (SWAT) is an open-source semi-distributed and physically based catchment-scale hydrological model that has been used extensively to simulate the impacts of land use and land cover changes on hydrologic processes (Perra et al., 2018; Wangpimool et al., 2013). However, its application remains challenging in groundwater-dominated catchments. Although modified versions of the SWAT model, with modifications to the groundwater module (Guse et al., 2016; Nguyen & Dietrich, 2018; Pfannerstill et al., 2014), or inclusion of coupled groundwater-surface water (Aliyari et al., 2019; Bailey et al., 2016; Chunn et al., 2019; Liu et al., 2020) have been tested, there has been little attempt to capture how different representations of processes, parametrisations, calibration and uncertainty analysis (CUA) techniques, affect the model performance. However, this is essential for understanding the application of models for use in NFM assessment.

The objective of this study is to fully assess a modelling framework for NFM assessment in two medium-scale lowland catchments with complex hydrology within the Thames basin, UK. By doing so, the study aims to highlight the importance of implementing such a framework in order to understand to what extent the selected catchment model can be used to assess land-based NFM, the impact of different model configurations, as well as the uncertainties and the sensitivity of the model to changes in soil and land management practice.

The paper is organised into four sections. Section 2 presents the study areas and describes the datasets, modelling framework and the model used as well as how it is calibrated, validated and applied for testing NFM scenarios. Section 3 presents the results for the studied catchments on the model performance, changes in parameters and processes according to the landscape properties and flooding events, and the effects of the tested NFM scenarios. Section 4 discusses the meanings

and implications of the results and finally, Section 5 summarises the main findings and outlook.

2 | MATERIALS AND METHODS

2.1 | Study areas

The modelling study was carried out for two lowland catchments, the Pang and the Blackwater, sub-catchments of the Thames Basin in southern UK (Figure 1). The study has been conducted as part of the LANDWISE project, for which a number of catchments were targeted for hydrological modelling which was performed at small, medium and large scales. The Pang and Blackwater catchments were selected for medium-scale modelling.

The Pang is mainly rural with intensive farming of wheat and barley cultivated in winter and spring, and oil-seed rape. The eastern part of the Blackwater is heavily urbanised compared to the western part which remains largely underdeveloped.

The bedrock geology of the Pang is mainly Cretaceous Chalk with high permeability, but downstream the Chalk is covered by low-permeability Palaeogene mudstone bedrock (Allen et al., 1997; Peters & van Lanen, 2005). In the Blackwater, the bedrock geology is more diversified, composed of mainly Chalk and London Clay, overlain by a series of relatively thin sands and clays.

The choice of Blackwater was guided by the fact that it has different landscape properties, compared to the other selected catchments. In fact, apart from having its upper part on the Whitewater catchment underlain by Chalk, most of the Blackwater catchment has different bedrock geology, and related soils, which gives an opportunity to test the model and the NFM effectiveness in different landscapes. In addition, heavily urbanised catchments like Blackwater are rarely ideal; assessing the suitability of such catchments was part of the underlying rationale of the LANDWISE project.

The Pang presents a particular feature known as the "Blue Pool" spring complex located near Bucklebury—a perennial spring discharging on average 0.1 and 0.2 $m^3 s^{-1}$ of groundwater to streamflow in the summer and
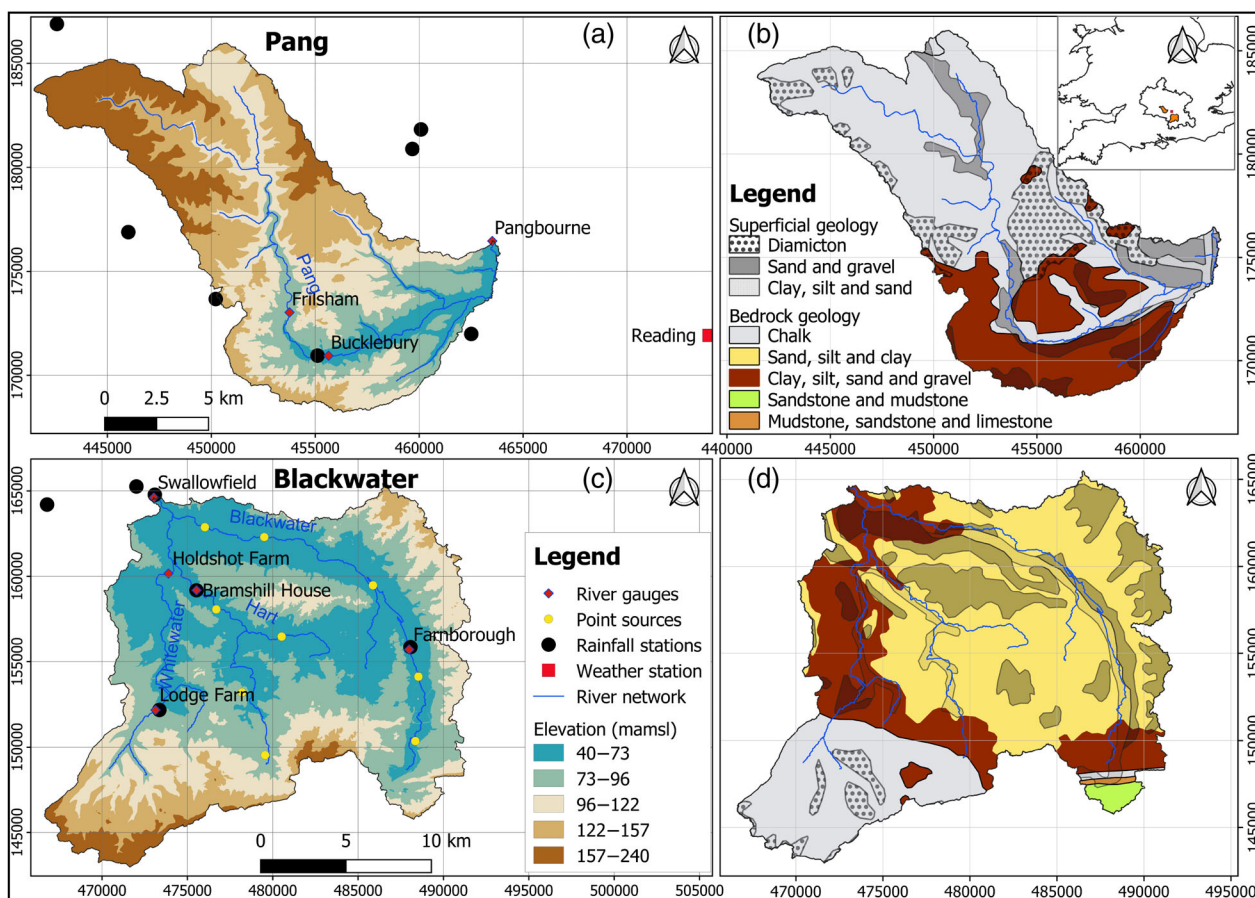

FIGURE 1 Elevation with the location of hydrometeorological stations and geology of the Pang (a,b) and Blackwater catchments (c,d).

TABLE 1 Hydrological characteristics of the Pang and Blackwater and their sub-catchments.

| Catchments | Sub-catchments | Hydrometric station coordinates | | Area (km^2) | MAQ (m^{3}s^{-1}) | BFI | POT threshold (m^{3}s^{-1}) | Percentage of time flow equalled or exceeded POT (%) | |
		Longitude	Latitude					2005–2008	2012–2014
Pang	Pang at Frilsham	453,762	173,027	89.5	0.22	0.95	-		
	Pang at Bucklebury	455,637	170,937	151.9	0.23	0.88	-		
	Pang at Pangbourne	463,517	176,459	170.5	0.65	0.87	1.199	5	30
Blackwater	Whitewater at Lodge Farm	473,127	152,172	46	0.39	0.94	0.995	0	21
	Whitewater at Holdshot Farm	473,887	160,167	102.8	0.84	-	-		
	Hart at Bramshill House	475,532	159,182	83.7	0.8	0.65	4.962	84	2
	Blackwater at Farnborough	487,992	155,702	38	0.5	0.7	-		
	Blackwater at Swallowfield	473,032	164,612	358	3.2	0.67	10.82	2	8

Note: BFI and POT are obtained from the National River Flow Archive (NRFA, 2022) while the sub-catchments are delineated with the SWAT model using the stations' coordinated provided in the NRFA.

Abbreviations: BFI, Baseflow Index; MAQ, mean annual discharge; POT, peak over threshold.

winter, respectively (Bradford, 2002; Griffiths et al., 2006). The Pang has experienced a series of flood events in recent years both from intensive rainfall (e.g., 2007) and groundwater emergence (e.g., 2000/2001, 2002/2003 and 2013/2014) (Hughes et al., 2011; Muchan et al., 2015). The 2000/2001, 2002/2003 and 2013/2014 groundwater floods related to accumulated wetness over 90–100 days, although in the case of 2000/2001, groundwater levels were already much higher than normal, while the 2007 summer flooding was related to a large convection storm.

Three streamflow gauging stations on the River Pang were used: Frilsham, Bucklebury, and Pangbourne (NRFA, 2022). Backwater effects have been observed at the confluence of the Pang and Thames, which result in drowning of the Pangbourne gauge and uncertainty in recorded high flows. Five stream gauge stations are used in the Blackwater: two on the Whitewater at Lodge Farm at the edge of the Chalk outcrop and downstream at Holdshot Farm; one on the Hart at Bramshill House; and two on the Blackwater at Farnborough and Swallowfield (Figure 1). Sewage effluent is a major component (up to 40%) of the Blackwater flows (NRFA, 2022).

The hydrological features of the nested sub-catchments are summarised in Table 1.

2.2 | Modelling framework

The modelling framework integrates the catchment model with different processes, configurations and methods as well as input data and information from different sources on elevation, weather, soil, land use and river flows (Figure 2). It allows the evaluation of whether the catchment hydrological system with NFM-related processes can be adequately simulated from available datasets while considering uncertainties. This involved several steps (Figure 2). Therefore, first, two model setups are constructed to test the influence of using slope bands (measurement of the percentage of slope inclination) since this is known to affect model performance. Second, the constructed models are calibrated based on two different options of using the time series of observed flow, which has different flooding events, while considering uncertainties. Third, each model verification is performed to assess how close the processes are to the reality and
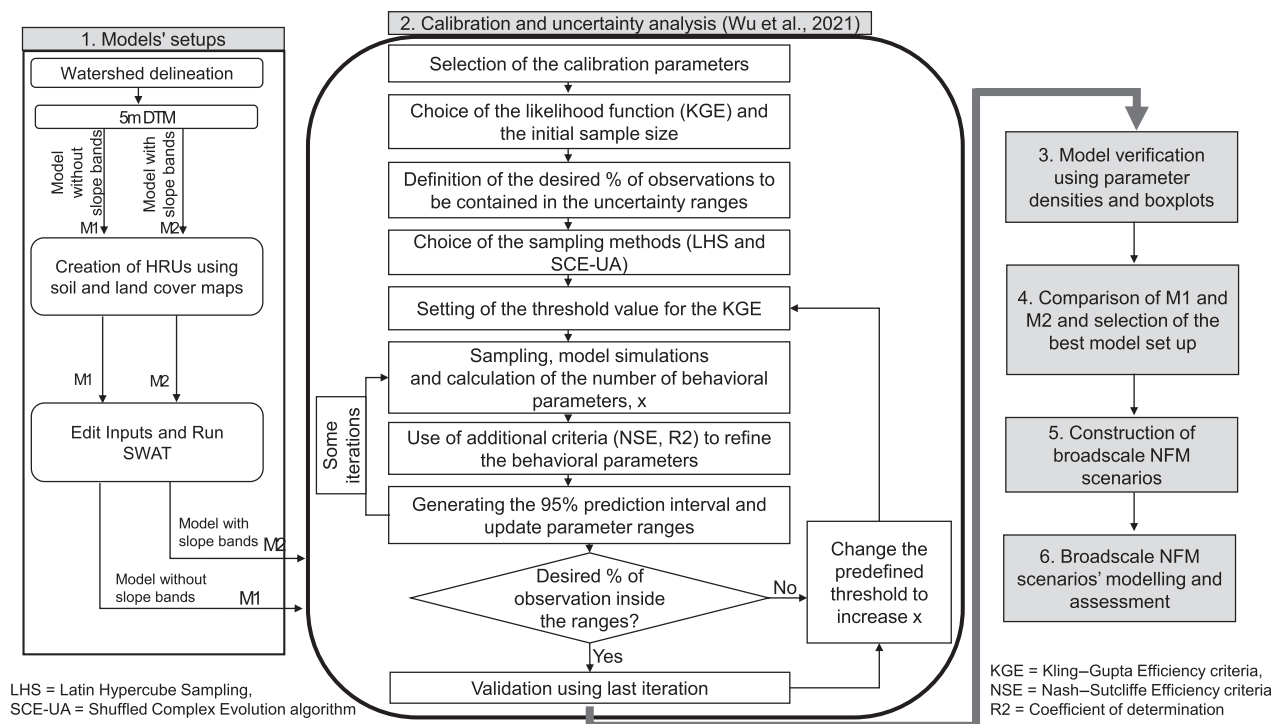
FIGURE 2 Modelling framework for evaluating model suitability for NFM assessment; M1, and M2 are model setups (1 for model without slope bands and 2 indicates the inclusion of slope bands).

the variations and changes in parameters. Fourth, the calibrated models are compared to select the best model set up. Fifth, the NFM scenarios are constructed and finally they are modelled and assessed in the last step 6.

For this study, the SWAT is used (see Section 1 of Supporting Information).

2.3 | Datasets

Daily rainfall station datasets from England's Environment Agency (EA) were collected between 2001 and 2014 used to set up and run the model. Further climate variables were obtained from the University of Reading meteorological observatory. Soil data are from the National Soil Map from England and Wales—NATMAP (Cranfield University, 2013). For the elevation model, the EA continuous digital terrain model (DTM) of 2 m resolution, resampled to 5 m due to computational issues, was used. Landcover data are from the UK Centre for Ecology and Hydrology land cover polygons for 2015.

River flow data from the UK National River Flow Archive (NRFA, 2022) were used for the calibration and validation of the models. Point source sewage effluent discharge from the Blackwater catchment is obtained from Thames Water. The integration of point source data in the model has helped to improve the baseflow simulation at the involved outlets.

2.4 | Model setups

The SWAT model was implemented via the QSWAT3 v1.1 plugin integrated into QGIS 3.10.

Two model setups per catchment were modelled (Table 2) to test the impacts of using slope bands in the models since catchment models are known to be particularly sensitive to these factors. Each of the two models is set up with the 5 m DTM, one with slope bands (measurement of the percentage of slope inclination) and another without (Figure 2, Step 1). The slope bands were defined according to the British Land Capability classification system. Based on the minimum and maximum altitude of the target catchment, we considered three aggregated slope classes: 0–3° (0–5.24%), 3–11° (5.24–19.4%) and > 11° (19.4%) corresponding to gently sloping, moderately-to-strongly sloping, and moderately-to-very deep sloping.

The soil and land cover maps used in this study and the delineated sub-basins as well as the geology maps (for information) are summarised in Figure S1.

The next step after HRU creation is importing the weather data and writing data files for each subbasin and each HRU. The SWAT model distributes the weather data to each sub-basin by assigning the data of the closest station to the sub-basin centroid. Subsequently, weather data of a subbasin are assigned to each of the HRUs within it.

TABLE 2 Characteristics of the model setups.

Catchments	Model setup	Elevation model	Slope bands	Threshold (km²)	Number of subbasins	Number of HRUs
Pang	PGM1[a]	5 m	No	4.5	17	837
	PGM2	5 m	Yes	4.5	17	2248
Blackwater	BWM1	5 m	No	9	17	1367
	BWM2	5 m	Yes	9	17	3561

[a]PG stands for Pang and BW for Blackwater; M1 and M2 represent models without and with slope bands, respectively.

TABLE 3 Parameters used for the SWAT calibration.

Parameter	Description	Effect of hydrological processes
ALPHA_BF[a]	Baseflow alpha factor (days)	Index of shallow groundwater flow response to changes in recharge
ALPHA_BF_D	Baseflow alpha factor for deep aquifer (days)	Index of deep aquifer groundwater flow response to changes in recharge
GW_DELAY	Groundwater delay (days)	Controls the timing of the recharge and outflow of the groundwater
GWQMN	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)	Allows to adjust the amount of water from the shallow aquifer that goes back to the river channel as baseflow
GW_REVAP	Groundwater "revap" coefficient	Determines the amount of water loss from the aquifer by re-evaporation
REVAPMN	Threshold depth of water in the shallow aquifer for "revap" to occur (mm)	Allows to control the amount of water available for re-evaporation
RCHRG_DP	Deep aquifer percolation fraction (-)	Determines the recharge of water to the deep aquifer and, indirectly, soil water storage and surface runoff generation
CN2[a]	SCS runoff curve number f (-)	Controls directly the infiltration of water into the soil, or the surface runoff generation
ESCO	Soil evaporation compensation factor (-)	Allow to adjust the water loss via evapotranspiration
SOL_BD[a]	Moist bulk density (Mg.m³)	Indictor of soil compaction, affects the soil moisture
SOL_AWC[a]	Available water capacity of the soil layer (mm)	Soil moisture available to plants
SOL_K[a]	Saturated hydraulic conductivity (mm h^{-1})	Controls the rate of water movement through the soil
SURLAG	Surface runoff lag time (-)	Controls the total amount of surface runoff entering the river channel on a given day
OV_N[a]	Manning's "n" value for overland flow (-)	Defines the catchment roughness
CH_N2	Manning's "n" value for the main channel (-)	Defines the river channel roughness
CH_K2	Effective hydraulic conductivity in main channel alluvium (mm h^{-1})	Controls water loss by seepage to the groundwater
ALPHA_BNK	Baseflow alpha factor for bank storage	Control water storage in the river channel
CANMX_FR	Maximum canopy storage for forest areas (mm)	Storage of water from rainfall interception by forest and its loss via evaporation
CANMX_RNG	Maximum canopy storage for rangeland (mm)	Storage of water from rainfall interception by rangeland and its loss via evaporation
CANMX_PAST	Maximum canopy storage for pasture (mm)	Storage of water from rainfall interception by pasture and its loss via evaporation
CANMX_AGR	Maximum canopy storage for agricultural areas (mm)	Storage of water from rainfall interception by agricultural land and its loss via evaporation

[a]Parameters for which a relative change is applied to an initial value during the calibration, whereas an absolute change was used for the remaining parameters.

2.5 | Model calibration and assessment

Calibration parameters were selected based on their relation to NFM processes. In total, 16 parameters for each group of subbasins, and 5 catchment-wide parameters, were selected (Table 3) resulting in a total of 53 parameters for the Pang and 85 for the Blackwater. Catchment-wide parameters are the baseflow alpha factor for the deep aquifer (ALPHA_BF_D) and the maximum canopy storage which is assumed to vary similarly across the subbasins but differently according to land cover classes such as forest (CANMX_FR), pasture (CANMX_PST), rangeland (CANMX_RNG), and agricultural land (CANMX_AGR). The values of CANMX_FR, CANMX_PST, CANMX_RNG and CANMX_AGR are selected based on previous research (André et al., 2008; Pearce et al., 1980; Rutter et al., 1975; Zou et al., 2015).

The SWAT model was calibrated using the simultaneous multi-site calibration technique (Leta et al., 2017) coupled with the multicriteria sequential calibration and uncertainty analysis (MS-CUA) approach proposed by Wu et al. (2021) using the Python package SPOTPY (Houska et al., 2015). This involved the definition of groups of subbasins based on the number of streamflow gauging stations used (Table 4).

Two calibration-validation breakdowns are tested to consider the different types of flood events that appear in the relatively short record: (i) 2012–2014 calibration which includes a groundwater flood and 2005–2008 validation which has a flood event from intensive rainfall; and (ii) 2005–2008 calibration and 2012–2014 validation, with 2 years warm up in each calibration. The two periods were defined based on the availability of the datasets required by the model using the longest continuous periods available. Considering alternate periods for calibration and validation allows to determine whether the same model is able to simulate floods originating from rising groundwater as well as from intensive rainfall and with the same parameters.

Predictive uncertainty at 95% confidence interval was estimated and the model performance evaluated using P-factor and R-factor (Abbaspour et al., 2017; Camargos et al., 2018; Section 2 of Supporting Information).

Verification was applied to check that the estimated parameters, simulated processes and their spatial–temporal variations are realistic (Figure 2, Step 3). Water balance components were also analysed and compared between calibrations. Finally, for each catchment, the best model set up was selected based on the quality of the simulated hydrograph (Figure 2, Step 4).

2.6 | NFM scenarios generation and assessment of effects

To assess the extent to which the best model setup can be used for NFM assessment, two broadscale hypothetical scenarios were generated (Figure 2, Step 5). These scenarios consist of "what if" assumptions and focused on the change in the water balance components and subsequently river discharge when all the land cover, except water and urban areas, is converted to: (i) broadleaf woodland and (ii) conventional agricultural land. The use of "what if" scenarios is a common technique in sensitivity analysis and one that allows the behaviour of the model to be interpreted and explained in more detail (Cloke et al., 2008). Their use is important to understand the sensitivity of the model to land use changes, which is therefore primordial for the assessment and understanding of the effects of more realistic NFM scenarios. It provides critical evidence, for instance, on how much increasing the extent of afforestation to its maximum extent can reduce flood risk.

The broadscale NFM effects in each catchment were assessed by looking at changes in the 95% prediction intervals (Figure 2, Step 6). Differences between large floods, small floods and high-flow pulses, considered as flows with exceedance probability less or equal to 2%, between 2% and 10% and between 10% and 40% respectively, were assessed. These flows were computed from the 95% prediction uncertainties of the calibrated model and each modelled scenario. The Kolmogorov–Smirnov

TABLE 4 Groups of subbasins used for the multisite calibration of the Blackwater and Pang catchments.

Catchments	Sub-catchments	Group	Subbasins
Pang	Pang at Frilsham	PI	1, 2, 3, 4, 7, 9, 12
	Pang at Bucklebury	PII	16
	Pang at Pangbourne	PIII	5, 6, 8, 10, 11, 13, 14, 15, 17
Blackwater	Blackwater at Swallowfield	BI	1, 2, 3, 4, 6, 8, 9
	Whitewater at Holdshot farm	BII	5, 13, 15
	Hart at Bramshill House	BIII	7, 10, 12
	Blackwater at Farnborough	BIV	11
	Whitewater at Lodge Farm	BV	14, 16, 17

8 of 23 | WILEY—**CIWEM** Chartered Institution of Water and Environmental Management Journal of Flood Risk Management

BADJANA ET AL.

test was applied to assess whether the data from the calibrated model and the modelled scenario come from different distributions. Finally, we used the bootstrap sampling via the R package "PairwiseCI" v0.1-27 (Schaarschmidt & Gerhard, 2019) to estimate the 95% confidence interval of the difference in medians when the data come from different distributions.

3 | RESULTS

For the two catchments, the integration of slope bands in the model is not necessary as this involves the integration and calibration of more parameters but does not markedly improve the quality of the simulated hydrograph.

Thus, models without slope bands, PGM1 and BWM1 for the Pang and Blackwater, respectively, were selected as the best ones.

3.1 | Modelling results for the pang catchment

3.1.1 | Model performance

For the 2012–2014 calibration, the model performs poorly at Frilsham, (Figure 3a) and Bucklebury (Figure 3b) with P-factor less the 0.5 and R-factor equal to 0.7 whereas it improves markedly at Pangbourne (P-factor = 0.72, R-factor = 0.75) where the hydrograph dynamics and peak
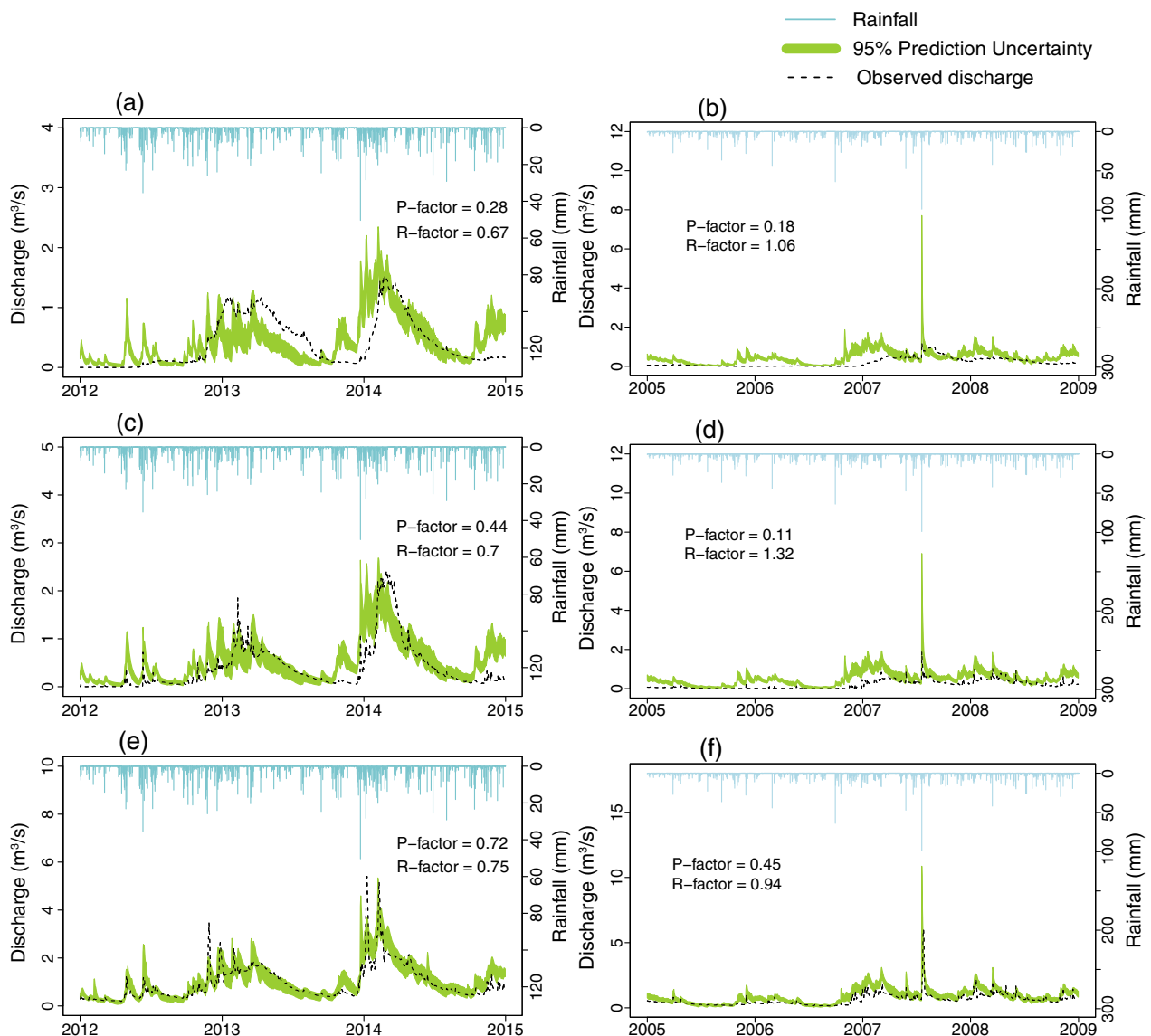

FIGURE 3 Model performance during the 2012–2014 calibration (with groundwater flood) and 2005–2008 validation (with rainwater flood) for the Pang at Frilsham (a,b), Bucklebury (c,d) and Pangbourne (e,f).

flows are simulated relatively well (Figure 3c). At Frilsham, the model reacts earlier to rainfall inputs, and the modelled hydrograph is flashier than the observed response of the catchment. During the validation period (2005–2008), the model performance drops drastically, overestimating flows at Frilsham (Figure 3d) and Bucklebury (Figure 3e) with P-factor less than 0.2 at the two stations although relatively better results are obtained in Pangbourne (Figure 3f) with a P-factor of 0.45. At all stations, the modelled flows are higher than recorded flows for the July 2007 peak, although it should be noted that the peak flow at Pangbourne, in particular, is highly uncertain due to extensive overspill into the neighbouring Sulham Brook.

The 2005–2008 calibration shows poor model performance at Frilsham and Bucklebury (P-factor ~ 0.2, R-factor >0.6) and an improvement at Pangbourne (P-factor = 0.5). However, the performance seems better compared to when the 2012–2014 calibration model is validated for 2005–2008. At Frilsham, the model is very responsive to rainfall events compared to the observed catchment response and underestimates the flows during the wet conditions of 2007 (Figure 4a). A similar performance is observed at Bucklebury (Figure 4b) but with less bias than Frilsham. Model performance improved downstream at Pangbourne (Figure 4c). The model error related to the overestimation of the 2007 flood event is much less than that obtained when the model is
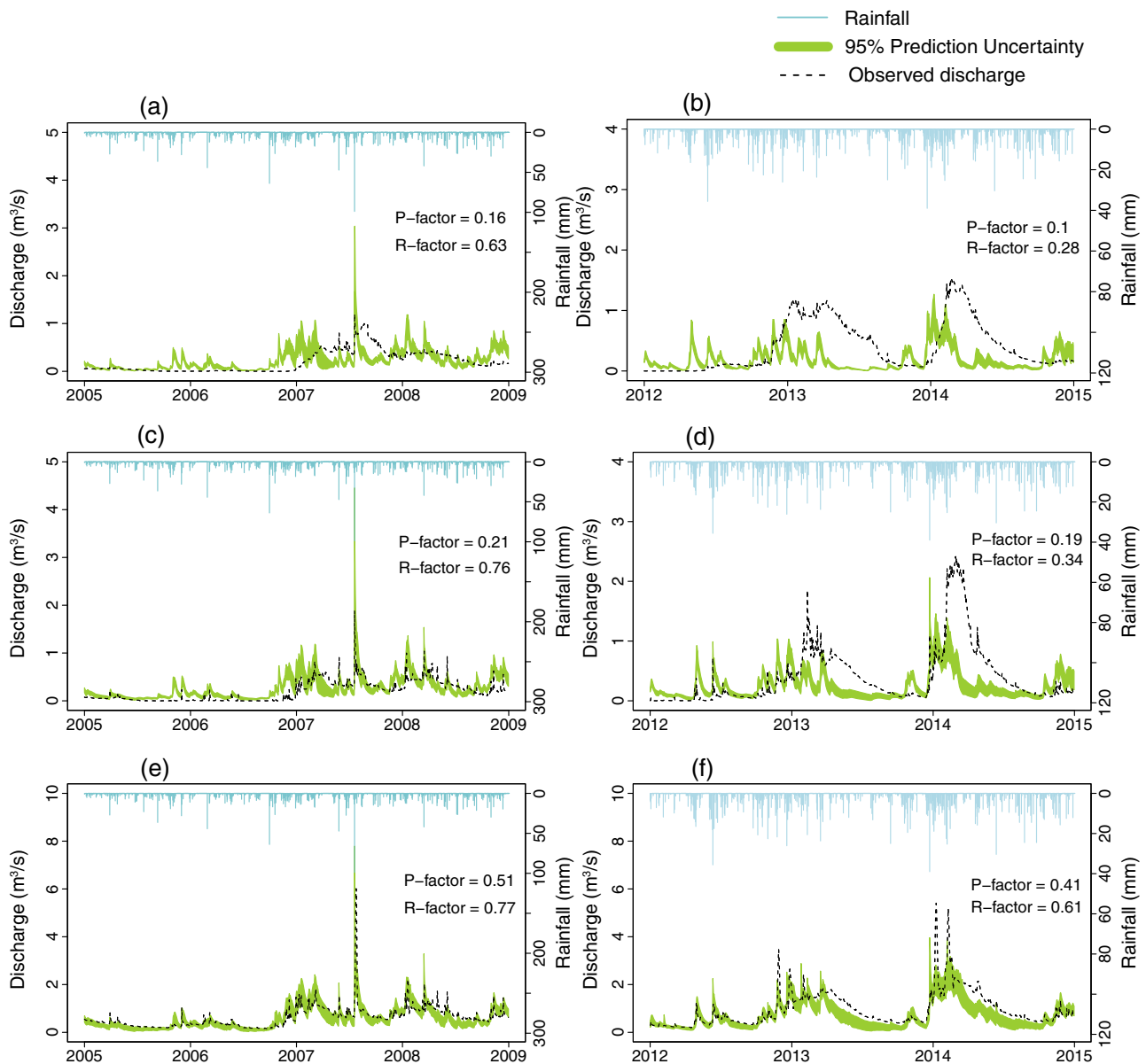

FIGURE 4 Model performance during the 2005–2008 calibration (with rainwater flood) and 2012–2014 validation (with groundwater flood) for the Pang at Frilsham (a,b), Bucklebury (c,d) and Pangbourne (e,f).

calibrated over 2012–2014 and validated for 2005–2008. During the validation period, the performance of the model drastically dropped at all outlets with a marked underestimation of streamflow at all outlets with the values of P-factor that are close to 0 (Figure 4d,e,f).

3.1.2 | Verification of models from 2005 to 2008 and 2012 to 2014 calibrations

The distributions of ESCO, SURLAG, OV_N, ALPHA_BF and REVAPMN do not change systematically (Figure 5e-l) between the two calibrations. The other parameters (other panels in Figure 5) demonstrate clear changes in their distributions between the Pang at Frilsham, Bucklebury and Pangbourne. Additionally, there is a similarity in the distributions of most parameters as well as SOL_BD and SOL_AWC (Figure 5c,d) for the sub-catchments of the Pang at Bucklebury and Pangbourne.

There are important changes in parameters related to infiltration and soil water: CN2, SOL_K and SOL_AWC

(Figure 5a,b,d), those related to groundwater dynamics: GW_DELAY, GWQMN, RCHRG_DP, REVAPMN and GW_REVAP (Figure 5i–m), and to river channel: CH_N2, CH_K2 and ALPHA_BNK (Figure 5n–p). Parameters such as SOL_BD (Figure 5c), ESCO, SUR-LAG, OV_N, and ALPHA_BF (Figure 5e–h) vary less. For the calibration over 2005–2008, for which an improved model performance and the minimisation of the streamflow overestimation were observed, CN2, RCHRG_DP and REVAPMN have decreased in PI and PII. This leads to more infiltration of water in the soil, more percolation to the deep aquifer and more water in the shallow aquifer that can move to the root zone. CH_N2 parameter has also decreased in the three subbasin groups (Figure 5n) while SOL_AWC and GW_REVAP parameters values increase, suggesting reduced channel roughness and more soil water storage and evaporation from the shallow aquifer. Moreover, the analysis of the simulated hydrographs of both calibrations for the Pang at Frilsham, where the differences are drastic, reveals that the changes occurred importantly in CN2, SOL_K,
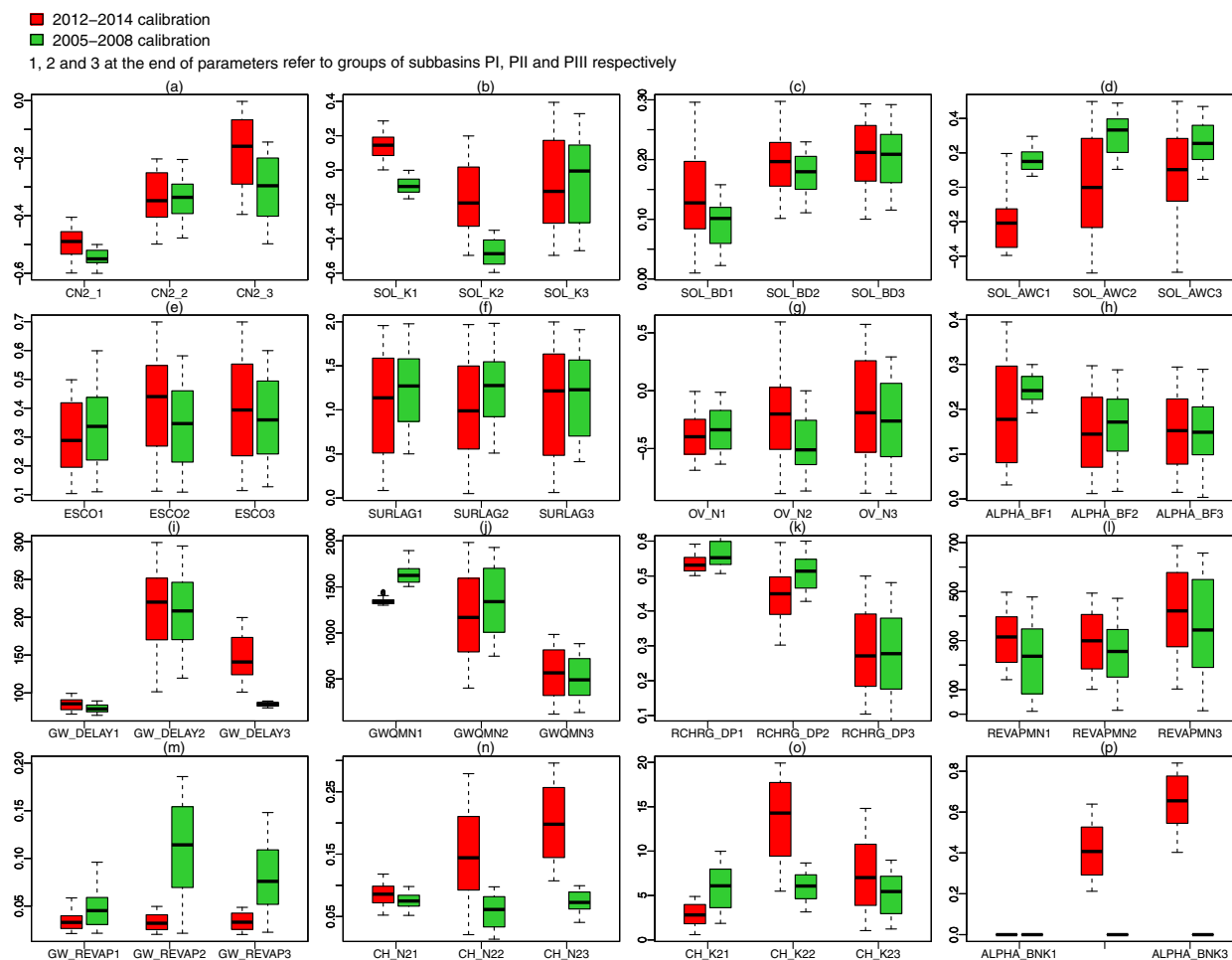

FIGURE 5 Model parameters from the two calibrations on the Pang catchment.

SOL_AWC, GWQMN, ALPHA_BF and CH_K2 for the 2005–2008 calibration compared to the one of 2012–2014.

3.1.3 | Verification of the average water balances

A comparative analysis of the average water balances shows that the most important differences between the two calibrations reside in the surface runoff (SURQ) and groundwater contribution to streamflow (GWQ) which decreased for the 2005–2008 calibration, reducing the water yield (WYLD). At the same time, a marked increase in the soil water storage (SW) occurred while the percolation (PERC), lateral flow (LATQ) and evapo-transpiration (ET) did not change much. In terms of percentage, different components of the water balance seem to be constant throughout the year, without any noticeable variability from 1 month to another (Figure 6).

Further analysis of the spatial variability of the simulated processes and their changes between the two calibrations show that the annual average simulated ET varies according to the land use; in a similar fashion for both calibrations (Figure 7a,b). It ranges between 200 to 800 mm and is reduced by values of 0–30 mm when the model is run with 2005–2008 parameters compared to 2012–2014 (Figure 7c). The generated surface runoff (SURQ_GEN) varies from 0–350 mm/year and is less than 10 mm for a large part of the catchment especially on the upper chalk, ranging from 10 to 50 mm on the lower clay. From the catchment coverage, SURQ_GEN above 50 mm, and more precisely between 50 and 100 mm, is higher for the 2012–2014 calibration (Figure 7d) than that of 2005–2008 (Figure 7e). The reduction in the SURQ_GEN observed for the model with 2005–2008 parameters is between 0 and 10 mm and in the upper part of the catchment (Figure 7f), but varies from a reduction of up to 40 mm and an increase of up to 10 mm elsewhere. Groundwater recharge (GW_RCHG) follows similar spatial patterns for both calibrations (Figure 7gh) and varies between 0 and 700 mm. High recharge occurs on chalk soils, and reduced recharge on soils with clay and impeded drainage. There is a reduction in recharge ranging between 20 and 40 mm for most of the catchment for the 2005–2008 calibration (Figure 7i) except on the lower part where an increase of up to 40 mm is obtained.
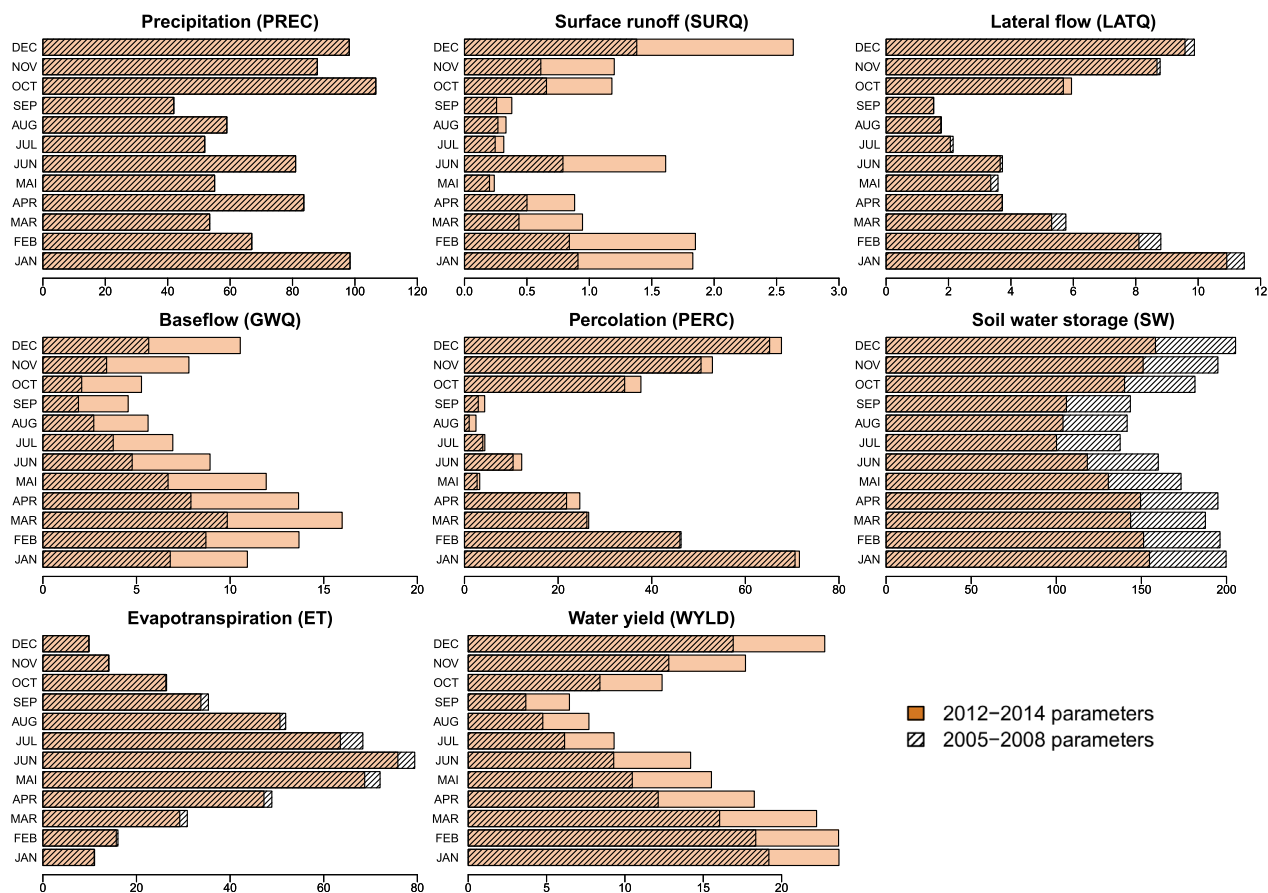

FIGURE 6 Ensemble average water balances of the 2012–2014 and 2005–2008 calibrations.
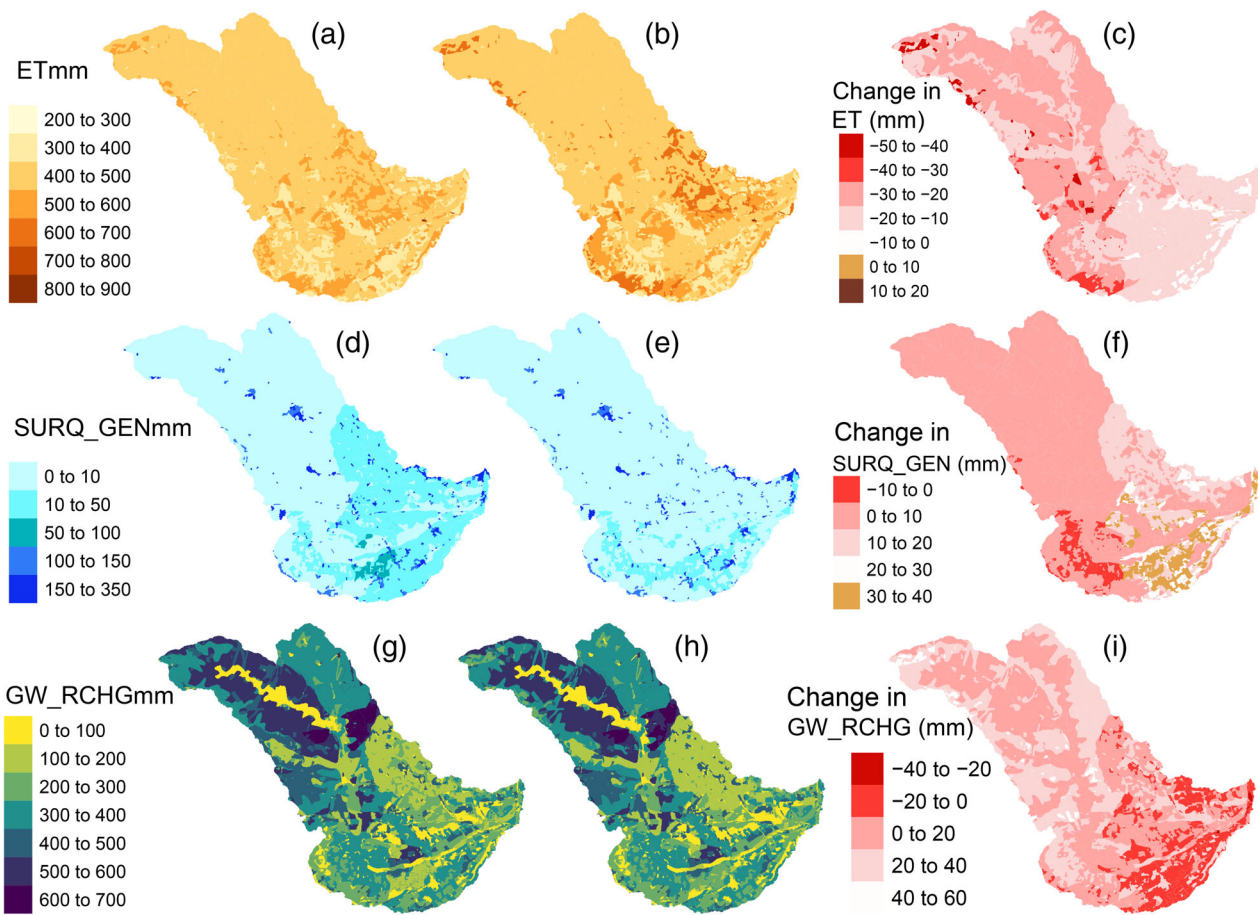
FIGURE 7 Averages for the 2012–2014 and 2005–2008 calibrations and the change between the two of the simulated evapotranspiration (a–c), surface runoff (d–f) and groundwater recharge (g–i) respectively.

3.2 | Modelling results for the Blackwater catchment

3.2.1 | Model performance

The 2012–2014 calibration and 2005–2008 validation indicate in general poor values of P-factor (between 0 and 0.7) and R-factor (from 0.2 to 1.7). The visual inspection shows that the model performs relatively better for most of the sub-catchments except on the Whitewater at Lodge Farm underlain by the chalk where the model barely captures the dynamics of the observed flows and has higher uncertainties during the calibration and validation (Figure 8a,b). On the same Whitewater sub-catchment downstream on the clay bedrock geology at Holdshot Farm, the quality of the model responsiveness improves but it overestimates the baseflow and peak flows during the calibration and validation (Figure 8c,d). Missing data during the validation period makes it challenging to gain further insights into the model performance. For the Hart at Bramshill House and the Blackwater at Farnborough where data are available only for the calibration period,

the model performs relatively better but hardly captures certain peak flows (Figure 8e,f). The best performance of the model is obtained for the Blackwater at Swallowfield (Figure 8i,j).

For the 2005–2008 calibration and 2012–2014 validation, the values of P-factor (0–0.6) and R-factor (0.4–2) at most stations, except at Lodge Farm during the calibration, still indicate an overall poor fit. However, in general, an improved model performance over the two periods without any drastic differences compared to 2012–2014 calibration/2005–2008 validation can be observed. The most important change can be observed for the Whitewater at Lodge Farm, where a larger prediction uncertainty is obtained (Figure 9a,b). For the Whitewater at Holdshot Farm, there is no noteworthy change in the performance criteria during the calibration and validation (Figure 9c,d) but the model produces higher peak flows compared to 2012–2014 calibration/2005–2008 validation. This is also observed on the Hart at Bramshill House (Figure 9e,f), the Blackwater at Swallowfield (Figure 9g,h).
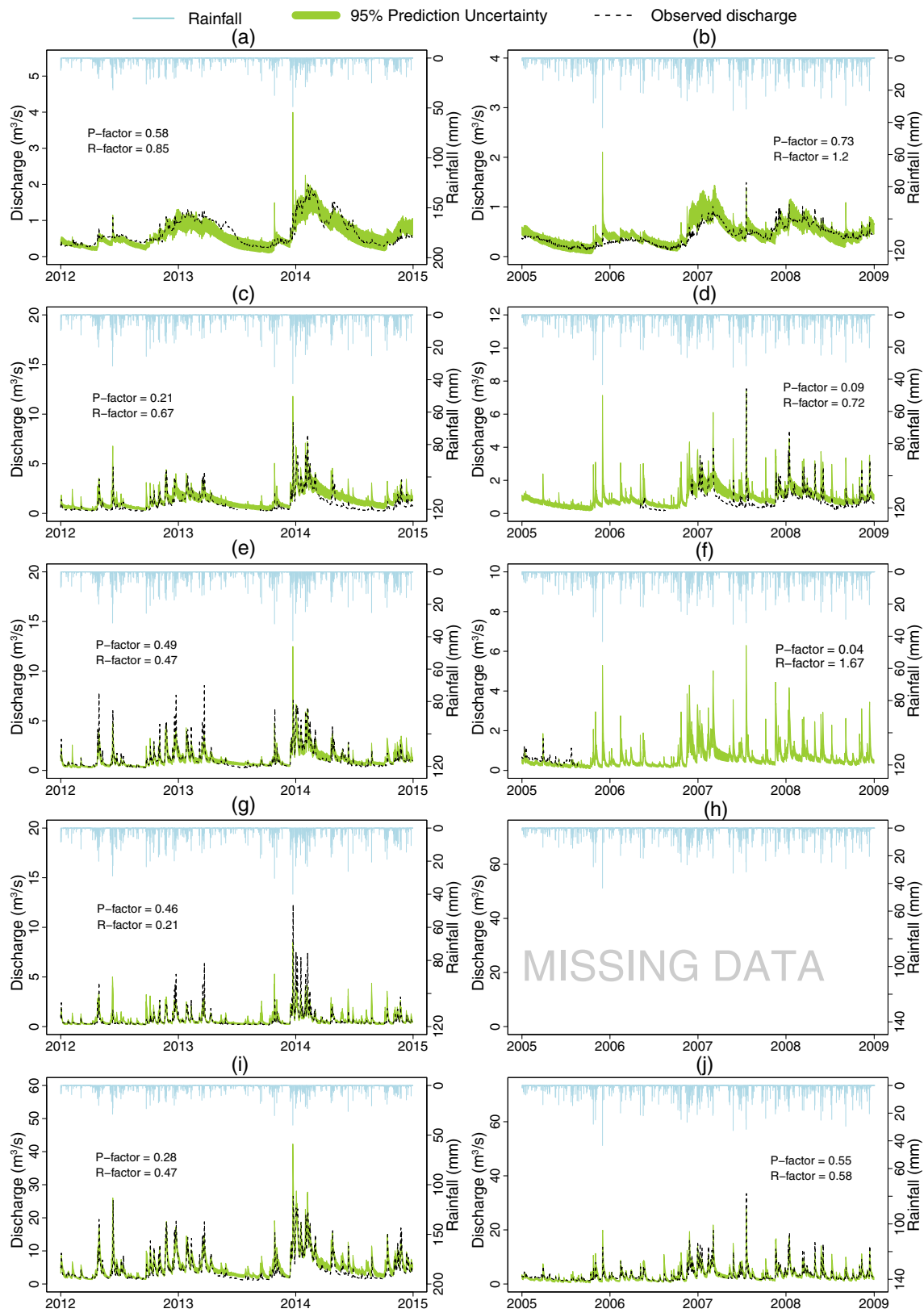
FIGURE 8 Model performance during the 2012–2014 calibration (with groundwater flood) and 2005–2008 validation (with rainwater flood) for the Whitewater at Lodge Farm (a,b), the Whitewater at Holdshot Farm (c,d), the Hart at Bramshill House (e,f), the Blackwater at Farnborough (g,h), and the Blackwater at Swallowfield (h,i).
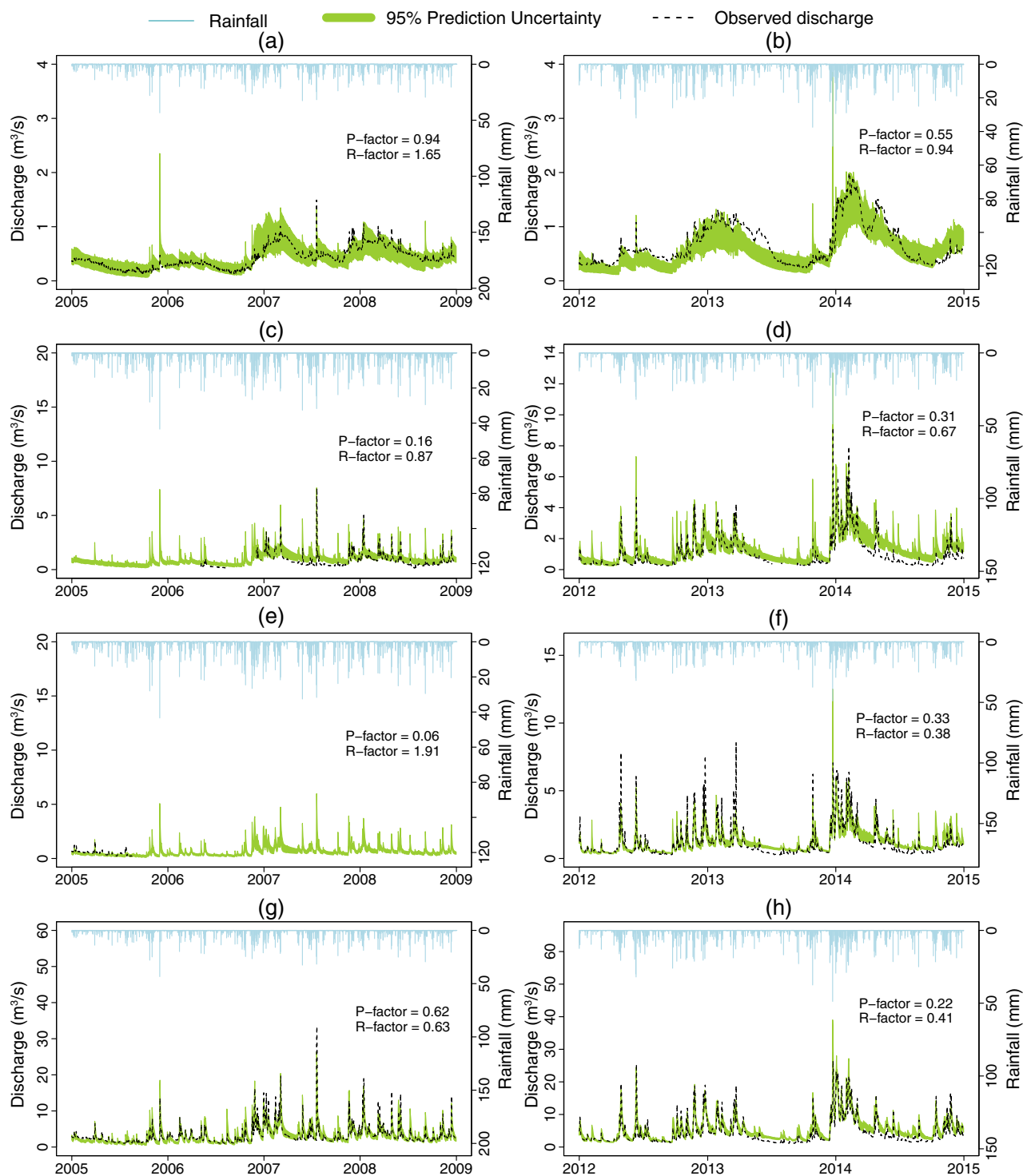
FIGURE 9 Model performance during the 2005–2008 calibration (with rainwater flood) and 2012–2014 validation (with groundwater flood) for the Whitewater at Lodge Farm (a,b), the Whitewater at Holdshot Farm (c,d), the Hart at Bramshill House (e,f) and the Blackwater at Swallowfield (g,h); data for Blackwater at Farnborough is missing.

3.2.2 | Verification of parameters

For most of the groups of subbasins, the distributions of the parameter values do not change between calibrations markedly except for BIII where the median value of

SOL_AWC has dropped from approximately 0.2 for 2012–2014 calibration to around −0.1 for 2005–2008 calibration (Figure 10d) suggesting an increase in the water available for surface runoff and baseflow. An increase in the median value of CN2 can also be observed for the
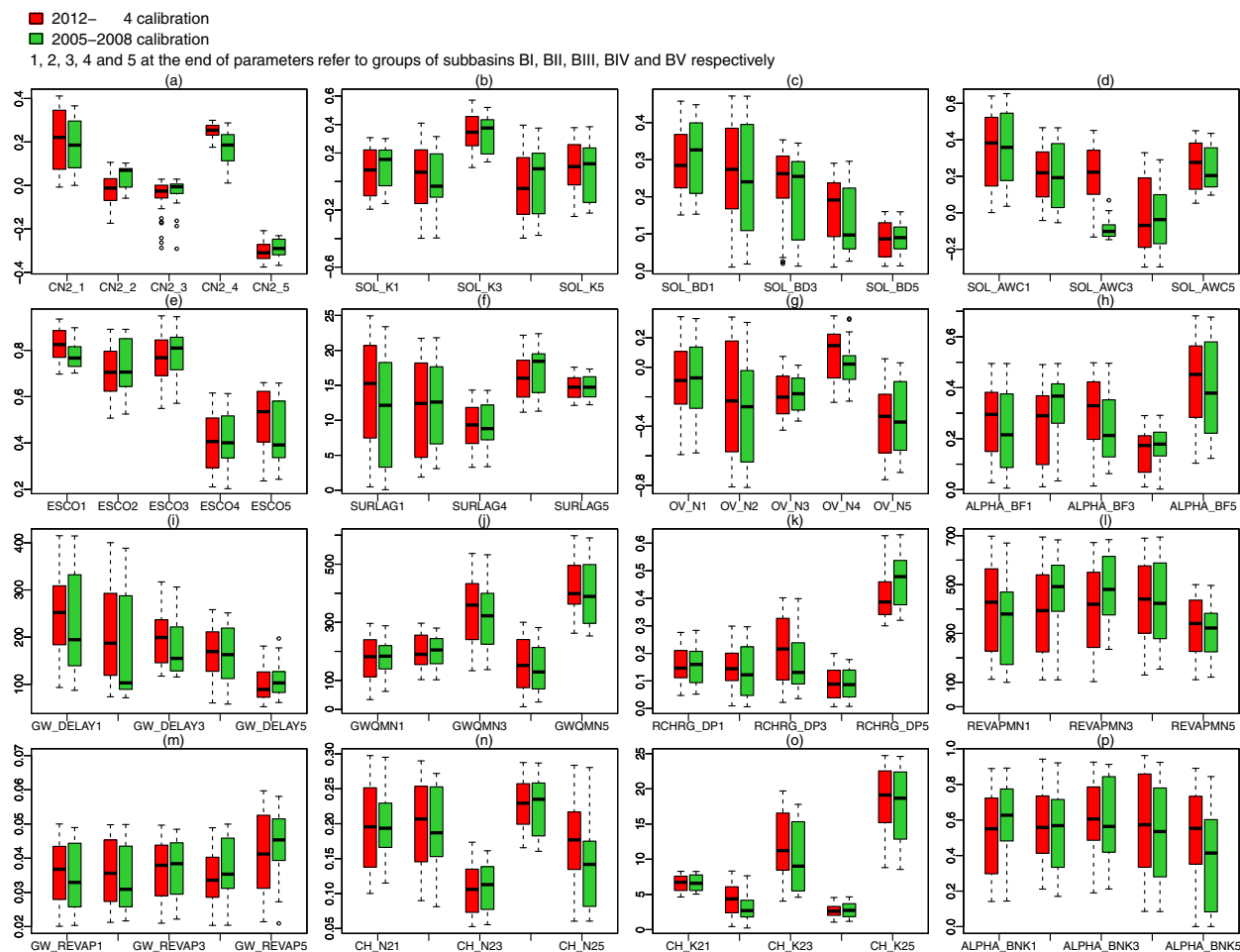
FIGURE 10 Model parameters from two calibrations on the Blackwater catchment.

2005–2008 calibration for BII for the Whitewater at Holdshot Farm.

3.2.3 | Verification of parameter distribution densities across the landscapes

Selecting the parameter set from the 2012–2014 calibration, the parameters' probability densities for the groups of subbasins show a consistent variation across landscape characteristics (Figure 11). For several parameters related to infiltration, soil water and groundwater, a clear difference can be observed in the densities for the Whitewater at Lodge Farm compared to those of other sub-catchments. For instance, the lowest values for CN2 and higher RCHRG_DP, implying higher infiltration and groundwater recharge, are observed on the sub-catchment underlain by the chalk. Conversely, distributions centred on higher CN2 and lower RCHRG_DP values are observed both on the Blackwater at Farnborough and Swallowfield indicating higher surface runoff and lower recharge, respectively. Furthermore, a peak in the probability density at higher ALPHA_BF is observed

for the Whitewater at lodge Farm indicating a more rapid land response to recharge compared to other sub-catchments. This response is also consistent with the observed density of the GW_DELAY, which is lower for the Whitewater at Lodge farm.

The densities of the maximum canopy storage parameter, CANMX, which was calibrated catchment-wide and plotted according to land cover types, indicate that higher interception occurs on forest and rangeland, with values of around 1.5 and 1.25 mm, respectively, while agricultural land and pasture have lower values (Figure 12). In addition, agricultural land and pasture tend to have similar maximum canopy storage capacity.

3.3 | Test of the broadscale NFM scenarios with the calibrated models

3.3.1 | Blanket conversion to broadleaf woodland

The broadscale land conversion to broadleaf woodland except for areas covered by water and urban areas in the
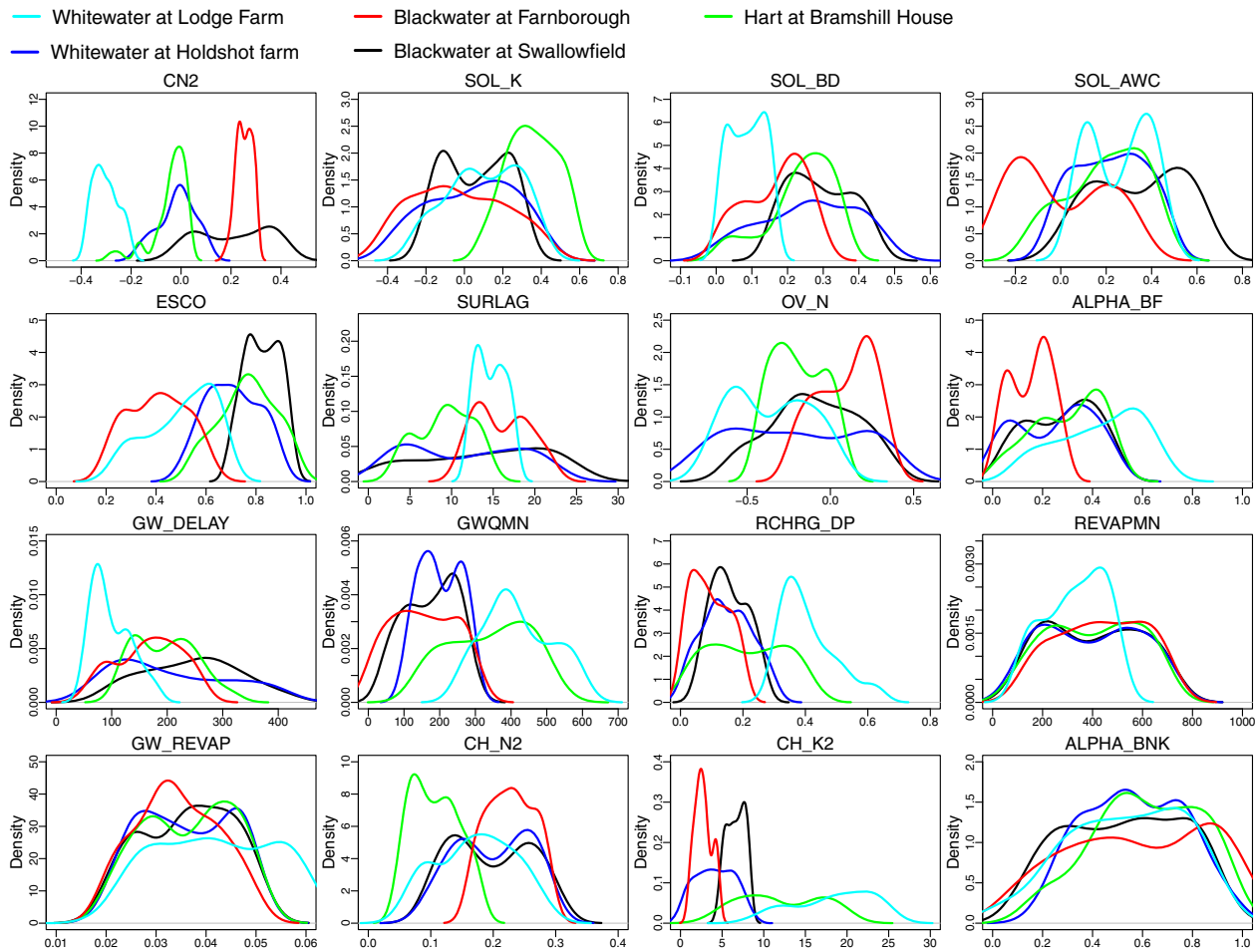
FIGURE 11 Probability densities of the calibration parameters for different sub-catchments (groups of subbasins) within the Blackwater catchment.
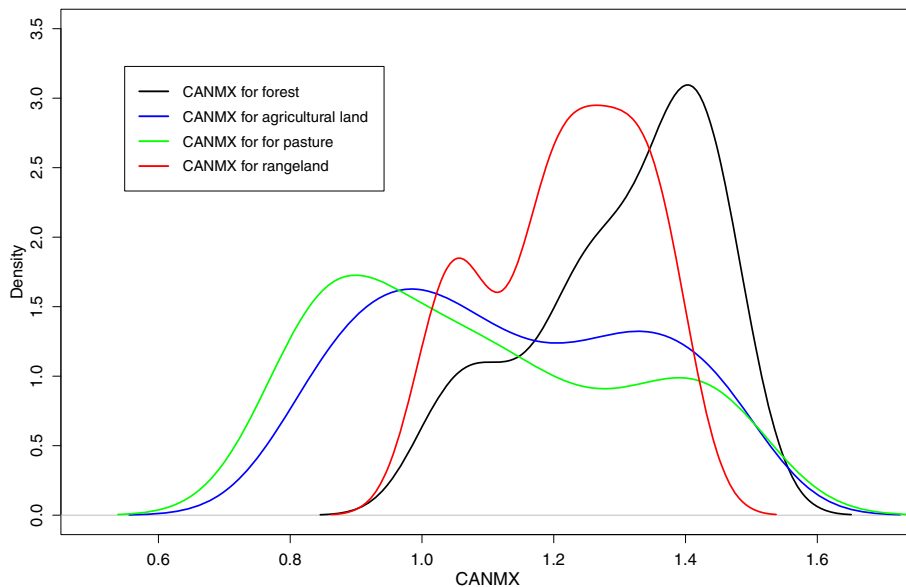

FIGURE 12 Density curves of the CANMX parameter distribution for different land use within the Blackwater catchment.

Pang (i.e., from 16% to 96% of catchment covered by broadleaf woodland) shows a shift in the 95% streamflow prediction uncertainty indicating a reduction in the streamflow (Figure 13a). The Kolmogorov–Smornov test gives the evidence that there are significant differences in the distributions of the large floods, small floods and
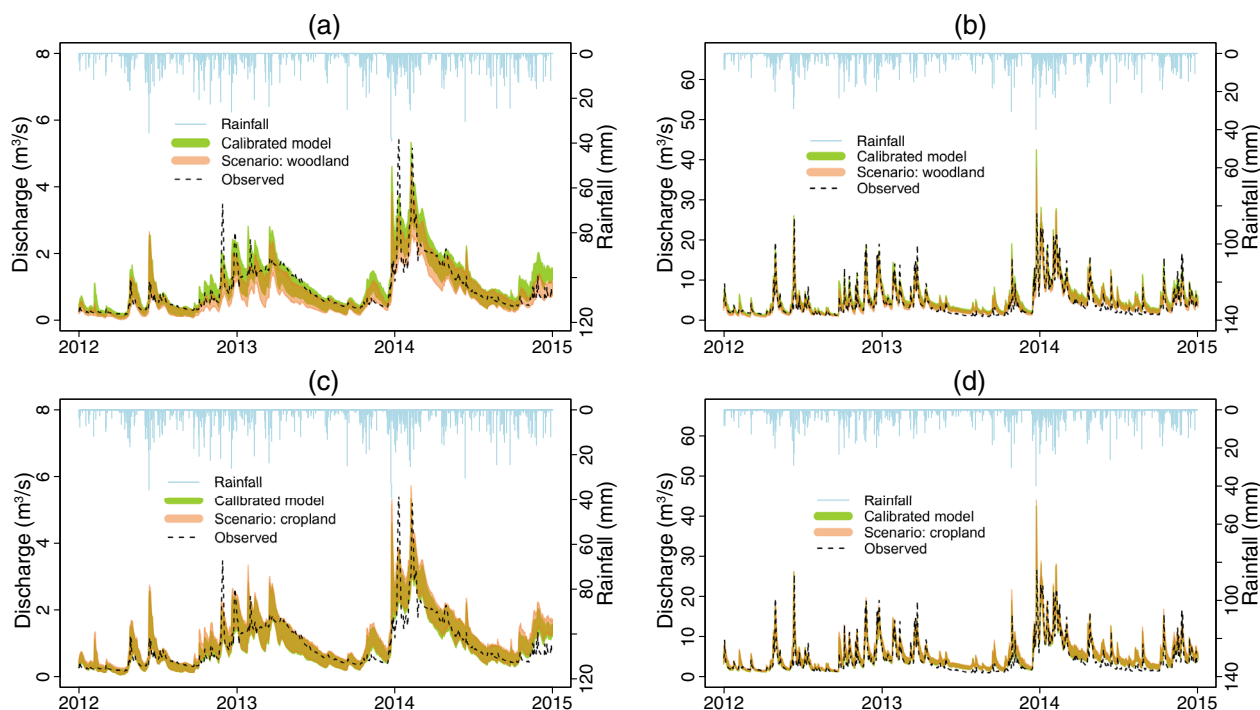
Journal of Flood Risk Management—WILEY

FIGURE 13 Effect of blanket land conversion to broadleaf woodland in (a) the Pang and (b) the Blackwater catchments; and blanket land conversion to cropland in (c) the Pang and (d) the Blackwater catchments between 2012 and 2014.

high-flow pulses extracted from the calibration and the modelled scenario. Estimations reveal a reduction of the medians between 12% and 14% of large floods, 15% and 16% of small floods, and 21% and 28% of high-flow pulses following woodland planting (Figure 14a–c).

Similar shift towards a reduction in the streamflow is observed in the Blackwater catchment following the blanket conversion to broadleaf woodland (i.e., from 15% to 74% of the catchment covered by woodland) (Figure 13b). Statistically significant differences in the distributions of the large floods, small floods and high-flow pulses extracted from the calibration and the modelled scenario were also obtained with reductions between 4% and 15%, 12% and 16%, and 17% and 29% in the medians of large floods, small floods and high-flow pulses, respectively (Figure 14d–f).

3.3.2 | Blanket conversion to cropland

The broadscale land conversion to cropland in the Pang at Pangbourne (from 54% to 96% of the catchment) and in the Blackwater (from 21% to 74% of the catchment) show an increasing shift of the 95% streamflow prediction uncertainty in the two catchments (Figure 13c,d). The Kolmogorov–Smirnov test shows that there is sufficient evidence to conclude that the large floods, small floods and high-flow pulses extracted from the calibration and

the modelled scenario in the two catchments come from different distributions.

For the Pang catchment, the is an increase between 5% and 7%, 5% and 6%, and 7% to 8% in the medians of large floods, small floods and high-flow pulses, respectively (Figure 14g–i). For the Blackwater at Swallowfield, the change is balanced between a reduction from 0 to 1% and an increase between 0% and 10% of the median of large floods while an increase between 5% and 8% and 8% to 9% are obtained in the medians of small floods and high-flow pulses respectively (Figure 14j–l).

4 | DISCUSSION

4.1 | Variation of SWAT model performance as affected by landscape characteristics, processes configuration and calibration techniques

For NFM modelling in a catchment with heterogenous landscapes, a model that can account for the heterogeneities of soil and land use and perform well across landscapes is essential.

In this study, the model performance varies across the catchments depending on landscape characteristics in terms of soil and geology, modelling technique as well as the cause of the flooding event (i.e., pluvial or
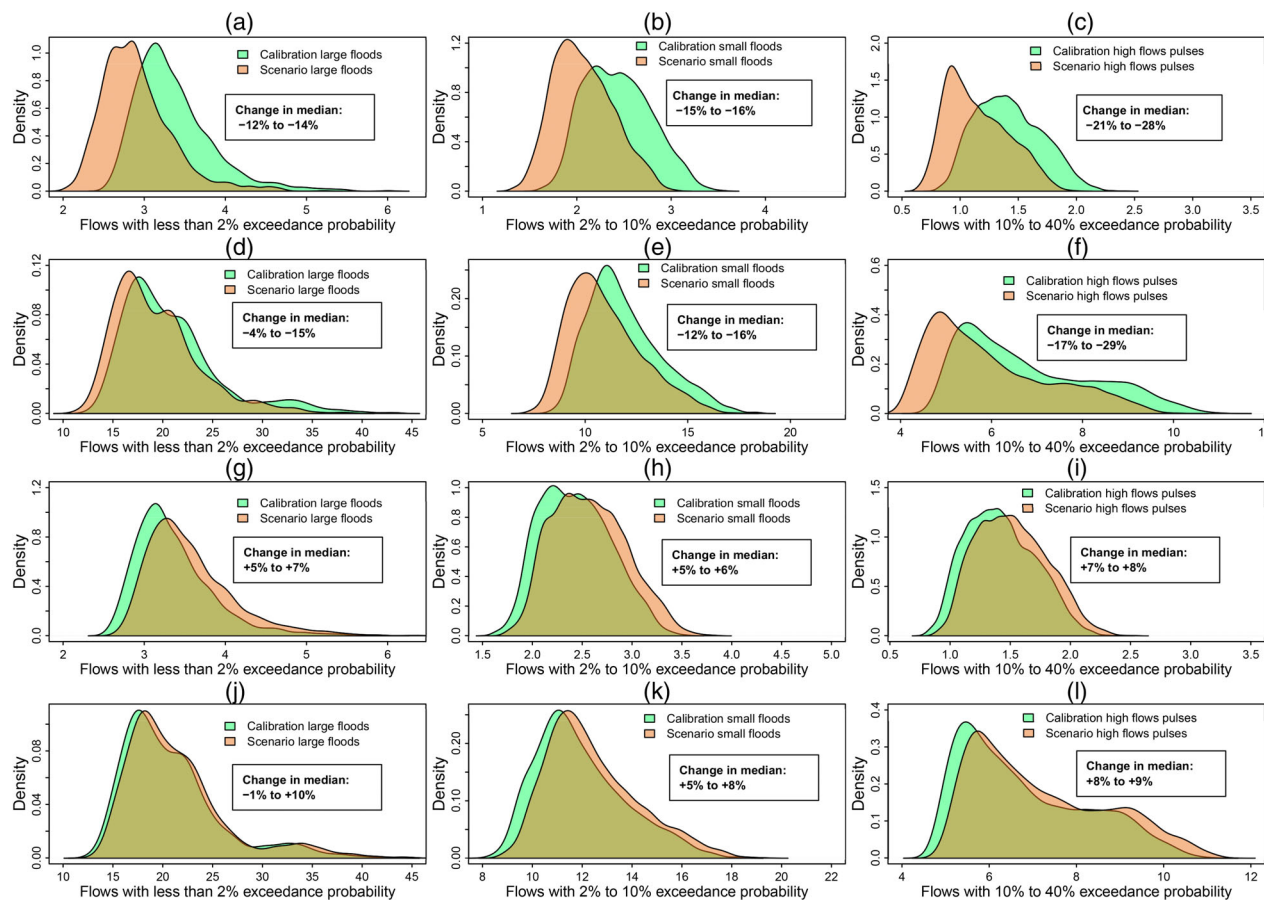
FIGURE 14 Changes in large floods, small floods and high-flow pulses from the broadleaf woodland scenario in the Pang (a–c) and Blackwater (d–f) catchments; and from the cropland scenario in the Pang (g–i) and Blackwater (j–l) catchments.

groundwater flooding). The poor performance of the model upstream in the Pang catchment at Frilsham and Bucklebury compared to Pangbourne reflects the complexity of the catchment hydrological system mainly governed by regional groundwater dynamics. The groundwater catchment has been shown to differ from the topographic catchment in the Pang, as regional groundwater flow is south-east towards the deeply incised River Thames (Jackson et al., 2006a; Jackson, Wheater, et al., 2006; Wheater et al., 2007) and the Pang acting as a drain that "switches on" when groundwater levels are high. The effect will be most pronounced in the upper Pang and thus helps explain why model performance at Frilsham and Bucklebury is poor, and, moreover, why model parameters calibrated against a period of high groundwater levels (2012–2014) are unable to reproduce groundwater baseflow during a time of low/average groundwater levels (2003–2008). Whereas streamflow at Frilsham is almost exclusively from the Chalk aquifer, streamflow at Pangbourne also has a surface runoff contribution from the Palaeogene clay deposits, which may further explain why model

performance at Pangbourne is better than at Frilsham and Bucklebury. SWAT conceptualises each HRU as an independent aquifer system (comprising a shallow unconfined aquifer and a deeper confined aquifer). To simulate hydrology in the Pang in a way that respects our understanding of the catchment, it is necessary to model lateral groundwater flow between HRUs as well as with neighbouring catchments. While the SWAT conceptualisation of the aquifer as two stacked aquifers is not completely consistent with the understanding of the hydrogeology of the Pang (Jackson et al., 2006a), it does allow groundwater to be lost from the system, accounting for the flow of groundwater beneath the Pang surface water catchment to discharge to the River Thames. The weak performance of the model predominately related to the simulation of groundwater can also be seen in the parameter shifts between the two calibration periods (Figure 5). In the 2005–2008 calibration, SOL_AWC increases, indicating a decrease in baseflow; GWQMN and RCHRG_DP increase, implying a reduction in baseflow generation and an increase in recharge; and REVAPMN decreases, meaning more water is allowed to

move back from the shallow aquifer to the unsaturated zone and is subsequently lost via evapotranspiration.

The fact that the simultaneous multisite calibration technique helped to achieve reliable results across different landscapes in the study catchments provides compelling support to previous work (Brighenti et al., 2019; Leta et al., 2017), which demonstrated the advantages of this calibration technique. As concluded by Leta et al. (2017), in catchments with high spatial heterogeneities in soils and landscapes as well as geology, and subsequently spatial variations in streamflow generation processes, this simultaneous multisite calibration should be preferred depending on the scale of the catchment. If the outlets are hydrologically connected, the simultaneous multisite calibration can be coupled with a sequential approach that consists of narrowing the parameter intervals from upstream to downstream as used in this study. This has the advantage of minimising the identifiability issue and quickly achieving the calibration.

4.2 | Performance of SWAT model for areas with different streamflow generation processes

Due to the variation of the SWAT model performance in different landscapes, it is important to understand where it performs well, and the reasons behind this, in order to guide future applications involving the model. The superior performance of the model at the Pang at Pangbourne and at sub-catchments of the Blackwater with a smaller groundwater contribution—that is, excluding gauges on the Whitewater—confirms that the model performs well in areas where streamflow generation is dominated by surface and shallow subsurface processes. In order to improve the performance of the model in catchments underlain by Chalk, SWAT would need to be coupled with a regional groundwater flow model. Although groundwater flow models of Chalk aquifers—including of the Pang (Jackson et al., 2011)—work well at a coarser temporal resolution as water resources models, modelling flood peaks on a daily time step is more challenging (Collins et al., 2020). In part, this is due to the complexity of interactions between matrix and fractures that govern flow through the Chalk unsaturated zone (Collins et al., 2020; Ireson et al., 2009; Jackson et al., 2006a; Rahman & Rosolem, 2017). These authors proposed different methods to improve water flow simulation in the unsaturated zone of Chalk including a decoupled soil layer approach (Mathias et al., 2006), a coupled discrete soil layer approach (Brouyère, 2006), and a multi-layered approach which represents both soil and weathered chalk (Ireson et al., 2009) with the application of the dual porosity Richards equation to groundwater flow through the matrix. These approaches add a considerable number of parameters which are difficult to measure and increase the computational time, making the robust uncertainty analysis used in this study more challenging to apply. This suggests that NFM modelling in Chalk catchments may have a number of obstacles to overcome to reach the level of confidence required by stakeholders to invest in NFM schemes.

4.3 | Test of the effectiveness of NFM in the selected catchments: The effect of size of flood event

The reduction of peak flows from the blanket land conversion to broadleaf woodland and their increase resulting from conversion to cropland, are consistent with previous studies on the effects of catchment woodland on the hydrological cycle and flood risks (Bosch & Hewlett, 1982; Thomas & Nisbet, 2007), as well as on the impacts of crop-land expansion on the streamflow (Rogger et al., 2017). Moreover, the relatively modest effect of broadscale land conversion to broadleaf woodland on the large flood events suggests that a worthwhile reduction of large flood flows in medium-sized catchments will require very extensive afforestation. Buechel et al. (2022) found similar effects of afforestation on high and very high flows in twelve catchments ranging in size from 500 to 10,000 km². They also showed that the planting extent has a stronger control on streamflow than the location of planting within the catchment. Crooks et al. (2000) tested two scenarios consisting of 100% grassland and 100% forest (50% deciduous and 50% coniferous), respectively, over 10,000 km² of Thames basin and found a decrease in flood frequency. Moreover, results of the EUropean River Flood Occurrence & Total Risk Assessment System (EUROTAS) project have shown for the Thames basin that a realistic scenario for urbanisation of 2% of the catchment leads to an increase of flood discharge within the uncertainty of the modelling process, and that the effects of urbanisation on flood runoff are localised (Samuels, 2001a, 2001b).

If the afforestation requires a substantial catchment area for a sensible reduction in peak flow, a small change in the flood flow can change the flood-risk of individual stakeholders and therefore is important in terms of flood risk reduction.

Results demonstrate that the model setup, process configuration, parametrisation as well as CUA used in this study enables NFM assessment, especially in areas where streamflow generation is dominated by surface and shallow subsurface processes. Such a consistent modelling framework provides means for investigating

20 of 23 | WILEY—**CIWEM** Chartered Institution of Water and Environmental Management Journal of Flood Risk Management

BADJANA ET AL.

the effects of different woodland types not only for NFM but also to guide the development of future tree planting proposals as raised by Cooper et al. (2021), for example, to avoid the occurrence of undesirable low flows. For the groundwater-dominated areas, challenges remain, requiring therefore a combination of techniques, analysis and reasoning including expert knowledge for a meaningful and objective conclusion regarding the effect of NFM.

The overall disappointing values of P-factor and R-factor, though somehow dependent on the threshold values of the efficiency criteria used for separating behavioural from non-behavioural simulations, indicate that the tested modelling framework only provides an indication of NFM effectiveness in the selected catchments particularly in areas dominated by surface and fast subsurface processes. For robust policy recommendation on NFM implementation in the selected catchments, a systematic improvement of the model performance will be necessary.

4.4 | NFM modelling at medium-scale catchments: Requirements and importance of the tested modelling framework

Overall, this study demonstrates that it is possible to reliably model land-based NFM in medium-scale catchments, but that success is highly dependent on the one hand on catchment landscape characteristics and on the other hand on the model choice, configuration, parameterisation, and CUA techniques. This reflects the importance of the tested modelling framework, which provides evidence on the feasibility of NFM modelling, using models like SWAT in complex landscapes with different hydrogeologic properties.

From the practitioners' point of view, the study can be seen as a supportive technical guideline for any NFM modelling and implementation in heterogenous medium-scale catchments. In fact, the proposed modelling framework allows its users to improve their understanding of catchments characteristics and NFM-related processes, to test the model suitability, identify its best set up and parametrisation and finally but not least account for uncertainties. The tested modelling framework also highlights the importance of model parametrisation, as well as CUA techniques, in catchments with heterogenous landscapes and complex hydrology.

5 | CONCLUSION

In this study, we tested how the SWAT model can be used to simulate NFM in medium-scale lowland catchments, while considering uncertainties by exploring the

effect of using different process configuration and calibration techniques. We found that in catchments with heterogenous soils, landscapes and geology, the model performance can be improved by using simultaneous multi-site CUA techniques compared to simple multisite calibration. Depending on whether stations are hydrologically connected, a cascading approach which adjusts parameters from upstream to downstream will enhance the calibration. The model performed better in areas where streamflow generation is dominated by surface and fast subsurface processes but poorly in groundwater-dominated areas. The use of catchment hydrological models to simulate groundwater-dominated catchments may help to understand the system functioning but does not offer the means to assess NFM measures given the uncertainties associated with the modelling results.

Apart from the feasibility of NFM modelling in medium-scale catchments, and based on the better performance in areas dominated by surface and interflow processes where the model performs better, it is likely that the NFM seems to be less effective on large floods in comparison to small floods and high-flow pulses, and the effectiveness is likely to vary according to storm conditions. However, further investigation is required to understand how the processes related to NFM change as a function of rainfall intensity, duration and frequency, and the interactions with groundwater.

Our methodological approach allows for a holistic assessment of the modelling framework to highlight the suitability and limitations of its components (data, methods and techniques). The proposed modelling framework provides the means to fully understand the functioning of the catchment system and how the NFM measures change the hydrological processes in areas where streamflow generation mechanisms are dominated by surface and fast subsurface processes. However, it does not provide means for accurately estimating the potential contribution of NFM in groundwater-dominated areas. We conclude that prior to hydrological impact modelling in the context of NFM, it is necessary to understand both the target flood generation mechanism in relation to catchment landscape properties and processes, and the model ability to simulate this, since the translation of landscape features related to NFM into tangible hydrological model configurations and parameters is not simple and straightforward. NFM modelling should only be used for policy recommendations for NFM implementation if the results can be assessed as robust and reliable. Our study provides an approach designed to ensure that this is the case and we hope that it also improves the perception of NFM modelling within the hydrological community. Our findings in particular highlight the need to further develop integrated hydrological simulation tools for groundwater-dominated catchments.

DATA AVAILABILITY STATEMENT

Part of the data that support the findings of this study are openly available from https://nrfa.ceh.ac.uk/, https://data.gov.uk/dataset/8275e71e-1516-42a1-bb0c-4fa73807fe2b/lidar-dtm-time-stamped-tiles, https://soilgrids.org/, https://www.bgs.ac.uk/geological-data/national-geological-repository/. The UK Centre for Ecology & Hydrology (UKCEH) land cover maps are available on request from https://www.ceh.ac.uk/ukceh-land-cover-maps. The National Soil Map and England and Wales (NATMAP) data are available from Cranfield University. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from http://www.landis.org.uk/data/natmap.cfm with the permission of Cranfield University.

ORCID

Hèou Maléki Badjana https://orcid.org/0000-0002-1420-2864
Sarah Collins https://orcid.org/0000-0001-5124-9699

REFERENCES

Abbaspour, K., Vaghefi, S., & Srinivasan, R. (2017). A guideline for successful calibration and uncertainty analysis for soil and water assessment: A review of papers from the 2016 international SWAT conference. *Water*, *10*(1), 6. https://doi.org/10.3390/w10010006

Aliyari, F., Bailey, R. T., Tasdighi, A., Dozier, A., Arabi, M., & Zeiler, K. (2019). Coupled SWAT-MODFLOW model for large-scale mixed agro-urban river basins. *Environmental Modelling & Software*, *115*, 200–210. https://doi.org/10.1016/j.envsoft.2019.02.014

Allen, D. J., Brewerton, L. J., Coleby, L. M., Gibbs, B. R., Lewis, M. A., MacDonald, A. M., Wagstaff, S. J., & Williams, A. T. (1997). The physical properties of major aquifers in England and Wales (technical report WD/97/34; Environment Agency R&D Publication 8, p. 312). *British Geological Survey* http://nora.nerc.ac.uk/id/eprint/13137/1/WD97034.pdf

André, F., Jonard, M., & Ponette, Q. (2008). Precipitation water storage capacity in a temperate mixed oak–beech canopy. *Hydrological Processes*, *22*(20), 4130–4141. https://doi.org/10.1002/hyp.7013

Ascott, M. J., Marchant, B. P., Macdonald, D., McKenzie, A. A., & Bloomfield, J. P. (2017). Improved understanding of spatio-temporal controls on regional scale groundwater flooding using hydrograph analysis and impulse response functions. *Hydrological Processes*, *31*(25), 4586–4599. https://doi.org/10.1002/hyp.11380

Bailey, R. T., Wible, T. C., Arabi, M., Records, R. M., & Ditty, J. (2016). Assessing regional-scale spatio-temporal patterns of groundwater–surface water interactions using a coupled SWAT-MODFLOW model. *Hydrological Processes*, *30*(23), 4420–4433. https://doi.org/10.1002/hyp.10933

Bosch, J. M., & Hewlett, J. D. (1982). A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. *Journal of Hydrology*, *55*(1), 3–23. https://doi.org/10.1016/0022-1694(82)90117-2

Bradford, R. B. (2002). Controls on the discharge of chalk streams of the Berkshire downs, UK. *Science of the Total Environment*, *282–283*, 65–80. https://doi.org/10.1016/S0048-9697(01)00954-8

Brighenti, T. M., Bonumá, N. B., Grison, F., Mota, A. d. A., Kobiyama, M., & Chaffe, P. L. B. (2019). Two calibration methods for modeling streamflow and suspended sediment with the swat model. *Ecological Engineering*, *127*, 103–113. https://doi.org/10.1016/j.ecoleng.2018.11.007

Brouyère, S. (2006). Modelling the migration of contaminants through variably saturated dual-porosity, dual-permeability chalk. *Journal of Contaminant Hydrology*, *82*(3–4), 195–219. https://doi.org/10.1016/j.jconhyd.2005.10.004

Buechel, M., Slater, L., & Dadson, S. (2022). Hydrological impact of widespread afforestation in Great Britain using a large ensemble of modelled scenarios. *Communications Earth & Environment*, *3*(1), 1–10. https://doi.org/10.1038/s43247-021-00334-0

Camargos, C., Julich, S., Houska, T., Bach, M., & Breuer, L. (2018). Effects of input data content on the uncertainty of simulating water resources. *Water*, *10*(5), 621. https://doi.org/10.3390/w10050621

Chunn, D., Faramarzi, M., Smerdon, B., & Alessi, D. (2019). Application of an integrated SWAT–MODFLOW model to evaluate potential impacts of climate change and water withdrawals on groundwater–surface water interactions in west-Central Alberta. *Water*, *11*(1), 110. https://doi.org/10.3390/w11010110

Cloke, H. L., Pappenberger, F., & Renaud, J.-P. (2008). Multi-method global sensitivity analysis (MMGSA) for modelling floodplain hydrological processes. *Hydrological Processes*, *22*(11), 1660–1674. https://doi.org/10.1002/hyp.6734

Collins, S. L., Christelis, V., Jackson, C. R., Mansour, M. M., Macdonald, D. M. J., & Barkwith, A. K. A. P. (2020). Towards integrated flood inundation modelling in groundwater-dominated catchments. *Journal of Hydrology*, *591*, 125755. https://doi.org/10.1016/j.jhydrol.2020.125755

Connelly, A., Snow, A., Carter, J., & Lauwerijssen, R. (2020). What approaches exist to evaluate the effectiveness of UK-relevant natural flood management measures? *A Systematic Map Protocol. Environmental Evidence*, *9*(1), 11–13. https://doi.org/10.1186/s13750-020-00192-x

Cooper, M. M. D., Patil, S. D., Nisbet, T. R., Thomas, H., Smith, A. R., & McDonald, M. A. (2021). Role of forested land for natural flood management in the UK: A review. *WIREs Water*, *8*(5), e1541. https://doi.org/10.1002/wat2.1541

Cranfield University. (2013). NATMAP—National Soil map from England and Wales. http://www.landis.org.uk/data/natmap.cfmz

Crooks, S., Davies, H., & Goodsell, G. (2000). Rainfall—Runoff modelling and the impact of land use change in the Thames catchment. In *Paper presented at the European Conference On Advances In Flood Research, Potsdam Institute for Climate Impact Research* (pp. 111–126). Potsdam Institute for Climate Impact Research.

Dadson, S. J., Hall, J. W., Murgatroyd, A., Acreman, M., Bates, P., Beven, K., Heathwaite, L., Holden, J., Holman, I. P., Lane, S. N., O'Connell, E., Penning-Rowsell, E., Reynard, N., Sear, D., Thorne, C., & Wilby, R. (2017). A restatement of the natural science evidence concerning catchment-based 'natural' flood management in the UK. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *473*(2199), 20160706. https://doi.org/10.1098/rspa.2016.0706

Ellis, N., Anderson, K., & Brazier, R. (2021). Mainstreaming natural flood management: A proposed research framework derived from a critical evaluation of current knowledge. *Progress in Physical Geography: Earth and Environment*, *45*(6), 819–841. https://doi.org/10.1177/0309133321997299

Griffiths, J., Binley, A., Crook, N., Nutter, J., Young, A., & Fletcher, S. (2006). Streamflow generation in the Pang and Lambourn catchments, Berkshire, UK. *Journal of Hydrology*, *330*(1), 71–83. https://doi.org/10.1016/j.jhydrol.2006.04.044

Guse, B., Pfannerstill, M., Gafurov, A., Fohrer, N., & Gupta, H. (2016). Demasking the integrated information of discharge: Advancing sensitivity analysis to consider different hydrological components and their rates of change. *Water Resources Research*, *52*(11), 8724–8743. https://doi.org/10.1002/2016WR018894

Hankin, B., Metcalfe, P., Johnson, D., Chappell, N. A., Page, T., Craigen, I., Lamb, R., & Beven, K. (2017). Strategies for Testing the Impact of Natural Flood Risk Management Measures. In T. Hromadka & P. Rao (Eds.), *Flood Risk Management*. IntechOpen. https://doi.org/10.5772/intechopen.68677

Houska, T., Kraft, P., Chamorro-Chavez, A., & Breuer, L. (2015). SPOTting model parameters using a ready-made python package. *PLoS One*, *10*(12), e0145180. https://doi.org/10.1371/journal.pone.0145180

Hughes, A. G., Vounaki, T., Peach, D. W., Ireson, A. M., Jackson, C. R., Butler, A. P., Bloomfield, J. P., Finch, J., & Wheater, H. S. (2011). Flood risk from groundwater: Examples from a chalk catchment in southern England: Flood risk from groundwater. *Journal of Flood Risk Management*, *4*(3), 143–155. https://doi.org/10.1111/j.1753-318X.2011.01095.x

Ireson, A. M., Mathias, S. A., Wheater, H. S., Butler, A. P., & Finch, J. (2009). A model for flow in the chalk unsaturated zone incorporating progressive weathering. *Journal of Hydrology*, *365*(3–4), 244–260. https://doi.org/10.1016/j.jhydrol.2008.11.043

Jackson, B. M., Wheater, H. S., Mathias, S. A., McIntyre, N., & Butler, A. P. (2006). A simple model of variable residence time flow and nutrient transport in the chalk. *Journal of Hydrology*, *330*(1), 221–234. https://doi.org/10.1016/j.jhydrol.2006.04.045

Jackson, C. R., Hughes, A. G., & Mansour, M. M. (2006). Numerical modelling of groundwater flow to Thames water utility Ltd's Gatehampton source, (Commissioned Report CR/06/205C). British Geological Survey.

Jackson, C. R., Meister, R., & Prudhomme, C. (2011). Modelling the effects of climate change and its uncertainty on UK chalk groundwater resources from an ensemble of global climate model projections. *Journal of Hydrology*, *399*(1), 12–28. https://doi.org/10.1016/j.jhydrol.2010.12.028

Kay, A. L., Old, G. H., Bell, V. A., Davies, H. N., & Trill, E. J. (2019). An assessment of the potential for natural flood management to offset climate change impacts. *Environmental Research Letters*, *14*(4), 044017. https://doi.org/10.1088/1748-9326/aafdbe

Lane, R. A., Coxon, G., Freer, J. E., Wagener, T., Johnes, P. J., Bloomfield, J. P., Greene, S., Macleod, C. J. A., & Reaney, S. M. (2019). Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. *Hydrology and Earth System Sciences*, *23*(10), 4011–4032. https://doi.org/10.5194/hess-23-4011-2019

Leta, O. T., van Griensven, A., & Bauwens, W. (2017). Effect of single and multisite calibration techniques on the parameter estimation, performance, and output of a SWAT model of a spatially heterogeneous catchment. *Journal of Hydrologic Engineering*, *22*(3), 05016036. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001471

Liu, W., Park, S., Bailey, R. T., Molina-Navarro, E., Andersen, H. E., Thodsen, H., Nielsen, A., Jeppesen, E., Jensen, J. S., Jensen, J. B., & Trolle, D. (2020). Quantifying the streamflow response to groundwater abstractions for irrigation or drinking water at catchment scale using SWAT and SWAT–MODFLOW. *Environmental Sciences Europe*, *32*(1), 113. https://doi.org/10.1186/s12302-020-00395-6

Macdonald, D., Dixon, A., Newell, A., & Hallaways, A. (2012). Groundwater flooding within an urbanised flood plain: Groundwater flooding within urbanised flood plain. *Journal of Flood Risk Management*, *5*(1), 68–80. https://doi.org/10.1111/j.1753-318X.2011.01127.x

Mathias, S. A., Butler, A. P., Jackson, B. M., & Wheater, H. S. (2006). Transient simulations of flow and transport in the chalk unsaturated zone. *Journal of Hydrology*, *330*(1–2), 10–28. https://doi.org/10.1016/j.jhydrol.2006.04.010

McLean, L., Beevers, L. C., Pender, G., Haynes, H., & Wilkinson, M. (2013). Natural flood management in the United Kingdom: Developing a conceptual management tool. In *35th IAHR world congress 2013: The wise find pleasure in water: Meandering through water science and engineering* (Vol. *6*, pp. 4996–5004). IAHR.

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-km2 scale catchment in the United Kingdom. *Hydrological Processes*, *31*(9), 1734–1748. https://doi.org/10.1002/hyp.11140

Muchan, K., Lewis, M., Hannaford, J., & Parry, S. (2015). The winter storms of 2013/2014 in the UK: Hydrological responses and impacts. *Weather*, *70*(2), 55–61. https://doi.org/10.1002/wea.2469

National River Flow Archive. (2022). National River Flow Archive. May 18, 2022. http://nrfa.ceh.ac.uk/data/station/info/39001

Nguyen, V. T., & Dietrich, J. (2018). Modification of the SWAT model to simulate regional groundwater flow using a multicell aquifer. *Hydrological Processes*, *32*(7), 939–953. https://doi.org/10.1002/hyp.11466

Pearce, A. J., Gash, J. H. C., & Stewart, J. B. (1980). Rainfall interception in a forest stand estimated from grassland meteorological data. *Journal of Hydrology*, *46*(1), 147–163. https://doi.org/10.1016/0022-1694(80)90040-2

Perra, E., Piras, M., Deidda, R., Paniconi, C., Mascaro, G., Vivoni, E. R., Cau, P., Marras, P. A., Ludwig, R., & Meyer, S. (2018). Multimodel assessment of climate change-induced hydrologic impacts for a Mediterranean catchment. *Hydrology and Earth System Sciences*, *22*(7), 4125–4143. https://doi.org/10.5194/hess-22-4125-2018

Peters, E., & van Lanen, H. A. J. (2005). Separation of base flow from streamflow using groundwater levels—Illustrated for the pang catchment (UK). *Hydrological Processes*, *19*(4), 921–936. https://doi.org/10.1002/hyp.5548

Pfannerstill, M., Guse, B., & Fohrer, N. (2014). A multi-storage ground-water concept for the SWAT model to emphasize nonlinear groundwater dynamics in lowland catchments. *Hydrological Processes*, *28*(22), 5599–5612. https://doi.org/10.1002/hyp.10062

Rahman, M., & Rosolem, R. (2017). Towards a simple representation of chalk hydrology in land surface modelling. *Hydrology and Earth System Sciences*, *21*(1), 459–471. https://doi.org/10.5194/hess-21-459-2017

Rogger, M., Agnoletti, M., Alaoui, A., Bathurst, J. C., Bodner, G., Borga, M., Chaplot, V., Gallart, F., Glatzel, G., Hall, J., Holden, J., Holko, L., Horn, R., Kiss, A., Kohnová, S., Leitinger, G., Lennartz, B., Parajka, J., Perdigão, R., ... Blöschl, G. (2017). Land use change impacts on floods at the catchment scale: Challenges and opportunities for future research. *Water Resources Research*, *53*(7), 5209–5219. https://doi.org/10.1002/2017WR020723

Rutter, A. J., Morton, A. J., & Robins, P. C. (1975). A predictive model of rainfall interception in forests. II. Generalization of the model and comparison with observations in some conifer-ous and hardwood stands. *Journal of Applied Ecology*, *12*(1), 367–380. https://doi.org/10.2307/2401739

Samuels, P. G. (2001a). *EUROTAS European River Flood Occurrence and Total Risk Assessment System Scientific Report to the European Commission* (EUROTAS Project Report Part 6; p. 60). https://doi.org/10.13140/2.1.2002.7682

Samuels, P. G. (2001b). The European River flood occurrence and total risk assessment system (EUROTAS). In *Paper presented at the British government Ministry of Agriculture, Fisheries and Food (MAFF) Annual Conference Of River And Coastal Engineers* (pp. 11.1.1–11.1.12). HR Wallingford. https://www.researchgate.net/publication/274279617_THE_EUROPEAN_RIVER_FLOOD_OCCURRENCE_AND_TOTAL_RISK_ASSESSMENT_SYSTEM_EUROTAS

Schaarschmidt, F., & Gerhard, D. (2019). Package 'pairwiseCI' (0.1-27) [Computer software]. https://cran.r-project.org/web/packages/pairwiseCI/pairwiseCI.pdf

SEPA. (2016). Natural flood management handbook. https://www.sepa.org.uk/media/163560/sepa-natural-flood-management-handbook1.pdf

Thomas, H., & Nisbet, T. R. (2007). An assessment of the impact of floodplain woodland on flood flows. *Water and Environment Journal*, *21*(2), 114–126. https://doi.org/10.1111/j.1747-6593.2006.00056.x

Wagener, T., Dadson, S. J., Hannah, D. M., Coxon, G., Beven, K., Bloomfield, J. P., Buytaert, W., Cloke, H., Bates, P., Holden, J.,

Parry, L., Lamb, R., Chappell, N. A., Fry, M., & Old, G. (2021). Knowledge gaps in our perceptual model of Great Britain's hydrology. *Hydrological Processes*, *35*(7), e14288. https://doi.org/10.1002/hyp.14288

Wangpimool, W., Pongput, K., Sukvibool, C., Sombatpanit, S., & Gassman, P. W. (2013). The effect of reforestation on stream flow in upper Nan river basin using soil and water assessment tool (SWAT) model. *International Soil and Water Conservation Research*, *1*(2), 53–63. https://doi.org/10.1016/S2095-6339(15)30039-3

Waylen, K. A., Holstead, K. L., Colley, K., & Hopkins, J. (2018). Challenges to enabling and implementing natural flood Man-agement in Scotland. *Journal of Flood Risk Management*, *11*(S2), S1078–S1089. https://doi.org/10.1111/jfr3.12301

Wheater, H. S., Peach, D., & Binley, A. (2007). Characterising groundwater-dominated lowland catchments: The UK lowland catchment research Programme (LOCAR). *Hydrology and Earth System Sciences*, *11*(1), 108–124. https://doi.org/10.5194/hess-11-108-2007

Wilkinson, M. E., Addy, S., Quinn, P. F., & Stutter, M. (2019). Natu-ral 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

Wu, H., Chen, B., Ye, X., Guo, H., Meng, X., & Zhang, B. (2021). An improved calibration and uncertainty analysis approach using a multicriteria sequential algorithm for hydrological modeling. *Scientific Reports*, *11*(1), 16954. https://doi.org/10.1038/s41598-021-96250-6

Zou, C. B., Caterina, G. L., Will, R. E., Stebler, E., & Turton, D. (2015). Canopy interception for a tallgrass prairie under Juniper encroachment. *PLoS One*, *10*(11), e0141422. https://doi.org/10.1371/journal.pone.0141422

SUPPORTING INFORMATION

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

> **How to cite this article:** Badjana, H. M., Cloke, H. L., Verhoef, A., Julich, S., Camargos, C., Collins, S., Macdonald, D. M. J., McGuire, P. C., & Clark, J. (2023). Can hydrological models assess the impact of natural flood management in groundwater-dominated catchments? *Journal of Flood Risk Management*, e12912. https://doi.org/10.1111/jfr3.12912