A comparison of rainfall-runoff modelling approaches for estimating impacts of rural land management on flood flows

Nataliya Bulygina, Neil McIntyre and Howard Wheater

ABSTRACT

There is a requirement for predictive tools to assist in land management and flood risk planning, and a variety of tools have been proposed recently. We compare four tools developed under various UK research programmes. The strengths and limitations of the tools are reviewed, model performances on historic data are assessed, and the methods are applied to estimating flood flows of 5- and 10-year return periods, and flow peaks under both recent land management conditions and speculative scenarios (grazing intensification and tree planting), using the Pontbren catchment, UK as a case study. Overall, the models agree on the direction of change, so that heavy grazing increases, and afforestation and tree strips decrease the flood flows. However, the estimated effects vary significantly between methods. It is concluded that method selection needs to carefully consider the type and scale of land management scenario being examined, and the sources of data available to support the modelling. Using an ensemble of suitable models is proposed as a useful way to represent a multi-expert opinion and to characterise the structural error associated with a single model.

Key words | flood flow, land use, Pontbren, rainfall-runoff modelling, regionalisation

Nataliya Bulygina (corresponding author)
Neil McIntyre
Howard Wheater
Department of Civil and Environmental
    Engineering,
Imperial College,
Skempton Building,
London, SW7 2AZ,
UK
E-mail n.bulygina@imperial.ac.uk

Howard Wheater
University of Saskatchewan,
National Hydrology Research Centre,
11 Innovation Boulevard,
Saskatoon, S7N 3H5,
Canada

INTRODUCTION

Management of land and management of water are strongly inter-dependent (e.g. DEFRA 2005), and integrated catchment management requires capacity for exploring hydrological impacts of land use and land management scenarios. For example, the interactions between land use and flooding, and land use and drought are of considerable practical interest (O'Connell et al. 2004). Interpretation of hydrological variability in terms of rural land management impacts has been difficult: comparative catchment studies are affected by differences in catchment and climate characteristics (e.g. geology, soil, topography, rainfall) (Kirby et al. 1991; Calder 1993; Robinson & Dupeyrat 2005) as well as data uncertainty (Beven et al. 2008). Where catchment inter-comparison reveals land management signals, this is not generally enough to predict responses under future land management scenarios; for

that, a predictive hydrological model is needed (McIntyre & Marshall 2010).

In hydrological modelling, representation of land management effects (for simplicity in this paper, 'land management' is used to include both changes to land use and the way a particular land use is managed) still remains a fundamental challenge. This is because there are few or no data on how the relevant changes affect either local-scale physical properties (for example, soil-plant hydrology) or catchment-scale response (O'Connell et al. 2004; Wheater et al. 2008; Parrott et al. 2009), and because modelling methods face fundamental challenges of methodology and data-support. While, in principle, physics-based rainfall-runoff models are able to represent small-scale processes and upscale them to arbitrary scales, in practice such models have limited capability for reliably estimating

catchment-scale land management effects due to the lack of knowledge about relevant local-scale hydrological properties and how they may change in response to land management (Ballard *et al.* 2010; Ballard 2011), as well as due to the difficulty of up-scaling (O'Connell *et al.* 2007). Physics-based models are also expensive to build and apply at catchment scales, often with no discernible return in terms of accuracy of predictions (McIntyre & Al Qurashi 2009). The application of simpler 'conceptual' rainfall-runoff models to predicting land management impacts generally relies on speculative changes to catchment-scale parameters (e.g. Packman *et al.* 2004; Rose & Rosolova 2007), or on regionalised parameter estimates which are highly uncertain (Bulygina *et al.* 2009, 2011). Various regionalisation exercises have identified relationships between land cover and conceptual model parameters (Arnold *et al.* 1996; Yadav *et al.* 2007; Merz & Bloschl 2009), including in some cases the estimation of uncertainty in the regression (e.g. McIntyre *et al.* 2004). By their nature, however, such regional models only include information on those land management changes that are most identifiable at catchment scales, for example urbanisation, forest cover and presence of surface reservoirs. A range of smaller-scale yet relevant land management options (e.g. Jackson *et al.* 2008; Ballard *et al.* 2010; Wilkinson *et al.* 2010) cannot be examined in this way.

Moreover, irrespective of the type of model, or the method used to estimate model parameters, there is uncertainty associated with assumptions in the model structure. Due to model structural uncertainty and errors in both physics-based and conceptual models, an individual deterministic prediction of land management effects may not be considered reliable, therefore ideally multiple perceptions about a hydrological system would be inspected, in the form of different perceptual models and/or parameterisations. This may be approached by using an ensemble of significantly different models, reflecting different expert opinions on the best way to conceptualise the system and its changes. The ensemble of predictions may be treated in two ways: (1) a model 'averaged' prediction is derived, potentially including weights to reflect the relative believability of the models, for example, related to their previous performances (Neuman 2003; Ye *et al.* 2004); and/or (2) the prediction ensemble is explicitly reported without weighting or averaging (IPCC 2007; Huisman *et al.* 2009).

The first approach, while being simpler to report and interpret, requires the chosen models and their weights to be adequately representative of the joint probability distribution of models to give an accurate estimate of the average. When such conditions are not met the 'averaged' prediction might perform less skillfully than the most skillful model in the ensemble (Winter & Nychka 2010). The second approach, recognising that any statistics such as the average may be poor estimates, illustrates the full sample of results, and allows propagation of model structural uncertainty into the forecasts representing all available (perhaps, diverse) interpretations of system behaviour.

In this paper we estimate land management effects on flood flows using four alternative methods developed under various British national flooding risk estimation programmes: the Flood Risk Management Research Consortium (FRMRC: Pender *et al.* 2005; Wheater *et al.* 2008), the Flood Risk from Extreme Events national research programme (FREE: Bulygina *et al.* 2011), the Flood Management Research and Development programme (FMRD: Calver *et al.* 2005), and the Catchment Flood Management Plan programme (CFMP: Hess *et al.* 2010). Each method uses a different model and a different approach for estimation of parameter values, as described in the next section. The methods developed under the FRMRC and FREE programmes (hereafter called 'Method 1' and 'Method 2' respectively) are probabilistic to account for parameter uncertainty, while the methods developed under FMRD and CFMP (hereafter called 'Method 3' and 'Method 4' respectively) are deterministic. None of the methods requires streamflow measurements for the parameter estimation (although in Method 1 any available runoff measurements may be integrated), so that they may be applied to ungauged catchments and land management scenario analysis. The objective of this paper is to critically review these different approaches to land management impacts analysis, including assessing their underlying principles, their performance on test gauged catchments, and reviewing the range of predictions they produce when used for scenario analysis. Subcatchments of the 12.5 km² Pontbren catchment in Wales, UK, are used as the test catchments. The paper is organised as follows: the next section describes and lists assumptions, limitations and strengths of the four chosen modelling methods; the following section introduces the

Pontbren catchment and land management scenarios; the penultimate section provides results of the study and discusses the findings; and the final section concludes the study.

REVIEW OF FOUR MODELLING APPROACHES TO LAND MANAGEMENT EFFECTS ESTIMATION

In this section, the assumptions, limitations and strengths of the above-mentioned four methods of land management effects estimation are described and discussed. The general issues with the methods are covered here – additional assumptions which are specific to the case study are covered later in the paper.

Method 1: a meta-modelling approach

The procedure is based on using a detailed, physics-based model to formalise local hydrological knowledge about processes and hydrologically-relevant catchment properties (Wheater *et al.* 2008; Ballard 2011). It is based on the hydrological model up-scaling ideas of Ewen (1997), and on the emulation of complex models by reduced order models (e.g. Young & Ratto 2009). A catchment is divided into hydrological response units that form a basis for a semi-distributed model construction. Hydrological process understanding is developed for each unit type and implemented as a high resolution physics-based model. The soil component is based on Richards' equation and can represent fine scale soil vertical structure (at 1 cm resolution), and can include effects such as soil compaction, feed-backs from tree-planting on soil structure, and field drainage. The models may be developed entirely from prior knowledge, or may also be conditioned on available local and catchment scale observations such as soil physical properties, soil moisture, field scale overland, drain flow, interception and groundwater measurements, and streamflow. This results in a library of physics-based models, representing different hydrological and land management conditions. The outputs from these models are used to identify conceptual models which, while being simple and computationally tractable relative to the original model, nevertheless reproduce the key flood responses. The unit-scale conceptual models are assembled into a catchment semi-distributed model via a

channel routing function (Orellana *et al.* 2008; Wheater *et al.* 2008). Parameter uncertainty is considered using Monte Carlo methods. Method 1 is based on the following assumptions and has the following limitations and strengths.

Main assumptions of Method 1

- The physics-based model has no significant structural error so that all uncertainty can be represented as parameter uncertainty.
- Response units can be treated as hydrologically disconnected units, shedding water directly to the channel network.
- The chosen conceptual model is capable of adequately capturing the physics-based model response (an ensemble of conceptual models could be used, but this has not yet been applied).

Main limitations of Method 1

- The computational expense limits the size and number of the response units considered, and the number of Monte Carlo samples used to represent uncertainty.
- If the physics-based model uncertainty is to be fully constrained, it is expected that a significant amount of data will be needed (for example, soil moisture, drain and overland flows, physical characteristics and their change). The capability of this approach when relying entirely on prior knowledge is a topic of ongoing research.

Main strengths of Method 1

- Prediction uncertainty is estimated.
- A physics-based model allows physical changes to the soil and vegetation properties to be translated directly to changes in model parameters, rather than relying on speculative changes or regionalised values.
- The method can be applied to any land management change. This includes the ability to represent the effect of spatial positioning of small-scale land management features (e.g. tree strips or localised surface water storage features).
- The meta-model is conditioned on local-scale information (via being fitted to the physics-based model

outputs), and potentially also on response unit and catchment-scale information (via conditioning of the meta-model on any available flow observations).

Method 2: a Bayesian approach using regionalised indices

This procedure conditions a hydrological model using regionalised values of flow indices to summarise expected hydrological system behaviour depending on catchment geology, soils, and land management. For example, Bulygina *et al.* (2011) used the Base Flow Index (BFI) estimated by the UK HOST (Hydrology of Soil Types) soil classification system (Boorman *et al.* 1995) and Curve Number (CN) estimated using the USDA's Soil Conservation Service soil and land management classification system (USDA 1986). The indices are used to condition parameters of a conceptual hydrological model using Bayes' equations, giving posterior parameter distributions for a given land management scenario. This is described fully by Bulygina *et al.* (2010, 2011), and may be summarised as follows. The posterior likelihood of a sampled parameter set is proportional to the consistency of the simulated hydrological response (as measured using the selected indices) with the response indices predicted for the same catchment by regionalisation. The consistency is measured on a scale defined by the probability distribution of the index – for example, an index with large variance would produce only a small difference in posterior likelihood between parameter sets and thus higher parameter uncertainty. If more than one index is used, then their joint probability distribution is used to define the parameter set likelihood. A large sample of parameter sets and associated likelihoods are used to define the posterior distribution. The posterior distribution of parameter values is propagated to uncertainty in the predictions. The approach is based on the following assumptions, and has the following limitations and strengths.

Main assumptions of Method 2

- Change in hydrological response due to the relevant land management changes can be captured by the chosen regionalised indices. In particular, it may be questioned whether the catchments and data sets used to produce the regional relationships are consistent with the catchment and scenarios being assessed. For example, Bulygina *et al.* (2011) address the question of how relevant the CN system is to UK soil types.
- The response can be captured by the chosen rainfall-runoff model structure.

Main limitations of this Bayesian approach

- Small-scale land management changes are difficult to evaluate, because regionalised indices are typically derived from catchment-scale data.
- Regionalised data on land management effects are not available for the UK; validity of (transformed) relationships from the USA is an implicit assumption.
- Estimates of the joint probability distribution of the regionalised indices are required (which in the case of a multivariate normal distribution would be described by the expected values and a covariance matrix); because this information is seldom given in available databases, this may require some judgement on behalf of the modeller.

Main strengths of this Bayesian approach

- Relative to a physics-based approach, the parameter estimation is straightforward and computationally efficient, because of the explicitly defined likelihood function and relative simplicity of the model.
- Prediction uncertainty is estimated.
- The method has the capability to represent a large variety of land management scenarios, for example different agricultural land uses and management practices, different types of forest and urbanisation.

Method 3: the CFMP land management tool

Method 3 estimates change in catchment scale daily maximum runoff depth for different return periods depending on climate, soil type, and land management type (Hess *et al.* 2010). The method was primarily devised as a screening tool to select policy units that were most sensitive to land

management change. In development of the tool, a continuous-time, daily water balance model, WaSim (Hess & Counsell 2000), was used to estimate antecedent conditions for the period 1961–2006 for different 'policy units' (a combination of agroclimatic zone, soil type, land cover and land management type) with parameters defined by relevant soil physical properties and land cover/management types (Hess & Counsell 2000). WaSim divides the unsaturated zone into three compartments, the upper 0.15 m layer, the active root zone and layer below the root zone. Soil water moves from one layer to the layer below only when its water content exceeds field capacity. The rate of drainage is a function of the relative saturation of the layer and the hydraulic properties of the soil. Water draining out of the lower layer is taken to be potential recharge. The infiltrated water is assumed to be irrelevant for flood flow. Surface runoff is comprised of two components; runoff due to intense rainfall (infiltration excess) and runoff due to saturated soil. Infiltration excess is estimated using the CN method, and, additionally, any rain falling on saturated soil is assumed to run off. Any precipitation that does not run off is assumed to infiltrate. WaSim is spatially lumped in that it does not explicitly consider spatial variability of inputs or outputs, however it accounts for soil and land use heterogeneity by running lumped models in parallel and simple weighted averaging of the output runoff. WaSim has been used to simulate 46 years of daily runoff depth for each of 46,600 combinations of the following: soil types (28 in total, based on the HOST classes), land cover types (5), field condition classes (5) and agroclimatic zones (68) (Smith 1976). The combinations that are not considered plausible are omitted. Then the (log-normal) frequency distribution of annual maxima daily flows is derived for each combination. Hence the change in runoff associated with moving from one policy unit to another can be estimated. The CFMP tool reports this change for the 5-, 10-, 50-, 75- and 100-year return periods, but does not report the long-term hydrographs. There is an option to enter an event hydrograph of a known return period for a baseline land management, recommended to be based on the Revitalised Flood Estimation Handbook methodology (Kjeldsen 2007), so that the CFMP tool can estimate the change in peak flow associated with a land management scenario by scaling the peak according to the change in daily runoff depth. The approach is based on

the following assumptions and has the following limitations and strengths.

Main assumptions of Method 3

- The combination of climate, land management and soil type under investigation can be adequately captured by the 68 agroclimatic areas of Smith (1976).
- The physically-based soil hydrology parameters in the WaSim model can be adequately characterised by the most extensive soil series within the HOST soil class.
- Land management types are uniformly distributed across compatible HOST soil types within a policy unit (this assumption is required to facilitate the spatially lumped model). WaSim can be parameterised using physical properties representative for each HOST type soils thereby not requiring parameter calibration.
- As for Method 2 above, the CN system, originating from data from the mid-west USA, can adequately represent conditions in England and Wales (see Bulygina *et al.* 2011).
- Catchment hydrological response can be captured by the lumped WaSim model, and the impacts of the spatial distribution for responses within the catchment, are negligible.

Main limitations of the CFMP tool

- Continuous hydrological predictions are not available, thereby not allowing land management effects estimation on different hydrograph aspects (i.e. time to peak, peak flow rate).
- The results given are for T-year daily maxima, rather than sub-daily maxima.
- Uncertainty in predictions is not quantified.

Main strengths of the CFMP tool

- Computational efficiency and ease of use (implemented as a MS Excel spreadsheet).
- A reasonably large number of scenarios (improved grassland, cereals, horticulture/non-cereal, semi-natural vegetation, and woodland; covering a range of return periods) can be explored for any catchment in England and Wales.

Method 4: the FRMD approach using regionalised parameters

The regionalisation method developed by Calver *et al.* (2005) estimates parameters of the continuous, spatially lumped, hourly Probability Distributed Model (PDM) (Moore 2007). Rainfall inputs to the soil store are first multiplied by a rainfall correction factor (a model parameter). The soil store can be depleted through evapotranspiration at a rate proportional to the moisture content of the store. Although the general structure of the model is a distribution of soil storage capacities, the shape parameter of the distribution is fixed to zero so that it is a simple bucket model, which generates effective rainfall only when the bucket overflows. Generated runoff is split between two routing stores: (linear) fast flow store and (cubic) slow flow store. Using multiple regression, the model parameters are related to catchment properties such as catchment altitude, slope, drainage density and length, soil type distribution, presence of lakes and reservoirs, and proportion of urban areas and grassland (Calver *et al.* 2005). An exception is the percentage of effective rainfall that goes to the fast store, which is fixed to the standard percentage runoff (SPR) derived from the HOST classification (Boorman *et al.* 1995). In principle, the parameter covariance might be used to estimate prediction uncertainty associated with the regression (e.g. McIntyre *et al.* 2004), however this information was not available, and so we apply the model deterministically using only the best-estimate parameters. To allow the impacts of intensification of land use to be analysed, Packman *et al.* (2004) used expert judgement to modify the BFI and SPR of soils to represent degradation of the soil conditions. The corresponding SPR and BFI values from the HOST database are used to estimate the PDM parameters. The approach is based on the following assumptions and has the following limitations and strengths.

Main assumptions of Method 4

- The 39 catchments used to develop the regression equations provide sufficient information about the variability of responses.

- The lack of consideration of parameter inter-dependence in the multiple linear regressions does not prevent useful parameter sets being estimated.
- Hourly catchment response can be captured, and land management effects estimated using the four-parameter lumped model.

Main weaknesses of the regression-based regionalisation

- Uncertainty in predictions is not quantified (parameter error covariance is not estimated).
- Only two types of land management are directly represented by inputs to the regression equations: grassland, and urban area. To account for an additional land management type – heavy grazing – speculative changes to BFI and SPR are made.

Regression-based regionalisation strengths

- Computationally inexpensive.
- The simple model and published regression equations allow the method to be implemented easily without necessarily having access to the original tool.

APPLICATION OF THE MODELLING APPROACHES: THE PONTBREN CASE STUDY

The four methods are applied to the Pontbren catchment to test performance in replicating observed flows, and to evaluate effects of different land management change scenarios, in terms of changes in flood frequency distributions, and peak flows.

The Pontbren catchment description

The models are applied to five gauged subcatchments of the Pontbren catchment in Powys, Wales, UK (Marshall *et al.* 2009), the largest of which has an area of $5.8 \, \text{km}^2$. Elevations in the catchment range from 170 to 438 m, and slopes are typically steep, on average $6°$. The land at Pontbren is almost exclusively grazed grassland, which

occupies approximately 88% of the land. Woodland occupies 7% of the land area, and the remaining 5% is crops, roofs, paved areas, private gardens and open water. The annual precipitation measured at Pontbren from 1 April 2007 to 31 March 2008 is between 1,200 and 1,450 mm depending on the gauge location, and potential evapotranspiration, estimated by the MORECS (Met Office Rainfall and Evaporation Calculation System) model (Hough & Jones 1997) for grassland is 450 mm, so that the ratio of rainfall to potential evapotranspiration is between 2.7 and 3.2. Soil types are dominated by low permeability silty clay loams (Table 1). Typically 20 cm depth of top-soil overlies a deep layer of relatively impermeable subsoil. On the upper part of the catchment, the topsoil has significant peat content.

The Pontbren catchment was maintained as an experimental catchment over the period 2004–2009, to investigate the effects of land management on flood runoff. It is used for this inter-comparison because high resolution rainfall-runoff data are available, and because the hydrological response and its links with land management in the catchment are relatively well-understood (Jackson et al. 2008; McIntyre & Marshall 2008, 2010; Wheater et al. 2008; Bulygina et al. 2009; Marshall et al. 2009; Ballard 2011; McIntyre et al. 2011) and hence it provides a good basis for assessing applicability of different models. Also, Method 1 requires the physics-based models already set up for this catchment

(Jackson et al. 2008). Finally, the land management at Pontbren is representative of upland land management over most of Wales and over much of the UK (Marshall et al. 2009).

Data used for this study are 10-minute resolution rainfall from the rain gauge at site 3 (Figure 1), which has the longest record, daily MORECS evapotranspiration rates, and 15-minute resolution streamflow data from five bed-mounted acoustic Doppler velocity meters (gauges 2, 5, 6, 7, and 9 in Figure 1) (McIntyre & Marshall 2008). The contributing areas at each of the five gauges are given in Table 2 along with other catchment properties derived from the LCM2000 land use maps and NSRI (National Soil Resources Institute) soil data base. Although another five flow gauges exist (Figure 1), their data are considered less accurate (McIntyre & Marshall 2008) and so are neglected here.

Twenty years (January 1989–December 2008) of climate data from Cefn Coch (approximately 3 km from Pontbren) are used to drive the models and estimate flood frequency distributions, which are used to assess the differences in flood frequency estimates between the four methods. The data from Cefn Coch are used because long-term climate data from within the Pontbren catchment are not available. The data from Cefn Coch consist of hourly rainfall from a tipping bucket gauge and daily evapotranspiration data calculated from climate data using the MORECS model (Hough & Jones 1997).

Table 1 | Dominant soil series and types of the Pontbren catchment

| Soil series	Broad texture group	Soil water regime	Soil parent material	HOST class	Area fraction
Hiraethog	Thin peat over loamy over lithoskeletal	Seasonally waterlogged, topsoils wet for most of autumn, winter and spring	Mudstone and sandstone	15	0.12
Manod, Denbigh	Fine loamy over lithoskeletal	Well drained, moderately permeable, subsoils rarely wet	Mudstones and sandstones	17	0.13
Sannan	Fine silty	Slight seasonal waterlogging, subsoils slowly permeable	Glacial till with siliceous stones	18	0.04
Cegin	Fine silty	Slowly permeable, seasonally wet	Glacial till with siliceous stones	24	0.34
Wilcocks	Peaty surface layer over loamy	Seasonally waterlogged, topsoils wet for most of autumn, winter and spring, subsoils wet for most of year	Glacial till with siliceous stones	26	0.28
Crowdy, Winter Hill	Deep peat	Permanently waterlogged	Humified peat	29	0.07
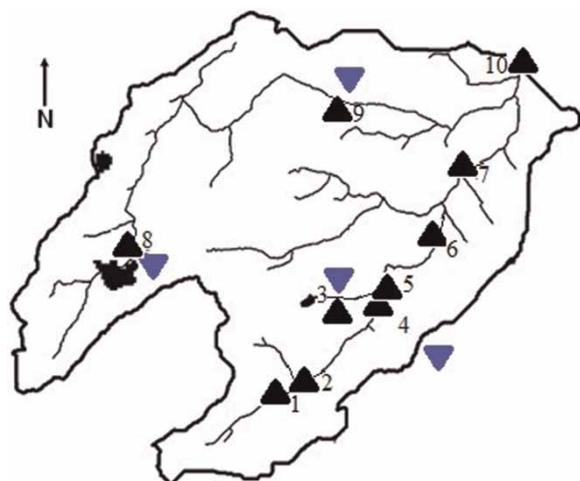
Figure 1 │ Pontbren instrumentation locations: ▲n Stream and drain flow monitoring site; ▼ Rain gauge, black areas next to gauge 8, north of gauge 8 and next to gauge 3 are lakes.

Performance assessment under existing land management

Assessing the suitability of the methods (summarised in Table 3) for predicting impacts of changed land management begins with assessment under unchanged conditions, and the ability of the models to replicate the range of observed flow responses observed at the five Pontbren gauges from the 45-month period April 2005–December 2008. This test is restricted to the meta-modelling, regionalisation, and regression-based parameterisation approaches: the CFMP method is not tested in this way because it evaluates changes in daily flows for different return periods, and it does not report time series of flow.

Four measures of performance are considered:

1. Nash–Sutcliffe efficiency (NS).
2. Nash–Sutcliffe efficiency for log-transformed flows (NSlog).

3. Relative percentage bias in total runoff volume (Bias).
4. Prediction bound width for probabilistic models (Pwidth).

The Bias measure concerns the total runoff amount, and characterises the partitioning between total evapotranspiration, runoff and storage. The NS applied to untransformed flows quantifies the match between the observed and predicted flow time variability at the appropriate model resolution time scale. The NS puts higher emphasis on fitting high flows due to the use of squared residuals in the calculation, whereas applied to log-transformed flows the NS tends to match low flow performance due to the log functional properties. The prediction bound width describes the precision of probabilistic model predictions, complementing the NS efficiencies that characterise prediction accuracy. Since the first two methods are probabilistic, using 30 samples to account for model parameter uncertainty, the NS, NSlog, and Bias measures are also probabilistic, having values determined for each individual model parameter sample. For the probabilistic methods, therefore, expected values and standard deviations of the probabilistic NS, NSlog and Bias (as well as Pwidth) measures are reported for comparison with the deterministic values derived using the deterministic methods.

Land management scenarios

Four different scenarios were considered: (1) current land management, (2) heavy grazing over the whole catchment area, (3) complete afforestation with deciduous trees, and (4) deciduous tree strips on fields under improved grassland. The tree strips are assumed to occupy 12% of field area, which is typical within Pontbren (Jackson et al. 2008). Method 1 is capable of explicit positioning of the strip at

Table 2 │ Properties of the five gauged subcatchments

Gauge no.	Catchment area, km²	Improved grassland, %	Unimproved grassland, %	Open water, %	Woodland, %	Average slope,⁰	BFIHOST
2	1.29	68.0	28.8	0.01	3.2	5.0	0.26
5	2.39	70.4	21.6	0.34	7.7	4.7	0.27
6	3.17	77.0	16.6	0.26	6.1	5.0	0.28
7	5.77	79.5	15.8	0.14	4.5	5.3	0.30
9	4.06	13.2	83.1	2.6	1.1	7.0	0.34

Table 3 | Summary of models used to represent the Pontbren hydrological response

Method	Soil moisture model summary	Routing model summary	No. of parameters to estimate	Spatial unit scale	Time step	Reference
Method 1: Meta-modelling	Physics-based model (Richards' equation) at field scale, upscaled to a conceptual model	Two parallel routing stores, plus one-parameter channel model	6 × number of land management types	Field size	15 min	Wheater *et al.* (2008)
Method 2: Indices-based regionalisation	PDM model	Two parallel routing stores, plus one-parameter channel model	5 × number of soil (HOST) – land management combinations	100 × 100 m	15 min	Bulygina *et al.* (2011)
Method 3: CFMP tool	WaSim model (1D conceptual model parameterised using physical properties)	No routing	20 × number of soil (HOST)– land management combinations	Lumped with spatial weighting	1 day	Hess *et al.* (2010); Hess & Counsell (2000)
Method 4: Regression-based regionalisation	Bucket model (PDM with shape parameter = 0)	Two parallel routing stores	4	Lumped	1 hour	Calver *et al.* (2005)

the bottom of each field; while Methods 2 and 3, due to their limitations, treat the strips as an afforested area fraction. Method 4 is designed to represent three types of land management only: grassland, urban, and heavy grazing, and therefore it was not used for scenarios 3 and 4. Results are presented as relative changes from those under the existing land management.

Following performance assessment, the effects on T-year floods are assessed. Using the 20-year series of rainfall from Cefn Coch, a time-series of runoff was generated by Methods 1, 2 and 4. For the four land management scenarios listed above, a log-normal distribution was fitted to each simulated annual daily maxima series. In each case, the log-normal hypothesis was accepted on the 95% confidence level using the Kolmogorov–Smirnov test. The 5- and 10-year return period flood peaks were estimated using the fitted distribution (longer return periods were not assessed because only 20 years of data were available). Method 3 reports relative changes in daily T-year floods (Hess *et al.* 2010); hence results from all four methods could be compared.

To further evaluate land management impacts on peak flows, the largest simulated flow event (20 December 1991) was modelled under the four scenarios, using all methods but Method 3 (as it does not report time-series results). The effects were estimated at all five gauges, but

results are presented below only for gauges 6 and 9 because these two gauges have the most significantly different soil/ land management properties (see Table 2). Method 1 is demonstrated for gauge 6 only, because the physics-based models which underlie this approach were not developed for the soil types contributing to gauge 9 (Wheater *et al.* 2008).

RESULTS AND DISCUSSION

Performance assessment on observed flows

Performances of the three time-series modelling approaches (Methods 1, 2 and 4) were assessed using the flow records from April 2005 to December 2008. The performance metrics are shown in Table 4, where the best performance measures are highlighted in grey. Methods 1 and 2 give the highest NS values (when mean values are compared). Method 2 gives the highest NSlog values (except for gauge 7), while Method 1 gives the lowest NSlog values. Also, Method 2 has the lowest absolute Bias value (except gauge 7), while Method 4 has the highest Bias (i.e. over-prediction of runoff volume). Model predictions for one of the largest observed runoff events observed at Pontbren (18 January

Table 4 | Nash–Sutcliffe efficiencies (NS), biases, and average 95% prediction widths for considered approaches[a]

Gauge	2	5	6	7	9
Method 1: Meta-modelling approach					
NS	0.46 (0.08)	0.77 (0.01)[c]	0.84 (0.01)[c]	0.81 (0.01)[c]	–[b]
NSlog	−6.50 (4.3)	−28.40 (22.7)	−28.10 (20.1)	−2.80 (3.1)	–
Bias, %	39.8 (0.9)[d]	4.6 (0.7)[d]	−17.7 (0.6)[d]	−25.4 (0.6)[c]	–
Pwidth, m^3/s	0.01 (0.03)[c]	0.02 (0.04)[c]	0.02 (0.04)[c]	0.03 (0.05)[c]	–
Method 2: Indices-based regionalisation approach					
NS	0.80 (0.03)[c]	0.76 (0.03)[d]	0.79 (0.03)[d]	0.77 (0.03)[d]	0.84 (0.01)[c]
NSlog	0.60 (0.03)[c]	0.70 (0.02)[c]	0.72 (0.02)[c]	0.16 (0.05)[d]	0.70 (0.02)[c]
Bias, %	15.2 (1.3)[c]	1 (1.4)[c]	−8.7 (1.7)[c]	27.3 (2.6)[d]	−13.2 (1)[c]
Pwidth, m^3/s	0.02 (0.03)	0.03 (0.04)	0.05 (0.07)	0.08 (0.12)	0.03 (0.04)
Method 4: Regression-based approach					
NS	0.63	0.71	0.78	0.32	0.64
NSlog	0.52[d]	0.59[d]	0.68[d]	0.35[c]	0.64[d]
Bias, %	49.2	35.5	26.7	76.6	16.8

[a]Expected values and standard deviations (in parentheses) of NS, NSlog, and Bias are given for the probabilistic approaches.
[b]The meta-modelling was not developed for the area contributing to gauge 9.
[c]Best measures.
[d]Second-best measures.

2007) are shown in Figure 2. Methods 1 and 4 tend to over-predict the peak flow. Method 2 provides much wider prediction bounds than the meta-modelling approach (see also average prediction bound width for Methods 1 and 2 in Table 4). Method 3 is not assessed here, since it only provides relative changes for T-year floods. Interestingly, the performance of Methods 2 and 4 drops significantly for gauge 7 (according to NSlog and Bias measures), even though the area draining into gauge 7 is similar to the area draining to gauge 6 (Table 2). This might be because the bed-mounted acoustic Doppler velocity meter used as gauge 7 was less reliable, especially for low flows, than that used as gauge 6 (McIntyre & Marshall 2008). Based on Table 2, Method 1 provides the highest NS values (closely followed by Method 2), and Method 2 provides better performance measures than the other two methods for low flows and water balance.

Based on the NS and Bias measures, Methods 1 and 2, the meta-modelling and indices-based methods, generally performed better than Method 4, the regression-based method. However, the latter was developed for larger catchments (area greater than 10 km^2), and with a broader set of

applications in mind, including urban and lowland catchments, whereas the former two were developed specifically to look at upland rural land management. The performance differences may also arise from some of the theoretical limitations of using multiple regression to estimate model parameters for ungauged catchments: the regression approach of Method 4 estimates the rainfall-runoff model parameters independently of each other (see McIntyre *et al.* 2005), whereas the other two methods maintain the dependencies by using sampled parameter sets. Method 1 usually produces the best NS values, which is likely to be because its underlying physics-based models were trained primarily to fit high flows (Wheater *et al.* 2008). Method 2 gives close to the best performance values for all gauges, but gives quite large prediction bound widths, Pwidth, relative to Method 1 (Table 4), therefore its applicability might be questionable on the basis of low precision. Also, the applicability of Method 2 is limited by its reliance on prior estimates of the flow indices, which would typically come from national databases (in the case study, BFI and CN); whereas Method 1 allows locally specific land management features and practices, such as locations of tree strips and
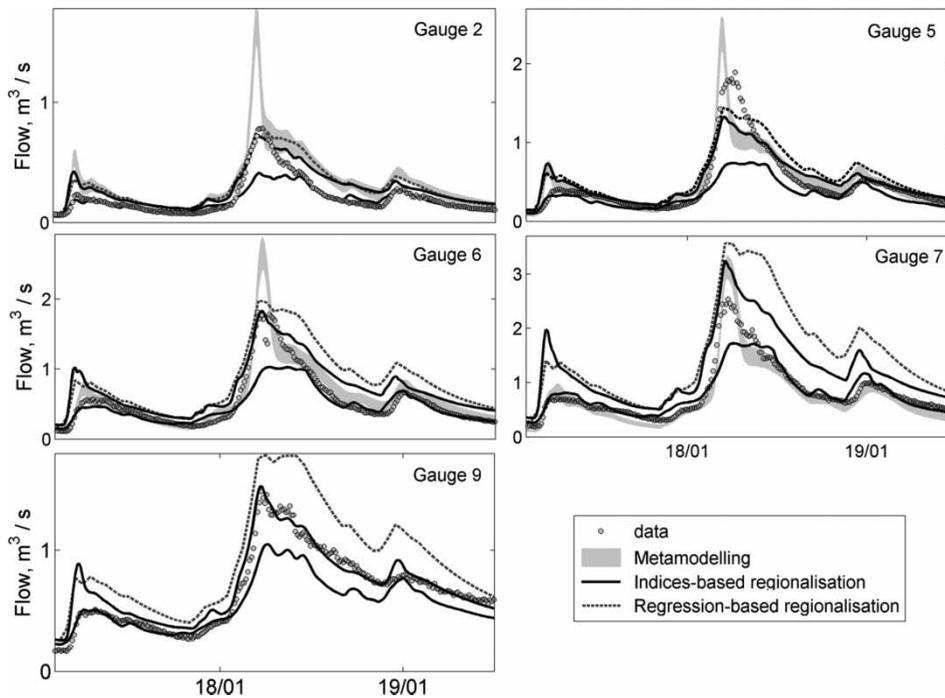
Figure 2 | Prediction comparison given by Methods 1, 2 and 4 for an event on 18 January 2007. Prediction is given either as a line (regression-based regionalisation), or as minimum-maximum prediction bounds (metamodelling and indices-based regionalisation).

surface storage features, to be explicitly included, albeit at greater computational expense and with the requirement for knowledge of local scale processes (Ballard 2011; Bulygina *et al.* 2011). Method 2, as applied here, also relies on the CN data, derived in the mid-western USA, with doubts about its applicability to UK soils and land management. Although it was also found to perform well on the Plynlimon catchments by Bulygina *et al.* (2011), clearly, more extensive validation tests are required before recommending widespread use of this index. This latter limitation – the assumption about the applicability of the CN data – also applies to Method 3, the time-series performance of which could not be tested as it reports only the T-year daily flood peaks.

Variability of predictions between the methods

To evaluate the variability of the methods in terms of their predictions of land management effects on flood peaks, changes to the 5- and 10-year return period floods were modelled, as well as changes to the largest simulated peak flow. For the probabilistic methods, the median out of all the samples is used for summarising results. The results generally show that heavy grazing increases flow magnitudes, while afforestation and tree strips decrease flow magnitudes (Table 5). For example, at gauge 6:

- Method 3 predicts the highest increase in peak daily flow under heavy grazing (28–34%, depending on return period), and Method 4 gives the lowest increase (3.9–4.1%);
- Method 1 predicts the highest decrease in maximum daily flow under afforestation (17.4–17.9%), and under tree strips (6.2–6.5%); and
- Method 2 provides estimates between those of Methods 1 and 3.

The estimated effects vary significantly between the methods. For example, if the area becomes heavily grazed, the 5-year return period maximum daily flow at gauge 6 is predicted to increase by between 4 and 34% depending on the method used; and is predicted to decrease by between 11 and 18% if the area becomes fully afforested with deciduous trees; and decrease between 1.5 and 6% if tree strips are introduced in the fields under improved grassland.

Table 5 | Median and 95% confidence intervals (shown in parentheses) for changes in maximum daily flows (%) at gauges 6 and 9, and in the largest simulated peak flow (December 1991) at gauge 6 under different land management change scenarios

Return period	Gauge 6		Gauge 9		Largest peak flow (gauge 6)
	5 years	10 years	5 years	10 years	
Peak flow[a]	27.7 mm/day	30.8 mm/day	23.4 mm/day	26 mm/day	0.5 mm/15 min
Heavy grazing					
Method 1: Meta-modelling	11.4 (−1; 21)	12.8 (0; 19)	−[b]		44.2 (35; 49)
Method 2: Indices-based regionalisation	10.3 (2; 21)	11 (2; 22)	13.5 (2; 22)	13.1 (1; 22)	17.8 (4; 36)
Method 3: CFMP tool	33.8	28.4	50.5	42	−[d]
Method 4: Regression-based regionalisation	4.1	3.9	4	3.8	4.2
Afforestation					
Method 1: Meta-modelling	−17.4 (−35; −5)	−17.9 (−34; −5)	−[b]		−48.4 (−64; −33)
Method 2: Indices-based regionalisation	−11.5 (−26; −3)	−11.7 (−27; −4)	−9.3 (−22; −3)	−10.1 (−22; −4)	−15.8 (−33; −2)
Method 3: CFMP tool	−18.1	−11.7	−14.5	−8.7	−[d]
Method 4: Regression-based regionalisation	−[c]				
Tree strips on improved grassland					
Method 1: Meta-modelling	−6.2 (−8; 2.4)	−6.5 (−8.4; −2.8)	−[b]		−7.8 (−17; −5)
Method 2: Indices-based regionalisation	−1.5 (−3; −0.2)	−1.5 (−3; −0.3)	−0.3 (−0.9; 0.1)	−0.4 (−0.9; 0)	−1.1 (−2.4; −0.2)
Method 3: CFMP tool	−2	−1.3	−1.1	−0.8	−[d]
Method 4: Regression-based regionalisation	−[c]				

[a]Estimated using Method 2 (median values).
[b]Method 1 was not developed for area contributing to gauge 9 (Wheater et al. 2008).
[c]Method 4 is insensitive to afforestation and tree strips (Calver et al. 2005).
[d]Method 3 does not provide continuous time predictions (Hess et al. 2010).

Furthermore, change estimates vary significantly across the samples taken within each probabilistic model run. All the results in Table 5 illustrate that the relative effects of change also depend on existing land management and soil distributions. Areas which are heavily grazed (i.e. the existing catchment under gauge 6) yield higher relative effects when afforestation and tree strips are implemented, while areas with relatively low grazing intensity (as under gauge 9) yield lower relative effects.

Methods 1, 2 and 4 predict that the relative effects tend to be similar for the 5- and 10-year return periods, whereas Method 3 predicts a significant decrease in effects for the more extreme flood. This may be because the land management change principally affects the linear routing part of the models in Methods 1, 2 and 4: the split between storm-flow and baseflow, and the residence times, rather than the soil moisture accounting model parameters. For example, this is clear in the sensitivity analysis done by Bulygina et al. (2011). In contrast, in Method 3, the impact of land management change is assumed to affect only the non-linear soil moisture model. In general, it has been difficult to identify relations between non-linear model parameters and land management (e.g. McIntyre & Marshall 2010), and hence the more evidence-based methods, Methods 1, 2 and 4, are expected to be more linear in their effects. For Methods 1 and 2, the small increase in effect going from $T = 5$ to $T = 10$ years is not significant: it can be explained by the probabilistic nature of the two

approaches, and the sampling error in the reported median value.

Land management effects on peak flow were evaluated for the largest simulated event in the 1989–2008 period using all but Method 3. This differs from the previous analysis of T-year floods in that it assesses largest flow at the 15 minute (or, hourly for Method 4) time resolution rather than daily (Figure 3). Scenario prediction bounds (maximum and minimum flows from the 30 samples) overlap each other for Method 2, signifying high uncertainty, which arises from the limited information in the two indices (BFI and CN) used. The bounds do not overlap for Method 1 signifying much lower uncertainty, and this method also has generally better accuracy for peak flows (Table 4). However, the application of Method 1 in this case has used local measurements to constrain the physics-based and catchment models (Wheater *et al.* 2008), and its accuracy and precision would be expected to decrease without the benefit of these local measurements. The Method 1 predictions are the most sensitive to the considered land management changes (Figure 3 and Table 5), suggesting that the other methods may be underestimating effects. As with the return period analysis, the predicted effects vary significantly between the methods: they range from a 4 to 44% increase in peak flow for the heavy grazing scenario, 16 to 48% decrease in peak flow for the afforestation scenario, and 1 to 8% decrease in peak flow under the tree strip scenario. Both

Method 1 and Method 2 estimate relative effects of land management on peak flow to be much larger than the effects on a 10-year return period maximum daily flow; while Method 4 evaluates the effects to be similar (Table 5). This may be linked to the differences in the spatial natures of the models, with the latter (lumped) model losing some of the locally intense runoff associated with the change. This may also explain why the latter method performs less well on high flows in general (Table 4).

The large difference in results between methods potentially presents a challenge for decision-makers, and to decide which method, if any, to trust, it is essential to understand the reasons for their differences. There are various potential reasons:

1. The literature describing the four approaches *does not rigorously define land management type*; for example, 'heavy grazing', 'forest in good condition' are general classifications, not explicitly defined by measurable properties. For the purpose of this comparison, we have interpreted the classifications as having a consistent physical meaning over all methods, whereas in fact it depends on the origins of the methods. For example, the meta-models of Method 1 were derived specifically for Pontbren based on the distribution of soil properties measured for each land management regime (Wheater *et al.* 2008); whereas the classification of land
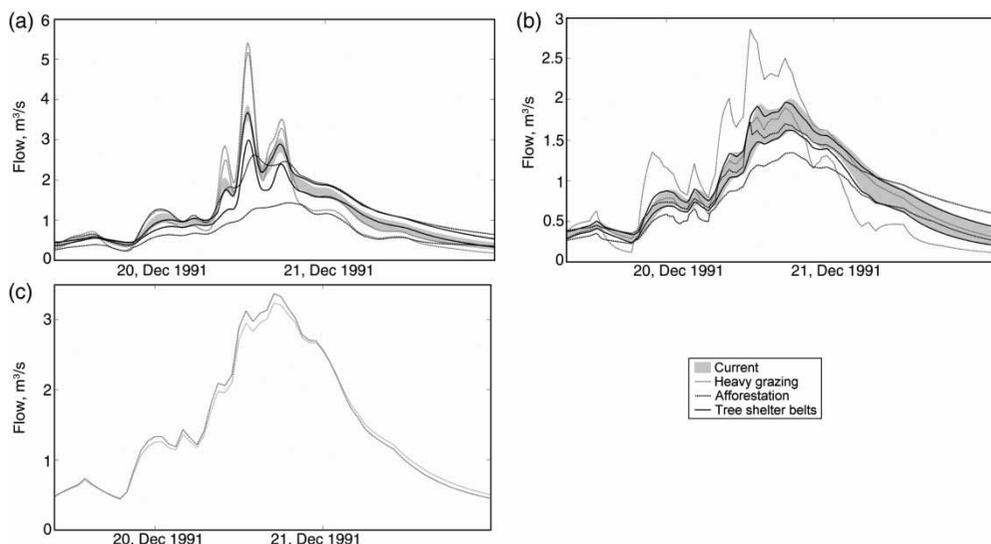

Figure 3 | Streamflow at gauge 6 for the largest simulated event (December 1991) as estimated by (a) Method 1 (minimum and maximum predictions), (b) Method 2, and (c) Method 4.

management used for the indices-based regionalisation of Method 2 and the CFMP tool of Method 3 is subjectively mapped to the CN classification system (Hess *et al.* 2010; Bulygina *et al.* 2011), while the regression-based regionalisation of Method 4 is based on speculative parameter modifications (Packman *et al.* 2004).

2. Closely related to the previous issue, each method draws on *different sources of information* to estimate the rainfall-runoff model parameters: a database imported from overseas (Methods 2 and 3), a very general regional model developed from only 39 catchments in the UK (Method 4); and up-scaling local evidence (Method 1). It is hardly surprising that the results are widely variable.

3. Each method uses *a different approach for identifying rainfall-runoff parameters* from the given information. Method 1 uses a GLUE-type approach to estimate parameter sets and propagate uncertainty through the up-scaling procedure; Method 2 takes a formal Bayesian approach including estimation of the information uncertainty and how it propagates to parameter set estimates and model predictions; Method 3 uses soil physical properties (without accounting for uncertainty) to parameterise its subsurface model, and CNs to estimate surface/subsurface flow volume, while Method 4 uses deterministically-applied regression relationships between catchment-scale properties and parameter values.

4. And finally, each method uses a *different model structure.* The major difference in the case study is the use of a field-scale semi-distributed model for Method 1, a gridded model for Method 2, and a lumped model for Method 3 and 4. The lumped models use the assumption that the distribution of runoff generation and travel times, which in general will tend to smooth out land management signals in peak flows, will not significantly affect results.

The decision about which method, if any, to adopt should aim to align the task at hand with the particular attributes of the methods. For example, Methods 1 and 2 were developed to be applied to spatially distributed problems, for assessing catchment scale impacts of local scale land management. Method 1 in particular has the ability to look at the full range of scales if adequate supporting small-scale information exists, whereas the evidence used in Method 2 originates largely from small catchments and

loses relevance at micro-scales. Method 3 is a nationally applicable tool designed to indicate likely effects on T-year daily floods of widespread (non-local) land management change; while Method 4 uses UK-based catchment-scale evidence although is more restricted in the range of land management scenarios that can be examined. In many cases, an ensemble or weighted average approach may be used to address the variability of results.

If the weighted average approach is followed a modeller has to assign a weight to each of the methods, which would normally be related to the confidence in the method for the task at hand. The confidence in the method may be assessed through its predictive quality on historical data, even though future predictions involve not only new climatic patterns, but also changed catchment conditions (due to land management change). The predictive quality assessment requires choice of an objective function, and its transformation into a relative reliability measure for use as a weight. Since the methods in this study did not use hydrological observations at the calibration stage (with the exception of Method 1), the multi-model weighting sometimes used in hydrology (Neuman 2003; Ye *et al.* 2004), which assumes that the weights are dependent on the number of observations used to calibrate the model and the number of model parameters, does not suit the needs. This calls for methods to specify model probabilities for ungauged catchment applications. Possible ways to address the issue would be to: (1) assign equal weights to the predictions and consider them as equally plausible (an ensemble approach); or (2) disregard methods that do not perform well on historic data, and weight the rest according to their relative performances, and from this derive confidence limits and a weighted average result. Using such a scheme for the Pontbren application, based on Table 4, Methods 1 and 2 would be assigned high relative weights, which is to be expected because these two methods and their associated models have been developed around Pontbren and other upland catchments, whereas Methods 3 and 4 were developed with more general national application in mind. The idea of using a weighted averaging scheme, however, does require more testing of the methods, ideally across multiple catchments that represent the relevant land management questions; more testing is needed before we are in a position to

make conclusions about the general reliability of the four tested methods.

CONCLUSIONS

Predicting the impacts of rural land management on flood flows has become central to integrated catchment management in the UK and elsewhere. Various methods have recently been developed to address this prediction challenge, based on rainfall-runoff simulation models. This paper assesses four of these methods, with the objective of providing guidance on choosing between the methods, and proposing some general priorities for improving tools for land management impacts analysis.

Four approaches to prediction developed under various UK flood risk estimation programmes were considered, employing different rainfall-runoff models and different strategies to estimate model parameter values. The four methods considered are: (1) the meta-modelling approach (Wheater *et al.* 2008; Ballard 2011); (2) the indices-based regionalisation (Bulygina *et al.* 2011); (3) the CFMP tool (Hess *et al.* 2010); and (4) the regression-based regionalisation (Calver *et al.* 2005).

The methods were assessed using the following assessment strategy: (1) review of the origins, assumptions, strengths and weaknesses behind each method; (2) performance assessment using an upland rural catchment which represents a sample of UK land management issues – the Pontbren catchment in Wales, UK; (3) evaluation of the variability of results between methods when applied to land management scenarios.

The performance assessment of the three methods which generate flow time-series (Methods 1, 2 and 4 above) using historical streamflow data from five gauges at Pontbren showed that the performances (for high flows, low flows and volume bias) were variable for all methods, with no method consistently producing performance which we would regard as good for all gauges and performance criteria. The first two methods, the indices-based and meta-modelling methods, generally performed better than the fourth, the regression-based method. This is related to the distributed nature of the former two methods, and the models they employ were selected specifically for this type

of upland application, whereas the latter was designed for more general and larger-scale use in the UK.

Changes in 5-year and 10-year daily maxima due to changes in land management were estimated, using all four methods, to evaluate variability between methods. The flood flows were estimated for current land management conditions and for three future scenarios: increase in stocking density, full afforestation with deciduous trees, and tree strips introduced at the bottom of all grazed fields. All four methods agreed on the direction of change: heavy grazing increases the flows, while afforestation and tree strips decrease the flows. At the same time, the estimated magnitude of change is highly variable across methods, and within the probabilistic methods.

The differences in the predictions made by the alternative methods lead to the questions: (1) what method is best for a particular task; and (2) whether and how the predictions should be combined and assigned relative weights related to their reliability. To answer the first question, careful consideration is needed of the origins and design of the methods in relation to the predictive task at hand, including their evidence base, space and time resolution, suitability of the model structure, the method used for parameter estimation, and performance history. To answer the second question, since model performance can only be evaluated on historical data, the relative reliability of methods for predicting the effects of future land management change cannot easily be resolved. Therefore, it might be recommended either: (1) to keep the whole model prediction range (a prediction ensemble); or (2) to exclude some models with low perceived reliability (as defined by judgement of a user, and/or performance on historic data) from the ensemble, potentially reducing the prediction range. The scope of this paper does not permit conclusions to be made about the general reliability of the four tested methods; but for a Pontbren-type problem it is proposed that meta-modelling approach combined with indices-based regionalisation are more applicable due to their origins, characteristics and performances on relevant historic data.

Taken overall, the results here provide an important and salutary illustration of the challenges in modelling the subtle effects of land management change. In particular, if the uncertainties illustrated above are to be reduced to provide more refined guidance to land management policy, a more

extensive set of detailed field data is required, from multiple catchments across the UK, together with continuing inter-comparison and refinement of these alternative modelling tools.

REFERENCES

Arnold, J. G., Williams, J. R., Srinivasan, R. & King, K. W. 1996 SWAT: Soil and Water Assessment Tool. USDA-ARS, Grassland, Soil and Water Research Laboratory, Temple, TX.

Ballard, C. 2011 The Role of Physics Based Models for Simulating Flood Flow Response to Rural Land Management Scenarios. PhD Thesis, Imperial College London, pp. 324. https://workspace.imperial.ac.uk/ewre/Public/BALLARD_Thesis_FINAL_2011[1].pdf

Ballard, C., McIntyre, N., Wheater, H., Holden, J. & Wallage, Z. 2010 Hydrological modelling of drained blanket peat. J. Hydrol. 407 (1–4), 81–92.

Beven, K., Young, P., Romanowitcz, R., O'Connell, E., Ewen, J., O'Donnell, G., Homan, I., Posthumus, H., Morris, J., Hollis, J., Rose, S., Lamb, R. & Archer, D. 2008 Analysis of Historical Data Sets to Look for Impacts of Land Use and Management Change on Flood Generation. DEFRA, London.

Boorman, D., Hollis, J. & Lilly, A. 1995 Hydrology of Soil Types: A Hydrologically-based Classification of the Soils of the United Kingdom. Institute of Hydrology, Wallingford.

Bulygina, N., McIntyre, N. & Wheater, H. 2009 Conditioning rainfall-runoff model parameters for ungaged catchments and land management impacts analysis. Hydrol. Earth Syst. Sci. 13 (6), 893–904.

Bulygina, N., McIntyre, N. & Wheater, H. 2011 Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes. Water Resour. Res. 47, W02503.

Bulygina, N., McIntyre, N., Wheater, H. & Ballard, C. 2010 Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes. Proceedings of the BHS International Hydrology Symposium, Newcastle University, UK, 19–23 July 2010.

Calder, I. R. 1993 The Balquhidder catchment water-balance and process experiment results in context – what do they reveal? J. Hydrol. 145 (3–4), 467–477.

Calver, A., Crooks, S., Jones, D., Kay, A., Kjeldsen, T. & Reynard, N. 2005 National River Catchment Flood Frequency Method using Continuous Simulation. DEFRA, London.

DEFRA 2005 Making Space for Water. Taking Forward a New Government Strategy for Flood and Coastal Erosion Risk Management in England. First Government Response to the Autumn 2004 Making Space for Water Consultation Exercise. Department of Environment, Food and Rural Affair, London, 44 pp.

Ewen, J. 1997 'Blueprint' for the UP modelling system for large scale hydrology. Hydrol. Earth Syst. Sci. 1, 55–69.

Hough, M. & Jones, R. 1997 The United Kingdom Meteorological Office rainfall and evapotranspiration calculation system: MORECS version 2.0 – an overview. Hydrol. Earth Syst. Sci. 1 (2), 227–239.

Hess, T. M. & Counsell, C. 2000 A Water Balance Simulation Model for Teaching and Learning – WaSim. Paper presented at the ICID British Section Irrigation and drainage research day, HR Wallingford. http://hdl.handle.net/1826/2455.

Hess, T., Holman, I., Rose, S., Rosolova, Z. & Parrott, A. 2010 Estimating the impact of rural land management changes on catchment runoff generation in England and Wales. Hydrol. Process. 24, 1357–1368.

Huisman, J., Breuer, H., Borman, H., Croke, B., Frede, H.-G., Graff, T., Hubrechts, L., Jakemen, A., Kite, G., Lanini, J., Leavesley, G., Lattenmaier, D., Lindstrom, F., Seibert, J., Sivapalan, M., Viney, N. & Willems, P. 2009 Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM) 3: scenario analysis. Adv. Water Resour. 32, 159–170.

IPCC Fourth Assessment Report 2007 Climate Change 2007. Available from: www.ipcc.ch/publications_and_data/ar4/wg2/en/contents.html.

Jackson, B., Chell, J., Francis, O., Frogbrook, Z., Marshall, M., McIntyre, N., Reynolds, B., Solloway, I. & Wheater, H. 2008 The impact of upland land management on flooding: insights from a multi-scale experimental and modelling programme. J. Flood Risk Manage. 1 (2), 71–80.

Kirby, C., Newson, M. D. & Gilman, K. 1991 Plynlimon Research: The First Two Decades. Institute of Hydrology, Wallingford.

Kjeldsen, T. 2007 The Revitalised FSR/FEH Rainfall-runoff Method – A User Handbook. Flood Estimation Handbook Supplementary Report No. 1, Centre for Ecology and Hydrology, Wallingford, UK. Available from: www.ceh.ac.uk/refh.

Marshall, M., Frogbrook, Z., Solloway, I., Reynolds, B., McIntyre, N., Wheater, H., Jackson, B. & Francis, O. 2009 The impact of upland land management on flooding; preliminary results from a multi-scale experimental programme. *Hydrol. Process.* **23** (3), 464–475.

McIntyre, N. & Marshall, M. 2008 Field verification of bed-mounted ADV meters. *Proceedings of the Institution of Civil Engineers-Water Management* **161**, 199–206.

McIntyre, N. & Al Qurashi, A. 2009 Performance of ten rainfall-runoff models applied to an arid catchment in Oman. *Environ. Model. Softw.* **24**, 726–738.

McIntyre, N. & Marshall, M. 2010 Identification of rural land management signals in runoff response. *Hydrol. Process.* **24** (24), 3521–3534.

McIntyre, N. R., Lee, H., Wheater, H. S. & Young, A. R. 2004 Tools and approaches for evaluating uncertainty in streamflow predictions in ungauged UK catchments. In: *Complexity and Integrated Resources Management* (C. Pahl-Wostl, S. Schmidt & T. Jakeman, eds). Proc. of IEMSS International Congress, Osnabrueck, Germany, June 2004.

McIntyre, N., Lee, H., Wheater, H., Young, A. & Wagener, T. 2005 Ensemble predictions of runoff in ungauged catchments. *Water Resour. Res.* **41**, W12434.

McIntyre, N., Young, P. C., Orellana, B., Marshall, M., Reynolds, B. & Wheater, H. S. 2011 Identification of nonlinearity in rainfall-flow response using data-based mechanistic modelling. *Water Resour. Res.* **47**, W03515.

Merz, R. & Bloschl, G. 2009 A regional analysis of event runoff coefficients with respect to climate and catchment characteristics in Austria. *Water Resour. Res.* **45**, W01405.

Moore, R. J. 2007 The PDM rainfall-runoff model. *Hydrol. Earth Syst. Sci.* **11** (1), 483–499.

Neuman, S. 2003 Maximum likelihood Bayesian averaging of uncertain model predictions. *SERRA* **17**, 291–305.

O'Connell, P. E., Beven, K., Carney, J., Clements, R., Ewen, J., Fowler, H., Harris, G., Hollis, J., Morris, J., O'Donnell, G., Packman, J., Parkin, A., Quinn, P., Rose, S., Shepherd, M. & Tellier, S. 2004 *Review of Impacts of Rural Land Use and Management on Flood Generation*. DEFRA, London.

O'Connell, E., Ewen, J., O'Donnell, G. & Quinn, P. 2007 Is there a link between agricultural land-use managment and flooding? *Hydrol. Earth Syst. Sci.* **11** (1), 96–107.

Orellana, B., Pechlivanidis, I. G., McIntyre, N., Wheater, H. & Wagener, T. 2008 A toolbox for identification of parsimonious semi-distributed rainfall-runoff models: application to the Upper Lee catchment. In: *International Congress on Environmental Modelling and Software* (M. Sànchez-Marrè, J. Béjar, J. Comas, A. Rizzoli & G. Guariso, eds). Barcelona, Spain, pp. 1–8.

Packman, J., Quinn, P., Hollis, J. & O'Connell, P. 2004 *Review of Impacts of Rural Land Use and Management on Flood Generation. Short Term Improvement to the FEH Rainfall-runoff Model: Technical Background*. DEFRA, London.

Parrott, A., Brooks, W., Harmar, O. & Pygott, K. 2009 Role of rural land use management in flood and coastal risk management. *J. Flood Risk Manage.* **2**, 272–284.

Pender, G., Bates, P., Wright, N., Hall, J. & Mason, D. 2005 *Next Generation Computer Modelling for the Prediction of Flood Level and Inundation Extent Part 1: Technical Specification of Research Needs*. Flood Risk Management Research Consortium. Available from: www.floodrisk.org.uk.

Robinson, M. & Dupeyrat, A. 2005 Effects of commercial timber harvesting on streamflow regimes in the Plynlimon catchments, mid-Wales. *Hydrol. Process.* **19** (6), 1213–1226.

Rose, S. & Rosolova, Z. 2007 *Ripon Land Management Project*. JBA Consulting, Skipton, UK.

Smith, L. P. 1976 *The Agricultural Climate of England and Wales: Areal Averages 1941–70*. HMSO, London.

USDA 1986 *Urban Hydrology for Small Watersheds*. United States Department of Agriculture.

Winter, C. & Nychka, D. 2010 Forecasting skill of model averages. *SERRA* **24** (5), 633–638.

Wheater, H., Reynolds, B., McIntyre, N., Marshall, M., Jackson, B., Frogbrook, Z., Solloway, I., Francis, O. & Chell, J. 2008 *Impacts of Land Management on Flood Risk: FRMRC WP2 at Pontbren*. Imperial College London, CEH Bangor. Available from: http://web.sbe.hw.ac.uk/frmrc/downloads/ur16_impacts_upland_land_management_wp2_2_v1_0.pdf.

Wilkinson, M. E., Quinn, P. F. & Welton, P. 2010 Runoff management during the September 2008 floods in the Belford catchment, Northumberland. *J. Flood Risk Manage.* **3** (4), 285–295.

Yadav, M., Wagener, T. & Gupta, H. 2007 Regionalization of constraints on expected watershed response behavior for improved predictions in ungaged basins. *Adv. Water Resour.* **30** (8), 1756–1774.

Ye, M., Neuman, S. P., Guadagnini, A. & Tartakovsky, D. M. 2004 Nonlocal and localized analyses of conditional mean transient flow in bounded, randomly heterogeneous porous media. *Water Resour. Res.* **40** (5), W05104.

Young, P. C. & Ratto, M. 2009 A unified approach to environmental systems modeling. *Stoch. Environ. Res. Risk. Assess.* **23**, 1037–1057.