The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments

Satish Bastola *, Conor Murphy, John Sweeney

ICARUS, Department of Geography, National University of Ireland Maynooth (NUIM), Co. Kildare, Ireland

ABSTRACT

This study attempts to assess the uncertainty in the hydrological impacts of climate change using a multi-model approach combining multiple emission scenarios, GCMs and conceptual rainfall–runoff models to quantify uncertainty in future impacts at the catchment scale. The uncertainties associated with hydrological models have traditionally been given less attention in impact assessments until relatively recently. In order to examine the role of hydrological model uncertainty (parameter and structural uncertainty) in climate change impact studies a multi-model approach based on the Generalised Likelihood Uncertainty Estimation (GLUE) and Bayesian Model Averaging (BMA) methods is presented. Six sets of regionalised climate scenarios derived from three GCMs, two emission scenarios, and four conceptual hydrological models were used within the GLUE framework to define the uncertainty envelop for future estimates of stream flow, while the GLUE output is also post processed using BMA, where the probability density function from each model at any given time is modelled by a gamma distribution with heteroscedastic variance. The investigation on four Irish catchments shows that the role of hydrological model uncertainty is remarkably high and should therefore be routinely considered in impact studies. Although, the GLUE and BMA approaches used here differ fundamentally in their underlying philosophy and representation of error, both methods show comparable performance in terms of ensemble spread and predictive coverage. Moreover, the median prediction for future stream flow shows progressive increases of winter discharge and progressive decreases in summer discharge over the coming century.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Conceptual rainfall–runoff (CRR) models forced with regional climate change scenarios downscaled from Global Climate Models (GCMs) are widely employed to assess the impacts of climate change at the catchment scale. This approach is subject to a range of uncertainties associated with future emissions of greenhouse gases, the response of the climate system to these changes at global and local scales, and uncertainties associated with impact models. These uncertainties then cascade through the climate change impact assessment methodology with potentially large uncertainties associated with critical future impacts at the local scale where key decisions are required in order to increase the resilience of water supply management and infrastructure to future changes. Given that uncertainty in modelling will not be significantly reduced in the short or medium term future, ensuring that potentially expensive and irreversible adaptation decisions made now are robust to the uncertainty in future climate change impacts means that considerable effort is required in investigating and quantifying sources of uncertainty.

Outputs from GCMs are based upon the fundamental laws of physics embodied within models and assumptions on the concentration of greenhouse gases in the atmosphere. As these GCMs differ in the way they simplify the climate system, and aggregate the process in space and time, future projections of water resources are dependent upon the choice of GCMs employed [36]. Utilisation of information from different models has been widely used to address these uncertainties. Giorgi and Mearns [21] introduced the Reliability Ensemble Averaging (REA) method for calculating uncertainty ranges from ensembles of different Atmosphere–Ocean General Circulation Models (AOGCMs). Similarly, Tebaldi et al. [46] extended the REA method and proposed a Bayesian statistical model that combines information from a multi-model ensemble of AOGCMs and observations to determine probability distributions of future temperature change on a regional scale. Several studies have used the output archived in Coupled Model Inter-comparison Projects to account for uncertainty in GCMs (e.g., [40]), while several others have used the output from perturbed physics ensembles to evaluate the uncertainties arising from GCM model formulation (e.g., [32]).

Output from GCMs reproduce the global and continental scale climate fairly well, however, they are inadequate in impact studies due to the differences in the spatial scale of the GCM and the

* Corresponding author. Tel.: +353 1 7086592.
 E-mail address: Satish.Bastola@nuim.ie (S. Bastola).

output needed for impact studies [54]. This limitation has been widely addressed through the use of regionalisation techniques to downscale large scale simulations from GCMs. In the last decade a number of methods have been employed, particularly empirical statistical downscaling and the deployment of Regional Climate Models (RCMs), with techniques differing in the way they reproduce various statistical characteristics of observed data [54].

In an attempt to quantify major sources of uncertainties associated with climate change impact assessment, New and Hulme [35] presented an approach to quantifying uncertainties associated with the estimation of future greenhouse gas emissions, the climate sensitivity, and limitations and unpredictability in GCMs. Similarly, Horton et al. [23] analysed the uncertainty induced by the use of different state of the art climate models on the prediction of climate-change impacts on the runoff regimes of 11 mountainous catchments in the swiss Alps.

However, most of the studies utilised a single hydrological model and ignored the modelling uncertainties associated with the structure of such models. Hydrological models are inherently imperfect because they abstract and simplify real patterns and processes that are themselves imperfectly known and understood. Experience with the calibration of hydrological models suggests that their parameters are inherently uncertain. Though many studies have addressed the issues of parameter uncertainty, very few have looked at the uncertainties related to model structure, particularly in the context of climate change assessments.

Since the role of uncertainties derived from hydrological modelling in impact assessment has received much less attention, this study attempts to identify the role of the selection and parameterisation of hydrological models on the overall uncertainty envelop involved in evaluating the impact of climate change on water resources at the catchment scale. The paper is structured as follows: Section 2 considers the sources of uncertainties in rainfall–runoff modelling and techniques employed to quantify prediction uncertainties. Section 3 provides an overview of the study basins and climate scenarios, the hydrological models used and methods employed to account for the different uncertainties that are associated with studying the impact of climate change on water resources. Results are outlined in Section 4.

2. Uncertainties in CRR models

Despite their acknowledge limitations, CRR models continue to be widely used for assessing the impacts of climate change on water resources and for projecting potential ranges of impacts from scenarios of future change. CRR models use relatively simple mathematical equations to conceptualise and aggregate the complex, spatially distributed, and highly interrelated water, energy, and vegetation processes in a watershed. Due to the randomness in nature and the lack of complete knowledge of the hydrological system, uncertainty is an unavoidable element in any hydrologic modelling study [8,22]. In hydrological modelling, uncertainty stems from a variety of sources such as; data uncertainty, parameter uncertainty, model structural uncertainty and state uncertainty.

An extensive review of the causes of uncertainty in hydrological modelling and various methods for assessing the uncertainty can be found in Melching [27]. The climate change/hydrological modelling literature has mostly focused on the prediction uncertainty arising from model parameters (e.g., [25,41]), despite the fact that uncertainties resulting from dependence on a single conceptual–mathematical model are typically much larger than those introduced through the inadequate choice of model parameter values (e.g., [14]). Larger differences in the model results are likely to occur when different model structures are used to simulate the

hydrological impact of the postulated climate changes thereby increasing the uncertainty of the future discharge prediction considerably (e.g., [24]).

To examine the impact of model structure error and complexity on model performance and modelling uncertainty, Butts et al. [12] used multi-model ensembles for the distributed hydrological model inter comparison study watersheds. Their work suggests the importance of considering uncertainty for streamflow forecasting and the utility of multiple model ensembles for considering model parametric and structural uncertainty. A review of the range of strategies for assessing structural uncertainties in environmental modelling is available in Refsgaard et al. [39]. These strategies can be broadly grouped into two depending upon weather or not target data is available. In the application of hydrological models in climate change impact assessment, the structure of the hydrological model cannot be assessed directly using observations. Therefore, the main strategy to account for modelling uncertainties is to extrapolate future conditions with multiple conceptual models.

2.1. Methods for assessing uncertainty

Among various methods for assessing the uncertainty of hydrological models, the Generalised Likelihood Uncertainty Estimation (GLUE) method developed by Beven and Binley [5] has been extensively used (e.g., [20]). The GLUE method is based on the premise that for a physically based hydrological model, no single optimum parameter set exists; rather a range of different sets of model parameter values may represent the process equally well. Different model structures, as well as different parameter sets in a particular model structure, can be easily combined within this framework. The technique is based on Monte Carlo simulation where a model is run a large number of times with different parameter sets. In GLUE, it is assumed that the error associated with a particular model (parameter set) will be similar in prediction to those found in calibration. More details on GLUE can be found in [5,20,30]. The major output of the GLUE method for assessing uncertainty is the prediction interval at each time step bounded by the lower and upper prediction limit. To examine the capability of the prediction intervals to capture the observed values, an index defined as the ratio of the number of the observations falling within their respective prediction intervals to the total number of observations is normally used (e.g., [30]). If prediction bounds are large enough to include most of the observations, it means that parameter variability alone can compensate for other sources of error, such as measurement and model structure errors and thus it can account for the total output uncertainty. The performance of median values (Q50) is also usually judged using the Nash Sutcliffe criterion. Furthermore, an average prediction interval defined by the average prediction bounds of a particular confidence level can be used as a measure to reflect the uncertainties in the modelling process.

Bayesian Model Averaging (BMA) is a standard statistical post processing tool. It can be used to account for model uncertainty by combining predictive distributions from different sources [38]. The application of BMA is growing in a multi-model ensemble of AOGCMs to produce mean and probabilistic climate change projections (e.g., [46,28]). In BMA the predictive probability density function (PDF) of any quantity of interest is a weighted average of PDFs centred on the individual forecasts, where the weights are equal to posterior probabilities of the models generating the forecasts and reflect the models relative contributions to predictive skill over the training period. The BMA weights can be used to assess the usefulness of ensemble members, and this can be used as a basis for selecting ensemble members for prediction. Ajami et al. [1] explored the use of the BMA scheme to develop more skilful and reliable probabilistic hydrologic predictions from multiple competing

predictions made by several hydrologic models. Ajami et al. [1] showed that the BMA scheme has the advantage of generating more skilful and equally reliable probabilistic predictions than the original ensemble.

3. Methodology

3.1. Study region and data

The area of focus for this study is the Republic of Ireland (Fig. 1). In particular, the impact of climate change on water resources at the catchments scale is investigated using four Irish catchments (see Fig. 1), namely the river Blackwater at Ballyduff (2302 km^2), the river Suck at Bellagill (1219 km^2), the Moy at Rahans (1803 km^2), and the Boyne at Slane (2452 km^2). These four catchments were selected so that they represent the diverse hydrological responses of different catchments located throughout the Republic of Ireland. Table 1 provides an overview of key catchment descriptors.

Six sets of statistically downscaled climate scenarios derived from three GCMs and two emission scenarios, namely A2 and B2, downscaled for Ireland by Fealy and Sweeney [17,18] were used
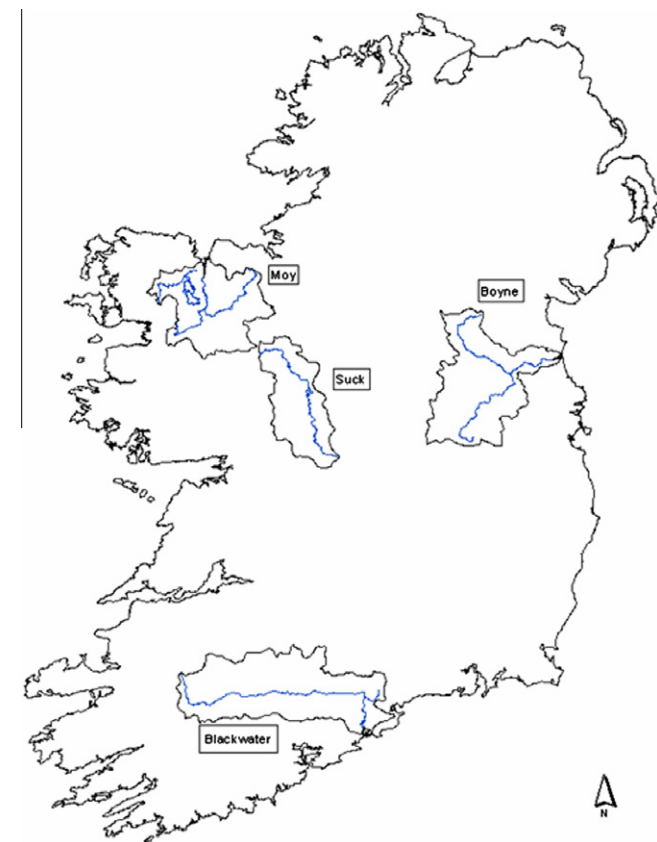
to characterise future climate evolutions. The GCMs considered included: HADCM3 from the Hadley Centre for Climate Prediction and Research (Met Office, UK); CCGCM2, from the Canadian Centre for Climate Modelling and Analysis (CCCMA; Canada) and CSIRO-Mk2 from the Commonwealth Science and Industrial Research Organisation (CSIRO, Australia). The A2 and B2 scenarios represent future emissions levels that could be considered 'medium–high' (A2 emission) and 'medium–low' (B2 emission). Though A2 and B2 encompass most of the range of the Special Report on Emissions Scenarios (SRES), the inclusion of A1F1 and B1, the high and low scenarios would allow a larger proportion of the range of future emissions to be included. A fully probabilistic assessment of future regional climate change and its impacts requires more GCMs and scenarios of radiative forcing. However, they are not readily available, because no climate modelling centre has performed GCM simulations for more than a few emissions scenarios and GCMs for Ireland. Though this limitation can be partly addressed using pattern scaling methodologies, which have been widely used to provide climate change projections for time periods and emission scenarios that have not been simulated by GCMs, the assumption is only weakly valid for precipitation [29], a primary input to hydrological models. Therefore in this study we only utilised the time series of downscaled data derived from three GCMs forced with two scenarios (SRES A2 and B2). The choice of only three GCMs is surely not reflective of the uncertainty in the input to the hydrological models. However, these GCMs were selected based on the availability of data for our region. Moreover, these three GCMs have been used extensively and appear in a range of peer reviewed literature and they represent a sample of the spread in estimated climate sensitivity.

The future potential evapotranspiration used is not a direct output of the GCMs, but is estimated based upon present climate using the Hargreaves method [56], a radiation based empirical model popularly used for the simulation of potential evapotranspiration, for each of the GCMs. Observed stream flow data from the Office of Public Works (available at http://www.opw.i.e./hydro/), and observed precipitation and temperature data from Met Éireann, the Irish National Meteorological Service were used.

3.2. CRR models selected

From among the large number of models that can be used for the purpose of modelling flow in catchments, we selected the following four conceptual rainfall–runoff models:

(a) TOPMODEL; a variable contributing area physically-conceived semi-distributed hydrological model. In TOPMODEL distributed predictions of catchment response are made based on a simple theory of hydrological similarity of points in a catchment. These points of hydrological similarity are identified by an index that is derived from catchment topography. TOPMODEL uses several assumptions to relate, in a simple way, the down slope flow rate at each point and the discharge at the catchment outlet which are as follows: (1) the water table is approximately parallel to the topographic surface; (2) the saturated hydraulic conductivity

Fig. 1. Location of case study catchments.

Table 1
Catchment descriptors.

River and gauge	Area (km^2)	S1085	SAAR (mm)	SAAPE (mm)	Forest (%)	Peat (%)	Pasture (%)	BH soils	Altbar
Boyne at Slane Castle	2452	0.68	890	504	0.04	0.05	0.89	0.69	90.90
Moy at Rahans	1803	0.74	1323	461	0.09	0.32	0.51	0.76	81.20
Suck at Bellagill	1219	0.39	1046	466	0.08	0.21	0.69	0.58	75.60
Blackwater at Ballyduff	2302	1.34	1200	516	0.14	0.05	0.79	0.62	165.60

falls off exponentially with depth; and (3) the water table is recharged at a spatially uniform, steady rate that is slow enough, relative to the response timescale of the watershed, to allow the assumption of a water table distribution that is always at equilibrium. These assumptions permit reconstruction of the spatial variability of catchment response to meteorological forcing solely from modelling of the response of the mean state. This quasi-stochastic approach is at once computationally efficient while still permitting dynamic representations of physical processes within the system. Detailed descriptions of TOPMODEL and its mathematical formulation can be found in Beven [6], and a review of TOP-MODEL applications can be found in Beven [7]. It is referred as TOP hereafter interchangeably. TOPMODEL is best suited to small/medium sized catchments ($\sim$500 km^2) with shallow soils and moderate topography which do not suffer from excessively long dry periods. In this study, the model is applied at a daily time scale and intended to estimate potential simulators (behavioural parameters) through model calibration rather than from measurement or physical reasoning. Therefore, TOPMODEL is applied to basins (<2500 km^2) larger than those in typical applications. Example applications where TOPMODEL or models similar in concept are employed for large catchments include the application from Xiong and Guo [55], and Franchini et al. [19]. Additionally, the BTOPMC model of Takeuchi et al. [43] and TOPNET of Bandaragoda et al. [4] enables the application of the TOPMODEL concept to larger basin (several thousand km^2) by dividing the basin into blocks or a smaller number of sub-basins and routing the flow from each of sub-basin or grid to the outlet (e.g., BTOPMC, TOPNET). In this study, the regional climate change scenarios are only available for a limited number of stations and therefore the information on the internal basin spatial variability of future changes restricted the application of distributed versions of TOPMODEL such as TOPNET and BTOPMC. In this study the original TOPMODEL is applied and will be considered for the multimodal simulation if simulated flow explains the variability in observation acceptably during both calibration and validation period.

(b) NAM; a conceptual lumped rainfall–runoff which was originally developed at the Institute of Hydrodynamics and Hydraulic Engineering at the Technical University of Denmark. The model has been applied in a large number of engineering projects covering various climatic regimes. The NAM model describes, in a simplified quantitative form, the behaviour of the different phases of the hydrological cycle, accounting for the water content in different mutually interrelated storages, namely surface zone storage, the root-zone storage, and the groundwater storage. The surface and interflow component of total runoff is routed through two linear reservoirs and the base flow is routed using a single reservoir. Each linear reservoir is characterised by a specific time constant. In the present application, the nine most important parameters of the NAM model were determined by calibration. The detail on the parameters and more detailed information regarding the NAM model can be found in Madsen [26].

(c) The HYdrologic MODel (HYMOD); also a conceptual and lumped model, was originally proposed by Boyle [11] in order to address the need for the development of models with complexity levels suitable for capturing typical and commonly measured hydrologic fluxes. The objective of HYMOD is to provide a research tool for scientific evaluation purposes (e.g., [51,52]). HYMOD uses a nonlinear tank connected with two series of linear tanks in parallel to model

rainfall excess mechanism. The model consists of two components, a slow flow component and a quick flow component. The slow flow component, modelled using one linear reservoir, represents the ground water. The quick flow component, which represents the surface flow, is modelled using three identical quick flow reservoirs. In the present application of HYMOD, a total of 5 parameters require to be estimated through model calibration.

(d) The TANK model; a conceptual model comprised of four vertical tanks with primary and secondary storage. For each basin, processes of infiltration, unsaturated and saturated flow, and through flow, are represented using a simple 'non-linear tank model' approach [42]. The model comprises of number of simple tanks with outlets arranged vertically one above other. Water in each tank partially discharges through a side outlet or outlets and partially infiltrates through a bottom outlet to the next lower tank. The output from the side outlets are the simulated runoffs. The outflow from the top tank, second, third and fourth tank is considered as surface runoff, intermediate runoff, sub-base runoff and base flow respectively. A total of 15 parameters are estimated through model calibration.

As the modelling objective is restricted to the prediction of discharge at the outlet gauging station of each catchment, the traditional lumped and semi-lumped conceptual models which are simpler, less data-demanding alternatives and cheaper to operate are used. Each of the models vary in the way they conceptualise the key hydrological processes and in complexity, primarily related to the number of parameters requiring calibration. Conceptual stores are common components of all lumped rainfall–runoff models used in this study. NAM and TANK use the linear store, i.e., output is proportional to the amount of stored water. HYMOD uses a nonlinear tank connected with two series of linear tanks in parallel. TOPMODEL uses the exponential store, i.e., output is related to storage through an exponential relationship. The exponential store is generally considered to be a tool for recession and base flow simulation but, as part of a rainfall–runoff model, it can also play an important role in the simulation of high flow events.

The models used also differ in the way they represent the spatial variability of response within the basin. TOPMODEL uses the topographic index derived from digital elevation model data to represent the spatial variability in the soil infiltration capacity. HYMOD uses a statistical distribution function to model spatial variability in soil infiltration capacity, whereas the TANK and NAM models do not take spatial variability within the basin into account. Actual evapotranspiration in all of the models is treated as a function of potential evapotranspiration and soil moisture storage. In the treatment of soil moisture accounting there are differences among the models. Consequently, the mechanisms of runoff generation are different among models. In HYMOD and TOPMODEL, the mechanism of quick runoff can be seen as a translation of the variable source-area concept of runoff generation: the greater the available water, the wetter the catchment; the larger the source area for surface runoff, the greater the part of the precipitation running off rapidly. All four models employ a single linear reservoir to model groundwater.

All four models have been applied in numerous applications and their potential for application to simulate flow under changed climate has been discussed previously, e.g., [44,15,13,2,33,45]. The models employed are independently developed by different researchers and organizations. The common assumption implicit in the application of conceptual models for climate change impact assessment is that the conceptual basis of the model enables the hydrological processes to realistically respond to changes in climatic input.

3.3. Estimation of prediction uncertainty

In order to examine the role of model uncertainty in climate change impact studies and include a full consideration of impact model uncertainty, we explored two methods, namely the Generalised Likelihood Uncertainty Estimation method (GLUE) and Bayesian Model Averaging (BMA).

3.3.1. GLUE

In the GLUE framework a set of behavioural predictions are extracted from the simulation based on the selected goodness-of-fit measure. The most common informal goodness-of-fit measure is based on the sum of squared errors (Eq. (1))

$$L(\theta_i|Y) = \left(1 - \sigma_i^2/\sigma_{obs}^2\right)^{N1} \qquad (1)$$

where $L(\theta_i|Y)$ is the likelihood measure for the ith model conditioned on the observations, Y, σ_i^2 is the error variance for the ith model (i.e., the combination of the model and the ith parameter set) and σ_{obs}^2 is the variance of the observations. The exponent $N1$ is an adjustable parameter that sets the relative weightings of the better and worse solutions.

The GLUE scheme, which is widely used to account for parameter uncertainty, is used to handle both parametrically and structurally different plausible models. A desired number of behavioural predictions from the selected hydrological models are ranked and likelihood weighted to characterise the parameter as well as structural uncertainty propagated through each of the hydrological models. As the number of behavioural simulations are not equal among models, a desired number of behavioural sets of model parameters should be sampled based on the prior probability attached to a model, i.e., random sampling of solutions from behavioural sets in proportion to the prior model probability. As this study assumes that all models are equally probable, only n number of behavioural solutions are randomly sampled from each model. This number n is selected as the minimum of the number of behavioural solutions among models, and the Nash–Sutcliffe efficiency measure (i.e., $N1 = 1$ in Eq. (1)) is used as an informal likelihood measure. Initially, the threshold value of 0.6 was selected, which was fine-tuned for each basin so that the prediction interval encapsulates as much observation as possible, and maintains a good population of behavioural solutions.

The implementation of the GLUE method to estimate prediction uncertainty associated with hydrological models can be expressed through the following procedure;

Step 1. Select K models that are structurally and/or parametrically different and choose the ranges of model parameters for each model.
Step 2. Select the likelihood measure and the threshold to differentiate between acceptable and unacceptable solutions.
Step 3. Run each of K sets of hydrological models with calibration data. At each run, a parameter set is randomly drawn (e.g., using simple random sampling, stratified random sampling, etc.) from the range of the model parameters assuming the parameter follows a uniform distribution over its range.
Step 4. From the number of behavioural solutions of the ith model (i.e., **NB**$_i$ where i = 1, M), obtained for the specified threshold value, sample randomly n number of behavioural solutions and repeat this for all selected models. The likelihood of the accepted solution derived from the set of K models is then rescaled so that their cumulative sum equals 1. Consequently, this rescaled likelihood (RL) is used to assign weight to each runoff prediction.

Step 5. Use the rescaled likelihood weights of the behavioural data set to assess parameter sensitivity and to compute prediction limits on hydrographs using:

$$P(\widehat{Z}_t < z) = \sum_{i=1}^{N} RL[f(\theta_i)|\widehat{Z}_{t,i} < z] \qquad (2)$$

where N is the number of behavioural models, i.e., $K * n$, P is the prediction quantile, θ_i is the ith set of model parameters, $\widehat{Z}_t$ is the value of the variable Z at time t simulated by the model $f(\theta_i)$, and RL is the rescaled likelihood measure.

In quantifying the prediction uncertainties, GLUE requires the user to take some subjective decisions by thinking of possible sources of uncertainties. These subjective choices are highly influential on the uncertainty estimation [31,48]). The prediction intervals estimated from GLUE are dependent upon the selection of models, definition of parameter ranges from which they are sampled, definition of likelihood measures used to evaluate, and method used to combine simulators. Moreover, they also implicitly depend upon input and observational error. By considering different parameter sets in GLUE, the modeller is considering the effect of parameter uncertainty. Furthermore, by considering different model structures and parameter sets the modeller is considering the effect of both parameter and model structural uncertainty. As parameter uncertainty alone can account for model structural uncertainty and vice versa, the effects of the different sources of approximation cannot be evaluated separately in GLUE in an additive way.

3.3.2. BMA

Bayesian Model Averaging provides a solution to the model selection problem by accounting for uncertainty about model forms or assumptions and by propagating this uncertainty to inferences about quantities of interest. In the situation in which several models $\{f_1,\ldots,f_k\}$ are theoretically possible, it is risky to base inference on the point estimates from a single model f_K. BMA allows us to account for this type of uncertainty as the predictive distribution of the quantity of interest, as shown in Eq. (3), is calculated as the average of the posterior predictive distribution of the quantity derived from each individual model weighted by the corresponding posterior model probability

$$p(\Delta|f_1,\ldots,f_k,D) = \sum_{k=1}^{K} p(\Delta|f_k,D)p(f_k|D) \qquad (3)$$

The posterior model probability, $p(f_k|D)$, of model f_k given the data, is given by Eq. (4)

$$p(f_k|D)\alpha p(D|f_k)p(f_k) \qquad (4)$$

where the constant of proportionality is chosen so that the posterior model probabilities add up to one. The prior probability, $p(f_k)$, in Eq. (4) presents the preference of model f_k before re-evaluation. Therefore, a model with better performance in history will have a greater weight in future application. Note that without any prior knowledge of model preference, the prior probability is assumed to have a uniform distribution among the N models. The quantity $p(D|f_k)$ is the integrated likelihood of model f_k.

The posterior mean and variance of Δ are as follows:

$$E[\Delta|f_1,\ldots,f_k,D] = \sum_{k=1}^{K} w_k \widehat{\Delta}_k \qquad (5)$$

$$Var[\Delta|f_1,\ldots,f_k,D] = \sum_{k=1}^{K} (Var(\Delta|D,f_k) + \widehat{\Delta}_k)w_k - E(\Delta|D)^2 \qquad (6)$$

where $\widehat{\Delta}_k = E(\Delta|D,f_k)$. Note that weight w_k has a value only between 0 and 1. A larger value indicates more preference on the prediction

by model f_k. In this application, the PDF from each model at time t is modelled by a gamma distribution (Eq. (7)) with heteroscedastic variance (Eq. (8))

$$p(\Delta|f_k) = \Delta^{\alpha_k - 1} e^{(\Delta/\beta_k)} / (\Gamma(\alpha_k)\theta^{\alpha_k}) \qquad (7)$$

$$\alpha = \mu_k^2/\sigma_k^2; \quad \beta_k = \sigma_k^2/\mu_k; \quad \mu_k = f_k; \quad \sigma_k^2 = b \cdot f_k + c \qquad (8)$$

$$l(w_1, \ldots, w_k | \sigma_1^2, \ldots, \sigma_k^2, \Delta) = \sum_{t=1}^{n} \log(w_1 p(\Delta|f_1) + \cdots + w_k p(\Delta|f_k)) \qquad (9)$$

where b and c in Eq. (8) are the coefficients that relates model output with model variances. As value of streamflow is nonzero and the distribution of daily streamflow is highly skewed, the PDF of from each model is modelled using gamma distribution. At each time step, the chosen PDF is centred on the individual forecasts with an associated variance that is heteroscedastic and directly depends on the actual stream flow prediction. The BMA parameters, i.e., weights and variances, were obtained from historical stream flow data (1971–1990) using Markov Chain Monte Carlo (MCMC) sampling (see [49]). This method uses the Differential Evolution Adaptive Metropolis (DREAM) algorithm developed by Vrugt et al. [50].

The probabilistic predictions of daily streamflow were derived based on each individual deterministic predictions obtained from each hydrological model and their weight and variances. The procedures used in this study to generate probabilistic predictions at each time step t are briefly described below.

Step 1. Select K models that can be structurally or parametrically different.

Step 2. Generate model prediction sets $\hat{y}_{i,k}$ ($i = 1, 2, \ldots, N$; $k = 1, 2 \ldots, K$).

Step 3. Calculate weights w_k and variance Var_k for each of the selected models.

Step 4. Generate new model-based prediction $\widehat{Y}$ using Eq. (5).

Step 5. Probabilistic predictions are made using mean weight (w_k) and variance parameters (Var_k) as follows:
 – Select an individual competing model (f_k) with probability proportional to its weight.
 – Sample from the probability distribution associated with the output from each individual model.
 – Repeat above two steps to sample a number of values that represent the distribution of streamflow at time t, and subsequently derive the uncertainty interval.

The BMA method is a statistical approach to conducting inference when faced with model selection uncertainty. The uncertainty of BMA depends upon the independence of the output from different models that are combined. It also depends on how the conditional probability density functions are associated with each prediction. BMA produces a complete PDF as a prediction and provides a quantification of the uncertainties. If one has a high enough number of ensemble members forming a sample of the hydrological model space, BMA will give a better approximation for the estimate of the hydrological modelling uncertainty. However, such situations are difficult to hold due to an under sampling of the hydrological model space and moreover the available conceptual models cannot be regarded as truly statistically independent because most of them use similar conceptualisations of dominant hydrological processes, a problem apparent for many environmental models, including GCMs.

In this study, the median predictions obtained from the GLUE method form the basis for implementing the BMA here, thereby incorporating deterministic predictions from four CRR models forced with observed rainfall data in order to make probabilistic predictions. Concerning the application of BMA, the model space could be expanded with output simulated from the use of different

likelihood functions, e.g., Duan et al. [16]. A wide range of statistical and hydrological objective functions are available, however, while different studies have tried to assess the suitability of different performance measures, it ultimately remains a subjective decision of the hydrologist to select one or more measures suitable for the task at hand. In this study, the likelihood based on Eq. (1) is used for the GLUE method and therefore to compare it with BMA, the output obtained from the same likelihood function is taken into account. The BMA variance (Eq. (6)), which contains two components: the between-model-variance and the within-model-variance, is essentially an uncertainty measure of the BMA prediction. This measure is a better description of predictive uncertainty than that which estimates uncertainty based only on the ensemble spread. In BMA, the uncertainty in model parameter values can be regarded as within-model uncertainty, and uncertainty in model choice can be regarded as between model uncertainty. Because the focus of this study is on the definition of uncertainty arising from a number of plausible models and not on the model selection problem we did not attempt to penalise the models depending upon the number of calibration parameters. This could have been incorporated within the BMA framework by assigning the prior probability of $P(f_k)$, in Eq. (4) that represents the preference of model f_k instead of sampling it from a uniform distribution.

The uncertainty of BMA depends upon the independency in the output from different models that are combined using BMA technique. It also depends on how the conditional probability density functions are associated with each prediction. BMA produces a complete PDF as a prediction and provides a quantification of the uncertainties. If one has a high enough number of ensemble members forming a sample of the hydrological model space, BMA will give a better approximation for the estimate of the hydrological modelling uncertainty.

3.4. Experiment design

In order to evaluate the role of hydrological model uncertainty in relation to the uncertainty envelope associated with the estimation of future impacts on stream flow, the response of the catchments to input from three GCMs forced with two GHG emission scenarios, evaluated from four hydrological models and their behavioural parameters sets was used. The results are presented for three benchmark periods in the future; 2020–2029 (2020s), 2050–2059 (2050s), and 2070–2079 (2070s). As it is difficult to attach preference to one scenario over the other, both scenarios (i.e., A2 and B2) were assumed equally likely. The predictions from the three GCMs were weighted based on the Climate Prediction Index [32] that reflects the ability of the GCM to reproduce observed climate data. This is done by multiplying the likelihood functions of the accepted solution and then rescaling it, similar to a probability measure, in order to make the cumulative sum equal to 1. Subsequently the simulated uncertainties are apportioned and assessed as follows:

 – HYDRO: the uncertainty in future simulations due to hydrological model structure and their parameters.
 – SCENE: the uncertainty in future simulations due to selection of emission scenario.
 – GCM: the uncertainty in future simulations due to the selection of climate models.
 – TOTAL: the total uncertainty in future simulations of stream flow from all combined sources.

To examine the performance of the prediction intervals in capturing the observed flows, an index defined as the ratio of the number of observations falling within their respective prediction intervals, to the total number of observations (hereafter referred

as count efficiency), and the average width of the prediction interval is used.

Concerning the application of BMA, the following four different probabilistic predictions were made by combining: (a) four median predictions obtained from the selected CRR models (referred to as HYDRO) forced with the selected GCM and the scenarios, (b) eight median predictions estimated from the CRR models forced with regional climate scenarios corresponding to A2 and B2 scenarios derived from the selected GCM (referred to as SCENE), (c) 12 median predictions estimated from the four CRR models forced with regional climate scenarios data derived from three GCMs (referred to as GCM A2 (B2)) and (d) 24 median predictions obtained from four CRR models forced with six regional climate scenarios (referred to as TOTAL). In all the experiment design, the BMA weight and variance parameters were estimated from the calibration period (1971–1990). However, depending upon the regional climate change scenario used, the BMA weight parameters were suitably modified. In this study, the calibrated BMA weight parameter was modified depending upon the ability of the GCM to reproduce climate data at a more regional scale. Furthermore, both scenarios were assumed equally likely in BMA, i.e., the weight parameter derived for each GCM is equally divided among A2 and B2 scenarios such that the total weight sums to one.

4. Results and discussion

4.1. Performance of GLUE and BMA under observations

The hydrological discharge simulation is carried out at a daily time step using the four conceptual models calibrated on observed data for the period of 1971–1990 and validated using the period of
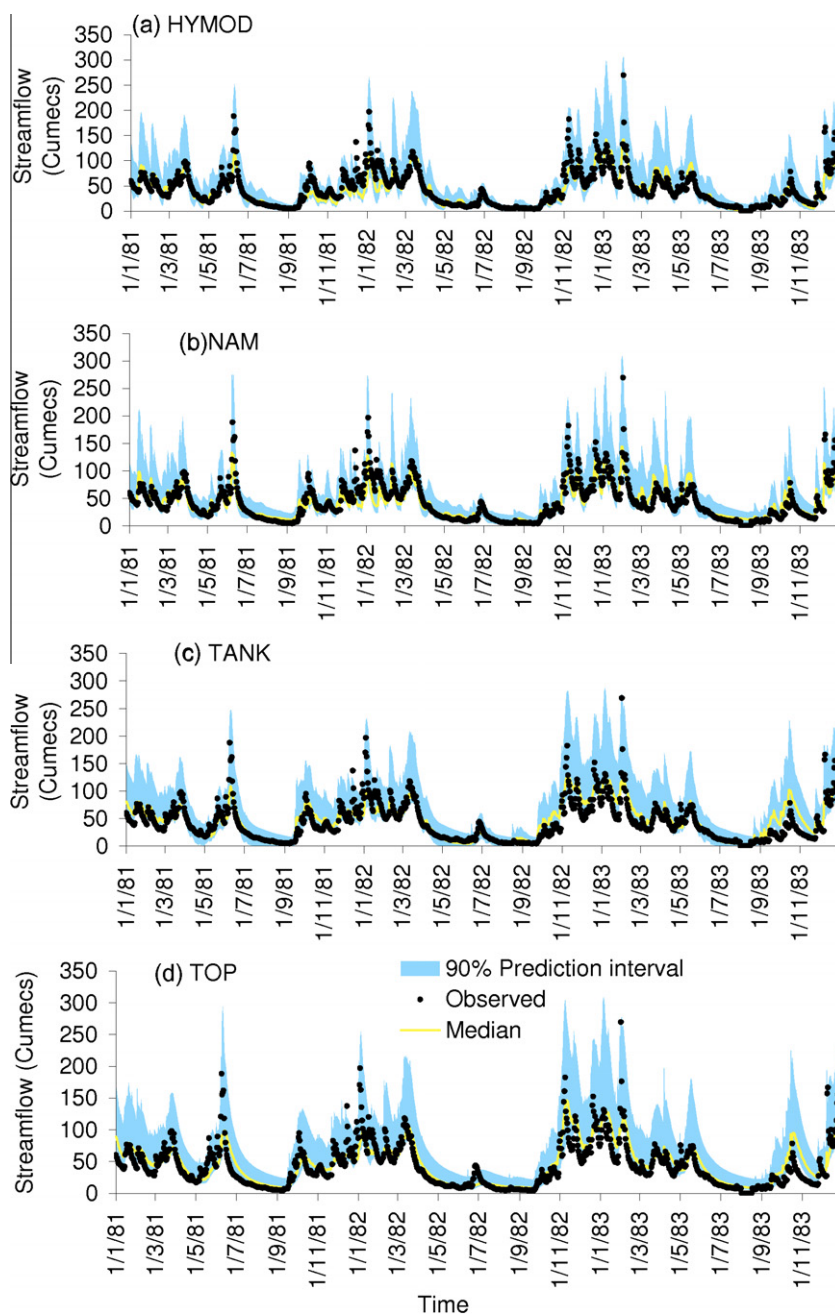

Fig. 2. Prediction intervals for Boyne basin including median and observed flow simulated from: (a) HYMOD, (b) NAM, (c) TANK, (d) TOP for the selected period (1981–1983).

Table 2
The performance of median prediction, the percentage of observation encapsulated within the prediction interval (CE), and the average width of the 90% prediction interval (PI) for each model during the calibration and validation period.

Sn	Period (calib/valid)	Basin (model)	NSE (median)		CE		PI (m³/s)	
			Calib	Valid	Calib	Valid	Calib	Valid
1	1971–1990/ 1991–2000	Moy (HYMOD)	0.77	0.66	0.68	0.56	30.5	33.0
2		Moy (NAM)	0.72	0.63	0.5S	0.52	25.7	27.7
3		Moy (TANK)	0.80	0.69	0.80	0.77	40.9	44.6
4		Moy (TOP)	0.80	0.70	0.72	0.70	34.0	37.5
5	1971–1990/ 1991–2000	Boyne (HYMOD)	0.79	0.76	0.80	0.83	28.2	29.4
6		Boyne (NAM)	0.76	0.74	0.77	0.7 N	23.8	25.1
7		Boyne (TANK)	0.70	0.73	0.67	0.75	25.6	27.1
8		Boyne (TOP)	0.69	0.68	0.52	0.57	23.3	24.7
9	1971–1990/ 1991–2000	Suck (HYMOD)	0.78	0.68	0.70	0.68	17.3	18.8
10		Suck (NAM)	0.72	0.63	0.56	0.51	14.7	15.9
11		Suck (TANK)	0.70	0.65	0.61	0.59	17.1	18.5
12		Suck (TOP)	0.68	0.60	0.34	0.31	12.7	14.1
13	1971–1990/ 1991–2000	Blackwater (HYMOD)	0.64	0.74	0.50	0.58	25.18	25.67
14		Blackwater (NAM)	0.63	0.72	0.31	0.40	15.62	16.13
15		Blackwater (TANK)	0.67	0.75	0.59	0.68	33.35	34.09
16		Black water (TOP)	0.64	0.71	0.33	0.31	21.77	22.69

1991–2000. The GLUE scheme was implemented to account for parametrically and structural different hydrological models. The number of behavioural predictions from each of the hydrological models was ranked and likelihood weighted to describe the parameter as well as structural uncertainty. Fig. 2(a–d) shows the prediction interval for the Boyne basin (1981–1983) for each hydrological model.

Table 2 shows the median NSE, the prediction interval (PI) and count efficiency (CE) for each model. It reveals that the PI and CE estimated from one model on a particular basin is different from the PI estimated on the same basin by a different model and on different basins. The GLUE estimated PI, CE and number of behavioural simulations (NB) for each individual model and basin depend on the threshold values. The selection of threshold values were made based on a sensitivity analysis where these measures, i.e., PI, CE and the number of behavioural simulation (NB), were estimated for different threshold values, namely NSE of 0.3, 0.5, and 0.7. For all models the PI, CE and NB increased with a decrease in value of the threshold and vice versa. However, the rate of decrease of PI, CE, and NB are (5%, 15%, 40%, respectively for PI, CE and NB) much smaller when moving the threshold value from 0.3 to 0.5 than when moving it from 0.5 to 0.7 (25%, 37% and 73%, respectively for PI, CE and NB). In this study, the threshold value of 0.6 was selected for Boyne and Moy and 0.5 for Blackwater and Suck basins.

This is done so that sufficient numbers of behavioural samples for each model can be obtained and at the same time the benefit in terms of improving the value of CE with decreasing threshold is small. Even with the decreased threshold (0.3), the 90% confidence interval could not encapsulate 90% of the observation data. The PI showed a tendency to grow wider with increasing discharge and with increasing variance in discharge. Moreover, it varied among hydrological models. In general, the prediction interval estimated from TOPMODEL and NAM are marginally smaller than the

PI estimated from HYMOD and TANK, with PI estimated from TANK being the widest. It is interesting to note that the TANK model has the highest number of parameters followed by NAM, HYMOD and TOPMODEL, respectively. Despite having fewer parameters, the PI estimated from HYMOD is, in most instances, bigger than NAM. Moreover, this comparison does not reveal any distinct relationship between the number of calibration parameters and the prediction interval or uncertainty in model prediction. For the Suck and Blackwater catchments, the PI simulated by TOPMODEL only encapsulated 30% of the observations, whereas the percentage of observations that are encapsulated within the PI are higher for HYMOD and TANK. This clearly indicates that the extent of uncertainty in prediction explained by model parameterization alone varies among models. Though the PIs estimated from different models show a general increase in count efficiency with wider PIs, the increase in CE is not proportionate with the increase in the PI, e.g., in the Boyne the PIs simulated by NAM and TOPMODEL are very similar but there is an apparent difference in the count efficiency for the PI resulting from these two models. Therefore, these four models, which differ in their conceptualisation of hydrological processes and their variability, produce apparently different simulations and descriptions of the uncertainty in the prediction. Therefore, both GLUE and BMA that weight model prediction based on model likelihood are utilised to address the model uncertainty. Concerning the application of BMA, the median prediction of the individual model obtained from the GLUE scheme, i.e., four individual time series of prediction obtained from each hydrological model, is processed. The probability density function from each model at any given time is modelled by a gamma distribution with heteroscedastic variance.

Fig. 3(a) shows the daily 90% PI (the results for only a three year is shown), derived using GLUE, for the Boyne when all four models and their behavioural parameter sets are taken into account. The width of the prediction interval – expressed in terms of cumecs – increased when different models are considered. In addition, the count efficiency of the prediction interval improved when different model structures are incorporated. Fig. 3(b) shows the daily PI for the Boyne when the median output from all four CRR models are combined using BMA. The prediction intervals estimated from GLUE are sharp as compared to BMA. Furthermore, posterior model output, estimated from BMA is more symmetrical than GLUE, which is skewed towards the lower bound.

Table 3 shows that the PIs estimated from GLUE are narrower than the same obtained from BMA. Consequently, a larger proportion of observations are encapsulated within the PI estimated from BMA than from GLUE for the selected threshold value. Furthermore, the median model performance obtained after processing different plausible predictions through BMA is better in terms of NSE than individual predictions. The inadequacies of the prediction interval estimated from GLUE in capturing the observations can be attributed to the subjectivity involved in the selection of threshold values and likelihood measures. The threshold value selected for implementation of GLUE will have effect on the PI and subsequently on the capability of the PI to capture the observed runoffs. Apart from that, the percentage of runoff observations bracketed by the prediction limits is still subject to a number of factors that affect the rainfall–runoff modelling efficiency. In addition, the uncertainty associated with the input data was not explicitly accounted for in both methods.

To implement the BMA scheme numerically, we have assumed that the model ensemble used in BMA are representative of the entire model space and that the individual ensemble members are independent of each other. Such assumptions of independence are unlikely to hold especially for environmental models. The empirical evidence that the prediction interval derived from BMA for the independent validation period encapsulated 90% of
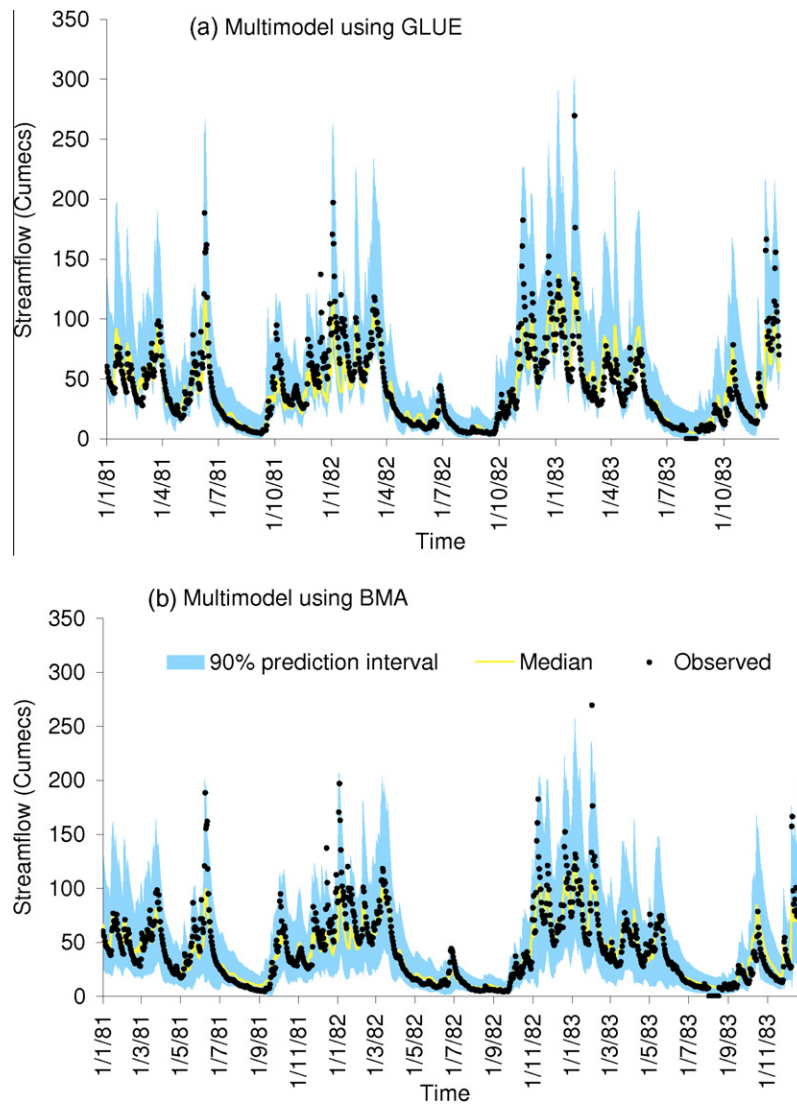
Fig. 3. Prediction intervals for Boyne basin including observed and median flow produced from multi-model ensemble of four selected models using (a) Generalised Uncertainty Estimation Method (GLUE) and (b) Bayesian Model Averaging (BMA) for the selected period (1981–1983).

Table 3
The performance of median prediction, the percentage of observation encapsulated within the prediction interval (count efficiency, i.e., CE), and the average spread of the 90% prediction interval (PI) for each model during the calibration and validation period.

Sn	Basin	Scheme	NSE (median)		CE		PI (m³/s)	
			Calib	Valid	Calib	Valid	Calib	Valid
1	Moy	GLUE	0.81	0.72	0.85	0.80	43.32	46.8
2	Boyne		0.80	0.78	0.90	0.92	31.8	33.4
3	Suck		0.79	0.69	0.74	0.70	19.2	20.9
4	Blackwater		0.66	0.74	0.68	0.76	36.52	37.32
5	Moy	BMA	0.90	0.79	0.97	0.92	64.03	66.76
6	Boyne		0.80	0.75	0.96	0.96	44.67	48.26
7	Suck		0.82	0.76	0.96	0.93	30.90	32.19
8	Blackwater		0.73	0.76	0.91	0.93	78.10	80.90

observation indicates that the models have explored the wider model space in the application shown in this paper. Furthermore, as the selected conceptual models can never cover all possibilities, but only a limited range, it is important to emphasise that the overall uncertainty of model predictions cannot be assessed in an absolute sense, only in a conditional or relative sense [9].

4.2. Contribution of CRR models to the envelope of future simulations

A common assumption implicit in most of climate change impact studies is that hydrological models calibrated over the historical period are valid for use in the future under a climate change regime. However, only a small number of studies have tested the validity of this assumption. Arnell [3] recommends that if model parameters are calibrated against a long time series of historical data containing both wet and dry periods then these parameters can be assumed to be valid under future climates, with a greater degree of confidence. A recent study from Vaze et al. [47] presents results from a modelling study carried out to investigate whether the calibrated parameter values for rainfall–runoff models based on historical observed data can be used to reliably predict runoff responses to changes in future climate inputs. The results, which involved four conceptual rainfall–runoff models and 61 catchments located within Australia, indicate that the models, when calibrated using more than 20 years of data, can generally be used for climate impact studies where the future mean annual rainfall is not more than 15% drier or 20% wetter than the mean annual rainfall in the model calibration period. Furthermore, their study also observed that it is generally more difficult for a model calibrated
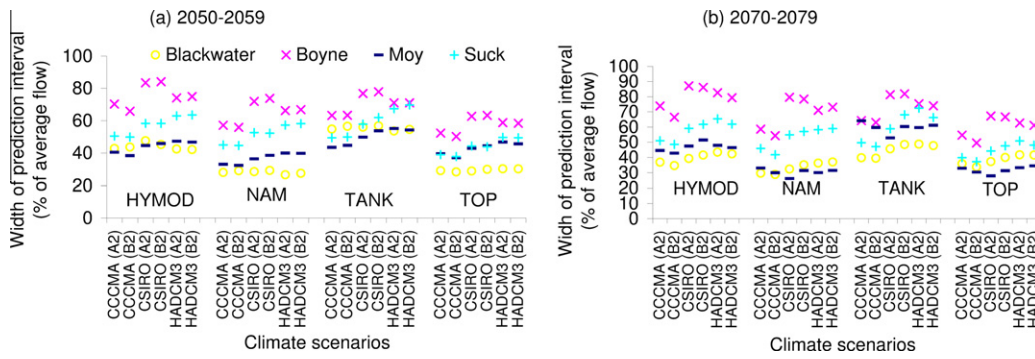
Fig. 4. Average width of the prediction interval simulated from behavioural set of model parameters derived for each of the selected four conceptual rainfall–runoff models for the period 2050–2059 and 2070–2079.

over a wet period to predict runoff over a dry period compared to a model calibrated over a dry period to predict runoff over a wet period. For Irish basins, the annual change in precipitation is mostly within 10% change. As the calibration period is sufficiently long to cover both wet and dry periods, i.e., the 1970s is a relatively dry decade whereas the 1980s are relatively wet; and the change in annual precipitation change compared to data that are used during model calibration is within 10%; the calibrated parameters were assumed valid for future simulation.

Fig. 4 shows the uncertainty in model prediction, expressed in terms of the average width of the prediction interval (% of long term average flow) arising from uncertainties associated with parameterization of the hydrological model for the period from 2050 to 2059 and from 2070 to 2079, when forced with six regional climate scenarios. It reveals that the prediction uncertainties arising from parameterization depend upon the characteristics of the CRR models, the regional climate scenarios and the type of catchment. On average, the PI (%), expressed in terms of observed flow, grew wider with time. The uncertainty in prediction for the future time period is highest for the Boyne catchment and smallest for the Blackwater, closely following the results obtained during model calibration. The variation of uncertainty among basins is likely to arise due to the variation in the applicability of CRR models and the variation in physical parameters of a basin, which plays an important role in characterising the response of a basin to a given input. In the selected basins, the PI (%) grew with a decrease in runoff coefficient and wetness index. Both of which tend to increase the nonlinearity in the basins response.

Fig. 5 illustrates the response of the Boyne basin simulated by the behavioural parameter sets of four different hydrological mod-

els, for two time periods, to climate scenarios derived from HAD-CM3 for the A2 scenario. The median estimates from the selected models are not significantly similar ($\alpha = 90\%$, two tail) with the exception of the median estimate from TOPMODEL and TANK. Consequently, the prediction intervals derived from the multi-model ensemble approach are wider in comparison to the estimates from each individual CRR model. It therefore shows the value of combining output from the selected models in explaining the uncertainties that the parameterization alone cannot account for. Fig. 6 shows the average width of the PI expressed in terms of the percentage of long term average flow for the three time periods, namely 2020s, 2050s and 2070s. The average width of the PI arising from uncertainties associated with the parameterization of CRR models is nearly 50% of the average flow and it increased, on average, to 70%, when different CRR model structures are included. However, this does not indicate that the role of hydrological model uncertainty is less in magnitude than parameter uncertainty as sources cannot be disintegrated as there is no true value of model parameter or structure that can be estimated from field measurements.

The width of the prediction interval nearly equalled the average flow when both scenarios are taken into account. It grew to nearly 120% of the average flow when three GCMs with A2 (B2) scenarios are included, and approached 140% of the average discharge when the total uncertainty was incorporated. Similarly, the uncertainties arising from the hydrological model varied among basin, climate scenario and the selected GCM highlighting the importance of conducting impacts assessment for individual catchments. In addition, the uncertainty derived from the choice of GCM is greater than that derived from emission scenario.
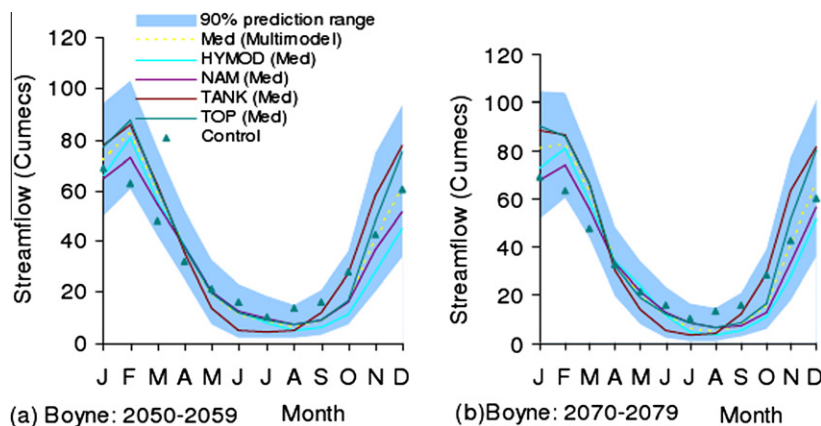
Fig. 5. The 90% prediction interval (shaded region) simulated by the behavioural parameters sets of four hydrological models forced with climate inputs from HADCM3 (A2) for two periods namely, 2050–2059 and 2070–2079 and median predictions from each individual hydrological model (continuous lines) for Boyne catchment.
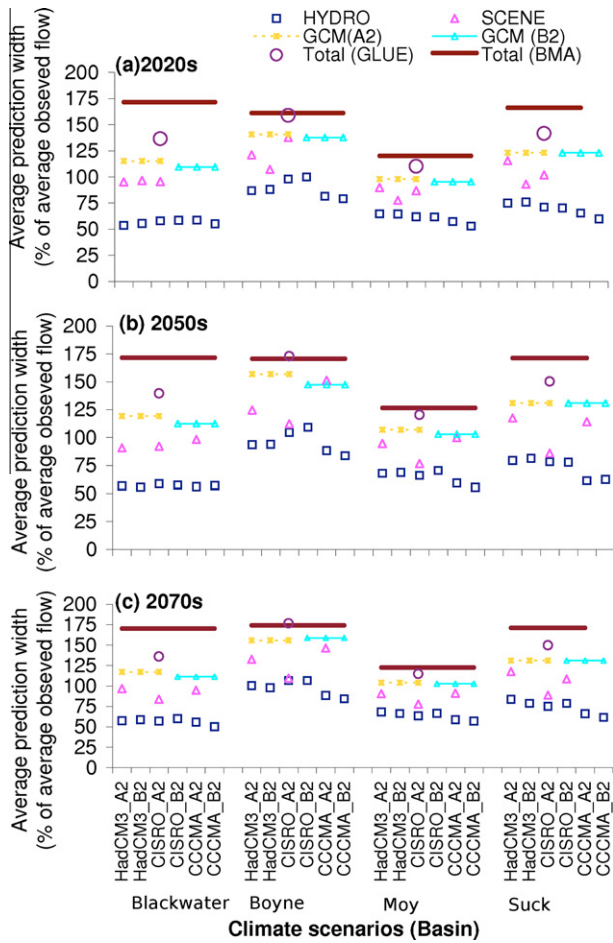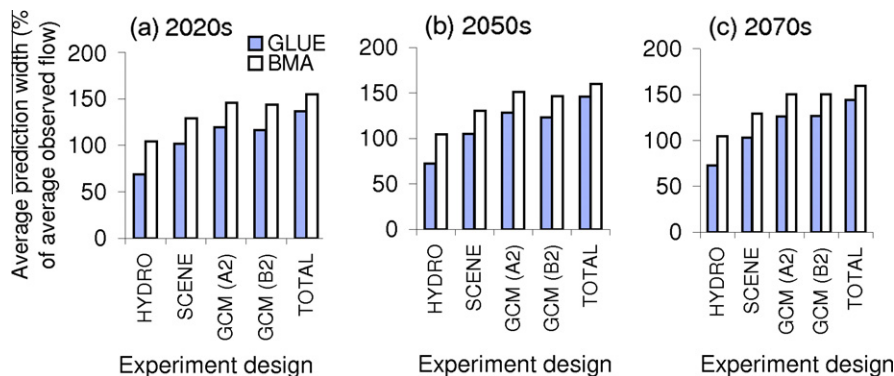
Fig. 6. The average width of the prediction interval (%) for three time period and for each of the five experimental design, i.e., uncertainty in the hydrological models (HYDRO), uncertainty in the selection of scenario and HYDRO (SCENE), uncertainty in the selection of Global Circulation Model (GCM) forced with A2 (B2) scenario and HYDRO (GCM (A2 (B2)), uncertainty in the selection of GCM, selection of scenario and uncertainty in hydrological models (TOTAL). The prediction uncertainty estimated from Bayesian Model Averaging (BMA) is also shown for the same time period.

However, the full range of emission scenario and GCM sensitivities are not sampled here and therefore results are only indicative, nonetheless they agree with the majority of research reported to date (e.g., [53,37]). The hydrological uncertainty seems to be fairly constant for each future time period, but the effect of GCM is

apparently different among three time periods. Fig. 7 shows the BMA estimates for five experiment design, namely for HYDRO, SCENE, GCM A2 (B2), and TOTAL for the three time periods and averaged across basin. The widths of the prediction interval (%) estimated from BMA are higher than the same estimated from GLUE by a factor of 1.4, 1.2, 1.2, 1.2, 1.1, respectively for HYDRO, SCENE, GCM A2 (B2) and TOTAL. Though both BMA and GLUE fairly much agree in magnitude for TOTAL, the role of hydrological model uncertainty is amplified by a factor of 1.4 when BMA is used. Both methods show that the role of hydrological model uncertainty is considerable and warrants routine inclusion in impacts assessment, particularly where robust adaptation decisions are required.

The effects of the different sources of uncertainty cannot be evaluated separately in an additive way as one can compensate for the other. Deconstruction of error into source components is difficult, particularly in hydrological modelling where the model is highly nonlinear and different sources of error may interact in a nonlinear way to produce deviation [10]. Therefore it is difficult to express each source of uncertainty individually with GLUE but, rather, it can be implicitly dealt with. In this study, while evaluating the source of uncertainty due to the selection of scenarios, the uncertainty in hydrological models (parameters and structure) is implicitly taken into account. The uncertainty in scenario selection is independently quantified for each of the three GCMs. Furthermore, while evaluating the sources of uncertainty due to GCM selection, as earlier, the hydrological model uncertainty is implicitly taken into account. The uncertainty in GCM is independently quantified for each of the two scenarios (A2 and B2). When evaluating the total envelop, the GCM uncertainty and scenario uncertainty cannot be treated as independent as the error in one can compensate for other. Therefore, while evaluating the total envelope, hydrological uncertainty, GCM uncertainty and scenarios uncertainty are taken into account by propagating the entire six regional climate scenarios through selected models using the behavioural parameters identified during model calibration. In all the five experiment designs the uncertainties in prediction are expressed as the cumulative effect.

Furthermore, Fig. 8 shows the total uncertainty envelope derived from six climate scenarios and four hydrological models for the four study basins and each future time period using both GLUE and BMA. Although the median prediction obtained from both GLUE and BMA are significantly similar, the upper 95% and lower 5% prediction quantile estimated from BMA are wider than the same estimated from GLUE. The smaller value in the GLUE PI in comparison to BMA can be attributed to the selection of a threshold value and likelihood measure. Furthermore, the GLUE implemented in this study uses a simplistic MC sampling scheme to sample parameters from their prior distributions. While this meth-

Fig. 7. The average width of the prediction interval (%), averaged across basin, estimated from Generalised Likelihood Uncertainty Estimation method (GLUE) and Bayesian Model Averaging (BMA) associated with various sources of uncertainties.
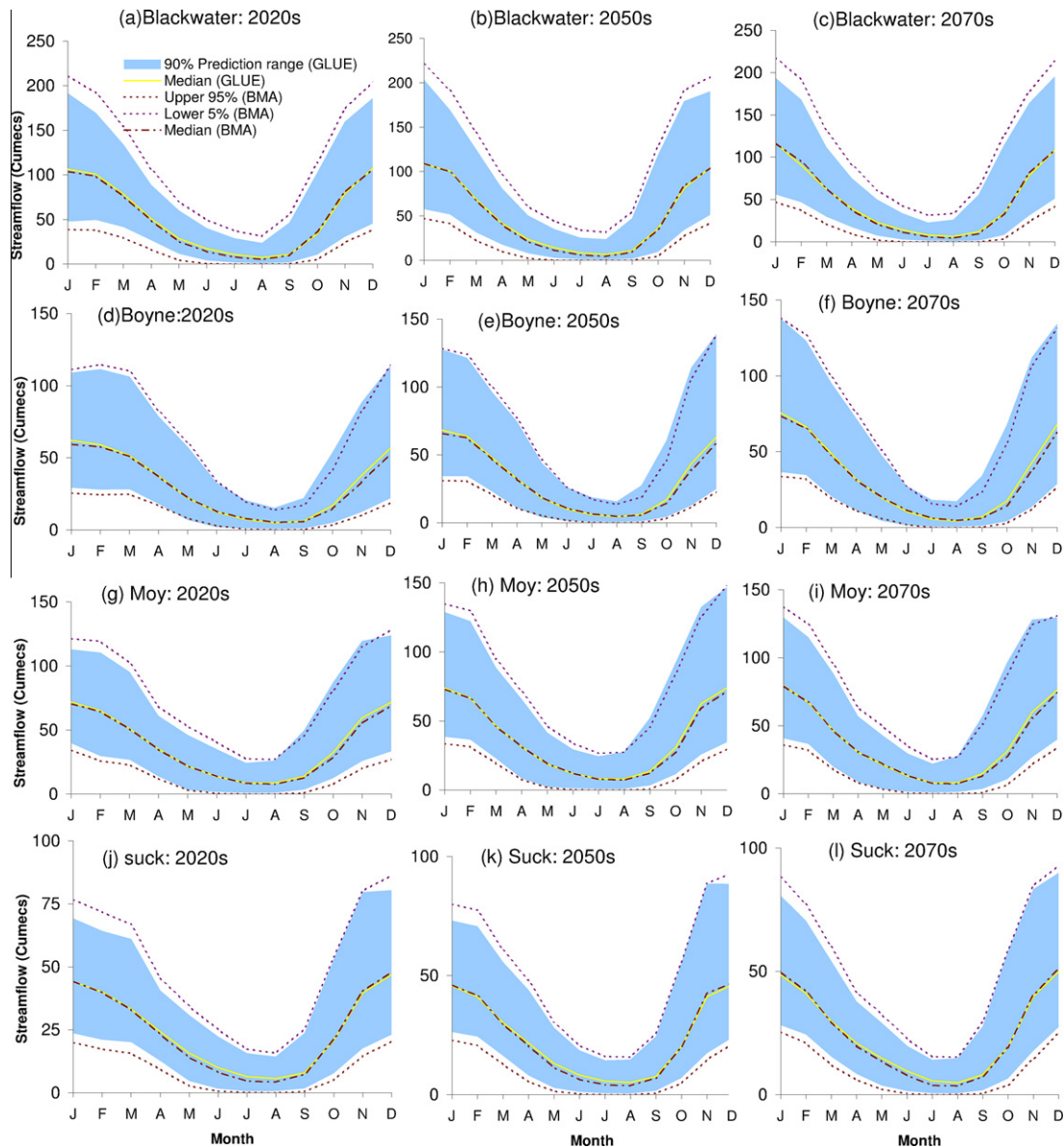
Fig. 8. The total uncertainty envelope derived from six climate scenarios and four hydrological models for the four study basins and for three time periods using Generalised Likelihood Uncertainty Estimation method (GLUE) and Bayesian Model Averaging (BMA).

od is adequate for simple models, complex models may require improved sampling strategies.

As the suitability of application of any selected model and the extent of nonlinearity in input–output relationship differs among basins, the uncertainty in prediction associated with parameter uncertainty and the selection of hydrological model is likely to be different among basin. The runoff coefficient for the Boyne and Suck are markedly lower than the Blackwater and Moy, and the nonlinear behaviours are common in basins that have low run-off coefficients (e.g., [34]). Moreover, the Boyne is the driest basin among the four selected. The physical characteristics of the basins, such as area, runoff coefficient and wetness therefore play a key role in explaining the variation in the PI among basins. The catchment characterised with largest catchment area, lowest runoff coefficient and lowest wetness index resulted in the widest prediction interval in all the five experiments designed in this study. Owing to the limited number of basins used in this study, we do not intend to make a regional relationship for the extrapolation of results beyond the basins used in this study. Further work is currently underway in relation to this.

Fig. 9 shows the percentage change in simulated monthly flow regime for each future time period. For all the basins, there is a tendency of an increase of flow in winter and decrease of flow in summer when moving from the 2020s to the 2070s and similar decreases in summer discharges as the century progresses, with associated implications for water management.

5. Conclusion

There is a cascade of uncertainty in climate change impact assessment that begins with the construction of future emission scenarios and ends in impact assessment. This study addresses the uncertainty in the projection of future water resources by incorporating four plausible yet conceptually diverse CRR models, forced with regional climate scenarios, using BMA and GLUE. In terms of the climate change signature there is a tendency of an increase of flow in winter and a decrease of flow in summer. As the magnitude of increases and decreases, as well as the uncertainty from each source considered vary among the basins selected it is
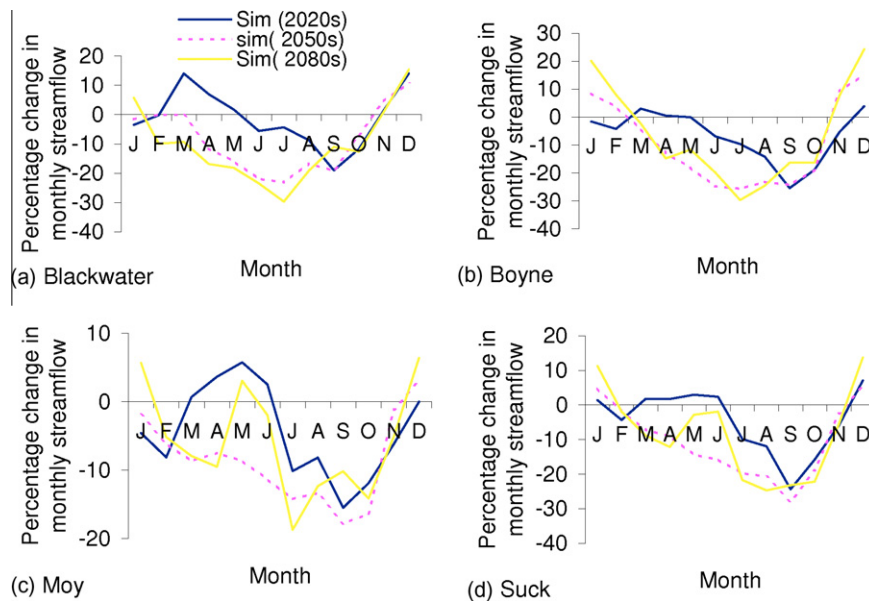
Fig. 9. Percentage change in monthly streamflow, median prediction derived from hydrological models forced with regional climate scenarios derived from three global climate models, and two emission scenarios using Generalised Likelihood Uncertainty Estimation method, in the Blackwater, Boyne, Moy and Suck catchments for three future time periods, namely,2020s, 2050s and 2070s.

critically important that a full impact assessment that accounts for the full range of uncertainties (including CRR model parameter and structure) be conducted where important decisions are required to be robust to climate change.

The uncertainties derived from the use of behavioural model parameters for each model is nearly 40% of the average river flow for calibration. The prediction interval (PI) and count efficiency (CE) estimated from GLUE varied among selected models and basins with no distinct relationship observed between the number of calibration parameters and the prediction interval. However, PI showed a tendency to grow wider with increasing discharge and with increasing variance in discharge. Among models, the prediction interval estimated from TOPMODEL and NAM are marginally smaller than the PI estimated from HYMOD and TANK, with PI estimated from TANK being the widest. It is interesting to note that the TANK model has the highest number of parameters followed by NAM, HYMOD and TOPMODEL, respectively. Despite having fewer parameters, the PI estimated from HYMOD is, in most instances, bigger than NAM. Though the PIs estimated from different models show a general increase in count efficiency (CE) with wider PIs, the increase in CE is not proportionate with the increase in PI. Furthermore, the 90% confidence prediction interval derived from GLUE did not encapsulate 90% of the observations. However, an improvement in the reliability of the prediction interval was apparent when the uncertainty in the selection of model structure was accounted for. The widths of the PI obtained from BMA are wider than obtained by each model and combination of entire models within the GLUE framework, for the selected threshold values and likelihood measures. The smaller value in the GLUE PI in comparison to BMA can be attributed to the selection of a threshold value and likelihood measure. Furthermore, the GLUE implemented in this study uses a simplistic MC sampling scheme to sample parameters from their prior distributions. While this method is adequate for simple models, complex models may require improved sampling strategies.

The same tool was further used to identify the role of the uncertainty in the hydrological models in the overall uncertainty envelopes by utilising six scenarios derived from three GCMs forced by two of the SRES emission scenarios representing and medium

high (A2) and medium low (B2) GHG evolution, to force each CRR model along with their behavioural sets of model parameters identified during calibration. The uncertainties derived from the use of behavioural model parameters for each individual model for future time period was nearly 50% of the average river flow. Moreover, prediction uncertainties from model parameterization varied among catchment, model and grew wider in time when moving from 2050s to 2070s. The PI increased to 70% when uncertainties in CRR model structure were accounted for. The width of the prediction interval nearly equalled the average flow when both scenarios are taken into account. It grew to nearly 120% of the average flow when three GCMs with A2 (B2) scenarios are included, and approached 140% of the average discharge when the total uncertainty was incorporated. The hydrological uncertainty appears to be fairly constant for each future time period, but the effect of GCM is apparently different among periods. Although the median prediction obtained from both GLUE and BMA are significantly similar, the upper 95% and lower 5% prediction quantile estimated from BMA are wider than the same estimated from GLUE. The width of PI estimated from BMA are higher than the same estimated from GLUE by a factor of 1.4, 1.2, 1.2, 1.2 and 1.1, respectively when model uncertainty, model uncertainty along with emission uncertainty, model uncertainty along with GCM uncertainty, and uncertainty including model, emission and GCMs are accounted for respectively. Similarly, the uncertainties arising from the hydrological model varied among basin, climate scenario and the selected GCM highlighting the importance of conducting impacts assessment for individual catchments. In addition, the uncertainty derived from the choice of GCM is greater than that derived from emission scenario.

This application therefore shows that hydrological model uncertainty has a significant role in the uncertainty envelopes of future climate change impacts and should be routinely considered in assessments, particularly where adaptation decisions are required to be robust to uncertainty. In addition to GLUE, BMA was also used to examine the uncertainties associated with future estimation of streamflow at the catchment scale. Using BMA probabilistic prediction were made by combining 24 median predictions (from GLUE) obtained from six climate scenarios and four hydro-

logical models. BMA is found to be a useful approach for application in climate change impact studies, allowing predictions from different models forced with input from different scenarios and GCMs to be combined in an efficient manner. Quantification of CRR uncertainty (parameter and structure) using BMA resulted in an uncertainty band that is apparently similar to the same estimated from GLUE. Clearly, any approach to modelling data that considers a set of competing models has merit. In our application, there were clear differences in individual predictions obtained from four models. Hence, use of BMA and or GLUE is likely to add value to a prediction by helping in avoiding predictions obtained with an inappropriate model and allowing a truer sense of uncertainty to be incorporated into future simulations, thereby increasing the information content available for decision making.

Acknowledgements

References

[1] Ajami NK, Duan Q, Sorooshian S. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour Res 2007;43:W01403. doi:10.1029/2005WR004745.
[2] Andersen HE, Kronvang B, Larsen SE, Hoffmann CC, Jensen TS, Rasmussen EK. Climate-change impacts on hydrology and nutrients in a Danish lowland river basin. Sci Total Environ 2006;365:223–37.
[3] Arnell NW. Hydrological impacts of climate change. In: Calow P, Petts GE, editors. The rivers handbook, vol. 2. Oxford: Blackwell; 1994. p. 173–85.
[4] Bandaragoda C, Tarboton DG, Woods R. Application of TOPNET in the distributed model intercomparison project. J Hydrol 2004;298:178–201.
[5] Beven K, Binley AM. The future of distributed models: model calibration and uncertainty prediction. Hydrol Process 1992;6:279–98.
[6] Beven K. TOPMODEL: a critique. Hydrol Process 1997;11:1069–85.
[7] Beven K. Rainfall–runoff modelling – the primer. Chichester, UK: John Wiley & Sons; 2000.
[8] Beven K. Towards a coherent philosophy for modelling the environment. Proc Roy Soc Lond A 2002;458(2026):2465–84.
[9] Beven K. Response to "Emergence of a new kind of relativism in environmental modelling: a commentary". Proc Roy Soc Lond 2004;A460:2147–51.
[10] Beven K, Lamb R, Quinn P, Romanowicz R, Freer J. TOPMODEL. In: Singh VP, editor. Computer models of watershed hydrology. Water Resources Publications; 1995. p. 627–68.
[11] Boyle D. Multicriteria calibration of hydrological models. PhD dissertation. Tucson, AZ: Department of Hydrology and Water Resources, University of Arizona; 2001.
[12] Butts MB, Payne JT, Kristensen M, Madsen H. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow prediction. J Hydrol 2004;298:242–66.
[13] Cameron D, Beven K, Naden P. Flood frequency estimation under climate change (with uncertainty). Hydrol Earth Syst Sci 2000;4(3):393–405.
[14] Carrera J, Neuman SP. Estimation of aquifer parameters under transient and steady-state conditions. Water Resour Res 1986;22(2):199–242.
[15] Dieterick BC, Lynch JA, Corbett ES. A calibration procedure using TOPMODEL to determine suitability for evaluating potential climate change effects on water yield. J Am Water Resour Assoc 1999;35(2):457–68.
[16] Duan Q, Ajami NK, Gao X, Sorooshian S. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Adv Water Resour 2007;30(5):1371–86.
[17] Fealy R, Sweeney J. Statistical downscaling of precipitation for a selection of sites in Ireland employing a generalised linear modelling approach. Int J Climatol 2007;27:2089–94.
[18] Fealy R, Sweeney J. Statistical downscaling of temperature, radiation and potential evapotranspiration to produce a multiple GCM ensemble mean for a selection of sites in Ireland. Irish Geogr 2008;41(1):1–27.
[19] Franchini M, Wendling J, Obled C, Todini E. Physical interpretation and sensitivity analysis of the TOPMODEL. J Hydrol 1996;175(1):293–338. 1002/joc.1506.
[20] Freer J, Beven K, Ambroise B. Bayesian uncertainty in runoff prediction and the value of data: an application of the GLUE approach. Water Resour Res 1996;32:2163–73.
[21] Giorgi F, Mearns LO. Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method. J Climate 2002;15:1141–58.
[22] Gupta H, Beven K, Wagener T. Model calibration and uncertainty estimation. In: Anderson MG, editor. Chichester, UK: John Wiley and Sons; 2003.
[23] Horton P, Schaefli B, Mezghani A, Hingray B, Musy A. Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty. Hydrol Process 2006;20:2091–109. doi:10.1002/ hyp.6197.
[24] Jiang T, Chen YD, Xu C-Y, Chen X, Chen X, Singh VP. Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China. J Hydrol 2007;336:316–33.
[25] Kuczera G, Parent E. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. J Hydrol 1998;211:69–85.
[26] Madsen H. Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. J Hydrol 2000;235:276–88.
[27] Melching C. Reliability estimation. In: Singh VP, editor. Computer models of watershed hydrology. Colorado, USA: Water Resource Publications; 1995. p. 69–118.
[28] Min SK, Simonis D, Hense A. Probabilistic climate change predictions applying Bayesian model averaging. Philos Trans Roy Soc A 2007;365:2103–16.
[29] Mitchell TD. Pattern scaling: an examination of the accuracy of the technique for describing future climates. Clim Change 2003;60:217–42.
[30] Montanari A. Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall–runoff simulations. Water Resour Res 2005;41:W08406. doi:10.1029/2004WR003826.
[31] Montanari A. What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology. Hydrol Process 2007;21:841–5.
[32] Murphy JM, Booth BBB, Collins M, Harris GR, Sexton DMH, Webb MJ. A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. Philos Trans Roy Soc A 2007;365:1993–2028.
[33] Najafi MR, Moradkhani H, Jung WI. Combined effect of global climate projection and hydrologic model uncertainties on the future changes of Streamflow. ASCE Conf Proc 2010;371(10):1. doi:10.1061/41114(371)10.
[34] Nemec J, Schaake J. Sensitivity of water resource systems to climate variation. Hydrol Sci J 1982;27(3):327–43.
[35] New MG, Hulme M. Representing uncertainties in climate change scenarios: a Monte Carlo approach. Integr Assess 2000;1:203–13.
[36] Prudhomme C, Davies H. Comparison of different sources of uncertainty on climate change impact studies in Great Britain. In: Climatic and anthropogenic impacts on the variability of water resources, FRIEND, Paris. Paris: UNESCO; 2007. p. 183–90.
[37] Prudhomme C, Jakob D, Svensson C. Uncertainty and climate change impact on the flood regime of small UK catchments. J Hydrol 2003;277:1–23.
[38] Raftery AE, Gneiting T, Balabdaoui F, Polakowski M. Using Bayesian model averaging to calibrate forecast ensembles. Mon Weather Rev 2005;133:1155–74.
[39] Refsgaard JC, Van der Sluijs JP, Brown J, Van der Keur P. A framework for dealing with uncertainty due to model structure error. Adv Water Resour 2006;29:1586–97.
[40] Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt K, et al., editors. Climate change 2007: the physical science basis. Cambridge University Press; 2007.
[41] Steele-Dunne S, Lynch P, McGrath R, Semmler T, Wang S, Hanafin J, et al. The impacts of climate change on hydrology in Ireland. J Hydrol 2008;356: 28–45.
[42] Sugawara M. Tank model. In: Singh VP, editor. Computer models of watershed hydrology. Littleton, Co.: Water Res. Publ.; 1995. p. 165–214.
[43] Takeuchi K, Ao TQ, Ishidaira H. For hydro-environmental simulation of a large ungauged basin—introduction of block-wise use of TOPMODEL and Muskingum–Cunge method. Hydrol Sci J 1999;44(4):633–46.
[44] Tanakamaru H, Kadoya M. Effects of climate change on the regional hydrological cycle of Japan. IAHS Publ. No. 212; 1993.
[45] Taye MT, Ntegeka V, Ogiramoi NP, Willems P. Assessment of climate change impact on hydrological extremes in two source regions of the Nile River Basin. Hydrol Earth Syst Sci Discuss 2010;7:5441–65.
[46] Tebaldi C, Smith RL, Nychka D, Mearns LO. Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multi-model ensembles. J Climate 2005;18:1524–40.
[47] Vaze J, Post DA, Chiew FHS, Perraud JM, Viney NR, Teng J. Climate nonstationarity-validity of calibrated rainfall–runoff models for use in climate change studies. J Hydrol 2011;394(3–4):447–57.
[48] Viola F, Noto LV, Cannarozzo M, Loggia GLa. Daily streamflow prediction with uncertainty in ephemeral catchments using the GLUE methodology. Phys Chem Earth 2009;34:701–6.
[49] Vrugt JA, Ter Braak CJF, Diks CGH, Higdon D, Robinson BA, Hyman JM. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. Int J Nonlinear Sci Numer Simul 2009;10:273–88.

[50] Vrugt JA, Diks CGH, Clark MP. Ensemble bayesian model averaging using Markov chain Monte Carlo sampling. Environ Fluid Dyn 2008;8: 579–95.

[51] Vrugt JA, Gupta HV, Bastidas LA, Bouten W, Sorooshian S. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resour Res 2003;39(8):1201. doi:10.1029/2002WR001642.

[52] Wagener T, Boyle DP, Lees MJ, Wheater HS, Gupta HV, Sorooshian S. A framework for development and application of hydrological models. Hydrol Earth Syst Sci 2001;5(1):13–26.

[53] Wilby RL, Harris I. A framework for assessing uncertainties in climate change impacts: low flow scenarios for the River Thames, UK. Water Resour Res 2006;42:W02419. doi:10.1029/2005WR004065.

[54] Wilby RL, Wigley TML. Downscaling general circulation model output: a review of methods and limitations. Progr Phys Geogr 1997;21:530–48.

[55] Xiong L, Guo S. Effects of the catchment runoff coefficient on the performance of TOPMODEL in rainfall–runoff modelling. Hydrol Process 2004;18:1823–36. doi:10.1002/hyp.1449.

[56] Xu CY, Singh VP. Evaluation and generalization of radiation based methods for calculating evaporation. Hydrol Process 2000;14:339–49.