

Integrating different types of information into hydrological model parameter estimation: Application to ungauged catchments and land use scenario analysis

Nataliya Bulygina,¹ Caroline Ballard,¹ Neil McIntyre,¹ Greg O'Donnell,² and Howard Wheeler^{1,3}

Received 28 July 2011; revised 2 May 2012; accepted 2 May 2012; published 15 June 2012.

[1] In hydrological modeling, two areas of application present particular challenges, first the modeling of ungauged catchments, and second the modeling of catchment nonstationarity; for example due to effects of land use change. The ungauged catchment problem requires that prior knowledge of the catchment is combined with evidence of behavior; for example from a regionalization exercise and/or spot flow measurements. Simulation of the effects of land use change requires that prior knowledge of the catchment is combined with information on the effects of that change on model parameters, generally in the absence of direct observations with which to condition the parameters. In both cases, ideally, all available sources of information about the behavior should be considered, and integrated in a way that maximizes the value of the information for model identification and uncertainty estimation. Using a formal Bayesian procedure, we combine three different sources of knowledge into a catchment scale conceptual model: (1) small-scale physical properties; (2) regionalized signatures of flow; and (3) available flow measurements. Applying the methodology to a distributed model for the Hodder catchment, UK, the physics-based information source contributed most to improving model performance, followed by peak flow times, and lastly the regionalized signatures. The flood frequency curve was evaluated under scenarios of land use change, and those changes that were significant relative to model uncertainty were identified.

Citation: Bulygina, N., C. Ballard, N. McIntyre, G. O'Donnell, and H. Wheeler (2012), Integrating different types of information into hydrological model parameter estimation: Application to ungauged catchments and land use scenario analysis, *Water Resour. Res.*, **48**, W06519, doi:10.1029/2011WR011207.

1. Introduction

[2] After a long history of hydrological model application to gauged catchments, simulating the hydrological fluxes in ungauged catchments remains a fundamental challenge. This includes both simulating fluxes in existing ungauged catchments, and predicting future fluxes; for example under climate and land use changes. One approach to this problem is a physics-based distributed model that, in theory, can be parameterized using local physical properties for both current and future conditions of change. However, this relies on assumptions that the model adequately captures all hydrological processes in a catchment of interest, and a large

number of parameters have to be accurately specified; it may result in poor parameter identifiability and insufficient prediction accuracy [Wheeler *et al.*, 1993]. Closely related to this, the effects of observable catchment characteristics or land use changes on physical properties are not generally well understood [O'Connell *et al.*, 2004; Parrott *et al.*, 2009]. Furthermore, the computational expense associated with physics-based models run at the catchment scale in general prohibits prediction uncertainty estimation on standard desktop computers [Jackson *et al.*, 2008; Ballard, 2011].

[3] Alternatively, conceptual, more parsimonious models can be implemented. Although this requires fewer parameters, the parameters correspond less directly to physical properties and so cannot be easily estimated directly from measurements or from literature. A common remedy to the problem is a regionalization strategy, where some knowledge is developed on how parameters vary spatially within a region. This is usually approached by relating model parameters in one way or another [McIntyre *et al.*, 2005a; Wagener and Wheeler, 2006] to some catchment descriptors (e.g., catchment area, steepness, soil permeability, geographical location) so that they can be estimated for any catchment for which the same descriptors may be estimated [e.g., Kapangaziwiri *et al.*, 2009]. The difficulty here is representing model parameter uncertainty in the catchment

¹Department of Civil and Environmental Engineering, Imperial College London, London, UK.

²School of Civil Engineering and Geosciences, Newcastle University, Newcastle-upon-Tyne, UK.

³School of Environment and Sustainability, University of Saskatchewan, Saskatoon, SK, Canada.

Corresponding author: N. Bulygina, Department of Civil and Environmental Engineering, Imperial College London, London, SW7 2AZ, UK. (n.bulygina@imperial.ac.uk)

descriptors-model parameter relationships. Alternatively, some *behavioral indices* (e.g., mean annual discharge, daily discharge standard deviation) are related to catchment descriptors and thus estimated for any catchment, then model parameters can be conditioned on these estimates [Yadav et al., 2007; Bulygina et al., 2009; Bulygina et al., 2011]. The potential advantages of the indices-based approach are: the regionalization step is not specific to any rainfall-runoff model; and a number of regional models linking flow indices to catchment properties are available [e.g., Boorman et al., 1995; USDA, 1986], hence avoiding, or at least reducing, the need to build new regional models; and the conditioning of parameter sets on indices maintains the dependencies between parameters as opposed to parameter regression methods which generally do not [McIntyre et al., 2005b]. Using either approach to regionalization, the knowledge about parameter variability over a region can be used to estimate parameter variability under future catchment changes, with the assumption that spatial information can substitute for a lack of temporal information [Wagener, 2007; Sivapalan et al., 2011].

[4] Although the computational burden of physics-based distributed models limits their suitability for catchment-scale modeling, they can still provide useful insights to assist our understanding of hydrological systems, particularly at the local scale. A method of introducing some of the power of *physics-based simulation* into a conceptual model for ungauged catchments application, while maintaining a computationally tractable model suitable for Monte Carlo analysis, has been proposed by Jackson et al. [2008] and Ballard [2011]. This builds on the upscaling concepts of Ewen [1996] as well as more general metamodeling concepts used in a wide range of scientific applications [Barton, 1998; Forsman and Grimvall, 2003; O'Hagan et al., 1999; Piñeros Garcet et al., 2006]. In the procedure of Jackson et al. [2008] and Ballard [2011], field-scale physics-based models are developed and used to simulate flow data for different types of hydrological response units. Simple conceptual models are fitted to the outputs. This provides a library of field-scale conceptual models, which can be integrated within a distributed modeling framework to simulate the hydrological response of ungauged catchments and explore effects of small scale changes on catchment scale response. Recognizing the considerable uncertainty in both the physics-based and metamodeling procedures, uncertainty is propagated using Monte Carlo methods.

[5] Another potential source of information comes from any *observations* which may be collected from the “ungauged” catchment, in particular it may be that introducing a relatively small number of catchment outlet flow observations into the model significantly reduces error and uncertainty [Seibert and Beven, 2009].

[6] Thus, arguably, there are three fundamentally different types of information that can be used to identify parameters for models of ungauged catchments: (1) physics-based information in the form of an output of a physics-based model; (2) empirical information derived from regional analysis of catchment response or from regional databases of flow indices; and (3) observations from the catchment being studied. All may be highly uncertain yet contain useful and complementary information. The use of the three types of information is implicit to many hydrological modeling studies

in that there is often some combination of physics-based conceptualization, regional data, and local observations utilized in the model building and estimation procedure. Typically, this is implemented by using prior knowledge based on physics and implicit or explicit use of regional knowledge to define a conceptual model and its prior parameter distributions. Where observations are available, this is generally followed by calibration, which in some cases includes the formal consideration of multiple observed variables [Kuczera and Mroczkowski, 1998; Wu et al., 2010]. However, the idea of treating the three different types of information source—physics-based models, regionalized data, and observations—in a consistent and complementary manner is an undeveloped area of research. Other research has used different sources of information in the model identification procedure, aiming at inference about errors, for example by evaluating inconsistency between prior models and regionalized data [Kapangaziwiri et al., 2009] or evaluating inconsistency across different observed variables and response modes [McIntyre et al., 2005b; Bulygina and Gupta, 2011].

[7] This paper describes a probabilistic framework for integrating the three sources of information. The paper focuses on assessing the value of the information for constraining prediction uncertainty toward observed values of streamflow. The potential application of the method to infer the nature and cause of prediction error is then discussed. The chosen prediction task is predicting the impacts of land use change on flood frequency using a spatially distributed conceptual rainfall-runoff model.

2. Method of Parameter Estimation

[8] Although the general approach described below can be applied to a much wider range of hydrologic modeling problems, the specifics of the method description apply to a typical setup of a spatially distributed conceptual rainfall-runoff model: the runoff time series from each of a large number of spatial runoff generating units are simulated and then integrated together using a channel routing model to give a time series of flow at the catchment outlet. The estimation problem therefore involves the parameters of the runoff generation units and also those of the channel routing model.

2.1. Probabilistic Setup for a Distributed Catchment Model

[9] We consider a conceptual distributed hydrological model structure S that is driven by input \mathbf{x} , requires parameters $\boldsymbol{\theta}$, and produces outputs \mathbf{y} (variables in bold denote vectors of values). Furthermore, we consider three types of information available about the hydrological behavior of a catchment of interest: (1) outputs \mathbf{F} from a physics-based model at the scale of the runoff generating unit, which correspond to a suitable parameter space for that model Φ ; (2) regionalized indices \mathbf{I} at runoff generating unit scale; and (3) a limited number of available flow observations \mathbf{D} at the catchment scale. Assuming an additive error structure, the conceptual model outputs and the available information are related via operators $f_F, f_I,$ and f_D as follows:

$$\mathbf{F} = f_F(\mathbf{y}) + \boldsymbol{\varepsilon}_F, \quad (1a)$$

$$\mathbf{I} = f_I(\mathbf{y}) + \boldsymbol{\varepsilon}_I, \quad (1b)$$

$$\mathbf{D} = f_D(\mathbf{y}) + \boldsymbol{\varepsilon}_D, \quad (1c)$$

where $\boldsymbol{\varepsilon}_F$, $\boldsymbol{\varepsilon}_I$, and $\boldsymbol{\varepsilon}_D$ are vectors of random variables representing error models characterized by a joint probability distribution with parameters $\boldsymbol{\psi}$. \mathbf{F} would typically be generated through multiple realizations of the physics-based model's parameter sets sampled from Φ , so that, if $\boldsymbol{\varphi}$ is a physics-based model parameter set, $\boldsymbol{\varphi} \sim p(\cdot|\Phi)$.

[10] A conceptual model parameter posterior given the three sources of information can then, initially, be presented as an integration over the physics-based model parameter space,

$$p(\boldsymbol{\theta}|\Phi, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) = \int p(\boldsymbol{\theta}|\boldsymbol{\varphi}, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) \cdot p(\boldsymbol{\varphi}|\Phi, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) d\boldsymbol{\varphi}. \quad (2)$$

The physics-based model parameter space descriptor Φ holds all necessary information regarding $\boldsymbol{\varphi}$, so that $p(\boldsymbol{\varphi}|\Phi, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) = p(\boldsymbol{\varphi}|\Phi)$. Therefore the above equation simplifies to

$$p(\boldsymbol{\theta}|\Phi, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) = \int p(\boldsymbol{\theta}|\boldsymbol{\varphi}, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) \cdot p(\boldsymbol{\varphi}|\Phi) d\boldsymbol{\varphi}. \quad (3)$$

This can be numerically approximated as follows:

$$p(\boldsymbol{\theta}|\Phi, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) \approx \frac{1}{M} \sum_{j=1}^M p(\boldsymbol{\theta}|\boldsymbol{\varphi}_j, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}), \quad (4)$$

where $\boldsymbol{\varphi}_j$ is the j th of M draws from $p(\cdot|\Phi)$.

[11] We denote runoff generating unit parameters by $\boldsymbol{\theta}_L$, and river routing unit parameters by $\boldsymbol{\theta}_R$, so that $\boldsymbol{\theta} = \{\boldsymbol{\theta}_L, \boldsymbol{\theta}_R\}$. Therefore the posterior from the right hand side of equation (4) can be factorized as

$$p(\boldsymbol{\theta}|\Phi_j, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) = p(\boldsymbol{\theta}_L|\boldsymbol{\varphi}_j, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) \cdot p(\boldsymbol{\theta}_R|\boldsymbol{\theta}_L, \boldsymbol{\varphi}_j, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}). \quad (5)$$

So far the theoretical development has been quite generic, but further steps depend on available information types. The rest of the section specifies the sources of information used in the case study and formulates likelihood functions.

2.2. Posterior Distribution Using Specific Types of Information

[12] Each runoff generating unit is classified as one of N response unit types according to a set of variables (soil type, land cover, and soil condition), which are perceived to be the most important controls on response or which must be controlled as part of land use scenario analysis. All less important variables are treated as random and contribute to model uncertainty. Models for each of the N response unit types are treated as acting independently of each other; and each model parameterized with a parameter set $\boldsymbol{\theta}_L^k$, where $k = 1, N$.

[13] For many applications, no unit-scale observations of hydrological fluxes or states are available, so that each parameter set $\boldsymbol{\theta}_L^k$ is restricted based on the relevant information from the physics-based model parameterized with $\boldsymbol{\varphi}_j^k$ and regionalized indices \mathbf{I}^k only. It may be the case that catchment scale flow measurements, represented by \mathbf{D} , can also

contribute information about some of the unit scale responses via a downscaling procedure. However, unit scale signals are likely to be considerably smoothed out by the channel routing [O'Connell *et al.*, 2004]. Therefore, $\boldsymbol{\theta}_L^k$ and \mathbf{D} are considered to be independent, and information provided by \mathbf{D} is only used to condition the channel model parameters $\boldsymbol{\theta}_R$. Then equation (5) is rewritten to include these assumptions:

$$p(\boldsymbol{\theta}|\boldsymbol{\varphi}_j, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) = \prod_{k=1}^N p(\boldsymbol{\theta}_L^k|\boldsymbol{\varphi}_j^k, \mathbf{I}^k, \mathbf{x}, S, \boldsymbol{\psi}) \times p(\boldsymbol{\theta}_R|\boldsymbol{\theta}_L, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}). \quad (6)$$

Using Bayes' law, equation (6) becomes

$$p(\boldsymbol{\theta}|\boldsymbol{\varphi}_j, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) = \prod_{k=1}^N \frac{p(\boldsymbol{\varphi}_j^k, \mathbf{I}^k|\boldsymbol{\theta}_L^k, \mathbf{x}, S, \boldsymbol{\psi}) \cdot p(\boldsymbol{\theta}_L^k|\mathbf{x}, S, \boldsymbol{\psi})}{p(\boldsymbol{\varphi}_j^k, \mathbf{I}^k|\mathbf{x}, S, \boldsymbol{\psi})} \cdot \frac{p(\mathbf{D}|\boldsymbol{\theta}_R, \boldsymbol{\theta}_L, \mathbf{x}, S, \boldsymbol{\psi}) \cdot p(\boldsymbol{\theta}_R|\boldsymbol{\theta}_L, \mathbf{x}, S, \boldsymbol{\psi})}{p(\mathbf{D}|\mathbf{x}, S, \boldsymbol{\psi})} \quad (7)$$

which can be rewritten as follows:

$$p(\boldsymbol{\theta}|\boldsymbol{\varphi}_j, \mathbf{I}, \mathbf{D}, \mathbf{x}, S, \boldsymbol{\psi}) = C \prod_{k=1}^N p(\boldsymbol{\varphi}_j^k, \mathbf{I}^k|\boldsymbol{\theta}_L^k, \mathbf{x}, S, \boldsymbol{\psi}) \cdot p(\mathbf{D}|\boldsymbol{\theta}_R, \boldsymbol{\theta}_L, \mathbf{x}, S, \boldsymbol{\psi}) p(\boldsymbol{\theta}|\mathbf{x}, S), \quad (8)$$

where C is a normalizing constant, and $p(\boldsymbol{\theta}|\mathbf{x}, S)$ is a prior distribution of the catchment scale model parameters, which in the absence of any other information is assumed to be a uniform distribution.

[14] The challenge is then to devise and assess suitable operator functions f_F , f_I , and f_D and their error distribution parameters $\boldsymbol{\psi}$, and the associated likelihood functions in equation (8).

2.3. Likelihood Functions for the Unit-Scale Runoff Generating Models

[15] To simplify notation we drop conditional dependence on inputs \mathbf{x} and conceptual model structure S . Below, based on the nature of information sources used, we assume that the physics-based and regionalized information are conditionally independent given runoff generating unit parameters, i.e., $p(\boldsymbol{\varphi}_j^k, \mathbf{I}^k|\boldsymbol{\theta}_L^k, \boldsymbol{\psi}) = p(\boldsymbol{\varphi}_j^k|\boldsymbol{\theta}_L^k, \boldsymbol{\psi}) \cdot p(\mathbf{I}^k|\boldsymbol{\theta}_L^k, \boldsymbol{\psi})$. The assumption that the errors in \mathbf{I} and $\boldsymbol{\varphi}$ are independent will not be valid if both contain information from the same origin; for example if a common soils database or elevation model has been used for both the physics-based modeling and the regionalization. If sufficient data were available to characterize it, such dependence could easily be accounted for by considering a joint distribution of the errors $\boldsymbol{\varepsilon}_F$ and $\boldsymbol{\varepsilon}_I$.

[16] The regionalized information \mathbf{I} in our case comes from two readily available indices: the Base Flow Index (BFI) [Boorman *et al.*, 1995] and the Soil Conservation Service Curve Number (CN) [USDA, 1986]. The unit scale base flow index is mainly influenced by local soil and geology, while the curve number is also influenced by small-scale land use effects. Bulygina *et al.* [2011] describe a Bayesian method of integrating these two indices together into the likelihood $p(\mathbf{I}^k|\boldsymbol{\theta}_L^k, \boldsymbol{\psi})$ so that the overall response of the conditioned model is consistent with the information and uncertainty in both indices. In summary, the method

involves taking a sample parameter set θ_L^k , running the run-off generating model, calculating BFI and CN using the model result according to their definitions in the above sources, and assigning a probability to the sample parameter set. This probability is in direct proportion to the joint probability density calculated from the bivariate normal distribution for \mathbf{I} , which is defined using regional information about BFI and CN published in the above sources. Thus, the method of Bulygina et al. [2011] is a formalization of the relatively well established procedure of conditioning model parameters on regionalized indices [Yadav et al., 2007].

[17] In the physics-based modeling procedure, small-scale high-resolution physics-based models are developed based on understanding of local hydrological processes, literature review, and local soil, land cover, and topographic databases. The physics-based models used in the case study below are described by Ballard et al. [2010] and Ballard [2011]. While, in theory, the physics-based models themselves could be used directly within the catchment model, their high spatial resolution and associated computational expense prohibits catchment-scale long-term evaluations. In particular, the use of Monte Carlo methods to estimate prediction uncertainty would be prohibitively time consuming. Therefore, instead of being used directly, the physics-based model outputs are used to support the conditioning of the faster-running conceptual model by quantifying the likelihood $p(\phi_j^k | \theta_L^k, \psi)$.

[18] A modeler can design the likelihood function to extract the information from the physics-based model which is perceived to be of value; for example the information which is missing or most uncertain from the regionalization, and/or information thought to be of most value for the particular application. Here we use a combination of two measures: (1) a measure that evaluates the fit between the peak flow rates for n largest rainfall events, and (2) a measure that evaluates the fit of the corresponding times to peak:

$$p(\phi_j^k | \theta_L^k, \sigma_Q, \sigma_T) = \prod_{i=1,n} N(Q_{\varphi_j^k}^i | Q_{\theta_L^k}^i, \sigma_Q) \prod_{i=1,n} N(T_{\varphi_j^k}^i | T_{\theta_L^k}^i, \sigma_T), \quad (9)$$

where $Q_{\varphi_j^k}^i$ and $Q_{\theta_L^k}^i$ are peak flows for the k th of n events modeled by the physics-based and conceptual model, respectively; $T_{\varphi_j^k}^i$ and $T_{\theta_L^k}^i$ are corresponding times to peak modeled by the physics-based and conceptual model, respectively; σ_Q and σ_T are standard deviations for peak flow and time to peak and are assumed to be constant over all n events; and N indicates that a normal distribution is assumed.

[19] The standard deviations characterize a tolerance to mismatches between physics-based and conceptual models, and are therefore hard to define objectively. While other choices may be made, we use Jeffrey's prior [Box and Tiao, 1992] to integrate out the likelihood dependence on σ_Q . We use a small number of time steps as a tolerance to time to peak mismatch.

[20] The likelihood is intended to complement the information provided by the regionalization on separation between slow and fast flows (BFI), and on event stormflow volumes (CN). Moreover, it reflects the interest in high

flows, and the fact that good high flow performance is unlikely to be achieved unless some specific high flow criteria are included in the conditioning [Wagener et al., 2004; Wagener and McIntyre, 2005]. ϕ_j^k could equally represent a sample from a set of physics-based model structures and associated parameter sets, thus recognizing uncertainty in the physics-based model structure. However, in the case study the uncertainty in the physics-based model is assumed to be sufficiently modeled by sampling only parameter sets.

2.4. Likelihood Function for the Channel Routing

[21] As this study is aimed at applications in poorly gauged catchments, the observations of a catchment scale response \mathbf{D} are restricted to several high flow peak arrival times. Although other choices can be made, the likelihood $p(\mathbf{D} | \theta_R, \theta_L, \psi)$ from equation (8) is defined using a product of uniform distributions U on $[-\Delta t; \Delta t]$, where Δt is the modeling time step:

$$p(\mathbf{D} | \theta_R, \theta_L, \psi) = \prod_{i=1,l} U_{[-\Delta t; \Delta t]}(\mathbf{P}_i - \mathbf{D}_i), \quad (10)$$

where \mathbf{P} and \mathbf{D} are vectors of the observed and modeled peak flow arrival times and l is the number of pairs considered. This allows the acceptance of only the channel routing parameters that result in peak arrival times that differ from the observed times by one time step at most.

[22] As explained in section 2.2, these flow data are assumed only to add information to the channel routing, and not to the response unit parameter distribution. This serves to limit the flow data requirements to those that are useful for estimating only θ_R , in this case a few measurements of time to peak. To further reduce flow data requirements, the other two sources of information could be introduced; for example via the regionalization of stream hydraulic parameters [e.g., Koren et al., 2004] or by relating the conceptual routing parameters to hydrodynamic models [e.g., Camacho and Lees, 2000; Dooge and O'Kane, 2003]. However, with the focus on modeling the impact of changes to the response units, elaboration was not considered appropriate.

3. Case Study—The Hodder Catchment

[23] The land use effect estimation strategy described in section 2 is applied to the Hodder catchment in northwest England. Land use changes have recently been made in the Hodder catchment under the Sustainable Catchment Management Programme of the region's water supplier, United Utilities [Ewen et al., 2010].

3.1. Catchment Description and Scenarios of Land Use Change

[24] The catchment area is approximately 261 km², with elevations varying between 40 and 544 m above sea level. Land cover in the Hodder catchment is predominantly grazed grassland consisting of both grassland that has been agriculturally improved by drainage, plowing, and fertilization, and grassland under rough grazing. There are also small areas of deciduous and coniferous forest, and arable farming (Table 1). There is a regulated reservoir (Stocks Reservoir) in the uplands with a contributing area of 37 km².

Table 1. Land Management and Associated Soil Condition [USDA, 1986] in the Hodder Catchment

	Deciduous Forest	Coniferous Forest	Improved Grassland	Rough Grazing	Other
Area %	4.2	2.8	60.7	30.5	1.8
Condition	Good	Good	Fair	Good	–

The annual rainfall is around 1500 mm in upland areas, decreasing to 1100 mm at lower elevations, and the ratio of long term precipitation to potential evapotranspiration ranges between 2.5 and 3.4 for the lower and upper parts of the catchment, respectively. The soils are dominated by blanket peatlands (some of which are drained) and slowly permeable soils (Table 2) assumed to be in either “Good” or “Fair” conditions (Table 1) [Ewen *et al.*, 2010]. For grazing, Good condition means insignificant soil compaction due to low stocking density, “Poor” condition means high soil compaction due to heavy grazing, and Fair condition means soil being moderately compacted (for other land use types, see definitions of soil conditions by USDA [1986]).

[25] Within the Hodder catchment there is a strong link between elevation, soil association, and land use, which restricts the plausible range of land use scenarios. Four potential changes to land management are investigated and given in Table 3. The patterns of land use changes for the four scenarios are shown in Figure 1.

3.2. Description of Hydrological Data

[26] Four tipping bucket rainfall gauges distributed over the Hodder catchment at different elevations provide rainfall observations (Figure 2). Streamflows are measured by a compound crump profile weir at Hodder Place, and water release is recorded at Stocks Reservoir. Rainfall is measured with 0.2 mm tips, and flow time series data are averaged over 15 min intervals. There is also a daily record of water abstractions in the catchment headwaters. Daily potential evapotranspiration (PE) rates for different land use types were estimated by the MORECS model, which is based on the Penman-Monteith equation [Hough and Jones, 1997].

[27] This paper focuses on two time periods: (1) model evaluation periods when spatially distributed rainfall and streamflow measurements at the Hodder gauge are available, and (2) a flood frequency evaluation period requiring long term rainfall and potential evapotranspiration data records. For the former, winter and summer data periods,

15 November 1999–15 March 2000, and 1 June 2000–31 August 2000, are selected on the basis of availability of gap-free data. One month warm-up periods, not included in the model conditioning or evaluation, are used to estimate the initial model state. The flood frequency estimation uses 10 years of gap-free rainfall observations recorded by a gauge located in the northwest part of the catchment, at an elevation of 167 m, from 15 November 1997 to 31 December 2007. Model performance evaluation is based on the shorter aforementioned periods, so flow measurements over this longer period are not needed. The times to peak flows for the three largest flood peaks in the winter period, 15 November 1999–15 March 2000, are used as large scale data **D**.

3.3. Model Description

[28] The Hodder catchment is represented as a set of 200 m × 200 m runoff generating response units, and HOST soil type, land use, soil condition, and flow direction are prescribed for each unit. It is assumed that the 200 m grid scale and 15 min time step adequately captures the hydrological response variability for the purpose of catchment scale modeling. The physics-based models for each soil type–land use combination present in the Hodder catchment are developed at this response unit scale [Ballard *et al.*, 2010, 2011]. Two different physics-based models were employed to represent the various runoff classes within the Hodder catchment. The first is a two-dimensional hillslope model that couples Richards’ equation [Hillel, 1971] for subsurface flow, the kinematic wave equation [Singh, 1996] for surface flow, an adapted version of the Rutter model [Valente *et al.*, 1997] to represent interception, and the Penman-Monteith equation for potential evaporation [Allen, 2006]. The model is used to represent all mineral soils and accounts for land use scenarios through changes in the distributions of model parameters. The second model is specific for the Winter Hill (peatland) soil types, and simulates peatland drainage management types, i.e., drainage through “grips,” blocked grips, and no drainage (intact soil). The model allows for complex drainage geometry and couples a one-dimensional Boussinesq model of the subsurface [Beven, 1981] with kinematic wave models for overland and drain flows. The model equations were integrated using a spatially implicit, error controlled forward time step solution [Ballard, 2011]. Alternative drainage management scenarios are implemented through model structural changes. The developed physics-based models

Table 2. HOST Soil Types in the Hodder (With Area Greater Than 3%)

HOST Type	Description	BFI _{HOST} ^a	Area %
4	Free draining permeable soils on hard but fissured rocks with high permeability but low to moderate storage capacity	0.79	4.2
10	Soils seasonally waterlogged by fluctuating groundwater and with relatively rapid lateral saturated conductivity	0.52	3.4
15	Permanently wet, peaty topped upland soils over relatively free draining permeable rocks	0.38	10.5
24	Slowly permeable, seasonally waterlogged soils over slowly permeable substrates with negligible storage capacity	0.31	35.7
26	Permanently wet, peaty topped upland soils over slowly permeable substrates with negligible storage capacity	0.24	18.1
29	Permanently wet upland blanket peat	0.23	21.2

^aDerived from the UK HOST soil classification [Boorman *et al.*, 1995].

Table 3. Scenarios of Land Management Change: Descriptions and Areas Affected

Scenario	Description	Area %
1	<i>Commercial forestry</i> (Sitka and Norway Spruce) in areas coinciding with Wilcocks soils, and under rough grazing. The forestry is restricted to elevations below 462 m due to a windthrown hazard [Miller <i>et al.</i> , 1987].	10.4
2	A reversion of the more marginal improved grassland back to rough grassland, coupled with a <i>de-intensification</i> of the remaining managed grassland. Areas classified as marginal are: areas above 150 m elevation, areas with slopes >11 deg and areas on Wilcocks and Belmont soils.	34.2
3	A conversion of rough grazing to <i>more intense grazing</i> (improved grassland) on Wilcocks soils with gentler slopes (<11 deg), requiring underdrainage, lime, and fertilizer application.	7.1
4	<i>Riparian deciduous tree planting</i> along major water ways (defined as having an upstream contributing area >2 km ²). Changes were not performed in areas coinciding with Winter Hill soils, which are in exposed locations and are ill-suited due to wetness.	6.6

have various structural limitations; for example they do not explicitly account for any possible flow convergence/divergence, soil macroporosity, or possible hydrological connectivity between runoff generating response units. These limitations were justified by the need for tractable physics-based models, the lack of detailed data to support more complex models, and by the need for a set of working assumptions as a starting point for the research [Ballard, 2011]. An underlying assumption in the method is that the uncertainty in \mathbf{F} due to the various limitations of the physics-based model is sufficiently represented by sampling

its parameter space Φ . The parameter space was restricted based on soils information from NSRI soils database [NSRI, 2011], a 5 m resolution DEM of the catchment, as well as supporting information from multiple literature sources. Full details of the parameter range selection are provided by Ballard (2011). M random samples are taken from Φ , which is defined by independent uniform distributions [Ballard, 2011].

[29] Due to the high computational expense, the physics-based models are used to represent only the most prevalent response unit types—all combinations of: (1) the four

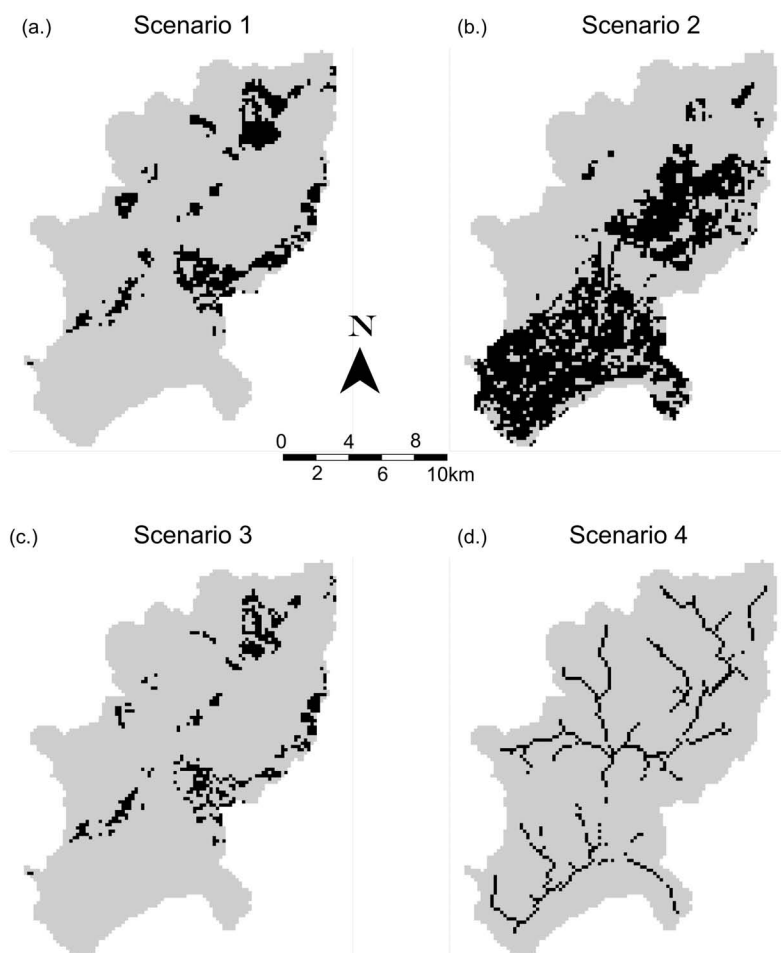


Figure 1. Patterns of land use change in the Hodder catchment. Black areas represent areas that undergo change under the four scenarios.

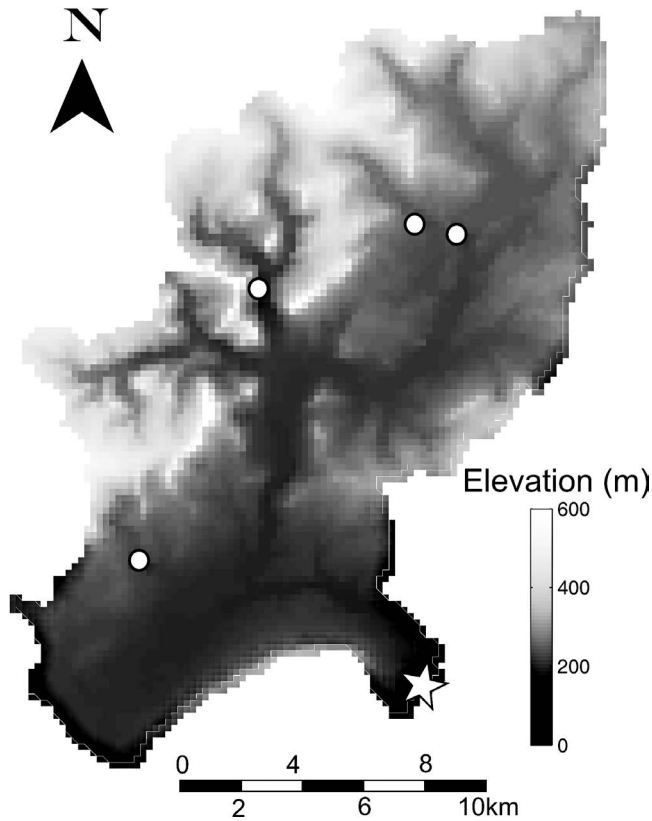


Figure 2. The Hodder catchment topography and instrumentation. White circles indicate locations of rain gauges, and white star indicates location of a compound crump weir at the Hodder Place (the Hodder catchment outlet).

dominant soil HOST types, 15, 24, 26, and 29; and (2) four dominant land use types: deciduous and coniferous forest, improved grassland, and rough grazing (applicable only to HOST classes 15, 24, and 26); and intact peat, drained peat, and peat with blocked drains (applicable only to HOST class 29). Thus 12 sets of M simulations were produced to represent F for these 12 combinations of response unit types. For other soil types (Table 2) and land use types (Table 1), parameters θ_L are constrained only by the regionalized indices. Also, the physics-based models do not distinguish between Good, Fair, and Poor soil conditions. Instead, the range of soil conditions within one soil type is represented in the physics-based model by random sampling of the soil property parameters within suitable prior ranges. Therefore information about the effect of soil condition comes only from the regionalized indices. Hence, there is scope to add more information into equation (9) by building and running more physics-based models, however the potential value seems unlikely to warrant the significant extra cost.

[30] The process of identifying the conceptual model structure is reported only briefly here—for a full description see *Ballard* [2011]. A conceptual model structure was chosen that consistently well captures the 15 min flows generated by the physics-based models at the 200 m \times 200 m scale over all the relevant combinations of soil type and land use. Several conceptual model structures were examined, and a catchment moisture deficit model [Evans and

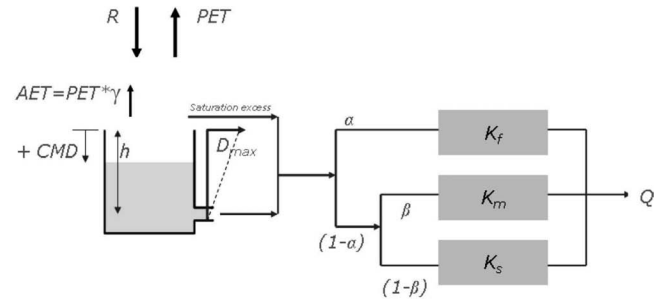


Figure 3. Local rainfall-runoff conceptual model structure (catchment moisture deficit model) used for the response units. It has a soil moisture deficit controlled runoff generation, and three runoff routing stores in parallel.

Jakeman, 1998] with three routing stores in parallel (Figure 3) was preferred. The catchment moisture deficit model is a conceptual store that drains at a linearly increasing rate as the catchment moisture deficit approaches zero, at which point the drainage rate is equal to the maximum drainage rate D_{max} . Drainage continues and evapotranspiration is equal to $\gamma \times PE$ until the catchment moisture deficit is greater than h , where γ is a proportionality constant. The soil moisture balance of each unit is applied over each 15 min time step and any saturation excess rainfall volume is added to the drainage volume to give an effective rainfall value that is applied uniformly over each time step. Such averaging over time steps may affect parameter and flow estimates at the 200 m \times 200 m scale [Kavetski and Fenicia, 2011], although the time step of 15 min is small compared to the average response time of the Hodder catchment (the order of 1000 min). The runoff (drainage plus excess rainfall) is split between the three linear stores according to two split coefficients, α and β , and routed with residence times K_f , K_m , and K_s ($K_f < K_m < K_s$). An exact solution to the linear routing equation is used. Each unit-scale conceptual model requires eight parameters γ , h , D_{max} , α , β , K_f , K_m , and K_s to be estimated, with prior parameter ranges given in Table 4.

[31] The gridded response units are connected with a stream network model, which is made up of a number of stream sections (one for each runoff generating response unit). A contributing area of 2 km² is assumed to initiate a first order channel section. We use a channel routing model with a spatially distributed celerity field described by *Maidment et al.* [1996]. Constant flow celerity c is assigned to each stream section based on its slope s and upstream contributing area A , so that $c = c_a \sqrt{s * A} / [\sqrt{s * A}]_a$, where c_a is a catchment averaged celerity and $[\sqrt{s * A}]_a$ is a catchment averaged slope-area combination. Therefore, when slope becomes steeper, and/or for larger contributing area, flow celerity becomes larger. The response of each stream unit is modeled as a lagged linear reservoir, where the lag is the grid's channel length divided by the corresponding estimate of celerity, and the ratio between the reservoir residence time and the response unit travel time is 4:1 [Maidment et al., 1996]. The reservoir adds some diffusion to otherwise pure flow translation. The ratio as well as the mathematical form for each unit celerity calculation is considered to be a part of the model conceptual structure, and therefore considered to be correct and fixed (see section 2.1). Alternatively, the ratio and the power in the

Table 4. Conceptual Model Prior Parameter Ranges

Parameter	D_{\max} (mm/15 min)	h (mm)	γ	α	β	K_f (15 min)	K_m (15 min)	K_s (15 min)	c_a (m s ⁻¹)
Range	0–0.2	50–100	0.5–2	0–1	0–1	1–15	15–60	60–1000	0.5–3

celerity calculations could be treated as additional uncertain parameters and be conditioned within the Bayesian scheme. However, in the case study, the only channel routing parameter to be treated as uncertain is the average flow celerity c_a (prior range is in Table 4). Recognizing that the nonlinearity in channel routing may significantly affect land use impacts on flood peaks and how these impacts vary spatially [O'Donnell *et al.*, 2011], a more sophisticated channel model would be a useful development of the case study.

[32] The prior parameter distribution for the conceptual model is based on 50,000 samples drawn from a uniform distribution (Latin hypercube method). The posterior parameter distribution, described in section 2, is approximated by 100 parameter sets drawn from the population of 50,000 using the importance sampling method [Doucet *et al.*, 2000]. The main reasons for the sample size choice are that the computational expense of a catchment scale model run prohibits larger samples, and the performance statistics were insensitive to doubling of the sample size. Then the posterior parameter space of the distributed model (θ in equation (2)) is represented by 100 response unit model parameter sets for each response unit type, as well as 100 average celerity values. Based on this, the whole catchment is parameterized in such a way that response units of the same type, within any one model run, have identical parameter values. This assumption avoids the variability of runoff between units of the same type being averaged out when integrating over the units, which in turn would lead to very low uncertainty over the 100 samples. Improving the spatial structure of parameter errors is another possibility for improving the application.

3.4. Underlying Assumptions

[33] The case study is based on the following assumptions:

[34] 1. The chosen model structure is assumed to be correct, so that model identification is reduced to model parameter identification.

[35] 2. The post-change response unit parameters were assumed not to depend on the prechange parameters.

[36] 3. All three sources of information—**F**, **I**, and **D**—are considered to be conditionally independent.

[37] 4. Available large-scale observations are presumed not to provide any information that would significantly constrain the response unit parameters.

[38] 5. Likelihoods are defined by simple distribution functions (normal and uniform) assuming that information sources are unbiased.

[39] 6. The applicability to the UK of the Curve Number system, which was developed for use in the United States, is presumed after Bulygina *et al.* [2011] and Hess *et al.* [2010].

[40] 7. Land use change is assumed not to cause significant changes to channel routing during flood flows, and it is

assumed that any change is captured by existing uncertainty in the routing parameters.

[41] 8. All used sources of information are consistent with the chosen model structure.

[42] An applicability of the assumptions above is discussed at the end of the paper.

3.5. Parameter Sensitivity and Identifiability

[43] The posterior parameter distributions of the response unit conceptual model are estimated for all combinations of soil type, soil condition, and land use currently present in the Hodder catchment, and under the land use scenarios. A series of tests is conducted to examine parameter identifiability, and the relative values of the regionalized and physics-based information. Specifically, the tests address the following questions:

[44] 1. How different are the posterior distributions from the prior (uniform) distributions?

[45] 2. How different are the posterior distributions derived using both regionalized and physics-based information sources from those derived using only the former?

[46] 3. How different are the posterior distributions derived using both information sources from those derived using only the physics-based information?

[47] The Kolmogorov-Smirnov statistic is used to quantify the differences between different pairs of distributions. The tests are conducted for three HOST soil types (15, 24, and 26), four land management types, (deciduous and coniferous forest, improved grassland, and rough grazing) and three soil condition types (poor, fair, and good). Therefore, to answer each of these questions, 36 Kolmogorov-Smirnov tests are conducted.

3.6. Estimation of Prediction Quality

[48] Since continuous measurements of catchment-scale rainfall and flow are available under recent land use, it is possible to evaluate model performance. Due to the probabilistic nature of the prediction, measures different from traditional deterministic measures of performance are required. Two main aspects of a probabilistic prediction are its reliability and sharpness, and these can be assessed using the following two measures (see section A1 for detailed descriptions):

[49] 1. An analog Nash-Sutcliffe efficiency introduced by Bulygina *et al.* [2009], which lumps both reliability and sharpness into one number.

[50] 2. QQ plots, described by Laio and Tamea [2007], which graphically compare the modeled and observed cumulative frequency distributions as well as allowing calculation of reliability index δ and a sharpness index π [Renard *et al.*, 2010; Bulygina and Gupta, 2011].

[51] The QQ plots, and reliability index α compare the predicted distribution of values with the observed distribution of values, thus requiring a measurement error structure. Making the common assumption about flow measurement errors being heteroscedastic, with variance increasing with flow [Sorooshian and Dracup, 1980; Thiemann *et al.*,

2001; *Bulygina and Gupta*, 2011], the flows are transformed using a Box-Cox transformation ($\lambda = 0.3$) to make the errors more homoscedastic.

3.7. Impact of Land Use on the Flood Frequency Curve

[52] Using the climate record from 1997 to 2007, and by sampling from the posterior distributions, 50 realizations of continuous-time flows are generated using the model under current (baseline) land use and another 50 are generated under each of the specified land use change scenarios. For each land use change scenario and each realization, only the parameters of response units that undergo land use change are different from the corresponding baseline response unit parameters: this helps to ensure that pre- and post-change responses are related.

[53] In the derivation of flood frequency from the simulated time series, the peaks-over-threshold method is preferred over the annual maxima method, as it makes better use of the data [*Beguaria*, 2005]. The number of peaks per year is set to a typical value of 3 [*Lang et al.*, 1999]. To help ensure peak independence, we consider only the highest peak flow per rainfall event, discarding any smaller peaks. The magnitudes of the peaks-over-threshold are fitted using a generalized Pareto distribution [*Cunnane*, 1979; *Naden*, 1992] with peak arrival time represented by a Poisson distribution, and the fitting is carried out using a method of probability-weighted moments [*Hosking and Wallis*, 1987]. Flood frequency curves are derived for each of the 50 realizations to represent uncertainty in T -year flow magnitudes. Flood return periods of up to only 10 years are considered, due to the limited duration of the available climate records. Having estimated the (uncertain) T -year return peak flows for the current land use (Table 1) and for the specified scenarios (Table 3), we calculate 50 relative differences between the scenario and current land use for each return period. The values are used to assess the significance of the change in T -year flood peaks, using a paired nonparametric Wilcoxon test [*Wilcoxon*, 1945; *Siegel*, 1956]. The test makes a single assumption that a distribution of differences between baseline and scenario peak flows is symmetric.

4. Results

4.1. Sensitivity and Identifiability Analysis

[54] Figure 4 shows the fraction of times the null hypothesis—that two specified parameter distributions are not different when altering specified input information—has been rejected. A value of 1.0 corresponds to a situation when all (36) marginal posterior distributions are different, and therefore the parameter is consistently sensitive to the considered change in input information; and vice versa a value of 0 means the parameter is consistently insensitive to the change. When both the physics-based and regionalized information sources are used, the maximum drainage rate D_{\max} , fast store residence time K_f , PE adjustment factor γ , and proportion of flow that goes into the fast store α are consistently well identifiable for the majority of soil type–land use–soil condition combinations (top row of data in Figure 4). The posterior distribution seems to be more affected by the physics-based information than by the regionalized information: the distribution conditioned only on the regionalized information is consistently different from that conditioned on both sources of information (middle row of data in Figure 4). Using information from only the physics-based modeling leads to distributions similar to the posterior distributions (bottom row of data).

[55] As well as the parameters of the response unit model, the channel routing parameter c_a is estimated. Based on three peak flow arrival times (all flows above $150 \text{ m}^3 \text{ s}^{-1}$), average celerity samples lie between 1.9 and 2.1 m s^{-1} and are relatively uniformly distributed between these bounds. For flow peaks less than $150 \text{ m}^3 \text{ s}^{-1}$, lower celerity values fit the arrival times better, but significantly delay the three arrival times considered, which are given more importance due to the focus on flooding. This result indicates, however, that there is scope to improve the reliability of the method, in particular for extrapolation to more extreme flood events, which may have even higher c_a values, by including a variable celerity routing model [*Singh*, 1996].

4.2. Model Performance

[56] Figures 5a–5d show the flow prediction bounds (90% confidence level) for a representative portion of the

Information	D_{\max}	h	γ	α	β	K_f	K_m	K_s
Prior distribution	1.00	0.39	0.69	0.69	0.47	1.00	0.19	0.19
Regionalised indices	0.89	0.58	0.75	0.75	0.67	1.00	0.33	0.28
Physics-based models	0.08	0.08	0.08	0.36	0.28	0.19	0.11	0.22

Figure 4. Data representing the fractions of times (out of the 36 land use and soil type combinations) that the parameter marginal probability distributions are significantly different (0.05 significance level) when the information used in the conditioning is changed. The posterior distribution derived by using all information sources is compared with (top row) the prior distribution, (middle row) the distribution obtained by using regionalized indices only, and (bottom row) the distribution obtained by using physics-based information only. The shading of each box indicates values between 0 and 0.25 (pure black), 0.25 and 0.5 (dark gray), 0.5 and 0.75 (light gray), and 0.75 and 1 (pure white).

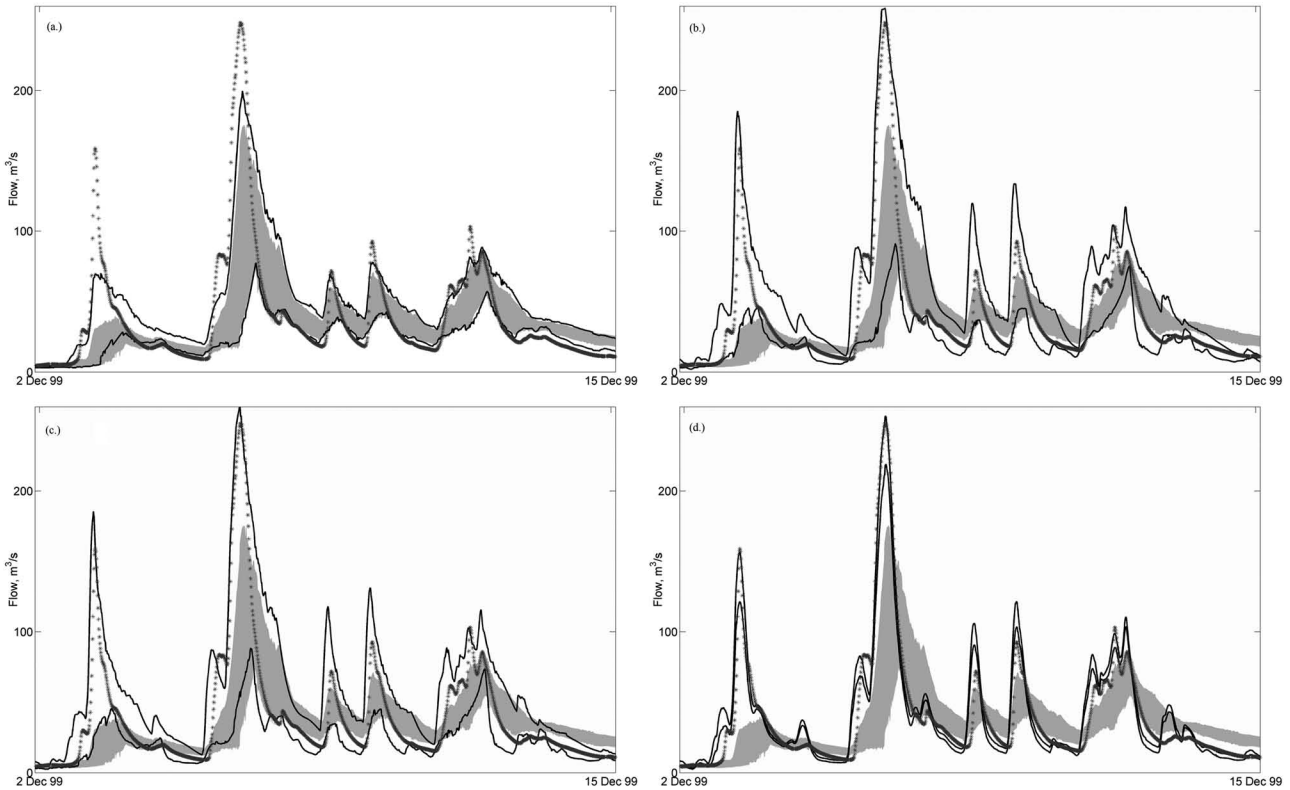


Figure 5. 90% flow prediction confidence intervals depending on information used for model conditioning. Gray area indicates prior prediction bounds. Dots are observed flows. Black lines are prediction bounds when: (a) only regionalized indices are used, (b) only physics-based information is used, (c) both regionalized indices and physics-based information are used, and (d) observations of time of flow peak, regionalized indices and physics-based information are used.

evaluation period depending on the amount of information used for the parameter conditioning: only prior, only regionalized, only physics-based, both regionalized and physics-based, and all three information sources. Visually, the physics-based information leads to larger changes in the prior bounds than the regionalized information. Further inclusion of the observed peak flow arrival times as a third source leads to a significant narrowing of the prediction bounds.

[57] The analog Nash-Sutcliffe efficiency NS^* and traditional Nash-Sutcliffe efficiency NS are given in Table 5 for both evaluation periods. These NS^* and NS values are calculated using the time series made up from the mean of the simulated flow distributions at each time step. The NS^* is given for unconditional (prior) model simulations NS^*_{prior} , the model simulations conditioned on all three available sources of information NS^* , conditional on regionalized indices only NS^*_I , conditional on physics-based information only NS^*_{Φ} , and conditional on regionalized and physics-based

information $NS^*_{I,\Phi}$. Conditioning the parameters on all three available information sources leads to large improvements in the NS^* performance statistic; for example in the winter period NS^* increases from 0.54 to 0.86. The physics-based information Φ leads to the largest improvement in NS^* , followed by information on peak flow arrival D , and lastly by the regionalized information I .

[58] The QQ plots for the winter and summer periods show that the flows overall tend to be underpredicted (the plots are concave) (Figure 6). Because the stream routing model is estimated to be representative of high flows rather than low flows, QQ plots that only take into account the high flows ($>50 \text{ m}^3 \text{ s}^{-1}$) are much closer to the ideal 1:1 line (Figure 6). The improved reliability (δ closer to the ideal value of 1) and sharpness (π increases) show that the high flows are represented better by the model than the overall flow response (Table 5). By inverting the estimated sharpness (see section A2), average relative prediction errors are 6.3% in winter and 17.9% in summer when all

Table 5. Model Evaluation Statistics for Winter and Summer Periods

Period	NS^a	NS^*_{prior}	NS^*_I	NS^*_{Φ}	$NS^*_{I,\Phi}$	NS^*	δ^b	π	δ_{50}^c	π_{50}
Winter	0.86	0.54	0.64	0.71	0.72	0.86	0.78	15.8	0.94	24.7
Summer	0.76	0.4	0.54	0.66	0.68	0.72	0.54	5.6	0.91	14.38

^a NS and NS^* are for the posterior predictions.

^bThe ideal value for δ and δ_{50} is 1, and the higher π and π_{50} , the better.

^cSubscript 50 denotes statistics for flows over $50 \text{ m}^3 \text{ s}^{-1}$.

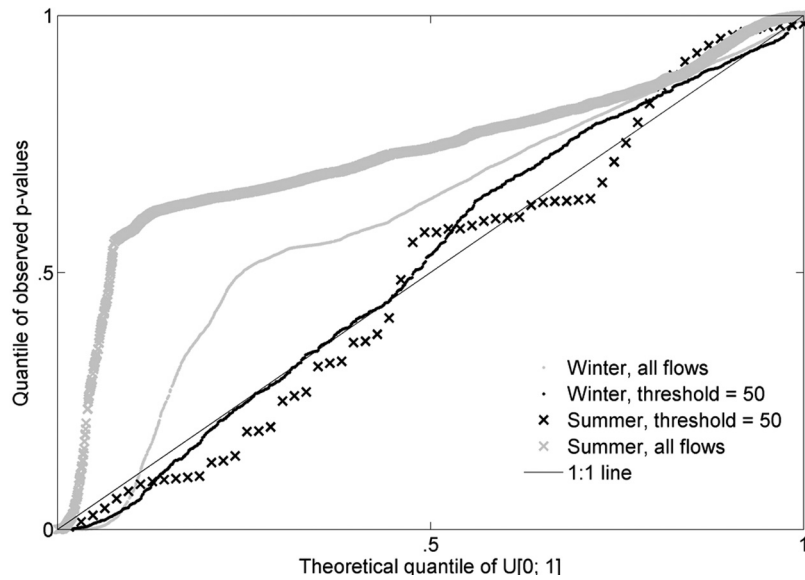


Figure 6. QQ plots for winter and summer evaluation periods when all information sources are used, comparing the cases when predictions are compared with: all observed flows, and only those observed flows above $50 \text{ m}^3 \text{ s}^{-1}$.

flows are used; or as 4% in winter, and 7% in summer when only high flows are considered. Furthermore, the QQ plots allow estimation of the percentage of observations falling within the 90% confidence intervals (such that 5% of observations fall below and 5% fall above the intervals). In Figure 6 this percentage is the x axis value that corresponds to $y = 0.05$ deducted from the x axis value that corresponds to $y = 0.95$, multiplied by 100. Hence, Figure 6 shows that the 90% confidence intervals cover 88% of all observed summer flows, 86% of all observed winter flows, 81% of the high summer flows, and 87% of the high winter flows.

4.3. Change in Flood Frequency Curve Under Scenarios of Land Use Change

[59] From the four considered scenarios, Table 6 shows that only two scenarios lead to changes in median flow peaks that the Wilcoxon tests deems significant at the 0.1 level: the partial afforestation with coniferous trees (scenario 1) for 1, 2, 5, and 10 year return period peak flows, and riparian deciduous tree planting (scenario 4) for 1 and 2 year return period peak flows. Figure 7 shows cumulative distributions of change for 1 year return period flow.

5. Discussion and Conclusions

[60] Maximizing the value of information within a framework of uncertainty reduction is recognized as a key objective for learning about the environment and has particular relevance to the hydrological challenge of modeling ungauged catchments and the closely related problem of predicting hydrology under land use change [Sivapalan, 2003; Beven, 2007; Wagener and Montanari, 2011]. Bayes’ method provides a framework for combining different sources of information into model estimation, and for evaluating the information in terms of how it contributes to model performance improvements and uncertainty reduction. The Bayes’ equation was applied in this paper to

develop a method for formally assimilating three different types of information—regionalized flow indices, small scale physics-based knowledge, and hydrological measurements—for identification and uncertainty analysis of conceptual rainfall-runoff models. In particular, the paper addresses the problem of prediction in ungauged catchments where measurements of flow response are scarce or nonexistent. While the use of physics-based constraints, regionalized data, and/or flow measurements in conceptual model development and parameter estimation is quite common, the formal combination of these information sources into one likelihood function, equation (7), is new.

Table 6. Relative T -Year Return Period Peak Flow Change (%) for Sampling Frequency $\lambda = 3$: Median, 90% Confidence Interval Lower and Upper Bounds^a

	1 yr	2 yr	5 yr	10 yr
<i>Scenario 1</i>				
Median	-3.0	-2.6	-2.6	-2.6
Lower bound	-5.6	-5.4	-5.0	-5.1
Upper bound	2.4	2.5	2.7	2.9
<i>Scenario 2</i>				
Median	-0.8	-0.9	-0.7	-0.2
Lower bound	-6.3	-5.9	-5.7	-5.9
Upper bound	2.4	2.4	1.9	1.9
<i>Scenario 3</i>				
Median	0.0	0.0	0.0	0.0
Lower bound	-2.0	-2.0	-1.9	-1.9
Upper bound	3.2	3.1	3.3	3.0
<i>Scenario 4</i>				
Median	-1.7	-1.5	-1.2	-1.1
Lower bound	-2.4	-2.3	-1.9	-1.9
Upper bound	-0.8	-0.6	-0.7	-0.2

^aThe bold numbers correspond to significant changes in medians at the 0.1 confidence level.

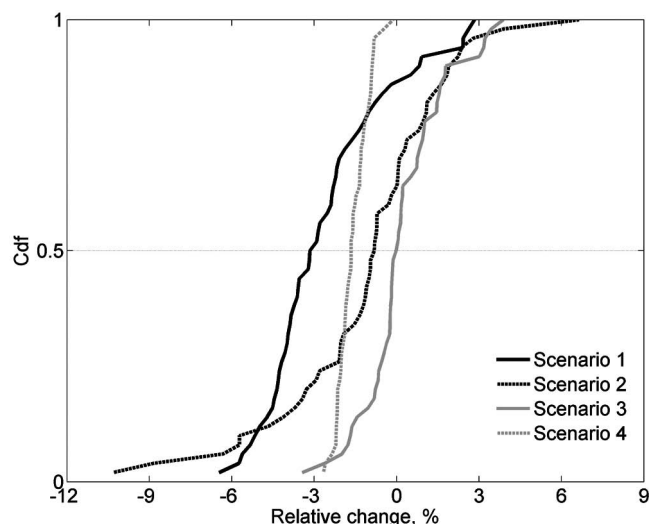


Figure 7. Relative percentage change in 1 year return period flow magnitude for the four scenarios of land use change considered.

[61] The case study used to demonstrate the strengths and limitations of the method was an investigation of land use change effects on the flood frequency curve of the Hodder catchment in northwest England. The application of this method to land use scenario analysis arguably holds significant advantages over other prediction methods: it accounts for natural variability in physical properties, and allows for explicit spatial positioning of land use interventions, thus improving on existing deterministic [Niehoff *et al.*, 2002] and spatially lumped [Hundeche and Bardossy, 2004] studies. Furthermore, the method mainly relies on regionalized knowledge about laboratory-scale physical properties and larger scale flow signatures, therefore reducing the data requirements of some previous land use change impact studies [Brath *et al.*, 2006; Eckhardt *et al.*, 2003; Jackson *et al.*, 2008; McIntyre and Marshall, 2010].

[62] The spatially distributed nature of the relevant changes called for a distributed model consisting of two parts: a spatially distributed grid of local runoff generating “response unit” models, and a channel routing model. Each response unit was classified according to combinations of soil type (HOST type), land use, and soil condition, and a prior model parameter space was specified for each class, hence allowing scenarios of future land use change in each grid square to be represented by changing its response unit class. The channel routing model was assumed to be stationary and represented by a simple lag and route model with constant parameters. The conceptual model parameter space was constrained using regionalized indices—Base Flow Index and Curve Number; and information derived from physics-based modeling—the peak flow rates and corresponding times of peaks of the largest events. The flow routing celerity was constrained based on a few observations of peak flow times and does not require flow magnitudes per se.

[63] The proposed parameter estimation strategy, as applied to the case study, is built on a series of eight assumptions (listed in section 3) that deserve some discussion.

[64] 1. The chosen model structure is assumed to be correct, so that model identification is reduced to model

parameter identification, although developing equation (2) and the subsequent theory to allow for alternative model structures would be straightforward.

[65] 2. Due to the lack of information about how Curve Number and how physical properties change (increase or decrease, and by how much), the post-change response unit parameters were assumed not to depend on the prechange parameters [for a discussion, see Bulygina *et al.*, 2011]. This is likely to result in overestimates of the uncertainty (since less information is employed in post-land use change conditioning) when considering the difference in peak flow going from pre- to post-land use change. A simpler option is to eliminate all pairs of realizations that give results that are opposite to expectations [Bulygina *et al.*, 2011], however this puts a lot of weight on prior expectations.

[66] 3. All three sources of information—**F**, **I**, and **D**—are considered to be conditionally independent, which is reflected in the likelihood functions. In the case study this is considered to be a reasonable assumption because the three sources did not rely on any common model, data, or prior perception. If they had, then in principle the likelihood function should be amended to reflect the bi- or tri-variate distribution of errors. However, estimating the a priori dependency between two significantly different types of information in practice would be difficult, and is not generally attempted in ungauged catchment problems.

[67] 4. Available large-scale observations are presumed not to provide any information that would significantly constrain the response unit parameters. Given the limited amount of observations typically available, the smoothing effect of the catchment routing, and the more local information provided by the other two sources, this is arguably reasonable. Furthermore, conditioning the response unit models on the catchment scale flow data would require the full distributed model to be run a large number of times (as opposed to conditioning on the other two sources of information, which required a large number of simulations only of each of the response unit types). However, in cases when flow observations representing different types of response units are available [e.g., Jackson *et al.*, 2008], their information content could be used in the response unit parameter restriction.

[68] 5. Likelihoods are defined by simple distribution functions (normal and uniform) assuming that information sources are unbiased. Because there was significant prior evidence to support the form of likelihood functions for **I** and **F** (the regressive nature of BFI_{HOST} and CN_{USDA} , and the exploration by Ballard [2011] into the relationship between the physics-based and conceptual model outputs), the paper focused on assessing the value of this information for achieving satisfactory model performance at the catchment scale rather than the problems of inference about bias in the model structure or its inputs, although there is clearly potential to do so.

[69] 6. The applicability to the UK of the Curve Number system, which was developed for use in the United States, is presumed after Bulygina *et al.* [2011] and Hess *et al.* [2010]. This has only been tested to a limited extent here and by Bulygina *et al.* [2011] and Holman *et al.* [2011] by showing that its inclusion slightly improves model performance, and more work on this is recommended.

[70] 7. Land use change is assumed not to cause significant changes to channel routing during flood flows, and it is

assumed that any change is captured by existing uncertainty in the routing parameters. This is not necessarily valid, especially if considering more extreme events than we have done here. For example, any increases in flood flows associated with land use change would increase the potential for channel erosion. While it would be interesting, addressing this assumption would require hydraulic and erosion models to be included.

[71] 8. It is possible that the multiple sources of information may be in conflict—in other words the model structure may be incapable of being consistent with all of them—either because of model structural error or because the errors in one or more of the information sources are not sufficiently represented by the associated error model. The conditioning could be approached in the manner of *Gupta et al.* [1998], which recognizes that lumping alternative sources of information into one likelihood is not necessarily helpful and instead treats model identification as a multiple objective problem. Alternatively, a formal method for treating information conflicts [e.g., *Fu and Kaplan*, 2011] could be applied.

[72] Some of these assumptions are considered necessary due to the lack of information, i.e., the independence of parameters between different response unit types, independence of information sources, and keeping channel routing parameters the same after land use interventions. Other assumptions are motivated largely by computational tractability, i.e., ignoring the influence of large scale flow observations on the response unit parameters. And finally, the adequacy of the two assumptions, that the conceptual model structure and the unvalidated likelihood functions are correct, could have been addressed within the Bayesian approach [*Thyer et al.*, 2009], but were excluded to keep the scope of the paper within reasonable bounds.

[73] Regarding the results obtained for the Hodder conceptual model, when posterior marginal parameter distributions are considered, the maximum drainable depth (h), and the medium and slow residence times (K_m and K_s) are found to be the least identifiable of the eight parameters (Figure 4), indicating a need for additional sources of information to further constrain the parameter space. The physics-based information restricts the response unit parameters the most (Figure 4). The model, conditioned on all three sources of information, provides predictions for winter and summer evaluation periods that are considered satisfactory, with probabilistic analog Nash-Sutcliffe efficiencies of 0.86 and 0.72 correspondingly. For comparison, the analog Nash-Sutcliffe efficiencies obtained from the prior parameter ranges were 0.54 and 0.4 correspondingly. The physics-based information improves the analog Nash-Sutcliffe efficiency the most, followed by the observations of time to peak, and the regionalized information (Table 5). Due to the constant celerity routing scheme used, the modeled response is too flashy at low flows, but captures the high flows well (Figure 4, Table 5).

[74] To illustrate the practical value of the method for predicting effects of land use change, four scenarios of localized land use changes are considered: coniferous afforestation, stocking density reduction, grazing intensification, and riparian deciduous tree planting (Table 3). Changes in 15 min T -year return period peak flows relative to the baseline condition are tested for significance given

the model uncertainty. The two tree planting scenarios are found to have statistically significant effects on flood peaks with median peak flow reduction between 2.6% and 3% for the coniferous afforestation, and between 1.5% and 1.7% for the riparian deciduous tree planting (Table 6). Although, according to the medians, the tree planting leads to flood flow reduction, the 90% confidence intervals (Table 6) show that increases in flow peaks might occur as well. This is partly due to the assumed independence of the pre- and post-change parameter sets, as previously explained, exacerbated by the fact that uncertainties in response can be high compared to the expected differences in response between unit types: the range in physical properties within one response unit type is considerable and hence there is overlap in properties between unit types; and the Curve Numbers do not vary widely between land uses under the same soil type and have high variance. Generally, the significance of the land use change effect decreases with increasing peak flow magnitude, i.e., increasing return period (Table 6). The achieved level of high flow prediction accuracy (4%–7% average prediction error during the evaluation period) meant that changes in flood flows under the scenarios of stocking density reduction and grazing intensification were not statistically significant. Although these results are site specific, they have practical significance for rural land use management in the UK. The results add to the growing evidence [*Bulygina et al.*, 2009, 2011; *Ewen et al.*, 2010; *O'Donnell et al.*, 2011] that relatively minor modifications to a rural landscape, which are likely to be considered practical in terms of the scale of change, have limited although positive effects on downstream flood risk.

[75] To summarize, the proposed method of combining multiple sources of information in a Bayesian framework improved flow prediction accuracy, reduced uncertainty, and, within the limitations of the case study application, highlighted the value of the different information types. The method was applied to the Hodder catchment in northwest England to upscale localized land use change effects to the catchment scale and to detect significant effects given the prediction uncertainty. It is expected that the application of the method would be revised for every specific application, to adapt to the information available and the type of information most relevant to the modeling task. To the best of our knowledge, this paper is the first attempt to formally combine the three distinct types of information—regionalized flow indices, physics-based knowledge via metamodeling, and flow measurements—using a formal Bayesian approach.

Appendix A: Performance Measures

A1. An Analog Nash-Sutcliffe Efficiency

[76] *Bulygina et al.* [2009] describe a probabilistic analog Nash-Sutcliffe efficiency NS^* for probabilistic predictions given by a sequence of random variables $\{X_t\}$ as

$$NS^* = \left\{ 1 - \frac{\sum_{t=1}^T (E[X_t] - x_t^0)^2}{\sum_{t=1}^T (x_t^0 - x^0)^2} \right\} - \frac{\sum_{t=1}^T Var[X_t]}{\sum_{t=1}^T (x_t^0 - x^0)^2}, \quad (A1)$$

where x_t^0 is an observed value at time t , x^0 is an average value for the observed data series, $Var[\cdot]$ denotes variance, $E[\cdot]$ denotes mathematical expectation, and T is the number of time steps in the sequence. In the current context, $\{X_t\}$ is the simulated time series of flow and x_t^0 is the time series of observed flow.

[77] The first part on the right-hand side of (A1) resembles the traditional Nash-Sutcliffe efficiency, when predictions at each time are characterized by mathematical expectations. And the second part reduces the traditional value according to spread in the predictions, so that the higher the prediction precision the closer the analog measure to the traditional one.

A2. QQ Plots, Reliability, and Sharpness Indices

[78] A probabilistic forecast of some quantity x_t is *reliable* if its predictive cumulative distribution F_t adequately approximates the true cumulative distribution of x_t . If the observations are consistent with F_t , the p -values $F(x_t)$ follow a uniform distribution on the interval $[0; 1]$, $U[0; 1]$ [Laio and Tamea, 2007]. This can be examined using a QQ plot that plots theoretical quantiles of $U[0; 1]$ versus quantiles of observed p -values. Deviations from the 1:1 line can be used to characterize model deficiencies [Laio and Tamea, 2007]:

[79] 1. If the graph crosses the 1:1 line then the prediction uncertainty is underpredicted.

[80] 2. If the graph is convex (u-shaped) then the true values are overpredicted.

[81] 3. If the graph is concave then the true values are underpredicted.

[82] Here, following Bulygina and Gupta [2011], we use an index adapted from Renard et al. [2010] to quantify reliability of the probability forecast:

$$\delta = 1 - 2 \cdot \frac{1}{T} \sum_{i=1}^T |p_{x(i)} - p_{x(i)}^*|, \quad (\text{A2})$$

where $p_{x(i)}$ and $p_{x(i)}^*$ are the i th observed (after reordering) and theoretical p -values of $x(i)$, T is the number of observations. The index δ characterizes the area between the p -value curve and 1:1 line, varying between 0 (not reliable) and 1 (perfectly reliable).

[83] The *resolution* is evaluated as an average relative precision of the predictions X_t [Renard et al., 2010]

$$\pi = \frac{1}{T} \sum_{t=1}^T \frac{E[X_t]}{\sqrt{Var[X_t]}}, \quad (\text{A3})$$

where $E[X_t]$ and $Var[X_t]$ are expected value and variance of the prediction X_t .

[84] **Acknowledgments.** This work was funded by the Natural Environment Research Council program ‘Flood Risk from Extreme Events’ (FREE; NE/F001134/1), and the Engineering and Physical Sciences Research Council program ‘Flood Risk Management Research Consortium’ (FRMRC Phase 2; EP/F020511/1).

References

Allen, R. G. (2006), Evaporation modeling: Potential, in *Encyclopedia of Hydrology Science*, 11 pp., John Wiley, New York.
Ballard, C. (2011), The role of physics based models for simulating flood flow response to rural land management scenarios, Ph.D. thesis, Imperial

College London, London, 324 pp.; [https://workspace.imperial.ac.uk/ewre/Public/BALLARD_Thesis_FINAL_2011\[1\].pdf](https://workspace.imperial.ac.uk/ewre/Public/BALLARD_Thesis_FINAL_2011[1].pdf).
Ballard, C., N. McIntyre, and H. Wheeler (2010), Peatland drain blocking—Can it reduce flood risk?, in *Proceedings of the BHS International Hydrology Symposium*, pp. 697–702, Newcastle Univ., Newcastle upon Tyne, U.K., 19–23 July.
Ballard, C., N. McIntyre, H. Wheeler, J. Holden, and Z. Wallage (2011), Hydrological modelling of drained blanket peat, *J. Hydrol.*, 407(1–4), 81–93.
Barton, R. R. (1998), Simulation metamodelling, in *Proceedings of the 30th Conference on Winter Simulation*, pp. 167–176, IEEE Computer Society, Washington, D.C.
Begueria, S. (2005), Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value, *J. Hydrol.*, 303, 215–230.
Beven, K. (1981), Kinematic subsurface stormflow, *Water Resour. Res.*, 17(5), 1419–1424.
Beven, K. (2007), Towards integrated environmental models of everywhere: Uncertainty, data and modelling as a learning process, *Hydrol. Earth Sys. Sci.*, 11(1), 460–467.
Boorman, D., J. Hollis, and A. Lilly (1995), *Hydrology of Soil Types: A Hydrologically-Based Classification of the Soils of the United Kingdom*, Institute of Hydrology, Wallingford, UK.
Box, G., and F. Tiao (1992), *Bayesian Inference in Statistical Analysis*, Wiley, New York.
Brath, A., A. Montanari, and G. Moretti (2006), Assessing the effects on flood frequency of land use change via hydrological simulation (with uncertainty), *J. Hydrol.*, 324, 141–153.
Bulygina, N., and H. Gupta (2011), Correcting the mathematical structure of a hydrological model via Bayesian data assimilation, *Water Resour. Res.*, 47, W05514, doi:10.1029/2010WR009614.
Bulygina, N., N. McIntyre, and H. Wheeler (2009), Conditioning rainfall-runoff model parameters for ungaged catchments and land management impacts analysis, *Hydrol. Earth Syst. Sci.*, 13(6), 893–904.
Bulygina, N., N. McIntyre, and H. Wheeler (2011), Bayesian conditioning of a rainfall-runoff model for predicting flows in ungaged catchments and under land use changes, *Water Resour. Res.*, 47, W02503, doi:10.1029/2010WR009240.
Camacho, L., and M. Lees (2000), Multilinear discrete lag cascade model for channel routing, *J. Hydrol.*, 266, 20–47.
Cunnane, C. (1979), A particular comparison of annual maxima and partial duration series methods of flood frequency prediction, *J. Hydrol.*, 18(3/4), 257–271.
Dooge, J. C. I., and J. P. O’Kane (2003), *Deterministic Methods in Systems Hydrology*, 309 pp., A. A. Balkema, Lisse, Netherlands.
Doucet, A., S. Godsill, and C. Andrieu (2000), On sequential Monte Carlo methods for Bayesian filtering, *Stat. Comput.*, 10, 197–208.
Eckhardt, K., L. Breuer, and H.-G. Frede (2003), Parameter uncertainty and the significance of simulated land use change effects, *J. Hydrol.*, 273, 164–176.
Evans, J., and A. Jakeman (1998), Development of a simple catchment-scale, rainfall-evapotranspiration-runoff model, *Environ. Model. Software*, 13, 385–393.
Ewen, J. (1996), ‘SAMP’ model for water and solute movement in unsaturated porous media involving thermodynamic subsystems and moving packets: I. Theory, *J. Hydrol.*, 182(1–4), 175–194.
Ewen, J., J. Geris, G. O’Donnell, W. Mayes, and E. O’Connell (2010), Multiscale experimentation, monitoring and analysis of long-term land use changes and flood risk - SC060092: Final Science Report, Newcastle University, Newcastle upon Tyne, UK.
Forsman, Å., and A. Grimvall (2003), Reduced models for efficient simulation of spatially integrated outputs of one-dimensional substance transport models, *Environ. Model. Software*, 18(4), 319–327.
Fu, G., and Z. Kapelan (2011), Fuzzy probabilistic design of water distribution networks, *Water Resour. Res.*, 47, W05538, doi:10.1029/2010WR009739.
Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, 34(4), 751–763.
Hess, T. M., I. P. Holman, S. C. Rose, Z. Rosolova, and A. Parrott (2010), Estimating the impact of rural land management changes and catchment runoff generation in England and Wales, *Hydrol. Processes*, 24(10), 1357–1368, doi:10.1002/hyp.7598.
Hillel, D. (1971), *Soil and water: Physical principles and processes*, in *Physiological Ecology*, 2, 288 pp., Academic, New York.

- Holman, I. P., T. M. Hess, and S. C. Rose (2011), A broad-scale assessment of the effect of improved soil management on catchment Baseflow Index, *Hydrol. Processes*, 25(16), 2563–2572, doi:10.1002/hyp.8131.
- Hosking, J., and J. Wallis (1987), Parameter and quantile estimation for the Generalised Pareto distribution, *Technometrics*, 29(3), 339–349.
- Hough, M., and R. Jones (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.
- Hundecha, Y., and A. Bardossy (2004), Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *J. Hydrol.*, 292, 281–295.
- Jackson, B. M., H. S. Wheeler, N. R. McIntyre, J. Chell, O. J. Francis, Z. Frogbrook, M. Marshall, B. Reynolds, and I. Solloway (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.
- Kapangaziwiri, E., D. A. Hughes, and T. Wagener (2009), Towards the development of a consistent uncertainty framework for hydrological predictions in South Africa, New approaches to hydrological predictions in data-sparse regions, *IAHS Publ.*, 333, 83–94.
- Kavetski, D. and F. Fenicia (2011), Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, *Water Resour. Res.*, 47, W11511, doi:10.1029/2011WR010748.
- Koren, V., S. Reed, M. Smith, Z. Zhang, and D.-J. Seo (2004), Hydrology laboratory research modeling system (HL-RMS) of the US national weather service, *J. Hydrol.*, 291, 297–318.
- Kuczera, G., and M. Mroczkowski (1998), Assessment of hydrologic parameter uncertainty and the worth of multiresponse data, *Water Resour. Res.*, 34(6), 1481–1489, doi:10.1029/98WR00496.
- Laio, F., and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrological variables, *Hydrol. Earth Syst. Sci.*, 11(4), 1267–1277.
- Lang, M., T. Ouada, and B. Bobee (1999), Towards operational guidelines for over-threshold modelling, *J. Hydrol.*, 225, 103–117.
- Maidment, D., F. Olivera, A. Calver, A. Eatherall, and W. Franczek (1996), Unit hydrograph derived from a spatially distributed velocity field, *Hydrol. Processes*, 10, 831–844.
- McIntyre, N., and M. Marshall (2010), Identification of rural land management signals in runoff response, *Hydrol. Processes*, 24(24), 3521–3534.
- McIntyre, N., B. Jackson, A. J. Wade, D. Butterfield, H. S. Wheeler (2005a), Sensitivity analysis of a catchment-scale nitrogen model, *J. Hydrol.*, 315, 71–92.
- McIntyre, N., H. Lee, H. Wheeler, A. Young, and T. Wagener (2005b), Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, 41, W12434, doi:10.1029/2005WR004289.
- Miller, K. F., Quine, C. P., and J. Hunt (1987), The assessment of wind exposure for forestry in upland Britain, *Forestry*, 60, 179–192.
- Naden, P. (1992), Analysis and use of peaks-over-threshold data in flood estimation, in *Flood and Flood Management*, edited by A. Saul, pp. 131–143, Kluwer Academic, Dordrecht.
- Niehoff, D., U. Fritsch, and A. Bronstert (2002), Land-use impacts on storm-runoff generation: scenarios of land-use change and simulation of hydrological response in a meso-scale catchment in SW-Germany, *J. Hydrol.*, 267, 80–93.
- NSRI (2011), LandIS Digital Soil Datasets, <http://www.landis.org.uk/data/index.cfm>. National Soils Research Institute, Cranfield.
- O'Connell, P. E. et al. (2004), *Review of Impacts of Rural Land Use and Management on Flood Generation*, pp. 1–146, DEFRA, London.
- O'Donnell, G., J. Ewen, and P. E. O'Connell (2011), Sensitivity maps for impacts of land management on an extreme flood in the Hodder catchment, UK, *Phys. Chem. Earth*, 36(13), 630–637.
- O'Hagan, A., M. C. Kennedy, and J. E. Oakley (1999), Uncertainty analysis and other inference tools for complex computer codes, *Bayesian Stat.*, 6, 503–524.
- Parrott, A., W. Brooks, O. Harmar, and K. Pygott (2009), Role of rural land use management in flood and coastal risk management, *J. Flood Risk Manage.*, 2, 272–284.
- Piñeros Garcet, J. D., A. Ordoñez, J. Roosen, and M. Vanclouster (2006), Metamodelling: Theory, concepts and application to nitrate leaching modelling, *Ecol. Model.*, 193(3–4), 629–644.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328.
- Seibert, J., and K. Beven (2009), Gauging the ungauged basin: How many-discharge measurements are needed?, *Hydrol. Earth Syst. Sci.*, 13, 883–892.
- Siegel, S. (1956), *Non-Parametric Statistics for Behavioral Sciences*, pp. 75–83, McGraw-Hill, New York.
- Singh, V. P. (1996), Kinematic wave modeling in water resources, in *Surface-Water Hydrology*, 1399 pp., John Wiley, Chichester, UK.
- Sivapalan, M. (2003), Prediction in ungauged basins: A grand challenge for theoretical hydrology, *Hydrol. Processes*, 177(15), 3163–3170.
- Sivapalan, M., M. A. Yaeger, C. J. Harman, X. Xu, and P. A. Troch (2011), Functional model of water balance variability at the catchment scale: 1. Evidence of hydrologic similarity and space-time symmetry, *Water Resour. Res.*, 47, W02522, doi:10.1029/2010WR009568.
- Sorooshian, S. and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resour. Res.*, 16(2), 430–442.
- Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Recursive Bayesian inference for hydrologic models, *Water Resour. Res.*, 37(10), 2521–2535.
- Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, 45, W00B14, doi:10.1029/2008WR006825.
- USDA (1986), Urban hydrology for small watersheds, *United States Department of Agriculture, TR 55*, 1–164.
- Valente, F., J. David, and J. Gash (1997), Modelling interception loss for two sparse eucalypt and pine forests in central Portugal using reformulated Rutter and Gash analytical models, *J. Hydrol.*, 190(1–2), 141–162.
- Wagener, T. (2007), Can we model the hydrological impacts of environmental change? *Hydrol. Processes*, 21, 3233–3236.
- Wagener, T., and N. McIntyre (2005), Identification of rainfall-runoff models for operational applications, *Hydrol. Sci. J.*, 50, 735–751.
- Wagener, T., and A. Montanari (2011), Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, *Water Resour. Res.*, 47, W06301, doi:10.1029/2010WR009469.
- Wagener, T., and H. Wheeler (2006), Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, *J. Hydrol.*, 320(1–2), 132–154.
- Wagener, T., H. Wheeler, and H. Gupta (2004), *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments*, Imperial College Press, London.
- Wheater, H. S., A. J. Jakeman, and K. J. Beven (1993), Progress and directions in rainfall-runoff modelling, in *Modelling Change in Environmental Systems*, edited by A. J. Jakeman, M. B. Beck, and M. J. McAleer, pp. 101–132, Wiley, New York.
- Wilcoxon, F. (1945), Individual comparison by ranking methods, *Biometrics Bull.*, 1(6), 80–83.
- Wu, W., J. S. Clark, and J. M. Vose (2010), Assimilating multi-source uncertainties of a parsimonious conceptual hydrological model using hierarchical Bayesian modeling, *J. Hydrol.*, 394(3–4), 436–446.
- Yadav, M., T. Wagener, and H. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. Water Resour.*, 30(8), 1756–1774.