CHAPTER 34

THE ROLE OF HYDROLOGICAL MODELLING UNCERTAINTIES IN CLIMATE CHANGE IMPACT ASSESSMENTS OF IRISH RIVER CATCHMENTS

SATISH BASTOLA, CONOR MURPHY and JOHN SWEENEY

First published in *Advances in Water Resources* (2011), 34(5), 562–576.

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 the 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.

Output 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 (Prudhomme et al., 2003). Utilization of information from different models has been widely used to address these uncertainties. Giorgi and Mearns (2002) 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 (2005) extended the REA method and proposed a Bayesian statistical model that combines information from a multimodal 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., Solomon et al., 2007), while several others have used the output from perturbed physics ensembles to evaluate the uncertainties arising from GCM model formulation (e.g.,Murphy et al. 2007).

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 output needed for impact studies (Wilby and Wigley, 1997). 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 (Wilby and Wigley, 1997, Khan et al., 2009).

In an attempt to quantify major sources of uncertainties associated with climate change impact assessment, New and Hulme (2000) presented an approach to quantifying uncertainites associated with the estimation of future greenhouse gas emissions, the climate sensitivity, and limitations and unpredictability in GCMs. Similarly, Horton et al. (2006) 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 utilized 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. Furthermore, experiences 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 follow: 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 employed and methods employed to account for the different uncertainties that are associated with studying the impact of climate change on water resource. Results are outlined in section 4.

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 conceptualize 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 (Beven, 2000; Gupta et al., 2003). 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 model and various methods for assessing the uncertainty can be found in Melching (1995). The climate change/hydrological modelling literature has mostly focused on the prediction uncertainty arising from model parameters (Kuczera and Parent, 1998; Steele-Dunne et al. 2009), 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., Carrera and Neuman 1986). 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., Jiang et al., 2007).

To examine the impact of model structure error and complexity on model performance and modelling uncertainty, Butts et al., 2004 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 to consider model parametric and structural uncertainty. A review on the range of strategies for assessing structural uncertainties in environmental modelling is available in Refsgaard et al., 2006. 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.

Methods for assessing uncertainty

Among various methods for assessing the uncertainty of hydrological models, the Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley, 1992) has been extensively used (e.g. Freer et al., 2004). 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 (Beven and Binley, 1992; Freer et al., 1996; Montanari, 2005). The major output of the GLUE method for assessing uncertainty is the prediction interval at each time step bounded by the lower prediction 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., Montanari, 2005). 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 (Raftery et al. 2005). The application of BMA is growing in a multimodel ensemble of AOGCMs to produce mean and probabilistic climate change projections (e.g., Tebaldi et al., 2005; Min et al., 2007). In BMA the predictive probability density function (PDF) of any quantity of interest is a weighted average of PDFs centered 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. Duan et al., 2007 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. Dual et al showed that the BMA scheme has the advantage of generating more skilful and equally reliable probabilistic predictions than the original ensemble.

Methodology

Study region and data

The area of focus for this study is 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 one provides an overview of key catchment descriptors.

Figure 1. Location of case study catchments

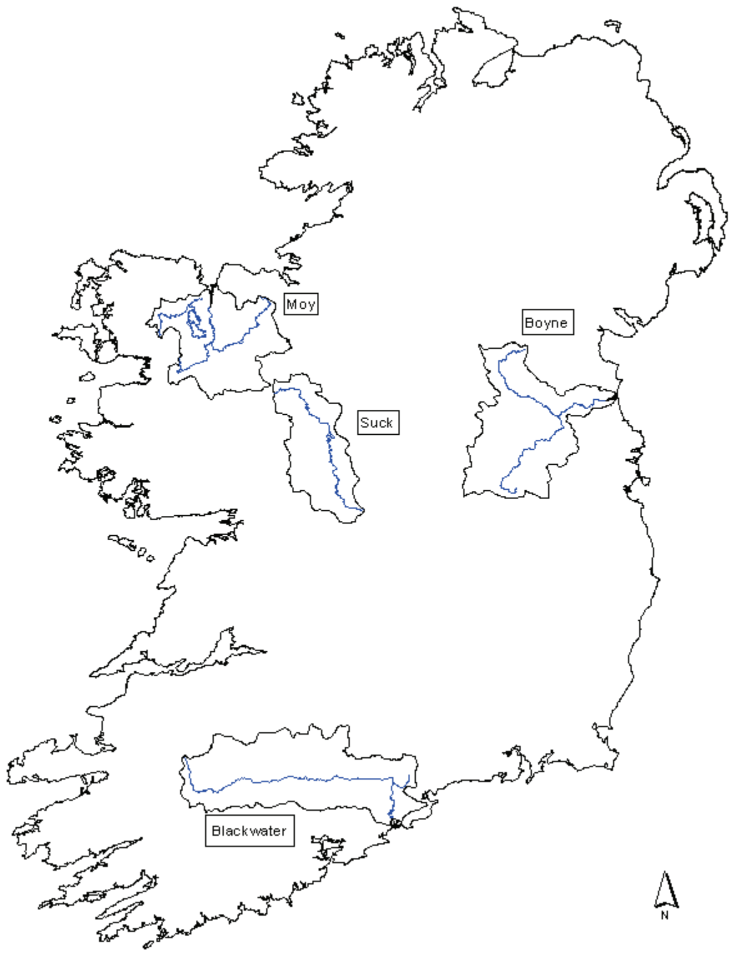


Table 1. Catchment Descriptors

River & Gauge	*Area (km^2)*	*S1085*	*SAAR (mm)*	*SAAPE (mm)*	*Forest %*	*Peat %*	*Pasture %*	*BFI 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

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 (2007) 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 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 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 (Mitchell, 2003), a primary input to hydrological models. Therefore in this study we only utilized the time series of downscaled data derived from three GCMs forced with two scenarios (e.g. A2, B2). The future potential evapotranspiration used is not a direct output of GCM, but is estimated based upon present climate using Hargreaves method, a radiation based empirical model popularly used for the simulation of potential evapotranspiration, for each of the GCMs. Furthermore, observed stream flow data from the Office of Public Works (available at http://www.opw.ie/hydro/), and observed precipitation and temperature data from Met Éireann, the Irish National Meteorological Service were used.

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 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 et al. (1995), and a review of TOPMODEL applications can be found in Beven (1997). It is referred as TOP hereafter interchangeably.

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 characterized 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 (2000).

c). The HYdrologic MODel (HYMOD); also a conceptual and lumped model, was originally proposed by Boyle (2001) 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., Wagener et al., 2001; Vrugt et al., 2003).

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 (Sugawara, 1995). A total of 15 parameters require to be estimated through model calibration

Each of these models varies in the way they conceptualize the key hydrological processes and in complexity, primarily related to the number of parameters requiring calibration. Among the four selected models, NAM and TANK describe the behaviour of each component of the hydrological cycle at the catchments level by using a group of conceptual elements. Conversely, TOPMODEL and HYMOD are both variable contributing area models. In TOPMODEL the spatial variability is taken into account through indices derived from topography whereas in HYMOD, the model spatial variability within basin is modelled using a probability distribution function. All four models employ a single linear reservoir to model groundwater.

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 Generalized Likelihood Uncertainty Estimation method (GLUE) and Bayesian Model Averaging (BMA).

<u>GLUE</u>

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 goodness-of-fit measure is based on the sum of squared errors (Eq. 1).

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

where $L(\theta_i | Y)$ is the likelihood measure for the *i*th model conditioned on the observations, Y, σ_i^2 is the error variance for the *i*th model (i.e. the combination of the model and the *i*th parameter set) and $\sigma_{obs}^{\,2}$ 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 entire selected hydrological models are ranked and likelihood weighted to characterize 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 the *n* numbers 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 (*N1* 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 i^{th} 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 is used to assign weight to each runoff prediction.
Step 5. Use the likelihood weights of the behavioural data set to assess parameter sensitivity and to compute prediction limits on hydrographs using:

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

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

The uncertainty bounds estimated by GLUE have been found to be sensitive to a number of factors such as the likelihood measure, and the threshold values employed (e.g., Viola et al., 2009). The increase in the value of *N1* will put more weight to the best simulation, thereby increasing the difference between good and bad solutions. Furthermore, if the threshold value grows, the width of the uncertainty bounds and percentage of data captured by prediction limits will decrease. Thus the choice of the threshold value is important since it strongly influences the size of the uncertainty bounds.

<u>BMA</u>
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 equation (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.

$$p(\Delta \mid f_1,..,f_K,D) = \sum_{k=1}^{K} p(\Delta \mid f_k, D)p(f_k \mid D) \quad (3)$$

The posterior model probability, $p(f_k \mid D)$, of model f_k given the data, is given by equation (4).

$$p(f_k \mid D)\alpha P(D \mid f_k)P(f_k) \quad (4)$$

where the constant of proportionality is chosen so that the posterior model probabilities add up to one. The prior probability, $P(f_k)$, in Eqn. (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:

$$E[\Delta \mid f_1....,f_k,D)] = \sum_{k=1}^{K} w_k \hat{\Delta}_k \quad (5)$$

$$Var[\Delta \mid f_1....,f_k,D)] = \sum_{k=1}^{K} (Var(\Delta \mid D, f_k) + \hat{\Delta}_k)w_k - E(\Delta \mid D)^2 \quad (6)$$

where $\hat{\Delta}_k = E(\Delta \mid D, f_k)$ (Raftery, 1993). 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 with heteroscedastic variance. 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., BMA weights and variances, were obtained from historical stream flow data (1971-1990) using Markov Chain Monte Carlo (MCMC) sampling. In this study, MCMC sampling was done with the Differential Evolution Adaptive Metropolis (DREAM) algorithm developed by Vrugt et al. (2008).

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 $\hat{y}_{i,k}\,(i=1,2,..N;k=1,2...K)$.

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

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

Step 5 Probabilistic predictions are maid using mean (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 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. 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 penalize 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.

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 scenarios, evaluated from four hydrological models and their behavioural parameters sets was used. The results are presented for three bench mark 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 (Murphy et al. 2004) 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 four CRR models (referred to as HYDRO) forced with the selected GCM and the scenario, b) eight median predictions estimated from the four 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)) 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 (1971-2000). 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.

Results

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 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 selections 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 lesser 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 basin.

This is done so that the 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.

Figure 2 Prediction intervals for Boyne basin including median and observed flow produced from; a) HYMOD , b) NAM, c) TANK, d)TOP for the selected period (1981-1983)

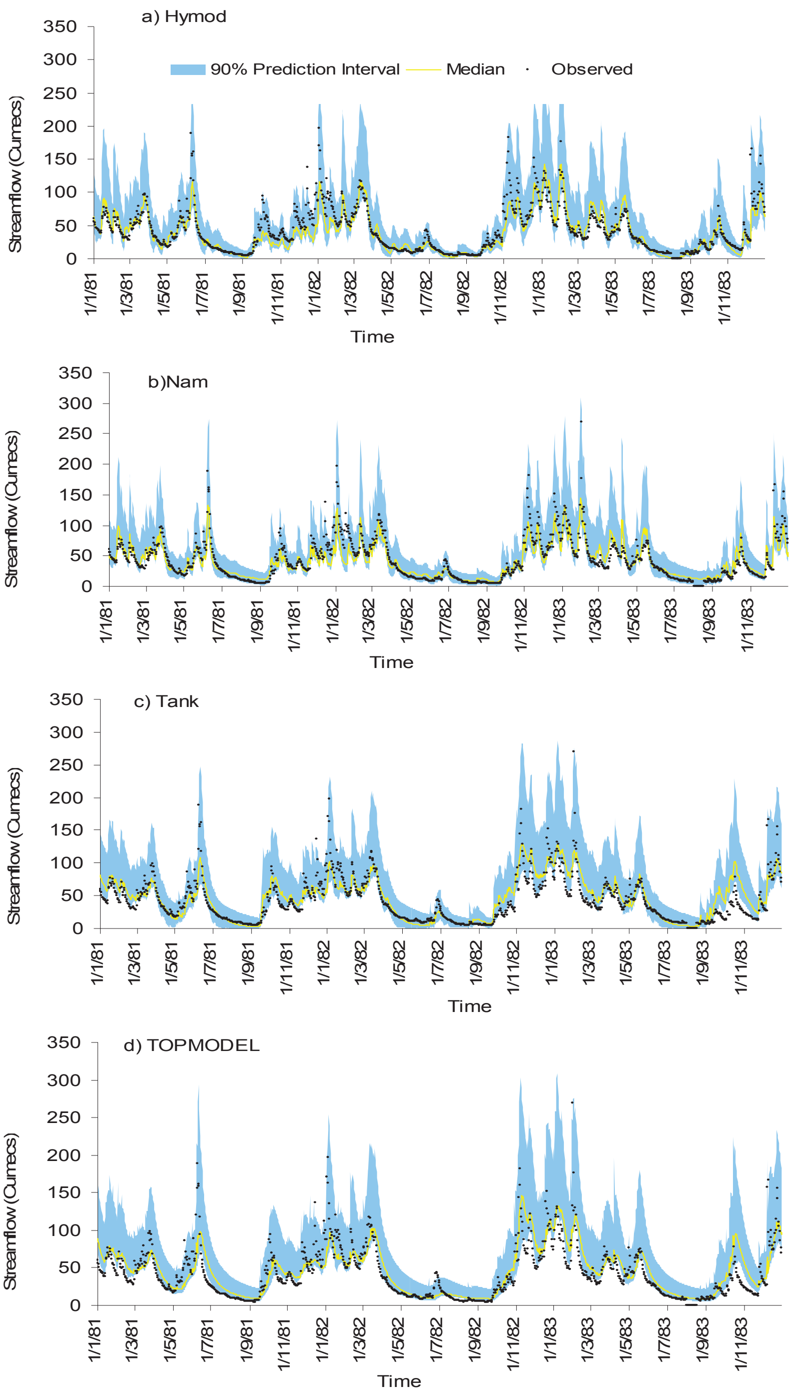


Table 2 The performance of median prediction, the percentage of observation encapsulated within the prediction interval (CE), and the average width of the 90% prediction interval (PI) for each model during the calibration and validation period

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	0.58	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	0.78	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		Blackwater (TOP)	0.64	0.71	0.33	0.31	21.77	22.69

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 conceptualization 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 utilized 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. The weight (Fig 3) and variance parameter of the BMA was estimated from 10 yrs of calibration data (1971-1981). The weight of HYMOD is apparently higher than that of the other three models.

Figure 3 Weight parameters for Bayesian Model Averaging (BMA) estimated from calibration period (1971-1990)

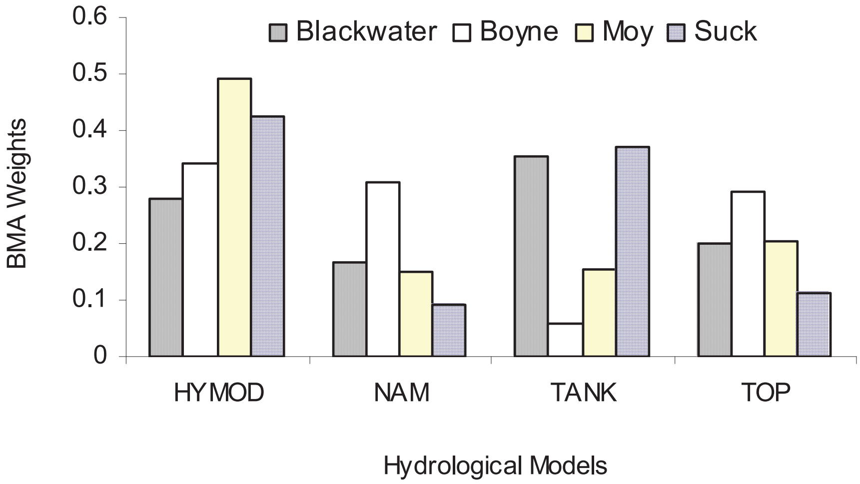
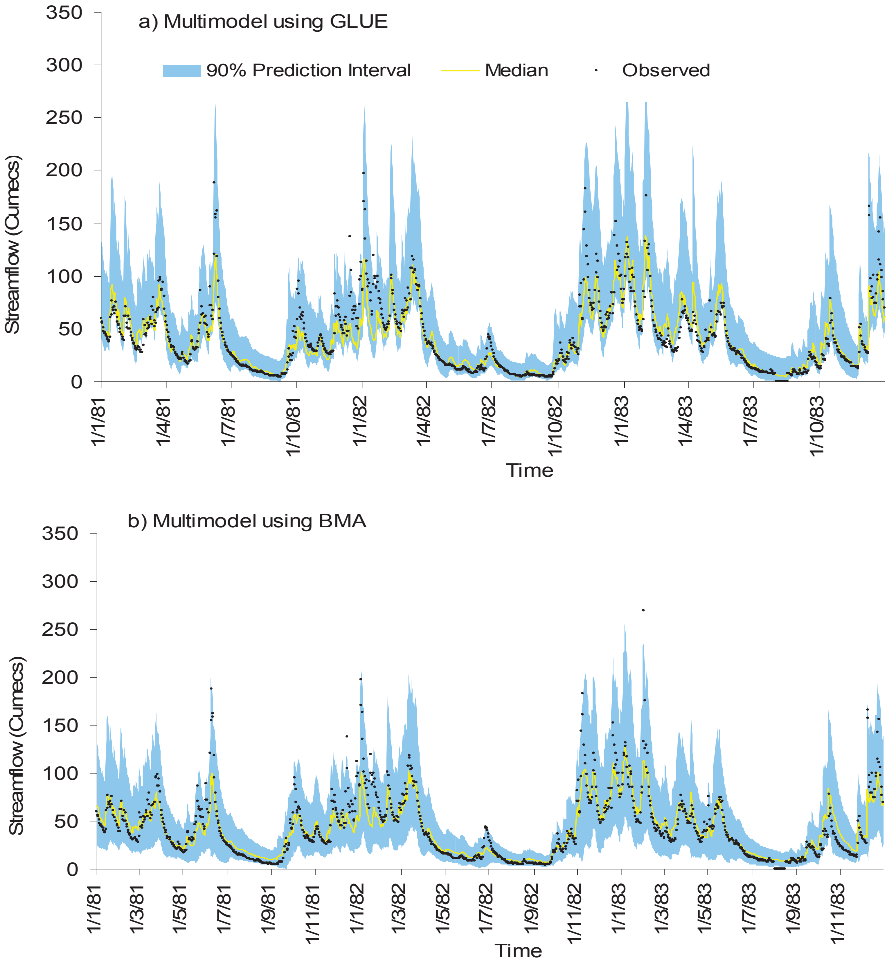


Fig 4 (a) shows the daily 90% PI (the results for only a three year is shown), derived using the GLUE, for the Boyne when all four models and their behavioural parameter sets are taken into account. The width of the prediction interval – expressed in terms of cumecs – increased when different models are considered. In addition, the count efficiency of the prediction interval improved when different model structures are incorporated. Fig 4(b) shows the daily PI for the Boyne when the median output from the 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.

Figure 4. Prediction interval for Boyne basin including median and observed flow produced from Multimodal ensemble of four selected models using a) Generalized Uncertainty Estimation Method (GLUE) and b) Bayesian Model Averaging (BMA) for the selected period (1981-1983)



The Table 3 shows that the PIs estimated from GLUE are narrower than the same obtained from BMA. Consequently, a larger proportion of observation 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 the 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.

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
4	Moy	BMA	0.90	0.79	0.97	0.92	64.03	66.76
5	Boyne		0.80	0.75	0.96	0.96	44.67	48.26
6	Suck		0.82	0.76	0.96	0.93	30.90	32.19
7	Blackwater		0.73	0.76	0.91	0.93	78.10	80.90

Contribution of CRR models to the envelope of future simulations

Fig. 5 shows the uncertainty in model prediction, expressed in terms of the Average Width of Prediction Interval (% of long term average flow) (AWPI) arising from uncertainties associated with parameterization of the hydrological model for the period from 1971 to 1990 and from 2050 to 2059, when forced with six regional climate scenarios. It reveals that the prediction uncertainties arising from parameterization depends upon the characteristics of 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 characterizing 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.

Figure 5. Average width of the prediction interval simulated from behavioural set of model parameters for four conceptual rainfall-runoff models for the period 1971-1990 and 2050-2059

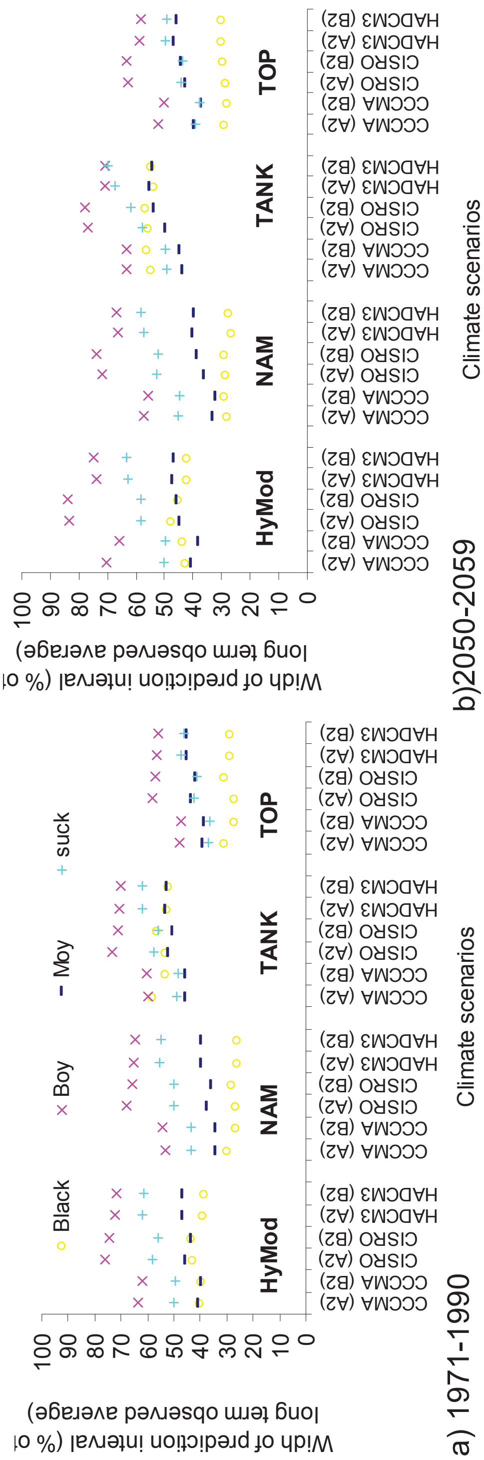


Fig 6 illustrates the response of the Boyne basin simulated by the behavioural parameter sets of four different hydrological models to climate scenarios derived from HADCM3 for the A2 scenario. The median estimates from the selected models are not significantly similar (α= 90%, two tail) with the exception of the median estimate from TOP and TANK. Consequently, the prediction intervals derived from GLUE are wider in comparison to the estimates from each individual CRR model.

<u>Figure 6. 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, 1971-1990 and 2050-2059 and median predictions from each individual hydrological model (continuous lines) for Boyne catchment.</u>

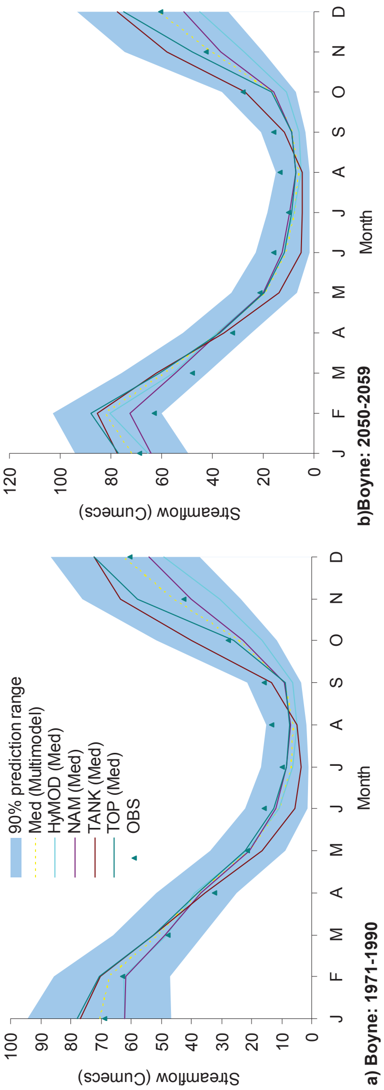


Fig 6 shows the seasonal prediction interval for the period from 1971 to 1990 and from 2050 to 2059. The prediction interval was constructed based on the behavioural predictions obtained from the models that were ranked and likelihood weighted to produce upper 95%, lower 5% and median 50% quantiles. For each basin, five different prediction intervals were calculated depending upon the sources of uncertainty included in the analysis. The 'hydrological model' scheme (referred as HYDRO hereafter) is assumed to quantify the uncertainty in hydrological model, the scheme 'scenario' (referred to as SCENE) is assumed to quantify the uncertainty arising from the selection of scenarios along with hydrological model, scheme GCM_A2 (GCM_B2) is assumed to quantify the uncertainty in GCM along with Hydrological model and the scheme 'TOTAL' is assumed to define the full consideration of impact model uncertainty.

Fig. 7 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 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 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 prediction interval nearly equalled the average flow when both scenarios are taken into account. It further grew to nearly 120% of the average flow when three GCMs with A2 (B2) scenarios are entertained, and nearly equalled to 140% of the average discharge when the total uncertainty is 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.

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. Wilby and Harris, 2006; Prudhomme and Davies, 2007). 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 8 shows the BMA estimates for five experiment design, namely for HYDRO, SCENE, GCM A2 (B2), and TOTAL (averaged across basin) for three time periods. The widths of the prediction interval (%) estimated from BMA are higher than the same estimated from GLUE. From the figure, it is apparent that the role of hydrological model uncertainty is considerable and warrants routine inclusion in impacts assessment, particularly where robust adaptation decisions are required.

Figure 7. The uncertainty in prediction for a) 2020s, b) 2050s and c) 2070s arising from 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.

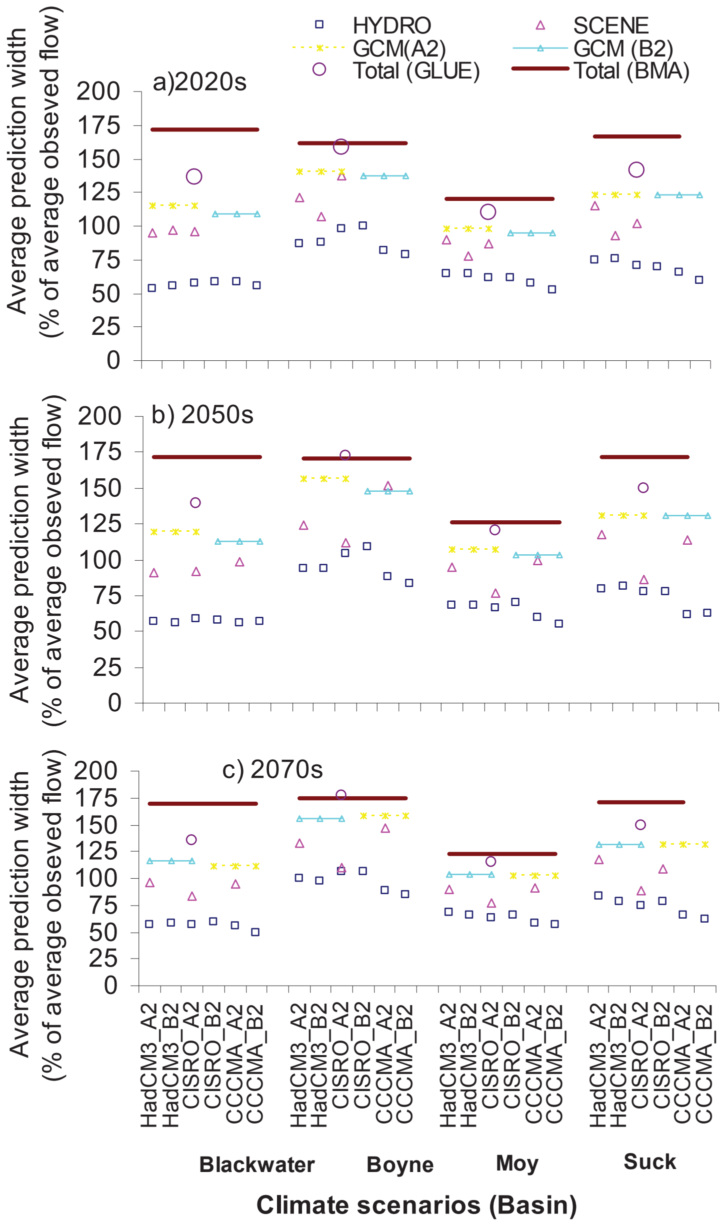
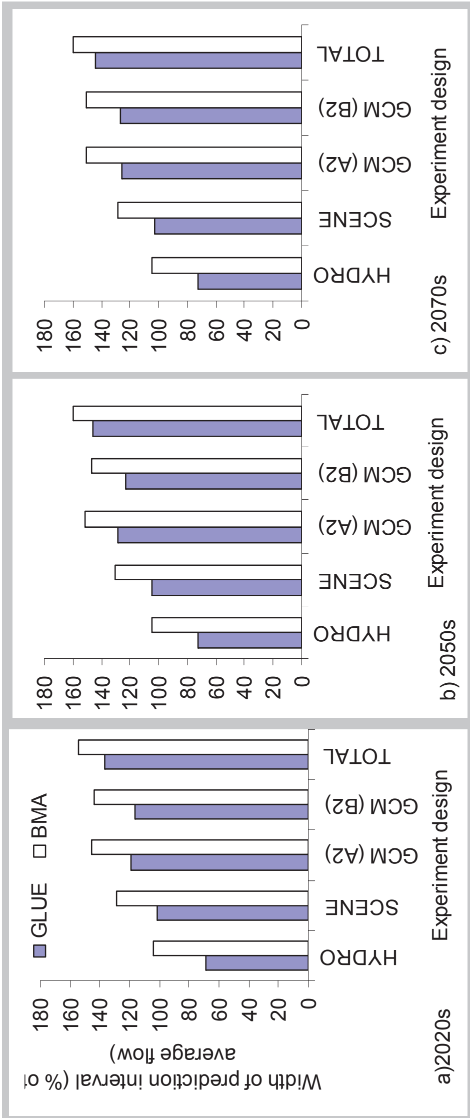


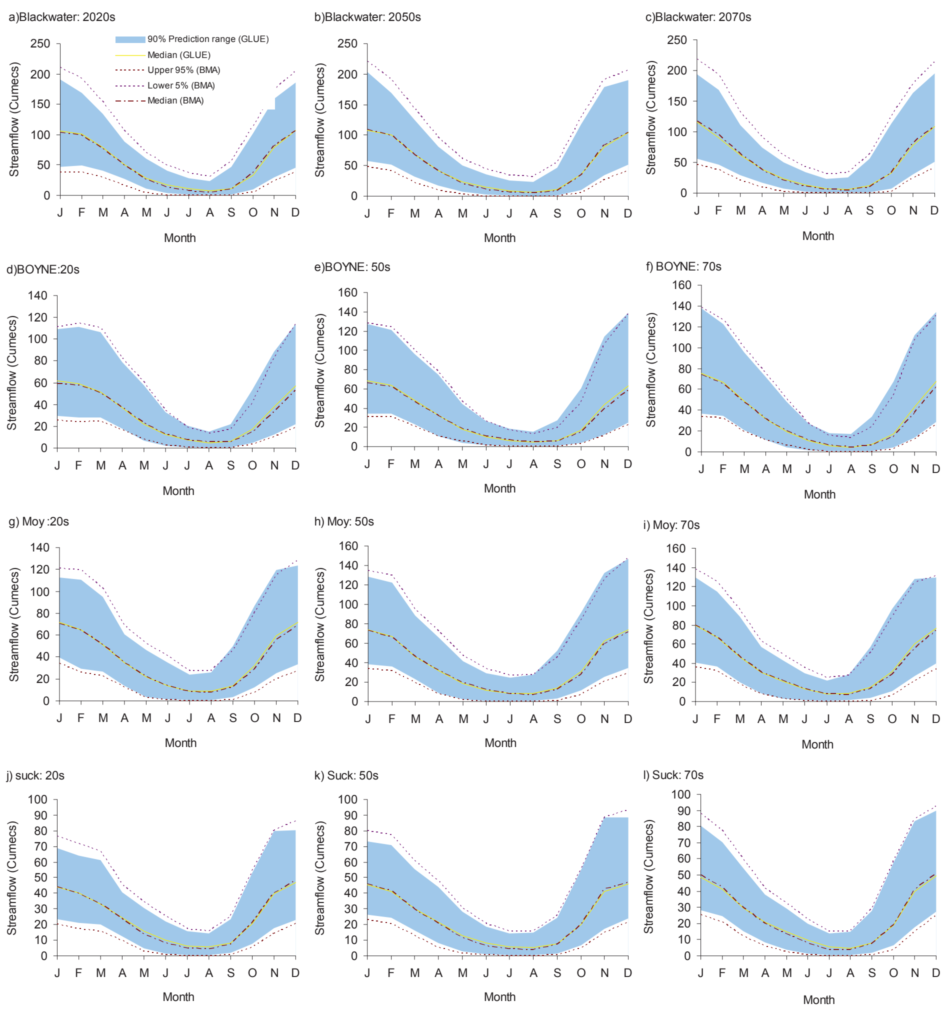
Figure 8. width of the Prediction interval (%), averaged across basin, estimated from Generalized Likelihood Uncertainty Estimation method (GLUE) and Bayesian Model Averaging (BMA) associated with various sources of uncertainties.



Similarly, Fig 9 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.

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 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 runoff coefficients (e.g., Nemec and Schaake, 1982).

Figure 9. Total uncertainty envelope derived from six climate scenarios and four hydrological models for the four study basins and for three time periods using Generalized Likelihood Uncertainty Estimation method (GLUE) and Bayesian Model Averaging (BMA).



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 characterized 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.

Figure 10 Percentage change in monthly streamflow, median perdiction derived from hydrological models forced with downscaled outut from three global climate models, and two emission scenarios using Generalized Likelyhood Uncertaity Estimation method, in the Blackwater, Boyne, Moy and Suck catchments for three future time periods, namely,2020s, 2050s and 2070s

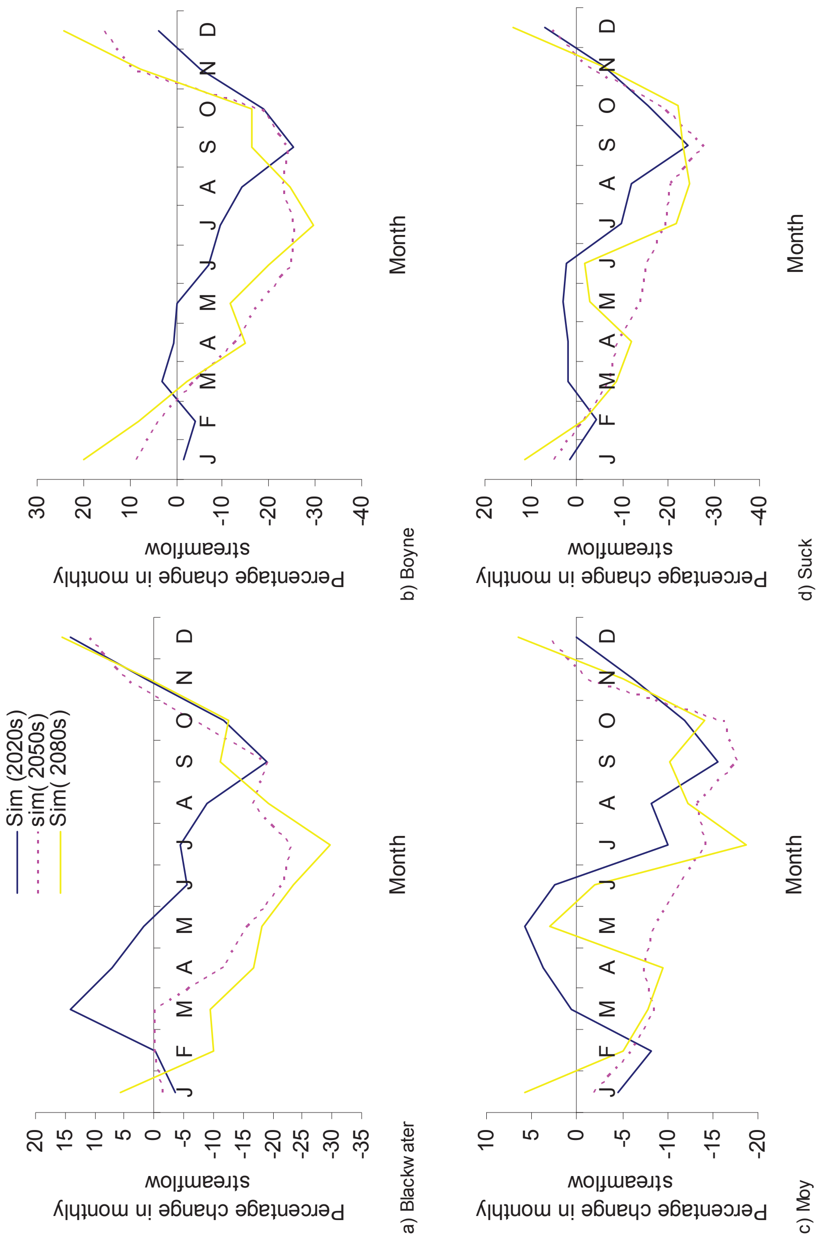


Fig 10 shows the percentage change in simulated monthly flow regime for each future time period. For all the basins, there is a tendency of an increase of flow in winter and decrease of flow in summer when moving from the 2020s to the 2070s and similar decreases in summer discharges as the century progresses, with associated implications for water management.

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 for 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 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 adapt to climate change.

The uncertainties derived from the use of behavioural model parameters for each model is nearly 40% of the average river flow. 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. Furthermore, the 90% confidence prediction interval 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 of 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 (e.g., Blasone et al., 2008).

The same tool was further used to identify the role of the uncertainty in the hydrological models in the overall uncertainty envelopes by utilizing 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 was nearly 50% of the average river flow. However, it increased to 70%, 100%, 120% and 140% respectively when uncertainty in CRR model structure, emissions scenarios, GCM and total uncertainty were accounted for.

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. BMA probabilistic predictions were made by combining 24 median predictions (from GLUE) obtained from six climate scenarios and four hydrological 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.

Acknowledgement. The authors are grateful for the financial support from Science Foundation Ireland's Research Frontiers Programme (RFP) under the project title 'Quantifying the cascade of uncertainty in climate change impacts for the water sector'.

References

Ajami, N.K., Duan, Q. and Sorooshian, S. (2007) An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, 43(1), W01403.

Beven, K. and Binley, A.M (1992) The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes*, 6, 279–298.

Beven, K., Lamb, R., Quinn, P., Romanowicz, R. and Freer, J. (1995) TOPMODEL In: Singh, V.P. ed. *Computer Models of Watershed Hydrology*. Highlands Ranch. CO: Water Resources publications, 627–668.

Beven, K (1997)TOPMODEL: a critique, *Hydrological Processes* 11, 1069–1085.

Beven, K (2000) *Rainfall-runoff modelling – The primer*. Chicester, UK: John Wiley & Sons.

Boyle, D. (2001) 'Multicriteria calibration of hydrological models'. Unpublished Ph.D. thesis. Department of Geography, University of Arizona.

Butts M.B., Payne, J.T., Kristensen, M. and Madsen, H. (2004) An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow prediction. *Journal of Hydrology,* 298, 242–266.

Carrera, J. and Neuman, S.P (1986) Estimation of aquifer parameters under transient and seady-state conditions. *Water Resources Research*, 22(2), 199–242.

Fealy, R. and Sweeney, J. (2007) Statistical downscaling of precipitation for a selection of sites in Ireland employing a generalised linear modelling approach. *International Journal of Climatology*, 27(15), 2083–2094.

Freer, J., Beven, K. and Ambroise, B. (1996) Bayesian uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resources Research*, 32(7), 2161–2174.

Giorgi F. and Mearns, L.O. (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method. *Journal of Climate*, 15, 1141–1158.

Gupta, H., Beven, K., and Wagener, T (2003) In: Anderson, M.G. ed., *Model calibration and uncertainty estimation.* Chichester, UK: John Wiley and Sons.

Horton, P., Schaefli, B., Mezghani, A., Hingray, B. and Musy, A. (2006) Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty. *Hydrological Processes*, 20(10), 2091–2109.

Jiang, T., Yongqin, D.C., Chong-yu, X., Xiaohong, C., Xi, C. and Singh, V.P. (2007) Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China. *Journal of Hydrology*, 336(3–4), 316–333.

Kuczera, G. and Parent, E. (1998) Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology*, 211(1–4), 69–-85.

Madsen, H. (2000) Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. *Journal of Hydrology*, 235(3–4), 276–288.

Min S.K, Simonis, D. and Hense, A. (2007) Probabilistic climate change predictions applying Bayesian model averaging. *Philosophical Transactions of the Royal Society A*, 365(1857), 2103–2116.

Mitchell, T.D. (2003) Pattern Scaling: An examination of the accuracy of the technique for describing future climates. *Climatic Change*, 60(3), 217–242.

Montanari, A. (2004) Uncertainty assessment in rainfall-runoff modelling: a review. Proceedings of the workshop on Statistical and Mathematical Methods for Hydrological Analysis, Naples, Italy.

Murphy, J. M., Booth, B.B.B., Collins, M., Harris, G.R., Sexton, D.M.H. and Webb, M.J. (2007) A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philosophical Transactions of the Royal Society A*, 365(1857), 1993–2028.

New, M.G. and Hulme, M. (2000) Representing uncertainties in climate change scenarios: a Monte Carlo approach. *Integrated Assessment*, 1(3), 203–213.

Prudhomme, C., Jakob, D. and Svensson, C. (2003) Uncertainty and climate change impact on the flood regime of small UK catchments. *Journal of Hydrology*, 277(1–2), 1–23.

Prudhomme, C. and Davies, H. (2007) Comparison of different sources of uncertainty on climate change impact studies in Great Britain. *Technical Document in Hydrology – UNESCO Paris*, 80, 183–190.

Raftery, A.E, Madigan, D. and Hoeting, J.A. (1997) Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437), 179–19.

Refsgaard, J.C., Van der Sluijs, J.P., Brown, J. and Van der Keur, P. (2006) A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, 29(11), 1586–1597.

Solomon, S., Qin, D., Manning, M., Marquis, M., Averyt, K., Tignor, M.M.B., Miller Jr., H.L. and Chen, Z. eds. (2007) *Climate Change 2007: The Physical Science Basis*. Cambridge: Cambridge: University Press, 2007.

Steele-Dunne, S., Lynch, P., McGrath, R., Semmler, T., Wang, S., Hanafin, J. and Nolan, P. (2008) The impacts of climate change on hydrology in Ireland. *Journal of Hydrology*, 356(1–2), 28–45.

Sugawara, M. (1995) Tank model In: Singh, V.P. ed., *Computer models of watershed hydrology*. Colorado: Water Resources Publications. 165–214.

Tebaldi, C., Smith, R.L., Nychka, D. and Mearns, L.O. (2005) Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multi-model ensembles. *Journal of Climate* 18(10), 1524–1540.

Viola, F., Noto, L.V., Cannarozzo, M. and La Loggia, G. (2009) Daily streamflow prediction with uncertainty in ephemeral catchments using the GLUE methodology, *Physics and Chemistry of the Earth, Parts A/B/C*, 34(10–12), 701–706.

Vrugt, J.A., Diks C.G.H., Clark, M.P. (2008) Ensemble bayesian model averaging using Markov chain Monte Carlo sampling. Environmental Fluid Mechanics, 8(5–6), 579–595.

Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Higdon, D., Robinson, B.A. and Hyman, J.M. (2009) Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10, 273–290.

Wilby, R.L., Harris, I. (2006) A framework for assessing uncertainties in climate change impacts: low flow scenarios for the River Thames, UK. *Water Resources Research*, 42, W02419.

Wilby, R.L. and Wigley, T.M.L. (1997) Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography,* 21(4), 530–548.

Nemec, J. and Schaake, J. (1982) Sensitivity of water resource systems to climate variation. *Hydrological Sciences Journal*, 27(3), 327–343.