Investigating Uncertainty in Global Hydrology Modelling

Katie A Smith, BSc. MSc. (Dunelm)

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

October 2015

Abstract

As projections of future climate raise concerns over water availability and extreme hydrological events, global hydrology models are increasingly being employed to better understand our global water resources and how they may be affected by climate change. Being a relatively recent development in hydrological science, global hydrology modelling has not yet undergone the same level of assessment and evaluation as catchment scale hydrology modelling. Until now, global hydrology models have presented just one deterministic model output for use in scientific research. Multi-model ensembles have compared these outputs for different global models, but the uncertainties within individual models have yet to be understood.

This study demonstrates a rigorous uncertainty investigation of the 123 parameters within the Mac-PDM global hydrology model over 21 global river catchments. Mac-PDM was selected for its relative simplicity amongst global hydrology models, and its suitability for application using high performance computer clusters. A new version of the model, Mac-PDM.14 is provided, with updated soil and vegetation classifications. This model is then subjected to a 100,000 parameter realisation Generalised Likelihood Uncertainty Estimation (GLUE) experiment, requiring 40 days of high performance computing time, and outputting over 2Tb of data. The top performing model parameterisation from this experiment provides an annual average error of 47% when compared to observed records, a 45% improvement over the previous version of the model, Mac-PDM.09. The soil parameters (field and saturation capacity) are shown to be the most sensitive parameters in the model. Given the computational expense of such an experiment, smaller sample sizes of parameter realisations are explored. Whilst the top performing parameterisation in a sample size as small as 1,000 can perform almost as well as that from 100,000 parameterisations, the

number of good parameterisations is fewer, and the range of model uncertainty may therefore be significantly underestimated.

Mac-PDM.14 is shown to have a lower mean absolute relative error than all models involved in both the Water and Global Change (WATCH) project and the Inter-Sectoral Impacts Model Intercomparison Project (ISI-MIP). Parameter uncertainty is compared to model uncertainty, and the uncertainty range between the models within the WATCH and ISI-MIP projects is comparable to the parameter uncertainty within Mac-PDM.14. Catchment specific calibrations of the global hydrology model are explored, and it is demonstrated that the model performance is improved by 22 to 92%, for the Niger and the Yangtze respectively, with catchment specific parameter values over a global calibration. Approximate Bayesian Rejection is applied to explore the catchment specific parameter values that result in good parameter performance. Few trends can be identified from this analysis, which suggests that Mac-PDM may be over-parameterised. Catchment specific calibrations in both high latitude and arid to semi-arid regions show significant improvement over global calibration, which indicate a deficiency in model structure; the addition of a glacier component to Mac-PDM is recommended. Model calibrations are validated using the ISI-MIP forcing dataset, and the best model performance gives an error of 44%. This is a betterment on the performance with the WATCH forcing data used in calibration, and so implies that models not need to be recalibrated every time new forcing datasets are employed.

This research highlights that the performance of global hydrology models can be significantly improved by running a parameter uncertainty assessment, and that in catchment scale studies, catchment specific calibration should be carefully considered. Furthermore, the uncertainty within individual global hydrology models is an important consideration that should not be overlooked as these models are increasingly included in ensembles and interdisciplinary studies.

Acknowledgements

Firstly, I am grateful to the School of Geography, University of Nottingham for providing my three-year PhD studentship. I would like to thank my supervisors, Simon Gosling and Michèle Clarke, for their support throughout my three years at Nottingham. Simon's extensive knowledge of the field of hydrological modelling and of the Mac-PDM.09 model has been invaluable. Michèle has broadened my knowledge of hydrology on a global scale and has enabled me to witness the great Brahmaputra River first hand! I would also like to thank many other members of the School of Geography's staff including Richard Field, Nick Mount, Suzanne McGowan, Chris Lavers and Charles Watkins for their open doors, as well as Ian Conway for all his help.

This PhD would not have been possible without the staff at the Global Rivers Data Centre for their substantial work compiling such a huge database, and also the USGS and the BWDB for their additional data. Furthermore, the work of all those involved in the WATCH and the ISI-MIP projects has been both central to this research, and incredibly inspiring. Furthermore, the exceptional facilities of the Nottingham High Performance Computer Cluster were integral to this research, and the support I received from Colin Bannister and his team was outstanding.

Thanks to all my wonderful friends at Nottingham, especially Laura, Flo, Alice, Heather, Jonathan and Sarah for making my time in the city unforgettable. Thanks also go to my climbing friends for swapping the chocolate for sport. My family deserve medals, as always, for their amazing support, and great thanks go to Alex for helping me through my final year.

Finally, thanks to the Centre for Ecology & Hydrology for hiring me, it's all been worthwhile!

Table of Contents

Table of Figures

Table of Tables

Chapter One:

Global Hydrology Modelling and Uncertainty

1.1 Introduction

Global Hydrology Models (GHMs) can help us understand the global hydrological cycle, and the ways in which it might be altered by climate change. The hydrological cycle is complex, made up of processes that drive the hydrosphere within the atmosphere, biosphere and lithosphere. These processes include precipitation, runoff, condensation, infiltration, interception, evaporation, transpiration, evapotranspiration, and groundwater flow. Studying the hydrological cycle is very difficult due to issues of both temporal and spatial scale. In addition, there are limitations in hydrological measurement techniques (Beven, 2012). Thus, models are required to provide a means of quantitative prediction that is required for decision making.

Hydrology models seek to describe the processes of the hydrological cycle as well as stores of water within the hydrological cycle. Dingman (2002) defines three types of model: "*Physical Models"* – which are tangible constructed representations of a portion of the natural world; "*Analog Models"* – which use observations of one process to simulate a physically analogous natural process. For example, the flow of electricity given by Ohm's Law is directly analogous to Darcy's Law of groundwater flow; and "*Mathematical Models"* – which are explicit sequential sets of equations and logical steps that convert numerical inputs, representing flow rates or states of storage, to numerical outputs, representing other flow rates of storage states. Dooge (1986) aptly refers to models, specifically simulation models, as a representation of a portion of the natural or human constructed world, "which is simpler than the prototype system and which can reproduce some but not all of the characteristics thereof."

Many modelling studies are carried out purely for research purposes, in order to gain better understanding of hydrological processes. It is through discrepancies in model output with observed data that model revisions are made, and hydrological understanding progresses.

However, the ultimate aim of hydrological modelling is to help inform decisions in water resource management, and to forecast and manage extreme events. This chapter will briefly review the history of hydrological science, and the development and progression of hydrology models. It will then introduce the issue of uncertainty, and how this influences modelling studies. Some examples of previous work that investigated uncertainties in hydrology modelling will be reviewed, before the research questions for this thesis are presented at the end the chapter.

1.2 Rainfall-Runoff Modelling

Hydrology models are often referred to as 'Rainfall-Runoff' models, since they use precipitation data to estimate runoff or river discharge. They vary hugely in complexity, from simple equations used to predict a single hydrograph peak, to extensive computer coded programs made up of suites of equations to describe sequences of hydrological processes. There are two broad categories of hydrological model: the simple "empirical" or "black box" models, those that seek to verify observations using past data, without much concern to the processes within the model; and the more complex "conceptual" or "physically-based" models which represent individual hydrological processes in series' of governing equations in an attempt to represent natural behaviour as we understand it. Both of these categories fall under the classification of "deterministic" models. Deterministic models are those which produce a fixed output, given a specific set of inputs, and have no random element to them. The converse of a deterministic model is a "stochastic" model, which does contain an element of randomness (e.g. when disaggregating precipitation data to a finer temporal resolution). Stochastic models may produce slightly different outputs even if the inputs are kept exactly the same. Many essentially deterministic models contain small stochastic subroutines such as this.

Most models use “variables” and “parameters” to drive them. Models seek to define a relationship between entities that are of interest. These entities are variables. Variables are generally time varying, and may be a model input, output, or the result of an equation that changes continuously over time. In rainfall runoff models, rainfall and runoff are variables, as is potential evapotranspiration which is the result of several process equations. Parameters however, are usually constants. They are values that are required within equations of the model in order to derive the relationships between variables. They are sometimes, physically meaningful, but also sometimes statistical scaling factors that do not have a tangible meaning. In rainfall runoff models, parameters may represent factors such as the height of a specific vegetation type, or the temperature at which snow begins to melt.

1.3 Hydrology Modelling: A Historic Review

A diagrammatic representation of the history of hydrology modelling, and the introduction of global hydrology models is shown in Figure 1.1. The origins of hydrological modelling can be traced back to the work of Mulvany (1851). Thomas James Mulvany developed the ‘rational method’, a simple equation which was used to predict the peak of a hydrograph. This was followed by the ‘event model’ by Édouard Imbeaux (1892) which was perhaps the first attempt to produce a distributed hydrological model. Imbeaux divided the Durance River in France into zones, and then estimated the travel time for the runoff from each zone to the outlet to produce a prediction of the hydrograph. This time-area concept was advanced in 1932 by Sherman who developed the “unitgraph”, which later became the ‘unit hydrograph’ (Sherman, 1949), and is still popular today.

The unit hydrograph is a simple method that does not require the division of the catchment into different time increments, but instead uses a transfer function to relate effective rainfall to total catchment runoff response in a unit of time.

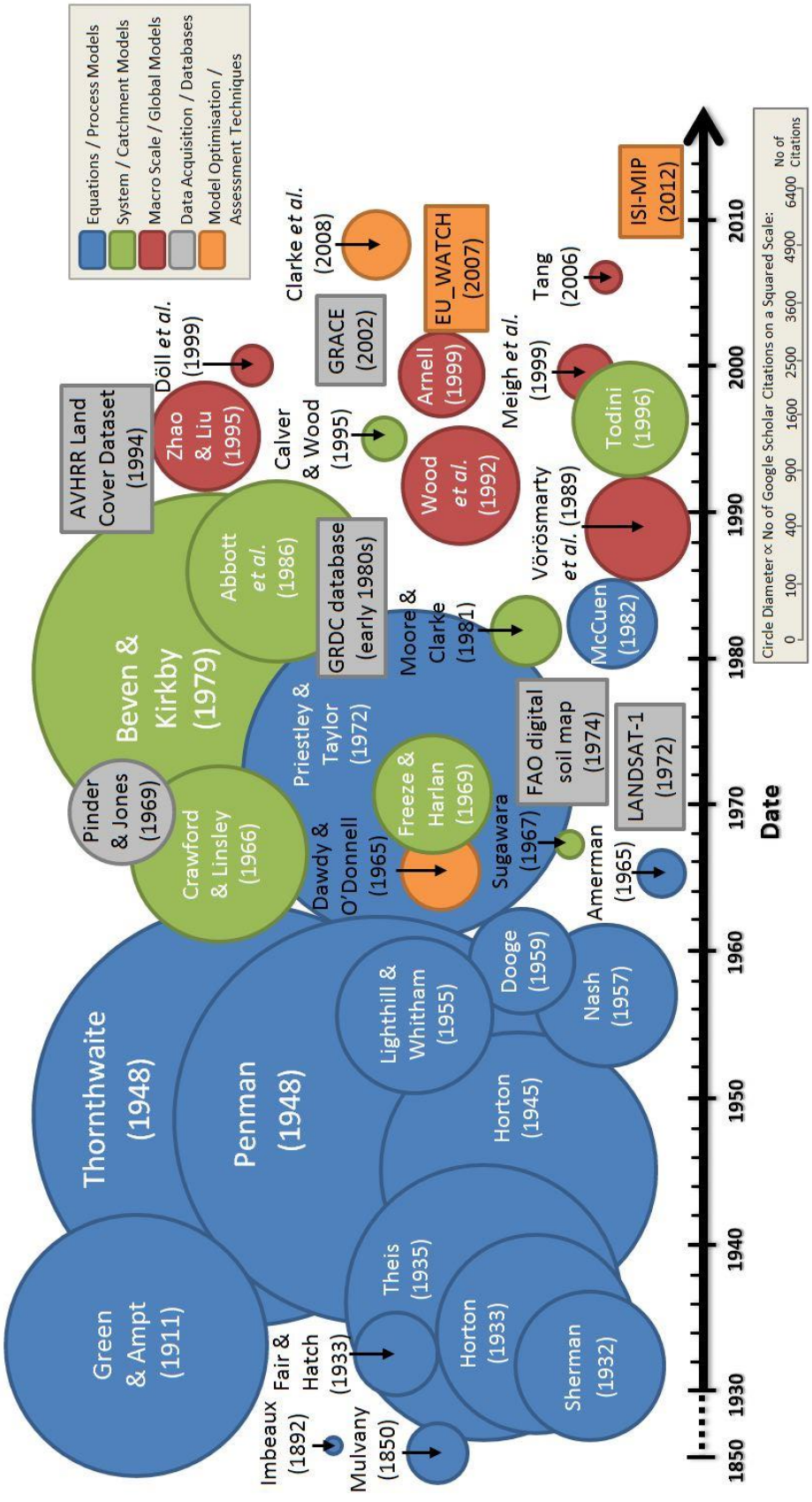


Figure 1.1 Select key papers and databases in the history of hydrological research. *Circle diameter is proportional to the number of paper citations. Rectangles indicate advances that cannot be associated with specific academic papers.*

However, the unit hydrograph method had a problem of linearity, as routing rainfall is a nonlinear problem that is influenced by rainfall intensity, soil properties, and antecedent conditions (Beven, 2012). Just a year after Sherman's paper, Robert Horton published a paper that went some way toward tackling this problem, which is still an issue in today's models. Horton (1933) developed a theory of infiltration to estimate rainfall excess and improve hydrograph separation techniques (Singh and Woolhiser, 2002). Horton's work on infiltration was preceded by the very well-known and still popular formula of Green and Ampt (1911). Horton is most famous for his final paper in 1945, which built upon his concept of infiltration excess overland flow, now known as "Hortonian overland flow" (Horton, 1945).

Alongside these works, Fair and Hatch (1933) developed a relationship to describe the permeability of soil, and Theis (1935) related heat-flow equations to groundwater problems (Kasenow, 2001). Theis' paper became the foundations of groundwater hydrology. Evapotranspiration was tackled by two equally popular papers by Thornthwaite (1948) and Penman (1948). Penman's work continued to develop the combined Penman-Monteith equation (Penman, 1956, Monteith, 1965, Allen et al., 1998), the modification of which is one of the two most popular potential evapotranspiration equations used in models today. The other is the simpler Priestley-Taylor equation (Priestley and Taylor, 1972) which requires less observational data. The mid-1950's saw some significant advances in hydrological research. The mathematicians Lighthill and Whitham (1955) established kinematic wave theory for flow routing in long rivers. Also applicable to traffic on long roads, this theory is a foundational mathematical development and is now a standard tool for modelling overland flow and other hydrologic processes (Singh and Woolhiser, 2002). Nash (1957) proposed the "instantaneous unit hydrograph" and Dooge (1959) developed the "generalized unit hydrograph". In 1965, Amerman introduced the "Unit Source Area" concept. This conceptual model categorises similar areas of a

catchment, by overlaying spatial databases in Geographic Information Systems (GIS), and assumes that if they are sufficiently alike, that they will respond in a similar manner. “Unit Source Areas” are now referred to as “Hydrologic Response Units” (HRUs), and are used in the popular catchment model SWAT (Spruill et al., 2000). Also in 1965, Dawdy and O’Donnell introduced model parameter optimisation in the need to progress away from trial and error approaches (Todini, 2007), this was a major advance in model assessment techniques.

1966 saw the arrival of the first model that attempted to simulate the hydrological cycle holistically; it was the Stanford Watershed Model (SWM), and is documented in Crawford and Linsley (1966). This was closely followed by the semi-distributed “tank” models by Sugawara (1967), and the work of Freeze and Harlan (1969) who conceived a three dimensional catchment model, which included all of the key hydrological processes such as precipitation, surface runoff, channel flow and their interactions with groundwater, evaporation, transpiration and more. This model was beyond the computational capabilities of the time, but became the foundations of the Système Hydrologique Europeen (SHE) model (Abbott et al., 1986).

The late 1960’s marked the beginning of a series of important advances in data measurement techniques and database releases. Tracers were first used to improve understanding of rainfall runoff processes by Pinder and Jones (1969). LANDSAT-1, the first civilian satellite to conduct scientific and exploratory studies of the Earth’s surface, was launched in 1972, and the FAO-UNESCO digital soil map of the world was released in 1974. The first 1-degree resolution land cover map was not released until 1994 however (the AVHRR Global Land Cover Dataset). The GRACE twin satellites were launched in 2002; this project uses gravity measurements to derive several indicators of hydrological mass balance, and has proven very valuable to hydrological research.

Research interest in catchment modelling increased dramatically in the 1980s, after TOPMODEL was released by Beven and Kirkby (1979). The SHE (Abbott et al., 1986), IHDM (Calver et al., 1995) and ARNO (Todini, 1996) models which were released throughout the 1980's and 1990's, are also popular catchment models, but have been significantly revised since their initial development. Rapid advances in computational power have allowed for the development from small scale catchment hydrology models to regional and global models to become possible. Vörösmarty et al. (1989) developed the first conceptually based macro-scale model. Many Global Hydrology Models (GHMs) have been developed since then including VIC (Wood et al., 1992), Xinanjiang (Zhao and Liu, 1992), GWAVA (Meigh et al., 1999), WaterGAP (Döll et al., 1999), MacPDM (Arnell, 1999) and DBH (Tang, 2006).

1.4 Global Hydrology Models: A Comparison and Critique

Interest in global and macro-scale hydrology modelling has increased substantially since the first macro-scale Water Balance Model (WBM) was proposed by Vörösmarty et al. (1989). The increase in published works on global or macro-scale hydrology models is presented in Figure 1.2. Currently, there are 8 popular GHMs in hydrological research, as detailed in Table 1.1. In this section, each of the 8 hydrological models will be briefly introduced and their similarities and differences discussed. Land surface models (LSMs) and coupled biosphere-hydrology models will also be mentioned, though they are not an essential component of this review.

There are several distinguishing features of GHMs, which are outlined in Table 1.1. These include: the soil moisture, evaporation and snowmelt scheme that they use; whether they model both water and energy balance; whether they consider anthropogenic factors, or whether they model "naturalised flows" (flows not including abstractions,

reservoir implications etc.); their routing method, if any; and the spatial and temporal resolution on which they operate.

GHMs model continental scale river basins and most operate on a 0.5 by 0.5 degree longitude-latitude grid scale at a (quasi-) daily time step. Global hydrology models are driven by General Circulation Models' (global climate models or GCMs) output data that provides estimates of variables such as precipitation, temperature, relative humidity and wind speed. Since land surface is very important in the behaviour of the hydrological cycle, gridded data on vegetation cover and soil types are often also required. Model outputs vary depending on the model's objective and most models output a variety of hydrological indicators. All hydrological models output either runoff or river discharge but additional outputs depend on the model; for example, Mac-PDM.09 outputs purely hydrological data, including extreme flow values and flow duration curve statistics, whilst WaterGAP outputs more socially driven information such as water availability, water withdrawals and water exploitation index.

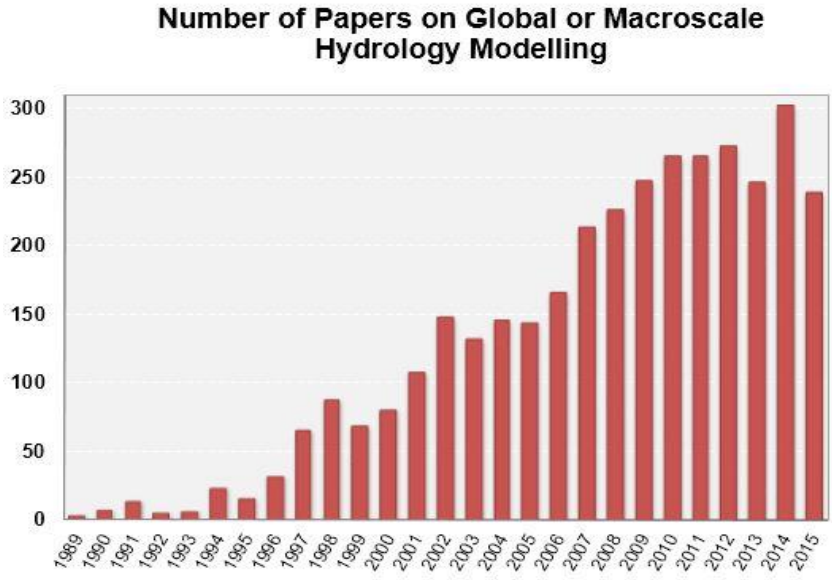


Figure 1.2 Number of papers published each year since 1989 under the search term *"Global hydrology" OR "Macroscale hydrology" AND model** in Google Scholar. *Search performed on 23rd January 2016.*

Table 1.1 Table of currently popular Global Hydrological Models

Model	Resolution	Time Step	Primary Motivation	Soil Moisture Scheme	ET Scheme	Snowmelt Scheme	Energy Balance?	Routing Method	Anthropogenic Factors	No of Parameters	Original References	Model Adaptations
GWAVA	0.1-0.5 deg	Daily	Water resources scarcity	PDM	Penman-Monteith	Degree-Day	No	Muskingham	Abstractions, reservoirs, aquifer use	5	Meigh et al (1999)	Dumont et al (2010), Dumont et al (2012)
H08	1 deg	6 hourly	Water use and water availability	Leaky Bucket	Bulk Equation	Energy Balance	Yes	Drainage Direction Map	Withdrawals from AQUASTAT, reservoirs	32	Hanasaki et al (2008a; 2008b)	Hanasaki et al (2010)
Mac-PDM	0.5 deg	Daily	Hydrological effects of climate change	PDM	Penman-Monteith	Degree-Day	No	None	None	15	Arnell (1999)	Gosling and Arnell (2011)
MPI-HM	0.5 deg	Daily	Hydrology model for a GCM	Improved Arno	Thronthwaite	Degree-Day	No	HD model	None	*	Hagemann and Dumenil (1998)	Hagemann and Dumenil Gates (2001), Hagemann and Gates (2003)
PCR-GLOBWB	0.5 deg	Daily	Hydrological effects of climate change	Improved Arno	Penman-Monteith	Energy Balance	Yes	Kinematic Wave	Reservoirs	44	Van Beek and Bierkens (2008)	Gruber et al (2011), Van Beek et al (2011)
Water GAP(2)	0.5 deg	Daily	Water use and water availability	After Bergstrom (1995)	Priestley-Taylor	Degree-Day	No	Drainage Direction Map	Irrigation, livestock. Households and industry, reservoirs	23	Alcamo et al (1997)	Doll et al (2003), Muller Schmied et al (2014)
WBM (plus)	0.5 deg / 30 min	Daily	Climate impacts on water balance	Bucket	Thornthwaite/ Hamon	Degree-Day	No	Muskingham Cunge	Irrigation, reservoirs	*	Vorosmary et al (1989)	Wisser et al (2010)
VIC	1/8 to 2 deg	Daily/ 3 hourly	Modelling of large scale river basins	Variable Infiltration Curve	Penman-Monteith	Energy-Balance	When snow is present	Drainage Direction Map	Irrigation, reservoirs	17	Liang et al (1994)	Liang et al (2003), Nijessen et al (2001a, 2001b), Andreadis et al (2009), Gao et al (2010)

* Literature includes multiple model configurations with varying numbers of parameters

1.4.1 WBM and WBM$_{plus}$

The Water Balance Model (WBM) was one of the first examples of a global hydrology model that existed separately to a GCM. First developed by Vörösmarty et al. (1989), WBM was designed as part of a global biogeochemistry study, and consisted of a water balance and water transport model. WBM originally operated on a 0.5° grid cell size, and used precipitation, temperature, potential evapotranspiration (PET), vegetation, soil and elevation data to predict soil moisture, evapotranspiration and runoff. WBM calculates soil moisture in a bucket scheme. During wet months, soil moisture can increase up to a maximum field capacity determined by the soil texture and rooting depth. During dry periods, soil water stocks are depleted according to a soil moisture retention function. For each wet month, soil moisture is calculated by incrementing antecedent values by the excess of available water over PET. This recharge may or may not be sufficient to bring the soil to field capacity at the end of the following wet season (Vörösmarty et al., 1989). Whenever field capacity is reached, excess water is transferred to subsurface pools, and runoff is generated as a linear function of the existing pool size (Vörösmarty et al., 1989). Vörösmarty et al. (1998) conducted an investigation into the potential evaporation functions used by the model, comparing 11 methods to determine their impact on predictions made by the global model. They determined that the simple empirical Hamon method was appropriate, whereas Thornthwaite had been used previously.

WBM has been developed into WBM$_{plus}$ (Vörösmarty et al., 1998, Rawlins et al., 2003, Federer et al., 2003) and is described in Wisser et al. (2010). WBM$_{plus}$ is a fully coupled water balance and transport model that simulates the vertical water exchange between the land surface and the atmosphere, and the horizontal water transport along a prescribed river network (Wisser et al., 2010). WBM$_{plus}$ can operate on a finer spatial resolution of 30 x 30 min (longitude x latitude), a daily

time step, and routes the runoff using the Muskingham-Cunge flood routing scheme. Compared to previous versions of WBM, WBM_{plus} includes modules that explicitly account for human activities such as irrigation water abstractions and reservoir operation that directly affect the water cycle process (Wisser et al., 2010). Döll et al. (2003) criticised the WBM model for the use of a correction factor in model validation, rather than parameter calibration. Fekete et al. (2004) suggest that WBM does not perform well in water-stressed semi-arid regions, where there is significant sensitivity of runoff to precipitation, and the processes are highly non-linear.

1.4.2 H08

H08 (Hanasaki et al., 2008a) is another model that uses a bucket soil moisture scheme, though in this case it is enhanced to a 'leaky bucket', where soil moisture can drain continuously, not just when the soil is at field capacity. The H08 leaky bucket is 15cm deep, uniformly across all vegetation and soil types. H08 was developed with the primary purpose to assess global water availability and use at a sub-annual timescale, thus H08 simulates both natural hydrological processes and major human activities related to water use. It consists of six sub models: land surface hydrology, river routing, crop growth, reservoir operation, water withdrawal, and environmental flow requirement (Chen et al., 2011). H08 is one of the few GHMs that simulates both energy and water balance on the land surface. Runoff is routed using the Total Runoff Integrating Pathways (TRIP) model, which provides a digital river map formed from flow direction. H08 includes a comprehensive crop module, which is similar to the Soil Water Integrated Model (SWIM), and simulates over 50 crop types. This model also simulates the operation of 452 reservoirs, totalling 4140km^3, each of which is ascribed its own operating rules which influences streamflow simulation. Environmental flow requirement is simulated in H08 as well as anthropogenic water withdrawal.

Hanasaki et al. (2008b) successfully applied the H08 model in a global water resources assessment. The model performed well in estimating the crop calendar and irrigation withdrawal and it highlighted regions of water stress that have been previously undetected. However, the original model assumes that water is only withdrawn from channels, excluding the significant abstractions from groundwater, lakes, ponds and glacial meltwater. Reservoirs smaller than 10^9 m^3 were also excluded from the model. Hanasaki et al. (2010) enhanced the model in order to estimate global virtual water flow. These changes included changing the spatial resolution from 1° to 0.5° longitude latitude, adding medium sized reservoirs ($3x10^6$ to $1x10^9$ m^3), and adding a conceptual water source, to represent deep groundwater, lakes, glaciers, water diversion and desalinization. This conceptual store (referred to as NNBW – non-local, non-renewable blue water) is however limitless, as the capacity of these sources is unknown, and is assumed to be available at all times in all places worldwide. This is unrealistic, but the process allowed for the comparison of geographical distribution of NNBW with estimated groundwater exploitation reports (Postel, 1999) for qualitative assessment and with aquifer withdrawal in the USA (Maupin and Barber, 2005) for quantitative assessment, which showed good results.

1.4.3 WaterGAP

The Water – Global Assessment and Prognosis (WaterGAP) model, sometimes referred to as the WaterGAP Global Hydrology Model (WGHM), is a water availability model that preceded H08. It was the first global model to compute both water use and availability on a basin scale (Alcamo et al., 2000), and the original model (WaterGAP1.0) is presented in Alcamo et al. (1997). The model takes into account basic socio-economic factors that lead to domestic, industrial and agricultural water use, and physical and climate factors that lead to river runoff and groundwater recharge (Alcamo et al., 1997). WaterGAP 1.0 was quickly

developed into WaterGAP 2.0 (Alcamo et al., 2000, Döll et al., 2003), WaterGAP 2.1 (Alcamo et al., 2003), and to the most recent WaterGAP 2.2 (Müller Schmied et al., 2014). These revisions are mostly synonymous with those of H08, and included algorithms of reservoir operation, groundwater recharge optimization, a variable flow velocity algorithm and consideration of the sources of water abstraction. Werth et al. (2009) also integrated water storage variation data from the GRACE satellite mission to reduce error in WaterGAP.

WaterGAP 2 consists of a global hydrology model and a global water use model which are linked in order to compute water stress indicators and to calculate the reduction of river discharge due to consumptive water use (Döll et al., 2003). The model operates on a 0.5° x 0.5° grid, forming 66896 cells worldwide, excluding Antarctica. Similarly to H08, runoff is routed using a global drainage direction map (DDM30). Soil moisture is modelled, taking into account the water content of the soil within the effective root zone, the effective precipitation, the actual evapotranspiration and the runoff from the land surface (Döll et al., 2003). Runoff is computed as a function of effective precipitation, and a calibrated runoff factor, which follows the approach of Bergström (1995) from the HBV model. With this approach, runoff increases with increasing soil wetness.

The water use part of the WaterGAP 2 model is divided into three sectors: domestic, industry and agriculture. The domestic and industry sectors take into account the effect of structural and technological changes on water use as a country develops, and the agricultural sector accounts for the effect of climate on irrigation water requirements (Alcamo et al., 2003). Alcamo et al. (1997) stated that despite calibration and testing against existing data, WaterGAP1.0 contains many limitations and so should only be used for the consideration of global scale trends, and not for individual watersheds. Whilst the model has progressed significantly since its initial development, WaterGAP 2

still needs work to improve its approach. WaterGAP 2 may be used to compare basins with catchment scale indicators such as total water withdrawals and total water availability. However, the reliability of modelled monthly flows needs to be improved to estimate critical high and low flow conditions (Alcamo et al., 2003).

1.4.4 GWAVA

GWAVA (the Global Water Availability Assessment model) is a third global hydrology model that is focused upon water use and availability, though it is used for prediction of water resources scarcity at continental and global scales (Dumont et al., 2010). Developed by Meigh et al. (1999), it applies monthly rainfall data to a probability distributed rainfall-runoff model (PDM, (Moore, 1985)) to generate monthly river flows and water availability statistics. The PDM model is utilised by many global hydrology models, as it allows a spatially variable distribution of soil moisture capacity described by a statistical probability distribution. This allows runoff to be generated in more than one part of a catchment, or grid cell, at any one time, rather than delaying runoff until the entire cell is saturated. This method is popular, as it enhances runoff production simulation without the requirement of additional data. However, the spatial allocation of soil moisture storage capacity is not influenced by vegetation or soil type. The PDM is described in detail in Moore (2007), and is also presented in Chapter 2 of this thesis.

Monthly water demands are estimated in GWAVA using population and per capita consumption data, combined with industrial and agricultural water requirements (Wallace and Gregory, 2002). GWAVA produces a water scarcity index, which is normalised to present -1 as little to no water to meet demands up to +1, representing more than sufficient water to meet demands. GWAVA takes groundwater into account as a water supply, and estimates groundwater availability as either a seasonally variable recharge or an aquifer yield (Wallace and Gregory, 2002). The model has been improved by Folwell and Farquharson

(2006) and Fung et al. (2006), and has been developed several times to include a water quality module (Dumont et al., 2010), a pollutant concentration module (Dumont et al., 2012). GWAVA has been applied to Eastern and Southern Africa, West Africa, the Caspian Sea Basin, South America, and the Ganges-Brahmaputra basin, and is currently being applied on a continental scale in Europe as well as globally (CEH, 2014).

1.4.5 Mac-PDM

The Macro-Scale – Probability Distributed Model (Mac-PDM) is a probability distributed model, as in GWAVA, that is designed to simulate the land surface hydrological dynamics of continental scale river basins with a daily water balance approach (Cloke and Hannah, 2011). It was first developed by Arnell (1999) as a simple macro-scale hydrological model which could be applied repeatedly over a large geographic domain without the need for calibration at the catchment scale. The model was significantly revised by Gosling and Arnell (2011) to produce the Mac-PDM.09 version of the model. These revisions included:

a) the ability to calculate average hydrological output from *n* model repetitions when forced with monthly data in order to account for model stochasticity;
b) the ability to read observed values of the coefficient of variation of daily rainfall, which was previously set as a constant 1.5, and;
c) the ability to read in daily climate data, rather than being forced by monthly data.

Mac-PDM.09 operates on a 0.5° x 0.5° grid and does not attempt to route runoff into channels. It therefore models runoff from two sources, quickflow and slowflow, in mm/day on the earth's surface, which if necessary can be converted runoff to catchment-wide discharge values for a given upstream contributing area. Mac-PDM also omits anthropogenic influences on hydrology, thus modelling 'naturalised' flows, that do not take account of water abstractions or reservoir

operational influences. The model uses gridded data for soil types and land cover classifications, and calculates evapotranspiration using the Penman-Monteith equation.

Despite the lack of anthropogenic consideration in Mac-PDM, the model has been applied in many studies of water availability (e.g. Fung et al., 2011) and water stress (e.g. New et al., 2011) which have used external calculations of water requirements and compared them to Mac-PDM runoff outputs. Mac-PDM has the advantage that it is a relatively simple model, and can therefore be used when running large ensembles of climate scenarios, such as those undertaken by Gosling et al. (2010) and Arnell and Gosling (2014).

1.4.6 MPI-HM

The Max-Planck Institute Hydrological discharge Model (MPI-HM) is another relatively simple macro-scale hydrology model. It consists of the simplified land surface (SL) scheme (Hagemann and Gates, 2003), which computes vertical water fluxes, and the hydrological discharge (HD) model (Hagemann and Gates, 2001), that globally simulates the lateral freshwater fluxes at the land surface (Chen et al., 2011). MPI-HM was developed by Hagemann and Dümenil (1997) in order to improve the hydrological balance module of the MPI ECHAM4 GCM. Several approaches to model structure for the HD model were explored by Hagemann and Dümenil (1997), the result of which was a three component model that uses runoff, drainage and grid cell inflow as inputs, to produce overland flow, baseflow and riverflow respectively. The sum of these three processes gives the outflow of the cell, which is routed using topography to create a flow direction map. The SL scheme is used to produce the inputs for the HD model, which comprises the main components of the hydrological cycle, including: separation of precipitation into rain and snow; snowmelt using the degree-day formula; potential evapotranspiration using the Thornthwaite formula;

and evapotranspiration, runoff, infiltration and drainage according to the ECHAM models (Hagemann and Gates, 2001).

Hagemann and Gates (2003) updated MPI-HM by developing an Improved Arno (IA) scheme to simulate soil moisture capacity. The original Arno scheme is very similar to the PDM, both of which assume that the soil water capacity distribution within the grid cell starts at zero and follows a continuous distribution defined by:

$$\frac{s}{S} = 1 - \left(1 - \frac{W_s}{W_{Smax}}\right)^b \qquad \text{(E1.1)}$$

where s/S is the percentage of the grid cell area S in which the soil water capacity is less than or equal to an assigned value W_s. W_{Smax} is the mean soil water capacity of a model grid cell. The parameter b defines the shape of the soil water capacity distribution curve.

The Improved Arno scheme adjusts this equation by allowing the specification of a minimum local (subgrid) soil water capacity W_{min} that is not necessarily zero. W_{max} is the maximum local soil water capacity, and W_{act} is the subgrid water content that corresponds to the fractional saturation of s/S of the grid cell, so that:

$$\frac{s}{S} = 1 - \left(\frac{w_{max} - w_{act}}{w_{max} - w_{min}}\right)^b \qquad \text{(E1.2)}$$

The Improved Arno scheme also allows the b parameter to be modified by an orographic shape parameter to account for the fact that on steep terrain, the probability of soil water capacities reaching saturation is higher (Hagemann and Gates, 2003).

1.4.7 PCR-GLOBWB

The PCRaster Global Water Balance Model (PCR-GLOBWB) also uses the Improved Arno scheme for soil moisture calculation. PCRaster (Wesseling et al., 1996) is the dynamic scripting language that the

model is coded in; it uses spatio-temporal operators with intrinsic functionality for constructing spatio-temporal models, and enables efficient manipulation of raster-based maps. Van Beek and Bierkens (2008) present the general outline of the model. Operating on a 0.5° x 0.5° grid scale, and daily time step, PCR-GLOBWB consists of two vertically stacked soil layers and an underlying groundwater layer, with subgrid parameterisation used to represent tall and short vegetation, surface water and the IA soil moisture scheme. Runoff is routed using kinematic wave theory, and the drainage network is taken from the drainage direction map DDM30 (Sperna Weiland et al., 2010). The model also calculates interception and snow storage (Wada et al., 2011).

Like Mac-PDM, PCR-GLOBWB calculates naturalised flows, and does not consider anthropogenic water use, however van Beek et al. (2011) adapted the model to include a surface water energy balance and reservoir operation scheme to calculate green and blue water availability. In part 2 of the same study, Wada et al. (2011) assessed global water stress at the monthly time scale. Wada et al. (2010) also used PCR-GLOBWB alongside the Global Groundwater Information System (GGIS) to estimate global depletion of groundwater resources. Gruber et al. (2011) added four regions with significant glacier mass to the model, in addition to the land masses of Greenland and Antarctica in a fluid mass motion experiment. Sperna Weiland et al. (2012) used PCR-GLOBWB to make a thorough global assessment of the effects of climate change on hydrological regimes and their associated uncertainties. Sperna Weiland et al. (2010), in their study on the usefulness of data from GCMs for hydrological studies, discovered that PCR-GLOBWB showed good results in comparison with observed river discharge data, however it performed less well in arid and mountainous areas.

1.4.8 VIC

The Variable Infiltration Capacity (VIC) model is perhaps the best known model in this review selection; however both VIC and H08 are technically Land Surface Models (LSM) rather than GHMs. Land surface models, which can be coupled to atmospheric models, tend to describe the vertical exchanges of heat, water and sometimes carbon in considerable details. GHMs however are traditionally more focused on water resources and lateral transfer of water (Haddeland et al., 2011). VIC has been extensively used for hydrological modelling, and has been included in the Water and Global Change (WATCH) project Water Model Intercomparison Project, as well as the Inter-Sectoral Impacts Model Intercomparison Project (discussed in section 1.6). VIC, originally developed by Liang et al. (1994), is a semi-distributed macro-scale model that balances both the water and surface energy budgets within a grid cell using a hybrid of physically based and conceptual components (Trambauer et al., 2013). VIC simulates sub-grid spatial variability in precipitation, land surface vegetation classes and soil infiltration capacity statistically (Nijssen et al., 2001b).

It has developed from a single layer, to a three or more soil layer model and has undergone several updates (Cherkauer et al., 2003, Bowling et al., 2004, Bowling and Lettenmaier, 2010). VIC been adapted to allow representation of water management effects (Haddeland et al., 2007, Haddeland et al., 2006a, Haddeland et al., 2006b, Zhao et al., 2013) including reservoir operation and irrigation diversions and return flows (Gao et al., 2010). The model can be run as either a water balance or a water and energy balance model, depending on the users' purpose. Running as just a water balance model simplifies the model and saves on computational expense. It uses the Penman-Monteith method of evapotranspiration calculation and the variable infiltration curve to account for the spatial heterogeneity of runoff generation, which follows the Arno conceptualisation (Gao et al., 2010). The runoff from each cell

is combined using a routing scheme to produce daily and accumulated monthly flows at selected points. The routing model allows for the explicit representation of reservoirs (Trambauer et al., 2013). Mishra et al. (2010) demonstrated that major historical drought events were successfully identified and reconstructed using VIC model simulations. Shrestha et al. (2013) showed that the VIC model performs well when run with observed climate data, however when driven with GCM-derived data, monthly maximum and minimum flow indicators showed significant differences with observed values, which raises a question on the capability of the model to predict extreme hydrological behaviour in the future.

1.4.9 Other Models

There are several other LSMs that have been applied in global hydrological research, including MATSIRO (Takata et al., 2003, Koirala et al., 2010), HTESSEL (Balsamo et al., 2009), ORCHIDEE (de Rosnay and Polcher, 1998), and JULES (Cox et al., 1999, Essery et al., 2003). All four of these models, along with VIC, have been included in the multi-model ensemble of Water and Global Change (WATCH) project (Haddeland et al., 2011), and all except HTESSEL have been used in the ISI-MIP fast track research project (Warszawski et al., 2014) (see section 1.7). Independently, MATSIRO has been applied in projecting global flood and drought risk with climate change (Hirabayashi et al., 2008), HTESSEL has been developed into a global flood alert system (Burek et al., 2012), and ORCHIDEE has been applied to simulate discharge in the Amazon (Guimberteau et al., 2012) and infiltration processes in west Africa (d'Orgeval et al., 2008). There are also coupled biosphere-hydrological models such as DBH, and WEB-DHM. Both DBH and WEB-DHM fully couple the biosphere scheme SiB2 with geomorphology based hydrological models (Wang et al., 2009b, Tang et al., 2007, Wang et al., 2009a).

Each of these different GHMs, as has been discussed, have different model structures, and so will output slightly different estimates of runoff or discharge from their simulations. This is due to the inherent uncertainty in hydrological modelling. All models - be they climate models, hydrological models or even economic models - are merely a set of equations attempting to express the behaviour and dynamics of the system under investigation. Therefore, as models are a manifestation of the authors understanding of this system, they are subject to errors, assumptions and uncertainties.

1.5 Uncertainty and its Origins

The word “uncertainty” is closely coupled with such negative terms as doubt, dubiety, scepticism, suspicion, mistrust and inconsistency (Kundzewicz, 1995). However, uncertainties are an important aspect of science, and needn’t be addressed with such negativity. Investigating, quantifying and presenting uncertainties can drastically improve our understanding of global change and can help relieve some of the scepticism surrounding modelling studies.

Firstly, an understanding of the origins of uncertainties is necessary. Figure 1.3 shows one classification of the differing types of uncertainty. According to Smith and Stern (2011), there are at least four varieties of uncertainty in studies of the impacts of global change:

Imprecision – or statistical uncertainty, is related to outcomes which we do not know precisely, but for which we believe robust, decision relevant probability statements can be provided;

Ambiguity – recognised ignorance, or scenario uncertainty, is related to outcomes for which we are not in a position to make probability statements;

Intractability – is related to computations known to be relevant to an outcome, but lying beyond the current mathematical or computational capacity to formulate or to execute faithfully;

Indeterminacy – is related to quantities relevant to policy-making for which no precise value exists. This applies, for instance, with respect to a model parameter that does not correspond to an actual physical quantity. It can also arise from the honest diversity of views among people, regarding the desirability of obtaining or avoiding a given outcome.

Imprecision is a challenge in communication. Science aims to quantify imprecision and reduce ambiguity, but there is not always a clear division between the two. Intractability makes reducing ambiguity difficult from technological constraints, and sadly indeterminacy involves seeking an answer that does not really exist.

In modelling studies, the origins of uncertainties can be visualised using the structure of a tree, with a dense network of roots and a broad canopy of leaves (see Figure 1.4). The roots of the tree represent all of the uncertain aspects that the modeller feeds into the model.

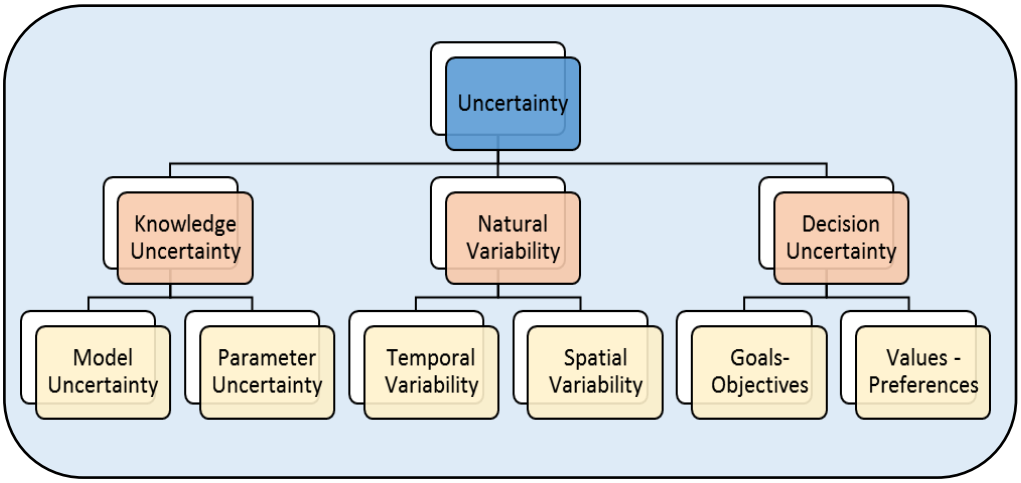


Figure 1.3 A classification of types of uncertainty. *After Loucks et al. (2005).*

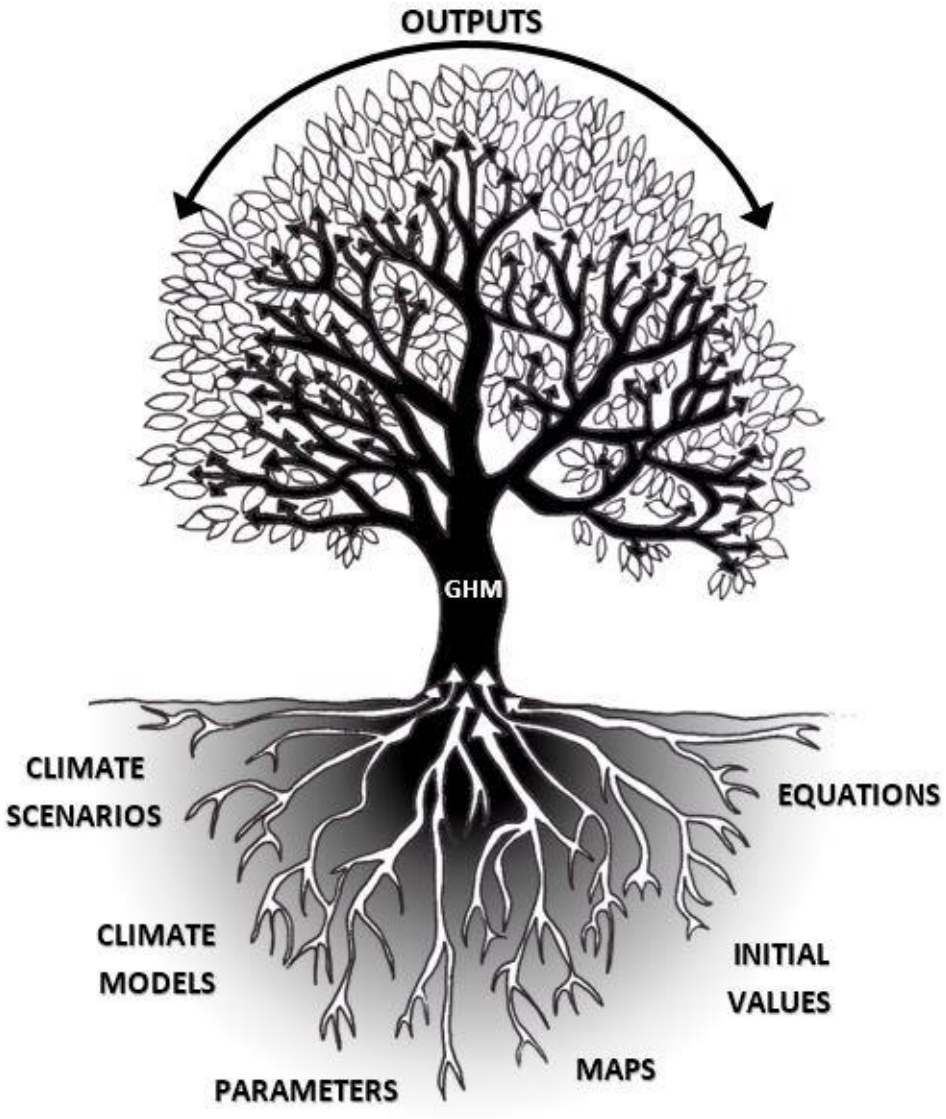


Figure 1.4 'The Modelling Tree'. A representation of the uncertainties within a single hydrological model.

In hydrology models, these things include the equations that govern the model (for example using the Penman-Monteith or the Priestley-Taylor potential evapotranspiration equations); the soil and vegetation maps; the values of parameters, such as vegetation height and field capacity; the choice of the climate model for input data; and, the climate scenario when projecting future change. Input data is a significant source of uncertainty, as different values for precipitation data and other climatological inputs will significantly impact the resultant runoff simulations. Input data for the past and present is referred to as "forcing data" and can be developed from observed records, individual climate

models, or ensemble climate model averages. Input data for climate projections is known as “driving data”.

All the uncertain aspects are fed into the model, are processed, and lead to a variation of model outputs. Each different decision made at the roots of the model tree will impact the output leading to a different result, and there are, therefore, a huge number of potential model “realisations” represented by the leaves of the tree.

As previously discussed, uncertainty has often been seen in a negative light (e.g. Pappenberger and Beven, 2006). In response to this, Juston et al. (2013) give seven reasons to be positive about uncertainty in hydrological modelling: 1- we learn about data, 2 - we learn about models, 3 - we produce more reliable and robust predictions, 4 - we learn about the value of additional data, 5 - we can engender trust by recognising and communicating uncertainties, 6 - we deepen academic understanding, and 7 - uncertainty estimation is getting easier.

1.6 Uncertainty Analyses in Hydrology Models

Recently, investigating the uncertainty in global hydrology models has received a lot of attention. To some extent, each of the sources of uncertainty discussed in section 1.5 has been considered in previous studies. However, research has largely focussed upon the model structure (structural uncertainty), which is addressed by comparing one model with another. Returning to the tree analogy, each tree represents a single hydrology model and the uncertainties within that model. In global hydrology research, as discussed in section 1.4, there are many models available, so there are many trees. This indicates the scale of the issue of uncertainty in global hydrology modelling; the potential number of model outputs is as broad as the number of leaves in a forest full of trees.

This section will review previous research on the uncertainties in global hydrology models. Starting with the work that has been done on multi-

model intercomparison projects (MIPs), this review will then go on to outline previous efforts of global hydrology model sensitivity analysis and parameter estimation experiments.

1.6.1 Multi-Model Ensembles and Model Intercomparison Projects

Studies investigating model structural uncertainty look at comparing different hydrology models with each other and often form a multi-model ensemble (MME). Multi-model intercomparison is a concept that has been implemented in climate science since 1990 when the Atmospheric Model Intercomparison Project (AMIP) was undertaken in order to provide a standard experimental protocol for atmospheric general circulation models. A framework was put forward for model diagnosis, validation and intercomparison (Tebaldi and Knutti, 2007). This was followed by the Coupled Model Intercomparison Project (CMIP, Meehl et al. (2000)), which has recently completed its 5th phase (CMIP5) with the World Climate Research Programme (Taylor et al., 2011). One example of a pioneering climate model ensemble project is that of climateprediction.net which was launched in 2003, and uses the idle computer power of participating members of the public to run ensembles of thousands of climate models with perturbed physics (adjusted parameters). Using this method, climateprediction.net have completed many projects, such as the BBC climate change project (Frame et al., 2009), and have many ongoing projects such as "weather@home", which will focus on how climate change may affect weather and the likelihood of extreme weather events. Other examples of climate MIPs are the EU ENSEMBLES project (of 2004-2009), and the ongoing QUMP (Quantifying Uncertainty in Model Predictions) which is run by the UK Met Office.

In the field of hydrology, the Hydrologic Ensemble Prediction Experiment (HEPEX) was launched in 2004 in order to explore a range of issues with hydrologic model uncertainty, including: sources of hydrological prediction errors; coupling meteorological model

ensembles with hydrology model ensembles; community use of ensemble forecasts; and how best to use ensembles as a decision making tool (Schaake et al., 2007). The same year, the Distributed Model Intercomparison Project was formulated to compare distributed models among themselves, and also to a lumped model (Smith et al., 2004). DMIP was run for the US, and included twelve catchment scale models, including VIC, SWAT and WATFLOOD. The project protocol specified three river catchments (the Elk, Illinois and Blue Rivers), provided the forcing data, catchment data (e.g. topography, soil texture and vegetation data), gauge data, and outlined the analysis strategy. The results of the project are published in a special edition of the Journal of Hydrology (vol. 298). One of the key findings was that factors such as model formulation, parameterization, and the skill of the modeller can have a bigger impact on simulation accuracy than whether or not the model was lumped or distributed (Reed et al., 2004). They also found that on average, calibrated models outperformed uncalibrated models during both the calibration and validation periods, and that defining reasonable parameters *a priori* from the physical characteristics of a watershed is more difficult than defining reasonable parameters for a conceptual lumped model through calibration (Reed et al., 2004). Finally, they found that models that combine techniques of conceptual rainfall-runoff and physically based distributed routing consistently showed the best performance.

More recently, Phase 2 of the Distributed Model Intercomparison Project (DMIP2) was completed. This project focused on the Oklahoma region and included 16 models: 14 distributed and 2 lumped. The two lumped models were used to define a robust benchmark for evaluating the improvement of distributed models compared to the lumped models (Smith et al., 2012a). The results of this experiment showed that distributed models can account for spatial variability in basin features and precipitation, while successfully preserving the water balance. They also found that the data used in calibrating the models must be

stationary and unbiased, and in general, distributed models provided an improvement on hydrograph simulations compared to lumped models (Smith et al., 2012b).

The DMIP studies evaluated model performance in comparison with observed data, and did not evaluate the deviations of model simulations with climate change projections. Velázquez et al. (2013) carried out such an investigation on four catchment models in two humid, mid-latitude catchments in Québec and Bavaria. Their results showed that the choice of hydrological model strongly influenced the response of hydrological indicators to climate change, especially in the case of low flows, whereas high flows showed less sensitivity to model choice. The choice of models was deliberately broad, ranging from conceptual and lumped to process-based and fully distributed, however a small sample of 4 models does not encompass the broad range of model structures in existence. Najafi et al. (2011) conducted a similar experiment, comparing three lumped and one distributed model, however they also investigated the uncertainty derived from the choice of climate model used for input data to drive the hydrological model. They found that the uncertainty derived from the choice of hydrology model was much smaller than that derived from the choice of climate model, except during the dry season, and concluded that the choice of hydrology model is important when assessing the impact of climate change on hydrology.

Gosling et al. (2011) conducted a comparative analysis of one global and six catchments scale hydrological models, for six catchments across the world. They used the models to project the impacts of climate change on annual average runoff, and extreme flows with seven different GCM inputs. In this study, each catchment model was used to simulate one of the six catchments, whilst the global scale model was used to model all catchments, and was compared to each catchment model individually. The results from this study agree with those of Najafi

et al. (2011) in that the differences in projected changes of mean annual runoff, as well as extreme high and low flow indicators between the global and catchment models, were generally relatively small in comparison to the range of projections across different GCMs.

The Water and Global Change (WATCH) - Water Model Intercomparison Project (WaterMIP) was a very comprehensive intercomparison project, and was the first international project to develop a multi-model ensemble for global hydrology models. It was coordinated by the Natural Environment Research Council - Centre for Ecology and Hydrology in the UK. Being either semi-or fully distributed across the globe, GHMs require significant computational resources, and so comparison of such models has not been feasible until recently. WaterMIP included 11 global models: 6 Land Surface Models and 5 Global Hydrology Models, which was later extended to 13 models. As part of the project WATCH released a forcing data set for the period 1901-2001, and a driving data set for the years 2001-2100. All data and modelling for the project was done on a 0.5° x 0.5° grid scale. The WATCH project consisted of a thorough analysis of global water availability, made up of many individual research projects that ranged from estimating water use in energy and manufacturing (Voß and Flörke, 2010), to investigating the processes that impact runoff generation in Northern Latitudes (Blyth, 2009).

One aspect of the WATCH project was to perform an intercomparison between the models.

Figure 1.5 shows the range of thirteen model outputs for six major river basins when simulating past runoff for the years 1985-1999. These graphs show that between models, runoff simulations can vary quite significantly and that there is more uncertainty in some catchments than in others. For example, the Brahmaputra River shows quite a small deviation about the ensemble mean, but the Murray-Darling demonstrates a wide range between the model simulations. In absolute

terms, the range of the Murray-Darling is smaller than that of other catchments, as it is a low flow catchment, but in percentage terms it shows significant uncertainty. These graphs also highlight that the range between models tends to be bigger during months of high flow, especially in very strongly seasonal catchments. Harding and Warnaas (2011) explain that interestingly, the two different types of models (GHMs and LSMs) did not group together in over-estimation and underestimation, except in areas where snow is a major influence. Haddeland et al. (2011) compared the models on an annual timescale with observed records, and showed that most models overestimate runoff in semiarid and arid basins. This can in part be explained by water abstractions in these areas, since the models were all set to simulate naturalised flows for this comparison project, but the overestimation could also be explained by the lack of the models' consideration of transmission loss along river channels and re-infiltration and subsequent evaporation of surface runoff.

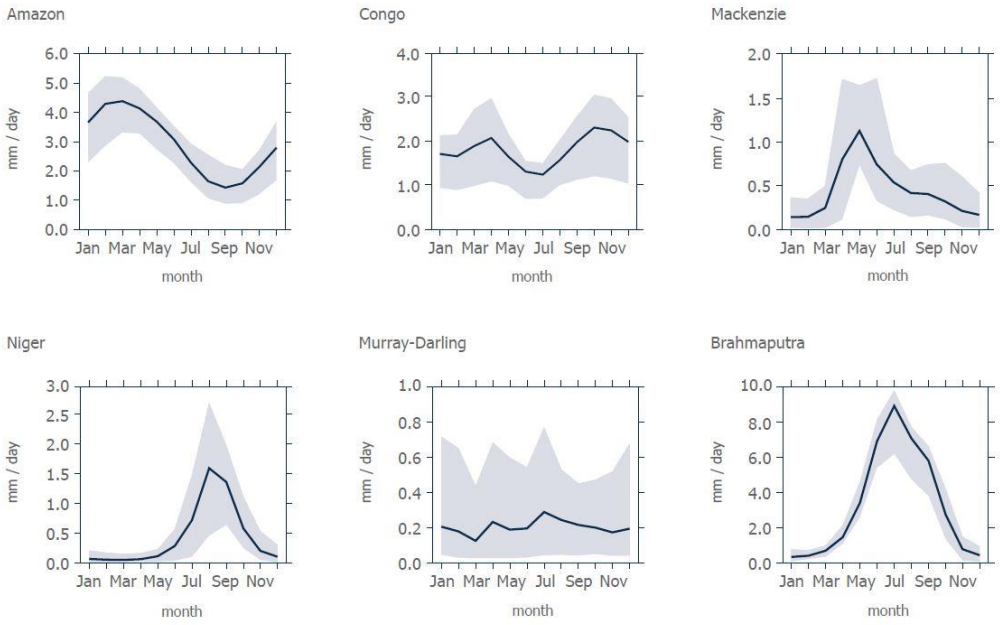


Figure 1.5 Multi-model total runoff monthly mean in mm per day for six of the world's major river basins for the period1985-1999. *The shaded area represents the range of the thirteen models. The continuous blue line is the ensemble mean (Harding and Warnaas, 2011).*

The global average runoff fractions of HTESSEL, JULES and MATSIRO were all lower than the other models, and GWAVA, LPJmL and MacPDM gave similar results, which were slightly higher than the others. Globally ORCHIDEE predicts the highest runoff fraction, and H08 and VIC are closest to the GHMs out of all of the LSMs.

WATCH also projected future global hydrology and assessed available water resources. One of their key findings was the map displayed in Figure 1.6. This map was an amalgamation of results from eight GHMs and 3 GCMs, and shows the projected changes in available water resources for the years 2071-2100 compared to 1971-2000. The map shows that Europe is the largest area projected to experience the largest proportional decline in water resources this century. The Murray-Darling catchment in Australia, as well as the Okavango in Africa and the Pearl River in China will also have their water supply halved by 2071-2100. The Mississippi will also see a significant decrease.

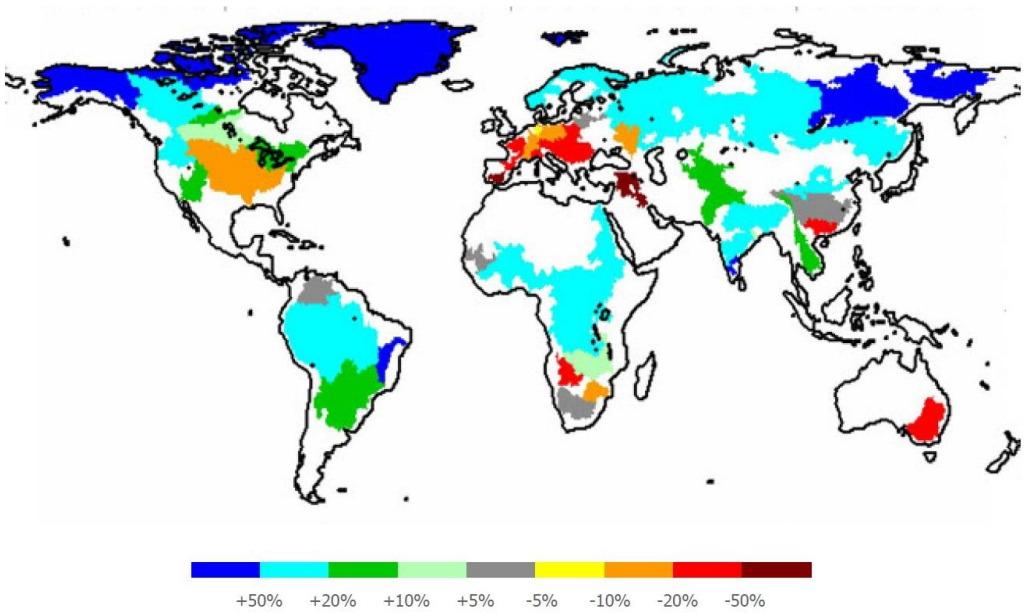


Figure 1.6 Changes (2071-2100 compared to 1971-2000) in available water resources projected by an ensemble of eight global hydrology models using data from three global climate models. *The available water resources were derived by taking into account the total runoff for selected large-scale river basins minus an estimate of the environmental flow requirements in the respective basins. Taken from (Harding and Warnaas, 2011)*

Many catchments are expected to have increased water supply, with the largest increases (>50%) in the high latitudes. South America, central Africa and all Russian catchments will have increased water supply. These are just a few of the research findings of the WATCH project.

WATCH ended in 2011 and soon afterwards, the Inter-Sectoral Impacts Model Intercomparison Project (ISI-MIP), which is coordinated by the Potsdam Institute for Climate Impact Research (PIK), began in 2012 with a one year fast-track project. The second phase, ISI-MIP2, was launched in May 2013 and is planned to last 4 years. The Fast Track (FT) project has contributed outcomes to the IPCC's Fifth Assessment Report (AR5), and results have been published in a special edition of PNAS (Vol. 111, issue 9, 2014). The ISI-MIP FT brought together 28 global impacts models from five different sectors: water, biomes, agriculture, health (malaria) and coastal infrastructure. The 12 hydrology models included in the study were: LPJmL, JULES, VIC, H08, WaterGAP, Mac-PDM.09, WBM, MPI-HM, PCR-GLOBWB, MATSIRO, DBH and ORCHIDEE (Warszawski et al., 2014). Much of the framework for ISI-MIP FT was already in place due to the WaterMIP project, which allowed for the speed of the one year project. Climate data made use of the CMIP5 GCMs, and covered the years 1960-2099 at 0.5° x 0.5° spatial resolution.

Prudhomme et al. (2014) investigated the uncertainties in hydrological drought projections for the 21st century. In the context of drought, their results contradict those of Gosling et al. (2011) and Najafi et al. (2011), as they determined that the uncertainty due to GHM choice is greater than that for global climate models, and that the different representations of terrestrial water cycle processes in GHMs are responsible for much larger uncertainty in response of hydrological drought to climate change than previously thought. The JULES model, which is the only model that accounts for the dynamic response of

plants to CO2 and climate, and so allows vegetation to grow in response to its environment, simulates little or no increase in drought frequency, whereas other models showed maximum drought severities up to and exceeding a 40% change in regional deficit index.

Figure 1.7 shows the change in Regional Deficit Index (% of area under drought conditions) from 7 GHMs and 5 GCMs for 17 regions across the world for 2070-2099 compared to 1976-2005. This figure shows that average changes vary between no change (Eastern Africa) and 28% increase (central Europe) with five regions projected to experience at least a 20% increase in Regional Deficit Index: South and Meso-America, Caribbean, and Central and Western Europe. The greatest uncertainty is in Eastern Europe, South and Southeast Asia and Eastern Africa). Figure 1.7 highlights the discrepancy between JULES and the other GHMs. Schewe et al. (2014) used an ensemble of 11 GHMs to assess water scarcity under climate change. They found that both GHMs and GCMs contributed to uncertainty in the ensemble projections, and that GHM uncertainty is particularly high in regions affected by declining water resources.

Figure 1.8 shows two maps which present the uncertainties in the study performed by Schewe et al. (2014). The top map shows the change in annual mean discharge at 2°C, and the darker the colour, the better the agreement among models. This shows that there is high confidence that there will be a significant reduction in discharge across the Mediterranean, and in southern America. On the other hand there is good agreement that there will be substantial increases in discharge in the high latitudes, in India and Bangladesh, and also across Ethiopia, Somalia and Kenya. The bottom map shows the ratio of GCM variance to total variance, so areas with high values (in blue), show where the GCM variance was higher than the GHM variance, whilst areas in red show where GHM variance was higher.

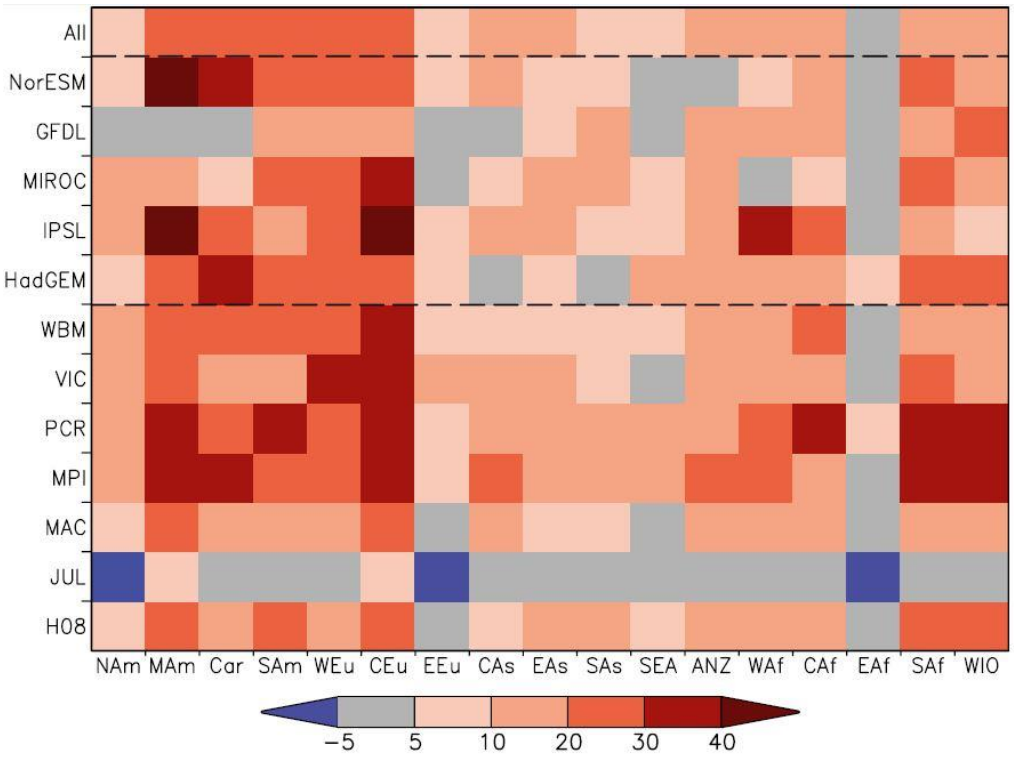


Figure 1.7 Mean percentage changes in regional deficit index (RDI) between 30 year simulations of reference (1976-2005) and future (2070-2099) under RCP8.5 for 17 world regions. Values are averaged over all of the MME members (All), by GCMs (central block) and by GHMs (bottom block). *Taken from Prudhomme et al. (2014) Regions read: North, Meso and South America (NAm, MAm, SAm), Caribbean (Car), Western, Central and Eastern Europe (WEu, CEu, EEu), Central, East, South and Southeast Asia (CAs, EAs, SAs, SEAs), Australia and New Zealand (ANZ), Western, Central, Southern and Eastern Africa (WAf, CAf, EAf, SAf) and Western Indian Ocean (WIO).*

The fact that the bottom map is predominantly blue indicates that on the whole, the variance in the GCMs outweighs that of the GHMs, but there are distinct regions where GHM uncertainty is high, which seem to be mostly in the tropics. These results, along with those of Prudhomme et al. (2014) suggest that for annual mean discharge, GHM uncertainty is generally smaller than GCM uncertainty, however for drought prediction, the choice of hydrology model plays a larger role in the model result.

Studies such as this give us information about the confidence we can place on models; however, in many cases, these comparisons are being made before the uncertainties within one model have been

thoroughly assessed. In the concept of the tree, MMEs compare one tree with another, but this is done by just plucking one leaf per tree and comparing them. The following subsection outlines current research progress in assessing the uncertainties within single models, the concept of comparing several leaves from the same tree.

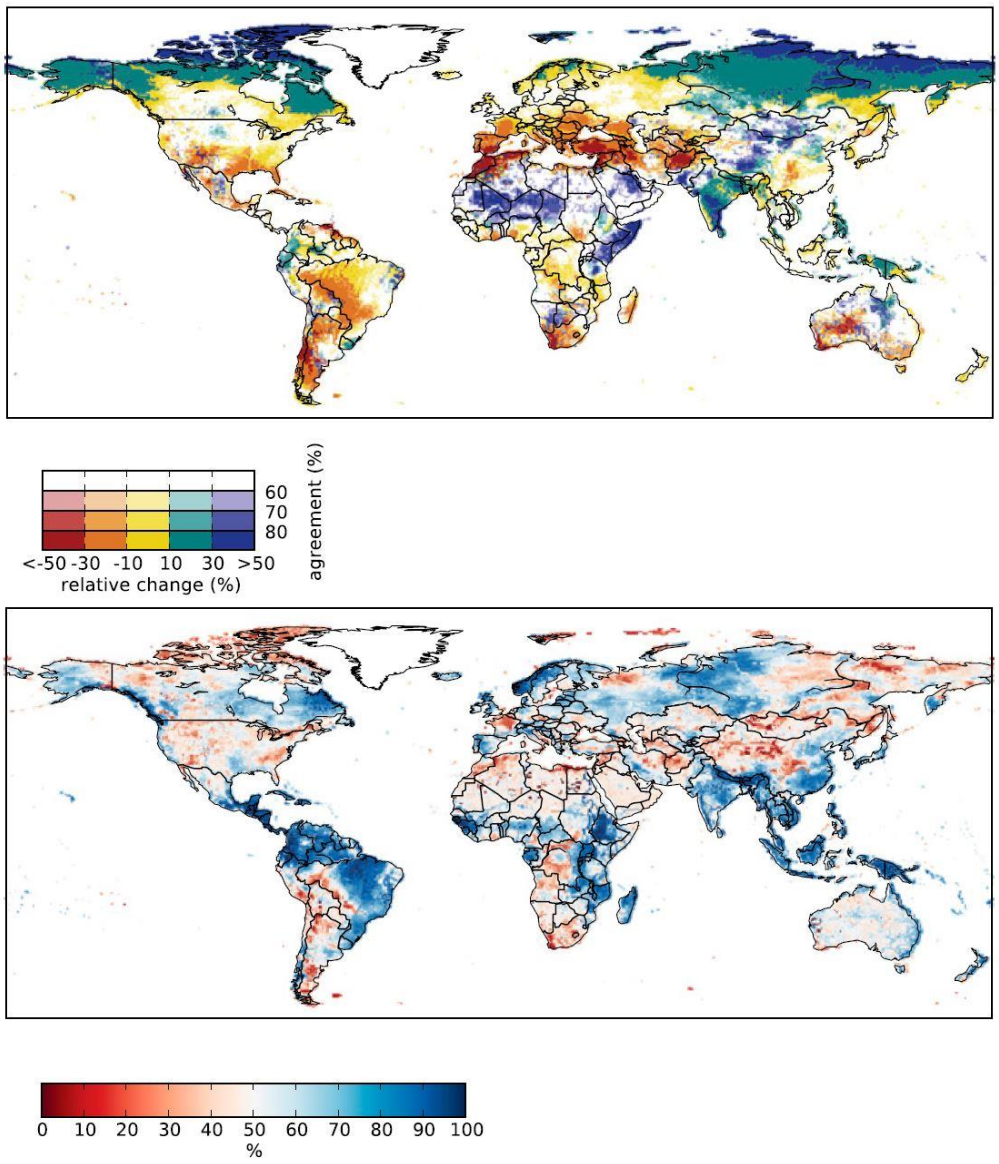


Figure 1.8 Relative change in annual discharge at 2°C compared with present day, under RCP8.5. *Upper: Colour hues show the multimodel mean change and saturation shows the agreement on the sign of change across all GHM-GCM combinations (% of models agreeing on sign of change). Lower: Ratio of GCM variance to total variance; in red areas, GHM variance predominates, in blue areas GCM variable predominates. GHM variance was computed across all GHMs for each GCM individually, and then averaged over all GCMs; vice versa for GCM variance. Greenland has been masked. Taken from Schewe et al. (2014).*

1.6.2 Single Model Uncertainty Estimation

As shown in Figure 1.4, there is significant uncertainty within each individual model. A few studies have attempted to address these uncertainties, though due to the computational requirements of such investigations most have been fairly basic. The advantage of MIPs, such as those that have been discussed, is that different GHMs are developed and run at different institutions, thus spreading the computational load. The most common method of assessing the uncertainties within a model is to run the model with many different set ups, e.g. using different equation settings, or with different model parameter values, which places significant demand on the researcher/research group and the computational facilities available to them.

Beven (2012) outlines the steps in developing a hydrological model as:

1. The Perceptual Model: Deciding on the processes
2. The Conceptual Model: Deciding on the equations
3. The Procedural Model: Getting the code to run on a computer
4. Model Calibration: Determining values for the parameters
5. Model Validation: Confirming applicability and accuracy

Steps 2 and 4 are significant sources of uncertainty. Deciding upon the equations to use in the model is often a subjective preference, but can be influenced by the amount of the data required. For example, the Penman-Monteith method of evapotranspiration calculation requires many more observed variables than the simpler Priestley-Taylor method. Since the availability of such data sets is becoming more easily accessible, the Penman-Monteith method is currently the more popular of the two in GHMs (see Table 1.1), but many catchment and global scale models still employ the Priestley-Taylor method.

Obtaining values for model parameters is mostly achieved using observed data sets. This is applicable to soil and vegetation

parameters, such as vegetation height and hydraulic conductivity. However, the spatial distribution of models, particularly for global hydrology models, requires that observations represent large areas. Therefore, parameter estimation often requires generalisation of observed values to the model's spatial scale. Even then it is of course impossible to measure parameter values for every grid square across the catchment, or even the globe, so parameters are ascribed to certain vegetation or soil classifications. The issue with this is that the modeller is then assuming that these classifications are consistent regardless of spatial location, in GHMs for example, crops in eastern Asia have the same physical properties (height, leaf area index, stomatal conductance etc.) as crops in the USA.

On top of this, many models contain parameters that do not have a physical meaning, such as the spatial variability of soil moisture capacity parameter (b, in equation 1.1), and so cannot be estimated using observed data. In this situation, parameter values may be taken from the literature, where other models may have used the same equations, or they may be estimated and then *optimised*. Optimisation is a method of model calibration. Calibration usually requires observations of the catchment response; the modeller will run repeated simulations of the model, adjusting the values of the parameters between each run and compare the results with the observed record. The modeller may do this manually, or may use a computerised algorithm until some 'best fit' parameter set has been discovered (Beven, 2012). This process can vary significantly in complexity, from a few parameters varied individually in tens of adjustments, to all parameters varied simultaneously in hundreds or thousands of adjustment sets (known as "model realisations"). Varying parameters in this way not only provides an optimum set (or several sets) of parameter values, but also provides an indication of the uncertainties in the choice of parameter values, as different sets of parameter values may produce similarly 'good' simulations when compared to

observations. This concept is known as “equifinality” (Beven, 2006a), and will be discussed further in Chapter 3.

In catchment hydrology, investigations into single model uncertainty have been fairly advanced and thorough. Catchment models have several advantages over global models that have enabled such rigorous analysis: they are usually quicker to run, and have smaller input and output files than global models; they cover a smaller spatial domain, so are likely to have fewer land cover classes and soil classifications to parameterise; and catchment modelling is in a more advanced stage of research than global hydrology modelling, due to its earlier introduction. Nevertheless, global hydrology models have been subjected to a range of calibration assessments and uncertainty analyses.

In 1998, Vörösmarty et al. investigated the impact of different potential evaporation (PE) functions on the Water Balance Model (WBM). They applied 11 different methods to simulate the annual streamflow for 679 gauged watersheds in the United States. The 11 methods covered both reference surface (e.g. Thornthwaite, Penman, and Hamon), and surface cover dependent (e.g. Priestley-Taylor and McNaughton and Black) algorithms. They found that for reference surface methods, simulated PE varied from approximately -100 to +100 mm yr^{-1}, whilst for surface cover dependent methods the range was much smaller (-50 to +50 mm yr^{-1}). Among individual methods, they found that by using different PE estimation methods, PE estimates can differ by hundreds of millimetres, with the largest differences seen in hotter, drier climates where PE is the highest. Vörösmarty et al. (1998) concluded that for contemporary climates, the Hamon method gives good results, however for climate change projections, the more theoretical surface cover dependent methods are more suited than the reference surface methods.

Gosling and Arnell (2011) also investigated the impact of potential evaporation method, along with parameter adjustments for the

parameters b and FC in the GHM Mac-PDM.09. As previously discussed, the b parameter defines the degree of variability in soil moisture capacity across the grid cell. FC represents the field capacity of each vegetation type, which is determined from the soil texture. Gosling and Arnell (2011) used just two PE methods, Priestley-Taylor (PT) and the Penman-Monteith (PM), and a low and a high value for each of the parameters such that b ranged from 0.3 to 0.8 (original value 0.5) and FC ranged from 1.2 to 0.8 (original value 1). They simulated 14 parameter perturbations. They found that of the three parameters investigated, simulated runoff was least sensitive to FC, with a less than 20% change. Runoff is slightly more sensitive to b, with the biggest effects in catchments with low runoff. By far the greatest influence on runoff was with the change in PE method. In dry regions, the PT method produced positive runoff anomalies of around 20-60% relative to the PM method, with the reverse trend in wet regions. This is not surprising given that the two methods include different meteorological variables, and that humidity is not present in the PT method (Gosling and Arnell, 2011).

The b parameter is also investigated in the calibration of the WaterGAP2 model by Alcamo et al. (2003), though in this study, the parameter is referred to as γ. Alcamo et al. (2003) attempted to calibrate the model to discharge stations at 724 locations across the globe for the years 1980-2010 (depending on data availability), by varying the b parameter. The aim was to limit the difference between the modelled and measured long-term average discharge over the calibration period to 1%. They found that by varying the b parameter between 0.3 and 3 (their estimate of the physically plausible range), a 1% difference was only achieved in 385 of the basins. In 201 of the basins, which were mostly snow-dominated areas, the model underestimated the discharge, perhaps due to measurement errors in the amount of snow, and in the other 138 basins, the model overestimated the discharge, due to transmission loss, and evaporation

from ponds. These errors were corrected using runoff coefficients using a multiple linear regression approach.

The variable infiltration parameter in VIC was targeted for calibration by Nijssen et al. (2001b), along with other soil hydrological properties of the depth of the second soil layer, the saturated hydraulic conductivities and the exponents for the unsaturated hydraulic conductivity in the first and second layers. Nijssen et al. (2001b) manually calibrated the model using five climatic zones and nine basins to match the total annual flow volume and the shape of the mean monthly hydrograph. The infiltration parameter and the depth of the second soil layer were ascribed a uniform value for all grid cells in a given climate zone. The remaining parameters were changed from their original spatially varying values using a regionally uniform multiplier.

They found that the infiltration parameter tended to be smallest in the arid climates, in an effort to reduce runoff production, and that the soil layer depth was smallest in the arctic. Nijssen et al. (2001b) then transferred the calibrated parameters to thirteen further basins, using the parameters for each cell from the relevant climatic zone calibration. This process was found to improve the simulated flow in six basins, gave little or no change in three basins, and resulted in worse simulations in four basins. Three of the four poorly modelled catchments were in the western Russian Arctic (Ob, Pechora and Severnaya Dvina), and had considerably higher precipitation than the basin immediately to the east (Yenisei), which was in the original 9, so they were recalibrated during the second round using an additional set of parameters. Once applied globally, the parameterisation led to an increase in global annual runoff of 9.4% and a reduction in evapotranspiration of 5%.

WaterGAP2 was again assessed by Müller Schmied et al. (2014) in an investigation into a variety of sources of uncertainty. Müller Schmied et al. (2014) considered five major sources of uncertainty: climate forcing,

land cover input, model structure, consideration of human water use and calibration (or no calibration). They developed a single variant of the WaterGAP2.2 model in each case, to investigate the sensitivity of the water fluxes and water storage variations of the variants compared to the standard version of the model. In the climate variant, the monthly dataset was adjusted from the standard WATCH forcing dataset was swapped for the CRU TS 3.2 with GPCC v6 monthly precipitation totals. In the land cover variant, the MODIS land cover data was swapped for the GLCC and CORINE datasets. Structural adjustments involved removing the various model improvements that have been implemented in the past decade, including the reservoir operation algorithm of Döll et al. (2009), and the variable flow velocity algorithm of Verzano et al. (2012). The no-calibration version of the model was an uncalibrated simulation with the standard version of WaterGAP 2.2. The calibration approach involved again adjusting the runoff coefficient (b/γ), within the limits of 0.1 and 5.0, and if necessary two additional correction factors. The no human water use variant reflected naturalised water flows and storages without the impact of human water use.

They found that the calibration of the model to 1319 gauging stations had the highest effect on the modelled water fluxes and led to the best fit of the modelled monthly and seasonal river discharge to the observed record. Adjusting the climate forcing had the second highest effect, and was stronger than that of alternative land cover inputs. The adjustments to the model structure showed that the modern version of the model has an improved fit to observed discharge. The structure affected globally averaged fluxes and storage values but the contribution of change is from a small number of grid cells. Human water use proved important for the global water storage trend, but the impacts on water fluxes were localised to areas of high water use.

A much more comprehensive approach was adopted by Wisser et al. (2010) in an assessment of the water balance model (WBM). Wisser et

al. (2010) carried out a Generalised Likelihood Uncertainty Estimation (GLUE, see (Beven and Binley (1992), Freer et al. (1996))) experiment which aimed to assess the impact of variations in model parameters on simulated discharge by randomly sampling the parameters within a predefined range, and by then running a large number of model simulations with different parameter sets. They applied this method to the Mississippi and Danube catchment for three model parameters, and found that the γ parameter, that determines the fraction of excess rainfall that fills a runoff detention pool or that becomes runoff instantly, was the most sensitive parameter, along with SF, that partitions precipitation to rainfall and snowfall. However, they concluded that the impact of variations in γ, SF and β (that controls the outflow from the runoff pool) on annual values of predicted discharge were minimal.

Sperna Weiland (2011) carried out a similar investigation on the PCR-GLOBWB model. Their experiment included 10 model parameters and used a Latin Hypercube Sampling method (see Chapter 3) to sample 250 model realisations. Sperna Weiland (2011) eliminated 95% of these 250 realisations, to leave the 12 remaining best parameter combinations for each of 5 river catchments (Amazon, MacKenzie, Mekong, Murray and Rhine). They did not investigate the sensitivities of individual model parameters in detail, and do not present the parameter values that produce the “best” model simulations when compared to observations. However, they determined that for all catchments except for the Amazon, the uncertainty ranges of the LHS ensemble enveloped the measured discharge data.

These papers reveal that most of the calibration and parameter uncertainty experiments carried out on global hydrology models have been focused on the potential evaporation method and the soil moisture storage capacity parameter. Perhaps due to computational constraints, in depth sensitivity and uncertainty analyses have yet to be carried out on global hydrology models. This should ideally be addressed before

GHMs are used in model intercomparison projects, as it seems premature to investigate the uncertainties derived from differing models before the uncertainties within a single model have been thoroughly examined. In the analogy of the tree, you want to be sure that you have plucked a leaf that is representative of that tree's canopy (see Figure 1.9), before you compare it with leaves from other trees.

Figure 1.9 Two leaves plucked from the same fig tree. *Image posted by Encanto Farms Nursery on the figs4funforum.websitetoolbox.com, permission granted.*

1.7 GHM Uncertainty in Policy Documents

Presenting uncertainty is very challenging, and in the context of policy documents where confidence in research findings is required for decision making, diagrams, graphs and language must be considered very carefully. The Intergovernmental Panel on Climate Change (IPCC) have been leaders in the challenge of uncertainty presentation.

They have been both praised and criticised for their use of verbal probability labels, such as *Likely* (66-100%), and *Extremely Likely* (95-100%) in their reports (e.g. Budescu et al., 2009). In hydrology projections, the IPCC have used novel mapping methods to display the agreement between multi-model ensembles. In the fourth assessment report (4AR) (IPCC, 2007), runoff maps were displayed with stippling in the regions where 80% of models agreed on the sign of change. Whilst

this method was quite easy to interpret, it only allowed for one distinct level of certainty to be displayed with a stipple/no stipple divide.

In the more recent fifth assessment report (5AR) (IPCC, 2014), the map presented by Schewe et al. (2014) (shown in the upper map of

Figure **1.8**) is reproduced. This map uses the colour scheme of Kaye et al. (2012), who recognised the need for care when producing bivariate maps, as they can be very difficult to interpret. This scheme uses both colour and saturation very effectively to display the pattern of change as well as grades of (un)certainty associated with the data. The IPCC 5AR mentions that GCM uncertainty, and scenario uncertainty, is generally higher than that of hydrology model uncertainty and hydrology model parameter uncertainty, however they refer to catchment hydrology models, and do not mention the uncertainties in GHMs. Similarly the IPCC Special Report on Extreme Events (SREX) (IPCC, 2012) discusses the uncertainties derived from hydrological model choice, but does not mention parameter uncertainty.

In the WATCH project, multi-model ensembles are well addressed, and are presented in the Outreach Report (Harding and Warnaas, 2011). The graphs shown in Figure 1.5 are taken from the Outreach Report, and they clearly demonstrate the range of model outputs that can be achieved from a multi-model ensemble. However, the spatial distribution of this uncertainty is not demonstrated, and further maps of water resources do no give uncertainty bounds. The Outreach Report states that “we must appreciate the uncertainty in model projections and we must maintain a culture of on-going model improvement”, and that “recognising potential [of the WaterMIP project] to improve models, to quantify uncertainty within them, and to provide a valuable framework for future global water-cycle work, [WaterMIP] quickly became a major output of WATCH”. However, the quantification of uncertainties is severely lacking in the Outreach Report, so decision makers would be required to sift through the projects archive of technical reports and

resultant peer reviewed journal articles. As with the IPCC reports, parameter uncertainties are not discussed in detail by the WATCH literature.

Another example of the use of hydrology models for policy is the AVOID project which sought to provide scientifically robust, policy-relevant answers to questions directly related to the UN Framework Convention on Climate Change (UNFCC) to “prevent dangerous anthropogenic interference with the climate system” (AVOID, 2014). The outputs from this project are divided into: Flyers and Presentations; Papers; Reports; and Media Kit. The Media Kit (which is aimed at journalists) does not mention uncertainty. Uncertainty in climate models is considered in the longer reports, but hydrology model uncertainty is not mentioned. For example, in the report on the implications of climate policy for avoided impacts on water and food security (Arnell et al., 2010), only one GHM was applied (Mac-PDM.09) and the parameter and structural uncertainties within the GHM are not discussed. By contrast, 21 GCMs were applied in this study, with the results from each GCM examined in detail.

Presenting uncertainties is quite a challenge. Policy makers have previously been presented with the results from one calibration of a hydrology model, or perhaps a range of up to a dozen models, as part of a multi-model ensemble. Further information on uncertainty estimation can sometimes be found deeper in the project literature, but it is not easy to come by. Parameter uncertainty experiments can contain several hundred or even several thousand model realisations, so choosing which models to present, and how to present them, is an important aspect of uncertainty studies. Presentation of perturbed parameter ensembles could be displayed in many ways including as a mean, a probability statement, or a total range. Appropriate representation of uncertainties is essential to maintain the usefulness of models and not induce doubt.

1.8 Summary

Hydrology modelling is an important means of understanding the hydrological cycle, and dates back to Mulvany's first attempt to predict the peak of a hydrograph in 1850. However, holistic catchment modelling was first introduced in the 1960's and several models quickly developed with advances in computational capabilities. Global hydrology modelling is especially computationally demanding, and did not take off until the 1990's. There are currently 8 commonly used global hydrology models, which vary quite significantly in structure and ultimate purpose. Being such a recent area of research, global hydrology modelling remains a very uncertain science. Studies to compare different GHMs have been undertaken in the past few years, but investigation into the uncertainties inherent within one global hydrology model has been neglected. Assessments of a full range of model parameters, including land cover and soil parameters have yet to be done.

1.9 Research Questions

The aim of this thesis is to address the issue of uncertainties *within* a global hydrology model by analysing parameter uncertainties.

Based upon this aim, three research questions have been developed:

Research Question 1: How can uncertainties within global hydrology models be assessed and quantified?

As previously discussed, this is a significant research gap in global hydrological science. Whilst common in smaller scale catchment models, uncertainty analysis in global models has been largely neglected. As such, this is the primary research question for this thesis. Uncertainty experiments are computationally demanding and so the feasibility of conducting such experiments as part of the calibration process is an important consideration. Different methods of uncertainty

estimation will be evaluated, and techniques to increase the efficiency of these methods will also be examined.

Research Question 2: What is the feasibility of including rigorous uncertainty estimation experiments in the global hydrology model calibration process?

Uncertainty experiments are notoriously computationally demanding. Whilst in an ideal world, all models would be thoroughly assessed and carefully calibrated before their publication, this may not be feasible. Currently, global hydrology models undergo basic calibration procedures and are then released for use in research. Following the findings of the first research question, which will demonstrate the methods that are available, this research will investigate the potential of applying such techniques to other models in order to determine the overall feasibility of uncertainty estimation experiments in the field of global hydrological research.

Research Question 3: To what extent are "global" hydrology models fit for purpose?

This research question seeks to use the findings of the first research question to query whether global hydrology models are being used in an appropriate way. Further questions that will help answer this include:

a. How can models be evaluated and validated?
b. How do global hydrology models perform in a catchment context?
c. Are the uncertainties in global hydrology models acceptable?

Ultimately, models may be highly uncertain, but they can still be useful.

1.10 Thesis Structure

This thesis is made up of five empirical chapters, followed by a discussion and a conclusion, as follows:

Chapter 2 - gives an in depth description of the chosen model, Mac-PDM.09. This chapter also introduces the study catchments and details work done updating the vegetation and soil maps used by the model.
Chapter 3 - explores methods for assessing uncertainties in modelling studies, and more specifically parameter estimation.
Chapter 4 – presents the results of a Generalised Likelihood Uncertainty Estimation experiment for Mac-PDM.09.
Chapter 5 - investigates the potential for using global hydrology models as catchment models.
Chapter 6 - applies the results from this experiment to an alternative input data set as a validation exercise, and then discusses the results from Chapters 4, 5 and 6 in the context of one other.
Chapter 7 - discusses some of the issues involved in global hydrology modelling and highlights some potential future research, including sensitivity analysis and the presentation of uncertainty to policy makers.
Chapter 8 - reviews and conclude this thesis.

Chapter Two:

The Macro-Scale—Probability-Distributed Moisture Model .09

2.1 Introduction

A global, or macro-scale, hydrology model (GHM) is capable of simulating the hydrology of the world without the constraints of catchment boundaries, and they are commonly applied on a gridded basis. In order to investigate the parametric uncertainties inherent in the field of global hydrological research, a GHM must first be selected for analysis. As discussed in Chapter 1, previous work has explored the uncertainties derived from different models, yet the uncertainties within one model have yet to be determined. Therefore, one model, Mac-PDM.09, has been selected for interrogation in these experiments. This chapter justifies the choice of the Mac-PDM.09 model over the other popular models available (see Table 1.1) and details on the model's structure and parameters are provided. Updates to the soil and vegetation maps, which constitute a new version of Mac-PDM (Mac-PDM.14), are described here. The study catchments that were investigated throughout this study are presented and the collection of both climatological data for model inputs and river discharge data for model validation is also reviewed in this chapter.

2.2 The Mac-PDM.09 Model

The Macro-scale−Probability-Distributed Moisture model (Mac-PDM) was chosen for this study. First developed by Arnell (1999), MacPDM was based upon the Probability Distributed Model (PDM) of Moore (1985). Since 1999, a revised version of Mac-PDM was presented by Arnell (2003), before the current version (Mac-PDM.09) was published by Gosling and Arnell (2011).

In comparison with many of the other global hydrology models available, Mac-PDM.09 is a relatively simple model, which makes it ideal for the uncertainty analyses in this study. Mac-PDM.09 focuses on natural hydrological processes and does not account for anthropogenic influences on global hydrology, or attempt to estimate water scarcity:

this limitation is discussed later (see Chapter 4.4). A full description of the model is given here, followed by a discussion of some of its limitations. The use of Mac-PDM.09 for this study is then justified.

Written in the FORTRAN programming language, Mac-PDM.09 runs on a daily time-step using either monthly or daily climate data (for monthly data Mac-PDM.09 disaggregates some variables to a daily time-step using a stochastic weather generator) (Gosling and Arnell, 2011). Daily input data has been applied throughout this study. The model is capable of using climate data from a variety of sources by employing the appropriate sub-routine on the start-up of the model. The climate input variables required are: precipitation; number of wet days (for monthly input data); temperature; relative humidity or vapour pressure; net radiation (or cloud cover); and wind speed. Soil and vegetation data are also required in the form of spatial gridded data. The model can run on a range of resolutions from 10 x 10 min to 2° x 2°; in this study it has been run on a grid of 0.5° by 0.5°, totalling 67420 cells of land globally.

The basic structure of the Mac-PDM.09 model is shown in Figure 2.1, and, like all other water balance models, can be described with the following equation:

$$S_t = P_t - AE_t - D_t - Q_t + S_{t-1} \qquad \text{(E2.1)}$$

Where P_t, AE_t, D_t and Q_t are precipitation, actual evaporation, delayed runoff and direct runoff during time interval t, respectively, and S_{t-1} and S_t are storage in the soil, lakes and wetlands at the beginning and end of the time interval (Arnell, 1999).

The following description of the components of the Mac-PDM.09 model is largely based upon that given by Gosling and Arnell (2011), with further information on the PDM soil moisture storage as described in Moore (2007).

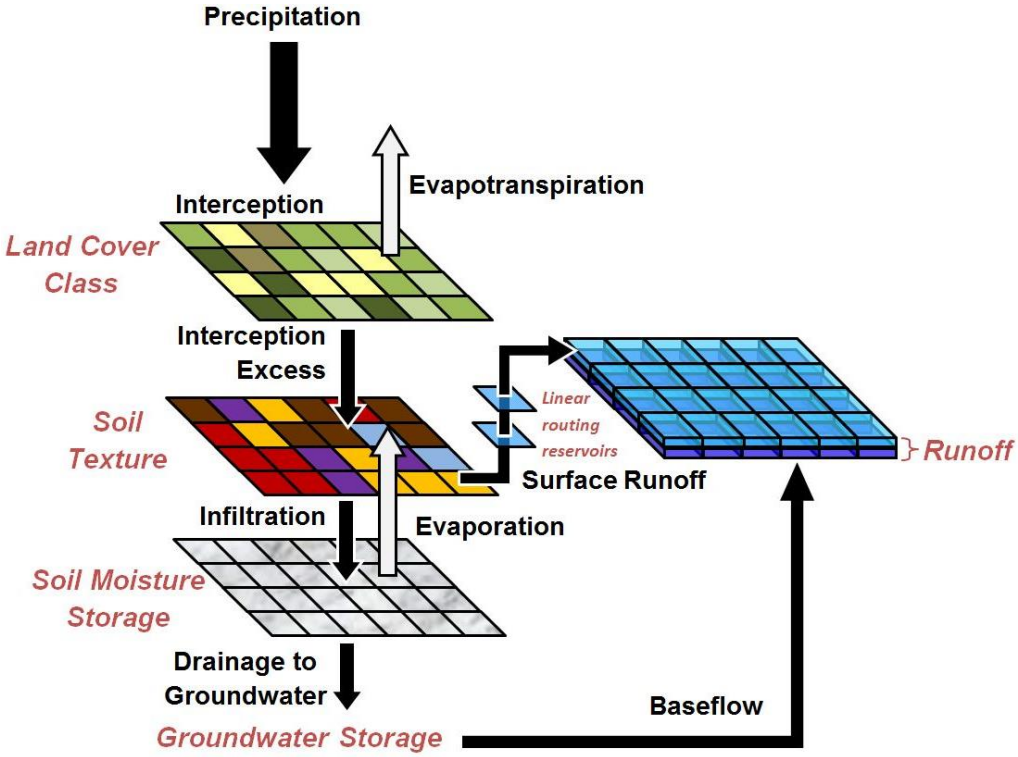


Figure 2.1 Schematic of the Mac-PDM.09 global hydrology model. *Arrows represent hydrological process equations. Black arrows indicate the routes of water to produce runoff.*

A list of the parameters used by Mac-PDM.09 is displayed in Table 2.1. Explanation of the use of these parameters by the model is integrated into the model description below.

2.2.1 Precipitation and Snowmelt

Below a certain temperature threshold (*thresh*, held at 0°C in this study) precipitation falls as snow, and snow that is stored on the land's surface begins to melt. Once snow begins to melt, it does so at a constant rate per degree per day, as defined by the model parameter *xmelt*. When downscaling monthly precipitation to daily precipitation, the parameters *CVrain* (coefficient of variation of daily rainfall), and *SDtemp* (standard deviation of daily temperature from the mean) are used. Mac-PDM.09 does not include a glacier component, nor the effect of the seasonal freezing and melting of permafrost. The model assumes that input precipitation is evenly distributed across each cell, the limitations of this are discussed later in this section.

Table 2.1 Table of Mac-PDM.09 Model Parameters

Parameter	Category	Brief Description	
b	Control	Soil Moisture Capacity Variability	
δ *(delta)*	Control	Interception Parameter	
grout	Control	Groundwater Routing Parameter	
S_{rout}	Control	Surface Routing Parameter	
fact	Control	Field and Saturation Capacity Scaling Factor	
thresh	Control	Temperature Threshold for Snowfall and Snowmelt	
xmelt	Control	Snow Melt Rate	(mm/day/°C)
fcpc	Soil	Soil Field Capacity	(%vol)
satpc	Soil	Soil Saturation Capacity	(%vol)
rootg	Veg.	Root Depth	(m)
rsc	Veg.	Leaf Stomatal Resistance	
capg (γ)	Veg.	Interception Parameter (max daily interception loss)	
rlai	Veg.	Leaf Area Index	
hc	Veg.	Vegetation Height	(m)
percov	Veg.	Percent Cover	(%)

2.2.2 Land Cover, Interception and Evaporation

In this study, Mac-PDM.09 uses 15 land cover classifications, which are used to define several parameters for the model. Vegetation type defines the amount of precipitation that is intercepted, as well as the potential evaporation rates, and the soil moisture storage capacity. Interception is defined using the following equation from Calder (1990):

$$I = \gamma[1 - exp(-\delta P)] \qquad \text{(E2.2)}$$

where: I is the amount of precipitation intercepted, P is precipitation, and γ and δ are the parameters *capg* and δ (*delta*) respectively.

The *delta* parameter is a constant value across the globe, whilst *capg* varies with land cover classification. Potential evapotranspiration is calculated using the Penman-Monteith method (Monteith, 1965), which requires two further calculations as inputs, alongside the leaf stomatal resistance (*rsc*) parameter. Firstly, the vegetation height (*hc*) parameter is used to calculate aerodynamic resistance (r_a) (see Allen, 2005 p. 181) for each vegetation type by:

$$r_a = \frac{\left(ln\left(\frac{z_u - d}{z_{om}}\right) - \varphi_m\right)\left(ln\left(\frac{z_T - d}{z_{oh}}\right) - \varphi_h\right)}{k^2 u_z} \qquad \text{(E2.3)}$$

where: z_u and z_T are the height above the ground surface (m) for the wind speed measurement and the air temperature measurement respectively (2m is used for both in this study), z_{om} is the roughness length (m) governing the transfer of momentum from the surface (0.123**hc* in this study), z_{oh} is an assumed roughness length (m) governing the transfer of sensible heat from the surface ($1/10$*z_{om} in this study), d is the zero plane displacement (m) of the logarithmic wind profile (height at which wind speed becomes near zero in the vegetation canopy, ($2/3$**hc* here), k is the von-Karman constant (0.41) (dimensionless) and u_z is the wind speed measurement at the z_u height.

The *rlai* (leaf area index) and *rsc* (leaf stomatal resistance) parameters are used to calculate the integrated canopy surface resistance (r_s), or bulk resistance, for each vegetation type. This equation calculates leaf surface resistance and upscales it to canopy resistance. The equation for this is based upon the work of Grant (1975) and is given as:

$$\frac{1}{r_s} = \frac{1 - K^{rlai}}{rsc} + \frac{K^{rlai}}{100} \qquad \text{(E2.4)}$$

where K is a radiation coefficient of 0.70.

The aerodynamic and canopy surface resistance values are then used in the Penman-Monteith equation to calculate potential evapotranspiration (mm/day) as:

$$PE = \frac{1000}{\lambda \rho_w}\left[\frac{\Delta R_n + 86.4\ \rho_a\, c_p\, (e_s - e)/\, r_a}{\Delta + \gamma\ [1 + r_s\ /\ r_a]}\right] \qquad \text{(E2.5)}$$

where: R_n is the net radiation (MJ/m^2/day), γ is the psychometric constant (0.66 hPa/°C), e is the vapour pressure (hPa), e_s is the saturation vapour pressure (hPa), r_a is aerodynamic resistance (s/m), r_s is the canopy resistance (s/m), ρ_a is the density of air (kg/m^3), c_p is the specific heat capacity of air (1.013 kJ/kg/°C), Δ represents the slope of the saturation vapour pressure temperature relationship, λ is the latent heat of vaporisation (MJ/kg), and ρ_w is the density of water (1000km/m^3).

Whilst this study uses the Penman-Monteith equation, the model has the capability of running with the Priestley-Taylor method (Priestley and Taylor, 1972). Mac-PDM.09 divides vegetation into ‘grass’ and ‘not grass’. For each ‘not grass’ land cover type, Mac-PDM.09 is ascribed a percentage of grass per cell (as per the model parameter *percov*), the remainder is taken up by the vegetation type itself. Again, the limitations of this will be discussed towards the end of this section.

2.2.3 Runoff Generation

Water that is not intercepted reaches the ground. If the soil is saturated, ‘quickflow’ is generated (surface runoff, but not necessarily overland flow), if not, water is infiltrated into the soil. Water leaves the soil either by evaporation or by drainage to groundwater and ‘slowflow’ (baseflow runoff generation). Actual evaporation is calculated as a linear function of potential evaporation and the soil moisture content, using the soil parameters *satpc* (saturation capacity) and *fcpc* (field capacity). Absolute soil moisture capacity is calculated by multiplying the

percentage values of *fcpc* and *satpc* by the rooting depth (*rootg*). The value of *rootg* over lithosols or organic soils is set to 0.1m. The *fact* parameter is used as a multiplication factor for modifying the modelled field capacity and saturation capacity. It is applied to calculate field and saturation capacity under grass and vegetation from the *fcpc, satpc*, and *rootg* inputs.

The soil moisture storage capacity is a very important part of the Mac-PDM model, and is based upon the PDM model of Moore (1985). The PDM dictates that the soil moisture storage capacity varies statistically across each cell, so that a variable proportion of the cell area is saturated at any given time, and 'quickflow' is generated from this part of the cell. This means that runoff can be generated from at least a part of the cell at almost any time, unlike other water balance models that require the entire catchment to be saturated before runoff is generated. Mac-PDM.09 therefore generates runoff more rapidly in response to smaller precipitation events (Gosling and Arnell, 2011), as is demonstrated in Figure 2.2.

Figure 2.2(a) depicts a model using a single storage tank of capacity c', which takes in precipitation, P, and loses water by evaporation, E. This store fills and spills, generating runoff, q', or empties and ceases to lose water by evaporation (Moore, 2007). Mac-PDM.09 allows the storage capacity to vary across a cell, so at any point c can be considered as a random variate with the probability density function $f(c)$, and that the proportion of the cell with depths in the range $(c, c + dc)$ is $f(c)dc$. If all of these stores were arranged in order of depth, with their open tops arranged at the same height, they would form a wedge shaped diagram as shown in Figure 2.2(b). In Figures 2.2(b) and 2.2(c) C^* depicts the water content of the store.

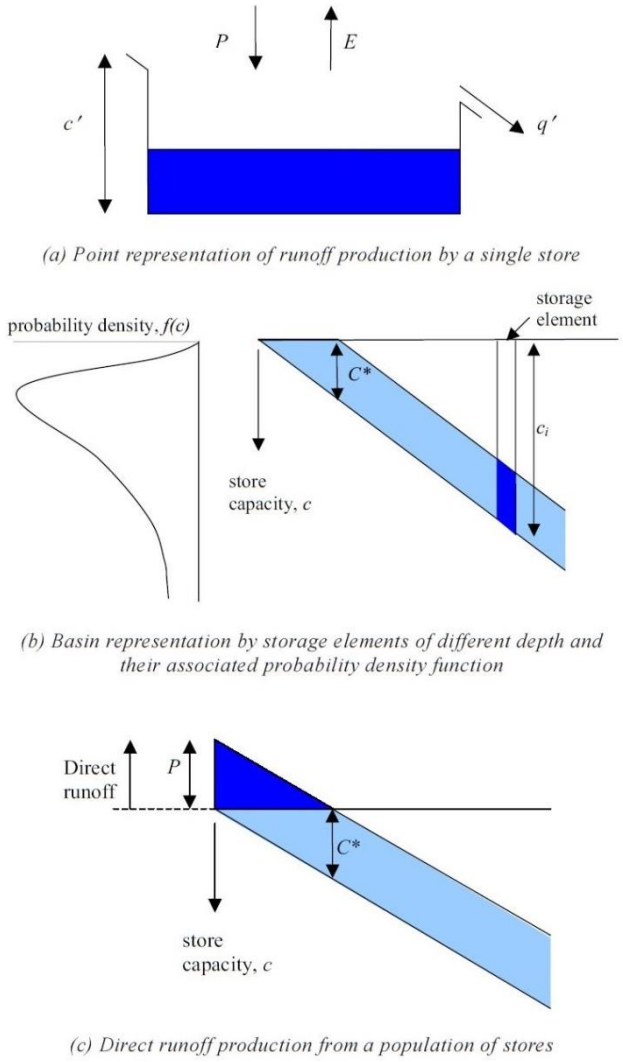


Figure 2.2 Definition diagrams for the probability-distributed interacting storage capacity component *(taken from Moore, 2007)*.

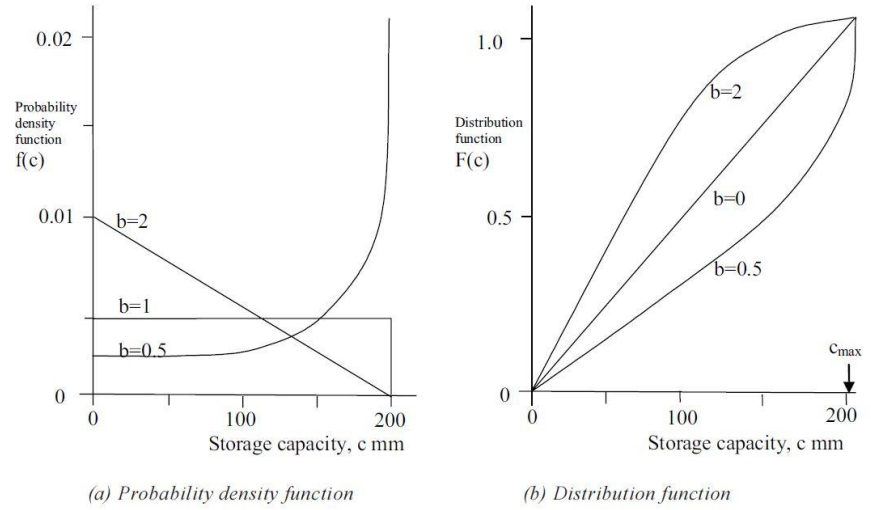


Figure 2.3 Pareto distributions of the b parameter of storage capacity

If precipitation fell at a net rate of P for a unit duration, on an initially dry cell, then stores will fill to a depth of P, unless they are of a lesser depth than P, in which case they would produce runoff during the interval of precipitation. Stores of the same depth as P would begin producing runoff at the end of the interval, so that the upper triangular area in Figure 2.2(c) denotes the depth produced from stores of a different depth over the unit interval (Moore, 2007). There is not necessarily the same number of stores of different depths, so actual runoff is calculated by weighting the depth produced by a store of a given depth by the frequency of its occurrence, as expressed by $f(c)$. Moore (2007) conducted trials on 5 different distributions for storage capacity (Pareto, rectangular, triangular, exponential, and lognormal), and decided upon the Pareto distribution of storage capacity which is now most widely used in applications of the PDM model. The Pareto distribution is employed in Mac-PDM.09, with the distribution function and probability density function as presented in equations 2.6 and 2.7.

$$F(c) = 1 - \left(1 - \frac{c}{c_{max}}\right)^{b} \qquad 0 \le c \le c_{max} \quad \text{(E2.6)}$$

$$f(c) = \frac{dF(c)}{dc} = \frac{b}{c_{max}}\left(1 - \frac{c}{c_{max}}\right)^{b-1} \qquad 0 \le c \le c_{max} \quad \text{(E2.7)}$$

where $F(c)$ is the proportion of the catchment with storage capacity less than c, c_{max} is the maximum storage capacity in the catchment, and b defines the degree of spatial variability. These functions are shown in Figure 2.3.

2.2.4 Runoff Routing

Mac-PDM.09 does not route runoff between cells, but, as previously mentioned, runoff is made from two sources, ‘quickflow’, and ‘slowflow’ (‘baseflow’). The quickflow (surface runoff) is routed through a cascade of two linear reservoirs to represent the delay and dispersion of runoff as it travels across the cell.

For each reservoir, the outflow q_s is the product of the surface water routing (S_{rout}) parameter and the surface storage, S_s;

$$q_s = S_{rout} \times S_s \quad \text{(E2.8)}$$

The baseflow is calculated using the groundwater routing parameter ($grout$) and groundwater storage using equation 2.9, which is derived from the non-linear storage form of the momentum equation given in Moore (2007):

$$bflow = grout \times \left(\frac{gstor}{100}\right)^3 \quad \text{(E2.9)}$$

Although the model runs at a daily time step, the routing parameters represent 'typical' rather than locally realistic hydraulic and geomorphological condition. Therefore, simulated daily runoff is very 'indicative' and monthly runoff is a more credible output (Gosling and Arnell, 2011). Monthly simulated runoff is used throughout this study.

2.2.5 Model Outputs

Mac-PDM.09 outputs a range of hydrological indicators, and these outputs depend on the temporal scale defined by the user. In this study, the "summary mode" was used, which outputs a table of 36 indicators for each grid cell. These include: average annual runoff, annual actual and potential evapotranspiration, annual rainfall and snowfall, average monthly runoff for January-December, the coefficient of variation (CV) of annual runoff , the mean and CV of maximum monthly and daily runoff, parameters of a GEV (generalized extreme value) distribution fitted by L-moments to average annual maximum monthly and daily runoff, and Q5, Q10, Q50, Q90 and Q95 (the flow exceeded 5, 10, 50, 90 and 95% of the time: Q5 is extreme high flow, and Q95 is extreme low flow). Each line of the model output, which describes a particular grid cell, is given a grid code. A separate text file then gives the

longitudinal and latitudinal location of each cell to enable the mapping of outputs, and further regional or catchment scale analysis.

2.2.6 Potential Limitations of the Mac-PDM.09 Model

There are several assumptions and methodological choices in the Mac-PDM.09 model that could potentially impact model output and introduce structural uncertainty to this study. A few of these, which will be discussed in turn here, are: the uniform distribution of precipitation across each cell; the uniform distribution of the *delta* parameter across the globe; the distribution of vegetation across each cell; the choice of the Penman-Monteith evapotranspiration equation; and the lack of a glacier component in the model.

It is unrealistic to have precipitation evenly distributed across a 0.5° x 0.5° area. Precipitation is more likely to be concentrated around areas of high relief, which during periods of snow cover could result in an influence on the time lag of snowmelt, which could also then influence the model output. The spatial distribution of precipitation in general though is unlikely to have a large influence of model output in this study for two reasons: because the model does not route runoff between cells (as will be discussed later in this section), and because the monthly runoff output is used in this study. There is the potential of slight under or over-catch of precipitation due to the extrapolation of catchment boundaries over a 0.5° x 0.5° grid, particularly if cells only part contain the catchment, as these are the areas with highest relief. However, the inclusion of cells with areas slightly outside of the catchments is likely to be balanced with the exclusion of cells with areas slightly inside the boundary. Precipitation distribution across each cell could be integrated into the model by assigning a distribution according to a Digital Terrain Model, but this would require significant revision of the model code.

In Mac-PDM.09, the *delta* parameter of interception is uniform across the globe whilst *capg* varies by vegetation type. The equation for

interception used is given in Equation 2.2. as: $I = \gamma[1\text{-}exp(-\delta P)]$, where γ and δ are the interception parameters *capg* and *delta* respectively and P is precipitation. The *capg* parameter is described as the maximum daily interception loss, whilst the remainder of the equation, within the square brackets, is described by Calder (1999) as the fraction of the day for which canopies remain wet during and following rainfall (the wet day fraction). Whilst in a physical sense, one might expect the wet day fraction to vary by vegetation type, this equation (with δ being constant across the globe) describes it as being dependent on precipitation. As the wet day fraction is multiplied by *capg* (γ), the interception equation as a whole varies by vegetation type, so we would expect there to be little detriment to the model output from keeping δ constant across all vegetation types. If the wet day fraction were to be a factor in the model without the association to the maximum daily interception loss, it might be worth investigating varying δ by vegetation type, but this is not presently the case in Mac-PDM.09.

The inclusion of the *percov* parameter, which describes the percentage grass in each cell, allows the model to have some variation in land cover across each cell. However, this is limiting in three ways: (1) there is only the option to have two land cover types in each cell (the specified vegetation type and grass); (2) there is only the option to have grass as the secondary land cover type; (3) the percentage cover of grass is not uniform across the globe, but it is uniform across each land cover type. Mac-PDM.09 accounts for combination vegetation types such as mixed forests, and wooded grasslands, which somewhat reduces the concern of only grass being available as a secondary land cover type. The uniformity of grass percentage across each vegetation type is rather unrealistic though, as not all areas of urban cover, for example, will have the same fraction of grass cover. A gridded map of grass percentage could be implemented to improve this aspect of the model, but this would again require significant alteration to the model,

as well as remote sensing work likely to introduce subjectivity. It is not known whether this would improve the model output.

The Priestley-Taylor and the Penman-Monteith equations for Potential Evapotranspiration (PET) are both available for use in the Mac-PDM.09 model code. Gosling and Arnell (2011) demonstrated the significant differences in model output depending on which equation was employed, especially in humid areas. The Priestley-Taylor equation is a simpler method that requires fewer input variables. However, the Priestley-Taylor equation does not include air humidity, and is a less physically meaningful method, that requires an additional model parameter. Priestley-Taylor is often applied when the necessary input data is not available for the Penman-Monteith method. In this study, the EU-WATCH project input data is used (as described in section 2.5 of this chapter), which provides all the necessary inputs for the Penman-Monteith equation. Therefore, the Penman-Monteith equation was used in this study, which also enabled the comparison of the model outputs throughout this study, with the results of Mac-PDM.09 runs from the WATCH and the ISI-MIP projects. Since the Mac-PDM.09 model is adjusted in this study, both through mapping changes and the calibration from the uncertainty analysis to produce Mac-PDM.14, changing the PET method as well would confound the results of a comparative study with the original version of the model (Mac-PDM.09).

The lack of a glacier component in Mac-PDM is a significant limitation of the model, especially in the global ‘water towers’ of the Himalayas and the Rocky mountains. In catchments that drain these areas, Mac-PDM is likely to underestimate runoff, and will simulate seasonal peak flows too early due to a lack of the delay in runoff caused by the locking up of precipitation in ice and snowpack. This limitation of the model is discussed further throughout this thesis.

2.2.7 Justification for the use of Mac-PDM in this Study

The experiment that forms the foundation of this study (see chapter 3) involves running a large ensemble comprised of members with different but plausible model parameterisations. The model has many advantages that make it ideal for this study, including: the ease of parameter perturbations; appropriate model outputs (notably the ability to output summary data: 30 year averaged monthly data rather than full time series data which requires substantial disk storage); its ability to be run quickly and efficiently; and its previous use in multi-model ensembles (Ludwig and Voss, 2009, Warszawski et al., 2014). When running Mac-PDM, the model parameters (which are detailed in Table 2.1) are described in three text files: the control file, the soil texture parameter file and the vegetation parameter file. This allows for easy alteration to the parameters and ensembles can be carried out using multiple versions of these text files, without needing to adjust the model code itself. The files are small in size (1KB for control and soil and 2KB for vegetation), which means that each parameterisation requires only 4KB of disk space for the variable input files. The following files can remain the same for each run: the climate forcing data which is 64GB, the 183KB model code, the 2.65MB 'files to read' file (which lists the climate input files), and the 3.08MB cell properties file (which assigns each grid cell across the globe a cell ID, a soil and vegetation type, and gives the area of the cell, since 0.5° x 0.5° cells are not the same size on the equator as at the poles). The model code was investigated for any additional parameters that may have been "hard-coded" into the model itself, but no physically meaningful extra parameters were found.

Outputs from Mac-PDM include global runoff average annual runoff, monthly average runoff and extreme flow indicators. In summary mode, each output file holds 21.99 MB of data, which is a manageable size for running a large ensemble, comprised of several thousand members. The output format is .txt which allows for easy analysis using software

such as Excel, MATLAB and ArcGIS. The Mac-PDM.09 model has previously been modified by Gosling et al. (2010) to facilitate its running on a campus grid. In this study, Mac-PDM.09 was run on a Condor system, which enables the use of idle computers across a campus or workplace to run model realisations. Since the climate data uses vast amounts of disk space (in this case the ClimGen data used which was 20GB; see section 2.5 for a description of climate forcing datasets including ClimGen), Gosling et al. (2010) developed the model to call the bytes of data that were required from the main server, using the 'files to read' file previously mentioned, rather than transfer the entire dataset to each processor at the start of a model run. This capability has allowed the model to be run in ensembles in several locations, such as at Oxford University by Fung et al. (2011), and at the University of Reading and the University of Nottingham for the WATCH (Ludwig and Voss, 2009) and ISI-MIP (Warszawski et al., 2014) projects.

The inclusion of Mac-PDM.09 in the WATCH and ISI-MIP projects will allow the results of this study to be compared to the results of these projects. This means that it will be possible to compare the uncertainties within a model with the uncertainties between models. The WATCH ensemble will be a fair comparison as the models were run with the same climate data as was used in this study (see section 2.5). These factors demonstrate that Mac-PDM is an appropriate model for use in this research. However, previous applications of the model have used out of date land cover classification maps (deFries et al., 1998, based upon AVHRR satellite data, 1984), and a 5 class soil texture map from FAO (1995). It was decided that these maps should be updated to a more recent land cover map, and the more detailed, and commonly used 12 class soil texture classification, in order to produce a new version of the model, Mac-PDM.14.

2.3 Mapping

2.3.1 Soil Texture Mapping

Soil texture is the relative content of sand, silt and clay in the soil (see Figure 2.4). It is important in hydrology modelling as it influences the amount of water and air that the soil holds, and the rate at which water can enter and move through it (FAO, 2014b). Mac-PDM.09 originally used a soil texture classification from the FAO (1995), which consisted of 5 soil texture classifications: sand, sandy loam, silt loam, clay loam, and clay and a sixth classification, lithosols. A map of this soil texture classification across the globe is shown in Figure 2.5.

The USDA Soil Conservation Service (1987) classification is the most commonly used in hydrology, and contains 12 textural classifications: sand, loamy sand, sandy loam, loam, silt loam, silt, sandy clay loam, clay loam, silty clay loam, sandy clay, silty clay, and clay. The relative proportions of sand silt and clay in each of these classifications is indicated in Figure 2.4. The FAO Digital Soil Map of the World (DSMW) remains an up to date resource, as despite its original publication in 1974, it has undergone several updates, the last of which was in 2007.The texture classification used in this study was defined using the FAO DSMW dominant soil unit map and database. The database contained a percentage sand, silt and clay measurement for each of the 117 soil units. These were then correlated to the USDA soil texture classifications (specified in Table 2.2), and graphically presented in Figure 2.4) to produce the updated map shown in Figure 2.6.

For 10 cases out of 117, the percentages could not be classified according to these criteria, so the nearest fit was ascribed. For 12 cases of the 117, percentage silt, sand and clay measurements were not given, in these cases the dominant soil texture for that major group was ascribed.

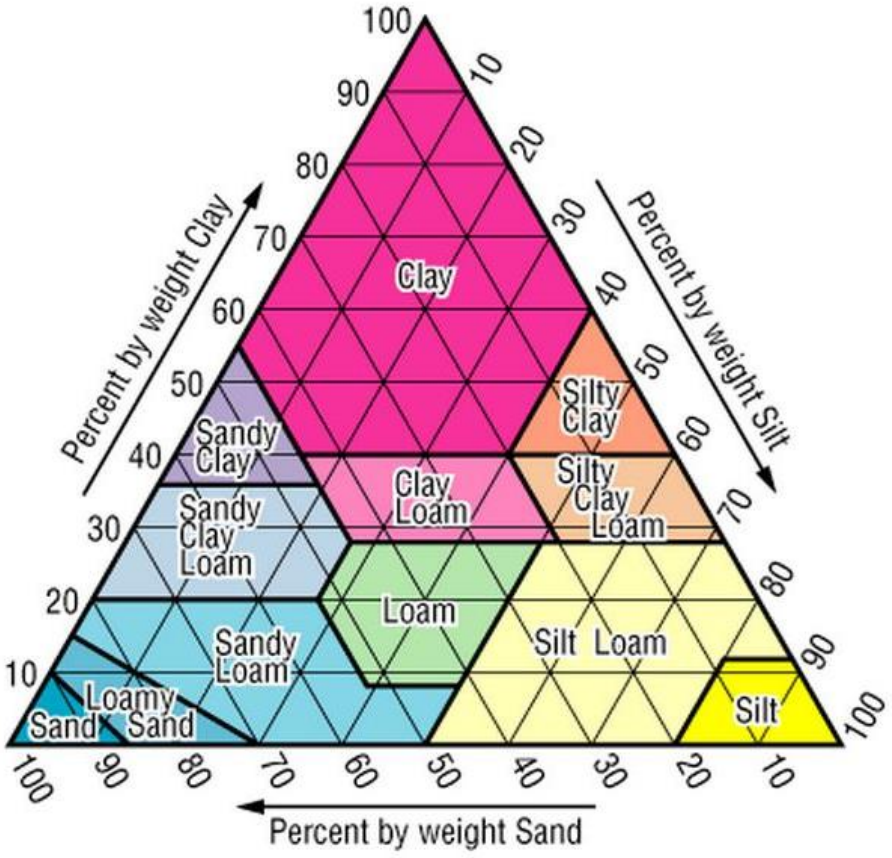


Figure 2.4 Soil texture classification triangle. *Taken from Soil Information for Environmental Modeling and Ecosystem Management (2014) website.*

Table 2.2 Soil texture classification table. *Figures taken from FAO (2014b) website.*

Soil Texture	% Sand	% Silt	% Clay
Sand	86-100	0-14	0-10
Loamy Sand	70-86	0-30	0-15
Sandy Loam	50-70	0-50	0-20
Loam	23-52	28-50	7-27
Silt Loam	20-50	74-88	0-27
Silt	0-20	88-100	0-12
Clay Loam	20-45	15-52	27-40
Sandy Clay Loam	45-80	0-28	20-35
Silty Clay Loam	0-20	40-73	27-40
Sandy Clay	45-60	0-20	35-55
Silty Clay	0-20	40-60	40-60
Clay	0-45	0-40	40-100

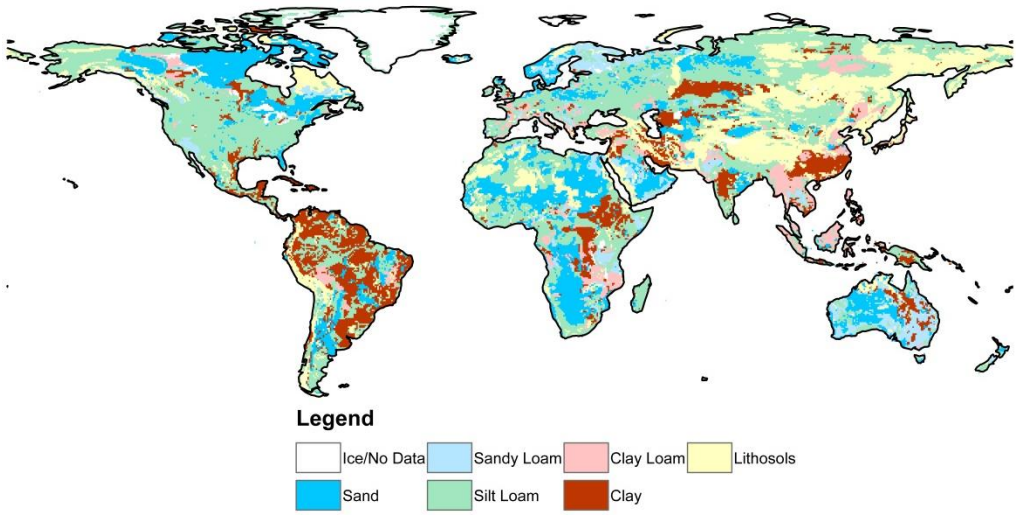


Figure 2.5 Previous Mac-PDM.09 soil texture classification

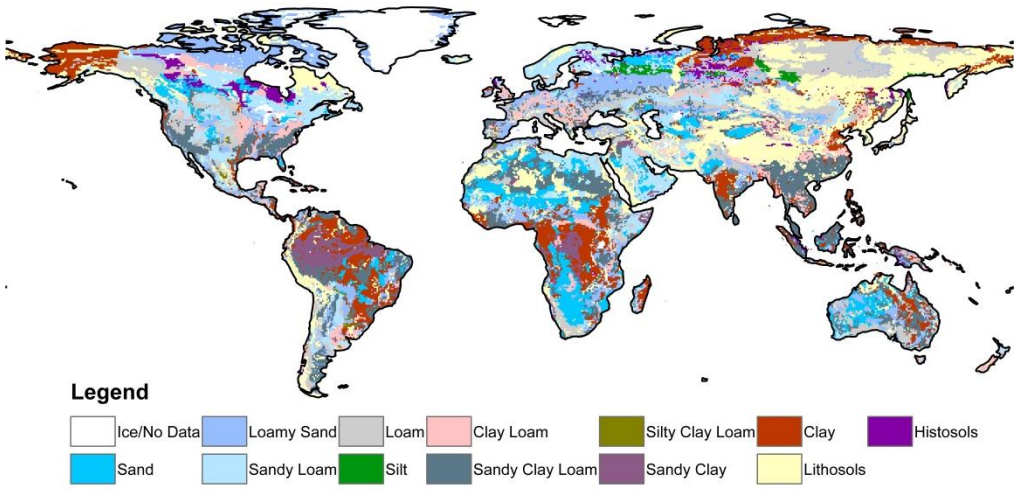


Figure 2.6 Updated Mac-PDM.09 soil texture classification

The classifications “Lithosols” and “Rock Debris” were ascribed the classification “Lithosols” (a 13th classification category) and the Histosols group were given their own classification “Histosols” (a 14th classification category). Glaciers, Salt Flats, Water Bodies and No Data were all given a “No Data” (0) value. Lithosols represent incredibly shallow soils or rocky areas, whilst histosols represent organic material, such as peat. Lithosols and histosols were included in the previous classification for Mac-PDM.09; however histosols were not actually present on the gridded map data.

The most noticeable difference between the previous soil texture map and the updated map in Figure 2.5 and Figure 2.6 respectively is the replacement of the silt loam classification. Interestingly, in the updated map silt loam does not appear at all, but is instead commonly replaced with clay, loam and also sandy loam.

The areas of lithosols remain the same, and histosols are introduced into areas of Canada and central Russia. Several large areas of clay are removed from Brazil and China, though clay is introduced to central Africa and Alaska. Areas of sand texture remain largely the same, though some are reclassified as loamy sand. Clay loam areas are reduced in size, and the large area over Burma is reclassified as sandy clay loam. Sandy clay loam is also introduced to large areas of the United States and Indonesia. Silt covers only minor areas of northern Russia. This update of the soil texture classification provides a much more diverse and realistic representation of global soils than applied in previous versions of the model (Gosling and Arnell, 2011, Fung et al., 2011, Hagemann et al., 2013).

2.3.2 Land Cover Mapping

Updating the land cover classification map was also a priority for Mac-PDM.09, since previous applications of the model have used the AVHRR satellite data from 1984, which is now more than 30 years out of date. Satellite data from AVHRR, MODIS, and many more have been used to develop several global land cover classification products which are detailed in Table 2.3. This table presents 7 readily available products, the last of which was not released in time for this research, but which demonstrates the continual advancement of land cover products.

Table 2.3 Details of available global land cover products

Name	Developer	Satellite Data	Date of Imagery	# of Classes	Reference
GLCF (AVHRR)	Global Land Cover Facility	AVHRR	1981-1994	14	deFries et al. (1998)
GLCF (MODIS)	Global Land Cover Facility	MODIS	2001-2012	17	Channan et al. (2014)
GLCC	United States Geological Survey	AVHRR	1992-1993	25	Loveland et al. (2000)
GLC2000	European Commission Joint Research Centre	SPOT4	2000	22	Bartholomé and Belward (2005)
GLCNMO	International Steering Committee for Global Mapping	MODIS (TERRA)	2003	20	Tateishi et al. (2008)
GlobCover 2009	European Space Agency	ENVISAT (MERIS)	2009	22	Arino et al. (2010)
GLC-SHARE	Food and Agriculture Organisation	Composite	2014	12	Latham et al. (2014)

Mac-PDM.09 used the classification from deFries et al. (1998) which consists of 14 land cover types: evergreen needle-leaf forests, evergreen broadleaf forests, deciduous needle-leaf, deciduous broadleaf forests, mixed forests, woodlands, wooded grasslands/shrublands, closed bushlands/shrublands, open shrublands, grassland, cropland, bare, mosses/lichens and water/ice. A map of this classification is shown in Figure 2.7. Upon investigation of current global land cover classifications, it became clear that the Mac-PDM.09 classification was lacking a few important land cover types: notably, artificial/urban areas and land that is permanently or regularly flooded. Mac-PDM.09 also lacks combination land covers that consist of more than one vegetation type (mosaics). GlobCover2009 presented the most appropriate and up-to-date land cover product available; however its classification system is over-complex for the parameterisation of a global hydrology model, and does not differentiate between open needle-leaved deciduous and open needle-leaved evergreen forest. Therefore, a new land cover classification was defined in order to keep the number of classification types to a minimum, whilst including important up-to-date information.

The list of new classification land cover types is given in Table 2.4. This table shows the previous AVHRR Mac-PDM.09 land cover classification and the GlobCover2009 classification, alongside the new classification system. The table demonstrates the combination of GlobCover2009 classifications to fit the new scheme, for example “irrigated cropland”, and “rain-fed cropland” were combined to develop an overall “cropland” classification.

As can be seen in this table, the “open needle-leaved deciduous or evergreen forest” (value 90) classification of GlobCover2009 needed to be divided into “Deciduous Needle-leaf Forest” and “Evergreen Needle-leaf Forest”, and this was done using the Global Land Cover 2000 dataset. However, a few of the cells with a value of 90 in GlobCover2009 had ambiguous classifications in GLC2000 (herbaceous cover, tree cover – burnt, and mosaic: cropland/tree cover/other natural vegetation). These cells were then referred back to the original Mac-PDM.09 land cover.

Again a few cells were classified as the ambiguous covers from Mac-PDM.09, woodland and mosses/lichens, and so were given the mosaic: trees/vegetation and sparse vegetation classifications respectively. The results of this updated classification system are displayed in Figure 2.8 (with the original classification shown in Figure 2.7). It is apparent from these maps that the classifications are more fragmented across the globe: in the previous map, there are distinct boundaries between land cover types, and whilst some are still evident in the new classification, there are generally more graded boundaries between types, with cell scattering of different types within areas with a dominant land cover. The removal of the “bushland” land cover classification is a distinct change; the large areas over Australia and South Africa are replaced with “Sparse Vegetation” and “Grassland”.

LCM2009_KAS		MacPDM09 (UMD) Equivalent		GlobCover2009 Equivalent		GLC2000 Equivalent	
Code	Description	Code	Description	Code	Description	Code	Description
0	No Data (Water, Ice, No Data)	0	Water	210; 220; 230	Water Bodies; Permanent Snow and Ice; No Data	23; 12	Water Bodies; Snow and Ice
1	Evergreen Needleleaf Forest	1	Evergreen Needleleaf	70; 90	Closed Needleleaved Evergreen Forest; Open Needleleaved Deciduous or Evergreen Forest	20	Tree Cover, needle-leaved, evergreen
2	Evergreen Broadleaf Forest	2	Evergreen Broadleaf	40	Closed to Open Broadleaved Evergreen or Semi-Deciduous Forest	16	Tree Cover, broadleaved, evergreen
3	Deciduous Needleleaf Forest	3	Deciduous Needleleaf	90	Open Needleleaved Deciduous or Evergreen Forest	19	Tree Cover, needle-leaved, deciduous
4	Deciduous Broadleaf Forest	4	Deciduous Broadleaf	50; 60	Closed Broadleaved Deciduous Forest; Open Broadleaved Deciduous Forest	14; 15	Tree Cover, broadleaved, deciduous, closed; Tree Cover, broadleaved, deciduous, open
5	Mixed Forest	5	Mixed Forests	100	Closed to Open Mixed Broadleaved and Needleleaved Forest	18	Tree cover, mixed leaf type
6	Mosaic: Trees/Vegetation	7	Wooded Grasslands	110; 120	Mosaic Forest-Shrubland/Grassland; Mosaic Grassland/Forest-Shrubland	8	Mosaic: Tree Cover/Other natural vegetation
7	Mosaic: Croplands/Vegetation	-	-	20; 30	Mosaic Croplands/Vegetation; Mosaic Vegetation/Croplands	6	Mosaic: Cropland/ Shrub and/or grass cover
8	Shrubland	9	Open Shrubland	130	Closed to Open Shrubland	10; 11	Shrub Cover, closed-open, deciduous; Shrub Cover, closed-open, evergreen
9	Grassland	10	Grass	140	Closed to Open Grassland		
10	Cropland	11	Cropland	11; 14	Irrigated Cropland; Rain-fed Cropland	3; 5	Cultivated and managed areas; Irrigated Agriculture
11	Sparse Vegetation	-	-	150	Sparse Vegetation	13	Sparse herbaceous or sparse shrub cover
12	Bare	12	Bare	200	Bare Areas	2	Bare Areas
13	Broadleaf Tree Cover Regularly or Permanently Flooded	-	-	160; 170	Closed to Open Broadleaved Forest Regularly Flooded (Fresh-Brackish Water); Closed to Open Broadleaved Forest Permanently Flooded (Saline-Brackish Water)	21; 22	Tree cover, regularly flooded, fresh water; Tree cover, regularly flooded, saline water
14	Vegetation Regularly Flooded	-	-	180	Closed to Open Vegetation Regularly Flooded	9	Regularly flooded shrub and/or herbaceous cover
15	Artificial Areas	-	-	190	Artificial Areas	1	Artificial surfaces and associated areas
		Ambiguous Land Covers		Ambiguous Land Covers		Ambiguous Land Covers	
		6	Woodland	90	Open Needleleaved Deciduous or Evergreen Forest	7	Mosaic: Cropland/ Tree cover/ Other Natural Vegetation
		8	Bushland			4	Herbaceous Cover, closed-open
		13	Mosses/Lichens			17	Tree cover, burnt

Table 2.4 Development of the new Mac-PDM.09 land cover classification using GlobCover2009, GLC2000 and the original UMD classification.

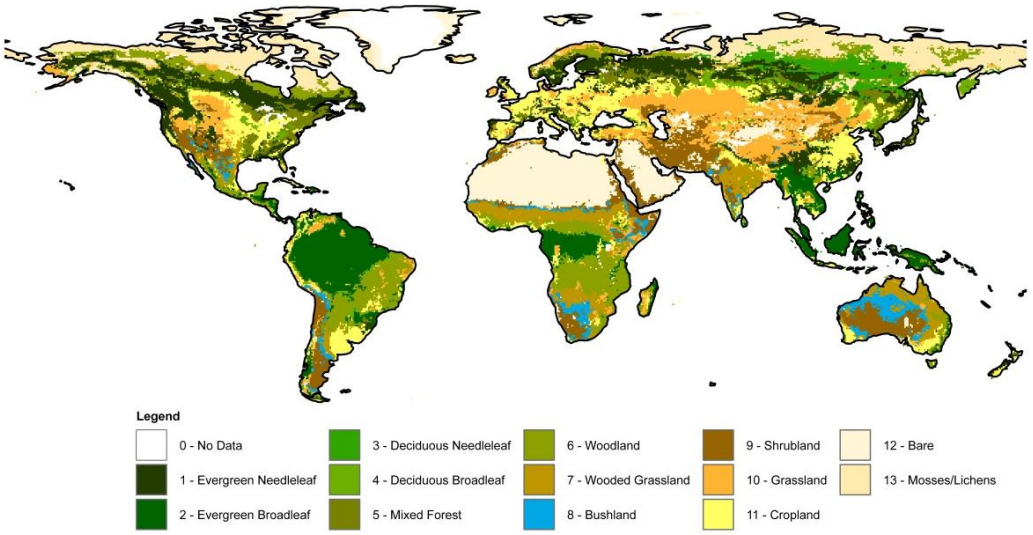


Figure 2.7 Previous Mac-PDM.09 land cover classification

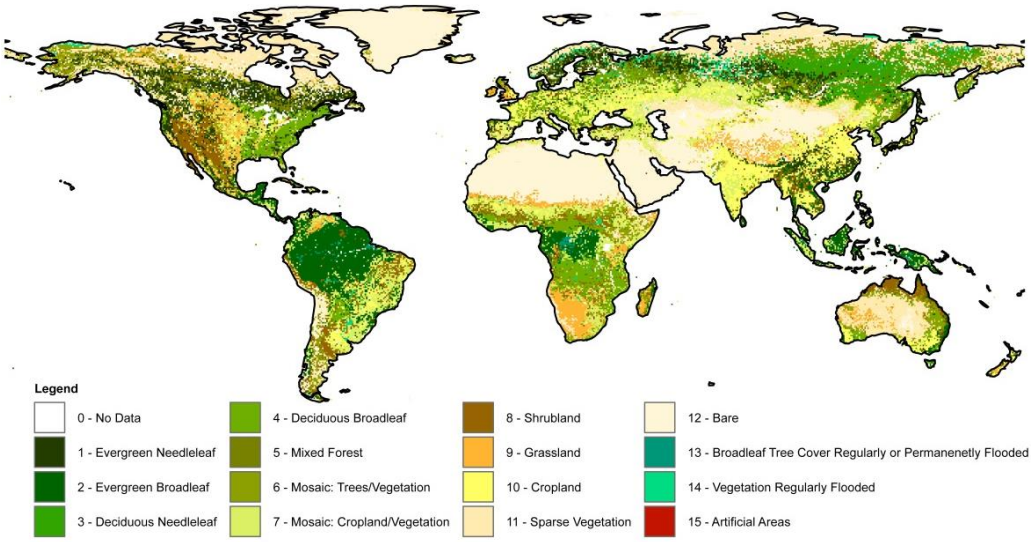


Figure 2.8 Updated Mac-PDM.09 land cover classification

There is a general reduction of grassland across the globe, especially across Central Asia, which is classified as sparse vegetation or bare in the new land cover classification. The Congo rainforest is significantly smaller, which may be in part due to deforestation between 1984 and 2009. Deforestation may be assumed due to the division of evergreen forest into eastern and western blocks by a band of cropland/vegetation mosaic; however, some of the evergreen forest reduction may be attributed to a reclassification of forest type to deciduous forests. India shows a distinct change from woodland/grassland to cropland and the

Middle East (excluding Saudi Arabia) sees a change from shrubland to sparse vegetation or bare. The forests of the Russian Arctic extend further north in the new land cover classification. Other than these distinct changes, which are mostly a result of the reclassification of vegetation schemes, the land covers are predominantly alike between maps. As previously mentioned though, fragmentation of the land cover types is apparent, especially the northerly bands of evergreen needle-leaf forest. The cropland of the USA is also interspersed with shrubland and grassland. This fragmentation of land cover types across the globe is more indicative of true global land cover at a 0.5° x 0.5° resolution.

A simulation experiment was conducted in order to compare outputs from the model using the new soil and vegetation maps with the original maps. As the model was not re-calibrated after the maps were changed, this experiment was primarily to check the model still yielded sensible outputs. The values used for the soil parameters are given in Table 2.5 and the land cover parameters are given in Table 2.6. Where the new classifications coincided with the original classifications, the parameter values were taken from the original input files (Arnell, 1999). The one exception to this was the sand parameters, which were matched to the source used to define the new soil classifications, taken from Saxton and Rawls (2006). For the vegetation parameters, those classifications that were retained from the Mac-PDM.09 classifications, were ascribed parameter values from the original input files, taken from Wilson and Henderson-Sellers (1985). Parameter values for the "Mosaic: trees/vegetation" classification were also available from this source. The other classifications were extrapolated from existing parameter values, taking physical meaning into consideration. For example, "artificial areas" was set to the same values as "bare", except for the $percov$ parameter (percentage grass in each cell), which was set a little higher to account for parks and gardens. The "broadleaf regularly or permanently flooded" classification was given parameter values between "evergreen broadleaf" and "deciduous broadleaf".

Table 2.5 Parameter values used for soil classifications in simulation experiment run. *Italicised round brackets indicate new values not from the original model documentation. Square brackets indicate original values that were not used.*

Soil Classification		
Type/Parameter	*fcpc*	*satpc*
Sand	10.0 [13.1]	46.0 [35.5]
Loamy Sand	*(12.0)*	*(46.0)*
Sandy Loam	20.0	41.3
Loam	*(28.0)*	*(46.0)*
Silt Loam	29.4	46.8
Silt	*(30)*	*(48.0)*
Clay Loam	33.1	50.4
Sandy Clay Loam	*(27.0)*	*(43.0)*
Silty Clay Loam	*(38.0)*	*(51.0)*
Sandy Clay	*(36.0)*	*(44.0)*
Silty Clay	*(14.0)*	*(52.0)*
Clay	48.3	54.4
Lithosols	27.0	50.0
Histosols	50.0	100.0

Table 2.6 Parameter values used for land cover classifications in simulation experiment run. *Italicised round brackets indicate new values not from the original model documentation.*

Land Cover Classification						
Type/Parameter	*rootg*	*rsc*	*capg*	*rlai*	*hc*	*percov*
Evergreen Needle-leaf	0.9	85	1.2	6	19.1	80
Evergreen Broadleaf	1.5	130	0.7	9	29.4	90
Deciduous Needle-leaf	0.9	85	1.0	4	10.0	80
Deciduous Broadleaf	1.2	100	0.6	5	14.9	80
Mixed Forest	1.1	100	0.8	6	18.0	80
Mosaic: Trees/Vegetation	*(1.1)*	*(100)*	*(0.8)*	*(6)*	*(18.0)*	*(25)*
Mosaic: Cropland/ Vegetation	*(0.9)*	*(90)*	*(0.6)*	*(4)*	*(7.0)*	*(15)*
Shrubland	0.6	80	1.0	2	1.4	25
Grassland	0.6	70	0.1	3	0.6	0
Cropland	1.2	100	0.6	5	14.9	10
Sparse Vegetation	*(0.2)*	*(100)*	*(0.2)*	*(1)*	*(0.2)*	*(90)*
Bare	0.1	100	0.0	0	0.0	90
Broadleaf Regularly or Permanently Flooded	*(1.3)*	*(110)*	*(0.6)*	*(7)*	*(22.0)*	*(85)*
Vegetation Regularly Flooded	*(0.6)*	*(90)*	*(0.8)*	*(4)*	*(5.0)*	*(15)*
Artificial Areas	*(0.1)*	*(100)*	*(0.0)*	*(0)*	*(0.0)*	*(95)*

The results of the comparison between model runs with the new and old soil and vegetation maps are presented in Figure 2.9. The main areas of difference are, unsurprisingly, some of the areas that have undergone the most dramatic changes in land cover classification. The areas that have seen the most significant increases in annual average runoff, of up to 13.5%, are Australia and Central Asia (Kazakhstan and Mongolia). In Central Asia this change was from grassland to sparse vegetation and bare, and in Australia the change was from bushland and shrubland to sparse vegetation and bare. Sparse vegetation and bare soils have less capacity to hold water than grassland, bushland and shrubland due to the lack of vegetation, and infiltration is less likely to occur, thus resulting in Hortonian overland flow. The lack of vegetation will also reduce interception. These physical factors contribute to the increase in runoff that can be seen in these areas.

There is also a band of slightly reduced runoff along the southern Sahara desert, where the land classification was altered from bare to grassland. This land cover change would increase the soil moisture storage capacity, increase infiltration, and increase interception, thus reducing runoff. The attribution of these results to physical processes provides confidence that the updated maps are performing well with the Mac-PDM model, and can now be used for an uncertainty assessment. This uncertainty analysis will, as an integral part of the process, investigate the appropriate parameter values of the newly classified land cover and soil types, it will also act to calibrate the model to the new land cover and soil texture classifications. This mapping work has produced a new version of the model: Mac-PDM.14.

2.4 Study Catchments

Since this research focuses on global hydrology, study catchments are required in order to validate the model against observed discharge data.

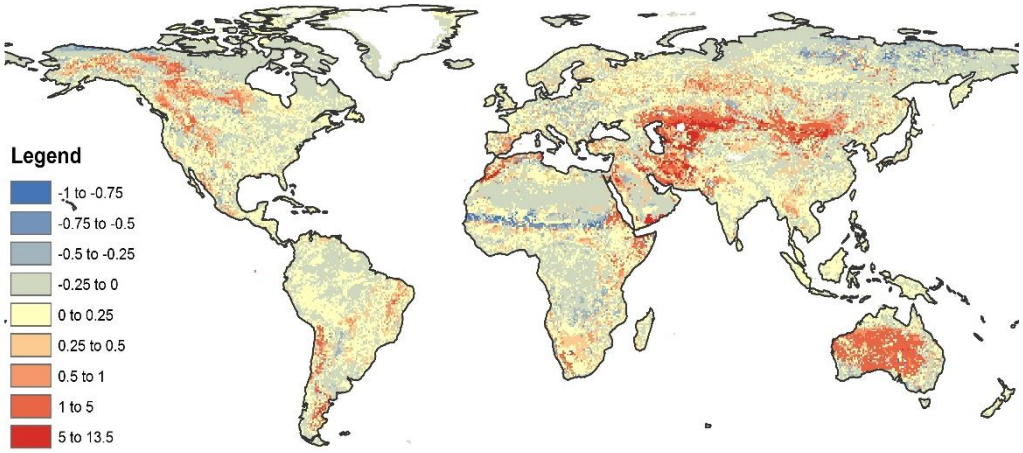


Figure 2.9 Percentage change in average annual runoff between runs using the original model maps and the updated maps for this study.

21 of the world's largest and most significant rivers were selected for investigation; the locations of these catchments is shown in Figure 2.10 and Table 2.7 gives descriptive statistics for each of the catchments. These study catchments were selected in order to represent the diversity of hydrological regimes across the globe.

Many factors were considered including catchment size, river length, discharge, rainfall and the location of the river outlet. The Amazon River has the largest catchment in the world at 6,869,000km^2 (Barthem et al., 2004), and has the highest average discharge of 220,800m^3/s. The River Nile is the longest river at 6,825km. The Danube flows through 19 countries, which makes it a significant management challenge. The Kolyma and the Amu Darya rivers are particularly dry in terms of simulated rainfall. The Murray Darling, despite its size has a very low river discharge. The Okavango has the lowest discharge of the selected catchments, but was chosen due to it being a large endorheic river basin (it does not flow out to the sea, but instead flows into the swampland of the Moremi Game Reserve. The Lena is the most northerly catchment which represents a snowmelt driven catchment, and the Murray Darling the most southerly.

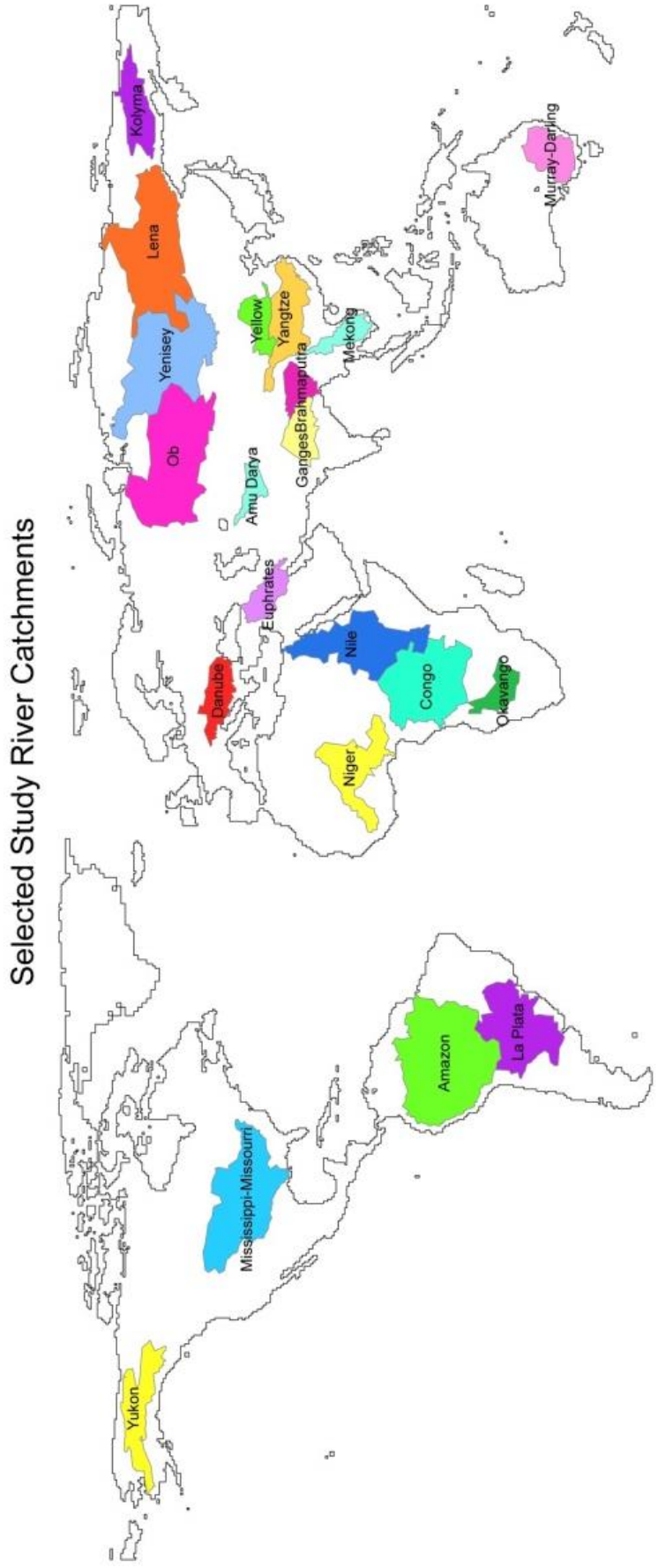


Figure 2.10 Map showing the spatial extent of the 21 selected river catchments.

Table 2.7 Descriptive statistics of the 21 chosen river catchments. *Average annual rainfall and runoff are Mac-PDM.09 (version 20) simulations for 1971-2000.*

River	Basin Area (km^2)	Length (km)	Ave Discharge (m^3/s)	No of Countries	Latitude of Mouth	Longitude of Mouth	Ave Annual Rainfall	Ave Annual Runoff	References
Amazon	6,869,000	6,400	220,800	7	-0.100	-49	2247	1247	Crist et al. (2014), Times Books (2009), Barthem et al. (2004)
Congo	3,691,000	4,374	39,640	11	-6.045	12.27	1541	797	Gupta (2007), Abe et al. (2004), Times Books (2009)
Yangtze	1,940,000	6,397	35,000	1	31.234	121.586	1058	461	Gupta (2007), Qu et al. (2004), Times Books (2009)
Brahmaputra	580,000	2,880	33,600	3	25.123	89.414	1839	1362	Gupta (2007), Singh et al. (2004) Times Books (2009)
La Plata	3,100,000	270	25,000	5	-35.316	-58.31	1331	449	Mugetti et al. (2004)
Yenisey	2,580,000	3,844	18,040	2	70.967	83.057	469	191	Gupta (2007), Times Books (2009)
Lena	2,472,000	4,337	16,200	1	72.627	129.221	381	142	Gupta (2007), Tsyban et al. (2005), Times Books (2009)
Mississippi	3,230,000	3,766	15,500	1	29.090	-89.151	889	943	Gupta (2007), Census Bureau (2009), Times Books (2009)
Ganges	952,000	2,525	15,000	2	22.05	90.5	1147	674	Gupta (2007), Jain et al. (2007), (Times Books (2009))

River	Basin Area (km^2)	Length (km)	Ave Discharge (m^3/s)	No of Countries	Latitude of Mouth	Longitude of Mouth	Ave Annual Rainfall	Ave Annual Runoff	References
Mekong	811,000	4,900	14,800	6	9.811	106.606	1637	830	Gupta (2007), Snidvongs and Teng (2006), Times Books (2009)
Ob	2,990,000	3,676	10,300	1	66.320	71.234	513	173	Gupta (2007), Tsyban et al. (2005),Times Books (2009)
Niger	2,113,200	4,023	8,500	8	5.1389	5.412	660	302	Lawson (1865), Abe et al. (2004)
Danube	816,000	2,857	6,855	19	45.133	29.454	806	323	Gupta (2007), Brilly (2010), Times Books (2009)
Yukon	850,000	3,187	6,120	2	62.256	-164.48	338	110	Gupta (2007), Census Bureau (2009), Times Books (2009)
Nile	3,349,000	6,825	2,810	10	30.1	31.06	598	221	Gupta (2007), Kliot (1993), Times Books (2009)
Kolyma	644,000	2,150	2,624	1	69.3	161.3	297	155	Tsyban et al. (2005)
Amu Darya	465,000	2,540	2,176	4	44.063	59.405	327	162	Alpas et al. (2011), Ellis et al. (1999)
Huang He (Yellow)	752,000	4,845	1,880	1	37.797	119.196	445	108	Gupta (2007), Hyndman (2008), Times Books (2009)
Euphrates	444,000	3,596	1,008	3	31.018	47.263	388	195	Cavendish (2006), Lowi (1995)
Murray Darling	1,061,469	2,740	391	1	-35.366	139.365	494	68	Gupta (2007), Gupta (2010), (Prideaux and Cooper, 2009), Times Books (2009)
Okavango	192,500	1,600	298	3	-19.146	23.001	525	121	Chauhan et al. (2014a), Mendelsohn and Obeid (2004)

The catchments were chosen in order to cover the globe as wholly as possible, including catchments from all continents, and all climates. The catchments' diversity across climatic types is shown in Figure 2.11. This figure shows that across the catchments, nearly all of the global climate classifications are included in this study. The Amazon, Congo and the Mekong rivers are "Tropical Humid". The Nile, Niger, Okavango and Euphrates rivers cover the "Dry Desert" and "Dry Steppe" environments. "Cool Humid" climates are represented by the Ob, Yenisey, Lena, Kolyma and Yukon catchments, and the Ganges, Brahmaputra, Yellow and Yangtze rivers are "Warm Humid". Since the model does not have a glacier component, the polar climates have not been considered in this study.

Several other global hydrology modelling studies have focussed on sets of catchments, and the majority of these have several catchments in common with those chosen in this study. Gosling et al. (2011) compared Mac-PDM.09 with catchment models for the Liard, Mekong, Okavango, Rio Grande, Xiangxi and Harper's Brook catchments. Kavetski et al. (2006) studied uncertainty in the VIC model over the Potomac and French Broad river catchments. Hagemann et al. (2011) looked at bias correction on the MPI-HM and LPJml models over the Mississippi, Amazon, Parana, Congo, Nile, Ganges, Brahmaputra, Murray, Yangtze, Amur, Danube, Baltic Sea, Kolyma, Ob, Lena, Yenisey, MacKenzie and Volga catchments. Sperna Weiland et al. (2010) applied the PCR-GLOBWB model to the Amazon, Brahmaputra, Congo, Danube, Ganges, Indus, Lena, Mackenzie, Mekong and Mississippi catchments. One final example is Döll et al. (2003), who mostly focussed on smaller catchments, but also included the Yenisey, Danube, Okavango and Mekong in their selection of 17 catchments when tuning and validating the WaterGAP model. This demonstrates that this selection of catchments is appropriate for global hydrology modelling, and should provide an adequate range of catchment behaviours for model testing.

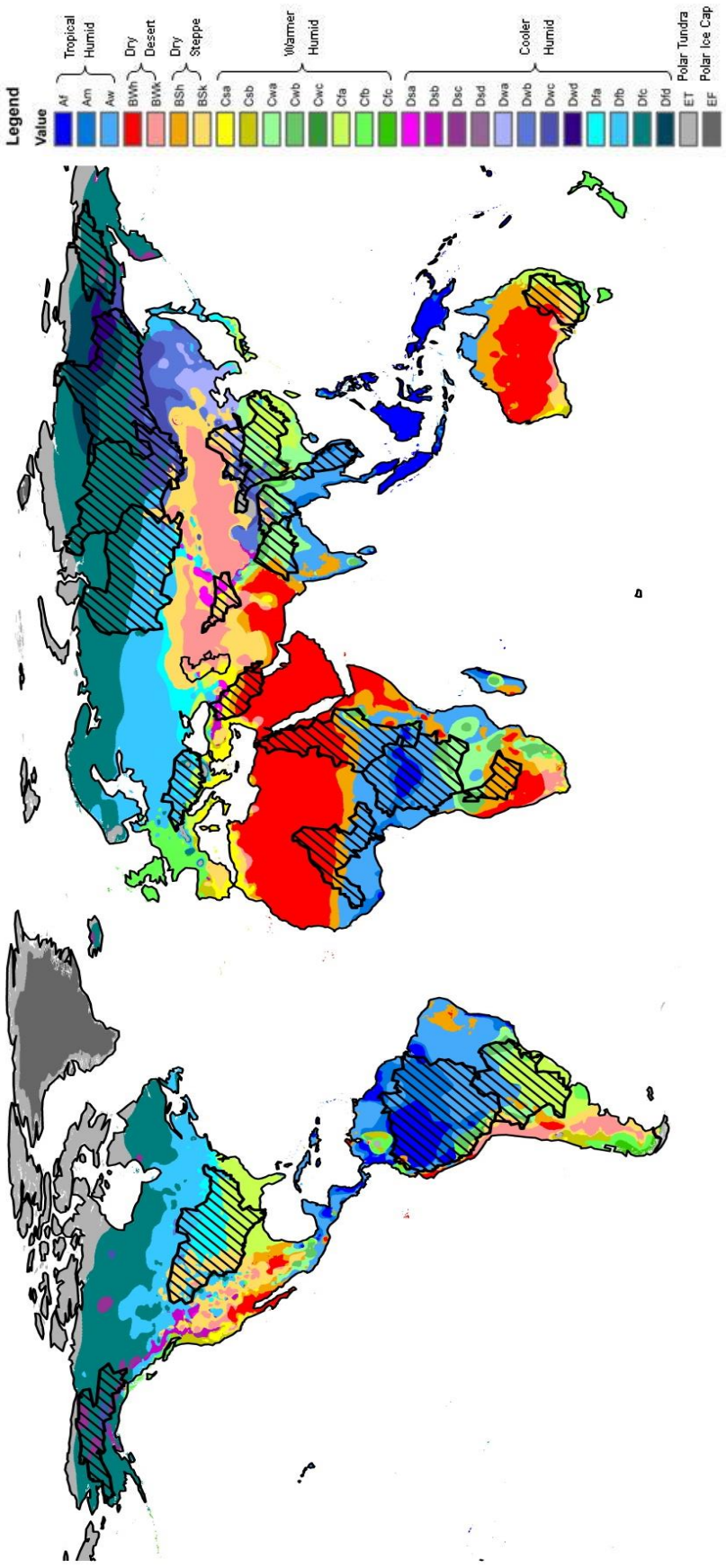


Figure 2.11 Köppen-Geiger Classification of global climates (Peel et al., 2007). *Chosen catchments shown in linear hatching*

2.5 Data Collection

This study seeks to assess the impacts of parameter uncertainty on model output; however both input data and validation data are also significant sources of uncertainty. Whilst this study will not go into detail on the uncertainties derived from these data sets, effort has been made to choose the best available data for use in this experiment.

2.5.1 Climate Forcing Data

Mac-PDM.09 has previously been applied using a variety of climate inputs or "forcing data". It has the capability of running using ClimGen, NCC, ClimatePrediction, CIAS, and WATCH data, as well as single catchment or multiple catchment data. NCC (Ngo-Duc et al., 2005) was developed for use by Land Surface Models (LSMs) in 2005 and covers the 53 year time period 1948-2001. The dataset is 6 hourly and 1° x 1°. It is based upon both the reanalysis products of NCEP/NCAR and the Climate Research Unit (CRU) observational data. ClimGen was developed by Tim Osborn of the Climate Research Unit and Tim Mitchell of the Tyndall Centre for Climate Change Research at the University of East Anglia (Osborne, 2009). This dataset uses pattern-scaling to generate monthly climate information based on climatological observations and outputs from GCM simulations. The forcing dataset covers the period 1961-1990, at a 0.5° x 0.5° resolution (Mitchell et al., 2004). The Climate Integrated Assessment System (CIAS) is a multi-institution modular and flexible integrated assessment system for modelling climate change (Warren et al., 2008). CIAS was developed with impacts models in mind, and the development of the system included assessing the ability of model outputs to be applied as inputs for impacts models. CIAS used the simple climate module, MAGICC, with the climate scenario downscaling module, DSM. The climate data developed is largely based upon the ClimGen data, and covers the time period 1901-2001 at 0.5° x 0.5°.

The WATCH forcing data was chosen for use in this study for several reasons. Firstly, Mac-PDM.09 was one of the models that took part in the WATCH project, so the model had a coded option for the WATCH input file format. The participation of Mac-PDM.09 in this project, and running the uncertainty assessment using the same data will allow the uncertainties of the parameter assessment to be compared to the model structural uncertainty derived from using different hydrology models. If a different climate input dataset were used, then this would contribute additional uncertainty and not allow for a fair comparison. Even with these reasons aside, the WATCH forcing data is an exceptional dataset, derived from the ERA-40 reanalysis project with sequential interpolation to a resolution of 0.5° x 0.5°, elevation corrections and monthly-scale adjustments based on CRU and GPCC monthly observations (Weedon et al., 2010).

The WATCH forcing data covers the period 1958-2001 and consists of eight variables, five at a 6-hourly time step (air temperature, pressure, specific humidity, wind speed, and long wave radiation flux) and three at a 3-hourly time step (short wave radiation flux, rainfall rate and snowfall rate). This data is also available at a daily time step, which was used in this study. Weedon et al. (2010) describe the key steps in the creation of the WATCH forcing data as:

1. Bilinear interpolation to the CRU half-degree grid,
2. Elevation correction of certain variables to account for differences in surface heights between the one- and half-degree grids, and
3. Adjustment of certain variables at the monthly scale via the CRU TS2.1 observations.

The data were compared to FLUXNET sites for additional validation, which showed close correspondence between the WATCH data and the observed data for all variables (Weedon et al., 2010).

2.5.2 River Discharge Validation Data

Acquisition of river discharge data for the validation of the Mac-PDM model was more challenging. There are two major global databases for runoff data, RivDis and the Global Rivers Data Centre (GRDC). RivDis (SAGE: Center for Sustainability and the Global Environment, 2014) is freely available online and contains records for over 3000 discharge stations. GRDC data is free of charge, but must be requested for specific stations after submitting a User Declaration. The GRDC database contains over 9000 records (GRDC, 2014), and was chosen as the primary resource for discharge data for this study. Records for all stations within the 21 study catchments were requested.

The spatial and temporal extent of these records is displayed in Figure 2.12. This map shows that the records are not evenly distributed either across or between catchments. The Mississippi River has by far the most records, which are also mostly over 50 years in length. The Danube, Murray Darling and Niger also have a good coverage of daily records, although the Niger catchment records are mostly in the upper reaches in Mali. In order to represent as much of the catchment as possible, discharge stations as near to the mouth of the river as possible were sought. Stations in the Lena catchment are sparse, however the map shows a few stations near the mouth that could be adequate. Catchments of concern were the Euphrates, Nile, Ganges and Brahmaputra. Several potential stations for each catchment were then selected and the data were analysed for length, period and integrity.

The thirty-year period 1971-2000 was selected for model validation, as the EU-WATCH forcing data did not extend beyond 2001 and a thirty-year period, as is common practice in climate averages (e.g. Met Office, 2015, NOAA, 2015), was deemed sufficient for model analysis. The best stations for each catchment were then selected.

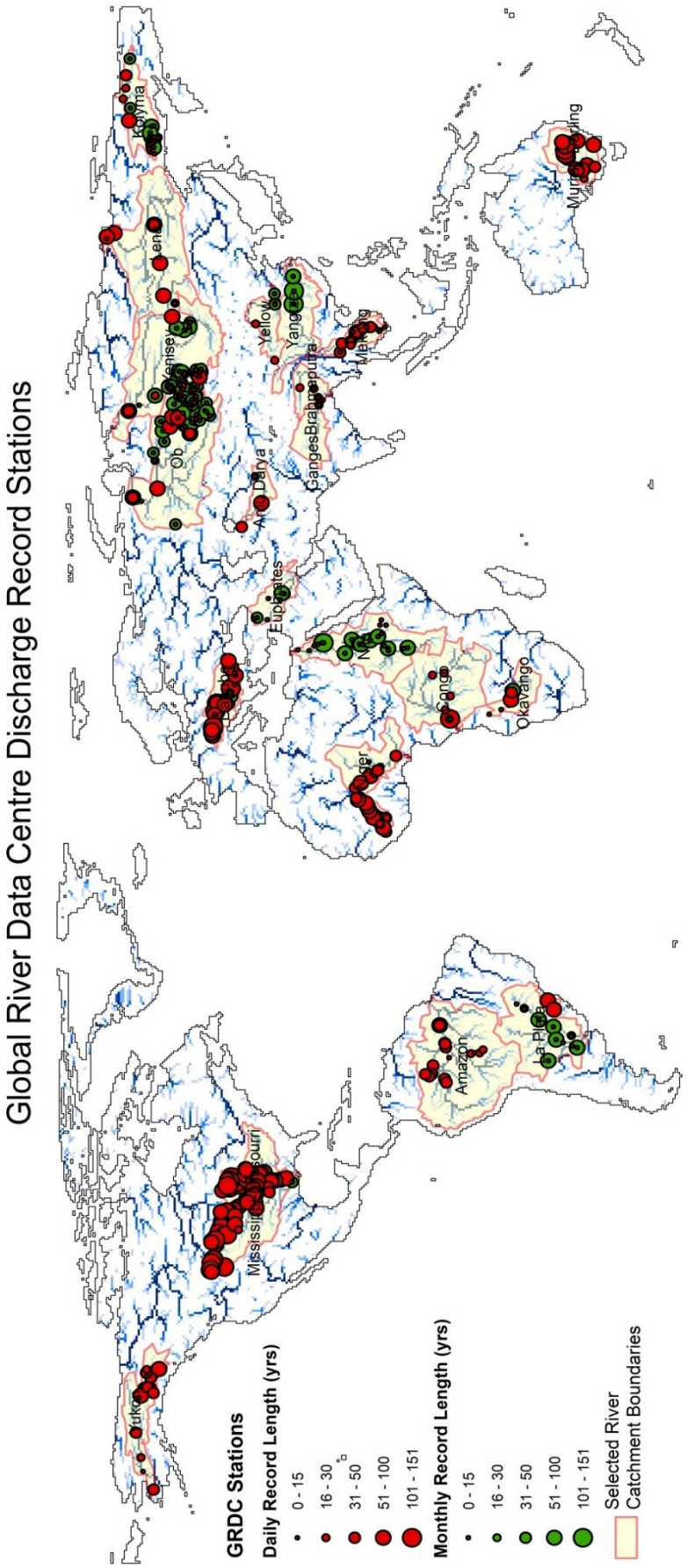


Figure 2.12 GRDC discharge station records for the 21 selected study catchments. *Green circles indicate monthly records and red circles indicate daily records. The size of the circles depicts the record length.*

Rivers that were highlighted as significantly lacking in data were the Euphrates, which had only 2 years of data, the Ganges, which had only 3 years of data, and the Nile, which had only 14 years. Fortunately discharge data for the Euphrates has been published by the United States Geological Survey (Saleh, 2010). This provided several records, the most appropriate of which yielded a 28.75 year record. Following a British Council funded visit by the author to Bangabandhu Sheikh Mujibar Rahman Agricultural University in Bangladesh, the Bangladesh Water Development Board (BWDB) kindly provided discharge data for several stations on the Ganges and the Brahmaputra Rivers. Since the Brahmaputra record from the BWDB was superior to the GRDC record, it was adopted for this catchment. Sadly, no additional records for the River Nile could be found, so the GRDC record of 14 years was retained. An overview of the records for each catchment is displayed in Table 2.8. These records will provide valuable data for the comparison of Mac-PDM.14 model outputs.

2.6 Summary

This chapter has outlined the preparatory work required before the uncertainty experiment could be carried out. The Mac-PDM model has been selected for use in this experiment and is considered to be a good choice for several reasons: (a) the ease of parameter perturbation; (b) its inclusion in the WATCH project and therefore ease of multi-model ensemble runs, as well as comparison with other hydrological models; (c) appropriate model outputs for analysis. The model structure was presented in detail and the model parameters were defined. Some adjustments to the maps used by the model were required. The soil texture map was updated from a 6 type classification to a 12 type, according to the USDA Soil Conservation Service Classification. The land cover map was also updated from a 1984 vegetation classification to a 2000-2009 map date.

Table 2.8 Discharge records chosen for validation of the Mac-PDM.09 model

River	Source	Station Code	Location	Contributing Area (km^2)	No of years available	% record complete
Amazon	GRDC	3629001	Obidos – Linigrafo, Brazil	4690962.5	30	98.33
Amu Darya	GRDC	2917110	Kerki, Turkmenistan	281930.9	19	66.667
Brahmaputra	BWDB	46.9L	Bahadurabad, Bangladesh	514170.4	30	94.167
Congo	GRDC	1147010	Kinshasha, Democratic Rep. of Congo	3634880.3	30	100
Danube	GRDC	6742900	Ceatal Izmail	777784.1	30	100
Euphrates	USGS	IRQ_E3	Downstream of Hindiya Barrage, Iraq	305192	28.75	89.17
Ganges	BWDB	90	Hardinge Bridge, Bangladesh	943236.6	30	94.44
Kolyma	GRDC	2998500	Sredne-Kolymsk, Russia	368675.5	30	98.889
La Plata	GRDC	3265601	Timbues, Argentina	2527003	24	78.61
Lena	GRDC	2903430	Stolb, Russia	2444557	30	100
Mekong	GRDC	2469260	Pakse, Laos	549046.8	23	76.67
Mississippi	GRDC	4127800	Vicksburg, Mississippi	2914994	30	100
Murray Darling	GRDC	520451	Burtundy, Australia	469893.3	30	100
Niger	GRDC	1834101	Lokoja, Niger	2069196.5	30	88.61
Nile	GRDC	1362600	Aswan Dam, Egypt	2764160	14	46.67
Ob	GRDC	2912600	Salekhard, Russia	2441938.8	30	96.67
Okavango	GRDC	1357100	Mohembo Mtaembo, Botswana	231011.5	30	100
Yangtze	GRDC	2181900	Datong, China	1679568.8	18	60
Yellow	GRDC	2180800	Huayuankou, China	732355.5	18	60
Yenisey	GRDC	2909150	Igarka, Russia	2419866.5	30	100
Yukon	GRDC	4103200	Pilot Station, Alaska	819519.1	26	70

The land cover types were defined using a combination of several popular methods that would allow for a detailed representation of global land cover without over-parameterisation. This work produced a new version of the model which will, from hereon in, be referred to as Mac-PDM.14. 21 of the world's largest river catchments have been chosen as study catchments in order to represent a wide range of environments and catchment behaviours. WATCH climate data is used to force the model for the years 1971-2000, and validation data has been acquired from the GRDC, USGS and BWDB.

The research presented in this chapter provides the foundations required to progress to the uncertainty experiment which will be introduced in Chapter 3. The results of this experiment are presented in Chapters 4 and 5.

Chapter Three:

Parameter Uncertainty in Global Hydrology Modelling Part 1

- Methods and Experimental Design

3.1 Introduction

> "As we know, there are known knowns, there are things we know we know. We also know that there are known unknowns, that is to say, we know there are some things we do not know. But there are also unknown unknowns, the one's we don't know we don't know"
>
> – Donald Rumsfeld

The issue of uncertainties in global hydrology models was presented in chapter 1. Previously, global models have been assessed for their structural uncertainties using multi-model ensembles, but have not been thoroughly assessed for parameter uncertainty. Several studies have investigated changing the values of a few parameters, the soil moisture storage capacity parameter being the focus of research (e.g. Gosling and Arnell, 2011, Alcamo et al., 2003, Nijssen et al., 2001a); however few have investigated more than a few parameters, and none have included all model parameters. The most comprehensive assessment has been by Sperna Weiland (2011) who ran 250 realisations of 10 model parameters, the results of this investigation were outlined in Chapter 1.6.

Following the work detailed in Chapter 2 which provided a new version of the model, Mac-PDM.14, this chapter outlines the methods used to carry out an extensive uncertainty analysis on this GHM. In this chapter, the distinction between sensitivity analysis and uncertainty estimation is drawn, and existing methods of uncertainty estimation are reviewed. Methods of parameter value sampling are presented, and the method used in this study is detailed. Parameter distributions are used in this study, so the technique of distribution definition is presented. The results of the uncertainty experiment are presented in Chapter 4.

3.2 Model Calibration, Parameter Estimation, Sensitivity Analysis and Uncertainty Estimation

The terms “model calibration”, “parameter estimation”, “sensitivity analysis” and “uncertainty analysis/estimation” are all used to describe very similar concepts in hydrology modelling. Model calibration can be defined as “the process of adjusting parameter values of a model to obtain a better fit between observed and predicted variables. [It] may be done manually or using an automatic calibration algorithm” (Beven, 2009). Both sensitivity analysis and uncertainty estimation contain this process, but they go further to understand the variation of outputs that different parameter values achieve. The simple, traditional approach to model calibration, whereby trial and error is used to adjust parameter values until the model output best meets observed data has some limitations, for example: calibration assumes that there is an optimum set of model parameter values; calibrated model parameter values may only be applicable to that particular model; the choice of method of comparison to the observed data will affect which parameter values are determined to perform best, and may be biased towards the calibrator’s specified use of the model (e.g. flood estimation); and adjustments to some parameters may impact the model output more than others, (Beven, 2012).

Sensitivity analysis and uncertainty estimation are both methods of assessing models’ responses to parameter values and structural changes, however they vary in their ultimate purpose. Sensitivity analysis can be defined as: “the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input” (Saltelli et al., 2008). Uncertainty analysis, on the other hand, focuses upon quantifying the uncertainty in model output. Tao (2008) states that sensitivity and uncertainty analyses are not explicitly related to model calibration, as some models may not require a formal calibration to estimate

parameters. In this case, uncertainty estimates may come from prior knowledge or past experience of the system; however, when calibration is required, this can be used as a posterior for uncertainty estimation. Saltelli et al. (2000) give six aims of sensitivity analysis, to determine:

1. if a model resembles the system or process under study;
2. the factors that mostly contribute to the output variability and that require additional research to strengthen the knowledge base;
3. the model parameters (or parts of the model itself) that are insignificant, and that can be eliminated from the final model;
4. if there is some region of the space of input factors for which the model variation is maximum;
5. the optimal region within the space of the factors for use in a subsequent calibration study;
6. if and which (group of) factors interact with each other.

Sensitivity analysis can be either local or global. Local sensitivity analysis (LSA) explores a local area of the parameter space, centred on nominal values; whereas global sensitivity analysis (GSA) extensively explores wide ranges of parameter space (Tao, 2008). GSA therefore comes with a much greater computational cost than LSA. However, derivative-based LSA requires more of the analyst's time to set up and carry out, which is difficult if the model parameters are uncertain or of unknown linearity (Saltelli et al., 2008, Wainwright et al., 2014). Common methods of sensitivity analysis include: one at a time (OAT) (Daniel, 1973, Daniel, 1958), the Morris method (Morris, 1991), principal component analysis (PCA) (Vajda et al., 1985), Monte Carlo (MC) analysis, Sobol' sensitivity indices (Sobol', 1993), and the Fourier Amplitude Sensitivity Test (FAST) (Cukier et al., 1973, Cukier et al., 1978). These methods are briefly described in turn here:

- *One-at-a-time* is a screening method that evaluates the effect of changing each parameter one by one on the model output. The output

of the perturbed parameter model is compared to a 'standard' value, usually in the middle of a set of parameter perturbation values.

- *The Morris method* is a global method variation of OAT that moves around the parameter space one parameter at a time, but does not return the previous parameter change back to its standard value. It is an economic method in that the number of experimental runs is proportional to the number of input parameters (Saltelli et al., 2000).
- *Principal Component Analysis* is a sophisticated method that uses linear sensitivity coefficients to extract meaningful kinetic information for several species of reactions at several time points (Saltelli et al., 2000). PCA uses eigenvectors and eigenvalues to reveal parts of the model that strongly interact, and their associated model response.
- *Monte Carlo analysis* uses randomly selected points in the parameter space to run the model, and then uses the results to determine uncertainty in model prediction, and the contribution of parameter inputs to this uncertainty. Monte Carlo is a sampling strategy that may be used in other forms of sensitivity or uncertainty analyses.
- *Sobol' analysis* produces sensitivity indices and identifies the influence of each parameter, interaction of parameters and their combination effects on the model outputs (Sobol', 1993). It is a popular method in hydrological model sensitivity analysis as it considers the interaction of model parameters (Qi et al., 2013).
- *FAST* is an alternative method to compute the same indices as the Sobol' method, however calculations are often limited to the first-order, or main effect.

Generally, sensitivity analysis is distinct from uncertainty analysis, though many studies have used a combined approach (e.g. Ratto et al., 2001, Kiczko et al., 2007). Uncertainty analysis aims to define the entire set of possible outcomes, along with their associated probabilities of occurrence. Sensitivity analysis however, as outlined above, aims to define the change in model output values that result from small changes in input values, and thus measures change in a localised region of the

parameter space (Loucks et al., 2005). Loucks et al. (2005) give five achievable outcomes of an uncertainty analysis:

1. a description of the range of potential outputs of the system at some probability level (e.g. the mean and standard deviation of the outputs).
2. an estimation of the probability that the output will exceed a specific threshold of performance measure target value.
3. the assignment of a reliability level to a function of the outputs, e.g. the range of function values that is likely to occur with some probability.
4. a description of the likelihood of different potential outputs of the system.
5. an estimate of the relative impacts of input variable uncertainties.

Methods of uncertainty analysis are discussed in more detail in section 3.4. Figure 3.1 shows the impact of both input data sensitivity and input data uncertainty on model output sensitivity. This figure demonstrates that input parameter uncertainty and model sensitivity combined can lead to high levels of output uncertainty.

3.3 A One-at-a-Time Sensitivity Analysis of the Mac-PDM.09 Model

Sensitivity analyses can differ hugely in complexity, especially between local and global methods. Since this study aims to focus on model uncertainty, a basic one-at-a-time sensitivity analysis was carried out at the very beginning of the study (using Mac-PDM.09, prior to the development of Mac-PDM.14), in order to understand the relative importance of each of the model parameters for the model output. This sensitivity analysis could also provide insight into whether it would be necessary to include all of the model parameters in the uncertainty experiment or not.

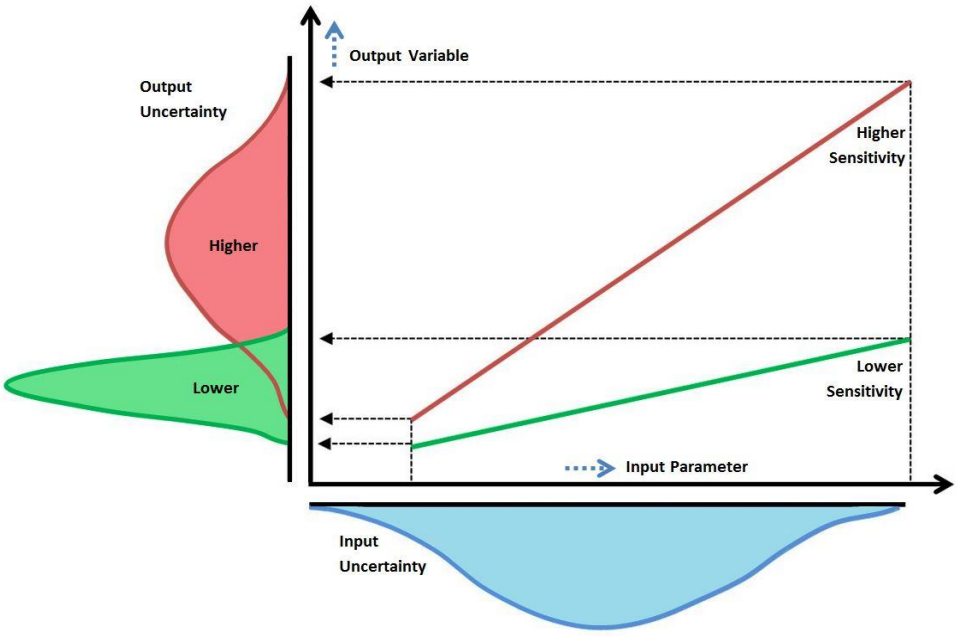


Figure 3.1 The relationship between model input parameter uncertainty and sensitivity to model output variable uncertainty. *After Loucks et al. (2005).*

The one-at-a-time sensitivity analysis varied each parameter systematically. The parameter values for the control file were varied by 0-200% of their base value (the original calibration value for Mac-PDM.09). Percentages of 0, 20, 40, 60, 80, 90, 95, 100 (BV), 105, 110, 120, 140, 160, 180 and 200% were used in each case. This required 14 model runs per parameter. The soil parameters were varied between 0 and 200% of their base values, changing all soil types at once (unless the increase took the value above a value of 100, in which case 100 was used), and were then varied between the values of 0 and 100 simultaneously at increments of 10, as well as one at a time while keeping the other soil types at their base values. This required 103 model runs per parameter. The vegetation parameters were also varied simultaneously by 0-200% of their base values and were then varied one by one. This required 210 models runs per parameter. The base values of the parameters are given in Table 3.1, 3.2 and 3.3. The *fact* parameter was excluded from this study, as well as the ultimate uncertainty experiment, as it is a scaling factor for the *fcpc* and *satpc* parameters. Thus *fact* was fixed at a value of 1.0 and the *fcpc* and *satpc* parameters were investigated individually instead.

Table 3.1 Base values of the Control File parameters

Parameter	*b*	δ	*grout*	*srout*	*xmelt*
Base Value	0.5	0.5	1.0	0.5	0.4

Table 3.3 Base values of the Soil Parameters

Soil Classification		
Type/Parameter	*fcpc*	*satpc*
Sand	13.1	35.5
Sandy Loam	20.0	41.3
Silt Loam	29.4	46.8
Clay Loam	33.1	50.4
Clay	48.3	54.4
Lithosols	27.0	50.0
Histosols	50.0	100.0

Table 3.2 Base values of the Land Cover parameters

Land Cover Classification						
Type/Parameter	*rootg*	*rsc*	*capg*	*rlai*	*hc*	*percov*
Evergreen Needleleaf	0.9	85	1.2	6	19.1	80
Evergreen Broadleaf	1.5	130	0.7	9	29.4	90
Deciduous Needleleaf	0.9	85	1.0	4	10.0	80
Deciduous Broadleaf	1.2	100	0.6	5	15.9	80
Mixed Forests	1.1	100	0.8	6	18.0	80
Woodlands	1.1	100	0.8	6	18.0	50
Wooded Grasslands	1.1	100	0.8	6	18.0	25
Closed Bushlands	0.9	80	1.0	3	1.7	40
Open Shrublands	0.6	80	1.0	2	1.4	25
Grass	0.6	70	0.1	3	0.6	0
Croplands	1.2	100	0.6	5	14.9	10
Bare	0.1	100	0.0	0	0.0	90
Mosses/Lichens	0.1	100	0.0	0	0.1	90

The results of this experiment are shown in Figure 3.2 and Figure 3.3. The percentage change in average annual runoff between the perturbed simulation and base value simulation, for the years 1971-2000, was averaged across the cells of each catchment (taking into account variations in grid cell size due to latitude), and the minimum and maximum changes across all 21 catchments were determined for each parameter perturbation.

The graphs in Figure 3.2 show this percentage change response across the varying parameter values for each model parameter. The range in colour fill areas indicates the range in response across all 21 study catchments. The runs that used a parameter value of 0, or 0% of the parameter base value were not included in these graphs, as division by 0 results in infinity; and this led to some extreme changes in the model output at values of 0. The model varies in sensitivity to changes in parameter values. The *satpc*, and *rootg* parameters show the highest levels of sensitivity, reaching just under a 1.2% increase in average annual runoff. The *rlai* parameter is the least sensitive, with a maximum change of + 0.0435%. Of the control file parameters, the *b* parameter is the most sensitive, showing a definite decrease in average annual runoff with reduction in the value of *b,* and a notable increase with larger values of *b*. The *delta* and *xmelt* parameters also show sensitivity, but the *grout* and *srout* parameters show little change when perturbed individually.

Of the soil types, *sandy loam* showed the highest sensitivity in perturbations of both *fcpc* and *satpc*. *Silt loam* was the second most sensitive in both parameters. The sensitivity of the model to the *fcpc* parameter is greatest at higher values; conversely, the sensitivity of the model to the *satpc* parameter is largest at lower values. This is due to the physical meaning of the parameters as field capacity and saturation capacity, and the logical requirement that saturation capacity be greater than field capacity for any given soil type.

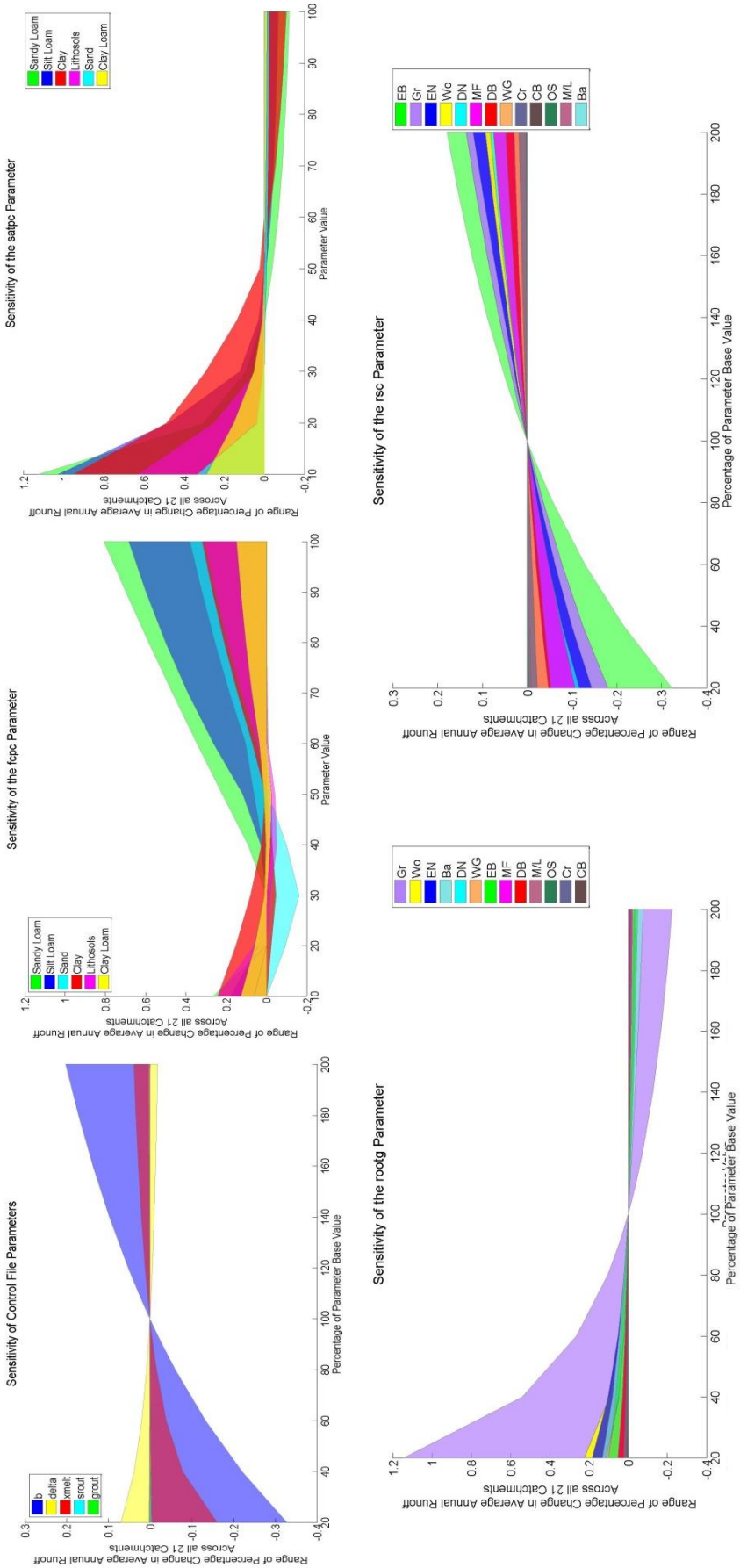


Figure 3.2 (continued on next page) *Vegetation classification codes read as follows: EN=Evergreen Needleleaf; EB=Evergreen Broadleaf; DN=Deciduous Needleleaf; DB=Deciduous Broadleaf; MF=Mixed Forests; Wo=Woodland; WG=Wooded Grassland; CB=Closed Bushland; OS=Open Shrubland; Gr=Grass; Cr = Cropland; Ba = Bare; M?l = Mosses/Lichens.*

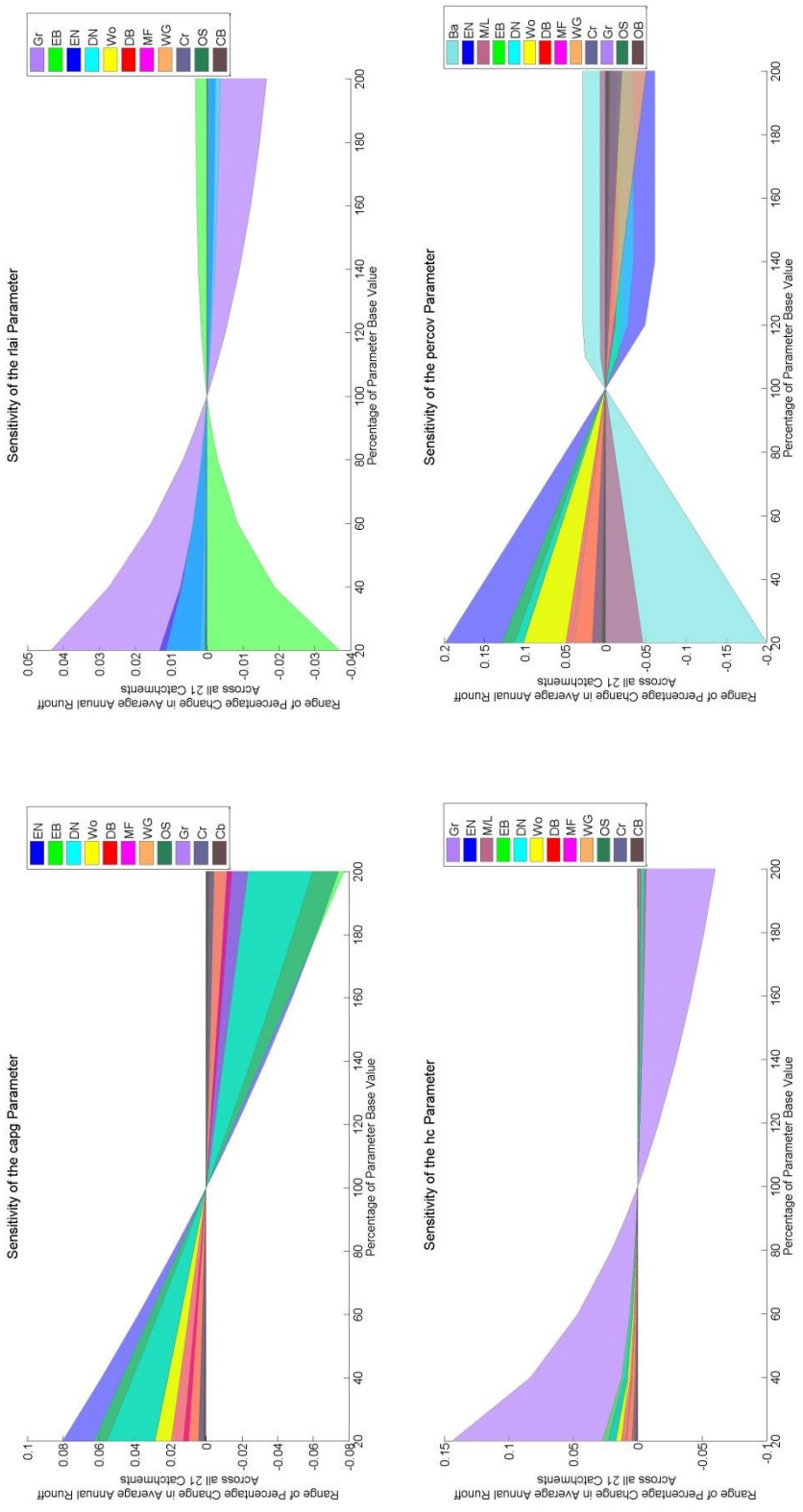


Figure 3.2 Sensitivity of the Mac-PDM.09 model to one-at-a-time parameter perturbations. *Control File and vegetation parameters were perturbed by +/- 100%. Soil parameters were perturbed between the values of 0 and 100. Ranges in colour fill indicates the range in response across the 21 study catchments, and vegetation class (soil and vegetation parameters). Parameter value runs using a value of 0 or 0% were discounted due to erroneous results derived from mathematical instability*

Of the vegetation parameters, *rootg* (root depth) shows the highest sensitivity to parameter perturbations. The vegetation classification with the highest sensitivity is *grass*, which showed an increase in average annual runoff of 1.145% with a value at 20% (0.12) of the base value, and also showed a decrease with an increase in value, reaching -0.223% at double the base value. *Grass* is also the most sensitive vegetation type for the *rlai* (relative leaf area index) and *hc* (vegetation height) parameters. *Evergreen Needleleaf* is the most sensitive vegetation type for the *capg* (interception parameter) and the *percov* (percent cover of grass) parameters. The *percov* graph shows steady rates of increase in change in average annual runoff as the parameter values are decreased for each vegetation type. The trend in average annual runoff with as *percov* parameter values are increased appears to be more complex, but can be explained by the fact that the parameter values were increased until they reached a value of 100 (as the parameter is expressed as a percentage, a value greater than 100 is not possible) after which they were kept at 100.

All vegetation parameter results show changes that differ in sign as they pass the base value mark (100%). None of the parameters give parabolic results, whereby the same model output could be achieved by more than one value of the parameter. The *capg* parameter shows a very linear trend, with sensitivity apparent with both increases and decreases in the parameter values. The *rlai* parameter is less linear, with much higher sensitivities to decreases in parameter value than increases in parameter value. Increases in the *rlai* parameter values above 120% show little change to the model sensitivity for all vegetation types except *grass.* A similar trend can be seen from the *rootg* parameter graph. Interestingly, for the *rlai* parameter, *Evergreen Broadleaf* shows the opposite trend to most other vegetation types, with average annual runoff increasing as *rlai* increases.

The *Mosses/Lichens* and *Bare* classifications were excluded from the *capg* and *rlai* sensitivity graphs as they began at base value of 0, and thus perturbation by percentage did not result in any change. *Bare* was also excluded from the *hc* graph for the same reason. In the actual experiments, both *Mosses/Lichens* and *Bare* were varied between 0 and 9 for the *rlai* parameter, which had no effect on the model output over any of the study catchments. For the *hc* parameter, *Bare* was varied between 0 and 1, which had a maximum effect of -0.126% at a value of 1. For *capg*, *Mosses/Lichens* and *bare* were both varied between 0 and 1, which had maximum effects of -0.11% and -0.068% respectively.

Figure 3.3 shows the variation of sensitivity between parameters and catchments. This figure was derived from the 15 model runs per parameter that varied the parameters from 0-200% of the original base value. In these model runs, for each parameter, the soil and vegetation types were all varied at the same time. So, for each parameter, with runs at 0, 20, 40, 60, 80, 90, 95, 100, 105, 110, 120, 140, 160, 180, and 200%, all vegetation and soil types were set at that percentage of the base value, whilst the other parameters remained at their base value. As with the previous graphs, the 0% values were excluded from this analysis. The figure shows the maximum response in each catchment from the 15 runs that were employed, the sign and size of which is indicated by colour. The most notable overall trend is that the *fcpc* and *satpc* parameters are the most sensitive. The *rootg* parameter is also very sensitive, giving significant increases in average annual runoff in most catchments. The *rsc* parameter shows the strongest negative response, with maximum change reducing the average annual runoff over most catchments. The strongest reduction in average annual runoff is for the *rsc* parameter in the Ob catchment, with a -0.37% decrease. It is also sensitive in the many other catchments, with no real trend in climatological zone.

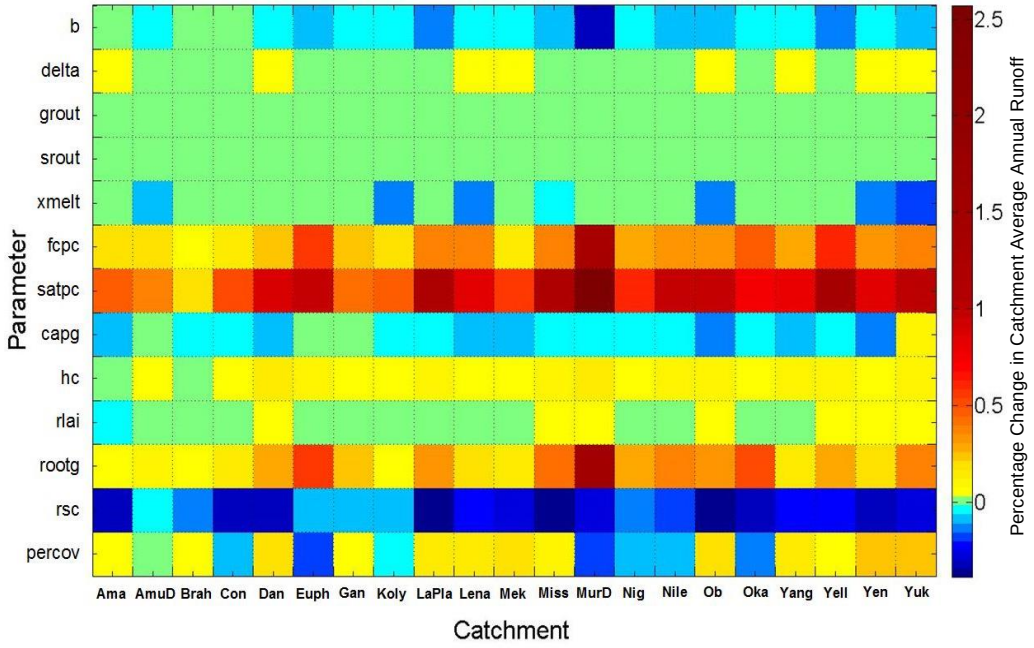


Figure 3.3 Maximum sensitivity of the Mac-PDM.09 model to parameter adjustments (values set at 0-200% of the base value) over each catchment. *Sensitivity is given as percentage change in catchment average annual runoff. For soil and vegetation parameters the results shown are from simultaneous perturbations of all soil and vegetation classes. Catchment codes read as follows: Ama=Amazon; AmuD=Amu Darya; Brah=Brahmaputra; Con=Congo; Dan=Danube; Euph=Euphrates; Gan=Ganges; Koly=Kolyma; LaPla=La Plata; Lena=Lena; Mek=Mekong; Miss=Mississippi; MurD=Murray Darling; Nig=Niger; Nile=Nile; Ob=Ob; Oka=Okavango; Yang=Yantze; Yell=Yellow; Yen=Yenisey; Yuk=Yukon.*

The strongest positive change in average annual runoff is for the *satpc* parameter over the Murray Darling catchment with an increase of 2.57% across all catchments. The *fcpc* shows a similar response, but with slightly lower increases than *satpc*. The high impact of changes in field and saturation capacity in the Murray Darling catchment is likely due to the fact that the catchment has the lowest average annual discharge of all 21 study catchments, and it receives very low annual precipitation.

As with the graphs in Figure 3.2, for the control file parameters, the *b* parameter shows the strongest trend, with decreases in average annual runoff, particularly in the Murray Darling and Kolyma catchments. The *xmelt* parameter shows sensitivity in the Yukon, Yenisey, Ob, Lena, Kolyma and Amu Darya catchments. This is unsurprising as *xmelt* defines the snow melt rate, and these are the catchments that have a

significant snowmelt contribution to their runoff. The *srout* and *grout* parameters are insensitive to change in this one-at-a-time analysis. The *delta* parameter and the *capg* parameter mirror each other, with *delta* showing slight increases in runoff over the Amazon, Danube, Lena, Mekong, Ob, Yangtze, Yenisey and Yukon catchments; whilst *capg* generally shows its strongest decreases in runoff over the same catchments (with the exception of the Yukon). The *capg* and *delta* parameters together define the amount of precipitation that is intercepted by vegetation, so it is reassuring that they show their strongest trends in the same catchments. The *rootg* parameter shows a fairly significant (1.53%) increase in the Murray Darling catchment, and also shows increases over the Euphrates and Okavango catchments. These catchments also experience decreases in average annual runoff with adjustment of the *percov* parameter. It is apparent that the Murray Darling, and Euphrates catchments are the most sensitive catchments to parameter perturbations. The La Plata, Ob, Lena, Mississippi and Yenisey can also be distinguished.

The fact that the soil parameters showed such significant sensitivity confirmed the requirement for an update of the model's soil classification system (see Chapter 2.3). Since these parameters have such a dramatic influence on the model output, it is necessary to define the soil textures across the world as accurately as possible. Similarly, the *Closed Bushland*, *Open Shrubland,* and *Mosses/Lichens* do not show significant sensitivity for any parameter changes, which might suggest obsolete vegetation types. This, coupled with the fact that the original vegetation map was significantly out of date, aided the decision to update the model's land cover classification system and map.

Whilst the results of this sensitivity assessment are very interesting and informative, they are merely a first step in model assessment. A one-at-a-time sensitivity analysis does not consider how the parameters interact with one another. It may seem that the *grout* and *srout*

parameters do not have much purpose in the model, as perturbations to these parameters do not alter the model output, however, it may be that these parameters interact with other parameters in the model, to have a secondary impact on model output. Therefore, a simultaneous parameter perturbation approach must be sought to achieve a comprehensive uncertainty assessment.

3.4 Methods of Parameter Uncertainty Analysis

This section will discuss five popular methods of simultaneous perturbed parameter uncertainty analysis: Generalised Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992), the Model-Independent Parameter Estimation & Uncertainty Analysis software package (PEST) (Doherty, 2010), the Shuffled Complex Evolution Metropolis Uncertainty Analysis (SCEM-UA) (Vrugt et al., 2002, Vrugt et al., 2003a), the differential evolution adaptive metropolis scheme (DREAM) (Vrugt et al., 2008, Vrugt et al., 2009a), and the Bayesian recursive estimation technique (BaRe) (Thiemann et al., 2001). Other methods, that are not discussed here in detail for the sake of brevity, include the Dynamic Identifiability Analysis Framework (DYNIA) (Wagener et al., 2003), the maximum likelihood Bayesian averaging method (MLBMA) (Neuman, 2003), dual state parameter estimation methods (Moradkhani et al., 2005a, Moradkhani et al., 2005b), and the simultaneous optimization and data assimilation algorithm (SODA) (Vrugt et al., 2005).

3.4.1 Generalised Likelihood Uncertainty Estimation

Generalised Likelihood Uncertainty Estimation (GLUE) is by far the most popular method of uncertainty analysis in hydrological modelling, and has been applied to numerous catchment scale models (e.g. Smith, 2011, McMichael et al., 2006, Cameron et al., 1999, Hossain et al., 2004). The GLUE methodology was developed by Beven and Binley (1992), and was inspired by Hornberger and Spear's (1981) method of

sensitivity analysis (Vrugt et al., 2009b). GLUE methodology aims to address the issue of “equifinality” in models. The equifinality concept originates from the notion that there can be no single correct or optimal model. Equifinality describes how different sets of model parameters may lead to an equally good model performance. A simple illustration of this would be to take a simple linear equation: $a + b + c = d$. If we had an observation of the value d that was 9, there are many possible combinations of a, b and c that could provide that answer. Using integers alone (0-9), there are 55 possible combinations that would result in the answer 9. In hydrology modelling, the same issue applies. Different sets of values may lead to similar model outputs, and using a Monte Carlo sample, one would expect to see both good and bad model outputs across a wide range of values for each model parameter, depending on the values of other parameters. This means that the ‘goodness’ of a model does not depend upon individual parameters, but on the whole set of parameter values, and the interactions between the parameters. Given that the structure of the model is adequate, unrealistic parameter combinations will lead to poor model results.

GLUE uses this theory to produce a set of ‘good’ models that are taken forward for use in model predictions and projections. GLUE uses prior distributions of parameter values to generate random sets of parameters using Monte Carlo simulation. The results of the model runs are then compared to observed data using a likelihood measure to assess the acceptability of each model based on the residuals. A specific likelihood measure is not defined, but is left for the modeller to determine according to their requirements. Models that reach a certain threshold in the likelihood measure are defined as “behavioural” and those that don’t, “non-behavioural”. When the model is used for projections, the behavioural models all contribute to the distribution of the projection, and are weighted according to their likelihood measure (Beven, 2012). Thus, there are several moments that introduce subjectivity in the GLUE process: when choosing feasible parameter

ranges and distributions; when defining a sampling strategy; when deciding upon a likelihood measure; and when determining the conditions upon which a model is accepted as behavioural or rejected as non-behavioural (Beven, 2012).

There has been significant debate in the literature surrounding the GLUE methodology, which has focused on the fact that GLUE is not formally Bayesian and is rather subjective in its approach. There have been three central debates in the literature, between those that believe GLUE is a useful working methodology for assessing uncertainty, and those that prefer to use more formal probabilistic approaches (Vrugt et al., 2009b). The provoking papers in these debates were “On undermining the science” (Beven, 2006b), “Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology” (Mantovan and Todini, 2006) , and “Pursuing the method of multiple working hypotheses for hydrological modeling” (Clark et al., 2011).

The “On undermining science” debate was initiated by Keith Beven (Beven, 2006b), who asked whether uncertainties in models are overestimated by GLUE or other uncertainty estimation techniques, whether showing the results of uncertainty analyses to users and stakeholders would undermine their confidence in science, and how uncertainties could be constrained in future to improve model results. He concluded that uncertainty analysis need not undermine science, but called for better evaluation of uncertainty in hydrological models. Several replies suggested that whilst uncertainty need not undermine science, the concept of uncertainty needs to be better defined, and methods of uncertainty analysis better developed (Todini and Mantovan, 2007, Hall et al., 2007). It was also suggested that uncertainty is all too often an afterthought in model development (Hall et al., 2007) and that uncertainties need to be made explicit in communications with end-users (Andréassian et al., 2007).

The “incoherence of GLUE” debate was sparked by Mantovan and Todini (2006), who challenged the use of “less formal likelihoods” which lose the learning properties of the Bayesian inferential approach. Beven et al (2007, 2008) maintained that GLUE is appropriate and coherent according the Bayes theorem in “*special cases where the modeller is prepared to make very strong assumptions about the nature of the modelling errors*”. This debate continued with further challenges by Mantovan et al. (2007), and concluded with Beven et al. (2008) demonstrating the flexibility of the GLUE approach in “*non-ideal cases*”. The more recent debate with Clarke et al. (2012, 2011, Beven et al., 2012) focussed on the superficial rejectionist nature of GLUE from a Bayesian perspective, and concluded with recognition of the need to continue improving the process of model development and evaluation.

It is clear from the extensive literature surrounding the GLUE methodology, and the many applications of GLUE in hydrology models, as well as other earth systems models, that it is a very popular and flexible approach to model uncertainty evaluation. It is also clear however, from the many exchanges between Professor Beven and other hydrologists, that there are two schools of thought regarding the application of formal and informal Bayesian methods, therefore a few of the Bayesian approaches to model uncertainty assessment will be discussed.

3.4.2 Bayesian Recursive Estimation

Bayesian Recursive Estimation (BaRe) (Thiemann et al., 2001) is one of the alternatives to GLUE using a formal Bayesian framework. It makes strong, explicit assumptions about the characteristics of errors in the observations, using an exponential power density error model (Liu and Gupta, 2007). BaRe defines prior probability distributions and parameter ranges, and samples them using Monte Carlo simulation as in GLUE. BaRe employs a recursive scheme for tracking the conditional probabilities associated with different parameter sets (Thiemann et al.,

2001). It predicts the outputs and the uncertainty in the outputs, and updates the probability of the model parameter sets as new data become available at the next time step. BaRe is a method that can estimate uncertainty even if historic observed data are not available for calibration, and is therefore useful for catchments that have only recently been gauged. However, BaRe does not separate out model structural and input data uncertainty, and as parameter estimation is the primary objective, uncertainty estimates are not updated after the posterior parameter distributions are obtained (Liu and Gupta, 2007).

3.4.3 Shuffled Complex Evolution Metropolis Algorithm

The Shuffled Complex Evolution Metropolis algorithm (SCEM-UA) (Vrugt et al., 2003b) is another formal Bayesian approach. It is a modified version of the SCE-UA algorithm developed by Duan et al. (1992), which combines the Metropolis algorithm, controlled random search, competitive evolution, and complex shuffling to update the parameter distribution and develop the posterior distribution. It uses Markov Chain Monte Carlo (MCMC) sampling to locate the high probability density region of the parameter space efficiently.

3.4.4 DREAM

The Differential Evolution Adaptive Metropolis (DREAM) (Vrugt et al., 2008, Vrugt et al., 2009a) algorithm is a development of SCEM-UA, which was specially designed to estimate the posterior density function of hydrologic model parameters in complex, high-dimensional sampling problems (Vrugt et al., 2008). It maintains a detailed balance and ergodicity which enables it to provide an exact Bayesian estimate of uncertainty (Vrugt et al., 2009b).

Vrugt et al. (2009b) compared the formal Bayesian method DREAM with less formal GLUE, for a hydrologic conceptual watershed model, HYMOD. They concluded that formal Bayesian approaches can generate very similar estimates of total predictive uncertainty to informal

Bayesian approaches. DREAM in their application showed a slightly smaller spread of streamflow prediction uncertainty bounds than GLUE, however GLUE reveals when no model can reproduce the observations given the available input data without compensation by a statistical error model or input adjustments. They found that GLUE cannot separate individual error sources and so it is difficult to identify structural deficiencies in the model. The DREAM method attempts to disentangle the different sources of uncertainty but suffers from interaction between individual error sources.

3.4.5 PEST

PEST is a model-independent parameter estimation and uncertainty analysis software package, that allows the user to undertake comprehensive linear and non-linear parameter and predictive uncertainty analysis alongside calibration, based on highly parameterised inversion (PEST, 2014). PEST can also identify the contributions of individual parameters to the uncertainty of prediction, and the worth of existing or new data in reducing predictive uncertainty (PEST, 2014). Ng et al. (2010) compared the GLUE and PEST methods for the hydrological model SWAT. They found that both analyses required some prior knowledge to be effective, which they obtained from deterministic calibration using a genetic algorithm. They found GLUE much more flexible, which makes it suitable for large complex models, but provides a greater level of subjectivity. PEST was found to be computationally frugal, and appropriate where the presence of local optima is not significant.

Ultimately, it seems there are advantages and disadvantages of all available approaches to uncertainty assessment. GLUE is a very flexible and straightforward approach to uncertainty assessment and has been well developed and defended in the 20 years since its inception (Beven and Binley, 2013). Therefore, in this study the underlying GLUE methodology has been applied to assess the

uncertainty in the Mac-PDM.14 model, and special care has been taken with regard to the subjective aspects of the analysis.

3.5 Defining Parameter Distributions

The first step in GLUE analysis is to identify the model parameters and define ranges and plausible distributions for sampling. This step required an extensive literature search for appropriate parameter values to estimate plausible parameter distributions. With the new soil and vegetation maps, there were 123 model parameters to define, including the control parameters. Seeking global values of these parameters was exceedingly difficult, so any estimates, be they local scale, regional, or global were included, collated then analysed to calculate a distribution.

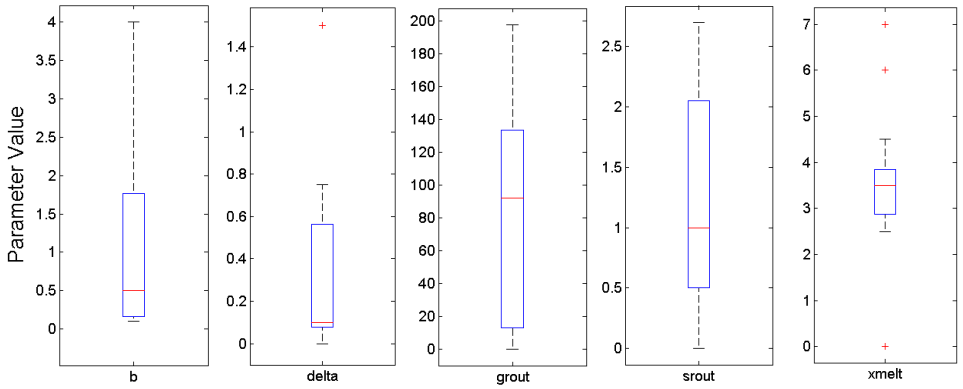
The main sources for parameter values in the literature are detailed in Table 3.4. This table shows that for any parameter that is not a vegetation parameter, it is very difficult to obtain observed data. For the control file parameters, this is mostly due to the fact that many of the parameters do not have a physical meaning, in which case values have been sought from modelling studies that use a similar model structure. For example, values of *grout* and *srout* were obtained from a report by CEH and BGS (2012) that used the GWAVA model. Since GWAVA contains the PDM model in its structure, several of the parameters are comparable with those in Mac-PDM.14. In Table 3.4, the *fcpc* and *satpc* parameters are mostly given the category of 'Generalised' in terms of origin and scale. This is because in many cases, these parameter values were taken from textbooks that contained tables of data on soil hydrology (e.g. Ward et al., 2000, Arnell, 2002, Shaw et al., 2011, Dingman, 2002). These textbooks did not specify the origins of the values, and so they cannot be specified as observed or modelled, nor local or global.

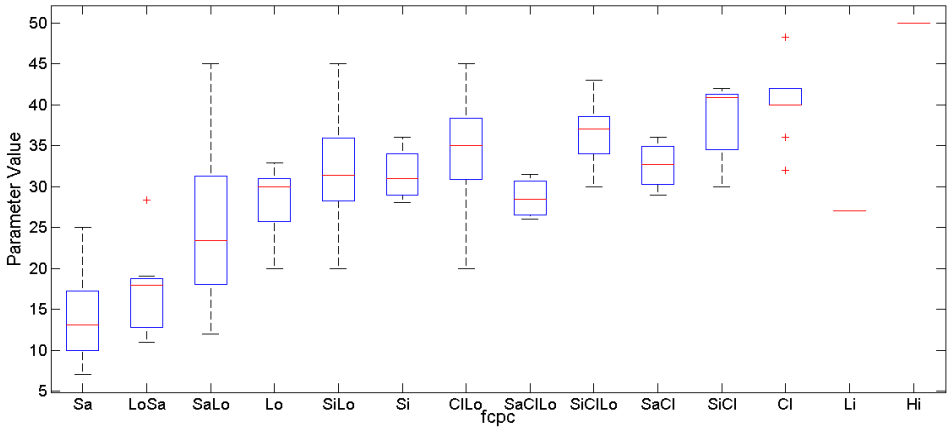
Table 3.4 Major sources of parameter values in the literature

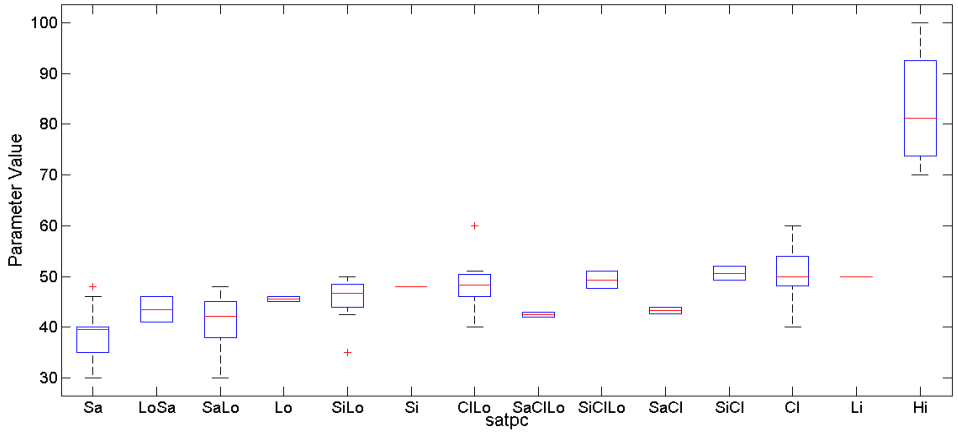
Parameter	Sources	Modelled/Observed	Location
b	Arnell (1999)[1], Moore (2007)[2], CEH and BGS (2012)[3], Wagener and Wheater (2006)[4]	Modelled[1,2,3,4]	Global[1,2], UK[3,4]
δ *(delta)*	Arnell (2002)[1], Calder (1990)[2], Calder (1986)[3], Calder (1998)[4]	Modelled[1,2,3,4]	UK[1,2], Generalised[3] UK & Malawi[4]
grout	CEH and BGS (2012)[1], Wagener and Wheater (2006)[2]	Modelled[1,2]	UK[1,2]
srout	CEH and BGS (2012)[1], Wagener and Wheater (2006)[2]	Modelled[1,2]	UK[1,2]
xmelt	Verdhen and Chahar (2012)[1], Martinec et al. (1998)[2]	Modelled[1,2]	Global[1,2]
fcpc	Arnell (1999)[1], Ward et al. (2000)[2], Cornell University (2010)[3], terraGIS (2007)[4], Arnell (2002)[5], Lambers et al. (2008)[6], Saxton and Rawls (2006)[7], Salter and Williams (1969)[8], Dingman (2002)[9]	Modelled[1,7], Generalised[2,3,4,5,6,8,9]	Global[1], Generalised[2,3,4,5,6,7,8,9]
satpc	Arnell (1999)[1], Ward et al. (2000)[2], Cornell University (2010)[3], Arnell (2002)[4], Saxton and Rawls (2006)[5], Shaw et al. (2011)[6], Clapp and Hornberger (1978)[7]	Generalised[2,3,4.6]	Global[1,2], Generalised[2,3,4,5,6,7]
rootg	Wilson and Henderson-Sellers (1985)[1], Breuer and Frede (2003)[2], Sellers et al. (1996)[3], LDAS (1999)[4,5], Breuer et al. (2003)[6]	Observed[1,2,4,6], Modelled[3,4]	Global[1,3,4,5], Local-Global[2,6]
rsc	Wilson and Henderson-Sellers (1985)[1], Arnell (2002)[2], Henderson-Sellers et al. (1995)[3], LDAS (1999)[4], Breuer et al. (2003)[5]	Observed[1,2], Modelled[3,4,5]	Global[1,2,3,4], Local-Global[5]
capg (γ)	Arnell (2002)[1], Calder (1998)[2], Calder (1986)[3]	Generalised[4], Modelled[2,3]	Generalised[4,3], UK & Malawi[2]
rlai	Wilson and Henderson-Sellers (1985)[1], Breuer and Frede (2003)[2], Scurlock et al. (2001)[3], Sellers et al. (1996)[4], Henderson-Sellers et al. (1995)[5], LDAS (1999)[6], Breuer et al. (2003)[7]	Observed[1,2,3,7], Modelled[3,4,6]	Global[1,3,4,5,6], Local-Global[2,7]
hc	Wilson and Henderson-Sellers (1985)[1], Breuer and Frede (2003)[2], Sellers et al. (1996)[3], LDAS (1999)[4,5], Breuer et al. (2003)[6]	Observed[1,2,4,6], Modelled[3,5]	Global[1,3,4,5], Local-Global[2,6]
percov	Henderson-Sellers et al. (1995)[1], LDAS (1999)[2], Zeng et al. (2000)[3], Hagemann (2002)[4]	Modelled[1,2,3,4]	Global[1,2,3,4]

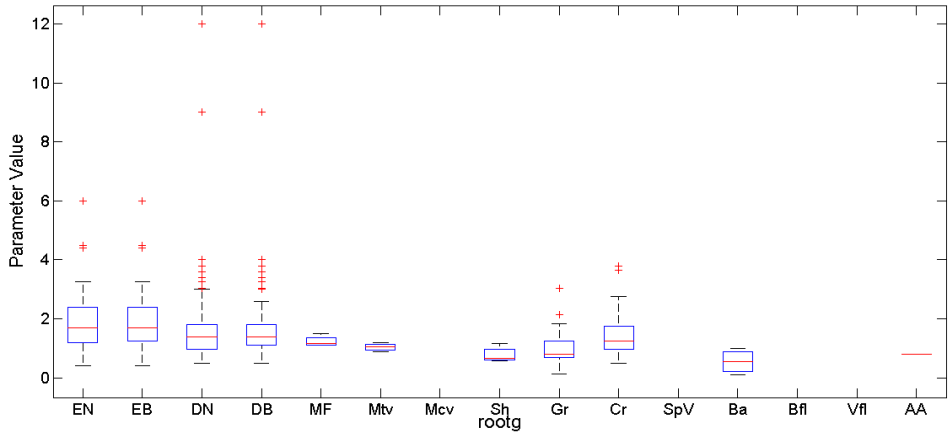
The dominant source of data for the vegetation parameters was the Plant Parameter Database (PlaPaDa) by Breuer and Frede (2003). PlaPaDa is an online database that collates parameters from across the literature for ecological and hydrological models. It contains more than 1300 values for 7 parameters: albedo, interception, leaf area index (LAI), plant height, rooting depth, stomatal conductance and base temperature. This is a very valuable resource and it provided 363 values for the $rlai$ parameter, 228 values for $rootg$, 91 values for the hc parameter, and 318 for rsc, A further paper by Breuer et al. (2003) provided additional data for stomatal resistance and rooting depth. The Land Data Assimilation System (LDAS) (LDAS, 1999) also provide a good database of parameter estimates from both observational data sources and land surface model simulations.

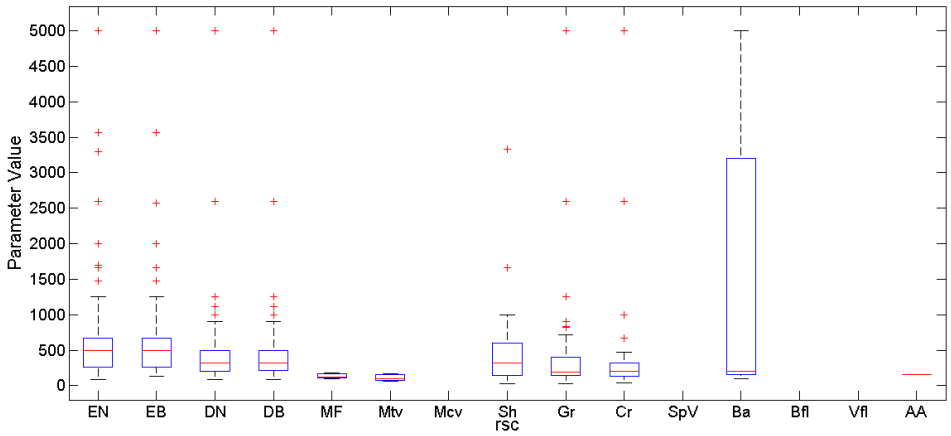
The parameter values were all collated, and were used to define distributions ready for sampling. Sampling was carried out using the software @RISK. Box plots presenting the data found in the literature search are given in Figure 3.4. @RISK was chosen due to its advanced and easy to implement sampling capabilities. It is primarily used in industry for decision making purposes, and is tailored to run models within the programme Excel, however it is possible to input parameter information, form distributions and perform sampling to produce a spreadsheet that can then be used to code an exterior model. @RISK has a library of over 50 distribution functions, including Normal, Uniform, Poisson, Extreme Value, Laplace and Log Logistic. The programme has an integrated BestFit® tool which selects the best distribution function for each parameter. This tool uses Maximum Likelihood Estimators (MLEs) to find the closest matching distribution to the data provided.

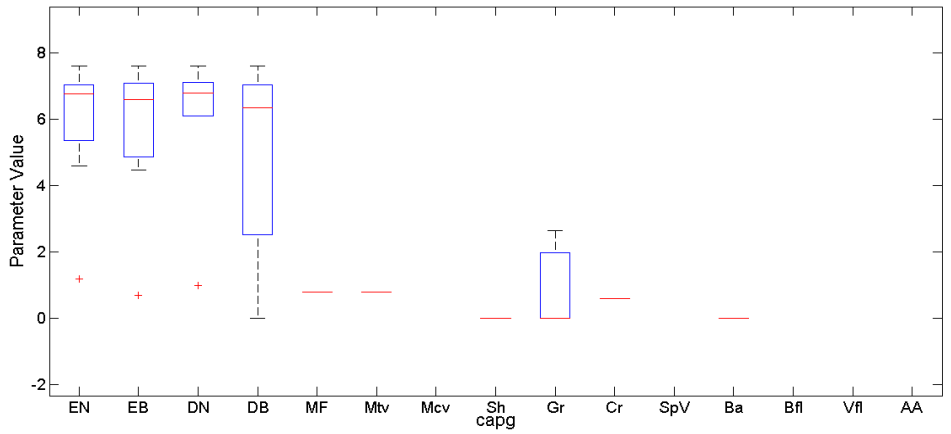

Parameter Value
b
delta
grout
srout
xmelt

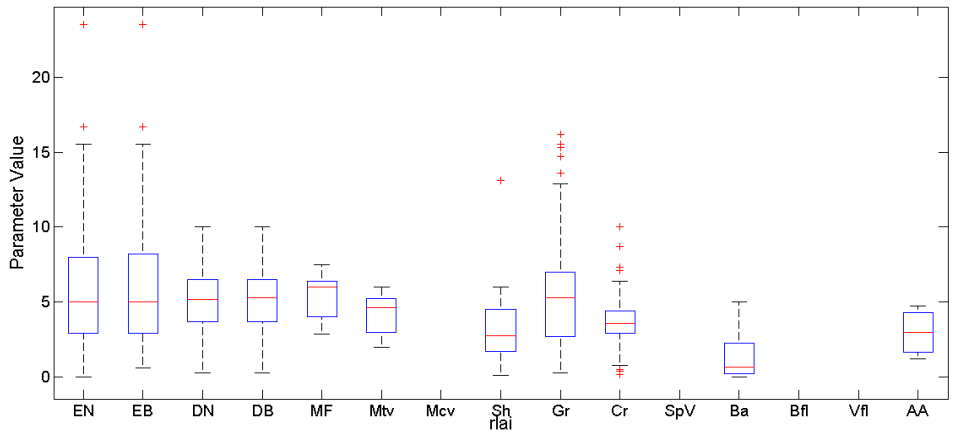

Parameter Value
Sa
LoSa
SaLo
Lo
SiLo
Si
ClLo
SaClLo
SiClLo
SaCl
SiCl
Cl
Li
Hi
fcpc

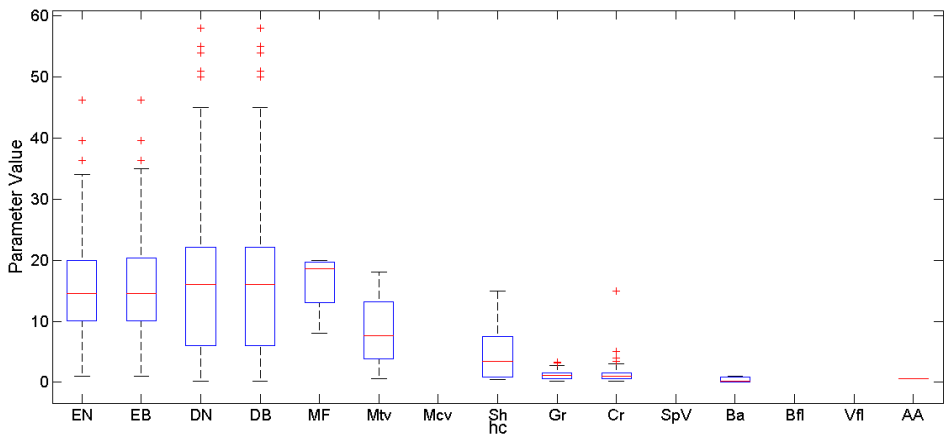

Parameter Value
Sa
LoSa
SaLo
Lo
SiLo
Si
ClLo
SaClLo
SiClLo
SaCl
SiCl
Cl
Li
Hi
satpc


Parameter Value
EN
EB
DN
DB
MF
Mtv
Mcv
Sh
Gr
Cr
SpV
Ba
Bfl
Vfl
AA
rootg


Parameter Value
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
EN EB DN DB MF Mtv Mcv Sh Gr Cr SpV Ba Bfl Vfl AA
rsc


Parameter Value
8
6
4
2
0
-2
EN EB DN DB MF Mtv Mcv Sh Gr Cr SpV Ba Bfl Vfl AA
capg


Parameter Value
20
15
10
5
0
EN EB DN DB MF Mtv Mcv Sh Gr Cr SpV Ba Bfl Vfl AA
rlai


Parameter Value
60
50
40
30
20
10
0
EN EB DN DB MF Mtv Mcv Sh Gr Cr SpV Ba Bfl Vfl AA
hc

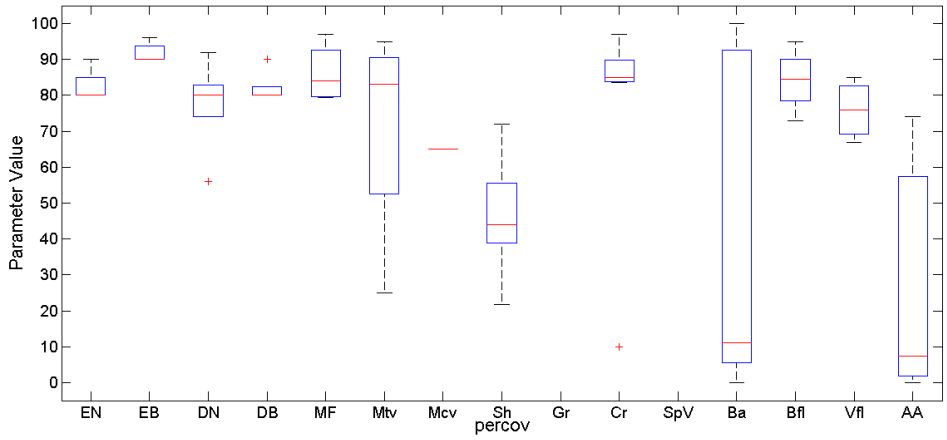


Figure 3.4 Box plots representing parameter values found in the literature. *Red lines indicate the sample mean, boxes represent the interquartile range, whiskers represent the remainder of the sample, except in cases with outliers, which are shown as red +'s. Samples with only one data point present as a red line. Soil classifications read as: Sa – Sand, LoSa – Loamy Sand, SaLo – Sandy Loam, Lo – Loam, SiLo – Silt Loam, Si – Silt, ClLo – Clay Loam, SaClLo – Sandy Clay Loam, SiClLo – Silty Clay Loam, SaCl – Sandy Clay, SiCl – Silty Clay, Cl – Clay, Li – Lithosols and Hi – Histosols. Vegetetation types read as: EN – Evergreen Needleleaf, EB – Evergreen Broadleaf, DN – Deciduous Needleleaf, DB – Deciduous Broadleaf, MF – Mixed Forest, Mtv – Mosaic: Trees/Vegetation, Mcv – Mosaic: Trees/Cropland, Sh – Shrubland, Gr – Grassland, Cr – Cropland, SpV – Sparse Vegetation, Ba – Bare, Bfl – Broadleaf trees regularly or permanently flooded, Vfl – Vegetation regularly flooded and AA – Artificial Areas.*

For any density distribution $f(x)$ with one parameter α, and a corresponding set of n sampled values X_i, an expression called the likelihood can be defined as:

$$L = \prod_{i=1}^{n} f(X_i, \alpha) \qquad \text{(E4.1)}$$

To find the MLE, maximise L with respect to α:

$$\frac{dL}{d\alpha} = 0 \qquad \text{(E4.2)}$$

And solve for α. This can be generalized to distributions with more than one parameter (Palisade Corporation, 2010). @RISK provides three statistical indicators of fitness: Chi-squared, Anderson-Darling (A-D) and Kolmogorov-Smirnov (K-S). The outcomes of the distribution fitting for the Mac-PDM.14 parameters, along with the statistical results of the fitting are shown in Table 3.5. Where less than 5 values were available from the literature, distributions could not be fitted, so uniform or triangular distributions were applied as appropriate. Each distribution

was also inspected for visual fit, and in some cases alternative distributions with close rankings were applied. Following the results of sensitivity analysis, minimum sampling values of 0 were replaced with 0.0001 to avoid extreme model response and infinity outputs. Distributions with long tails were also adjusted to truncate the minimum and maximum values to within a sensible range, slightly beyond the range of literature values. Table 3.5 shows these adjusted minimum and maximum values, along with the mean of the data values taken from the literature. This spreadsheet was then ready for use in sampling parameter values for the GLUE experiment.

3.6 Sampling Methods

GLUE traditionally uses a Monte Carlo technique for sampling the parameter space. Monte Carlo uses random number generation to sample the parameter space (Landau and Binder, 2005), and so requires a large number of samples to adequately fill the sample space, especially in a high dimensional sampling problem.

There is no 'rule of thumb' as to how many samples are required per dimension, but since Mac-PDM.14 has 123 model parameters to sample, it is likely that Monte Carlo sampling would require more samples than would be feasible in order to achieve a good sample. Therefore a more efficient sampling method was sought. Latin Hypercube sampling (McKay et al., 1979) is an alternative sampling technique that has been applied in GLUE experiments of catchment scale hydrological models before, for example the MIKE-SHE model (Christiaens and Feyen, 2002), and the SWAT model (Muleta and Nicklow, 2005). Latin Hypercube sampling (LHS) is inspired by the Latin square experimental design, and is designed to ensure that each value of a variable is represented regardless of its resultant importance (Cheng and Druzdzel, 2000). LHS requires that in a matrix of data, there be only one sample per column and row. Figure 3.5 demonstrates this concept for a two-dimensional 5 by 5 matrix.

5				4	
4					5
3		1			
2	2				
1			3		
Y/X	**1**	**2**	**3**	**4**	**5**

Figure 3.5 Latin hypercube sample of a 5 x 5 matrix (after Cheng and Druzdzel, 2000).

For each sample, [i,j], the sample values of X,Y are determined by:

$$X = F_X^{-1}((i - 1 + \varepsilon x)/n)$$

$$Y = F_Y^{-1}((j - 1 + \varepsilon y)/n),$$

where n is the sample size, εx and εy are random numbers, and Fx and Fy are the cumulative probability distribution functions of X and Y respectively. Figure 3.6 shows a comparison of Monte Carlo and Latin Hypercube sampling for a 2-dimensional grid of 8 samples. This demonstrates the space-filling properties of LHS.

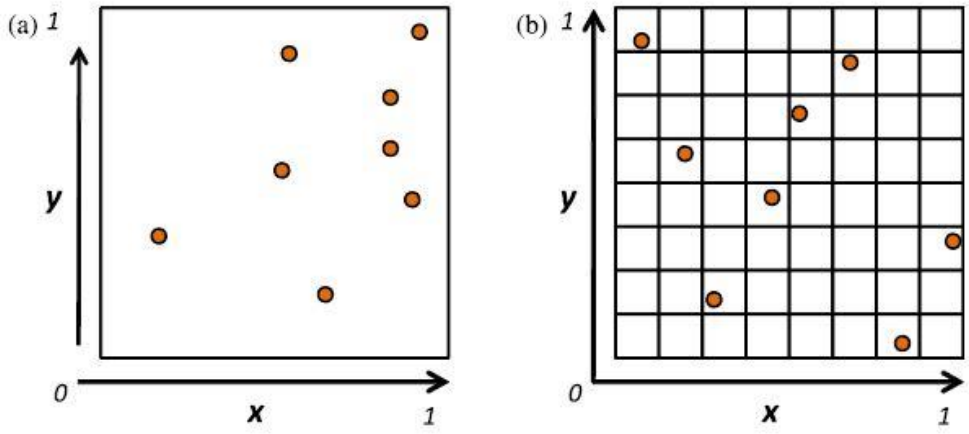


Figure 3.6 Comparison of sampling techniques a) Random Monte Carlo sampling and b) Latin Hypercube sampling. *(N=8 samples), taken from Oehler et al. (2012).*

Table 3.5 Parameters and their distributions as defined by @RISK. *Statistical tests read: value (rank). Vegetation codes given at end*

Parameter	Classification	Count	Distribution	Distribution Parameters	Base Value	Min	Mean	Max	Chi-Squared	A-D	K-S
b	-	11	Pareto	θ=0.60565, a=0.1	0.5	0.0001	1.044	5	0.0909 (#1)	-	0.2066 (#2)
delta	-	9	Exponential	β=0.38089, shift=-0.042321	0.5	0.0001	0.3803	3	0.1111 (#1)	0.3877 (#4)	0.2456 (#5)
grout	-	7	Uniform	-	1.0	0.0001	83.714	210	0.1429 (#1)	0.3282 (#8)	0.2103 (#8)
srout	-	7	Uniform	-	0.5	0.0001	1.22	3.5	0.1429 (#1)	0.2852 (#6)	0.1687 (#2)
xmelt	-	16	Laplace	μ=3.5, σ=1.3568	4.0	0.0001	3.503	10	0.8750 (#1)	0.4936 (#1)	0.1875 (#3)
fcpc	Sand	15	Extreme Value	a=11.658, b=3.7833	13.1	4	13.849	32	0.0000 (#1)	0.1905 (#4)	0.1211 (#8)
	Loamy sand	7	Pareto	θ=2.441, a=11	12.0	8	17.313	35	2.4286 (#13)	-	0.2665 (#10)
	Sandy loam	16	Triangular	min=12, most likely=12, max=50	20.0	12	25.083	50	0.4375 (#7)	-	0.1025 (#7)
	Loam	7	Extreme Value Min	a=29.993, b=2.9217	28.0	15	28.173	36	0.1429 (#1)	0.2349 (#1)	0.2044 (#1)
	Silt loam	16	Laplace	μ=31.4, σ=7.079	29.4	15	31.65	50	0.5 (#2)	0.2494 (#2)	0.1110 (#1)
	Silt	4	Uniform	-	30.0	18.48	31.5	47.88	-	-	-
	Clay loam	13	Laplace	μ=35, σ=6.6656	33.1	15	33.851	54	0.1538 (#2)	0.2444 (#4)	0.1154 (#3)
	Sandy clay loam	4	Uniform	-	27.0	17.16	28.587	41.895	-	-	-
	Silty clay loam	7	Normal	μ=36.55, σ=4.1139	38.0	23	36.55	49	0.1429 (#6)	0.1682 (#2)	0.1578 (#4)
	Sandy clay	4	Uniform	-	36.0	19.14	32.604	47.88	-	-	-
	Silty clay	5	Extreme Value Min	a=39.969 b=3.0288	41.0	23	37.968	48	0.2 (#2)	0.5137 (#3)	0.3362 (#5)
	Clay	10	Laplace	μ=40, σ=3.6071	48.3	28	40.149	52	0.4 (#2)	0.5771 (#1)	0.2988 (#6)
	Lithosols	1	Uniform	-	27.0	13.5	27	40.5	-	-	-
	Organic	1	Uniform	-	50.0	25	50	75	-	-	-
satpc	Sand	10	Uniform	-	35.5	25	38.55	52	0.4 (#1)	0.4093 (#7)	0.2545 (#6)
	Loamy Sand	2	Uniform	-	46.0	27.06	43.5	61.18	-	-	-
	Sandy Loam	10	Extreme Value Min		41.3	20	40.89	52	0.0000 (#1)	0.1417 (#1)	0.1108 (#1)

Parameter	Classification	Count	Distribution	Distribution Parameters	Base Value	Min	Mean	Max	Chi-Squared	A-D	K-S
satpc	Loam	2	Uniform	-	46.0	29.964	45.55	61.18	-	-	-
	Silt loam	10	Extreme Value Min	a=47.283, b=2.7047	46.8	30	45.54	53	0.4 (#2)	0.2241 (#1)	0.1285 (#1)
	Silt	1	Uniform	-	48.0	24	48	72	-	-	-
	Clay loam	10	Laplace	μ=48.25, σ=4.6952	50.4	33	48.66	65	0.0000 (#1)	0.2071 (#1)	0.1184 (#1)
	Sandy clay loam	2	Uniform	-	43.0	27.72	42.5	57.19	-	-	-
	Silty clay loam	2	Uniform	-	51.0	31.482	49.35	67.83	-	-	-
	Sandy clay	2	Uniform	-	44.0	28.116	43.3	58.52	-	-	-
	Silty clay	2	Uniform	-	52.0	32.472	50.6	69.16	-	-	-
	Clay	10	Laplace	μ=50, σ=4.8083	54.4	34	50.68	66	0.4 (#2)	0.4342 (#3)	0.2 (#4)
	Lithosols	1	Uniform	-	50.0	25	50	75	-	-	-
	Organic	4	Uniform	-	100	46.2	83.125	133	-	-	-
Rootg	EN	41	Log Logistic	γ=0.15036, β=1.5283, α=2.881	0.9	0.1504	2.0102	8	1.7317 (#1)	0.1620 (#1)	0.0699 (#1)
	EB	41	Log Logistic	γ=0.12059, β=1.5843, α=3.0334	1.5	0.15	2.0261	8	7.1951 (#7)	0.2038 (#1)	0.0794 (#1)
	DN	68	Exponential	β=1.329, shift=0.48046	0.9	0.481	1.8290	15	20.2353 (#4)	1.6211 (#4)	20.2353 (#5)
	DB	68	Log Logistic	γ=0.3249, β=1.0606, α=2.2227	1.2	0.325	1.8474	15	11.4118 (#1)	0.8605 (#1)	0.1153 (#1)
	MF	4	Triangular	min=0.726, most likely =1.2, max=15	1.1	0.726	1.2313	15	-	-	-
	Mtv	4	Triangular	min=0.57552, most likely=1.05, max=10	1.1	0.5755	1.043	10	-	-	-
	Mcv	0	Triangular	min=0.0001, most likely=1.2, max=5.054	0.9	0.0001	-	5.054	-	-	-
	Sh	8	Uniform	-	0.6	0.3	0.7811	5	0.5 (#7)	1.107 (#14)	0.3536 (#11)
	Gr	45	Extreme Value	a=0.71489, b=0.4025	0.6	0.0001	0.9498	4	20.6 (#6)	0.9293 (#2)	0.1360 (#2)
	Cr	90	Log Normal	μ=1.2388, σ=0.62937, shift=0.17317	1.2	0.1733	1.4117	5	3.6222 (#2)	0.1477 (#2)	0.0493 (#2)
	SpV	0	Uniform	-	0.2	0.0001	-	4	-	-	-

Parameter	Classification	Count	Distribution	Distribution Parameters	Base Value	Min	Mean	Max	Chi-Squared	A-D	K-S
rootg	Ba	3	Uniform	-	0.1	0.0001	0.55	2	-	-	-
	Bfl	0	Triangular	min=0.5, most likely=1.3, max=18	1.3	0.5	-	18	-	-	-
	Vfl	0	Triangular	min=0.0001, most likely=1, max=6	0.6	0.0001	-	6	-	-	-
	AA	1	Uniform	-	0.1	0.3985	0.79	1.1955	-	-	-
rsc	EN	64	Log Logistic	γ=67.822, β=388.06, α=1.7299	85	67.822	733.32	5500	6.3125 (#1)	0.3618 (#1)	0.0863 (#1)
	EB	61	Log Logistic	γ=110.6, β=326.32, α=1.5805	130	110.6	685.57	5500	9.9672 (#1)	0.6748 (#1)	0.0879 (#2)
	DN	94	Log Logistic	γ=65.713, β=241.68, α=1.8372	85	67.822	733.32	5500	12.2553 (#3)	0.4509 (#1)	0.0596 (#2)
	DB	94	Log Logistic	γ=68.956, β=239.47, α=1.8252	100	68.86	448.62	5500	10.6170 (#2)	0.5268 (#1)	0.0585 (#2)
	MF	3	Triangular	min=66, most likely=500, max=5500	100	66	133.33	5500	-	-	-
	Mtv	3	Triangular	min=42.9, most likely=400, max=5500	100	42.9	111.38	5500	-	-	-
	Mcv	0	Triangular	min=14, most likely=400, max=5500	100	14	-	5500	-	-	-
	Sh	37	Log Normal	μ=557.3, σ=867.55, shift=5.5415	80	15	556.67	4000	2.1622 (#3)	0.2064 (#1)	0.0686 (#2)
	Gr	88	Log Logistic	γ=18.847, β=199.05, α=1.8502	70	18.847	363.99	5500	12.0 (#2)	0.3989 (#1)	0.0777 (#1)
	Cr	69	Log Logistic	γ=29.93, β=165.02, α=1.8747	100	29.93	343.87	5500	3.4706 (#2)	0.2481 (#1)	0.0603 (#1)
	Spv	0	Uniform	-	100	15	-	5500	-	-	-
	Ba	5	Uniform	-	100	50	1615	5500	-	-	-
	Bfl	0	Triangular	min=60, mostly likely=500, max=5500	110	60	-	5500	-	-	-
	Vfl	0	Triangular	min=14, most likely=400, max=5500	90	14	-	5500	-	-	-
	AA	1	Triangular	min=50, most likely=100, max=4000	100	50	154.84	4000	-	-	-
capg	EN	8	Triangular	min= 0.5, most likely=6.5, max=10	1.2	0.5	5.8875	10	2.0 (#5)	-	0.3841 (#7)
	EB	7	Extreme Value Min	a=6.5668, b=1.3231	0.7	0.3	5.65	9.3	0.4129 (#2)	0.4938 (#2)	0.2190 (#1)

Parameter	Classification	Count	Distribution	Distribution Parameters	Base Value	Min	Mean	Max	Chi-Squared	A-D	K-S
capg	DN	6	Extreme Value Min	a=6.7691, b=1.1747	1	0.5	5.8983	9.2	0.6667 (#2)	0.7090 (#2)	0.2654 (#1)
	DB	7	Extreme Value Min	a=6.563, b=1.3349	0.6	0.15	5.6357	9.2	1.2857 (#2)	0.5012 (#2)	0.2211 (#1)
	MF	1	Uniform		0.8	0.2	0.8	10	-	-	-
	Mtv	1	Uniform		0.8	0.2	0.8	10	-	-	-
	Mcv	0	Uniform		0.6	0.2	-	8	-	-	-
	Sh	1	Uniform		1	0	0	12	-	-	-
	Gr	3	Uniform		0.1	0	0.8867	5	-	-	-
	Cr	1	Uniform		0.6	0.2	0.6	8	-	-	-
	SpV	0	Uniform		0.2	0	-	6	-	-	-
	Ba	1	Uniform		0	0	-	5	-	-	-
	Bfl	0	Uniform		0.6	0.2	-	10	-	-	-
	Vfl	0	Uniform		0.8	0.2	-	10	-	-	-
	AA	0	Uniform		0	0	-	5	-	-	-
rlai	EN	127	Inverse Gaussian	μ=8.1918, λ=29.208, shift=-2.0956	6	0.0001	6.0963	26	7.4016 (#2)	0.2024 (#2)	0.0359 (#1)
	EB	127	Weibull	α=1.3342, β=5.9855, shift=0.56309	9	0.563	6.0701	25	6.3780 (#5)	0.2049 (#1)	0.0344 (#2)
	DN	105	Normal	μ=5.1547, σ=2.1258	4	0.0001	5.1547	11	3.9143 (#3)	0.1900 (#1)	0.0409 (#1)
	DB	105	Normal	μ=5.1684, σ=2.1251	5	0.0001	5.1684	11	4.6 (#2)	0.2460 (#1)	0.0479 (#1)
	MF	5	Normal	μ=5.358, σ=1.7621	6	0.0001	5.358	11	0.2 (#1)	0.2324 (#4)	0.2390 (#1)
	Mtv	5	Extreme Value Min	a=4.8541, b=1.1556	6	0.05	4.1837	7.5	0.2 (#1)	0.1931 (#1)	0.1695 (#1)
	Mcv	0	Triangular	min=0.0001, most likely=4, max=18	4	0.0001	-	18	-	-	-
	Sh	25	Extreme Value	a=2.1049, b=1.7291	2	0.0001	3.1555	14	0.4 (#1)	0.3932 (#2)	0.1035 (#1)

Parameter	Classification	Count	Distribution	Distribution Parameters	Base Value	Min	Mean	Max	Chi-Squared	A-D	K-S
rlai	Gr	93	Weibull	α=1.4166, β=6.0462, shift=0.09393	3	0.0905	5.6216	19	16.8817 (#2)	1.0724 (#10)	0.1390 (#8)
	Cr	55	Laplace	μ=3.6, σ=1.9187	5	0.0001	3.7886	12	5.7091 (#1)	0.5087 (#1)	0.0814 (#1)
	SpV	0	Triangular	min=0.0001, most likely=1, max=10	1	0.0001	-	10	-	-	-
	Ba	12	Exponential	β=1.3414, shift=-0.1178	0	0	1.3414	7	0.3333 (#3)	0.3152 (#1)	0.1393 (#1)
	Bfl	0	Triangular	min=0.0001, most likely=5.5, max=25	7	0.0001	-	25	-	-	-
	Vfl	0	Triangular	min=0.0001, most likely=4, max=19	4	0.0001	-	19	-	-	-
	AA	3	Uniform	-	0	0	2.9761	6	-	-	-
hc	EN	131	Weibull	α=1.8033, β=17.673, shift=-0.28264	19.1	0.00016	15.476	50	14.3817 (#9)	1.0069 (#10)	0.0848 (#9)
	EB	131	Weibull	α=1.7189, β=17.956, shift=-0.89986	29.4	0.000114	15.967	50	12.9924 (#8)	0.8855 (#9)	0.0811 (#7)
	DN	99	Gamma	α=1.3122, β=12.222, shift=0.14938	10	0.149	16.188	75	13.1111 (#2)	1.0651 (#7)	0.1157 (#1)
	DB	99	Gamma	α=1.316, β=12.332, shift=0.14491	14.9	0.145	16.188	75	17.5556 (#3)	1.3214 (#8)	0.1308 (#2)
	MF	4	Triangular	min=0.0001, most likely=16, max=75	18	0.0001	16.313	75	-	-	-
	Mtv	4	Triangular	min=0.0001, most likely=8.5, max=75	18	0.0001	8.4887	75	-	-	-
	Mcv	0	Triangular	min=0.0001, most likely=6, max=25	7.0	0.0001	-	25	-	-	-
	Sh	14	Exponential	β=4.4013, shift=0.18562	1.4	0.1856	4.9014	25	2.2857 (#11)	0.4947 (#1)	0.1378 (#1)
	Gr	85	Weibull	α=1.6219, β=1.2975, shift=0.094239	0.6	0.094239	1.2587	16	34.7059 (#3)	1.1211 (#3)	0.1269 (#1)
	Cr	46	Log Logistic	γ=0.02, β=0.80913, α=1.6573	14.9	0.02	1.6783	16	11.0435 (#3)	0.3459 (#3)	0.1167)#2)
	SpV	0	Triangular	min=0.0001, most likely=0.2, max=5	0.2	0.0001	-	5	-	-	-
	Ba	3	Uniform	-	0	0	0.4	1.2	-	-	-

Parameter	Classification	Count	Distribution	Distribution Parameters	Base Value	Min	Mean	Max	Chi-Squared	A-D	K-S
	Bfl	0	Triangular	min=0.0001, most likely=16.1, max=75	22.0	0.0001	-	75	-	-	-
	Vfl	0	Triangular	min=0.0001, most likely value=3, max=25	5.0	0.0001	-	25	-	-	-
	AA	1	Uniform	-	0	0	0.6018	1.2	-	-	-
percov	EN	4	Uniform	-	80	75	82.5	95	-	-	-
	EB	5	Uniform	-	90	86	91.8	98	-	-	-
	DN	5	Uniform	-	80	50	77.6	95	-	-	-
	EB	5	Uniform	-	80	75	82	95	-	-	-
	MF	4	Uniform	-	80	75	86.125	98	-	-	-
	Mtv	4	Uniform	-	25	20	71.546	95	-	-	-
	Mcv	1	Uniform	-	15	10	65	90	-	-	-
	Sh	10	Uniform	-	25	16	45.598	77	0.4 (#8)	0.3346 (#6)	0.1711 (#7)
	Gr	7	-	-	0	-	-	-	-	-	-
	Cr	7	Uniform	-	10	5	76.791	85	3.5714 (#3)	2.712 (#13)	0.6161 (#10)
	SpV	0	Uniform	-	90	50	-	95	-	-	-
	Ba	5	Uniform	-	90	75	41.698	95	-	-	-
	Bfl	4	Uniform	-	85	70	84.25	98	-	-	-
	Vfl	3	Uniform	-	15	62	76	90	-	-	-
	AA	3	Uniform	-	95	75	27.145	99	-	-	-

Vegetation parameter codes read as: EN=Evergreen Needleleaf; EB=Evergreen Broadleaf; DN=Deciduous Needleleaf; DB=Deciduous Broadleaf; MF=Mixed Forest; Mtv=Mosaic:trees/vegetation; Mcv=Mosaic:cropland/vegetation; Sh=Shrubland; Gr=Grassland; Cr=Cropland; SpV=Sparse Vegetation; Ba=Bare; Bfl=Broadleaf Trees Regularly or Permanently Flooded; Vfl=Vegetation Regularly Flooded; AA=Artificial Areas.

In this study, @RISK was used to generate the Latin Hypercube sample of the 123 parameters, using the distributions identified in Table 3.5, with an ensemble size of 100,000 model runs. With the assigned parameter distributions the grid from which to take the samples is no longer evenly divided, as in Figure 3.6b, but is instead divided according to the area underneath the distribution curve. Some illustrations of this are given in Figure 3.7.

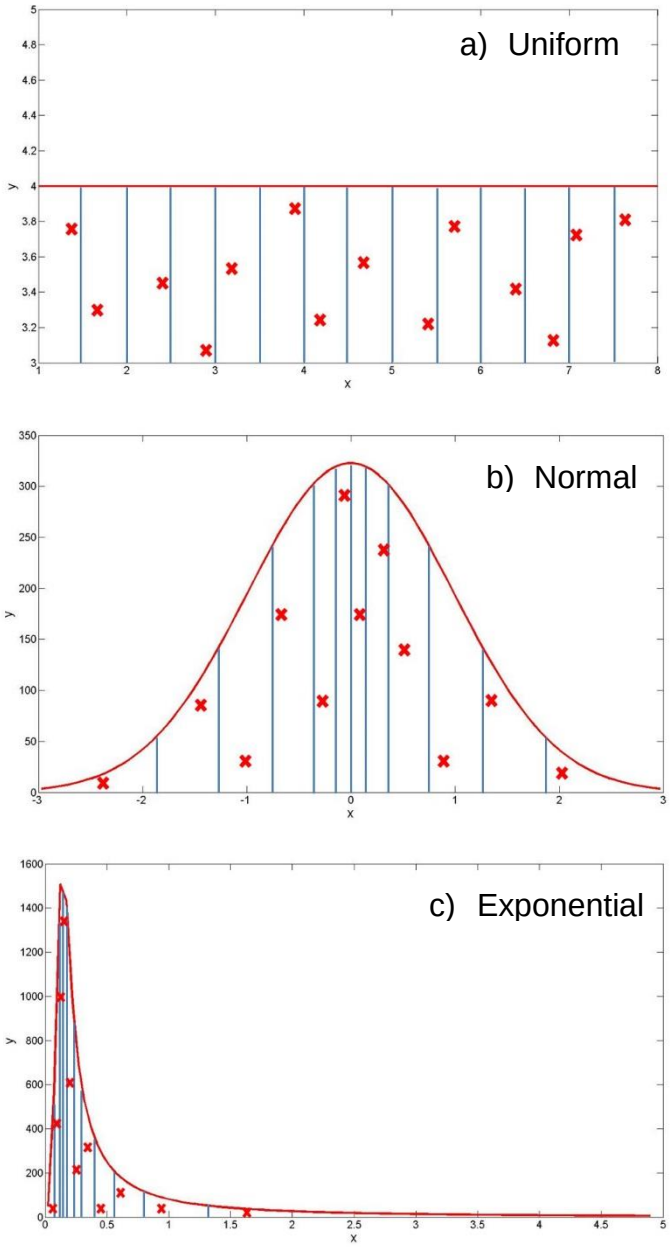


Figure 3.7 Illustrations of LHS samples under different distribution types.

This figure shows that where probability densities are high, the sampling space is more concentrated, allowing the sample to focus on the more likely region of the parameter space, whilst still sampling the full range appropriately.

Once the parameter values had been sampled by LHS, the model control, soil and vegetation files for each of the 100,000 parameter realisations were created using the sampled values, and the Mac-PDM.14 model was run on the Nottingham High Performance Computer Cluster. The 100,000 model runs took approximately 40 days to run, and output just over 2 terabytes of data. The post-processing of the model outputs to extract catchment averaged data took a further 10 days, and produced 479MB of data. The GLUE experiment outputs were assessed using a likelihood function. This process, along with the results of the experiment are discussed in Chapter 4.

3.7 Summary

This chapter has reviewed the definition of, and approaches for, sensitivity analysis and uncertainty assessment of numerical models. A one-at-a-time sensitivity analysis was performed on the Mac-PDM.09 model. This revealed that the soil parameters, field capacity and saturation capacity are the most sensitive parameters in the model when perturbed individually. The root depth parameter also shows significant sensitivity over grass. The results reinforced the need to update the soil and vegetation maps, which was described previously in Chapter 2.

Popular methods of uncertainty analysis were reviewed and critiqued. The GLUE technique was chosen for the assessment of the Mac-PDM.14 model. An extensive literature review was carried out in order to define the ranges and distributions of the parameter values in this experiment. The decision making software @RISK was employed to fit distributions to the parameter values and to sample the parameter

space using a Latin Hypercube Sampling (LHS) technique. LHS was employed due to its superior efficiency over the traditional Monte Carlo sampling technique. The results of this experiment will be presented in Chapter 4, after techniques of evaluating model performance are discussed.

Chapter Four:

Parameter Uncertainty in Global Hydrology Modelling

Part 2

- *Calibration and Results*

4.1 Introduction

Models vary in type, complexity and scale. However, regardless of their structure, the ultimate aim of models is to represent a physical system. In order to determine if they are adequate representations, models must be assessed for their skill in reproducing observed hydrological behaviour. Krause et al. (2005) give three reasons why a hydrologist needs to evaluate their model's performance: 1) to provide a quantitative measure of the models capability of reproducing historic and future catchment behaviour; 2) to provide a way of evaluating improvements to the model through adjustments to the parameters and structure, inclusion of additional data, and representation of important spatial and temporal characteristics of the catchment; and 3) to compare current modelling studies with previous efforts.

While chapter 4 outlined the methods and experimental design of a Generalised Likelihood Uncertainty Estimation (GLUE) experiment, this chapter will present the results. However it is first necessary to review available objective functions which are used in hydrology for the purpose of comparing model realisations with observed discharge data. This chapter then details the method chosen to compare the Mac-PDM.14 model realisations with the discharge data from study catchments.

4.2 Objective Functions and Likelihood Measures

Objective functions, likelihood measures, evaluation metrics, error measures, evaluation criteria, and 'goodness of fit' measures are all synonymous terms used to describe a numerical equation that can be applied to assess the skill of a model using observational data. There have been several reviews that have detailed multiple evaluation metrics, with the most comprehensive by Dawson et al. (2007) who developed a web-based toolbox that can be used to calculate multiple assessment criteria simultaneously. Other notable contributions include

Krause et al. (2005), Hauduc et al. (2011), Gupta et al. (1998) and Reusser et al. (2009). Table 4.1 gives a list of commonly used evaluation metrics in hydrological studies, and provides some example references of their application in this field. Each method has its own strengths and weaknesses, and each has conditions upon which it may be more or less suitable. A few of the most popular methods, and their characteristics, will be discussed here. Dawson et al. (2007) define three types of metric, as follows:

1) statistical parameters of observed and modelled time series datasets;
2) statistical parameters of the residual error between observed and modelled time series datasets; and
3) dimensionless coefficients that contrast model performance with accepted norms or recognised standards.

Within this framework, the residual error measures are the most diverse, but dimensionless coefficients are perhaps the most popular. The first of Dawson's categories includes basic measures such as mean, standard deviation, minimum, maximum and skewness. The second category, residual error measures, can be subdivided into absolute and relative measures. Absolute error measures define the error in the same units as the variables of interest (Hauduc et al., 2011), whilst relative errors measures express the error in terms of ratios and percentages. Absolute measures include absolute mean error (AME, eq. 1), peak difference (PDIFF, eq. 2), mean absolute and mean error (MAE, eq. 3 and ME, eq. 4), root mean square error and fourth root mean quadrupled error (RMSE, eq. 6 and R4MS4E, eq. 8), Akaike and Bayesian information criterion (AIC, eq. 9 and BIC, eq. 10), and number of sign changes (NSC, eq. 11). Relative measures include relative absolute error (RAE, eq. 12), percent error in peak (PEP, eq. 14), mean absolute relative error (MARE, eq. 15), median absolute percentage error (MAPE, eq. 16), mean relative error (MRE, eq. 17), mean squared relative error (MSRE, eq. 18) and relative volume error (RVE, eq. 19).

4.2.1 Absolute Residual Error Measures

Mean error (ME) can identify a systematic bias in a model, where the model systematically overestimates, or underestimates the observed. However, the errors in ME may counteract each other. Mean absolute error (MAE) avoids this, and defines the average magnitude (but not sign) of the error; underestimation or overestimation is not specified. Absolute maximum error (AME) gives the value of maximum error in the time series or dataset, which could be useful if the model is required to maintain a threshold of goodness, but it is sensitive to outliers in the residuals (Hauduc et al., 2011). Peak difference (PDIFF) is a metric that examines the agreement in the magnitude of the highest peak in the dataset. The peaks need not necessarily be in the same temporal location in the time series, but the metric is useful in determining the model's capability of producing similar ranges of forecast values to that of the observational data (Dawson et al., 2007).

Root mean square error (RMSE) is a very popular evaluation metric (Nayak et al., 2004, McLeod et al., 1987, Coulibaly and Baldwin, 2005, Dawson et al., 2007); it squares the residuals in order to avoid error compensation, and the root returns the metric to actual units. This metric emphasises larger errors, and therefore tends to focus on high flow events in the time series. The fourth root mean quadrupled error puts even more emphasis on large errors (Hauduc et al., 2011). RMSE and MAE are fairly comparable metrics, and Willmott and Matsuura (2005) assessed their abilities to describe average model performance error. Willmott and Matsuura (2005) determined that MAE is favourable over RMSE as it is unambiguous, and the most natural measure of average error magnitude. RMSE is based on the sum of squared errors, and so does not describe average error adequately.

Table 4.1 List of evaluation metrics used in hydrological modelling studies

No.	Abbr.	Name	Equation	Best	Worst	Example References
1	AME	Absolute Maximum Error	$\mathrm{AME} = \max(\lvert Q_i - \hat{Q}_i \rvert)$	0	∞	Gupta et al. (1998), Hauduc et al. (2011)
2	PDIFF	Peak Difference	$\mathrm{PDIFF} = \max(Q_i) - max(\hat{Q}_i)$	0	+/- ∞	Modarres and Ouarda (2013), Guse et al. (2014)
3	MAE	Mean Absolute Error	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} \lvert Q_i - \hat{Q}_i \rvert$	0	∞	Chang et al. (2004), Willmott and Matsuura (2005), Karunasinghe and Liong (2006),
4	ME	Mean Error	$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n} (Q_i - \hat{Q}_i)$	0	+/- ∞	Chang et al. (2001)
5	NMBE	Normalised Mean Bias Error	$\mathrm{NMBE} = \frac{\frac{1}{n}\sum_{i=1}^{n}(Q_i - \hat{Q}_i)}{\bar{Q}}$			Jain and Srinivasulu (2006)
6	RMSE	Root Mean Square Error	$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{n}}$	0	∞	Bowden et al. (2002), Alvisi et al. (2006), , Shukla and Lettenmaier (2011)
7	$\mathrm{RMSE_r}$ NRMSE	Relative Root Mean Square Error/Normalised Root Mean Square Error	$\mathrm{RMSE_r} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{n}}}{\bar{Q}}$			Fernando and Jayawardena (1998), Pebesma et al. (2005), Jain and Srinivasulu (2006)

No.	Abbr.	Name	Equation	Best	Worst	Example References
8	R4MS4E	Fourth Root Mean Quadrupled Error	$\mathrm{R4MS4E} = \sqrt[4]{\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^4}{n}}$	0	∞	Hauduc et al. (2011), Dehghani et al. (2014)
9	AIC	Akaike Information Criterion	$\mathrm{AIC} = -2\ln(ML) + 2p$ $\mathrm{AIC} = m\ln(ML) + 2p$	0	∞	Hipel (1981) (1), Hsu et al. (1995) (2), Mutua (1994), Engelhardt et al. (2012)
10	BIC	Bayesian Information Criterion	$\mathrm{BIC} = m\ln(ML) + p\ln(m)$	0	∞	Hsu et al. (1995), Marshall et al. (2005), Laio et al. (2009)
11	NSC	Number of Sign Changes	$\mathrm{SC} = Number\ of\ sign\ changes\ (of\ residuals)$	0	∞	Gupta et al. (1998), Dawson et al. (2007)
12	RAE	Relative Absolute Error	$\mathrm{RAE} = \frac{\sum_{i=1}^{n}\lvert Q_i - \hat{Q}_i\rvert}{\sum_{i=1}^{n}\lvert Q_i - \bar{Q}\rvert}$	0	∞	Hauduc et al. (2011), Modarres and Ouarda (2013)
13	ARE	Absolute Relative Error	$\mathrm{ARE} = \left\lvert\frac{(Q_i - \hat{Q}_i)}{Q_i}\right\rvert * 100$	0	∞	Hu et al. (2001), Dawson et al. (2007)
14	PEP	Percent Error in Peak/Error of Peak Discharge	$\mathrm{PEP} = \frac{max(Q_i) - max(\hat{Q}_i)}{max(Q_i)} * 100$	0	+/- ∞	Lin and Chen (2004), Dawson et al. (2007), Huang et al. (2008)
15	MARE	Mean Absolure Relative Eror	$\mathrm{MARE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\lvert Q_i - \hat{Q}_i\rvert}{Q_i}$	0	∞	Riad et al. (2004), (Jain and Srinivasulu (2006)), Wei et al. (2012)

No.	Abbr.	Name	Equation	Best	Worst	Example References
16	MdAPE	Median Absolute Percentage Error	$MdAPE = Median\left(\left\|\frac{Q_i - \hat{Q}_i}{Q_i}\right\|\right) * 100$	0	∞	Lin and Chen (2005), Dawson et al. (2007), Villarini et al. (2008)
17	MRE	Mean Relative Error	$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_i - \hat{Q}_i}{Q_i}\right)$	0	+/- ∞	Creutin et al. (1997), Dawson et al. (2006), Modarres and Ouarda (2013)
18	MSRE	Mean Squared Relative Error	$\mathrm{MSRE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_i - \hat{Q}_i}{Q_i}\right)^2$	0	∞	Dawson et al. (2007), Gebhart et al. (2013)
19	RVE	Relative Volume Error	$RVE = \frac{\sum_{i=1}^{n}(Q_i - Q_i)}{\sum_{i=1}^{n}Q_i}$	0	∞	Bergström et al. (2002), Lin and Chen (2005), Parajka et al. (2006), Akhtar et al. (2009)
20	Rsqr	Coefficient of Determination/Pearson Product Moment Correlation Coefficient	$RSqr = \left[\frac{\sum_{i=1}^{n}(Q_i - \bar{Q})(\hat{Q} - \tilde{Q})}{\sqrt{\sum_{i=1}^{n}(Q_i - \bar{Q})^2 \sum_{i=1}^{n}(\hat{Q} - \tilde{Q})^2}}\right]^2$	1	0	Vázquez et al. (2002), Giustolisi and Laucelli (2005), Krause et al. (2005), Dawson et al. (2007)
21	CE/NSE	Coefficient of Efficiency/ Nash Sutcliffe Efficiency	$\mathrm{CE} = 1 - \frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n}(Q_i - \bar{Q})^2}$	1	-∞	Vázquez et al. (2002), Krause et al. (2005), McCuen et al. (2006), Schaefli and Gupta (2007)

No.	Abbr.	Name	Equation	Best	Worst	Example References
22	IoAD	Index of Agreement	$IoAD = 1 - \frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n}(\lvert\hat{Q}_i - \bar{Q}\rvert - \lvert Q_i - \bar{Q}\rvert)^2}$	1	0	Willmott (1981), Krause et al. (2005), Dawson et al. (2007), Modarres and Ouarda (2013)
23	PI	Persistence Index	$PI = 1 - \frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n}(Q_i - Q_{i-1})^2}$	1	- ∞	Kitanidis and Bras (1980), Dawson et al. (2007), Schaefli and Gupta (2007)

where Q_i is the observed (expected) value, $\hat{Q}i$ is the modelled (forecast) values (where i = 1 to n *data points), $\bar{Q}i$ is the mean of the observed data, and $\tilde{Q}i$ is the mean of the modelled values. The best column indicates the most desirable outcome (the best score available), and the worst column the least desirable outcome. For the AIC and BIC metrics,* ML *is the maximum likelihood, p is the number of free parameters in the model, and* m *is the number of data points use in the calibration.*

The final metrics among the absolute residual error group are the Akaike and the Bayesian information criterion (AIC and BIC). These are quite unique metrics that are not often used in hydrological model evaluation. They use a more traditional evaluation metric within them, which is adjusted to the number of parameters in the model, and the number of data points used in the calibration. They both attempt to account for model complexity, and seek the minimal model that best explains the dataset. They quantify the relative performance of a model, assuming that a model with many parameters will closely fit the data, but not have many degrees of freedom, and will therefore have limited application. AIC and BIC give credit to simple models, and discourage over-fitting (Dawson et al., 2007).

4.2.2 Relative Residual Error Measures

Many of the relative residual error metrics are very similar, but each has its own characteristics. Relative absolute error (RAE) compares the total absolute error to the error that would result from a forecast of the mean of the observed values. It gives a ratio of the overall level of agreement between the modelled and observed data, and is influenced by the spread of the observed records. Mean relative error (MRE) is another measure that records the overall level of agreement between the modelled and observed datasets, however it does not make use of the mean of the observed record. In the same way as mean error (ME), MRE is a signed metric where over and underestimations of the observed data may cancel each other out. Mean squared relative error (MSRE) is essentially the same metric, but the square of the relative residuals makes it more sensitive to larger errors at lower magnitudes. Due to the potential cancelling out of errors in both of these metrics, the mean absolute relative error (MARE), and median absolute percentage error (MdAPE) metrics are more popular (Dawson et al., 2007). MARE is again an overall agreement metric, but it uses the absolute value of the residual, and then expresses it relative to the observed value. Since

it is not squared, it is less sensitive to the high errors that can occur at high magnitudes. MARE is often expressed as a percentage, and referred to as MAPE. MdAPE uses median, rather than mean, and is therefore less affected by outliers and skewed error distributions. Percent error in peak (PEP) and relative volume error (RVE) are more specific metrics that are commonly used for single event modelling.

4.2.3 Dimensionless Evaluation Metrics

The third of Dawson's categories, dimensionless evaluation metrics, include the very popular Nash Sutcliffe Efficiency metric (also known as Coefficient of Efficiency, CE eq. 21), the Coefficient of Determination (Rsqr, eq. 20), the Index of Determination (IoAd, eq. 22) and Persistence Index (PI, eq. 23). The Coefficient of Determination is the square of the Pearson's product-moment correlation coefficient (R^2). It describes the squared ratio between the covariance and the multiplied standard deviations of the observed and predicted variables (Krause et al., 2005). This efficiency metric only compares the dispersion of the predicted values with the dispersion of the observed values, and does not take the magnitude of the data into account. Thus, a model may significantly underestimate, or overestimate each of the observed records, but still result in a good r^2 value if the dispersion is of a similar magnitude.

The Nash Sutcliffe Efficiency metric (Nash and Sutcliffe, 1970) has been widely used in hydrology, and several papers have reviewed its capabilities of capturing goodness of fit (e.g. McCuen et al., 2006, Jain and Sudheer, 2008, Schaefli and Gupta, 2007, Criss and Winston, 2008). It is defined as one minus the sum of the absolute squared differences between the predicted and observed variables, normalised by the variance of the observed values during the specified time period (Krause et al., 2005). Since the Nash Sutcliffe efficiency (NSE) metric uses squares, larger values in the time series are over-emphasised, and errors in lower values are neglected. Also, as Schaefli and Gupta

(2007) explain, the NSE compares the performance of the model with that of the simplest imaginable model (one that's prediction is the mean of the observed variables, as does RAE). This means that, depending on the nature of the river's flow regime, the meaning of the value of NSE can differ widely; since for strongly seasonal time series', the NSE may be misleadingly high, whereas for catchments with a more constant mean, the model would need to explain the small fluctuations accurately to gain a high NSE value. Thus, Schaefli and Gupta (2007) suggest that a benchmark model is required in order to compare model performance across varying hydrologic regimes.

The Index of Agreement (IoAd) was proposed by Willmott (1981) and is one minus the ratio of the sum of squared error to potential error (potential error being the sum of the largest quantification that can be obtained for each individual forecast with respect to the mean of the observed dataset) (Dawson et al., 2007). IoAd is an improvement over the Coefficient of Determination (R^2), as it is sensitive to differences in the predicted to observed variances (Dawson et al., 2007). However, since the metric uses squares, again it is also sensitive to peak values over low values. Again, IoAd can give relatively high values for poor models, and the best models' IoAd scores are not significantly higher. The Persistence Index is one minus the ratio of the sum of the squared error to what the sum of squared error would have been if the forecast were the last observed value. This metric suffers similar interpretation issues as NSE, and should be compared to the performance of a benchmark model (Dawson et al., 2007).

4.2.4 Evaluation Metrics for Mac-PDM.14

Legates and McCabe (1999) established that correlation-based measures, such as the Coefficient of Determination, and the Nash Sutcliffe Efficiency metrics are not appropriate for the evaluation of model performance, due to the ability of poor models to have high correlation values, as well as the inherent difficulty of interpreting such

metrics. Therefore, absolute and relative residual error metrics were considered for this study. Squared error metrics were disregarded due to their bias towards peak flow simulation, which would focus model performance on the catchments with the highest flows, whilst disregarding the performance of the model in drier catchments. Mean Absolute Error is widely praised in the hydrological literature (e.g. Willmott and Matsuura, 2005, Legates and McCabe, 1999), however, due to the broad range in runoff values across the chosen catchments for this study, Mean Absolute Relative Error was used to evaluate Mac-PDM.14 to allow the errors from each catchment to be fairly included in the overall performance score.

4.3 Results

The results of the 100,000 model simulations were assessed against the observed records obtained from the GRDC, BWDB and USGS (see Chapter 2.5) using the Mean Absolute Relative Error (MARE) evaluation metric for each catchment as follows:

$$MARE_{MacPDM} = \frac{1}{13}\left(\left|\frac{\hat{Q}_{tot}-Q_{tot}}{Q_{tot}}\right| + \left(\sum_{i=1}^{12}\left|\frac{\hat{Q}m_i - Qm_i}{Qm_i}\right|\right)\right) \quad \text{(E 4.1)}$$

Where $\hat{Q}tot$ is the modelled average annual runoff for each catchment, $Qtot$ is the observed annual average runoff for each catchment, $\hat{Q}m_i$ is the modelled average monthly rainfall and Qm_i is the observed average monthly runoff for each catchment (i = Jan-Dec). The MARE statistics were then averaged across catchments to give a 'global' average for the 21 catchments.

4.3.1 Evaluation Metric Scores

The MARE values ranged from 0.9 to 7.9, which seemed surprisingly high (an error of 90% for the best model). Upon investigation of the catchment MARE statistics, it was found that Mac-PDM.14 model performance was especially poor in the Murray Darling and the Nile

catchments, with the lowest MARE values for these individual catchments being 2.5 and 3.7 respectively (compared to an average value of 0.2, std=0.16, for the remaining 19 catchments). The maximum MARE values for the Murray Darling and Nile catchments were 99.7 and 22.7 respectively (compared to an average value of 3.4, std=2.9, for the remaining catchments). These results suggest that the model is not good at simulating these catchments.

Figure 4.1 shows the top ranking MARE model (i.e. lowest MARE), when the MARE was averaged across all 21 catchments, and the MARE when averaged across 19 catchments excluding the Murray Darling and the Nile, as compared to the observed record. The observed records for Murray Darling and Nile rivers show very low flow year round, which is likely the result of extensive anthropogenic influences on the flow regime in the form of large dams and reservoirs, and abstractions for irrigation. The discharge station for the River Nile that was used for the observed record is positioned at the outflow of the Aswan Dam, which impounds Lake Nasser, with a total holding capacity of 5.97 billion cubic meters of water. At Aswan, the Nile has the lowest specific discharge of any river with a catchment greater than 1 million km^2, at 0.98 litres s^{-1} km^{-2} (Woodward et al., 2007). Of the estimated mean flow of 84 km^3, 18.5km^3 is allocated for abstraction by Sudan in the Nile Waters Treaty of 1959, 55.5 km^3 allocated to Egypt, and the remainder is subject to extensive losses through seepage and evaporation (Woodward et al., 2007, Chauhan et al., 2014b, Sene et al., 2001).

The Murray Darling is the fourth longest river system in the world, after the Amazon, Mississippi-Missouri and the Nile (Thoms et al., 2007). It is also one of the world's driest catchments, and recently experienced the "millennium drought" that lasted from 1995 to late 2009. Less than 10mm of rain has been recorded in a 12 month consecutive period in the Darling catchment five times (Thoms et al., 2007). In addition, only

about 5% of rainfall reaches the river system, and 44% of runoff is dedicated to irrigation (Ryan, 2009). Furthermore, nearly all of the river system is significantly degraded from its original state, with 10% of the rivers total length being classified as substantially modified, and 84.5% being moderately modified (Thoms et al., 2007). Due to the lack of routing in the model, the significantly high level of transmission losses that would be experienced in such a large dry catchment would not be accurately represented. This, coupled with the substantial human influence in this catchment, are the main factors in the poor performance of Mac-PDM.14 in the Murray Darling catchment.

Each of the study catchments is subjected to some degree of anthropogenic disturbance, and the implications of Mac-PDM simulating 'naturalised flows' is considered in more detail in the following section. However, due to the severity of the disturbances in these two catchments, and the resultant unfluctuating low flow of the observed records, the Murray Darling and the Nile catchments were excluded from further analyses of the performance of the model. Without the Murray Darling and the Nile, the values of MARE ranged from 0.47 to 2.58 across the 100,000 model realisations.

Figure 4.1 shows that for most of the river catchments, the removal of the Murray Darling and the Nile from the MARE score leads to a better performing top ranking model when compared with the observed records. For the Amazon and the Amu Darya however, the 21 catchment average top ranking model gives a better fit with the observed record. This is because the MARE was averaged across all catchments, so the significant improvement across the majority of catchments outweighs the worsening in others. From Figure 4.1, it can be seen that the model performs well in catchments with a strong seasonal flow regime, such as the Brahmaputra, Ganges, Mekong and Mississippi.

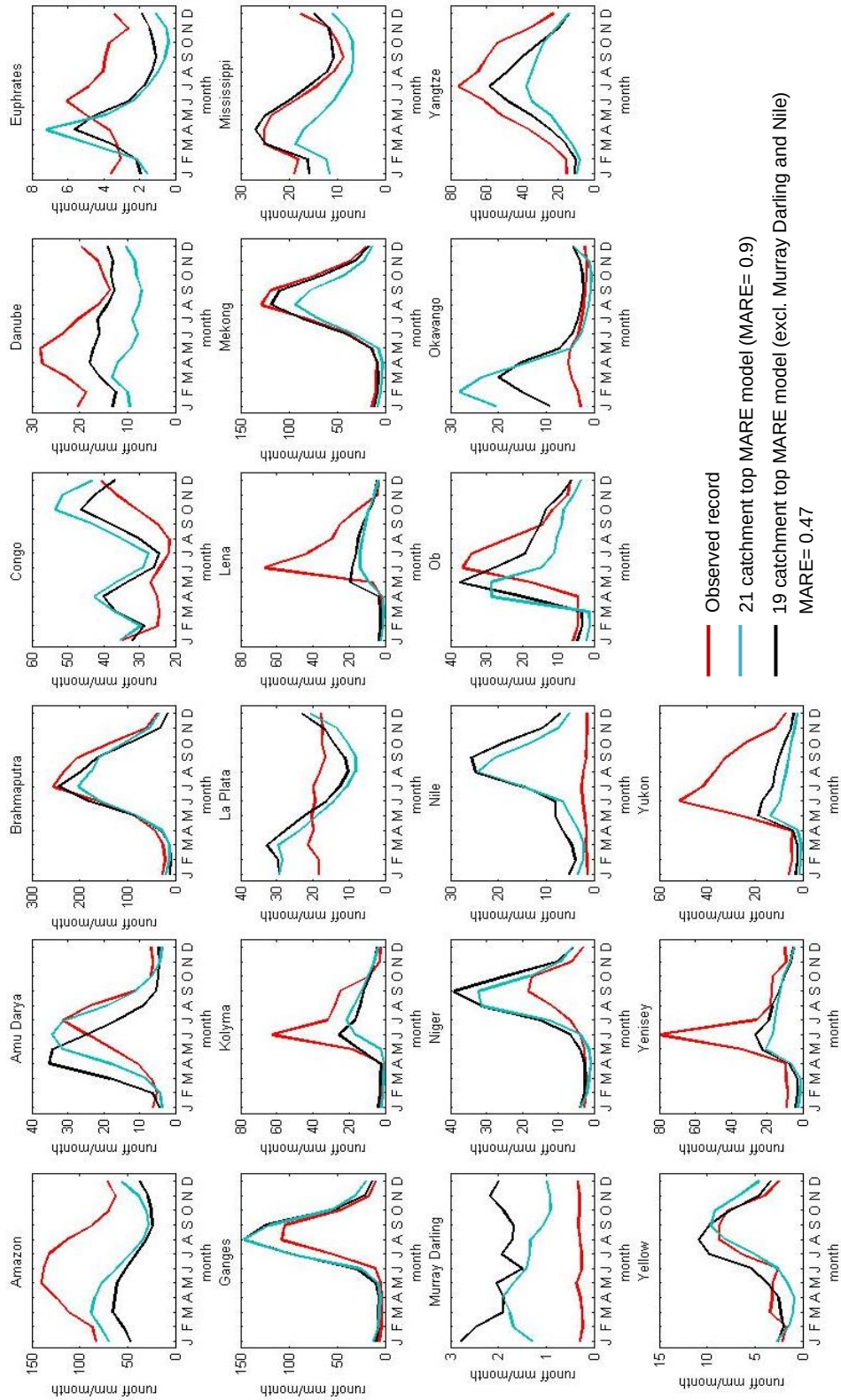


Figure 4.1 Top MARE rated models from the full 100,000 realisation ensemble, for 21 catchments with and 19 catchments without the Murray Darling and Nile, compared with observed records

The high latitude catchments of the Lena, Yenisey, and Yukon are not so well simulated, as the model underestimates the peak flows significantly; this is likely due to the lack of a glacier component in Mac-PDM, as well as the fact that the model does not consider the seasonal freezing and melting of permafrost.

A comparison of the top ranking model for the 19 catchments with the previous version of Mac-PDM (Mac-PDM.09), before the soil and vegetation classifications were updated, is shown in Figure 4.2. Mac-PDM.09 scored a MARE of 1.05 excluding the Murray Darling and the Nile catchments. Of the 100,000 model realisations 34,406 of the Mac-PDM.14 models scored a MARE lower than that of Mac-PDM.09, meaning that the updating of the maps and calibration of the model can easily improve the model performance. However, by studying the graphs in Figure 4.2, it is apparent that the top performing parameterisation of Mac-PDM.14 provides a betterment over Mac-PDM.09 in most, but not all of the catchments. For example in the Yenisey and Yukon, the underestimation of the peak flow is exaggerated. The top ranking MARE Mac-PDM.14 model performs significantly better than MacPDM.09 in the Euphrates catchment, the Okavango, and the Congo. Improvement is also evident in the Yellow, Mekong, Niger, Ganges, and in the Amu Darya, where the peak timing is still early, but the magnitude is a better fit with the observed.

Considering just the ‘top ranking’ model however does not provide an appropriate evaluation of the model performance; the main reason being the issue of equifinality, which is described in Chapter 3.4. It is also worth noting here that this model calibration only considers 19 of the world’s catchments, and that this model parameterisation may not be optimum for other catchments that have not been evaluated in this study.

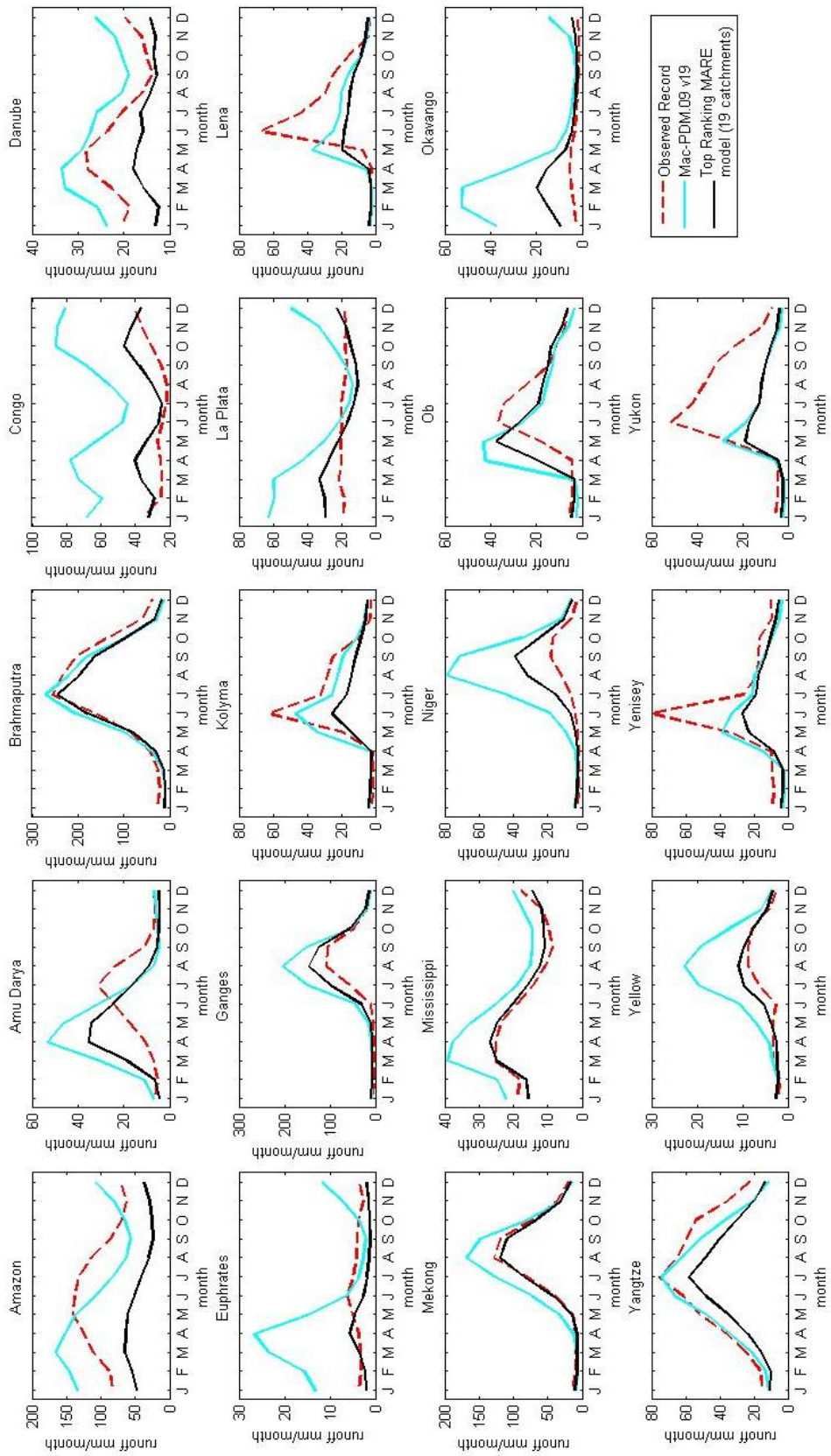


Figure 4.2 Top ranking model realisation for MARE averaged over 19 catchments compared to Mac-PDM.09

Therefore, groupings of model parameterisations may allow for a more realistic assessment of model output, taking equifinality into account, or the models may also be weighted according to their goodness-of-fit.

4.3.2 Using Evaluation Metrics and Likelihood Measures

The MARE evaluation metric (or objective function) was used to create a likelihood measure, which would enable the models to be weighted according to their goodness-of-fit and used to determine an ensemble weighted average. All models with a MARE value of less than 1 were considered to be 'behavioural'. This meant that all models that had an average error across the 19 catchments of less than 100% were included in the weighting. This left 25,532 model realisations, which was 26% of the ensemble. The likelihood measure was calculated by taking the reciprocal of the MARE value, and then dividing by the sum of the reciprocals for the 25,532 models, which made the likelihood measure values sum to 1. The simulated model values for January to December were then multiplied by the likelihood measure, and summed to give an ensemble weighted average.

This ensemble average is shown in the cyan dashed line in the graphs in Figure 4.3, compared with the observed records shown in red. These graphs also show the ranges of outputs from all of the model realisations that scored a MARE value of less than 1, 0.75 and 0.5. As previously mentioned, 25,532 models scored less than 1, 1,238 scored less than 0.75 and only 2 models scored less than 0.5. As with the top ranking models, these graphs confirm that the Mac-PDM.14 model performs best for highly seasonal catchments, with good fits in the Brahmaputra, the Ganges, Mekong, Mississippi and Yangtze. The high latitude catchments show a significant underestimation of the peak flow, even with a fairly relaxed model acceptance of a MARE less than or equal to 1. For the majority of catchments, this acceptance limit encompasses the observed record.

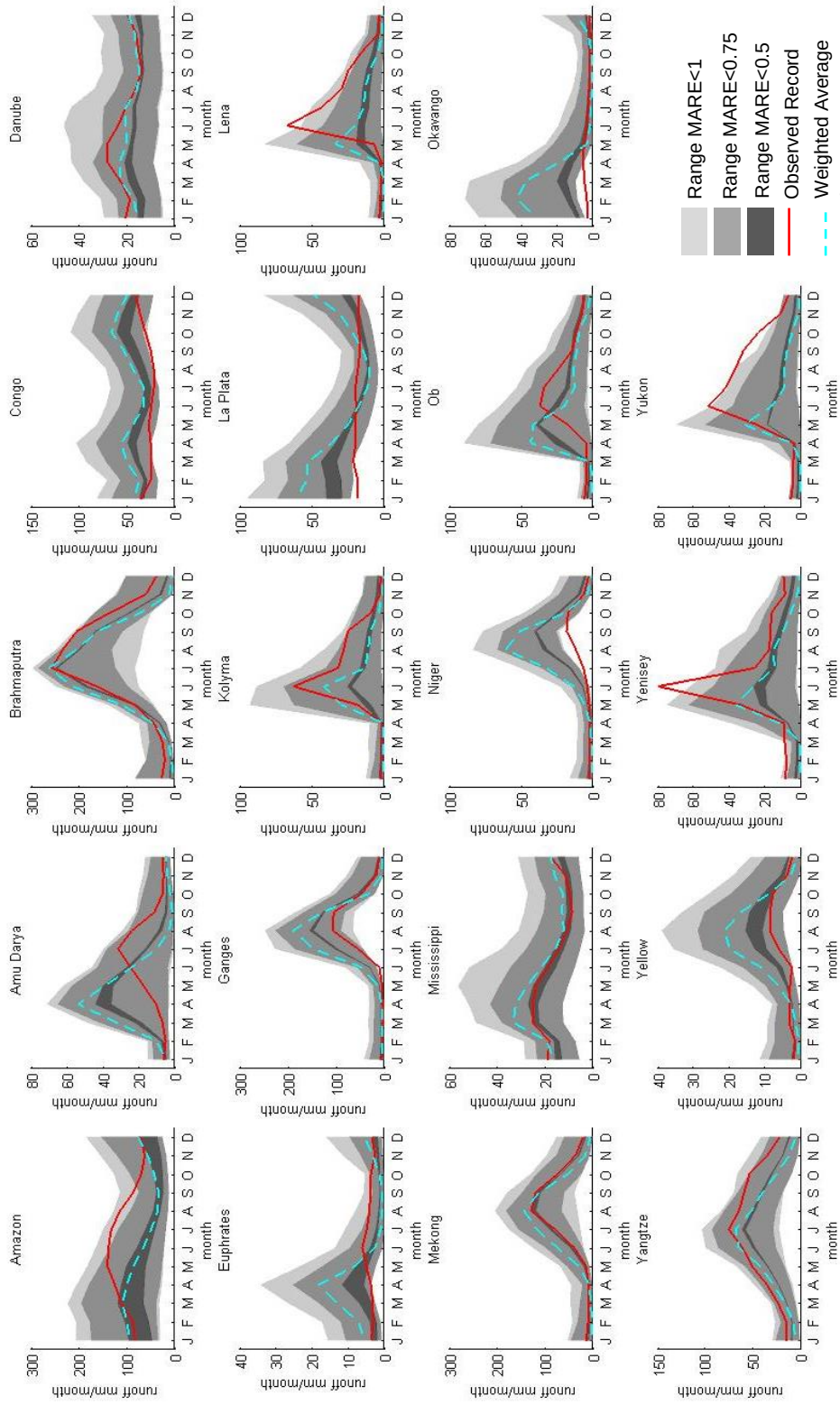


Figure 4.3 Mac-PDM.14 model ensemble results for 19 study catchments compared to observed records for the years 1971-2000.

The June peak flows of the Lena and Yenisey catchments are underestimated, as is the magnitude of the receding annual limb of the Yukon catchment in September-October. The low flows of the Okavango January-March, the La Plata in January and the Niger in June-September are overestimated by the model. This again may be due to abstractions, which is discussed further in the following section. The peak flows of the Amazon River and the Amu Darya are simulated several months too early, the Ob also shows an early peak by two or three months, and the high latitude catchments, the Lena, Yenisey, and Yukon, peak one month too soon. The premature simulation of peak flows is likely due to the fact that the model does not route the runoff, which in very large catchments can cause a significant delay in runoff production from precipitation. Delayed peaks are likely due to the lack of a snowmelt module in the model, so water that should be held in frozen stores is counted as runoff for earlier months. It is noticeable that the weighted average line deviates from the range of MARE<0.5, and provides a higher estimate of runoff than the two models that scored <0.5 for all of the catchments.

The use of likelihood measures is a very subjective approach and is one of the criticised aspects of the GLUE methodology (e.g. Mantovan and Todini, 2006). The influence of deciding which models are classified as “behavioural”, and which are rejected upon the model output is significant and the impacts of this on Mac-PDM.14 is demonstrated in Figure 4.4, which shows a weighted average when the limit of behavioural models was set to MARE values of between 0.5 and 1 at increments of 0.1. These graphs show that a steady change in output can be seen as the number of models included as behavioural is reduced, however this change does not always trend towards the observed record. The Amazon, Danube, Lena, Yangtze, Yellow, Yenisey and Yukon all give worse results with fewer models included.

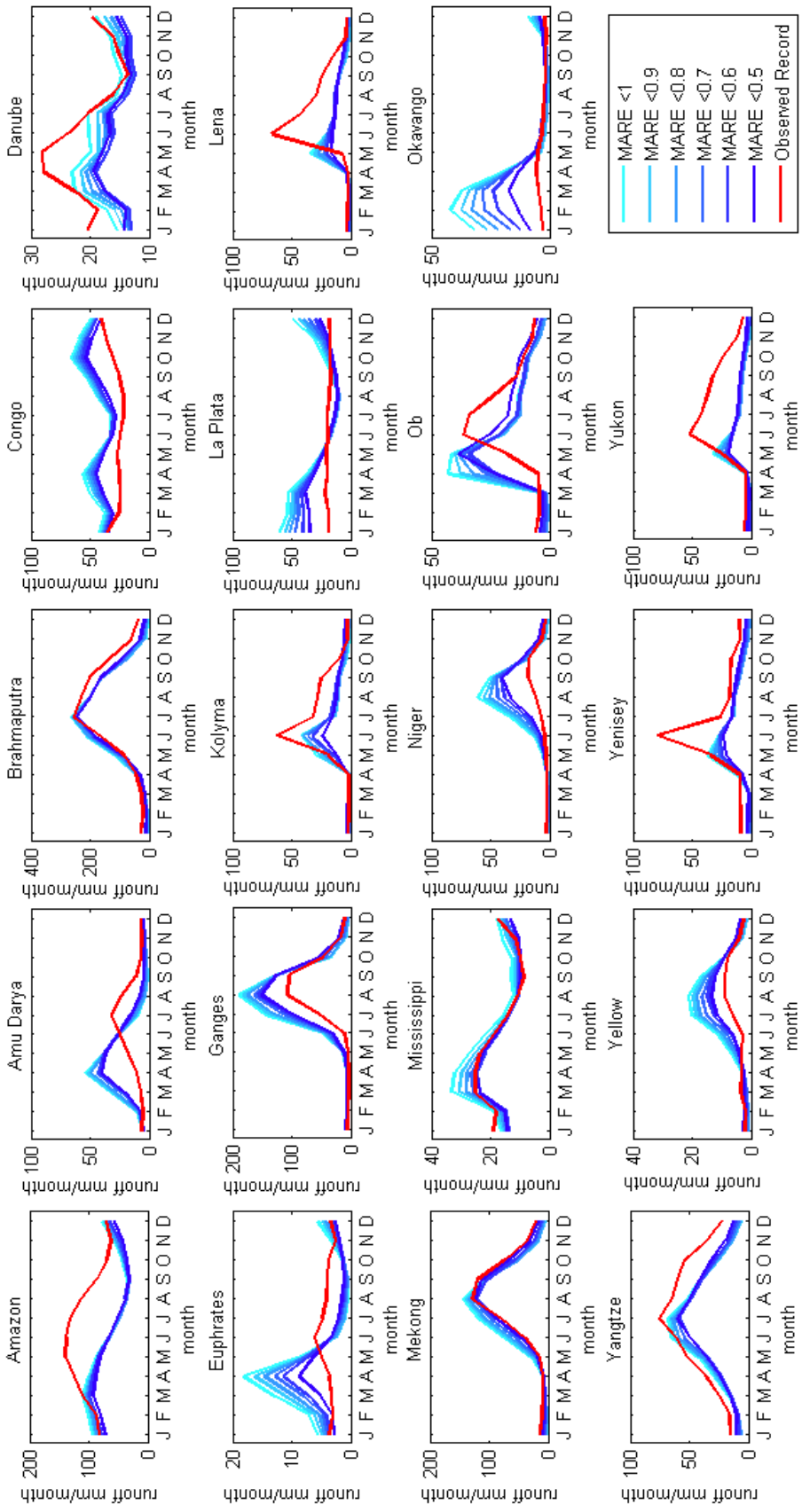


Figure 4.4 Impact of behavioural limit adjustment on likelihood measure based weighted average model output

In each catchment, the reduction of behavioural models leads to a reduction in simulated average runoff, which suggests that this analysis is favouring the accurate modelling of those catchments that are overestimated by Mac-PDM.09 over those catchments that are being underestimated.

Since the model is ranked using MARE, averaged across all catchments, it is possible that in order to maximise the performance of Mac-PDM.14 in some difficult catchments, where the model does not perform well, the performance of Mac-PDM.14 in other catchments is being reduced. The possibility of using Mac-PDM.14 calibrated for individual catchments is the focus of the next chapter, Chapter 5. For this reason, it seems the presentation of results through ranges, or fans, such as in Figure 4.3, provides a more informative representation of model outputs than using a weighted average by means of a likelihood measure. This is because a full range of potential model outputs is presented. For example, the Yukon catchment shown with weighted averages in Figure 4.4 indicates that after May, Mac-PDM.14 is incapable of simulating the high flows of June-October, however from Figure 4.3 it can be seen that some of the models with a MARE of <1 come quite close to the observed record.

The graphs in Figure 4.3 might lead one to believe that the model is more uncertain with higher flows, as the ranges of the model outputs is widest during periods of peak flow. As MARE is a relative metric, this was investigated further by averaging the Absolute Relative Error per month for each catchment. The results of this investigation are shown in Figure 4.5, which reveal that the volume of runoff does not determine the amount of error in the model. For example, in the Brahmaputra catchment, which has its widest range of model outputs in July, the months with the largest Relative Error are actually December and January. The width of the model output ranges is reasonably wide in

these months in Figure 4.3, and the low runoff value means that this error is more significant per mm than the error in July.

The graphs in Figure 4.5 give a good indicator as to how the model performs through both high and low flows. The Kolyma and Mississippi catchments are modelled with a consistent level of accuracy year round (with MARE of ~0.5 and 0.2 respectively). The Brahmaputra and Yangtze are best simulated during high flows, whilst the Amu Darya, Congo, Euphrates, Ganges, Niger, Okavango and Yellow Rivers are best simulated during low flows. The Lena and the Ob catchments are simulated consistently, except for the peaks in May and April respectively, which are dramatically underestimated.

It is important to note here that MARE does not account for the sign of change, and that understanding of over and under-estimation must be interpreted from the visual inspection of additional graphs (in this case those in Figure 4.3). To this end, the values of the MARE score on the y-axis of the graphs in Figure 4.5 help explain why the weighted averages reduce the runoff values as the MARE behavioural limit is made stricter. With the exception of the peak in the Lena catchment MARE in May, the largest of the MARE scores are all associated with overestimations of the observed record. The highest is the Okavango in February, which has an average MARE score of 11.74 for all models with overall MARE <1. The Ob in April is an overestimation, with average MARE of 8.8 for overall MARE <1, and similarly, the high errors of the Amu Darya, Euphrates, Ganges, Niger and Yellow are all associated with overestimations of the observed record. Catchments that are underestimated, such as the Amazon, Danube, Kolyma, Yangtze, Yenisey and Yukon all show small MARE values that do not exceed 1 for any individual month. Overall, by comparing the months of high and low flows with the months of greatest error, it seems that Mac-PDM.14 is a fairly balanced model, and shows no significant trend towards modelling either high or low flows with more accuracy.

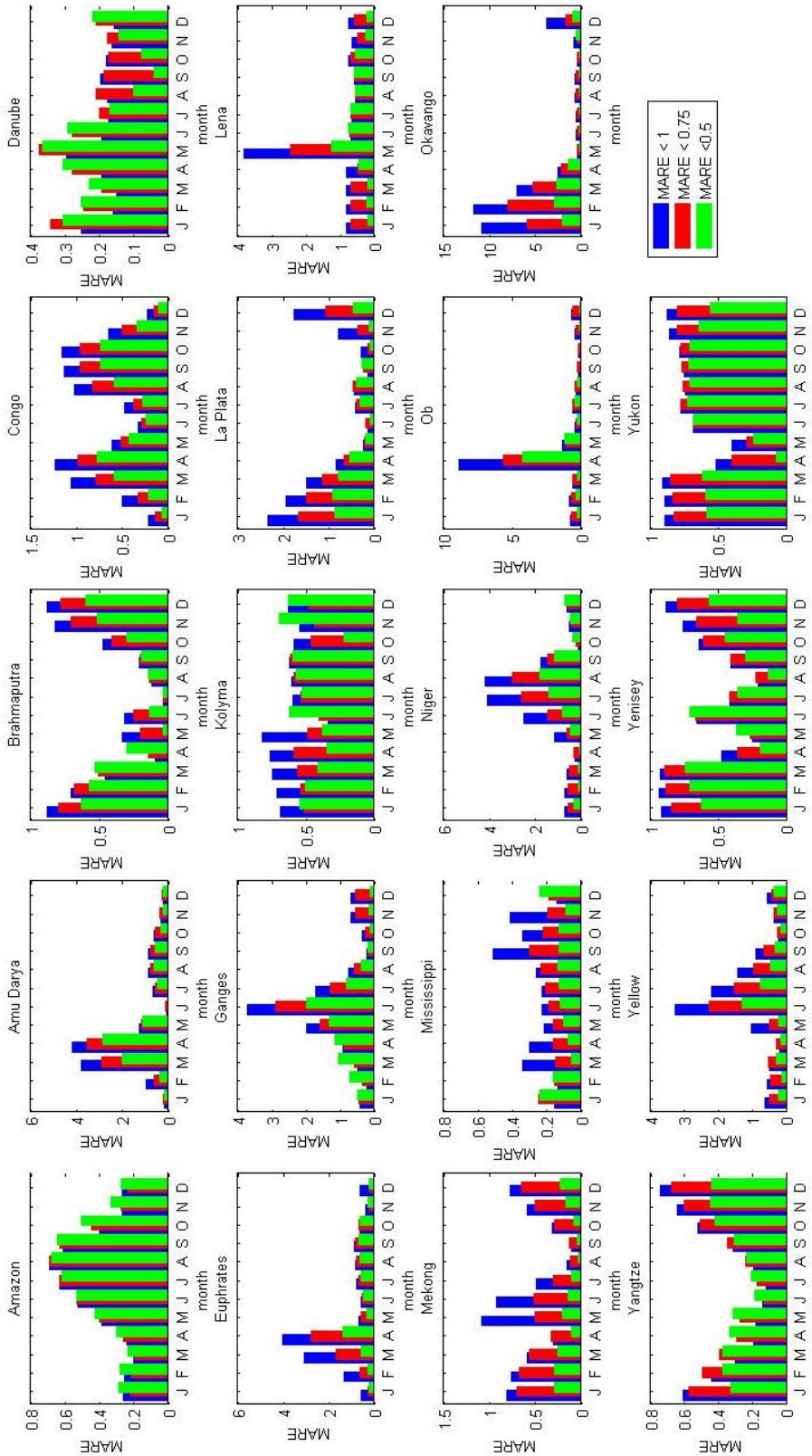


Figure 4.5 Monthly MARE scores for models with catchment averaged MARE <1, 0.7 and 0.5.

4.4 Naturalised Flow Modelling

As previously discussed, each of the catchments has to some extent been modified or is subject to abstractions. Modelling such influences on a catchment is incredibly difficult, primarily due to lack of data, though some hydrological models do take anthropogenic impacts into account; for example, WaterGAP 2 estimates domestic, industrial and agricultural water use, but the results are highly uncertain (Alcamo et al., 2003). GWAVA also considers water use and availability but is used for continental and global scale investigations of water resources scarcity (e.g. Dumont et al., 2010). There have been several research efforts attempting to quantify the volumes of abstractions on a global scale (e.g. Shiklomanov and Rodda, 2003, Alcamo et al., 2007, Shen et al., 2008) alongside databases such as the commonly used AQUASTAT which provides country based water withdrawal data (FAO, 2014a).

Shen et al. (2008) used the AQUASTAT database, alongside an irrigation map and an urban/rural population data set, to develop a geographic distribution of current (2008) water withdrawals for each sector, domestic, industrial and agricultural. They estimated a global total withdrawal for all sectors of 3824.3 km^3 $year^{-1}$. A map of the spatial distribution of these withdrawals is given in Figure 4.6. This map shows that areas with the greatest water withdrawals are India, China and Japan with values up to 7km^3 $year^{-1}$ per grid square (0.5 x 0.5 deg). Europe and the USA also show substantial abstractions. This would suggest that of the chosen study catchments, the Brahmaputra, Ganges, Yangtze and Yellow Rivers would show the greatest discrepancies through the modelling of naturalised flows, however the Brahmaputra and the Ganges rivers are two of the best simulated catchments by the Mac-PDM.14 model.

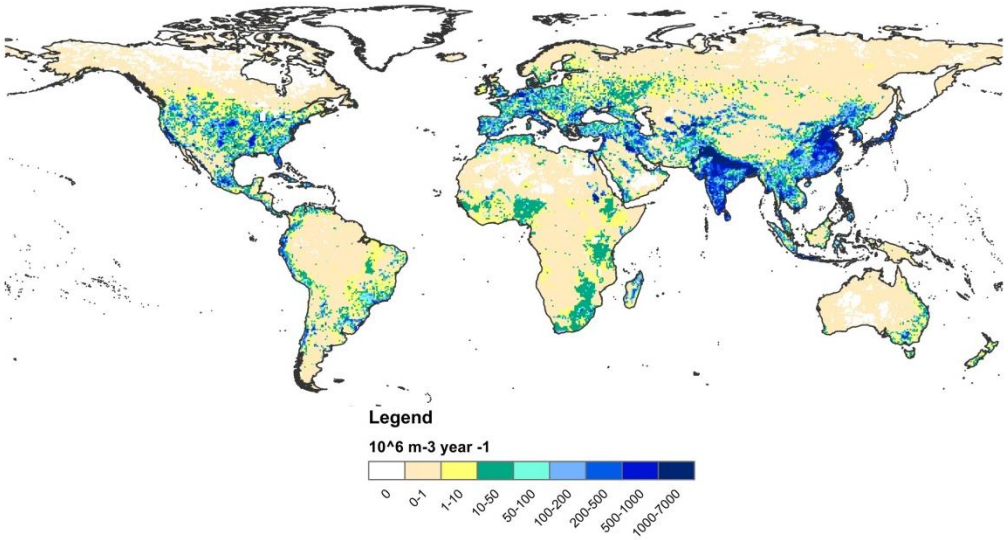


Figure 4.6 Distribution of total water withdrawal (domestic, agricultural and industrial) estimated for the year 2008. *Data kindly provided by Prof. Yanjun Shen, Chinese Academy of Sciences.*

The Nile and the Murray Darling catchments however, show very poor simulations, which were suspected to be the result of abstractions, have very low abstraction levels according to Shen et al. (2008).

Rather than attempting to correct the discharge records to account for the removal of water by abstractions, the University of New Hampshire (UNH) with the Global Rivers Data Centre (GRDC) have instead produced composite runoff fields that combine their datasets of observed discharge records with simulations from the Water Balance Model (WBM) (Fekete et al., 1999, Fekete and Vorosmarty, 2011). Whilst this data does not explicitly represent ‘naturalised flows’, as it is constrained by observed discharge data, the data does represent corrected, spatially distributed runoff for comparison with modelled runoff. Davie et al. (2013) carried out a simple validation of the Global Hydrology Models that took part in the ISI-MIP project (Inter-Sectoral Impacts Model Intercomparison Project) using the UNH/GRDC composite data set. They found that the models tended to predict higher runoff than the GRDC data set.

The composite data set was used to derive monthly average runoff for each catchment to allow for a comparison with the observed record, and with the Mac-PDM.14 adjusted parameter ensemble outputs. The results of this analysis are shown in Figure 4.7. The Niger and the Yellow rivers were excluded from the ranking of GLUE realisations when compared to the composite data, as the composite data showed such negligible flow in these catchments that the MARE values were unreasonable (ranging from 58.9-636.9 with all 21 catchments, but from 2.25-12.71 excluding the Niger and the Yellow). The top ranking model when compared to the composite data scored a MARE of 2.25, which is not much lower than the MARE of the worst model, when compared with the observed record (the best being 0.4733).

A noticeable difference in runoff can be seen between the composite data and the observed record in all catchments. In the aforementioned Brahmaputra, Ganges, and Yangtze Rivers, where abstractions are estimated to be significant (Oki et al., 2001, e.g. Mekonnen and Hoekstra, 2011), the composite data actually shows a significant reduction in runoff compared with the observed record, which is unexpected. The Nile and the Murray Darling catchments however, for which Mac-PDM.14 gave unreasonable results when compared to observed, showed higher flows with the same temporal fluctuations that can be seen the Mac-PDM.14 simulations. The composite runoff for the Euphrates is much higher than the observed record, and indeed higher than the top performing Mac-PDM.14 simulations. The La Plata catchment also shows significantly higher runoff for the composite data in the months December-March. In most other catchments, the composite data is not dramatically different from the observed or the Mac-PDM.14 simulations. It is interesting that the mistiming of the peak flow in the Amazon catchment is still an issue with the composite data, whilst in the Amu Darya the peak of the composite data is predicted to be 4 months prior to that of the observed record.

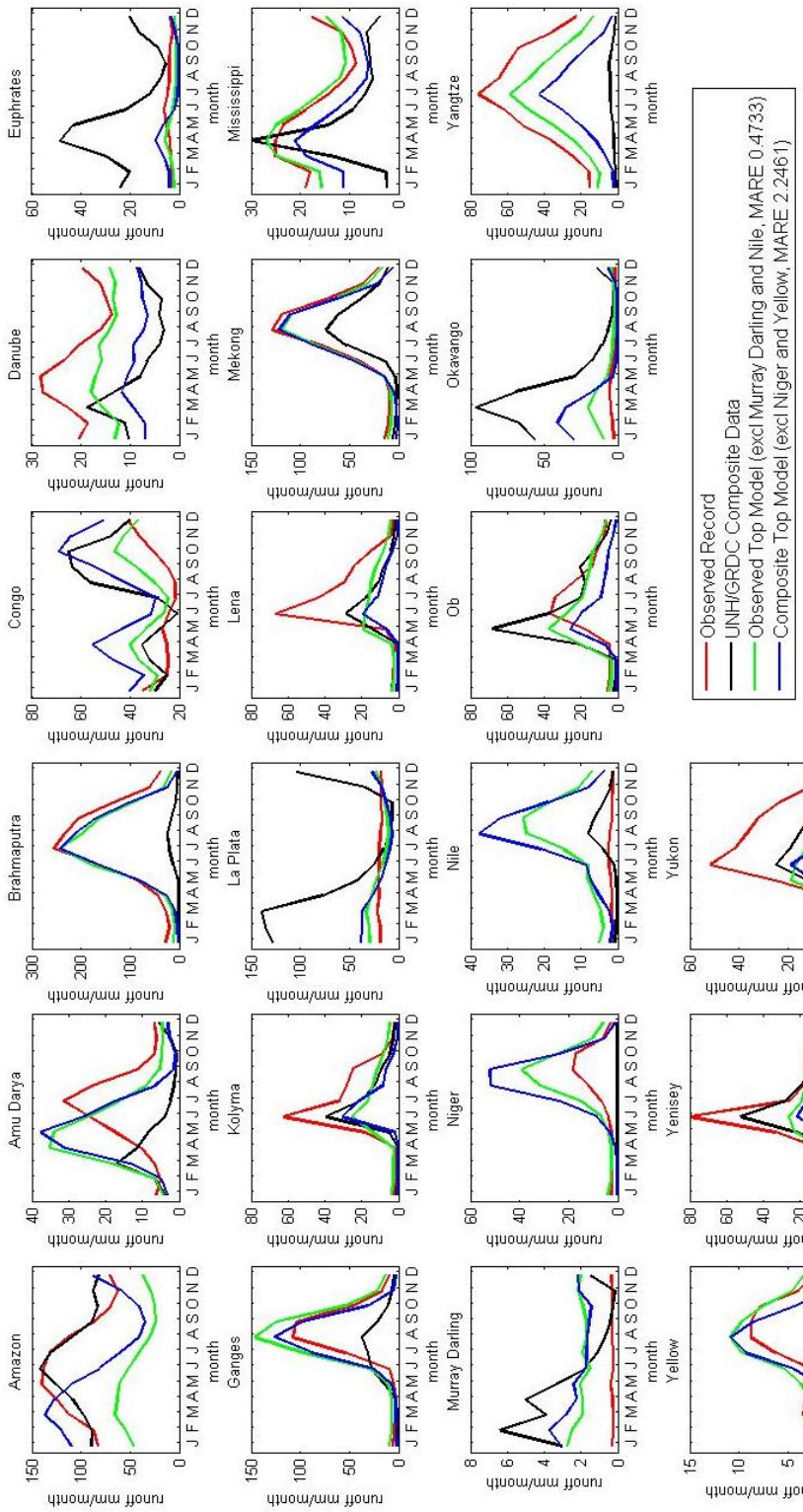


Figure 4.7 Comparison of the UNH/GRDC composite data set with observed discharge record derived runoff and the associated top ranking realisations of the Mac-PDM.14 parameter ensemble.

Fekete and Vorosmarty (2011) explicitly state that the use of the composite data for validation is not recommended due to the fact that it is a combination of observed and modelled discharge. Therefore, whilst this analysis is relevant, the composite data will not be taken forward for use in further analyses of Mac-PDM.14.

Naturalised flow modelling is common in hydrological research, and so comparisons of model output with observed records will not often provide strong agreement. Accurate datasets of abstractions and alterations to the timing of peak flows on a global scale, which are not yet available, are required for meaningful comparisons of observed discharge with model outputs.

4.5 Parameter Uncertainty vs Model Uncertainty

Previous research has investigated the structural uncertainty derived from different GHMs (e.g. Warszawski et al., 2014, Schewe et al., 2014), however the parameter uncertainties within these models has not yet been well assessed. It was therefore considered appropriate to determine whether the magnitude of uncertainties from different sources are comparable.

The EU-WATCH project ran a Multi-Model Ensemble (MME) of 11 global models: 6 Land Surface Models and 5 Global Hydrology Models. The output data for 10 of these models (GWAVA, H08, Htessel, Jules, LPJml, MATSIRO, MPI-HM, Mac-PDM.09, Orchidee, and WaterGAP) is available on the FTP website hosted by the Centre for Ecology and Hydrology. Data from H08 was incomplete. The remaining model data was downloaded and analysed for comparison with the Mac-PDM.14 parameter ensemble. Each model that participated in EU-WATCH was run with the WATCH Forcing Data, which enabled a fair comparison with the Mac-PDM.14 parameter ensemble which was run using the same input data. The EU-WATCH project ran all of the models under naturalised flow options. Even those that had the option to estimate

anthropogenic influences on runoff had those modules switched off to make for a fair comparison.

The results of the WATCH MME, along with the Mac-PDM.14 parameter ensemble, are shown in Figure 4.8. At first glance it is apparent that the models all show very similar results. The Murray Darling and the Nile catchments were included in Figure 4.8, as they show that despite a poor performance when compared to the observed record, the range of different models all show similar outputs. This indicates that either the observed records are unreliable for these catchments, or that all of the hydrology models exhibit similar flaws in their attempts to simulate the Murray Darling and the Nile catchments. Again it is the catchments with very strong seasonal cycles in the hydrological regime that are modelled with more confidence. For example, the models are all very close together for the monsoon-impacted catchments of the Brahmaputra, the Ganges, the Mekong, the Niger and the Yangtze. The models show more variation in the rivers with more even flow distributions throughout the year: the Congo, the Danube, and the Murray Darling.

In terms of comparison with parameter uncertainty, the range between models is rarely larger than the range between parameter realisations with a MARE <1, and the majority actually lie within boundaries of the models that had an overall MARE of <0.75. The Orchidee model tended to have the highest runoff simulations, with the biggest discrepancies between Orchidee and the other models being apparent in the Danube, Mississippi, and Yellow rivers. LPJml also shows high runoff particularly in the Kolyma, Ob, Okavango and Yukon catchments. MPI-HM seems to simulate very high peak flows, with notable peaks exhibited in the Danube, Lena, Mississippi and Yenisey catchments. None of the models appear to simulate particularly low runoff values, although despite the high peaks in some catchments, MPI-HM shows the lowest runoff in the Brahmaputra, Ganges, Nile and Yangtze.

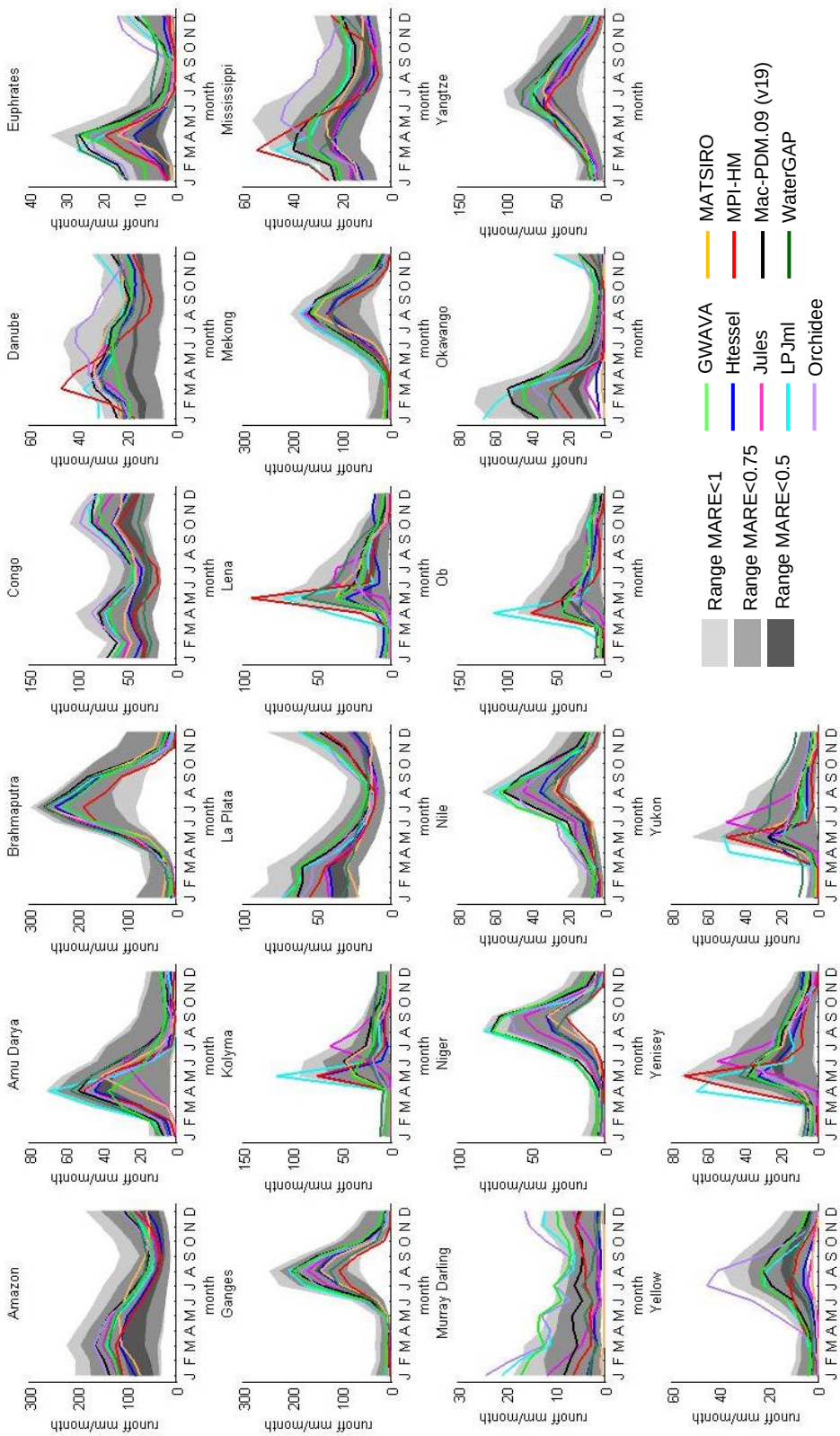


Figure 4.8 Mac-PDM.14 parameter ensemble compared to WATCH multi-model ensemble

The previous version of the Mac-PDM model (Mac-PDM.09, shown in black), seems to fit fairly centrally among the other models.

These results indicate that, depending on the limit of acceptability determined by the modeller (the number of models accepted as behavioural), the range of uncertainty that is derived from parameter uncertainty, is quite similar to the range of uncertainty derived from the type of model employed. This means that the full range of uncertainty from both parameter and structural uncertainty would be rather larger than the ranges shown in Figure 4.8, as the parameter uncertainty of Orchidee will push the upper uncertainty bounds towards higher predictions, and those models that simulate runoff to be lower than Mac-PDM will push the lower boundaries wider if they were assessed for their parameter uncertainty.

The bar chart in Figure 4.9 shows a comparison of the performances of each model in the MME using the MARE metric that was employed in the Mac-PDM.14 parameter ensemble analysis. This chart shows that on average, the models have a similar uncertainty level to the Mac-PDM.14 ensemble results. The top performing Mac-PDM.14 model from the ensemble is shown on the far left, with a MARE for the 12 months of 0.47. The ensemble average for all models that performed better than an overall MARE of 0.75 gave a MARE of 0.67.

Of the MME models, MATSIRO performed the best when assessed against the 19 catchments, with a MARE of 0.5, and LPJml performed least well with a MARE of 1.6. None of the models performed better than the top ranking GLUE realisation of Mac-PDM.14. Htessel, Jules, and MATSIRO were the only models to perform better than the Mac-PDM.14 ensemble average. These models, as well as MPI-HM and WaterGAP, performed better than the version of Mac-PDM used by the WATCH project (labelled Mac-PDM in the graph).

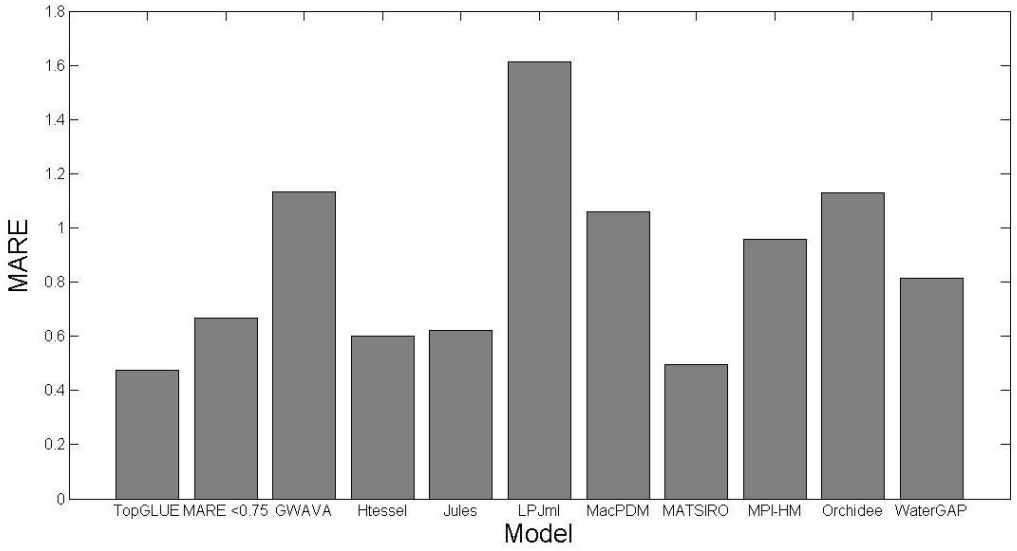


Figure 4.9 Bar chart showing the MARE for each EU-WATCH MME model when compared to observed records (MARE calculated for Jan-Dec for 19 catchments). Far left bars are the top ranking Mac-PDM.14 GLUE model and the weighted average for all parameter ensemble models with MARE <0.75.

4.6 Assessment Feasibility

This analysis of the parameter uncertainty of a Global Hydrology Model has been undertaken to determine whether this approach could be included in the calibration process in the development of GHMs. The choice of the number of model realisations to run is another of the subjective steps of an uncertainty analysis, on top of the choice of evaluation metric, sampling strategy and limit of acceptability, and depends on the computational resources available to the modeller as well as the number of parameters being assessed. This study investigated 123 model parameters, and ran an ensemble of 100,000 parameter realisations. Using the University of Nottingham High Performance Computer Cluster (HPC) this took approximately 40 days, and output just over 2 terabytes of data. Whilst the model does not require significant amounts of RAM to run (less than 2.5GB), the HPC offered the ability to run several hundred model realisations at once, and to queue the jobs, which allowed efficient transitions when model runs had completed. The analysis of the data took additional time and hard disk space. Without access to such computational power,

assessing the parameter uncertainty of a model could take considerably longer; therefore, several smaller parameter ensembles were run in order to investigate the impact of ensemble size on the performance assessment of the model. Ensembles of 10,000, 5,000 and 1,000 model realisations were run, using the same parameter distributions and sampling technique as the 100,000 ensemble of model realisations. The mean absolute relative error (MARE) metric was used, as before, excluding the Murray Darling and Nile catchments.

The distributions of the MARE scores for each parameter ensemble are shown in Figure 4.10. These histograms show that the MARE scores are similarly distributed across the different ensemble sizes, ranging from ~0.5 to ~2.4, with peaks between 1 and 1.1. This demonstrates that reducing the ensemble size does not mean that the modeller is less likely to obtain a “good” model. However, the smaller the sample size, the fewer “good” models there are to choose from, as is demonstrated in Table 4.2. The statistics in the table show that only the 100,000 model ensemble achieved a MARE of <0.5. Therefore, in order to account for the issue of equifinality, the modeller might decide to relax the limit of acceptability for smaller sample sizes, rather than just accept the one or two models that meet a stricter criterion.

The best model for each ensemble had MARE values of 0.59, 0.55, 0.56 and 0.47 for the 1, 5, 10 and 100 thousand model ensembles respectively. This shows that increasing the number of realisations does reduce the MARE of the best model slightly. The top ranking model from each ensemble is shown in Figure 4.11. The graphs in Figure 4.11 show that the best models for each realisation ensemble give very similar results, and that the model outputs do not progress towards the observed record with more model realisations. For example, in the Danube catchment, the 100,000 realisation ensemble performs much better than the 10,000 realisation ensemble, but the 1,000 and 5,000 realisation ensembles are better than the 100,000.

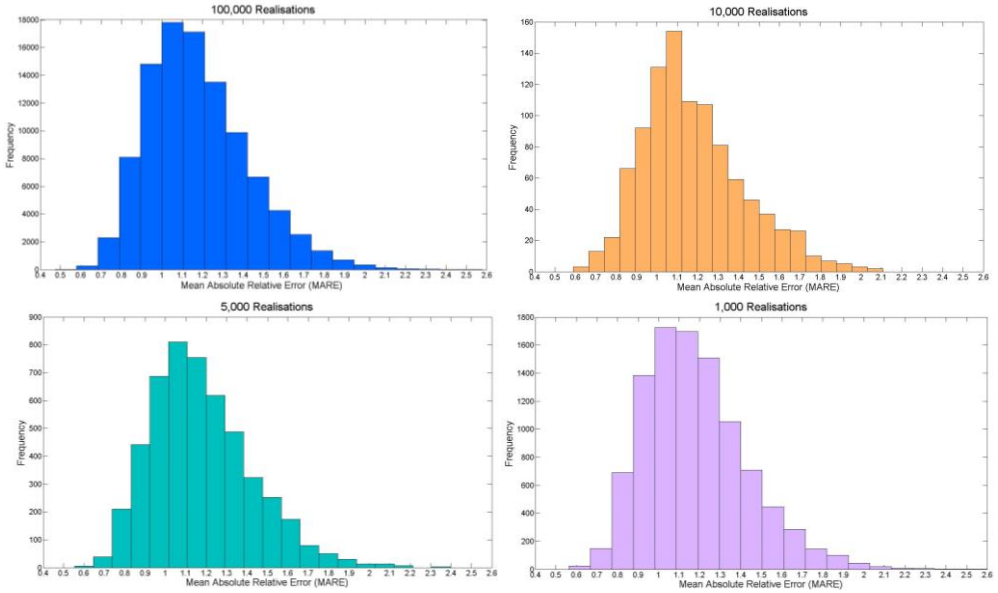


Figure 4.10 Histograms of MARE scores for the 100, 10, 5 and 1 thousand realisation runs.

Table 4.2 Number of model realisations that achieved MARE scores of 0.5 to 1 for each of the 4 ensemble sizes

Number of Realisations in Ensemble	*Number (and percentage of ensemble size) of Model Realisations with a MARE less than or equal to:*					
	0.5	0.6	0.7	0.8	0.9	1
100,000	2 (0.002)	40 (0.4)	1238 (1.238)	3092 (3.092)	10656 (10.656)	25532 (25.532)
10,000	0 (0)	5 (0.05)	44 (0.44)	282 (2.82)	1102 (11.02)	2533 (25.33)
5,000	0 (0)	2 (0.04)	20 (0.4)	163 (3.26)	583 (11.66)	1267 (25.34)
1,000	0 (0)	1 (0.1)	6 (0.6)	33 (3.3)	109 (10.9)	242 (24.2)

Similarly, in the Congo catchment, the 100,000 model realisation ensemble performs much better than the 5,000 and 10,000 model ensembles, but the 1,000 ensemble also performs well. The 100,000 realisation ensemble only performs noticeably better than the other ensembles in the Euphrates, Mississippi, Niger, Ob and Okavango catchments.

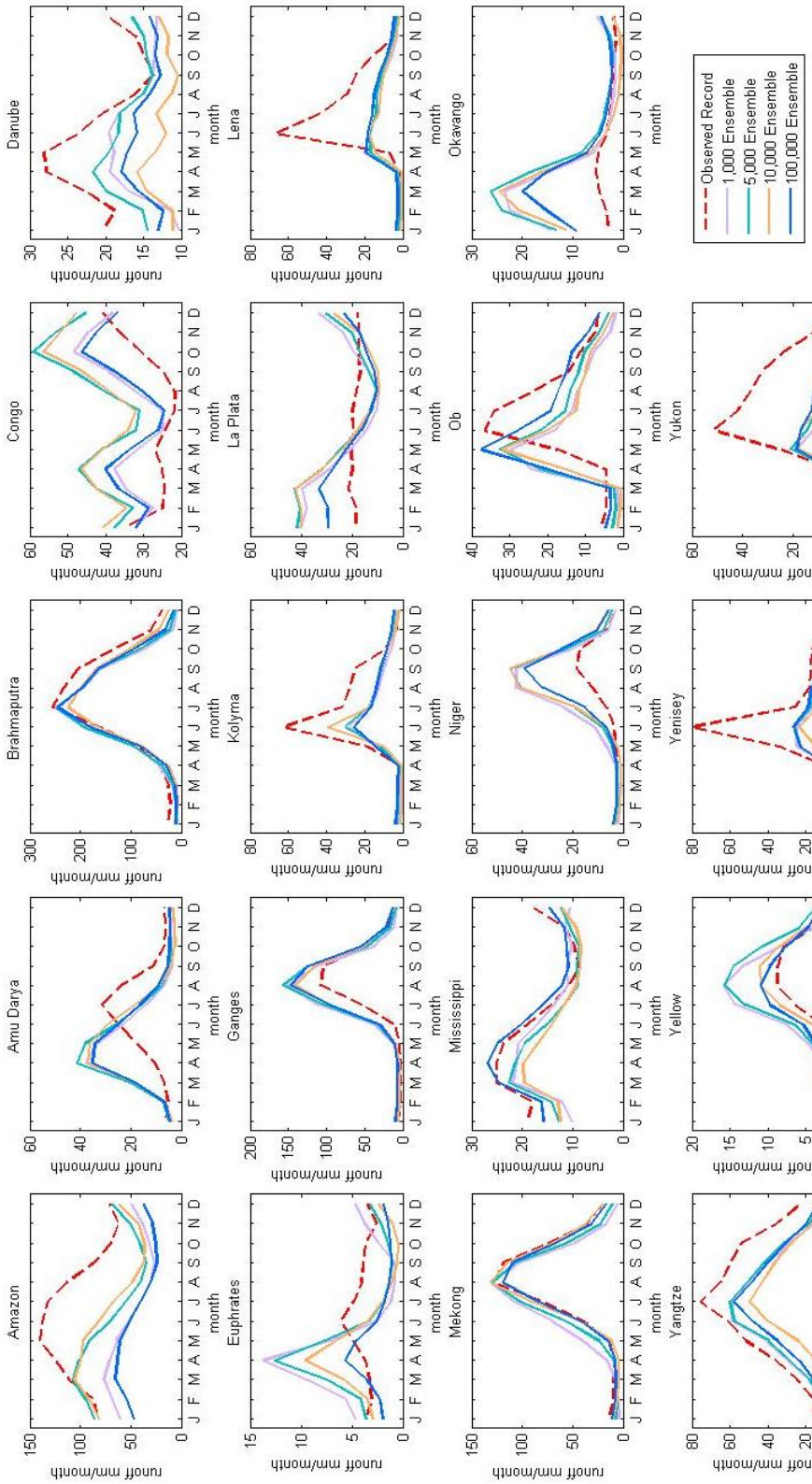


Figure 4.11 Top ranking models for the 1, 5, 10 and 100 thousand realisation ensembles

These results are similar to those of the weighted average results, where we would expect to see a trend towards the observed record as we "improve" the model, but the results are instead rather erratic, and inconstant between catchments. This is likely due to the attempt to seek a "globally good" model, which is pushing to increase the runoff in catchments such as the Amazon, Lena, Yenisey, and Yukon, whilst simultaneously attempting to reduce the simulated runoff in the Ganges, Congo, Niger and Yellow Rivers. This raises the question *"is it really possible, or indeed sensible, to simulate global runoff using one set of model parameters?"* This is addressed further in Chapter 6.

The outputs of the ensemble size experiment were analysed further to assess the impact of ensemble size on the best model output and the range of outputs within an acceptability limit (MARE <0.75). Findings for 6 of the 19 catchments are shown in Figure 4.12 which demonstrates that although the top performing models give very similar outputs (as highlighted in the graphs in Figure 4.11), the ranges of model outputs within a specified limit of acceptability vary significantly with ensemble size. With a MARE < 0.75, the model ensembles had 1238, 116, 58 and 18 models accepted as behavioural for the 100, 10, 5 and 1 thousand realisation ensembles respectively. It is evident from the graphs in Figure 4.12 that the smaller the ensemble size, the smaller the range of model outputs with a MARE < 0.75. What this means is that if a modeller uses smaller ensembles, they may get a similarly good top ranking model to that achieved from a larger ensemble, but they would be underestimating the parameter uncertainty within the model significantly.

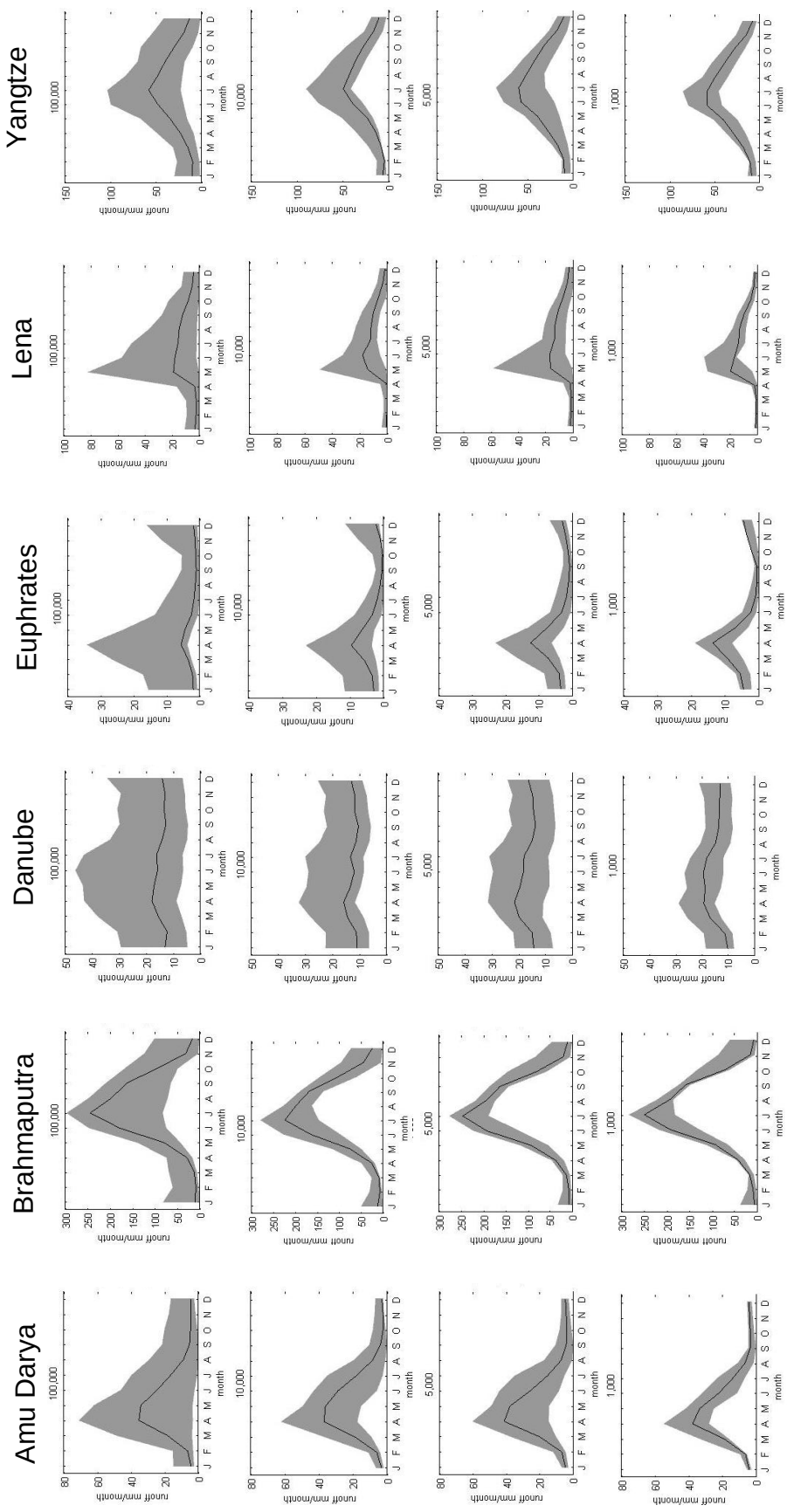


Figure 4.12 Top ranking models and range of models with MARE <0.75 for the 4 ensemble sizes over 6 example catchments

4.7 Summary

This chapter has presented a Generalised Likelihood Uncertainty Estimation (GLUE) analysis of the Mac-PDM.14 model, which simulated and analysed 100,000 model parameter perturbations. It was determined that the Mac-PDM.14 model does not perform well for the Murray Darling and Nile catchments, and when these were excluded, leaving 19 study catchments, the model realisation with the best performance across these catchments had a Mean Absolute Relative Error (MARE) of 0.47 (an average error of 47%). A likelihood measure was used to calculate weighted averages for each catchment which used more information from the parameter ensemble than looking at just the top performing model. It was found that for most catchments, constraining the limit of acceptability (the number of models taken as 'behavioural') provided a better fit with the observed values, but due to the attempt to match so many catchments with different hydrological regimes, with Mac-PDM.14 overestimating some and underestimating others, not all catchments gave better results. The MARE per month for each catchment was investigated, which revealed that Mac-PDM.14 does not systematically favour the more accurate simulation of either high or low flows, but that highest error varied temporally between catchments.

The Mac-PDM.14 outputs were compared to composite runoff data from the University of New Hampshire/Global Rivers Data Centre in an attempt to account for some of the impacts of abstractions and dams on the observed record. The results of this comparison showed that Mac-PDM.14 was better at simulating the observed record than the composite data, with the best model compared to the composite data having an average MARE of 2.25. The composite data indicated that the Murray Darling and the Nile simulations are more reasonable than the observed record suggests, but that the Brahmaputra and Ganges, that showed excellent results when compared to the observed record,

significantly overestimated the composite data. Since the composite data is a combination of modelled runoff and observed discharge records, the data could not be used for model validation, due to circularity of argument.

The parameter uncertainty of Mac-PDM.14 was then compared to model structural uncertainty, using data from the EU-WATCH MME. The outputs of 9 models were analysed, and plotted against the ranges of outputs from Mac-PDM.14 with different thresholds of MARE. The results of this indicated that the range of outputs from different models closely reflected a range of a parameter ensemble of MARE<0.75. Due to the subjectivity of deciding upon an evaluation metric, and determining a limit of acceptability in a parameter ensemble, it cannot be said whether parameter uncertainty is higher or lower than the uncertainty derived from employing different models; however the results are comparable, and the models mostly simulate similar seasonal runoff cycles. The MARE range across the MME models was 0.5 to 1.61; the range or MARE values across the entire parameter ensemble was 0.47 to 2.58.

Finally, since this uncertainty analysis involved an arduous 100,000 model realisations, several smaller experiments were run in order to compare the results and determine whether uncertainty experiments need to be so rigorous, or whether smaller scale studies could provide adequate insight into the parameter uncertainty of a model. It was found that with ensemble sizes of 1, 5, 10 and 100 thousand realisations, the top performing models all gave good performance, with a slight reduction in MARE values as the ensemble size was increased (the ensembles gave MARE values of 0.59, 0.55, 0.56 and 0.47 respectively). The distribution of errors across the ensembles was also very similar.

The main issue with smaller ensembles was that, despite the top performing model giving a good MARE score, there were far fewer

models with good scores. For example in the 100,000 model ensemble, 40 models scored a MARE <0.6, whilst of the 1,000 model ensemble only 1 model scored a MARE <0.6. This means that if the modeller wishes to choose a fairly strict limit of acceptability, the range of uncertainty in the model output may be misleadingly low. The misperception of uncertainty from small ensembles has also been demonstrated by using a set MARE threshold, whereby the range of model outputs is significantly larger for larger ensembles. This shows that a modeller should be as rigorous as their computational capacity and budget allows, and should consider the fact that the uncertainty ranges found from a parameter ensemble will likely only be a subset of the true uncertainty range.

The next chapter will review the potential of using Mac-PDM.14 as a catchment model, and will investigate the parameter values that result in good model realisations.

Chapter Five:

Parameter Estimation and Global Models as Catchment Models

5.1 Introduction

The Generalised Likelihood Uncertainty Estimation (GLUE) experiment presented in Chapter 4 indicated that the Mac-PDM.14 model is better at simulating some catchments than others, and that the drive to define a 'globally good' model led to conflicting results. The catchments that were overestimated by Mac-PDM.14 had higher errors than the catchments that were underestimated, and therefore, model realisations that sought the lowest overall error were biased towards accurate simulation of these catchments. This raised the question, *"how well could the Mac-PDM.14 model perform when calibrated for individual catchments, and how much better could these outputs be, over those of a globally calibrated model output?"* Therefore, this chapter seeks to determine whether it is indeed sensible to simulate global runoff using one set of model parameters across the entire global domain, or whether it may be more sensible to use global hydrology models as catchment models. The parameter values that produce both a 'good' global, and a 'good' catchment model are evaluated.

5.2 Catchment Calibration of a Global Hydrology Model

In order to rank the model realisations for a 'globally good' model, the Mean Absolute Relative Error (MARE) evaluation metric was used. The 30 year average annual runoff, as well as the monthly runoff statistics for each catchment, were compared to the observed record, and then averaged across catchments to give a score across all catchments. Going back to the individual catchment MARE scores, before averaging them, facilitates an assessment of how well Mac-PDM.14 could perform if it were to be employed for a single catchment.

Figure 5.1 shows the percentage of models within the 100,000 realisation ensemble that scored a MARE below several intervals up to a value of 1 for each catchment. Here, the taller the bar and the more green/blue that is visible, the better the results. It is immediately

apparent that the Mac-PDM.14 model performs better in some catchments than others. For example, in the Amazon and the Yangtze, all of the 100,000 model realisations present a MARE <1. The Brahmaputra, Kolyma, Yenisey and Yukon also come very close to this, whereas the Okavango only has 13 model realisations <1, and the Niger only 561. Figure 5.1 also indicates consistency across the model realisations by the length of the colour bars; for example, the Brahmaputra River shows 89% of its models giving a MARE between 0.4-0.5, and the Amazon has 62% of its models between 0.3 and 0.4. The Mississippi, Congo and Euphrates show a more even distribution of MARE values across the ensemble. This perhaps suggests that the model is less sensitive to parameter perturbations in the Brahmaputra, Amazon Yangtze and Yukon catchments. The Yangtze catchment gives the model best performance, with one model giving a MARE value of <0.05, and more than half the models scoring less than 0.3. However, from the graph it can be seen that the Danube and the Mississippi give the highest number of models with MARE <0.2.

Figure 5.2 shows the ranges in model outputs across the top 20 models when globally-calibrated, compared to the ranges of the top 20 models when calibrated against each catchment individually, for the 19 catchments. Table 5.1 shows the monthly and annual average MARE values for the best globally-calibrated and catchment-calibrated model realisations, and shows which performed better. These graphs show that catchment-calibrated models perform significantly better than the globally calibrated models, and that improvements are seen in all catchments. Table 5.2 shows that the annual average MARE for each catchment is improved when employing the catchment-calibrated model over the globally-calibrated model for all 19 catchments.

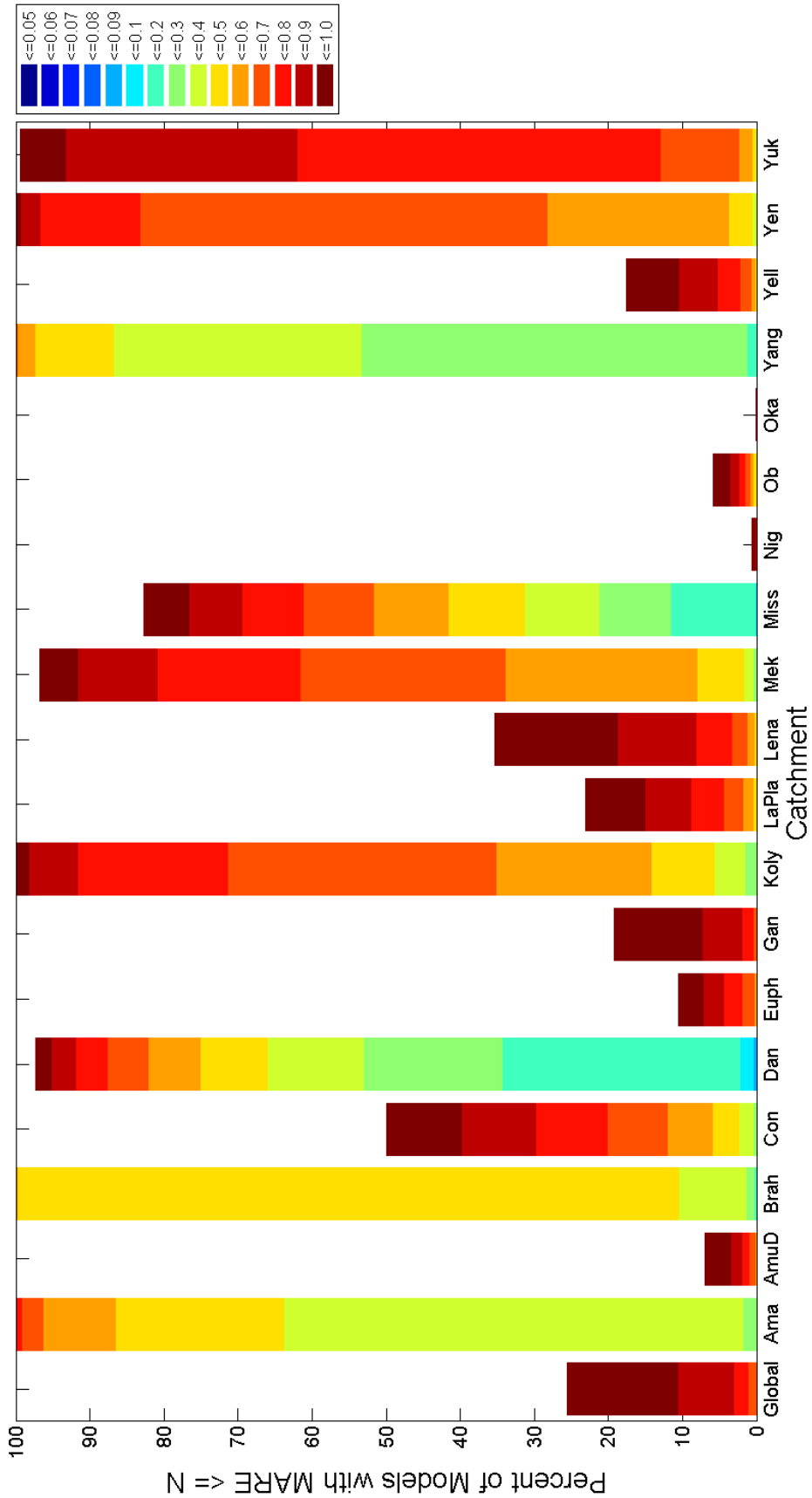


Figure 5.1 Percentage of models with MARE less than set values between 0.05 and 1 for each catchment and globally. Catchment codes are listed in the caption of figure 4.3

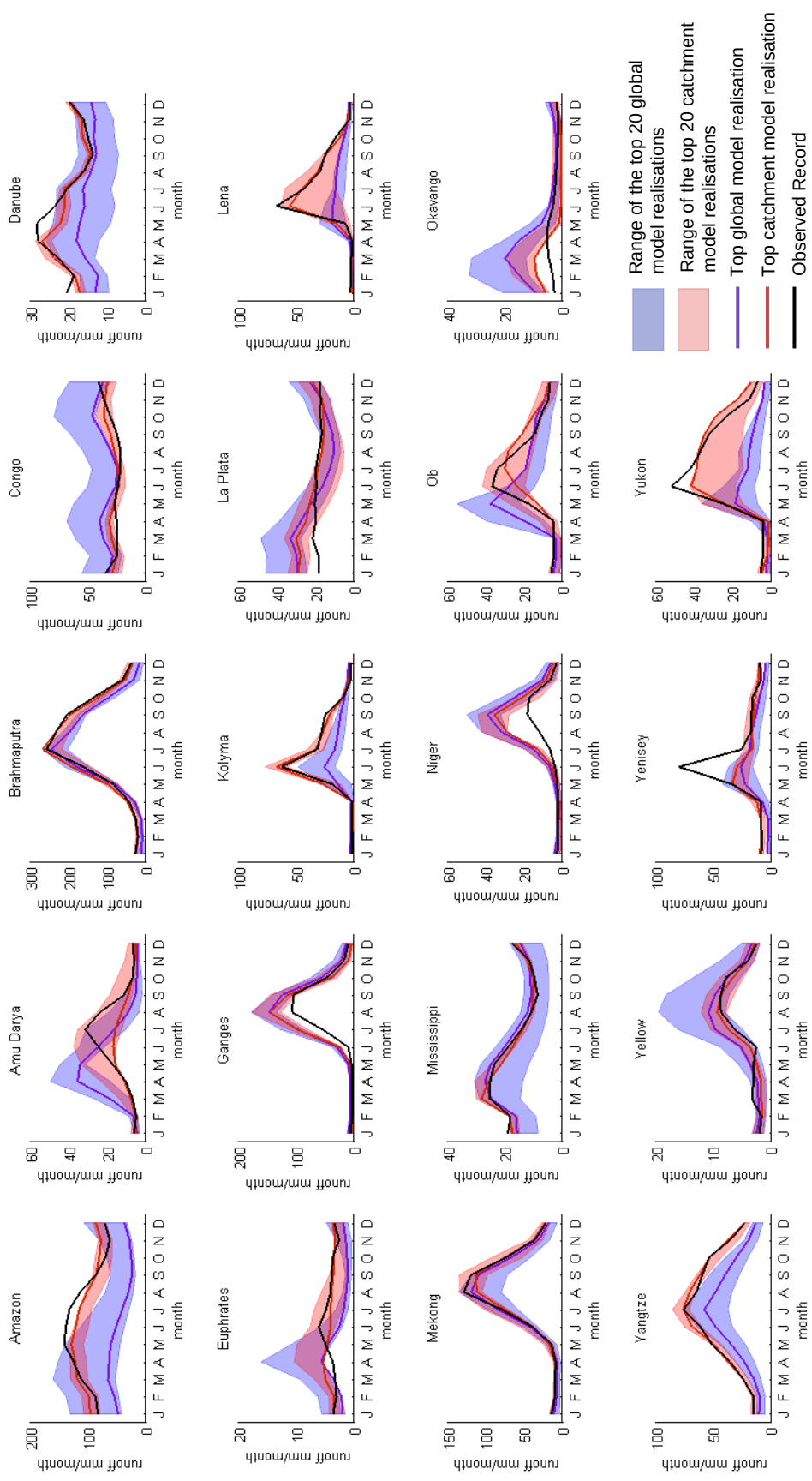


Figure 5.2 Comparison of the top 20 globally-calibrated models with the top 20 catchment-calibrated models, for the 19 catchments.

Table 5.1 MARE Values for the best globally-calibrated and catchment-calibrated model realisations showing where the catchment-calibrated model outperformed the globally-calibrated model.

	January		February		March		April		May		June		July		August		September		October		November		December		Average Annual	
	Global	Catch	Global	Catch	Global	Catch	Global	Catch	Global	Catch	Global	Catch	Global	Catch	Global	Catch	Global	Catch	Global	Catch	Global	Catch	Global	Catch	Global	Catch
Ama	0.153	0.1653	0.0955	0.1252	0.2082	0.0644	0.3231	0.0246	0.4489	0.0836	0.5572	0.1232	0.6425	0.1271	0.7022	0.1067	0.6797	0.0067	0.5648	0.1614	0.3889	0.2413	0.2104	0.2241	0.4145	0.1211
AmuD	0.4075	0.0851	0.1749	0.0783	0.2188	0.1431	0.768	0.0634	0.8135	0.0899	0.405	0.311	0.0136	0.4951	0.1991	0.4673	0.0115	0.1887	0.0141	0.0275	0.3121	0.0689	0.4573	0.1537	0.3163	0.181
Brah	0.2501	0.0703	0.4122	0.0474	0.5312	0.0569	0.3713	0.1259	0.0688	0.1698	0.034	0.188	0.2064	0.0424	0.2439	0.0012	0.214	0.0404	0.2114	0.0864	0.114	0.0996	0.0952	0.0909	0.2294	0.0849
Con	0.018	0.1655	0.187	0.0034	0.5163	0.2079	0.6961	0.2462	0.359	0.0097	0.1952	0.0356	0.2651	0.0859	0.6166	0.248	0.7116	0.2315	0.7545	0.1808	0.4282	0.0334	0.0656	0.1824	0.4011	0.1359
Dan	0.5325	0.1625	0.4653	0.0183	0.4182	0.0471	0.566	0.043	0.6759	0.2158	0.6623	0.1111	0.5595	0.0377	0.5291	0.0155	0.483	0.0553	0.4604	0.0907	0.4782	0.0386	0.4723	0.0182	0.5252	0.0711
Euph	0.575	0.0205	0.3	0.1487	0.435	0.5209	0.9824	0.4588	0.2056	0.0591	0.6129	0.3343	0.6764	0.1662	0.7675	0.0382	0.8593	0.0796	0.8999	0.0487	0.8173	0.2714	0.6909	0.0337	0.6519	0.1817
Gan	1.1344	0.4329	0.8687	0.2499	0.7865	0.0321	0.6595	0.0154	0.5913	0.3583	1.4587	1.4669	0.794	0.8061	0.3714	0.3622	0.1464	0.089	0.203	0.1839	0.8687	0.4668	0.9803	0.5653	0.7386	0.4191
Koly	0.0013	0.0017	0.3598	0.1056	0.6176	0.2092	0.781	0.2687	0.8777	0.454	0.7341	0.0729	0.332	0.0268	0.3794	0.0446	0.489	0.1565	0.0618	0.0089	1.105	0.4843	0.5459	0.1907	0.5237	0.1687
LaPla	0.5909	0.5774	0.553	0.5139	0.3902	0.3995	0.1897	0.2518	0.0683	0.1384	0.2677	0.045	0.4697	0.0553	0.551	0.0828	0.5119	0.0698	0.4098	0.0968	0.2337	0.054	0.1581	0.2126	0.3662	0.2081
Lena	0.376	0.3501	0.587	0.4159	0.7093	0.412	0.5258	0.2006	0.2185	1.0832	0.794	0.1703	0.6812	0.0435	0.5747	0.0631	0.5616	0.0335	0.4081	0.2491	0.5305	0.2812	0.0988	0.0756	0.5055	0.2815
Mek	0.444	0.0455	0.566	0.0453	0.6738	0.0555	0.7333	0.1118	0.4817	0.1915	0.1249	0.0281	0.1405	0.0575	0.2749	0.1335	0.3449	0.0496	0.4066	0.0186	0.3824	0.0356	0.3534	0.0786	0.4105	0.0709
Miss	0.3926	0.0914	0.3202	0.0231	0.2582	0.1218	0.3278	0.0242	0.4248	0.1018	0.4369	0.0823	0.3808	0.0278	0.3591	0.0067	0.2258	0.1746	0.3058	0.048	0.3103	0.0599	0.3824	0.0757	0.3437	0.0698
Nig	0.3751	0.1825	0.1502	0.065	0.4241	0.1162	0.636	0.1799	0.5178	0.1291	0.0331	0.3924	0.9134	0.9502	1.6561	1.4515	0.7655	0.9405	0.0486	0.2059	0.5172	0.3359	1.0379	0.5788	0.5896	0.4607
Ob	0.6389	0.0134	0.74	0.0316	0.7094	0.002	5.4401	0.1561	0.6358	0.279	0.6034	0.3504	0.6788	0.1185	0.566	0.18	0.3762	0.473	0.2664	0.3516	0.2295	0.2747	0.4596	0.0891	0.9454	0.1933
Oka	6.5568	1.1149	7.4471	1.6829	4.0126	1.0572	1.5743	0.0725	0.043	0.7708	0.1606	0.8278	0.3097	0.8438	0.4877	0.8616	0.6162	0.8542	0.6888	0.6812	0.387	0.2878	1.1536	0.1123	1.9531	0.7639
Yang	0.3805	0.0303	0.5163	0.0023	0.5165	0.0022	0.4906	0.0474	0.5385	0.0425	0.43	0.1146	0.5028	0.0056	0.4636	0.0925	0.4715	0.0359	0.4764	0.0119	0.4115	0.0592	0.3825	0.0054	0.4651	0.0375
Yell	0.2974	0.0474	0.1742	0.0369	0.703	0.4782	0.7152	0.4361	0.5145	0.1966	0.0249	0.4205	0.2085	0.1842	0.0766	0.0699	0.0961	0.0094	0.1669	0.1518	0.6939	0.006	0.8316	0.1479	0.3752	0.1821
Yen	0.7043	0.1384	0.8378	0.0493	0.9018	0.0162	0.3381	0.3468	0.3873	0.0021	0.7932	0.6286	0.4086	0.344	0.1957	0.0429	0.2743	0.0946	0.3549	0.1473	0.2108	0.2615	0.5383	0.0831	0.4954	0.1796
Yuk	0.77	0.1023	0.8269	0.458	0.8804	0.677	0.5253	0.5388	0.4673	0.0235	0.8202	0.1977	0.8006	0.0713	0.8128	0.0279	0.8193	0.0747	0.7952	0.2342	0.7139	0.5292	0.688	0.4287	0.7433	0.2803

The timing of the peak flows in the Amazon, Amu Darya and Ob catchments, which were simulated several months too early using the globally-calibrated models, are much improved. Figure 5.2 shows that the overestimation by the globally calibrated models of the peak flows in the Congo, Euphrates, La Plata, Okavango and Yellow, are all reduced in the catchment calibrated models, giving a much better fit with the observed record. Table 5.2 indicates that the Okavango catchment gives worse results from the catchment-calibrated model than the globally-calibrated model for the months May-September, however these are the lower flow months, and the significant improvement in the simulation of December-April outweighs the small decrease in performance over the summer months. The underestimations of the Amazon, Danube, Kolyma, Lena, Yangtze and Yukon are also improved, with the catchment simulations giving higher runoff outputs.

Despite the improvement in timing of the peak flow in the Amu Darya catchment, the magnitude of the peak flow is better simulated by the globally-calibrated model. The Brahmaputra, which was already well simulated, shows little change between the catchment-calibrated and the globally-calibrated models, and the Ganges, Niger and Yenisey show only a little improvement using catchment-calibrations. The Mekong, Mississippi and Yellow were fairly well simulated by the global top performing model, but the range of outputs (and hence the uncertainty) is reduced significantly when catchment specific model realisations are applied. Table 5.2 shows that the Danube, Mekong, Mississippi and Yangtze catchments had improvements in MARE when using the catchment-calibrations over all months of the year. The Euphrates, Kolyma, Lena, Yellow and Yukon show improvement in all but one month.

In most catchments, the observed record is within the range of the top 20 catchment calibrated models for the majority of the year. In the Amu Darya, Congo, Lena, Ob, and Yangtze, the observed record did not fit

within the 20 top global models, but does within the catchment models. In the Amazon, Danube, Euphrates, Ganges, La Plata, Niger, Okavango, Yenisey and Yukon however, there remain months that are not well simulated by even the catchment-calibrated models. However, the observational record is an average of the 30 year period of 1971-2000, within which the runoff may have varied significantly, so some consideration of the natural variability of the catchment runoff should be taken.

5.2.1 Natural Variability

Figure 5.3 shows a similar evaluation of the model's overall 'goodness' in each catchment, but investigates the range of readings within the observed record. Since the comparison of the models with observed records has so far been focused upon matching the 30 year average, this has not taken account of how variable the runoff in the catchments may be.

One model that seems distant from the average may still be well within the range of 'natural variability' (the minimum and maximum observed values within the 30 year record), whilst another that seems to better fit the observed average may not be within the full range of the observed record if the river flow is very consistent year on year.

Figure 5.3 shows the number of model realisations within the ensemble of 100,000 models that achieve 0-12 months within the observed minimum and maximum runoff values. The Danube, Kolyma, Mississippi and Yangtze perform the best, with 51,094 of the model realisations lying within the range of natural variability for the Danube catchment for all 12 months of the year. For the Mississippi, 27,143 models fit all 12 months, and the Kolyma 20,342. The Congo exhibits the largest number of models that do not fit any of the months, at 51,406; however more of the models manage to fit a lower number of months than the Yukon, where 98,824 of the models could not fit more

than 2 months of the year within the range of the observed values. Again the Brahmaputra stands out as being a consistently well simulated catchment across the ensemble, with 72,449 of the realisations achieving 6 months within the observational range. Figure 5.3 shows which catchments give the best results when natural variability is taken into account, however the actual range of natural variability is not presented, so it may be that some catchments have wider bands of variability, and hence are easier to model. The true ranges of natural variability are shown in the hydrographs of Figure 5.4, alongside the ranges of the top 20 catchment-calibrated and globally-calibrated model realisations.

These graphs show that all catchments have significant ranges between the minimum and maximum observed values, and that in general, the largest ranges are seen during months of highest runoff. In addition, the maximum observed values are often further away from the 30 year mean than the minimum observed values. The Murray Darling and the Nile catchments were included in this set of graphs, as the very large range in the observational values for the Murray Darling show that the outputs of the Mac-PDM.14 model, when calibrated towards that specific catchment, are not as bad as the global top performing model presented in Chapter 4 (see Fig 4.1). In fact, the top 20 catchment models for the Murray Darling River lie well within the maximum observed values of the catchment record. The River Nile however, has an incredibly low range across the minimum and maximum values of observations and so the model still performs very poorly in this catchment.

From these graphs it can be seen that where the range of the top 20 catchment models did not fit the mean observed record in the Euphrates, Ganges, La Plata, and Yukon catchments, the model outputs are within the range of individual observations.

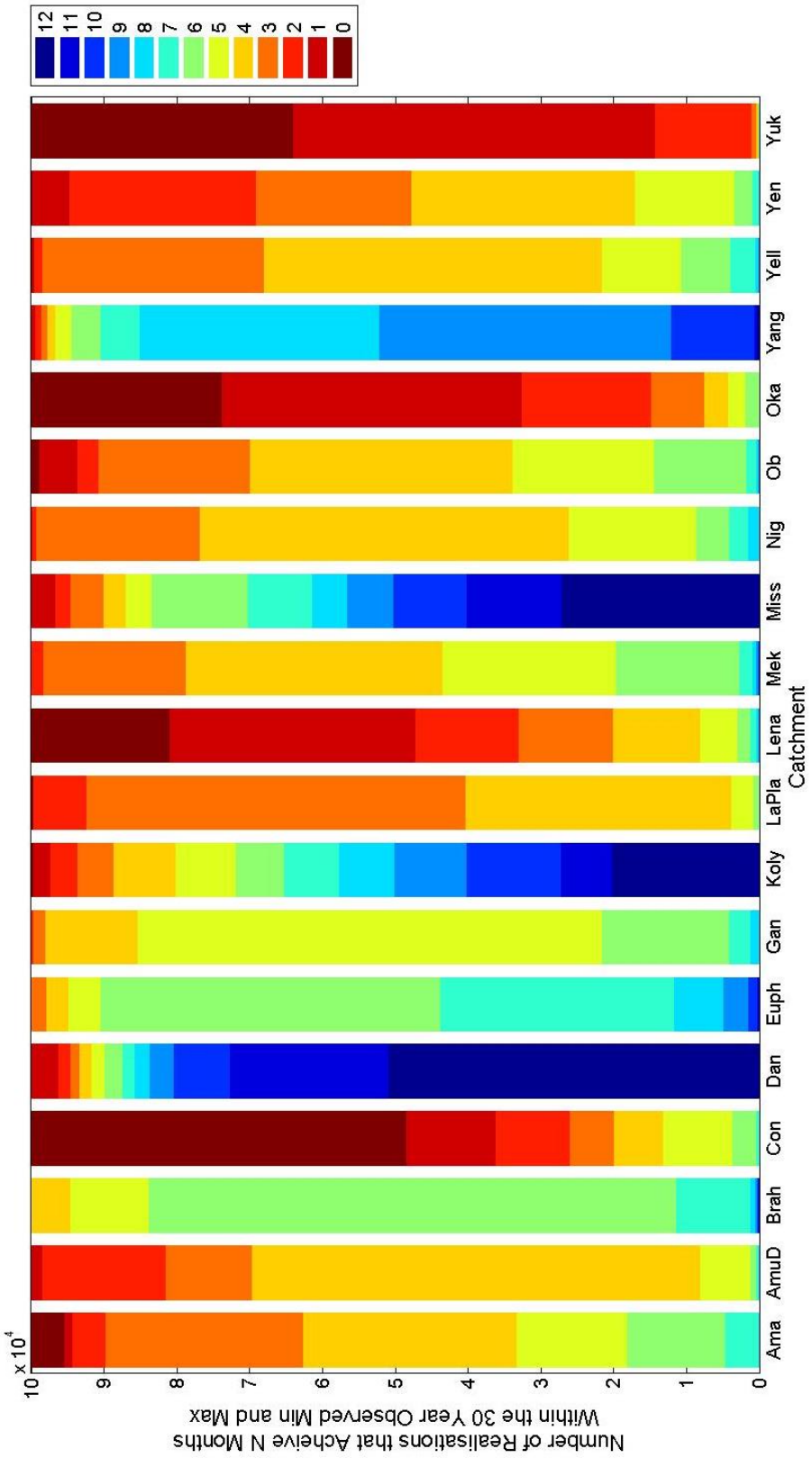


Figure 5.3 Number of Models with N months within the 30 year minimum and maximum observations. Catchment codes are listed in the caption of figure 4.3

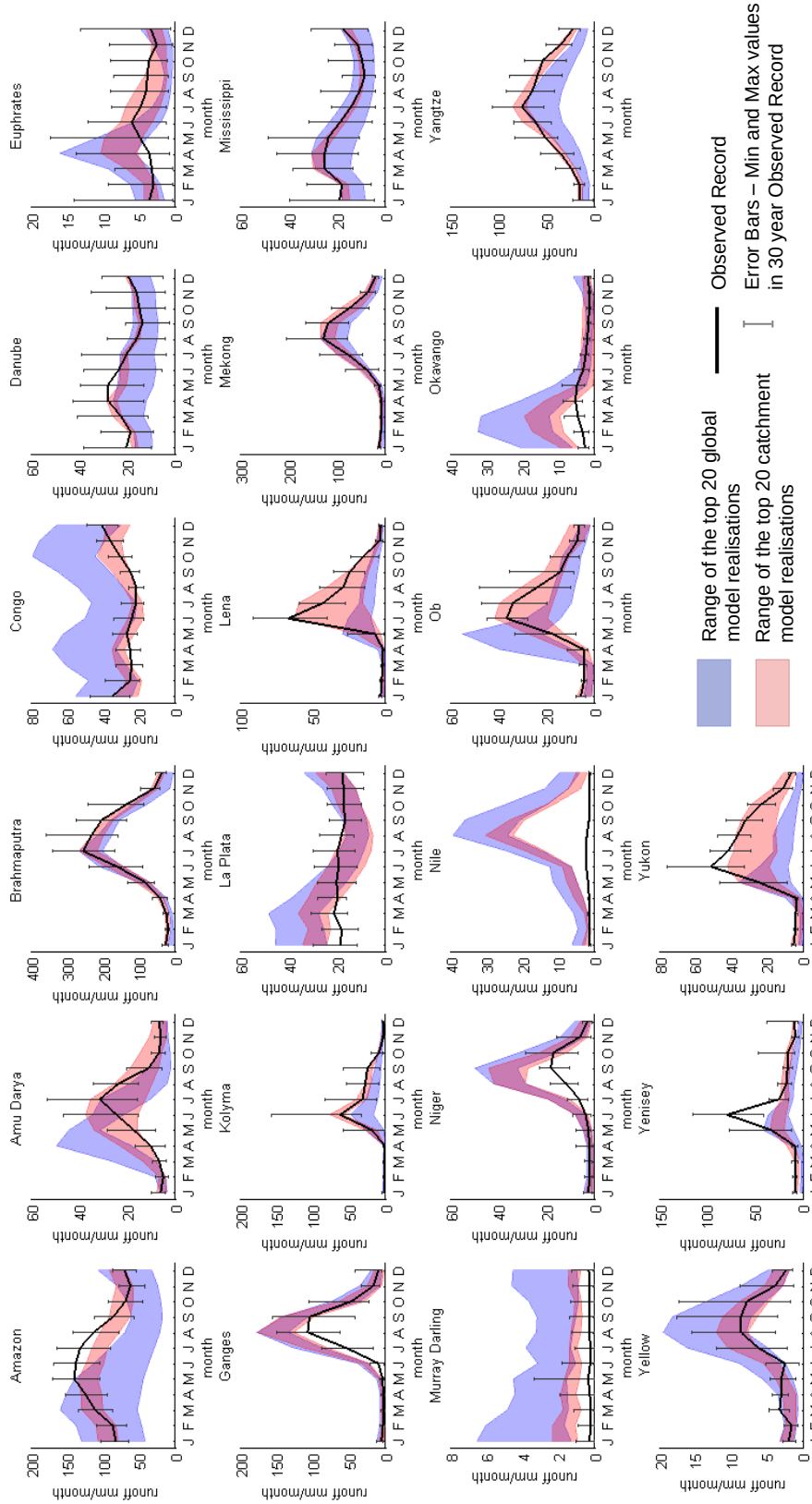


Figure 5.4 Ranges of the top 20 globally-calibrated and catchment-calibrated realisations with respect to natural variability in the observed record (30 year min and max values)

The only remaining simulations that lie outside of the range of the minimum and maximum observations are the Niger in August and September, the Okavango in February and March, and the Yenisey peak runoff in June. The eastern Arctic catchments, the Kolyma, Lena and Yenisey all show very similar observational records, with a large peak in the runoff in June, and particularly high observational ranges for that month. The range in observational values varies, and Figure 5.5 shows the mean of the minimum-maximum ranges relative to the mean value of each month for each catchment. This graph shows that the Murray Darling catchment has by far the largest observational range, of 513%, and that the Nile has the lowest at 23%. The Euphrates, Kolyma, Ganges, Niger and Mississippi have ranges greater than 150% whilst the Amazon, Brahmaputra, Congo, La Plata, Mekong, Yangtze and Yukon all have ranges less than 100%. The Brahmaputra, which is the best simulated catchment, has a relatively small observational range of 81%.

It is apparent from these results that the model performs significantly better when calibrated to individual catchments than when calibrated as a global model, particularly when natural variability is taken into account, so the next questions are *"why?"*, *"what is it about the parameter values for these realisations that make them better in each catchment?"* and *"can any similarities be found among the catchments with similar hydrological regimes?"* In order to investigate this, the parameter values of the "good" models for each catchment needed to be examined, and compared to those of the globally "good" models. Firstly however, the globally "good" model parameters are compared to the parameter values of the original model calibration.

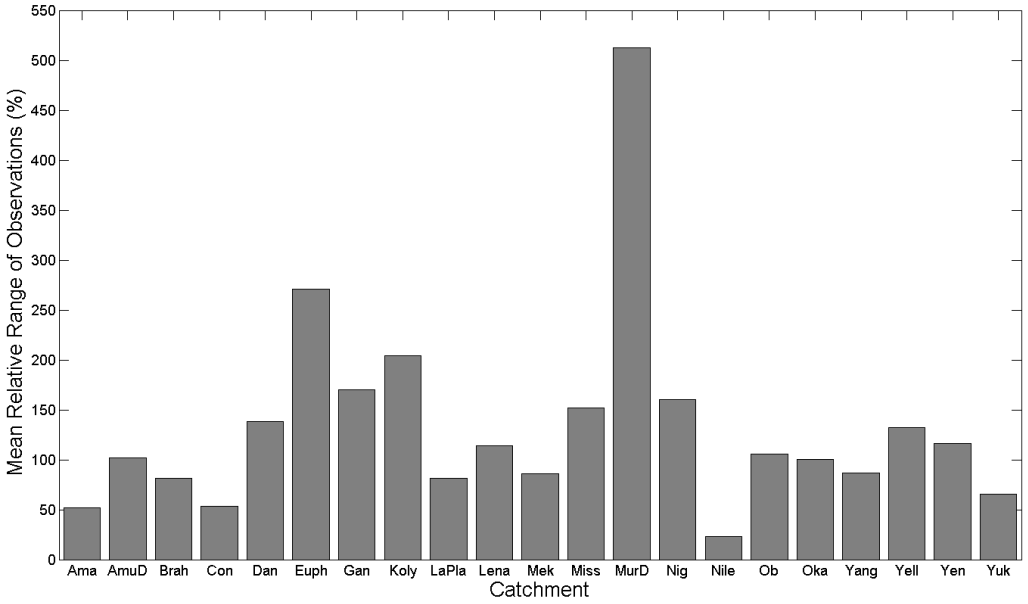


Figure 5.5 Mean relative range of observations for each catchment

5.3 Parameter Estimation of a Global Mac-PDM.14

The top 20 globally-calibrated model realisations were extracted, and the ranges of values for each parameter across the 20 realisations were calculated. This was then standardised against the minimum and maximum values of each parameter for the entire ensemble (100,000 models) to facilitate comparison between parameters (see Figure 5.6).

Figure 5.6 shows some interesting results as many of the bars do not encompass the original calibration value, and very few are centred on the original calibration value. The width of the bars, to some extent, indicate a confidence in the parameter value, as bars with a narrow range suggest that the parameter must be within the tightly constrained range to produce a good model. However, those with wide ranges show that the value of the parameter is uncertain, and may be insensitive or may depend on other parameter values. This behaviour is a sign of equifinality in the model. The b parameter for example shows a certain deviation away from the original model calibration, as the top 20 models are associated with a tightly constrained range much less than the original calibration.

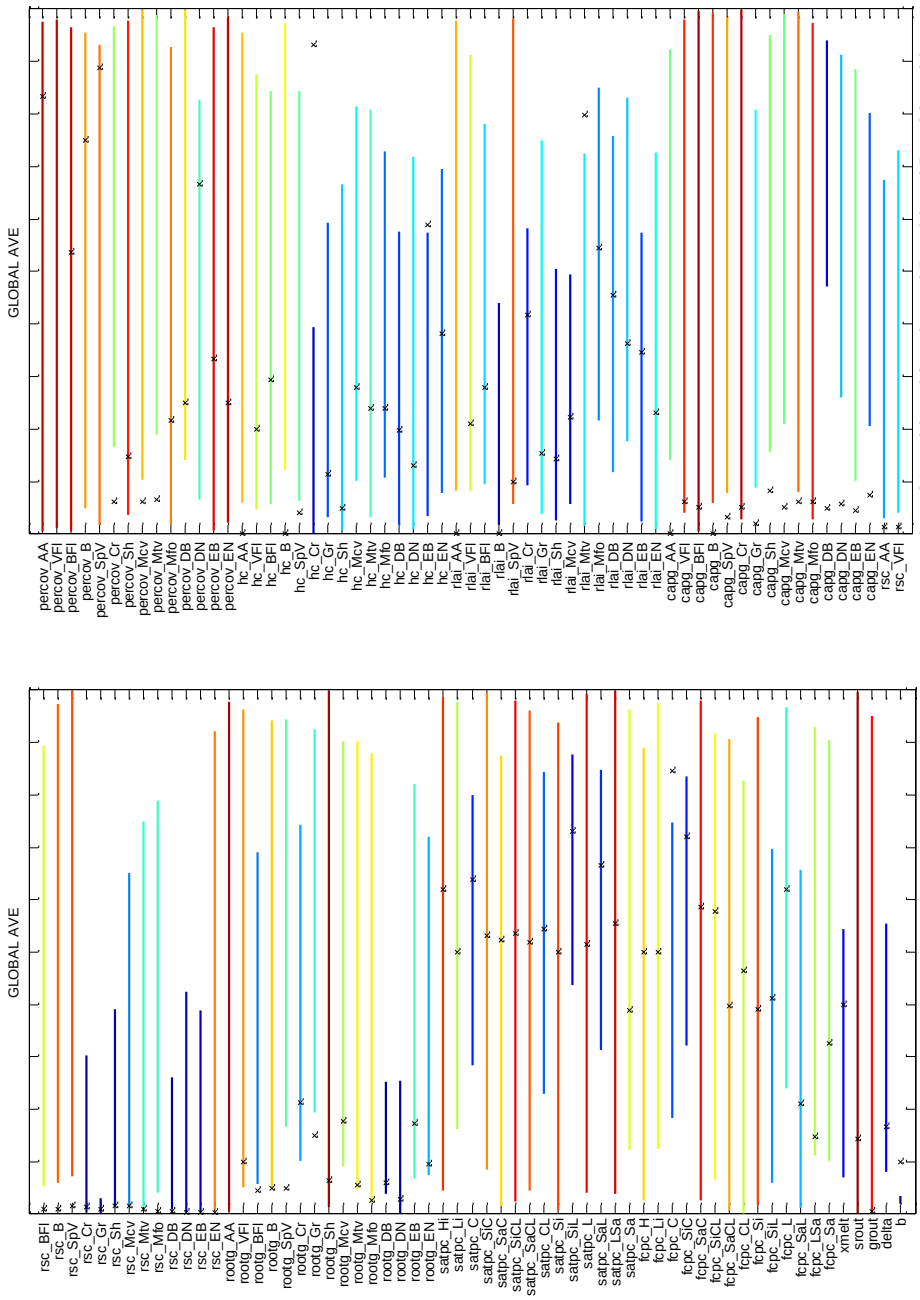


Figure 5.6 Range of top 20 global model realisation parameter values, standardised on a 0-1 scale using the range of sampled parameter space. *x indicates original model calibration value, colours qualitatively indicate the width of the range, but are not set to a scale*

The *fcpc* and *satpc* parameters show ranges that cover the original model calibration well, except for *fcpc_C* (clay) which indicates an original model calibration that was too high. The *rootg* parameters were sampled to values much higher than the original model calibration, and most vegetation types covered a large proportion of the range sampled, however the deciduous needle-leaf (*DN*) and deciduous broadleaf (*DB*) parameters were more closely constrained to the lower end of the sampled space. The grassland (*GR*), sparse vegetation (*SpV*) and broadleaf forest regularly flooded (*BFl*) ranges all lie beyond that of the original model calibration values.

The *rsc* parameter ranges lie mostly above the original model calibration for all vegetation types, except grassland, which is well constrained. The *capg* parameter ranges also lie above the original calibration values. There is a notable stepped progression towards higher parameter values across the forest parameter values (evergreen needle-leaf, evergreen broadleaf, deciduous needle-leaf and deciduous broadleaf), which suggests that deciduous canopies intercept more precipitation in the top performing models than in the original calibration, and more than the evergreen canopies. The *rlai* and *hc* parameters have relatively narrow ranges for the majority of the vegetation types, when compared to other parameters. The original model calibration values of *rlai* fit well with the ranges produced by the parameter ensemble, and a stepped progression towards higher parameter values can be seen again for the forested vegetation types. The original model calibration values for the *hc* parameters also mostly lie within the ranges of the parameter ensemble, except for Evergreen Broadleaf (*EB*) and Cropland (*Cr*) which were previously overestimated, and Bare (*B*) and Artificial Areas (*AA*) which were underestimated. The *percov* parameter values for the top 20 global models all show wide ranges, which suggests that this remains an uncertain parameter.

5.4 Catchment Specific Parameter Estimation

The top catchment-calibrated 20 model realisations for each catchment (which were used to produce the graphs in Figure 5.4) were analysed to find the ranges of parameter values for each catchment, which were then plotted in floating bar graphs. All 21 catchments were included in this study, as the parameter values of the best models of the poorly performing Murray Darling and Nile catchments may provide some information as to the reasoning behind the model performance. This resulted in 122 graphs, so only a selection that showed the most interesting results are presented in Figure 5.7. For example, those that showed significant deviation from the original model calibration value, or those that showed distinct variation between catchments. In these graphs, deviation from the dashed line shows that the ensemble produced parameter values for 'good' models that were distinctly different from the original model calibration value. The crosses show the mean of the parameter values for the top 20 models, so obvious variation in the crosses between catchments may suggest that different catchments require different values to produce a good model.

5.4.1 Control File Parameters

The highest variations between catchments is in the control file parameters: b (soil moisture capacity variability), $delta$ (an interception parameter), $grout$ (groundwater routing parameter), $srout$ (surface water routing parameter), and $xmelt$ (snow melt rate). For the majority of catchments, the b parameter value for the top 20 models was lower than that of the original model calibration, and this is to be expected from the strong response that could be seen in the global average model parameters shown in Figure 5.6. However, for the Euphrates, Kolyma, Yenisey and Yukon the mean value of b was higher than the original model calibration value. The bars for these catchments are particularly wide, which suggests that although the average is higher, there is a large amount of uncertainty in the b parameter values. The

Danube, Murray Darling, Niger, Nile and Okavango show particularly low values of *b* and have tightly constrained ranges. The *delta* parameter graph shows that most of the catchments closely agree with the original model calibration value, however, the most northerly catchments, the Kolyma, Lena, and Yukon display a trend to lower values of *delta*. The Ob and the Yenisey catchments are also northerly, and show a lower *delta* value than other catchments, but it is a less significant trend.

The *grout* parameter shows significant variation between catchments and the ranges of values among the top 20 model realisations is inconsistent. It is also noticeable that the original calibration value is very low (value of 1), and the parameter space sampled includes a maximum of 210. The Kolyma and the Lena show very low values of *grout*, and have narrow ranges. The Danube and Ob also show low *grout* values, but with less certainty. The Ganges, Mekong, Okavango and Yangtze show very uncertain parameter ranges, which span the entirety of the sampled space. The *srout* parameter graph is highly variable across catchments, with the Amazon and the Ganges showing the lowest values with narrow ranges. The Danube shows a high value of *srout*. The lack of overlap between the Danube and the Ganges and Amazon catchment values suggests that this is a parameter that should be considered for catchment specific calibration, as a globally defined value may not be sufficient.

The *xmelt* parameter graph shows a strong response from the Kolyma and the Lena catchments, as well as the Ob. The remaining catchments lie with mean values just below the original model calibration value, except the Mississippi whose mean lies just above the original value. This suggests that snow melt rate is slower in the more northerly catchments, which is to be expected, and so demonstrates that this parameter has appropriate physical meaning within the model. Results among the remaining soil and vegetation parameters were less

conclusive, but some deviations in parameter values in certain catchments could be seen for some soil and land cover types.

5.4.2 Soil Parameters

The *fcpc* (field capacity) parameter values for many of the soil types were quite variable across catchments, but most were centred about the original calibration value. The *fcpc* values for sand show all catchments conforming to the original calibration value, except for the Okavango catchment in which the mean value of the top 20 models was 10% higher. The results show loamy sand should be given a higher parameter value across all catchments and that clay should have a much lower parameter value for all catchments. Similar results could be seen for the *satpc* (saturation capacity) parameter, whereby sand was overestimated in the original model calibration. Again, many of the graphs showed variation across catchments, with lithosols giving the most significant differences, showing a particularly low value for the Amu Darya catchment.

5.4.3 Rootg Parameter

The *rootg* (root depth) parameter graphs show fairly consistent results across catchments, and all land cover classifications show a preference for higher values of the parameter than the original model calibration values. The grassland graph is shown in Figure 5.7 as this shows an interesting result for the Okavango catchment; a particularly high parameter value with a range that does not overlap with the low values of the Yangtze.

5.4.4 Rsc Parameter

The graphs for the *rsc* parameter showed that there was high agreement between catchments on the parameter value. However, for eight of the land cover types, the parameter value was much higher than that of the original model calibration value, whilst the remaining

seven land cover types showed values that were still a little higher, but were in the lower values of the parameter space sampled. This is indicated by the difference in the grassland and sparse vegetation graphs in Figure 5.7.

5.4.5 Capg Parameter

The *capg* parameter, which is the second interception parameter in Mac-PDM.14, showed strong variation in parameter values between catchments for many of the vegetation types. The strongest differentiation between catchments could be seen for the mosaic: cropland/vegetation parameter type (as shown in Figure 5.7), where the Euphrates gives the highest value and the Danube the lowest. The La Plata, Nile, Yellow and Okavango rivers also show high parameter values. However, unlike the *xmelt* parameter, and the *rootg* grassland parameter, the *capg* ranges overlap for all catchments, suggesting that although optimum solutions may benefit from single catchment calibration, it is not as essential for this parameter as for *xmelt*. The evergreen broadleaf vegetation type for *capg* also showed catchment differentiation, with the Congo showing a higher parameter value than the other catchments.

5.4.6 Rlai Parameter

The catchments show good agreement on parameter values for the *rlai* parameters, for which the mean of the top 20 models lies not far above or below the original model calibration value. The only exception is for bare and artificial areas, where the original model calibration value was 0, and the model realisations give values of ~1.5 and ~3 respectively.

5.4.7 Hc Parameter

The *hc* (vegetation height) parameter graphs also showed good agreement between catchments, with the bare and the grassland land cover types showing the most variability. The cropland vegetation type

showed an optimum parameter value of ~1m, and none of the catchments ranges reached the original model calibration value of 14.9m. This original value is far too high to be physically meaningful, and a value approximating 1 seems more appropriate.

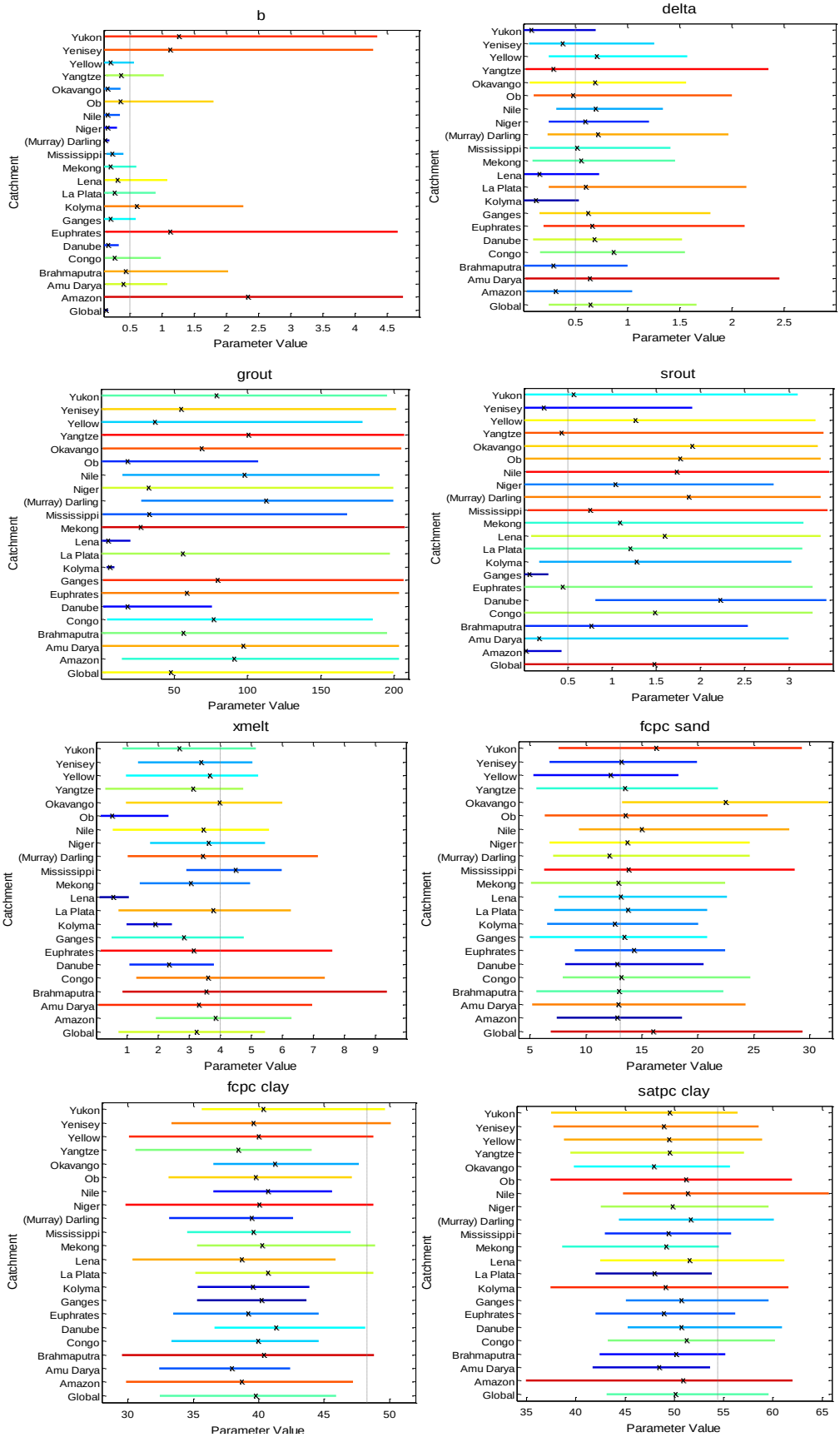
5.4.8 Percov Parameter

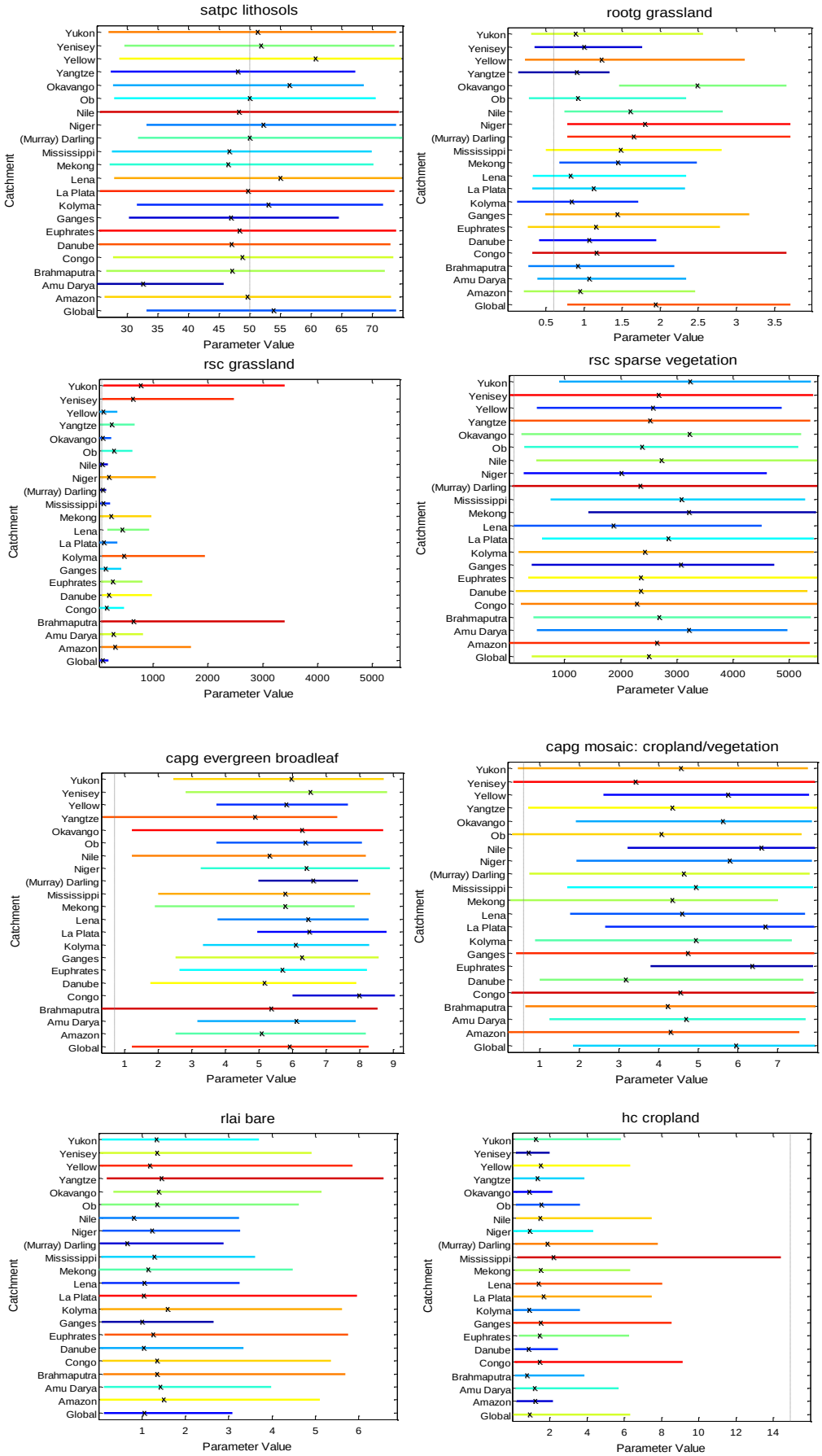
The *percov* (percent cover) graphs exhibited some variation between catchments, but not significantly; all range bars overlapped. The Congo catchment showed a slightly higher value of *percov* than other catchments for the evergreen broadleaf land cover classification. The mosaic: cropland/vegetation graphs showed the greatest variability (and is shown in Figure 5.7, with the Danube giving a very low mean parameter value (27.3%), and the Nile, Euphrates and La Plata giving the highest values (76.7, 75.9 and 74.5% respectively).

5.4.9 Summary of Catchment Specific Parameters

In summary, the control file parameters showed the greatest variations between catchments, particularly the routing parameters and the snow melt rate parameter. The soil parameters showed agreement across catchments, but sometimes provided a different value to the original model calibration value. The root depth, vegetation height, leaf area index and stomatal resistance parameters mostly showed little variation between catchments, but the interception parameter *capg* showed more variable results. The catchments that most often deviated from the others were the Kolyma, Lena, Ob, Danube, and Okavango.

Figure 5.7 shows the ranges and mean values of the parameters for the top 20 model realisations for each catchment, but Figure 5.3 shows that for most catchments, many more than 20 models showed good agreement with the range of values within the 30 year observed record. Therefore, the parameter values for a larger group of model realisations were explored further, and this was done using Approximate Bayesian Computation theory.


b
delta
grout
srout
xmelt
fcpc sand
fcpc clay
satpc clay
Catchment
Parameter Value
Yukon
Yenisey
Yellow
Yangtze
Okavango
Ob
Nile
Niger
(Murray) Darling
Mississippi
Mekong
Lena
La Plata
Kolyma
Ganges
Euphrates
Danube
Congo
Brahmaputra
Amu Darya
Amazon
Global


satpc lithosols
rootg grassland
rsc grassland
rsc sparse vegetation
capg evergreen broadleaf
capg mosaic: cropland/vegetation
rlai bare
hc cropland
Catchment
Parameter Value
Yukon
Yenisey
Yellow
Yangtze
Okavango
Ob
Nile
Niger
(Murray) Darling
Mississippi
Mekong
Lena
La Plata
Kolyma
Ganges
Euphrates
Danube
Congo
Brahmaputra
Amu Darya
Amazon
Global

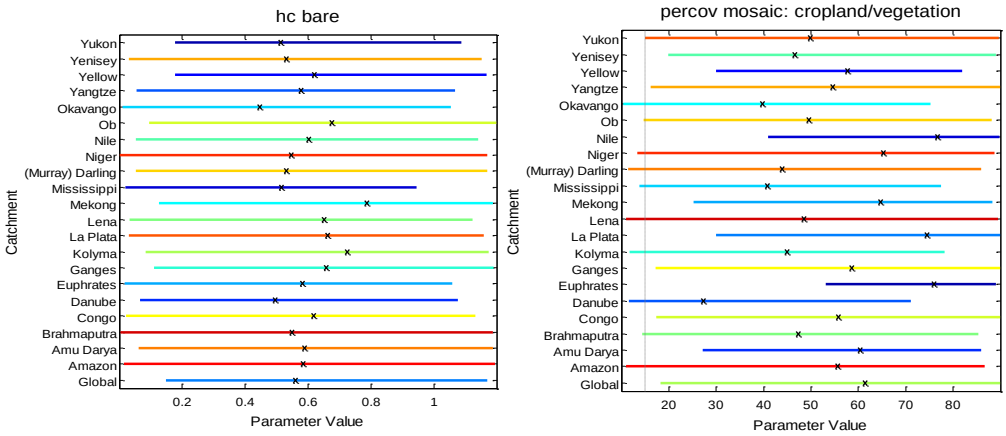


Figure 5.7 A selection of catchment parameter value plots for the top 20 catchment-calibrated model realisations for each catchment. *Dashed line indicates the original model calibration value; crosses indicate the mean of the parameter values from the top 20 realisations; colours help display the length of the line, with narrowest in blue and longest in red. Colour bar not set to a scale.*

5.5 Approximate Bayesian Rejection for Parameter Estimation

Bayes theorem is a rule for updating the prior probability of a hypothesis when additional information becomes available. Bayes' rule can be written as:

$$p\left(\theta|\tilde{Y}\right) = \frac{p(\theta)\,p(\tilde{Y}|\theta)}{p(\tilde{Y})} \qquad \text{(EQ 5.1)}$$

where $\mathrm{p}(\theta|\tilde{\mathrm{Y}})$ is the posterior parameter distribution, $\mathrm{p}(\theta)$ is the prior distribution, $\mathrm{L}(\theta|\tilde{\mathrm{Y}}) \equiv \mathrm{p}(\tilde{\mathrm{Y}}|\theta)$ denotes the likelihood function, and $\mathrm{p}(\tilde{\mathrm{Y}})$ represents the evidence (or normalisation constant) (Sadegh and Vrugt, 2013). The Approximate Bayesian Computation (ABC) methodology, or likelihood-free inference algorithms, relax the need for an explicit likelihood function, $\mathrm{L}(\theta|\tilde{\mathrm{Y}})$, and instead use summary statistics to measure the distance of each model simulation to the data (Sadegh and Vrugt, 2013). ABC seeks to determine a posterior distribution of parameter values between the observed and simulated data that have a distance smaller than an error tolerance value. The Approximate Rejection Algorithm, which is drawn from in this study, is the most basic form of an ABC algorithm and is written as:

ABC-REJ ALGORITHM

1) Draw $\theta \sim p(\theta)$
2) Simulate Y from the model $Y \sim \eta(\theta)$
3) Accept θ if $\rho\left(\tilde{Y}, Y(\theta')\right) \leq \delta$

In words, a sample point θ is taken from the prior distribution, $p(\theta)$. This is used to simulate the output of the model, $Y \sim \eta(\theta)$, which is then compared to the simulated data, Y, using the distance function $\rho(\tilde{Y}, Y(\theta'))$. If this distance function is smaller than a tolerance value, δ, then the simulation is close enough to the observations and is accepted as being in the posterior distribution, $p(\theta | \rho\, \tilde{Y}, Y(\theta')) \leq \delta)$, (Vrugt and Sadegh, 2013). Accepted values of θ are not from the true posterior distribution, but rather from an approximation to it (Wilkinson, 2013). The choice of distance function is a subjective decision, like it is in GLUE, and it needs to be carefully considered to reduce the loss of information. Common examples from genetic applications of ABC include Canberra, Euclidean and Manhattan distance, alongside the hydrological indicators outlined in Table 5.1, such as Nash-Sutcliffe Efficiency and Root Mean Square Error.

Sadegh and Vrugt (2013) discuss two immediate similarities between GLUE and ABC: 1) that the distance function is similar to the informal likelihood measure used in GLUE to differentiate between behavioural and non-behavioural models, and 2) that the sampler used in the ABC-REJ algorithm to sample from the prior distribution is similar to Latin Hypercube sampling, which is commonly employed in GLUE. Since there are many similarities in the way the modelling experiment is run, the 100,000 realisation model ensemble that was run as a GLUE experiment could be used to explore the parameter results using simple Bayesian theory.

The top 1,000 of the model realisations were extracted for each catchment, and the distributions of the parameter values among those top models were plotted and compared to the distribution given to the 100,000 member ensemble as parameter input sample distributions. This provided a means of identifying if any new information was gained by comparing the prior (the 100,000 member Latin Hypercube Sampled parameter distribution) to the posterior distribution (the top 1000 models realisation parameter values). This provides some analytical advantages over the method used to produce the graphs in Figure 5.7: distributions show whether the mean value of the top performing models is a true optima, or whether it is just the mid value of an insensitive parameter; and the comparison with the prior distribution indicates whether the mean value, or the range of values could have been influenced by the allocation of a strong prior distribution. However, the comparison between catchments is less easy to distinguish as the distributions must be plotted on separate graphs in order to be seen clearly.

The results of this investigation confirm, but extend, the findings of the top 20 model realisation parameter ranges and means. Again, the control file parameters showed distinct results, and so all catchment results are displayed in Figure 5.8 to Figure 5.12. The b parameter showed consistent goodness at low parameter values, though some catchments gave lower values than others; for example the Murray Darling catchment showed much lower values than the Amazon. The b parameter simulates the variability of soil moisture content across a catchment, so it is understandable that the very dry Murray Darling catchment provides an exceptionally low value.

The *delta* parameter also shows a diversification of preferential values between catchments; with the Congo showing high peak values (approx. 1.6), and several other catchments showing peaks at near zero values, such as the Kolyma and Yukon catchments. The *delta*

parameter is an interception parameter that works alongside the *capg* parameter in the interception equation by Calder (1990). The higher the value of *delta,* the more precipitation that is intercepted by the vegetation. Many catchments, such as the Congo, La Plata and Niger showed higher values than the prior, implying that for most catchments, the prior was inaccurate. However, the arctic catchments of the Kolyma and Yukon returned low values, which could be expected due to the fact these are the two catchments with the lowest precipitation rates of the 21 study catchments. The tropical catchments of the Brahmaputra and the Amazon also show low values, which perhaps could be explained by much higher precipitation rates, which push the *delta* value towards lower values to produce a reasonable overall interception amount.

The *grout* and *srout* parameters exhibit a strong response in some catchments and none in others. This could not be seen clearly from the range plots in Figure 5.7. The only catchments that did not exhibit the same trend between *grout* and *srout* were the Danube and the Kolyma, which showed trends towards lower values for the *grout* parameter but showed less or no real trend for *srout*. This suggests that quick flow routing is more dominant in these catchments than slow flow routing. The distinct trend towards low values in many catchments for the *grout* and the *srout* parameters, when compared to the insensitivity of the results for the globally calibrated parameter values, demonstrate the need for consideration of catchment model calibration. However, this study does not indicate whether a low parameter value for all catchments would suffice. The seeming insensitivity of many of the catchments to parameter values would need to be investigated further. This is not the case for the *delta* parameter though, which shows distinctly different parameter values between catchments, with a level of certainty, particularly as the graph peaks deviate from the sampled (prior) mean value.

The *xmelt* parameter, shown in Figure 5.12, show that the majority of catchments maintain the peak value of the prior at 3.5, however several catchments, the Amu Darya, the Ob, the Lena, and to some extent the Ganges, show a value near 0. The Ob and the Lena would be expected to show lower (slower) melt rate values, as they are northerly catchments, it is surprising though that the Kolyma, Yenisey and Yukon do not also give low *xmelt* values.

In the semi-arid Amu Darya catchment, snowmelt contributes 69% of its mean annual flow, predominantly from seasonal snowmelt (Savoskul and Smakhtin, 2013). Whilst the Mac-PDM model does not have a physically-based numerical representation of seasonal snowmelt or glacial meltwater specifically, snowmelt may still play an important role in the simulation of this catchment. The Ganges catchment, which also showed a peak at the prior value, but showed some trend towards lower values, also has a snowmelt contribution. In the Ganges this contribution is less distinct than in the Amu Darya catchment, where snowmelt accounts for approximately 7% of the mean annual flow (Savoskul and Smakhtin, 2013).

Figure 5.13 shows a selection of additional distribution graphs to demonstrate the more distinct trends in some of the other parameters. In each of these cases, some catchments showed a deviation away from the prior distribution, and different responses could be seen between catchments. In most cases, not all of the catchments deviated from the prior, as shown in the *fcpc* sand parameter, where only the Okavango showed a higher peak than the other catchments. In some parameters where a uniform prior distribution was used, such as the *capg* mosaic:cropland/vegetation parameter, some catchments showed no response, while others showed an obvious trend. It may be that the catchments that showed no response are insensitive to perturbations in these parameters.

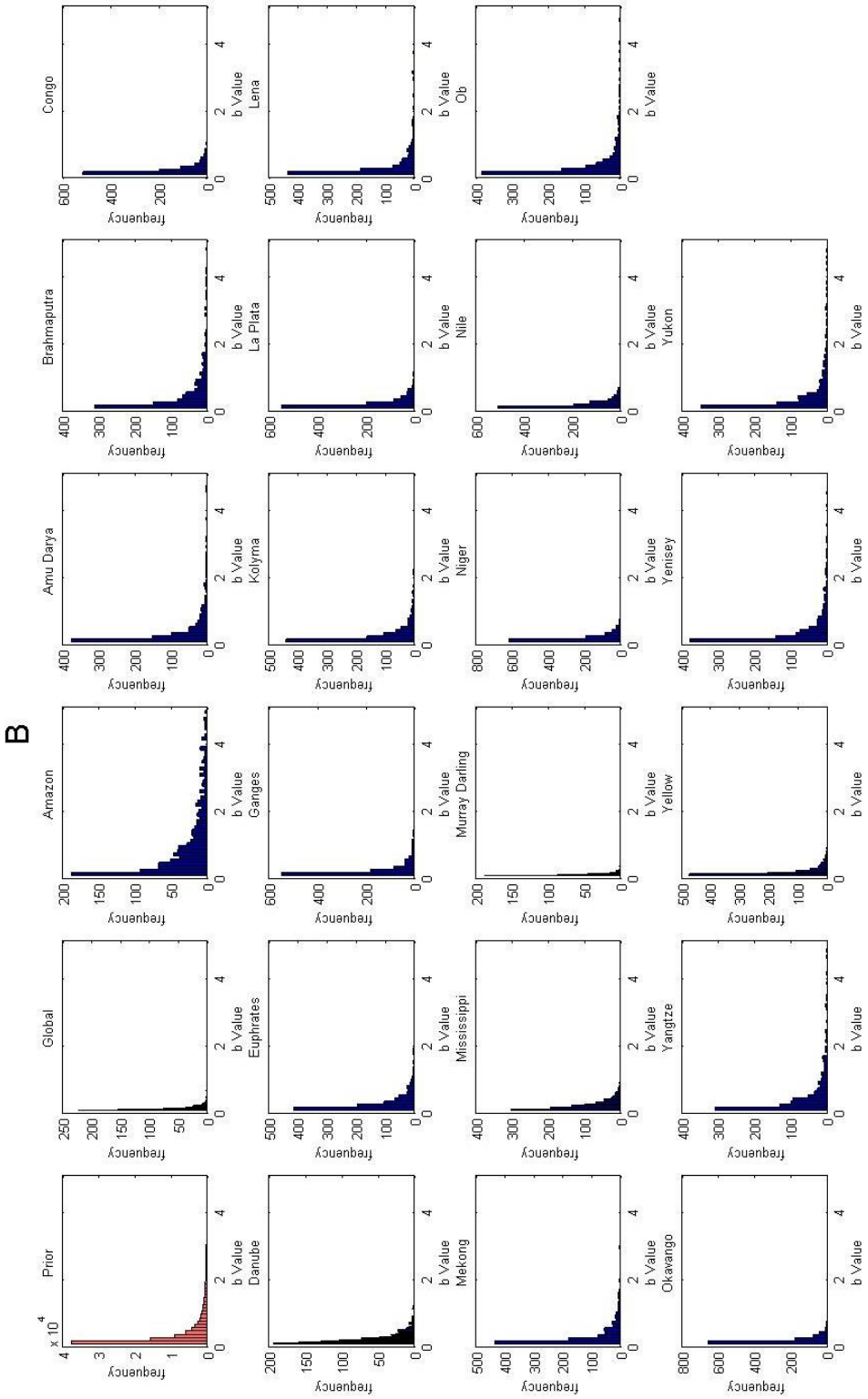


Figure 5.8 Prior and posterior distributions for the *b* model parameter

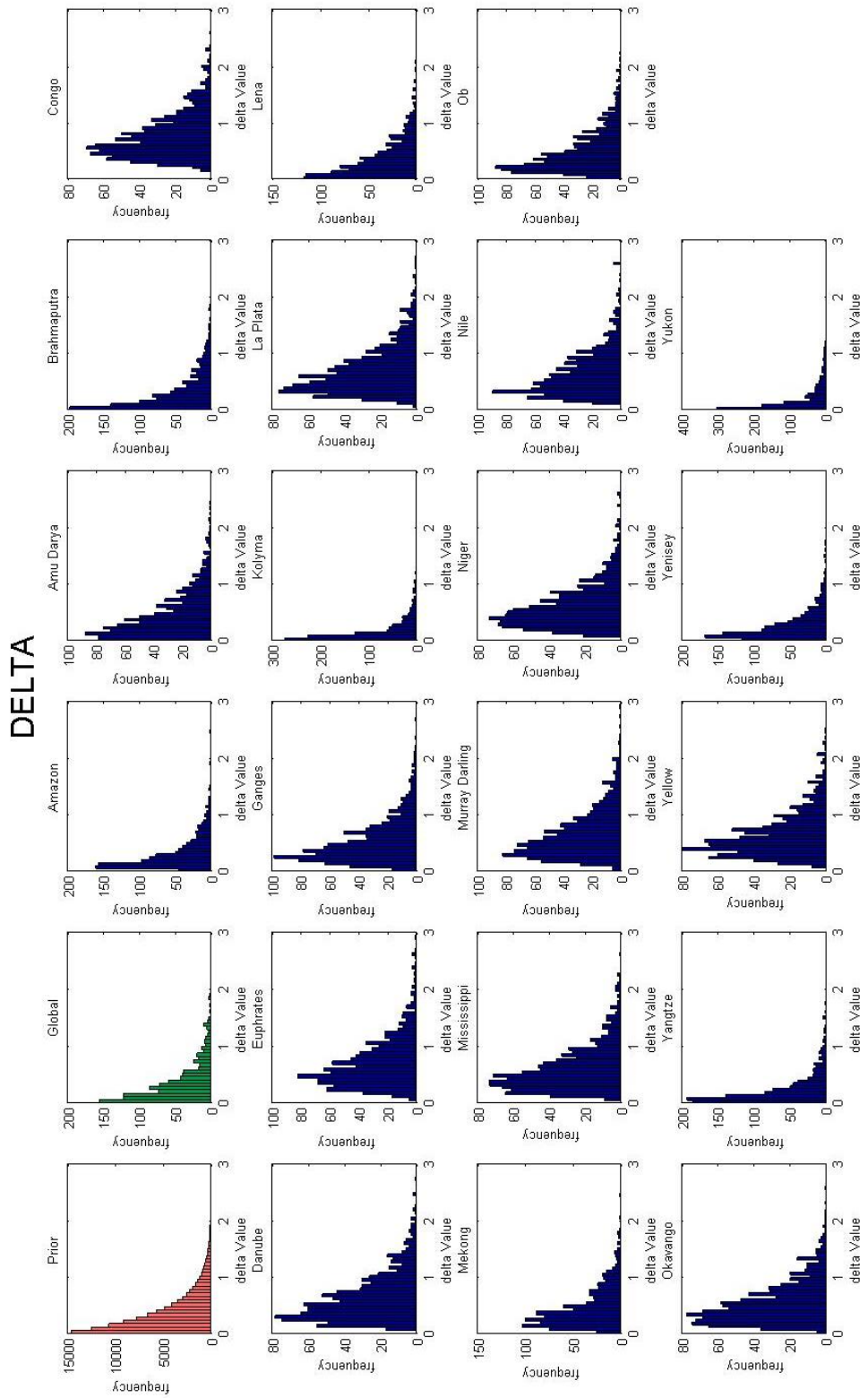


Figure 5.9 Prior and posterior distributions for the *delta* model parameter

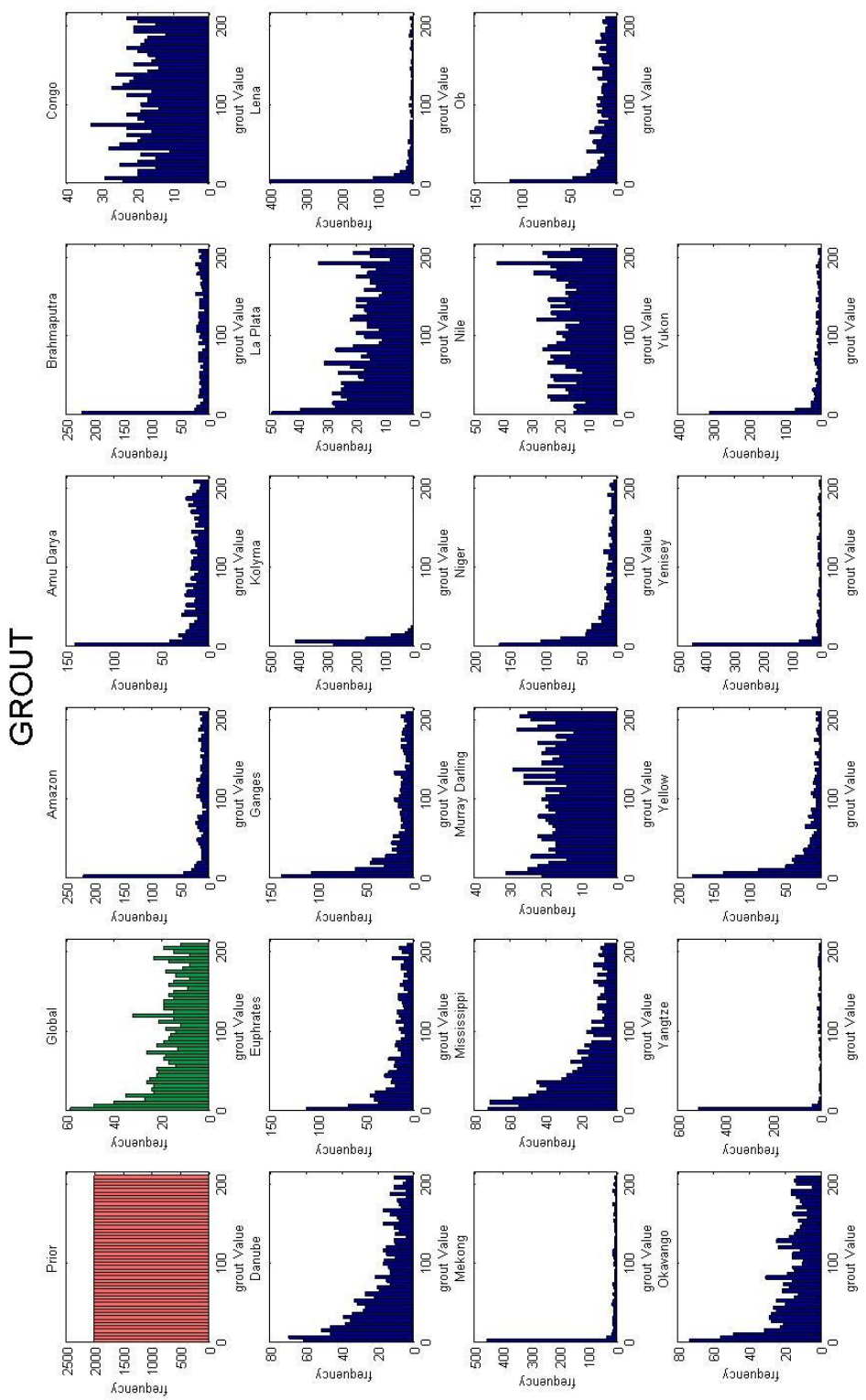


Figure 5.10 Prior and posterior distributions for the *grout* model parameter

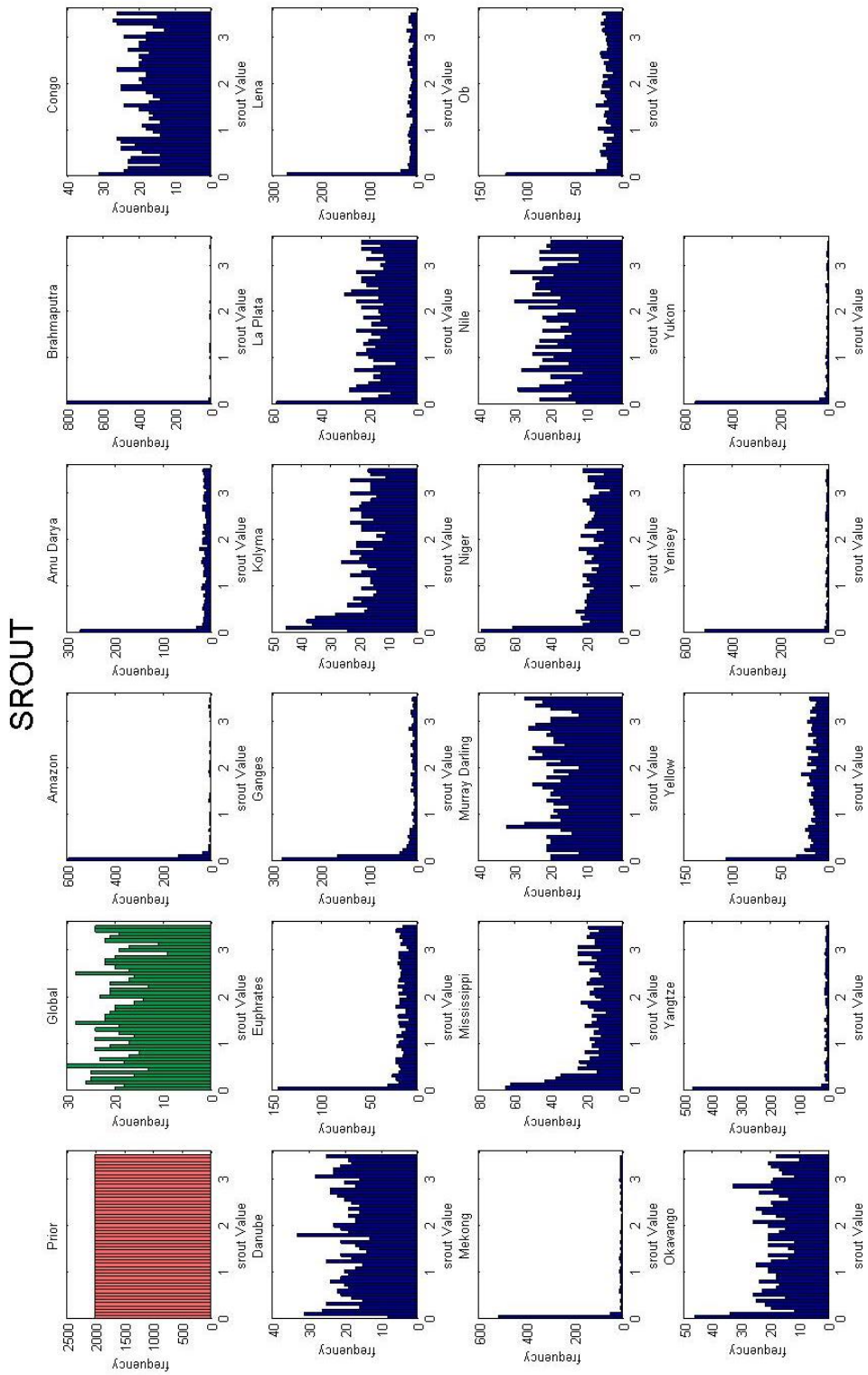


Figure 5.11 Prior and posterior distributions for the *srout* model parameter

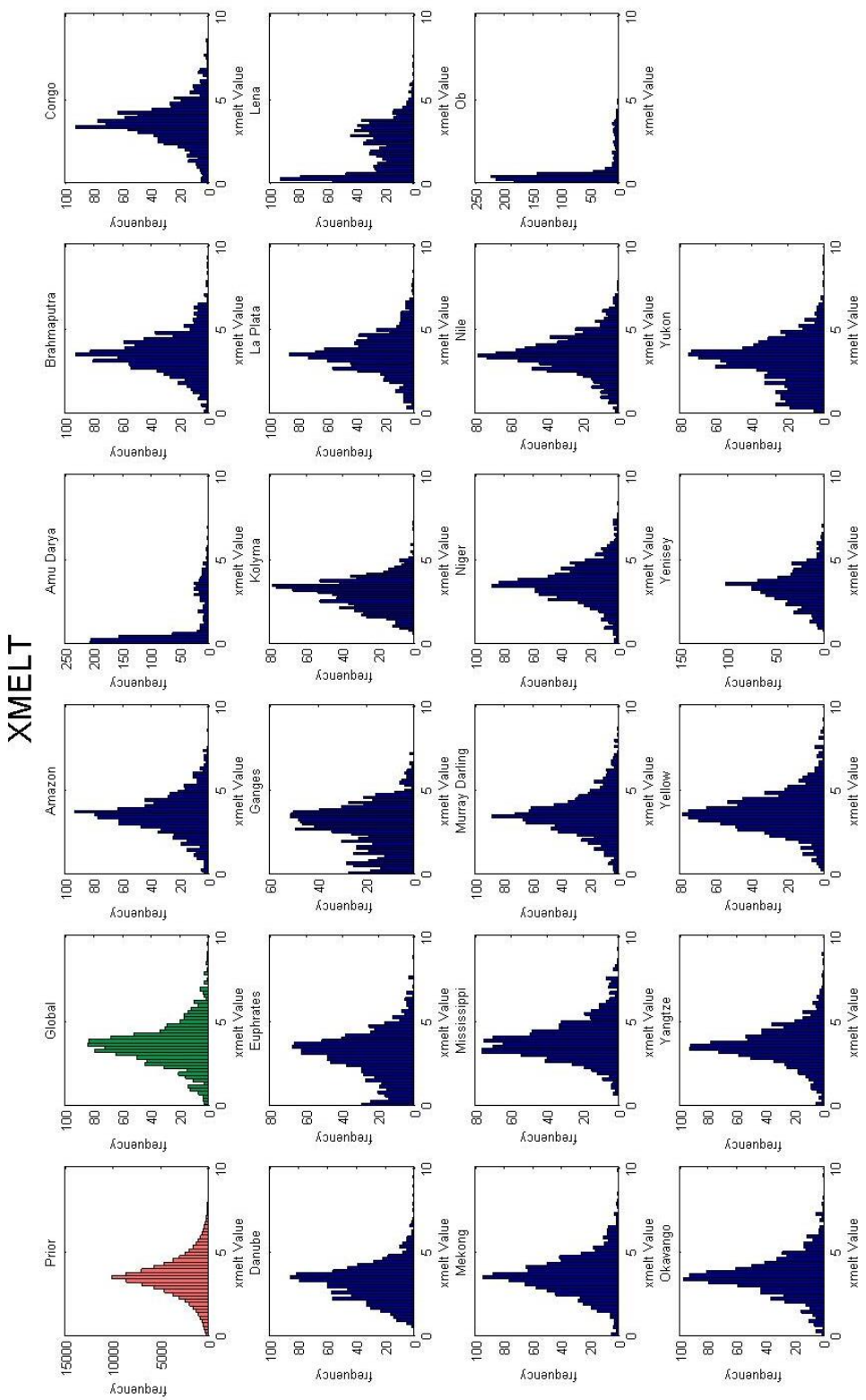


Figure 5.12 Prior and posterior distributions for the *xmelt* model parameter

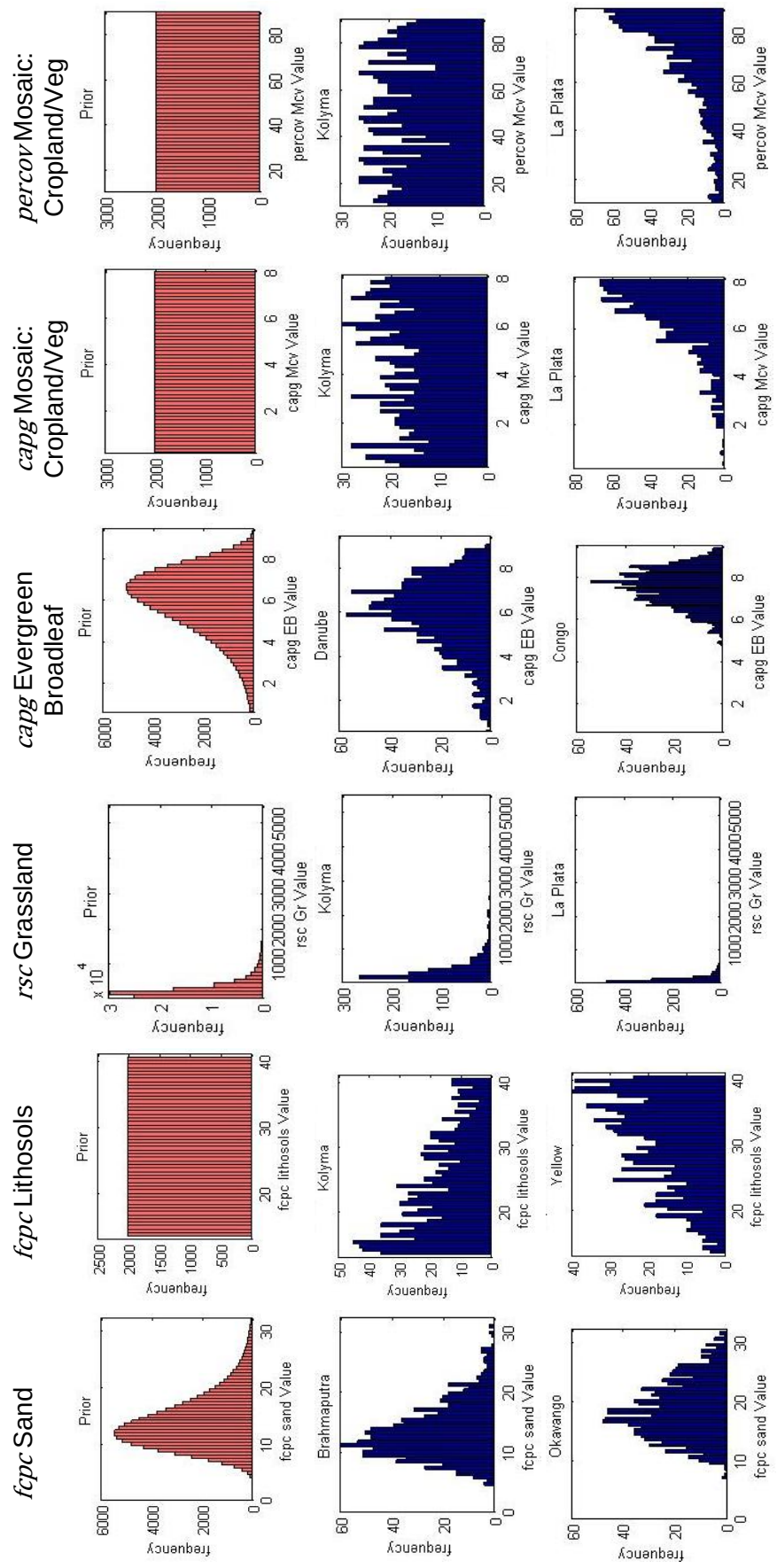


Figure 5.13 Prior and posterior distributions of select other parameters to demonstrate catchment variation

Table 5.2 summarises the trends seen in all distribution plots and compares them to the trends found in the range plots, in order to ascertain whether the conclusions from the previous plots were accurate. Soil and vegetation types that are not given in this table showed results that agreed with the prior distributions, and showed no differentiation in parameter values between catchments.

The results indicate that whilst many of the posterior distributions agreed with the prior distributions, it is not a straightforward process defining parameter values for each catchment. The results in section 5.2 demonstrate that sets of good parameter values have been identified for each catchment, and that these perform far better than globally-calibrated model realisations. However, through parameter estimation, it is difficult to ascertain what exact values could be used for each parameter. Whilst the posterior distributions of the top 1,000 model realisations show trends towards certain values, the ranges of parameter values in the top 20 models are often broad. This suggests that these parameter sets demonstrate equifinal behaviour (Beven, 2012). It would not be possible at this stage to pick the optimum parameter values individually for each parameter and compile them to create a set of parameters, but there are sets of specific parameter values that perform well.

The tendency of the posterior parameter values to trend towards the prior distribution suggests insensitivity of these parameters. Additionally, it suggests that regardless of the value of that parameter, the model will produce a good calibration, and therefore the posterior distributions of the good models are the same as that of the prior distribution. Alternatively, it could just suggest that the prior distribution was correct. Several potentially insensitive parameters and equifinality issues indicate that the Mac-PDM model is likely over-parameterised.

Table 5.2 Summary table of trends shown in distribution plots and their agreement with the previous representation as range plots

Parameter	Sub-Class	Trends in Distribution Graphs	Agreement with Range Plots
b	-	**Murray Darling** shows low peak, **Amazon** higher	Agrees, suggests sensitivity
delta	-	Many catchments give higher value than prior, especially **Congo, Euphrates** and **Yellow. Kolyma, Yangtze, Yenisey** and **Yukon** peak lower than prior	Agrees, suggests sensitivity influenced by catchment precipitation amounts
grout	-	Many catchments show strong trend to low values	Demonstrates sensitivity, less apparent on range plots
srout	-	Many catchments show strong trend to low values	Demonstrates sensitivity, less apparent on range plots
xmelt	-	**Amu Darya, Lena** and **Ob** show strong low value trend	Agrees, suggests sensitivity particularly in snowmelt driven catchments
fcpc	sand	**Okavango** shows higher peak value than other catchments	Agrees with range plot. Peak of distribution lower (at ~19%) than mean of range plot at 22.52%
	lithosols	**Kolyma** trends to low value, **Yellow** to high values, others unresponsive	Agrees with range plot, some sensitivity across catchments
satpc	sandy clay loam	**Murray Darling** shows slight trend toward higher values	Agrees with range plot
	lithosols	**Kolyma** and **Yellow** show preference to high parameter values	**Yellow** shows high value in range plots, **Kolyma** indistinct. Ranges show false low for **Amu Darya** - possible equifinality
rootg	grassland	**Okavango, Niger, Mississippi** and **Murray Darling** higher than prior, **Yellow, Yangtze** and **Kolyma** lower than others	Agrees with range plot
	bare	**Euphrates** slight trend to higher values	Range plot could be read with false confidence, as means agree (~1) but distributions are uniform – equifinal behaviour
rsc	grassland	**Brahmaputra, Kolyma, Yenisey** and **Yukon** high (match prior distribution), **Okavango, Murray Darling, Mississippi** and **Nile** show lower values	Agrees with range plot, sensitive parameter

Parameter	Sub-Class	Trends in Distribution Graphs	Agreement with Range Plots
capg	evergreen broadleaf	**Congo** shows strong response to high values	Agrees with range plot
	mosaic: cropland/vegetation	Strong trends to high parameter values from the **Congo, Euphrates, La Plata, Niger, Nile** and **Yellow** catchments. Gentler trend for **Mississippi**, and **Okavango. Danube** shows slight low value trend	All agree except **Congo**, which does not show a notably high value on range plot. Sensitive only in some catchments
	shrubland	**Murray Darling, Niger** and **Nile** trend to upper values	Range plots less distinct. Possible equifinality
	cropland	**Ganges, La Plata** and **Yellow** trend to high values	Range plot shows varied response. Possible equifinality
	broadleaf regularly or permanently inundated	**Congo** trends to high values	Agrees with range plot. Sensitive only in this catchment
rlai	evergreen broadleaf	**Congo** trends to low values	Agrees with range plot
	mosaic: cropland/vegetation	**La Plata** shows trend towards low values	Agrees with range plot
	grassland	**Congo, Danube, Mississippi, Okavango** slightly lower than prior, **Ob, Kolyma, Yenisey** and **Yukon** higher values	Not evident in range plot. Potential sensitivity by climatic regime
percov	evergreen broadleaf	**Congo** trends to higher values	Agrees with range plot
	deciduous broadleaf	**Congo** trends to higher values	Congo high on range plot but not exceptionally, **Yellow** shows a narrower range high value, but is uniform in distribution plots
	mosaic: cropland/vegetation	**Congo, Euphrates, La Plata, Niger, Nile and Yellow** catchments all trend to high values, **Danube** slight trend to low values – as with *capg* mosaic: cropland/vegetation	Mostly agree, but **Yellow** not notably high on range plot. Strong sensitivity in some catchments
	shrubland	**Niger** and **Nile** show trend to high values, **Murray Darling** and **Congo** show slight trend to high values	Agrees with range plot (though not strong trends on range plot)
	cropland	**Ganges** and **Yellow** show strong trend to high values, **La Plata** shows gentler trend	Agrees with range plot. Sensitive in some catchments

More analysis would need to be carried out to determine whether the parameters are indeed insensitive or not, and to examine the interactions between parameters. This analysis was considered by way of model emulation, but it was deemed beyond the scope of this PhD study. Model emulation is considered further in the Discussion Chapter (Chapter 7).

The parameter values of the top performing models for each catchment, including the global best model parameter values from the GLUE experiment are given in the Appendix of this thesis. These are the parameterisations that were used for the line plots in Figure 5.2.

5.6 Mac-PDM.14 in a Catchment Modelling Context

With the retention of the results of the 100,000 model parameter ensemble, alongside gauged discharge records, Mac-PDM.14 could be calibrated for any global catchment with a small amount of analysis. This analysis would certainly be less time consuming than calibrating a catchment model each time. The question then is: *"can a catchment calibration of a global model perform as well as a catchment model?"* It would not be appropriate to compare the goodness-of-fit of the Mac-PDM.14 model calibrations with catchment models without using the same input data. Therefore, in order to answer this question, catchment models would need to be acquired, calibrated, and run with either the WATCH or the ISI-MIP input data, which was beyond the scope of this PhD study.

Whilst there have been several inter-comparisons of hydrology models within their respective scales (e.g. Refsgaard and Knudsen, 1996, Smith et al., 2004, Haddeland et al., 2011, Warszawski et al., 2014), the first comparison of a global scale model with catchment models was in 2011 (Gosling et al., 2011). In this study, Mac-PDM.09 was compared to six individual catchment models, however the authors compared the results of a global calibration of Mac-PDM.09 with the catchment

models under climate change scenarios. Thompson et al. (2013) extended this analysis by comparing Mac-PDM.09 with the catchment models MIKE-SHE and SLURP for several locations in the Mekong catchment and in this case again projections of the models under climate change scenarios were compared. Piniewski et al. (2014) compared the results of the catchment scale model SWAT, and the global scale model WaterGAP, on projections of environmental flow indicators for the Narew basin in Poland. This study also focussed on climate change projections, and so the performance of global models has not yet been compared to catchment models over historic periods. Specific catchment calibrations of global models have also never before been considered.

An obvious advantage of catchment scale hydrology models over global models is the finer resolution at which they operate. Global hydrology models most often run at a resolution of 0.5x0.5 degrees, however some can be run at finer resolutions, for example GWAVA can be run at 0.1 x 0.1 degrees (Dumont et al., 2010). Catchment models, on the other hand, can operate at finer resolutions, from tens of meters, such as the 30m SWAT model (Chaubey et al., 2005), to as fine as 2m, which has been used for applications of TOPMODEL (Lane et al., 2004). The main reason for the coarser resolution of global hydrology models is the availability of climate input data. Therefore, the application of catchment models on a global scale, would likely suffer the same issue, as regional climate model data is not available for all global catchments. On the other hand, global hydrology models are often used to simulate catchment hydrology (Loos et al., 2009, Moors et al., 2011, Ogata et al., 2012, Siderius et al., 2013, Green et al., 2014), and specific catchment calibrations could be used to dramatically improve the performance of the models in these circumstances.

Calibrations of GHMs have until now used sets of catchments to carry out the calibration and then developed a globally averaged

parameterisation. For example, Mac-PDM.09 was calibrated using a suite of 50 worldwide catchments (Gosling and Arnell, 2011) and WaterGAP2 was calibrated against 724 discharge stations across the globe (Alcamo et al., 2003). This method does not recognise or address the concept that different catchments may require, or at least benefit from, different parameterisations. When catchment models are employed, they are recalibrated to the catchment in question each time, whilst global models lump all catchments together with globally uniform parameter values. The results from this chapter suggest that parameter values differentiated by catchment could provide much better model results than globally averaged parameter values. Whilst this would be difficult to employ for global scale studies, it would be relatively straightforward when global hydrology models are employed for catchment scale research studies.

5.7 Summary

This chapter has investigated the parameter values of Mac-PDM.14 in detail. During this study it became apparent that seeking alternate parameterisations of the model for each study catchment yielded significantly better results than applying a global set of parameter values. The months of peak runoff that were underestimated by the globally calibrated models were better simulated by catchment calibrated models. Similarly, the timing of peak flows that were simulated several months too early by the global models were improved with catchment calibration. The natural variability of the observed record was considered, as the performance of the model may have been misrepresented in some catchments if variability was high. This was found to be the case for the Murray Darling catchment; where the ranges of the top 20 models both catchment-calibrated and globally-calibrated did not cover the observed mean values, but the simulated runoff was within the bounds of the variability of the 30 year runoff record. The Nile however, for which the model performs particularly

poorly, has a low variability within the observed record, so this analysis showed that the model results are still poor for the Nile catchment. When catchment calibrations of Mac-PDM.14 are considered alongside the natural variability of the catchment, the model is shown to output very good results, with only 5 of the total 240 months (12 months over 20 catchments, excluding the Nile) lying outside of the ranges of observed monthly runoff values.

The parameter values of the catchment model calibrations were investigated to see if there were any significant differences between catchments that would suggest a global calibration would be insufficient. The ranges of values of the top 20 model realisations for each catchment were investigated. The control file parameters (*b*, *delta*, *grout*, *srout* and *xmelt*) showed the most variation between catchments. The soil parameters, *fcpc* and *satpc*, showed agreement across parameters, but a deviation from the original calibrated value. The interception parameter *capg* also showed some variation between catchments. Approximate Bayesian Rejection was then used to assess the parameter values further, using distributions of the top 1,000 model realisations (top 1%) for each catchment as an acceptance limit. Distributions showed more detail than the range plots, and the control files again showed the most interesting results, as well as the *capg* parameters. Only a few catchments for a few parameters showed distinguishable parameter values for different catchments. Little consistency with climatic regime, latitude or easily distinguishably hydrologic characteristics could be identified to explain this behaviour.

These results suggest that the model may be over-parameterised as many parameters seemed insensitive to perturbations. The sensitivity of parameters would need to be assessed in more detail, perhaps using model emulation (e.g. Lee et al., 2011) to identify parameter interactions and investigate equifinality (Beven and Freer, 2001). The parameter values of the top globally-calibrated model, as well as the top

catchment-calibrated model, for each of the catchments are presented for future applications of Mac-PDM.14. The use of global models as catchment models has been discussed. A comparison of the performance of catchment specific calibrated Mac-PDM.14 with a catchment model for historic periods would be an innovative step forward from this research.

The next chapter validates the global and catchment calibrations of Mac-PDM.14 with an alternative set of climate input data, and compares the performance of Mac-PDM.14 with the results of the recent ISI-MIP multi-model ensemble.

Chapter Six:

Mac-PDM.14 Model Validation

6.1 Introduction

This chapter presents the results of a validation of the Mac-PDM.14 model with an alternate climate dataset. The findings presented in Chapters 4 and 5 demonstrate that the uncertainty experiment and subsequent calibrations of the model have significantly improved the model's performance when tested against observed data. However, due to the nature of the calibration process, it is expected to perform better as the model is 'trained' to the observed record. Therefore, it is necessary to validate the model with a different set of climate input data, either over a different historic time period, or an alternate modelled climate data set of the same period, derived using a different method. Validation gives the new calibrations of the model a measure of credibility and, if the results are satisfactory, demonstrates that the model can be taken forward and applied using newly available input datasets. This chapter presents the validation datasets available, and compares the results of the top performing model calibrations, run with the original WATCH data, with those run with the validation dataset. The performance of the model run with the validation dataset is then compared to other models from a new multi-model ensemble project, ISI-MIP (Warszawski et al., 2014). Finally, the top 20 performing catchment individual model calibrations from the GLUE ensemble, run with the validation dataset, are reviewed.

6.2 Validation Datasets

Observational datasets that could efficiently be applied to the Mac-PDM.14 model for validation were available from two sources: the WATCH project data that was used for calibration, applied for a different time period for validation purposes; and the ISI-MIP (Inter-Sectoral Impacts Model Inter-comparison Project) data, which could be applied for any time period, to provide a test for the model. The relative merits and limitations of each dataset will be discussed briefly in turn.

The WATCH data, as described in Chapter 2 and by Weedon et al. (2010), was derived using a reanalysis of the ERA-40 data for the period 1958 to 2001. The processing procedure used bilinear interpolation of the variables from the 1 degree ERA-40 grid to the 0.5 degree CRU land sea mask. Furthermore, monthly averages were interpolated, temperatures were corrected for elevation, and the data was then bias corrected using CRU TS2 data. Since the ERA-40 reanalysis data was not available prior to 1958, in order to develop a full twentieth century dataset, the WATCH project generated weather data by reordering the ERA-40 data a year at a time using a weather generator. By this process, the statistical characteristics of the data for the years 1901-1957 were the same as 1957-2001, but the timing of particular weather events were not accurate for any particular location (Weedon et al., 2010). Therefore, using the data prior to 1957 would not be a good test for the model since the data would produce very similar annual average time series to those already employed. Accounting for the necessary 5 year 'spin up' period of Mac-PDM.14, the only additional years in the reanalysis dataset would be the 8 year period 1962-1970. This is a short time period, which again would not provide a very comprehensive validation assessment for the model. Additionally, only 15 of the 21 discharge stations chosen for the catchment calibration provide data prior to 1965.

The ISI-MIP dataset covers the period 1960-2000 and was derived using a trend preserving, statistical bias correction approach developed specifically for the project (Warszawski et al., 2014). The ISI-MIP data is based upon the WATCH forcing dataset, but has adjusted the monthly mean and daily variability of simulated climate data to observations, whilst preserving the long-term climate signal. The full details of the correction method are provided by Hempel et al.(2013). In order to determine whether the ISI-MIP data was sufficiently different from the WATCH dataset for the years 1971-2000 to provide an adequate validation test for the Mac-PDM.14 model, the mean monthly rainfall

was plotted, and is shown in Figure 6.1. Investigating the difference between specific rainfall events was considered to be unnecessary as the mean monthly outputs have been the focus of this study, and the majority of global hydrology simulation studies to date.

The graphs in Figure 6.1 show that although the overall shape of the monthly data for each catchment is broadly the same, there are some distinct differences between the WATCH and the ISI-MIP datasets. The very seasonal catchments such as the Brahmaputra, Ganges, La Plata, Lena, Niger and Yenisey rivers show the most similar rainfall data between WATCH and ISI-MIP, but the catchments with more complex annual rainfall patterns show deviations. In particular, the rainfall in January, February and March in the Amazon is much higher when estimated by the WATCH dataset, than the ISI-MIP data. Similarly June, July and August in the Ob catchment show higher rainfall in the WATCH data. The ISI-MIP data shows a high rainfall peak in the Yellow river in August, which is not present in the WATCH dataset, and the Murray Darling shows a very different record between the two. The peak flow in the Euphrates catchment is reached earlier in the ISI-MIP dataset, with the peak in February, as opposed to the peak in April for the WATCH data. Neither dataset seems to show consistently higher rainfall than the other, with several overlaps apparent in most catchments. These graphs suggest that the datasets show the similarities that would be expected of two datasets covering a long period of time for the same catchments, whilst showing enough differences to be able to test the models behaviour using different input data.

6.3 Ensemble Performance with Validation Data

The 1,238 model realisations from the GLUE experiment that resulted in a mean absolute relative error (MARE) less than 0.75 were run with the ISI-MIP data, and the model outputs were again assessed using MARE.

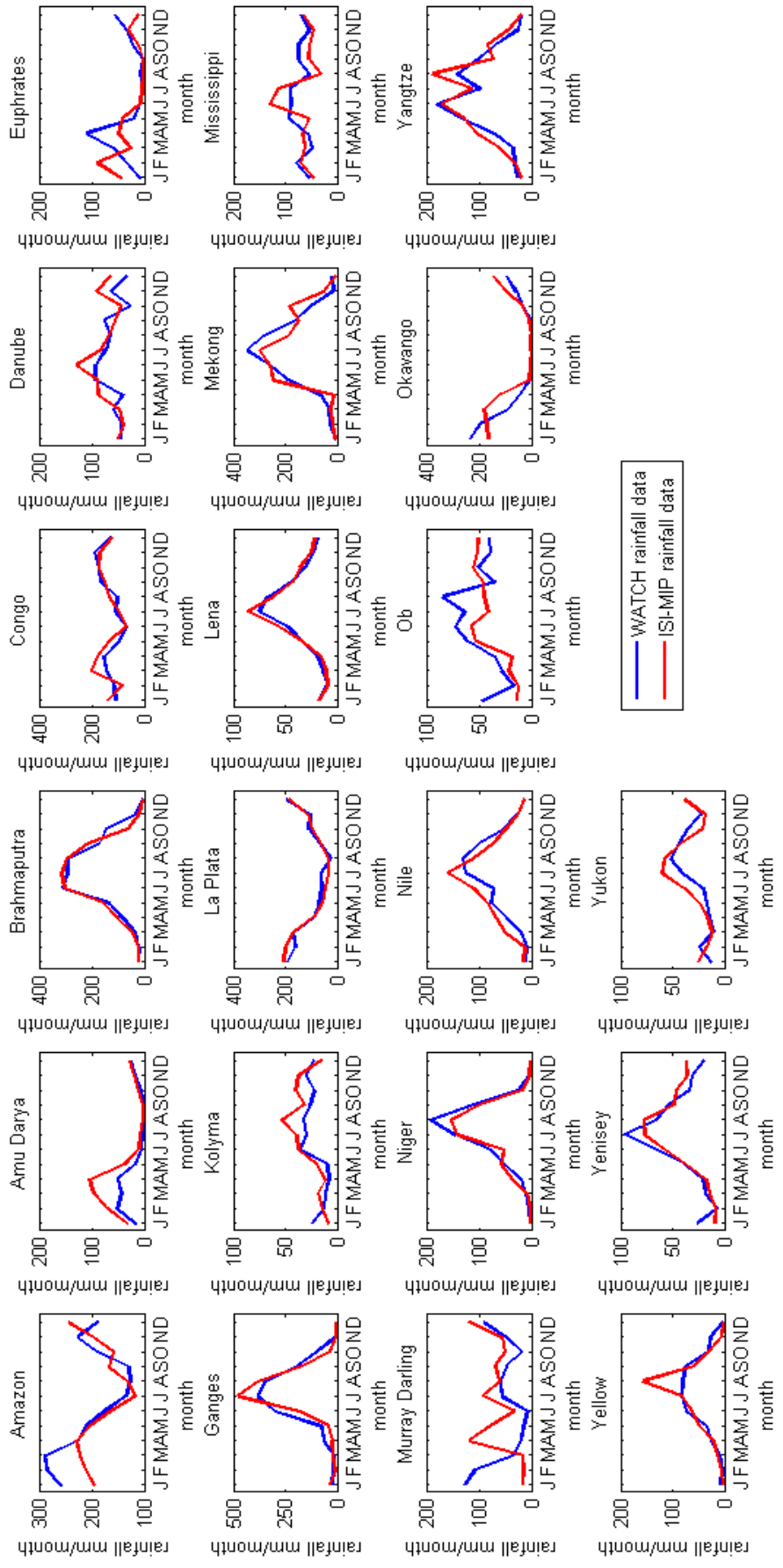


Figure 6.1 Comparison of the average monthly rainfall calculated from the daily WATCH and ISI-MIP input data sets, 1971-2000.

A set of three 'best' models were then plotted, as shown in Figure 6.2. As before, the Nile and the Murray Darling have been excluded due to their poor model performance. These top three models were: the original top model realisation from the GLUE experiment run with the WATCH input data; this same model realisation run with the ISI-MIP input data; and the top ranking (lowest MARE scored) model realisation out of the 1,238 GLUE model realisations run with the ISI-MIP data. This best ISI-MIP model realisation happened to be the 7th best from the GLUE experiment with the WATCH data. The MARE of the original top model realisation (of the 100,000 runs) had a value of 0.47 when run with the WATCH data. With the ISI-MIP data, this model realisation had a MARE of 0.45; so the model actually performed better with the ISI-MIP data. Of the 1,238, model run with the ISI-MIP data, the best MARE value (from the 7th best GLUE model as previously mentioned) was 0.44. This shows that overall, despite having calibrated the model to the WATCH data, the performance of the model is certainly satisfactory when applied using an alternative input dataset.

The results in Figure 6.2 show that, despite the difference in rainfall input, the overall shape of the runoff output was similar between the WATCH and the ISI-MIP input datasets. This can be expected, as most global scale catchments are complex systems with storage and vegetation feedbacks with the atmosphere, so runoff may not directly reflect rainfall patterns, particularly when considering a 30 year average. There are differences however, and there are many catchments where the best of the 1,238 models driven with the ISI-MIP data shows a significant betterment over the WATCH model; for example in the Amazon, where the runoff is less of an underestimation than with the WATCH dataset. Similarly, the Danube shows runoff levels much nearer the observed record. In a few cases, such as the Yellow river, the WATCH data provided an overestimation of the observed record, and the ISI-MIP data has heightened that

overestimation. The January to March flows of the La Plata are an additional example of this.

The best GLUE ensemble model, when run with the ISI-MIP data is presented in these graphs, and shows the results if the newly calibrated model parameter sets were to be applied with a different dataset, without further calibration, as is common practice with many GHMs employed by the hydrological modelling community. It is quite unlikely that when passed to a new model user, they would be able to run 1,238 model parameterisations to find a new optimum model calibration with the new input dataset. However, as that was possible in this case, it was carried out in order to determine how well the model could perform with the ISI-MIP data. Therefore, the yellow lines on the graphs in Figure 6.2 would be what we might expect a user to produce in his or her own application of Mac-PDM.14 (and not the turquoise lines, which would require significant resources to define), so this is a robust test of whether the newly calibrated Mac-PDM.14 is 'valid'.

Figure 6.3 shows the MARE values for each catchment for each of the 3 top models shown in Figure 6.2. We would expect the model to perform slightly worse than the WATCH optimum model with the ISI-MIP data, since the model has not been calibrated with this dataset. This is the case in several of the catchments, especially the Amazon and the Congo, which is apparent in Figure 6.2, but there are in fact many catchments where the model performs better with the ISI-MIP data than it did with the WATCH data. The underestimation of the Danube catchment is less pronounced with the ISI-MIP data, and the overestimations of the high peaks in the Niger and the Okavango runoff are also reduced. These results suggest that the input data of the ISI-MIP project is driving more realistic simulations of runoff than the WATCH dataset in these catchments.

The Brahmaputra stands out in Figure 6.3 as being worse with the ISI-MIP data, whilst the Ganges shows a distinct improvement. Neither of

these results are evident in Figure 6.2, perhaps due to the fact these are the two catchments with the highest flow magnitude. Figure 6.3 shows that of the 19 catchments considered, 11 showed a decrease in model performance with the ISI-MIP data, while 8 showed an increase. When applying the best of the 1,238 ISI-MIP driven model realisations instead of the top GLUE model realisation, 10 of the catchments showed a better MARE value than the WATCH driven model. Whilst the average MARE value across all catchments was better for both ISI-MIP driven model realisations included in these graphs, Figure 6.3 shows that achieving a good global model calibration is a trade-off between catchments. Section 6.5 of this chapter gives the results of the validation of the catchment-calibrated model realisations taken from Chapter 5 of this thesis.

The uncertainty in the simulations that used the WATCH and the ISI-MIP datasets were explored by plotting the ranges in the simulations from the 1,238 model realisations. These were the same 1,238 model realisations run with both the WATCH data, as from the original GLUE experiment, and run with the ISI-MIP data for comparison and validation. Figure 6.4 displays the results. Again, where we might have expected to see the range of outputs from the ISI-MIP data to be wider than those of the WATCH data, we see that the results using this new dataset are actually significantly better. For all catchments, the upper limit of the range of outputs is lower when using the ISI-MIP data, than the WATCH data. There are only a few instances where the observed record is not encompassed by any of the top 1,238 model realisations, and only in the Yukon does this cover more months for the ISI-MIP data than for the WATCH data.

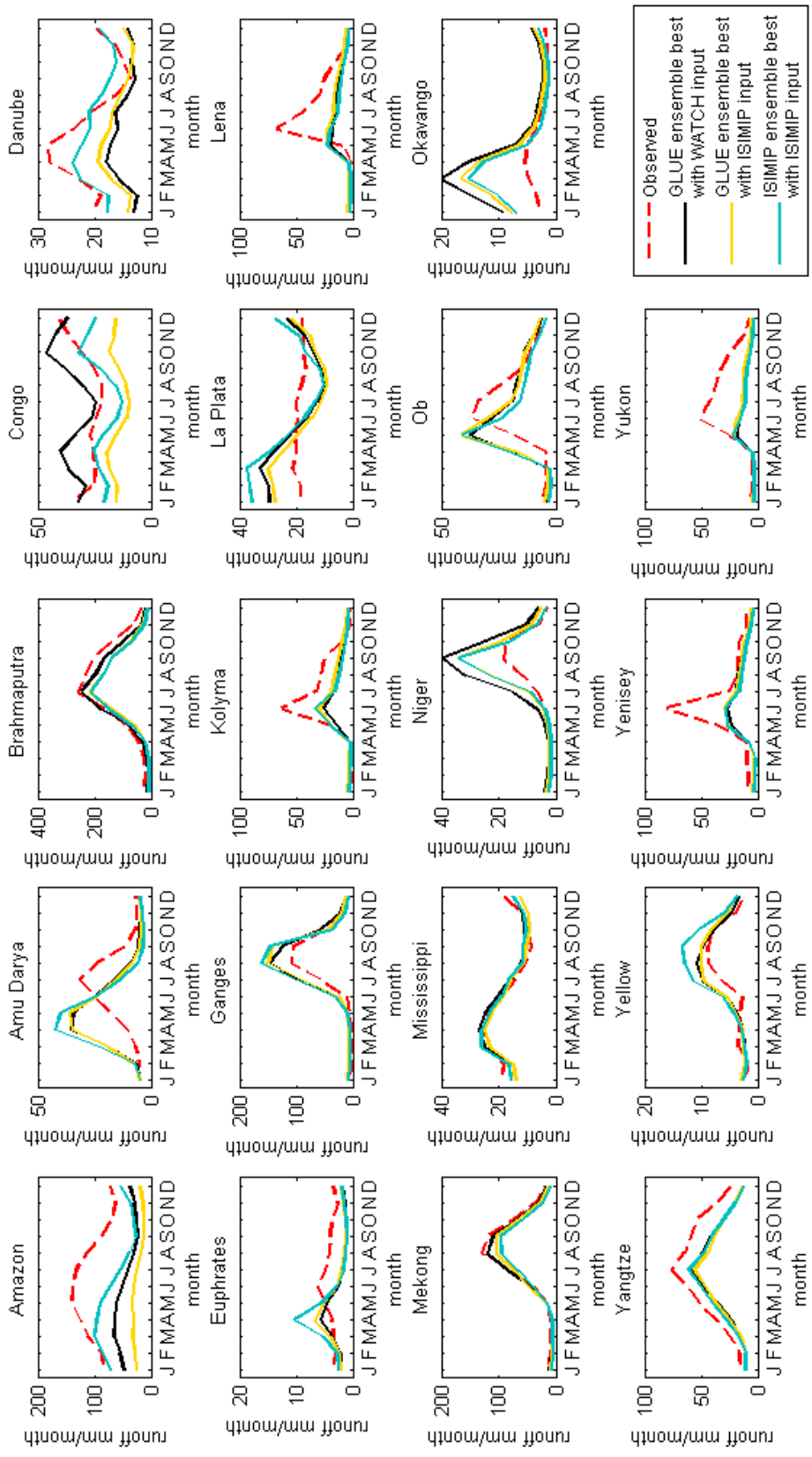


Figure 6.2 Catchment hydrographs of the top performing GLUE model realisation with WATCH and ISI-MIP input data, alongside the ISI-MIP top performing model realisation

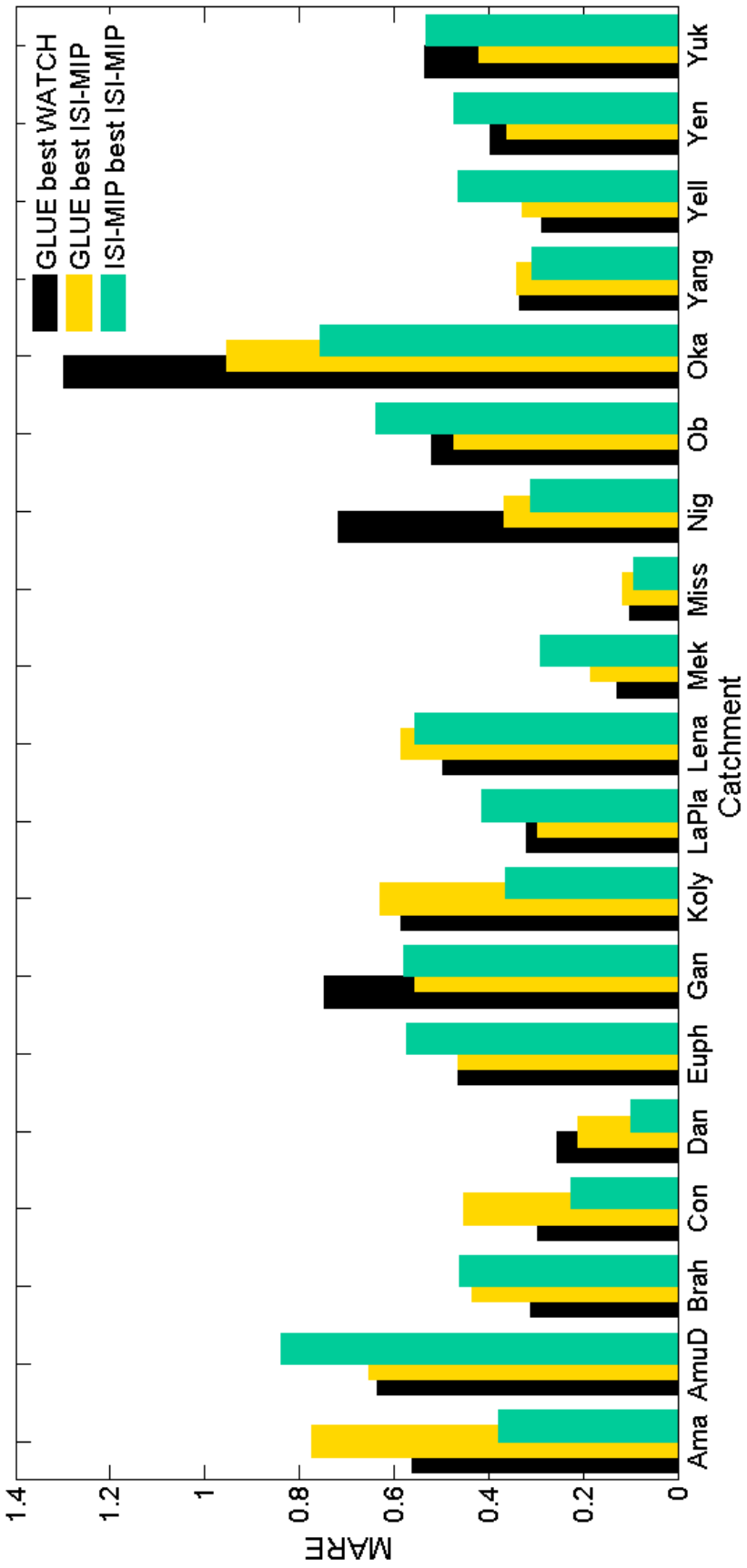


Figure 6.3 Catchment MARE values for the top performing model realisation from the GLUE ensemble (of 100,000 models) run with both WATCH data and ISI-MIP data, alongside the top performing model realisation from the ISI-MIP ensemble (of 1,238 models) run with ISI-MIP data.

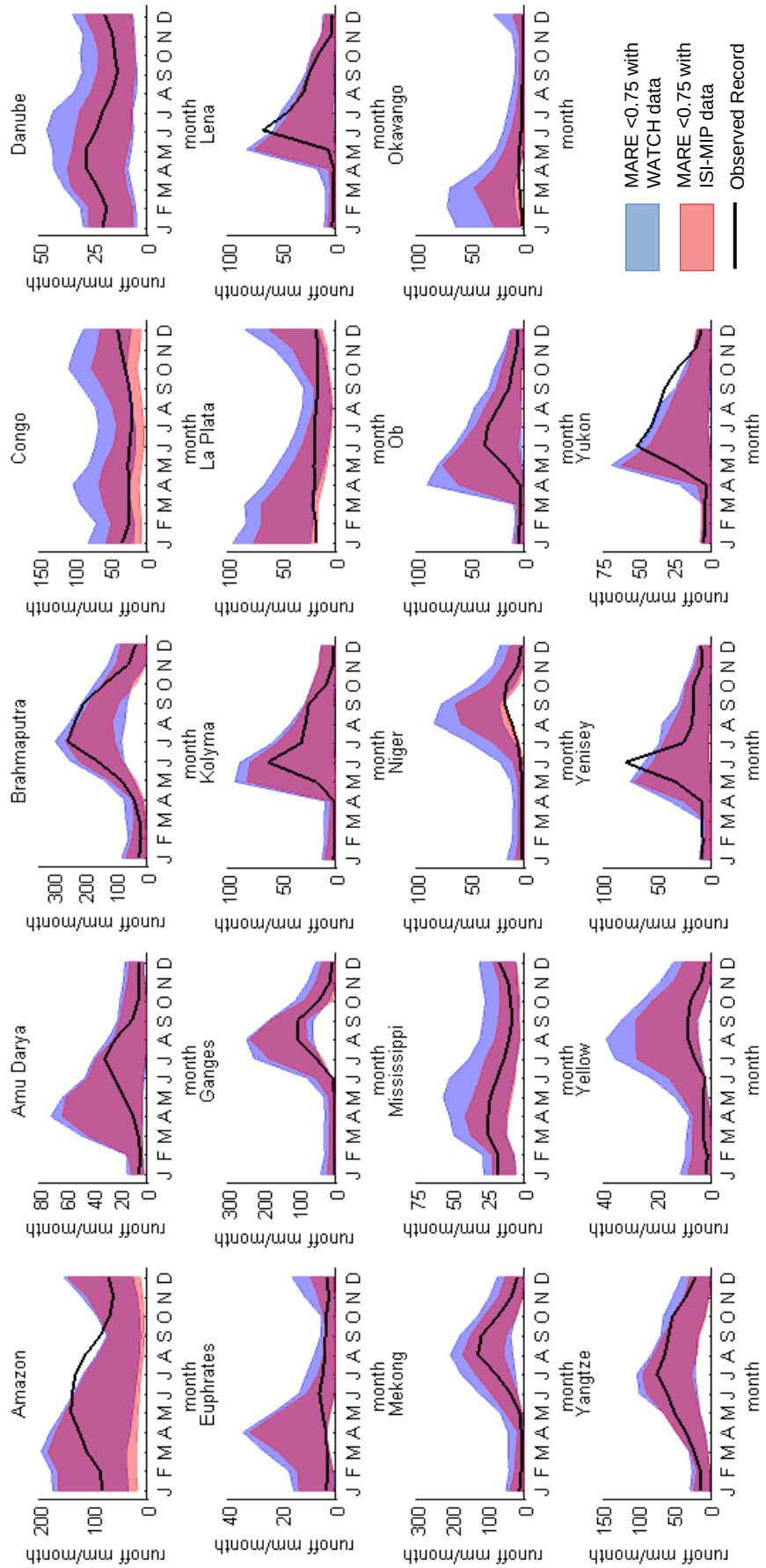


Figure 6.4 Ranges of outputs from GLUE ensemble models with MARE <0.75, run with WATCH input data and ISI-MIP input data

The uncertainty ranges of the Danube, Mekong, Mississippi, Ob and Yangtze are all significantly reduced with the ISI-MIP data when compared with the WATCH data. In the Congo catchment, the observed record sits more centrally within the range of the ISI-MIP driven model outputs, whereas for the WATCH driven models the observed record lies at the very bottom of the model simulation range, suggesting systematic overestimation. This is an important result, as modellers may be tempted to bias correct or post-process the results of their model runs to account for such overestimation. However, Figure 6.4 suggests that such results can be due to the input data, and that correction could lead to worse results when different, potentially better, input data is applied. These results show that we can have confidence when applying ensemble-based model calibrations with alternative climate input datasets, particularly when a range of model realisations are considered.

6.4 Mac-PDM.14 and the ISI-MIP MME

As in Chapter 4 with the WATCH multi-model ensemble (MME), the top performing ISI-MIP data driven model has been compared with the ISI-MIP multi-model ensemble members. The openly available outputs of each of the model runs from the ISI-MIP project were downloaded through the project website (ISI-MIP, 2015). Table 6.1 shows those models that were included in the WATCH and the ISI-MIP projects, which provided discharge data available for download. The ISI-MIP models provided data for the period 1971-2004, but 1971-2000 was used in this study to allow for comparison with the modelling carried out for this thesis. The MARE over each of the 19 river catchments (excluding the Nile and the Murray Darling) was calculated for the ISI-MIP models. As indicated in Table 6.1, not all of the Global Hydrology and Land Surface Models (GHMs and LSMs) participated in both WATCH and ISI-MIP and provided data that could be used in this study; with only Jules, LPJ-ML, Mac-PDM.09, MATSIRO, and MPI-HM

providing data from both projects. Orchidee was used in the ISI-MIP project, but output data was only available post-2000. The results of the MARE scores averaged over the 19 catchments are shown in Figure 6.5, which shows both the top performing model realisation from the GLUE ensemble, that had been calibrated with the WATCH data but run with the ISI-MIP data, and the top performing model realisation from the smaller ensemble of 1,238 model realisations, which was essentially recalibrated to the ISI-MIP data. These two models are indicated by the yellow and the green lines in Figure 6.2 respectively. When compared with the graph for the WATCH ensemble (shown in Figure 4.9), the models in the ISI-MIP multi-model ensemble (MME) had higher MARE values than the WATCH MME, and so did not perform as well. From the ISI-MIP MME, no models scored a MARE lower than WBM at 0.72, whilst the WATCH MME best (MATSIRO) gave a value of 0.5. Interestingly, MATSIRO scores a worse value within the ISI-MIP MME, with a MARE of 0.95.

Table 6.1 Participation of models in the WATCH and the ISI-MIP projects. *X indicates models which participated in the project but which could not provide data applicable for comparison in this study.*

Model	WATCH	ISI-MIP
DBH		✓
GWAVA	✓	
H08		✓
HTESSEL	✓	
Jules	✓	✓
LPJml	✓	✓
Mac-PDM.09	✓	✓
MATSIRO	✓	✓
MPI-HM	✓	✓
Orchidee	✓	x
PCR-GLOBWB		✓
VIC		✓
WaterGAP	✓	x
WBM		✓

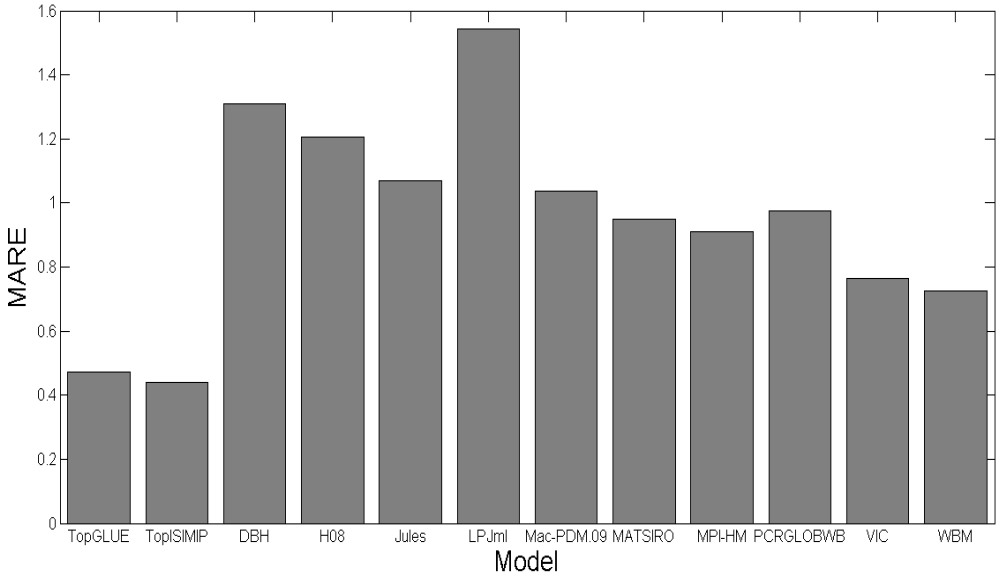


Figure 6.5 MARE scores of the ISI-MIP multimodel ensemble compared with the top performing Mac-PDM.14 model calibrations with WATCH and ISI-MIP data (MARE calculated for Jan-Dec for 19 catchments). *The TopGLUE model realisation is the top performing model from the GLUE ensemble, driven with ISI-MIP data, and the TopISIMIP model realisation is the top perfoming model from the ensemble of 1,238 realisations, also driven with ISI-MIP data.*

Whilst the top performing Mac-PDM model from the GLUE ensemble scored only slightly better than the other WATCH MME results, Figure 6.5 shows that Mac-PDM.14 has a much lower MARE than any of the other participating models in the ISI-MIP project. As previously mentioned, the MARE of the WATCH calibrated model run with the ISI-MIP data was 0.45, whilst the ISI-MIP calibrated model scored a slightly better 0.44. The WATCH calibrated model run with the WATCH data had a MARE of 0.47. This shows again that despite the calibration data set used, the model performed better with the ISI-MIP data. However, it performed even better when calibrated to the ISI-MIP data. It is interesting to see here that the original version of Mac-PDM (Mac-PDM.09) scored a MARE of 1.04, which highlights the significant improvement to the model with the updated land cover and soil maps, as well as the GLUE experimental calibration method. The MME results of the WATCH and ISI-MIP projects are considered in more detail in

relation to catchment-calibrated model validation in Section 6.5 of this chapter, and shown in the graphs in Figure 6.9.

The graphs in Figure 6.6 show the range of the 1,238 GLUE ensemble models with a MARE <0.75, run with the ISI-MIP data (herein referred to as the parameter ensemble) compared to the ISI-MIP ensemble members. These graphs are remarkably similar to those seen in Chapter 4 (figure 4.8), which reinforces the conclusion from that chapter: that the range of parameter uncertainty of Mac-PDM.14 with an acceptable error limit (MARE <0.75) is comparable to the structural uncertainty from using different models for most catchments. The Amazon seems to be an exception to this, as the parameter ensemble range is much larger than the range of the ISI-MIP ensemble members. The Congo also shows a discrepancy between parameter and model structural uncertainty, as the ISI-MIP ensemble members show a high runoff magnitude in the top part, and above, the parameter realisations range. The observed record values from Figure 6.4 shows that the ISI-MIP ensemble runs are mostly overestimating the runoff, which is observed to be primarily below a value of 50 mm per month. Between the ISI-MIP models, few conclusions can be drawn, except that DBH shows consistently high runoff values, whilst VIC gives simulations in the middle of the ISI-MIP ensemble range (except for in the Congo where it gives a better simulated runoff, lower than the other models). WBM simulates values that are close to VIC, and these are the two models with the best MARE scores. The Murray Darling is a catchment where the ISI-MIP ensemble displays higher values of runoff than the parameter ensemble. In Chapter 4 (Figure 4.7) it was demonstrated that Mac-PDM.14 overestimated the runoff on the Murray Darling considerably, so these results show that the other models in the ISI-MIP project have the same problem, particularly the Land Surface Models LPJml and JULES.

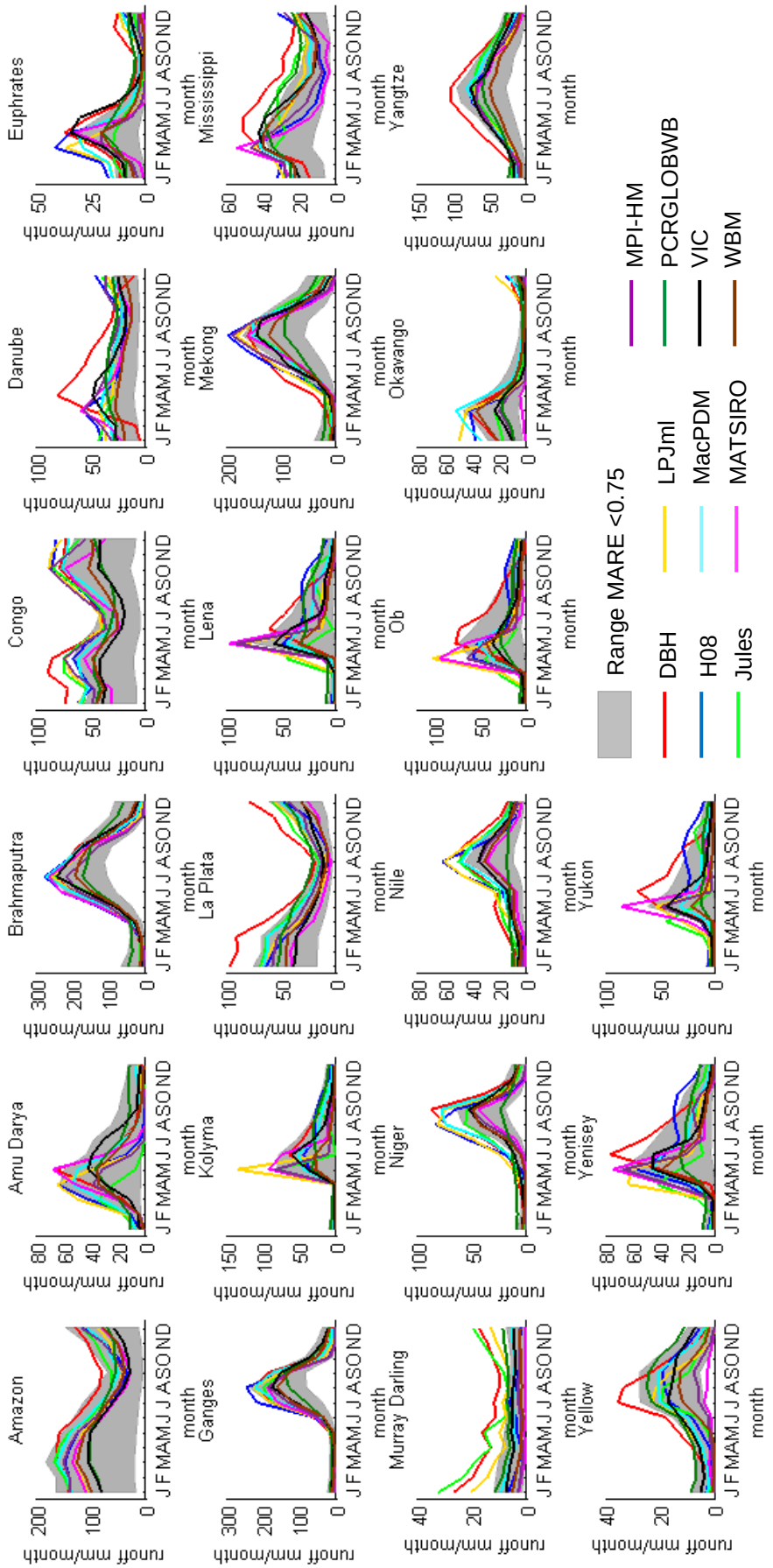


Figure 6.6 Range of MARE <0.75 runs using ISI-MIP input data, compared with ISI-MIP MME model results.

6.5 Catchment Validation

In this subsection, the use of global models as catchment models (the principle introduced in Chapter 5) is validated. The top 20 models for each catchment, identified from the GLUE ensemble of 100,000, were run with the ISI-MIP climate input data. The results of this, compared to the top 20 models with the WATCH data, are shown in Figure 6.7. The ranges displayed on these graphs were derived using the same 20 model realisations with the two different climate input datasets. The graphs in Figure 6.7 also show the top performing global calibration of the model with WATCH data, and the top performing global ISI-MIP model, which are not the same model realisation.

Immediately apparent from these graphs is the gap between the WATCH and the ISI-MIP driven models for the Congo catchment. The WATCH driven models provide better simulations compared to the observed data. The results in Figure 6.4 show that using a globally-calibrated model, the ensemble of 1,238 models contained the observed record within the range of model outputs, which suggests that the globally calibrated model was capable of simulating runoff in the Congo; however the differentiation between the catchment-calibrated model outputs with the WATCH and the ISI-MIP input data indicates that the catchment-calibration of Mac-PDM.14 is sensitive to input data, particularly in the Congo. The ISI-MIP data driven simulations seem to perform worse in the Amazon as well. However, there are many catchments where the ISI-MIP data performs better than the WATCH data, where it gets closer to, or reaches, the observed record where the WATCH data does not. For example, the June peak of the Lena is just missed by the WATCH driven runs, but is encompassed by the ISI-MIP driven runs. The same is the case with the May peak in the Danube which is underestimated by the WATCH driven models. Also in the Ganges, the Niger and the Okavango, where the WATCH data

overestimated the observed record, the range of ISI-MIP runs reach, or come closer to, the observations.

Another noteworthy result is the difference in certainty between the catchments. It is important to note the scale of the graphs is not the same so the absolute uncertainty (in mm) is difficult to determine, but the proportional uncertainty can still be gauged. It can be seen that the Brahmaputra, the Danube, the Ganges, the Kolyma, the Mekong, the Mississippi, the Yangtze and the Yenisey all have small ranges of model output, which suggests that the model simulations are quite certain. The Congo, the La Plata, the Lena, the Ob and the Okavango however, show less certain results, with wider ranges across the top 20 catchment realisations. The ranges of model output do not seem to differ significantly from the WATCH to the ISI-MIP input data, except in the Euphrates, Lena and Ob, where the peak runoff months show a slightly wider range with the ISI-MIP data that with the WATCH data. The Amazon also shows a wider range of model outputs with ISI-MIP input data for the months January-July.

The graphs in Figure 6.8 show the difference between the top performing catchment-calibrated and globally-calibrated models. Here the results show that the catchment models are significantly better than the global models for both the WATCH and the ISI-MIP data, except for the Congo and the Niger, where the difference is between the ISI-MIP and the WATCH models, and the catchment and the global models perform similarly well. The magnitude of the runoff is notably improved with catchment calibration in the Amazon, the Kolyma, the Lena, the Yangtze and the Yukon. The shape of the monthly runoff series is better simulated by the catchment calibrations in the Amu Darya. In most catchments, the best performing catchment models using the WATCH and the ISI-MIP data show very similar results, with only the Amazon, Congo, Danube, Niger and Okavango showing notable differences.

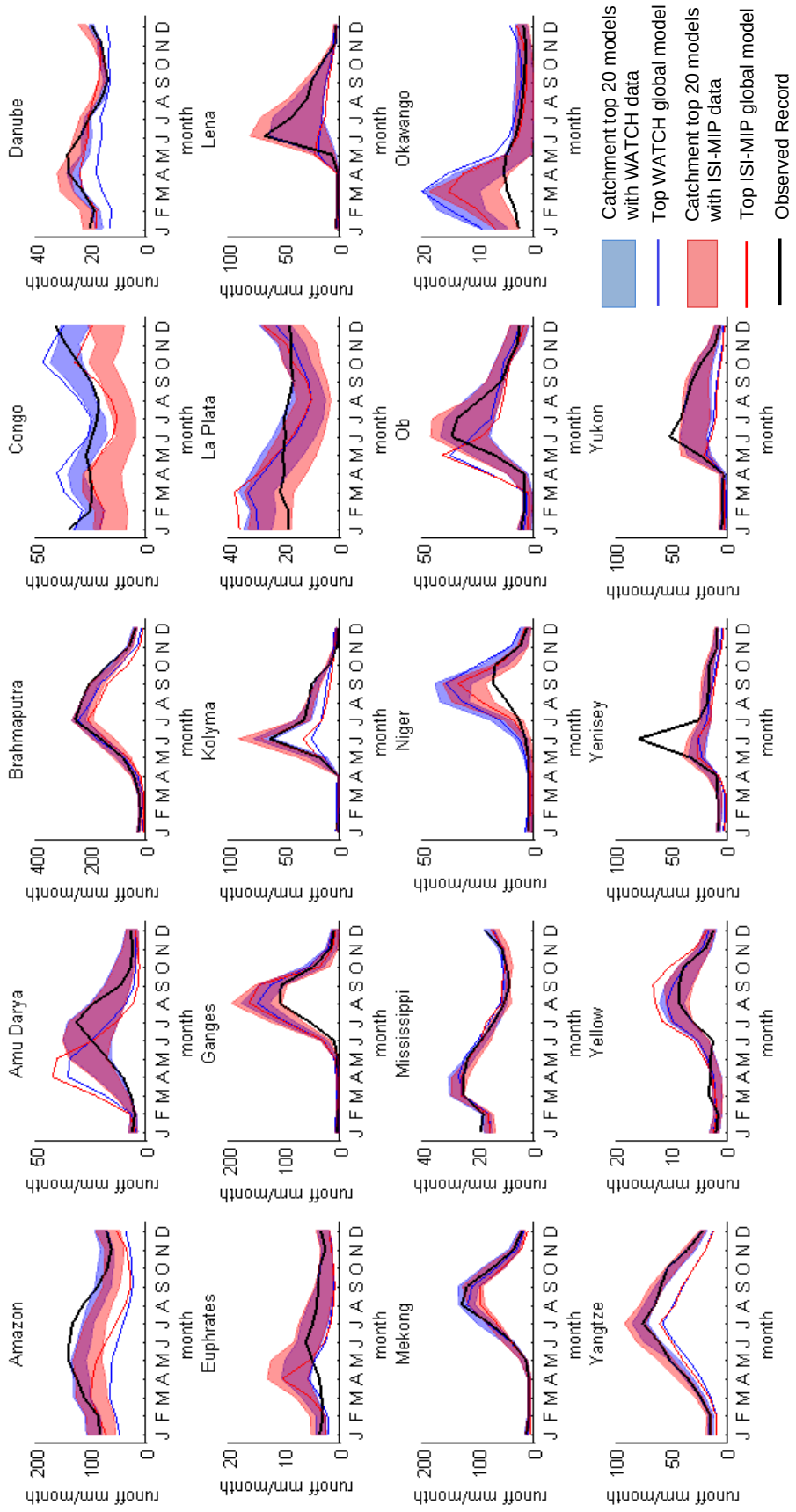


Figure 6.7 Catchment hydrographs showing the ranges of the top 20 catchment models from the GLUE ensemble, run with WATCH and ISIMIP data.

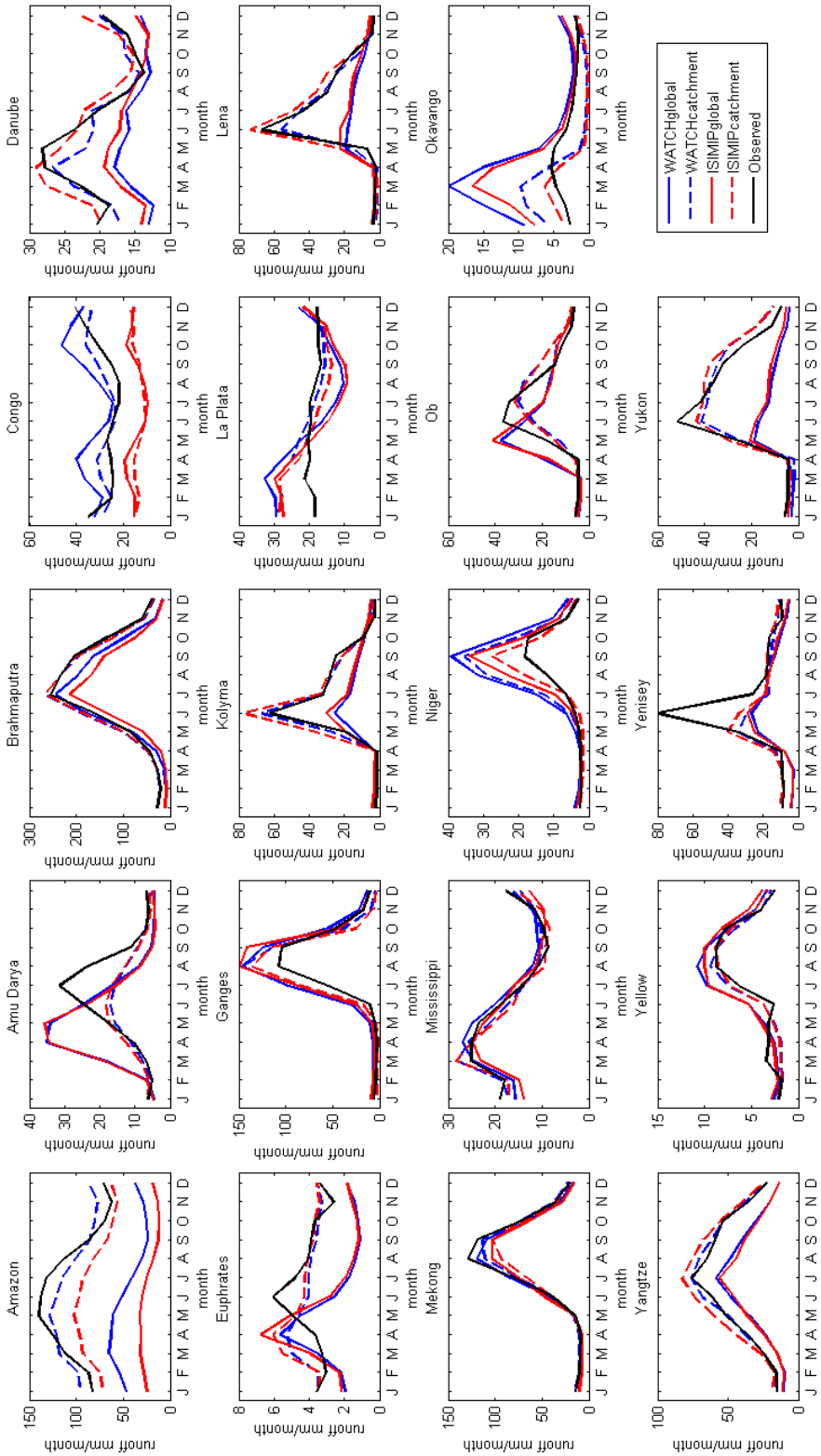
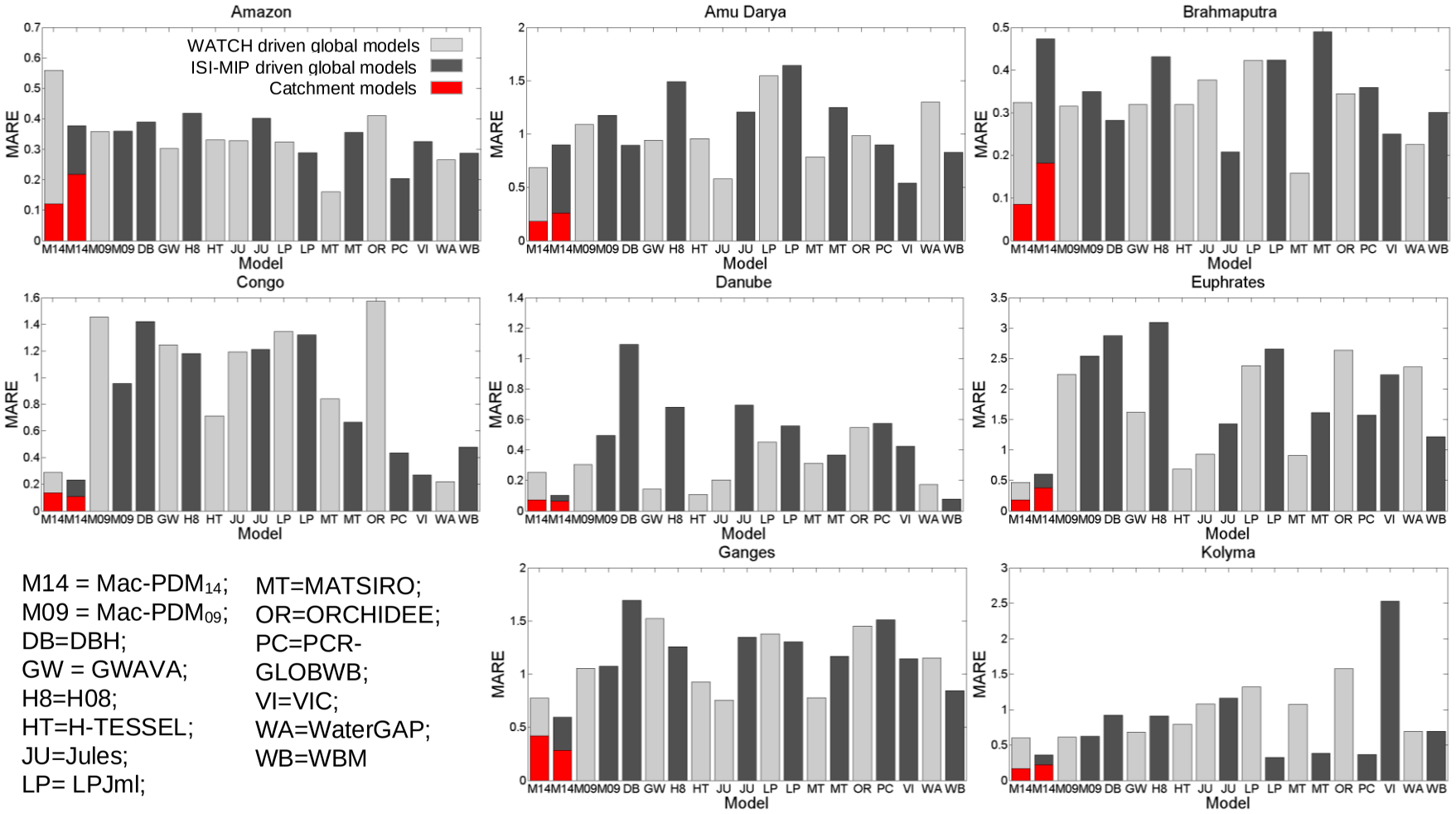


Figure 6.8 Catchment hydrographs for the top performing globally-calibrated and catchment-calibrated model realisations for the WATCH and the ISI-MIP input data

The error of the WATCH and the ISI-MIP driven realisations of Mac-PDM.14 are compared to the other WATCH and ISI-MIP multi-model ensemble members for each catchment in Figure 6.9. In these graphs, the left-hand two bars represent the WATCH and ISI-MIP driven models respectively. For these two bars, the height of the grey bar is the MARE of the top performing global calibration of the model, and the red bar is the MARE of the top performing catchment calibration of the model. Note that the catchment-calibrated models use the same model parameter values, whilst for the globally-calibrated models, the WATCH driven model is the top performing model of the GLUE 100,000 realisation ensemble, and the ISI-MIP driven model is the top performing model of the 1,238 model realisations re-run with the ISI-MIP data.

The graphs in Figure 6.9 show the performance of each model from the multi-model ensemble in more detail. Interestingly, as shown in Figure 6.3, when averaged over all the months of each catchment, sometimes the WATCH driven Mac-PDM.14 performs better, and sometimes the ISI-MIP driven Mac-PDM.14 does. The catchment calibrated model always performs better than the globally calibrated model, though not noticeably in the Mississippi. The most significant betterments between the catchment and the globally calibrated models are in the Amazon, the Brahmaputra and the Yangtze.

Overall, the global calibrations of Mac-PDM.14 seem to show comparable results to the ISI-MIP and WATCH MME models. The Amazon catchment shows particularly bad results for globally-calibrated Mac-PDM.14 with WATCH data in comparison to the other WATCH and ISI-MIP MME models. Also, the previous version of Mac-PDM (Mac-PDM.09) performs better than the global calibration of Mac-PDM.14. This is likely due to the fact the global calibration is attempting to match the other 18 catchments and therefore is trained away from the optimum calibration for the Amazon.



M14 = Mac-PDM$_{14}$;
M09 = Mac-PDM$_{09}$;
DB=DBH;
GW = GWAVA;
H8=H08;
HT=H-TESSEL;
JU=Jules;
LP= LPJml;
MT=MATSIRO;
OR=ORCHIDEE;
PC=PCR-GLOBWB;
VI=VIC;
WA=WaterGAP;
WB=WBM

Figure 6.9 Comparison of the MARE of catchment and global best WATCH and ISI-MIP driven model realisations with MMEs.

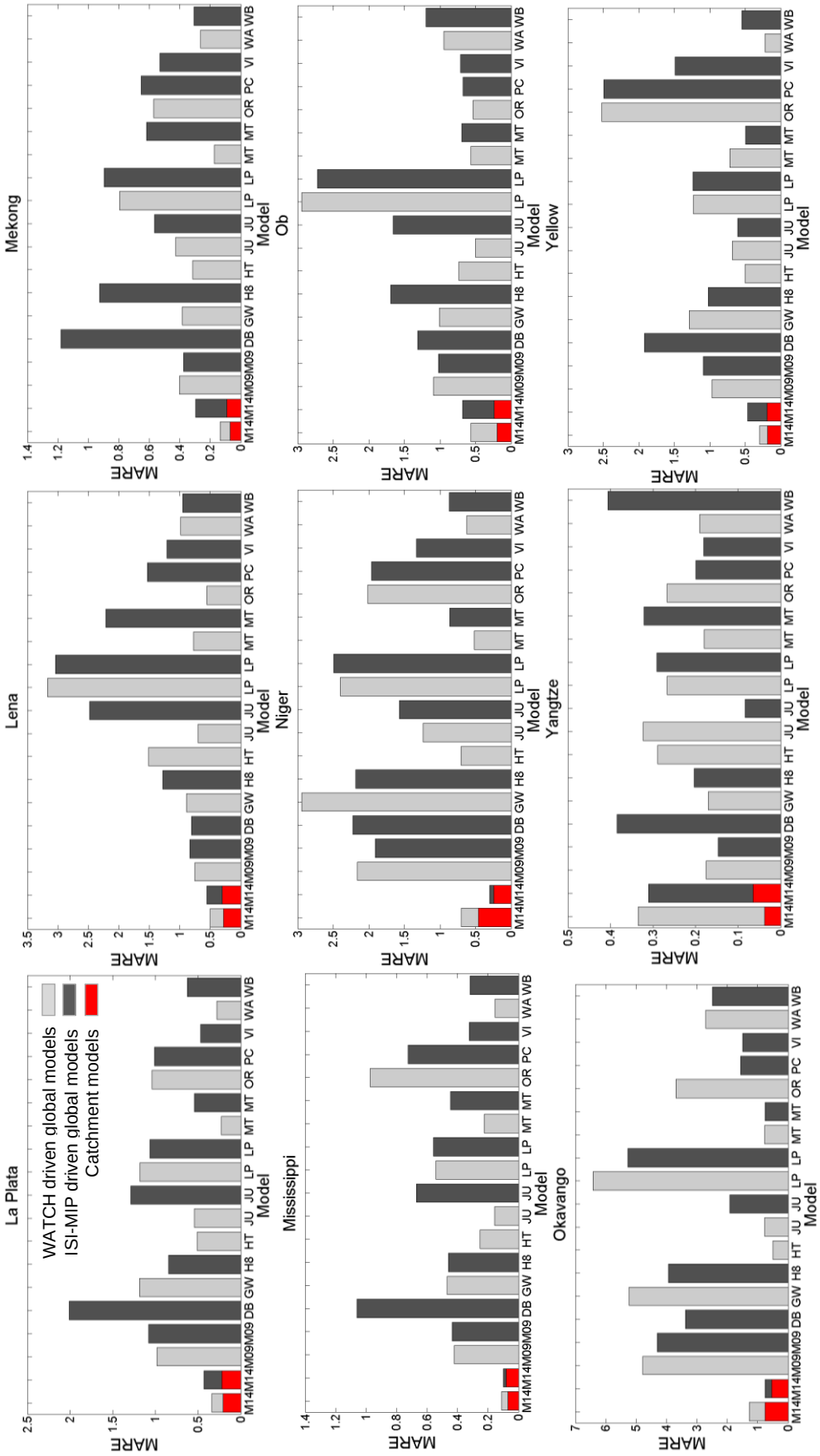


Figure 6.9 Cont.

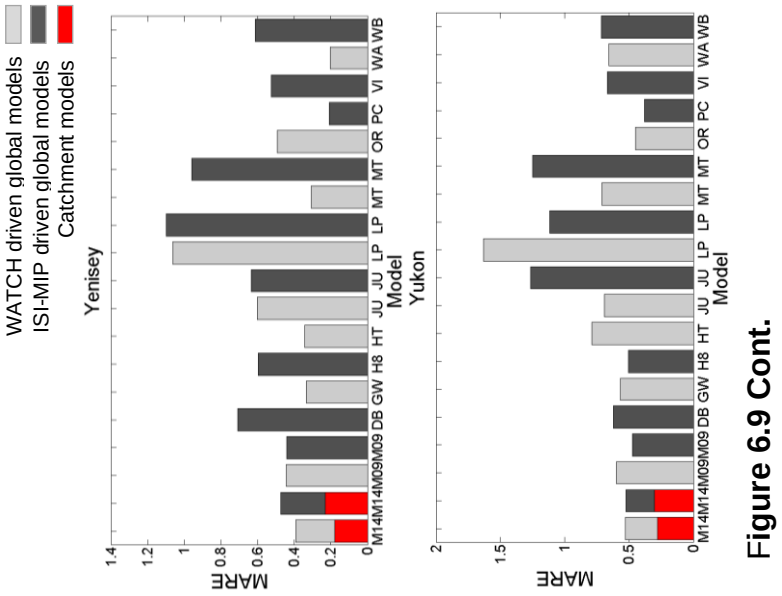


Figure 6.9 Cont.

Mac-PDM.09 outperforming Mac-PDM.14 is the case in several of the catchments. This is explored in more detail in section 6.6 of this chapter.

For each catchment, different models from the WATCH and ISI-MIP multi-model ensemble perform the best. The results of the models for each catchment show closer competition with Mac-PDM.14, than when averaged across all catchments, as was shown in Figure 6.5. The global calibrations of Mac-PDM.14 only performed better than all other models (including Mac-PDM.09) in 6 out of the 19 catchments: the Euphrates, Ganges, Lena, Mekong, Mississippi and the Niger. The Mississippi and Niger catchments showed the best results with the ISI-MIP data, while the remaining four showed the best MARE when driven with the WATCH data. The catchment calibrations of Mac-PDM.14 performed better than the models in all catchments except the Okavango, where H-TESSEL performed very well. Out of the catchment calibrations of Mac-PDM.14, 15 catchments had better results with the WATCH input data than the ISI-MIP input data. MATSIRO performed the best in the Amazon, Brahmaputra, Kolyma, and Yangtze. WaterGAP performed best in the Congo, La Plata, Yellow, and Yenisey.

VIC performed best in the Amu Darya, WBM in the Danube, Jules in the Ob, and PCR-GLOBWB in the Yukon.

From the scales of Figure 6.6, it can be seen that some catchments exhibit more certainty than others; the Okavango shows high levels of uncertainty, as does the Ob, Lena, Euphrates, Niger and Kolyma. This selection of catchments matches those that showed a wide range of model outputs among the top 20 model realisations in Figure 6.7, so reinforces that these catchments exhibit higher levels of uncertainty than the others.

6.6 Calibration and Validation

As has been identified from Figure 6.9, in some catchments the calibration process made the model perform worse than the original version of the model (Mac-PDM.09). Conversely, in some catchments, the use of the ISI-MIP climate input data made the model perform better, which is unexpected. This leads to the questions:

1. *"does calibrating a model globally improve the model performance over the majority of the catchments?",*
2. *"can a model calibrated to one climate input dataset be sensibly implemented using a different dataset?"*, and finally,
3. *"does the expected improvement in model performance from calibration, outweigh the expected decrease in performance due to the application of a non-calibrated climate input dataset?"*

These questions will help address a serious issue in global hydrology modelling, of whether models must be recalibrated to each new set of climate input data. Figure 6.10 shows that in the context of Mac-PDM, the changes in MARE are due to both calibration (the move from Mac-PDM.09 to Mac-PDM.14), and validation (the move from WATCH to ISI-MIP input data). This figure demonstrates the performance of the top performing, globally-calibrated model from the 100,000 member GLUE ensemble.

The change in model performance due to calibration is shown in Figure 6.10 by the green arrows. Downward pointing arrows show the expected decrease in MARE values from Mac-PDM.09 to Mac-PDM.14. This expected change is shown by full shading of the arrows. The unexpected decrease in model performance is indicated where the green arrows have an upward direction, and are shaded with hatching. The first question "*does calibrating a model globally improve the model performance over the majority of the catchments?"* can be answered by studying these green arrows. 16 of the 19 catchments (the 21 study catchments, excluding the Murray Darling and the Nile catchments which have previously shown erroneous results) show a decrease in MARE. This suggests that calibrating the model does improve the model performance for the majority of catchments. The most significant decrease in MARE, and so the most significant increase in model performance, was observed in the Congo catchment, with a decrease in MARE of 80%. The Euphrates, Mississippi, Okavango, Yellow, Niger, Mekong, and La Plata all showed substantial reductions in MARE, of greater than 65%. The three catchments that showed a significant increase in MARE were the Yangtze, Amazon and Brahmaputra, with increases of 91, 56 and 2.7% respectively. This is a significant reduction in MARE for the Yangtze catchment, so a catchment calibration for this catchment would be an important consideration.

The change in model performance due to input climate data is shown by the red arrows in Figure 6.10. In this case, the expected change would be an increase in MARE, as the ISI-MIP data that the model is not calibrated with is likely to cause a decrease in model performance. Therefore, the upward arrows have a solid red fill, and the downward arrows have a hatched red fill. Here, 9 of the 19 catchments show an increase in MARE with the ISI-MIP data over the WATCH data.

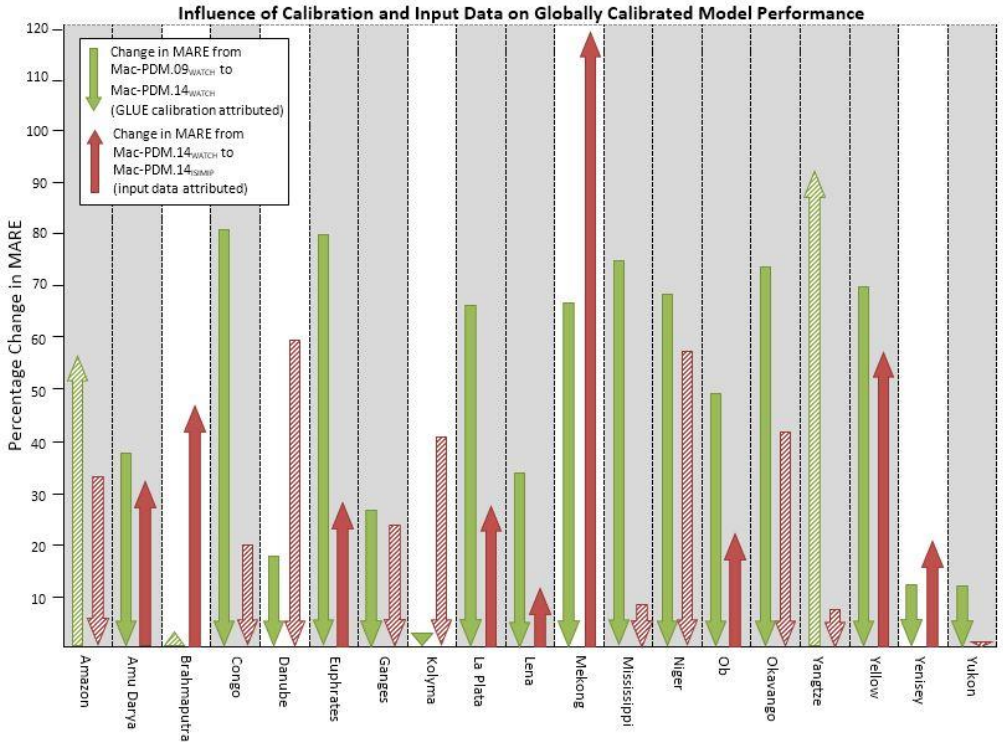


Figure 6.10 Percentage changes in model performance (MARE) for each study catchment due to the calibration process (from Mac-PDM.09 to Mac-PDM.14) and due to the use of alternative climate input data (WATCH to ISI-MIP). *Error due to calibration shown in green and climate input data shown in red. Upward arrows indicate an increase in MARE value, downward arrows indicate a decrease in MARE value. Solid arrows indicate the expected direction of change, whilst hatch arrows indicate unexpected change. Grey shading is assigned to catchments where the change due to calibration is stronger than the change due to input data type (validation).*

The largest increase in MARE with the ISI-MIP data was in the Mekong catchment, an increase of 119%. The Yellow and the Brahmaputra also show large increases at 57 and 46% respectively. This leaves 10 catchments where the MARE actually decreased with the ISI-MIP data, the most notable of which are the Danube and the Niger with 59 and 57% reductions in MARE respectively. These results show that the second question: "*can a model calibrated to one climate input dataset be sensibly implemented using a different dataset?"* can also be affirmed, as the results in general are much better than could be expected.

The final question: *"does the expected improvement in model performance from calibration, outweigh the expected decrease in*

performance due to the application of a non-calibrated climate input dataset?" can be considered by looking at the differences between the green and the red arrows for each catchment. The shaded background of the figure indicates which catchments show a larger change from the calibration procure than that of the change in input data. Here, 14 of the catchments show a larger change from calibration than data input. In the Amazon and the Yangtze though, the direction of the calibration arrow was towards a worse MARE value. In order to determine that the beneficial impact of the calibration procedure outweighs the detriment of the change in input data, the height of a downward facing green arrow must be larger than that of an upward facing red arrow; downward facing red arrows may also be included regardless of their length as they show an improvement. 14 of the 19 catchments show an overall good result, whereby the decrease in MARE from calibration process was not outweighed by an increase in MARE from the change in input data. The Amazon and the Yangtze were not included in this grouping, but were replaced with the Danube and the Kolyma, where the input data arrow (in red) was larger than the green calibration arrow, but because it was downward facing, showing a betterment in MARE, the model performed better overall than Mac-PDM.09.

Studying the overall percentage change from Mac-PDM.09$_{WATCH}$ to Mac-PDM.14$_{ISI-MIP}$, 15 of the 19 catchments showed an improvement in MARE. The Mekong is included in this list because although the percentage change from the input data is greater than the percentage change from the calibration effect, the total difference still resulted in a better model (MARE went from 0.4 to 0.14 to 0.3 for Mac-PDM.09$_{WATCH}$, MacPDM.14$_{WATCH}$ and Mac-PDM.14$_{ISI-MIP}$ respectively). The four catchments that showed a worse result were the Yangtze, and the Brahmaputra, with high increases in MARE of 76 and 50% respectively, and the Yenisey and Amazon with small increases of 6 and 5% respectively. In the Yangtze, this was definitely the result of the calibration procedure, whilst in the Brahmaputra this was due to the ISI-

MIP data. The greatest overall improvement was in the Niger catchment, which showed an 86% decrease in MARE from 2.16 to 0.3. 8 of the 19 catchment showed improvements greater than 50%. These results are a solid conclusion that the model need not be recalibrated for each new dataset.

6.7 Alternative Evaluation Metrics

So far, this study has focused only at the MARE evaluation metric. This was chosen as the metric is a good all round tool, that does not place emphasis on high or low flows, and can be applied on summarised data (calculated over few data points), such as the output of Mac-PDM used in this study. MARE also allowed straightforward comparison of model performance between catchments. 4 alternative metrics have been applied to 6 instances of Mac-PDM for each catchment, each using the 30 year average values of Jan-Dec:

1. Mac-PDM.09 run as part of the WATCH project, with WATCH input data (globally calibrated),
2. Mac-PDM.09 run as part of the ISI-MIP project, with ISI-MIP input data (globally calibrated),
3. The top performing GLUE ensemble (100,000 realisation) model Mac-PDM.14 run with WATCH data (globally calibrated),
4. The top performing ISI-MIP ensemble (1,238 realisation) model Mac-PDM.14 run with ISI-MIP data (globally calibrated),
5. The top performing GLUE ensemble model Mac-PDM.14 run with WATCH data (catchment calibrated), and
6. The top performing ISI-MIP ensemble model Mac-PDM.14 run with ISI-MIP data (catchment calibrated)

The metrics applied were the Nash Sutcliffe Efficiency criteria (Eq 6.1), Percent Bias (Eq 6.2), Root Mean Square Error (Eq 6.3), and the Standardised Effect Size (Eq. 6.4).

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n}(Q_i - \bar{Q})^2} \quad \text{(E 6.1)}$$

$$PBIAS = 100 * \left(\frac{\sum_{i=1}^{n}(\hat{Q}_i - Q_i)}{\sum_{i=1}^{n} Q_i}\right) \quad \text{(E6.2)}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{n}} \quad \text{(E6.3)}$$

$$SES = \left|\frac{(Q_i - \bar{Q})}{\sigma(Q_i)}\right| \quad \text{(E6.4)}$$

where $\hat{Q}_i$ is simulated runoff at time-step i, Q_i is observed runoff, $\bar{Q}$ is the mean of the observed record, and $\sigma(Q_i)$ is the standard deviation of the observed record. Graphs showing the results of these analyses are presented in Figure 6.11 to Figure 6.14. The NSE and RMSE metrics are not likely to be reliable indicators of goodness, due to the fact they were applied to only 12 time-series values (one for each month of the year, averaged over the 30 year period 1971-2000). The NSE metric is at its optimum at a value of 1, and NSE values of greater than 0.7 or 0.8 are commonly accepted as representing a 'good' model fit (e.g. Krause et al., 2005, Park and Ip, 2010). The metric may fall as low as -∞.

The incredibly low values of NSE in many of the catchments (shown in Figure 6.11) for the Mac-PDM.09 models is suggestive of an issue using the NSE metric in this situation. However, the results show significant improvements in model performance in the majority of catchments from Mac-PDM.09 to Mac-PDM.14. As with MARE, the Amazon is an exception to this, with the Mac-PDM.09 models showing a better model performance than the global calibrations of Mac-PDM.14. Also mirroring the results of the MARE evaluation metric, the Congo, Euphrates, La Plata and Okavango show the largest improvement as a result of the calibration process (from Mac-PDM.09 to Mac-PDM.14). The Lena and the Yukon show the greatest differences between the catchment calibrations of the model and the global models of both Mac-PDM.09 and Mac-PDM.14. The

Brahmaputra and the Mekong show especially good results from all models, and the Yangtze, Kolyma, Lena, Ganges, Danube, Mississippi, Yellow and Yukon catchments all show good results with the catchment calibrated models.

In the Percent Bias (PBIAS) graphs in Figure 6.12, the opposite trend is expected, as good model fit would be a value of 0, with +/- ∞ representing worse model results. These graphs again show generally better results from the Mac-PDM.14 models than the Mac-PDM.09 model, which reinforces the benefit of the model calibration process. Again, the Amazon is a notable exception, and here the Brahmaputra, Lena, Kolyma, Yangtze, Yenisey and Yukon show worse results from the global calibration of the model. The better scoring of the catchment calibrations of Mac-PDM.14, using the WATCH data over the ISI-MIP data, is considerable in the PBIAS scoring method. The Murray Darling and the Nile catchments show results orders of magnitude worse than the other catchments, a result that can be seen in the Nash Sutcliffe graphs as well, and was realised early on in the thesis with the MARE metric.

The Root Mean Square Error (RMSE) graphs in Figure 6.13 also indicate the best model performance at a value of 0, and show in general a slight betterment from Mac-PDM.09 to the global calibration of Mac-PDM.14, and then further improvement from the global calibration of Mac-PDM.14 to the catchment calibration. 14 of the 21 catchments showed better results from Mac-PDM.09 to Mac-PDM.14 global calibration, and all catchments showed improvement from the global calibration to the catchment calibration of Mac-PDM.14. The Yenisey is the only catchment where neither the global nor the catchment calibrations of Mac-PDM.14 perform better than Mac-PDM.09: this is consistent with the NSE and PBIAS evaluation metrics.

The Standardised Effect Size (SES) relates the model's error to the range of the observed values, and so takes into account the variability

of the catchments flow regime. Again the best score of SES is 0, and the metric can range up to +∞. The results of this score are again similar to those that we have already seen. In the majority of catchments, Mac-PDM.14 scores better than Mac-PDM.09, and the catchment calibration again scores better than the global calibration in all catchments. In this case, as was seen in MARE metric, but not NSE, PBIAS and RMSE, the Yenisey shows improvement with the catchment calibration of Mac-PDM.14 over Mac-PDM.09. Again the Yangtze and the Amazon show significantly worse results from Mac-PDM.14 than Mac-PDM.09, but the catchment calibrations perform better. The improvements from Mac-PDM.09 to Mac-PDM.14 are most notable in the Mississippi, Congo, Okavango, Niger and Euphrates.

The graphs in Figure 6.11 to Figure 6.14 demonstrate that despite some trade-offs between catchments (e.g. the Amazon), the calibration process significantly improved the results of the Mac-PDM model across a range of model evaluation metrics. Despite the fact that the model was calibrated using the MARE metric, the NSE, RMSE, PBIAS and SES scores all show significant improvements in the majority of catchments. Furthermore, the catchment calibrations of Mac-PDM.14 also show improvement over the global calibration in most catchments for all metrics. This is an important result, because it means that the choice of error metric that is used in the calibration process, which is an inevitably subjective choice, does not dramatically influence the results of the calibration. The calibration process has improved the model regardless of the error metric employed.

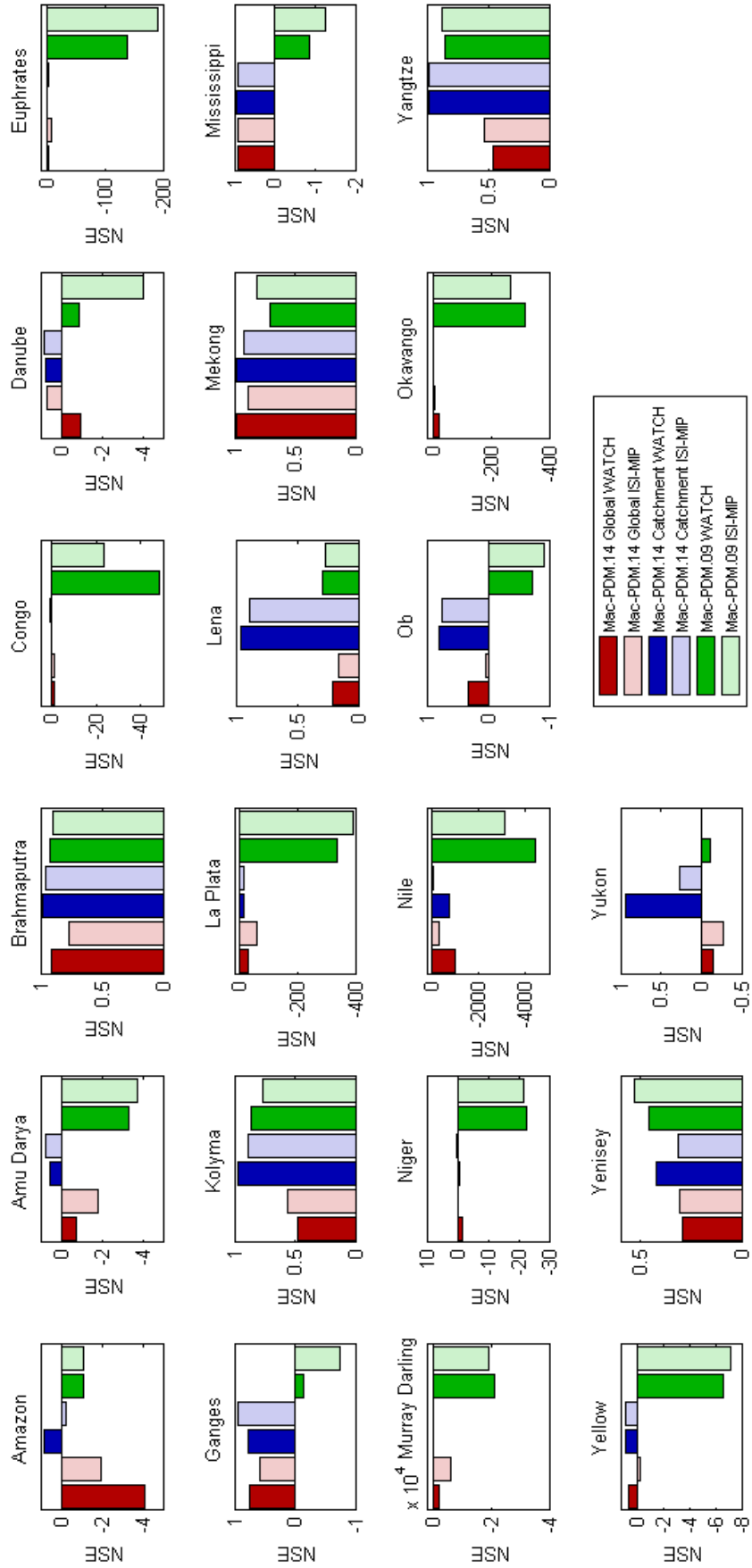


Figure 6.11 Nash Sutcliffe Efficiency scores for the top performing global and Mac-PDM.14 models run with WATCH and ISI-MIP input data, as well as the original Mac-PDM.09 model runs for each project.

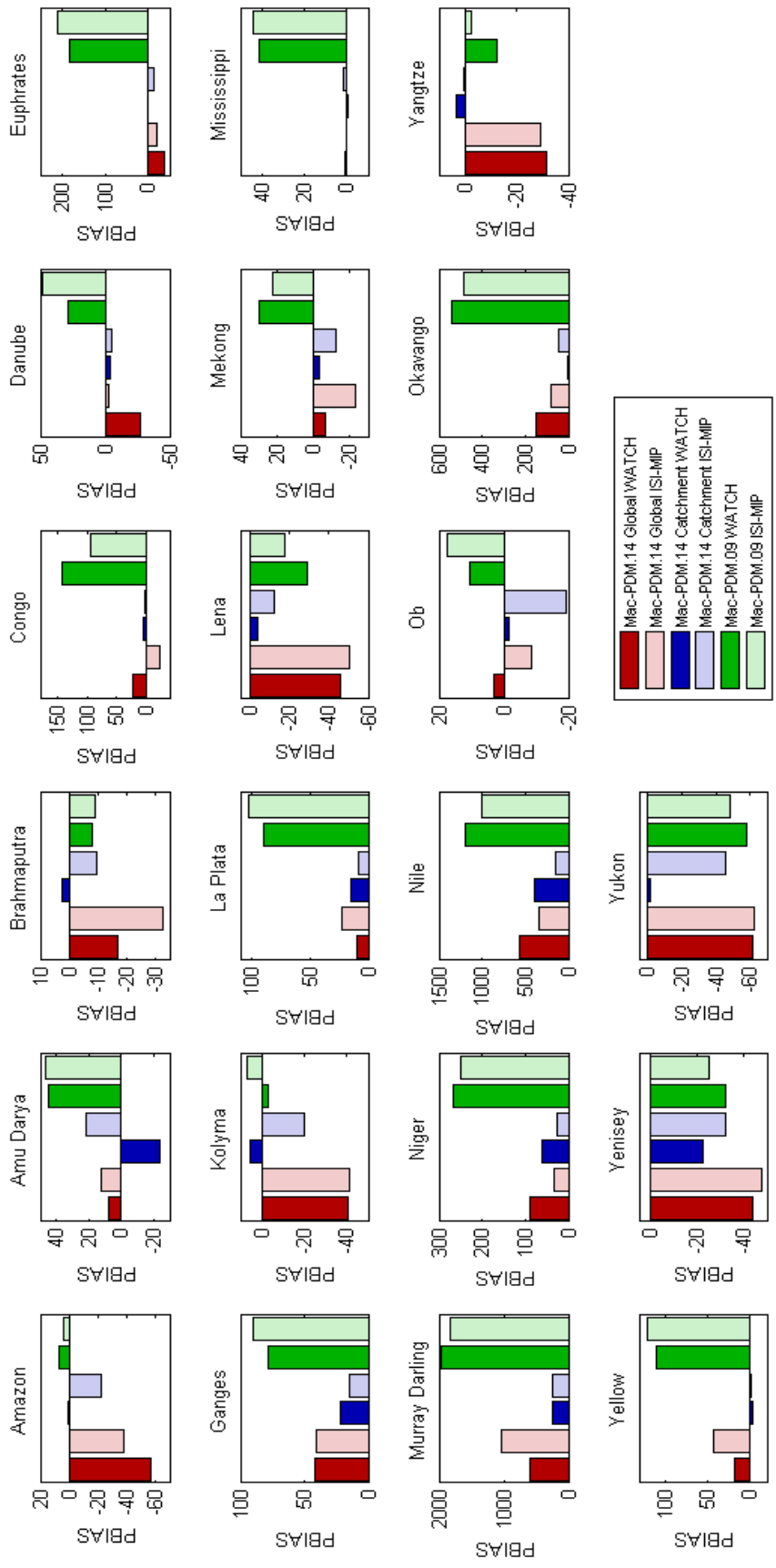


Figure 6.12 Percent Bias scores for the top performing global and Mac-PDM.14 models run with WATCH and ISI-MIP input data, as well as the original Mac-PDM.09 model runs for each project.

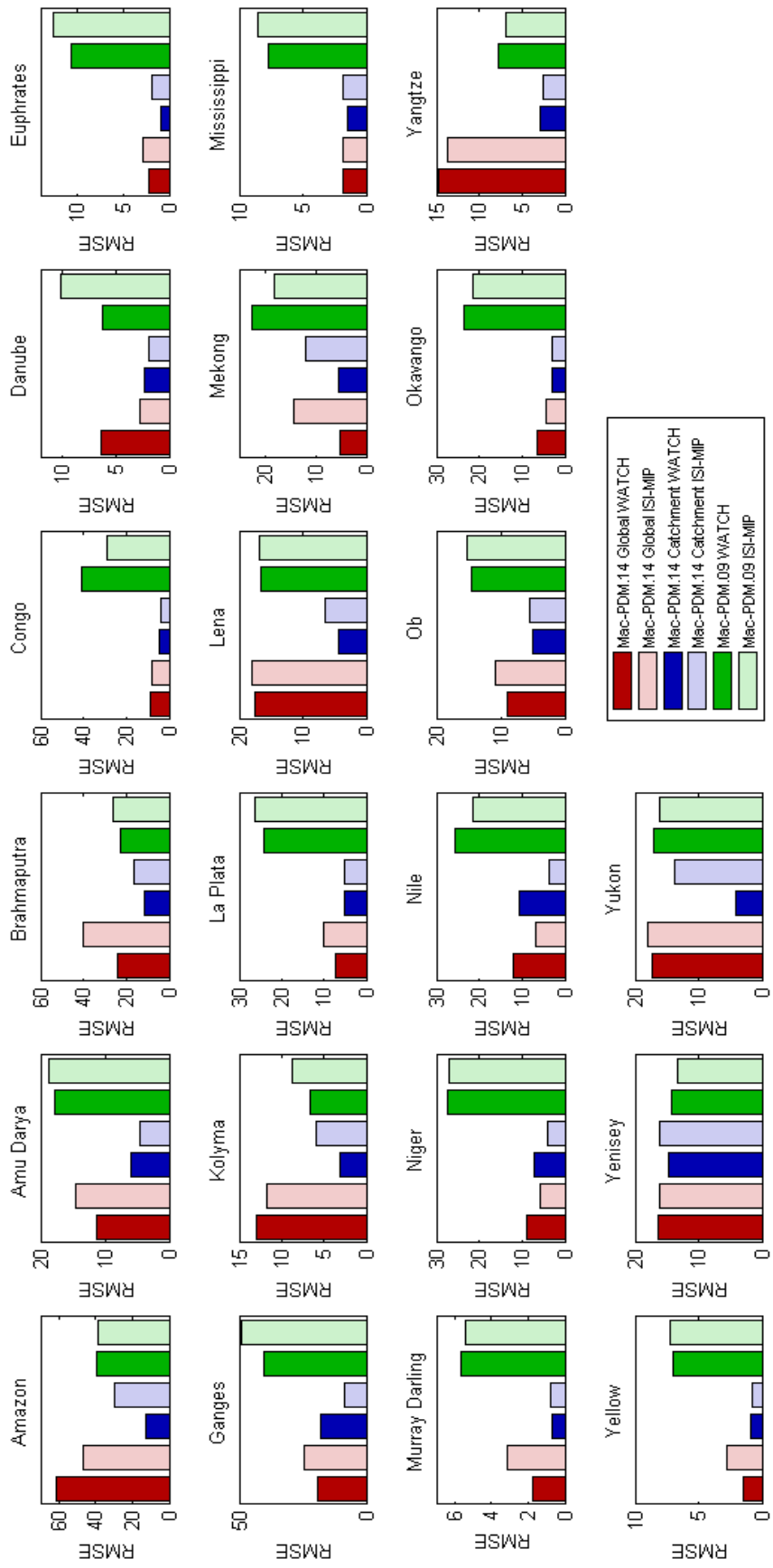


Figure 6.13 Root Mean Square Error scores for the top performing global and Mac-PDM.14 models run with WATCH and ISI-MIP input data, as well as the original Mac-PDM.09 model runs for each project.

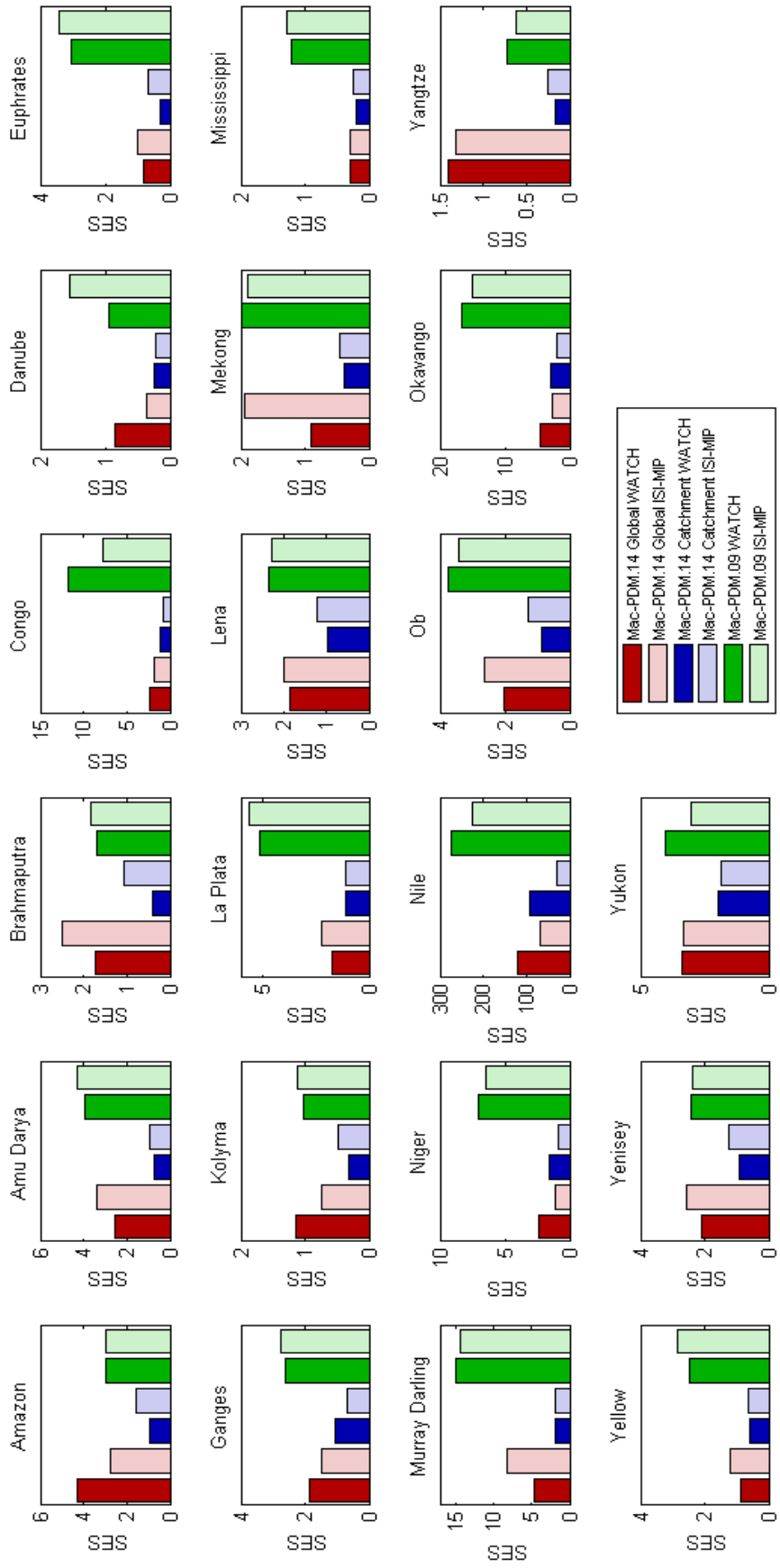


Figure 6.14 Standardised Effect Size scores for the top performing global and Mac-PDM.14 models run with WATCH and ISI-MIP input data, as well as the original Mac-PDM.09 model runs for each project.

6.8 Summary

This chapter has demonstrated that the updated, assessed and recalibrated Mac-PDM.14 model can be applied with alternative climate input datasets with good results. The ISI-MIP dataset for the period 1961-2000 was deemed different enough to the WATCH dataset to provide an adequate validation test for the model, so was applied to both global and catchment calibrated model realisations. The global model realisations with a MARE <0.75 were run with the ISI-MIP input data and, for the most part, the results were actually better than those when run with the WATCH data. There was a smaller range of uncertainty across the 1238 models that were run with the ISI-MIP data, than when run with the WATCH data for most catchments. When compared to the other individual members of the ISI-MIP multi model ensemble, Mac-PDM.14 performs much better, with a MARE score of 0.44, compared with the best MME model (WBM) at 0.72.

When catchment calibrations of the Mac-PDM.14 model are considered and investigated in detail, it is apparent that the specific catchment calibrations are more sensitive to input data. The model has been more finely tuned to the WATCH data, and the difference in the ISI-MIP data has more of a negative impact on the runoff outputs. However, in some catchments, the model again performs better with the ISI-MIP data. The catchment calibrations perform better than the global calibrations for all catchments, regardless of which climate input data is being used. Global calibrations of Mac-PDM.14 do not always give better outputs than all of the ISI-MIP multi-model ensemble members, but the catchment calibrations of Mac-PDM.14 do. It would be interesting in further research to see how the catchment calibrations of Mac-PDM.14 performed against catchment calibrations of the other models.

The influence of the calibration process upon the model performance in each catchment was assessed. This was compared to the model performance using the WATCH and the ISI-MIP data. It was found that

the influence of model calibration was greater than that of the input data, which suggests that the model need not be recalibrated to every new input dataset. Alternative model evaluation metrics were employed (Nash Sutcliffe, Percent Bias, Root Mean Square Error, and Standardised Effect Size). These metrics showed very similar results to the Mean Absolute Relative Error metric that was used for the calibration and evaluation. This indicates that the results of the calibration process are conclusive, and that the subjective choice of evaluation metric does not negate the improvements seen in the model performance.

These results demonstrate that this approach to uncertainty analysis and the subsequent calibration of a Global Hydrology Model can be both beneficial and useful, and will remain relevant when new climate input datasets become available.

Chapter Seven:

Discussion: Global Hydrology Modelling – Obstacles and Opportunities

7.1 Introduction

Several obstacles and opportunities in global hydrology modelling have been highlighted in this study and the following issues are addressed here:

1. Simulating naturalised flow by hydrology models and the impact that this may have on analysis of model performance and on decision making from the models outputs (see chapter 4).
2. The need for better process representation in Mac-PDM, and the implications of this on model results.
3. The application of global hydrology models as catchment models, and the potential for further work comparing global models with catchment hydrology models (see chapter 5).
4. The trade-off between catchments in a global model calibration, and the previously assumed need to recalibrate models to new climate input data (see chapter 6).
5. The issue of computational demand in uncertainty analysis (see chapter 4).
6. The possibility of applying model emulation techniques for sensitivity analysis, in order to reduce computational demand, and extend understanding of parameter interactions and optimum values (see chapter 5).
7. Assessing parameter uncertainty under climate change projections, and previous work on the presentation of such uncertainty from the literature in other scientific fields.
8. Gaining a deeper understanding of the full range of uncertainties in global hydrology modelling studies, and sources of additional uncertainty that could be investigated further.

7.2 Naturalised Flow Simulation

The Mac-PDM model, along with a number of other hydrological models (e.g. DBH, VIC, WBM, MPI-HM, WaterGAP, H08, PCR-GLOBWB which participated in the ISI-MIP project (Warszawski et al., 2014)) simulates

'naturalised runoff' or flow. However, these models are compared to observed discharge records which are subject to significant deviations from a natural flow regime, due to influences from abstraction and other human alterations to the flow. This issue was discussed and a potential solution to this problem was examined in Chapter 4 (see section 4.4). Mac-PDM.14 was compared with the UNH-GRDC Composite Runoff data, which attempts to provide a corrected observed time series for comparison with naturalised flow simulation. However, whilst the Composite Runoff data is useful for water resource assessments and validation of atmospheric models, the usage guidance of the dataset states that "the use of the composite runoff data for validation is not recommended (because it is a mixture of modelled and measured discharge)" (Fekete and Vorosmarty, 2011 pp. 18).

Some hydrological models such as GWAVA (Meigh et al., 1999) and WaterGAP (Alcamo et al., 1997) use abstraction data to simulate water availability, yet abstraction data is very difficult to obtain, especially in the detail required to produce simulations of discharge with certainty levels high enough to be deemed useful. Abstraction data and reservoir levels are often quite sensitive and are currently unavailable to the public. This leads to estimated abstraction datasets, often at a national scale (such as AQUASTAT, FAO (2014a)), which introduce significant uncertainty to modelling studies.

The difficulties that lead modellers to simulate naturalised river flow often encourage them to focus their efforts on catchments that demonstrate low levels of human influence on the flow regime, such as those in the UK benchmark catchment network (Bradford and Marsh, 2003). This is often used for catchment selection in UK hydrological research (e.g. Hannaford and Marsh, 2008, Hannaford and Marsh, 2006, Stahl et al., 2010). In small scale catchment studies the selection of catchments can often factor in this consideration. Global scale catchment modelling studies on the other hand, are less able to avoid

the impact of water withdrawals as the largest of the world's rivers are often the most modified. Until detailed abstraction data become available globally, this issue will remain a significant challenge for hydrological modellers.

7.3 Process Representation in Mac-PDM

The results of the global calibration of Mac-PDM indicate that there are some structural deficiencies in the model. Significant underestimations, overestimations and mistiming of runoff peaks in catchments with distinct climatic regimes suggest that the model may need additional components to simulate runoff in both high latitudes, and arid to semi-arid regions of the world.

The underestimation of peaks in high latitude catchments such as the Lena, Yenisey and Yukon, and the simulation of peaks too early in these catchments as well as the Amazon, Amu Darya and Ob could be improved with the inclusion of a glacier component in the model that also takes into account the seasonal freezing and melting of permafrost. This structural issue was identified when discussing the potential limitations of the Mac-PDM model in Chapter 2, and became apparent in the results from the GLUE experiment shown in Figure 4.1. The mistiming of the peaks is due to the fact that in reality, much of the autumn and winter runoff is locked up in ice stores, and isn't released until spring. This isn't represented in the model, and so the runoff is simulated according to the timing of the precipitation (see Figure 6.1 for precipitation input), which results in an early runoff peak. Whilst the model does include a simple degree day scheme for snowmelt, this does not account for the larger scale ice processes of permafrost and glaciers which dominate the runoff regime in many of the study catchments investigated in this thesis.

The sensitivity of the xmelt parameter values, particularly in high latitude catchments suggests that the model could be adjusting this

parameter in order to account for the lack of glacier representation. The routing parameters were also sensitive, possibly attempting to adjust the mistiming of the runoff peaks. Whilst amending the parameter values improves model performance when calibrated to individual catchments, this cannot improve the model when a global calibration is needed.

Similarly, the model does not accurately represent runoff in arid and semi-arid catchments such as the Okavango, Murray Darling, Niger and Nile. In all of these catchments, the model overestimates runoff significantly. The performance of the model in the Murray Darling and Nile catchments is confounded by the exceptional influence of the abstractions and reservoirs on the runoff, however the overestimation in the Niger and the Okavango indicate that process representation in arid and semi-arid regions could be improved. It was harder to determine which model parameters led to an improvement in model performance in these catchments, however transmission loss and evaporation are likely key processes in these areas that could be explored further.

7.4 Catchment Models or Global Models?

This study has largely focused on the use of a global model as a catchment model, by calibrating the chosen model specifically for each catchment. The improvement in model performance using this approach was significant. The largest improvement was seen in the Yangtze catchment with an 89% reduction in MARE, from 0.33 to 0.04. All catchments showed an improvement, with catchment specific calibration showing an improvement of 33% over global calibration in the Niger catchment. This raises the question whether global hydrology models could be used in the place of catchment models. Catchment hydrology models (CHMs) generally include more complex parameterisations than global hydrology models (GHMs), and they are time-consuming to calibrate (Gosling et al., 2011). However, catchment

models operate on a higher resolution than global models which increases usefulness of model outputs.

Global hydrology models have been applied in catchment specific research (e.g. Abdulla et al., 1996, Thompson et al., 2013, Weedon et al., 2014, Trambauer et al., 2013, Gain and Wada, 2014, Aus der Beek et al., 2011), which demonstrates that there is merit in the resolution vs model complexity trade-off between CHMs and GHMs. These applications of GHMs in catchment studies suggest that the significant improvements in model performance from catchment specific calibration found here are worth exploring further. In future a comparison of a chosen CHM with a GHM for the same catchment, following a similar calibration procedure, would be worthwhile. A comparison of GHMs and CHMs was carried out by Gosling et al. (2011), but this did not take into account the newly discovered improvement in Global models when calibrated specifically to each catchment. A comparison calibrating GHMs and CHMs in a similar way for each catchment would reveal how well both models performed, and may provide different cost options for distributed catchment modelling.

7.5 Model Calibration and Input Climate Data

Chapter 6 demonstrated the substantial benefit of calibrating the Mac-PDM model. However, when evaluating the globally calibrated model, there were some catchments where calibrating the model made the output runoff worse in comparison with the observed record. When the MARE metric, which was used to perform the calibration, was considered only 3 catchments showed a worse result than the original version of the model (Mac-PDM.09); but for the Yangtze catchment there was a substantial worsening, with an increase in MARE of 91.4%. This demonstrates an inevitable trade-off between catchments in the optimisation of a global hydrology model.

Other evaluation metrics were calculated including NSE, PBIAS, RMSE and SES (see Chapter 6). Whilst all showed improvement in global averaged results after calibration, the number of catchments that showed improvement was less clear, with 11, 11, 15 and 12 out of 19 catchments showing improvement for each metric respectively. RMSE and NSE have not been found to be appropriate measures for this study, due to the 30 year averaged output of Mac-PDM, so it is not surprising that the calibration did not give much improvement in these statistics. The improvement of more than half of the catchments with SES and PBIAS is reassuring though, showing that the calibration procedure does indeed make the model better overall, regardless of the metric considered. The catchment calibration of the models showed significant improvements from Mac-PDM.09 to Mac-PDM.14 over the majority of catchments for all metrics.

In the same analysis, the benefits of model calibration were evaluated alongside the effects of using an alternative input dataset. This study suggested that the calibration process had a much stronger influence on the model output than the input data. This leads to the question, is calibration to new datasets necessary? If a rigorous calibration process is carried out, does it make the model good enough to negate the need to recalibrate to different input data? In this study, the results suggested that the improvement from model calibration outweighed the impact of changing the model input data for the majority of catchments. This may not always be the case though. There are some similarities in the way the WATCH and the ISI-MIP forcing data were derived, so a significantly different input dataset may yield a more substantial deviation in the model output. Input data may also evolve over time, therefore if a large calibration experiment is deemed useful for several input datasets, will this only be the case for a certain period of time? It would be valuable to explore this further with multiple input datasets of varying origins, as the need to recalibrate models to input data for each

new research project uses up considerable amounts of project resources.

7.6 Computational Demand

The increasing computational demand is a significant challenge in hydrological research. As datasets are produced in higher spatial and temporal resolution, and as models become increasingly complex to best represent the physical world, advances in computing technology struggle to keep up with the demand. Many institutions now make use of high performance computer (HPC) clusters, such as the one at Nottingham used for this research, which consist of multiple computers connected in a network that can run iterative or batch computations in parallel and provide storage for large datasets. The cluster used in this study consisted of a total of 2,656 CPU compute cores capable of running at over 46 teraFLOPS. The cluster at Nottingham, known as Minerva, cost approximately £1 million to set up, and costs around £150,000 per year to maintain. This cost is well justified by the widespread use of the cluster from a broad variety of research fields across the university, and is easier to maintain than several facilities spread across several sites. Clusters at other universities vary in size: e.g. the 2,340 core Darwin Cluster at the University of Cambridge (University of Cambridge, 2009), the 800 core Aquila cluster at the University of Bath (Chapman, 2013), and the 208 core ALICE cluster at University of Leicester (University of Leicester, 2015).

Outside of universities, research institutions also have clusters. A good example is the JASMIN facility, funded by the Natural Environment Research Council (NERC) and the UK Space Agency (UKSA), which provides 3,500 compute cores and 13 petabytes of fast parallel disk storage (Bennett et al., 2014) for research by NERC scientists. So, whilst not all scientists have access to such systems as used in this study, high performance computing clusters are becoming increasingly popular. This means that the bar is set high for scientific research, and

that multiple model runs for model calibration can be expected, as well as multi-model ensembles for the assessment of model uncertainty.

7.7 Sensitivity Analysis – Model Emulation

The parameters of the Mac-PDM.14 model were investigated (see chapter 5) to determine whether any trends could be identified in the GLUE results that would indicate whether parameter values for different catchments could be linked to catchment characteristics. It was determined that no firm conclusions could be drawn from the results, which indicated issues with parameter insensitivity and equifinality, which in turn suggests that the Mac-PDM.14 model is over-parameterised. Methods of exploring this further were discussed with several statisticians and it was decided that the best route forward in this situation would be to use model emulation. Model emulation derives statistical relationships between model input and output in order to simulate huge numbers of model runs without actually running the full model. This is necessary in computationally demanding climate models, and allows the modeller to explore the parameter space in detail and to carry out a variance based sensitivity analysis (Saltelli et al., 2000).

Lee et al. (2011) describe the processes involved in Gaussian process emulation, the steps of which are summarised in Figure 7.1. O'Hagan (2006) provide a tutorial of the BACCO approach (Bayesian Analysis of Computer Code Outputs) to quantifying, analysing and reducing model uncertainty using Gaussian process model emulation, aimed at non-mathematicians. In this study, despite having already run 100,000 model runs, it was estimated that it would take approximately three months to derive a Gaussian emulation model for just one catchment and so this investigation was deemed outside the scope of this study. With increases in computing power this sort of thorough investigation is likely to be possible in the near future. If it were possible, the outputs of model emulation can bring great insight to the modeller. Sensitivity analysis of the emulations can produce spatial information about the

most important sources of uncertainty in model output (Lee et al., 2013).

Parameters can be ranked in order of importance for each area, and their contribution to uncertainty in the model output can be quantified over time. An example of the potential outputs from such a study is shown in Figure 7.2, where the results from a model emulation experiment on a global aerosol processes model (GLOMAP-mode) on the sensitivity of model inputs to CCN output are presented. This sort of analysis is a significant breakthrough in parameter uncertainty assessment; and in physically based models it enables efficient visualisation of the effects of specific physical processes on model output processes.

7.8 Climate Change Projections and the Cascade of Uncertainty

The opportunity of investigating the impacts of parameter uncertainty on climate change impacts projections is one potential next step. Having assessed the parameter uncertainty of the Mac-PDM.14 model, and validated the top selected parameterisations with alternative input climate data over a historic period, it would be a novel next step in global hydrology modelling to assess how this parameter uncertainty contributes to runoff projections under different climate projections. Parameter uncertainty in catchment hydrology model projections has been researched (e.g. Wilby, 2005, Wilby and Harris, 2006), as has global multi-model uncertainty on projections (Haddeland et al., 2011), but parameter uncertainty contribution to climate impacts projection uncertainty from global hydrology models has yet to be addressed.

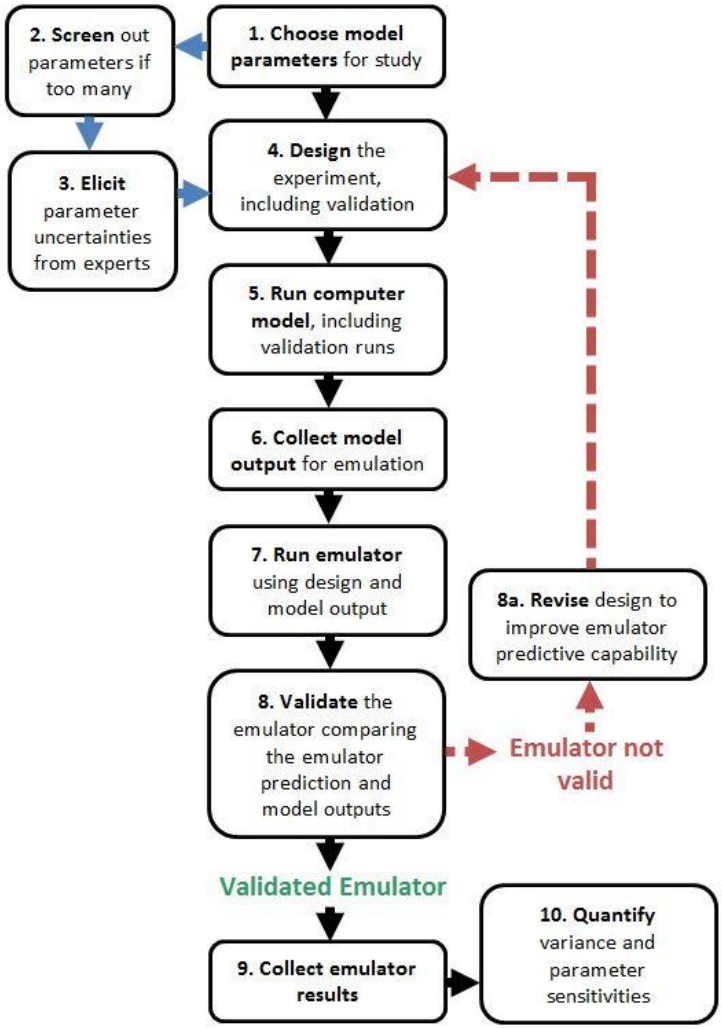


Figure 7.1 Flow chart of the basic steps in an emulation study. *After Lee et al. (2011).*

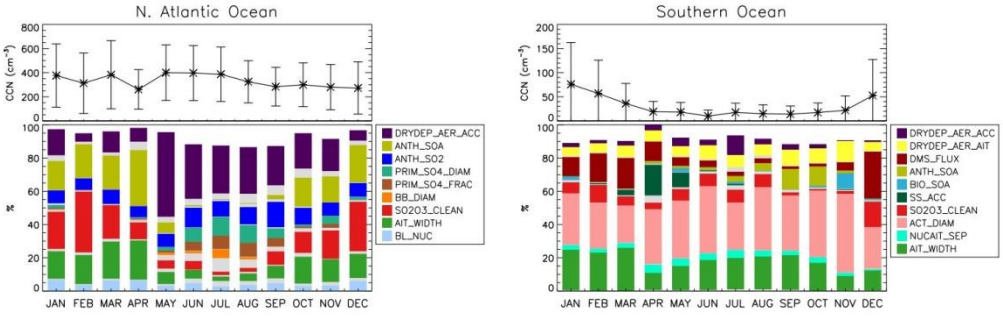


Figure 7.2 Example of the possibilities of model assessment outputs using model emulation (taken from Lee et al., 2013). "*Time series of mean emulator predicted CCN concentration with 2σ error bars (top graphs) and the main effect sensitivities (the percentage of CCN variance due to each parameter) (bottom graphs) across the year 2008 for different locations. Parameters with main effect < 5% are shown in grey. The white space filling the bars to 100% shows the fraction of variance due to interactions between the parameters, since with no interactions the main effect sum to 100%" (Lee et al., 2013).*

Wilby and Dessai (2010) presented a ‘cascade of uncertainty’ that demonstrated the propagation of uncertainty from climate scenarios, through greenhouse gas emissions, climate models, regional scenarios, impacts models, and local impacts, to adaptation responses (see Figure 7.3). This cascade is a ‘top-down’ assessment of climate risks and has a long history, beginning with the “CO_2 Pyramid” (Schneider, 1983). This was developed into the “uncertainty explosion” (see Figure 7.4) (Schneider and Kuntz-Duriseti, 2002) which was presented in the IPCC Third Assessment Report (McCarthy, 2001). Hawkins (2014) adopted the visualisation from Wilby and Dessai (2010) and applied it to the work of CMIP5 using actual data (see Figure 7.5). This study used three cascade levels to represent 1) the emissions pathway – the Representative Concentration Pathways (RCPs), 2) the different climate model results from the same forcing, and 3) the role of internal climate variability resulting from multiple realisations from the same forcing pathway (Hawkins, 2014). The same approach has also been applied to sea-ice extent from CMIP5 in Swart et al. (2015).

When impacts models are taken into account, the cascade by (Hawkins, 2014) represents only the first few levels of the cascade, and hydrology models and their realisations can be added as two additional levels to the bottom of this cascade. Alternatively, in an individual study of hydrology model uncertainty, the same cascade could be applied with the different climate model inputs in the place of the RCP input presented here. This cascade concept is a novel idea for easily interpretable presentation of uncertainty in complex multi-scenario, multi-model and multi-realisation climate projection studies.

7.9 Presenting Uncertainty

In addition to quantifying uncertainty, presenting uncertainty is a significant challenge in scientific research. The way in which scientific results are presented to decision makers determines which party is taking the most risk.

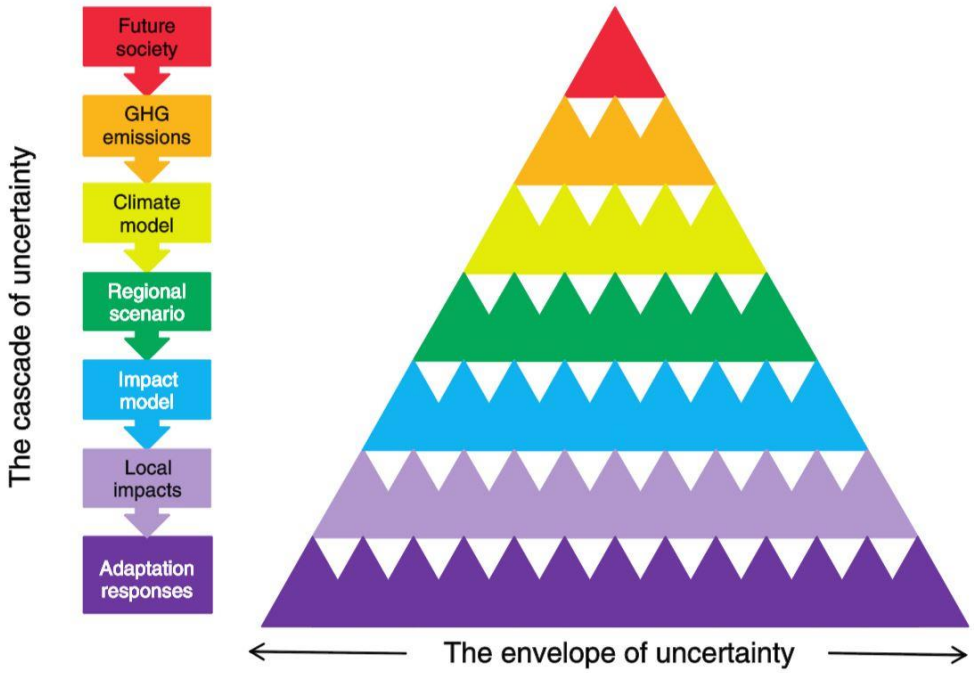


Figure 7.3 The Cascade of Uncertainty (taken from Wilby and Dessai, 2010). "*A cascade of uncertainty proceeds from different socio-economic and demographic pathways, their translation into concentrations of atmospheric greenhouse gas (GHG) concentrations, expressed climate outcomes in global and regional models, translation into local impacts on human and natural systems, and implied adaptation responses. The increasing number of triangles at each level symbolise the growing number of permutations and hence expanding envelope of uncertainty. For example, even relatively reliable hydrological models can yield very different results depending on the methods (and observed data) used for calibration" (Wilby and Dessai, 2010).*

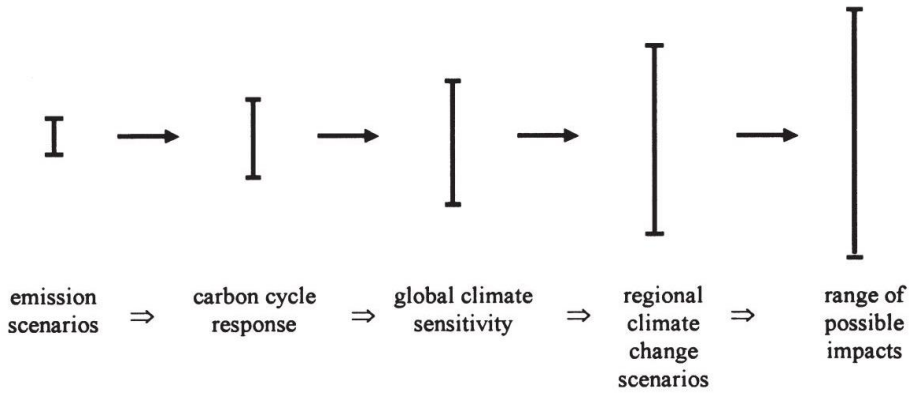


Figure 7.4 The "Uncertainty Explosion". *The ranges in major uncertainties typical in impact assessments, multiplies to encompass a comprehensive range of future consequences, including physical, economic, social, and policy responses (Modified after Jones (2000) and the "cascading pyramid of uncertainties" in Schneider (1983).*

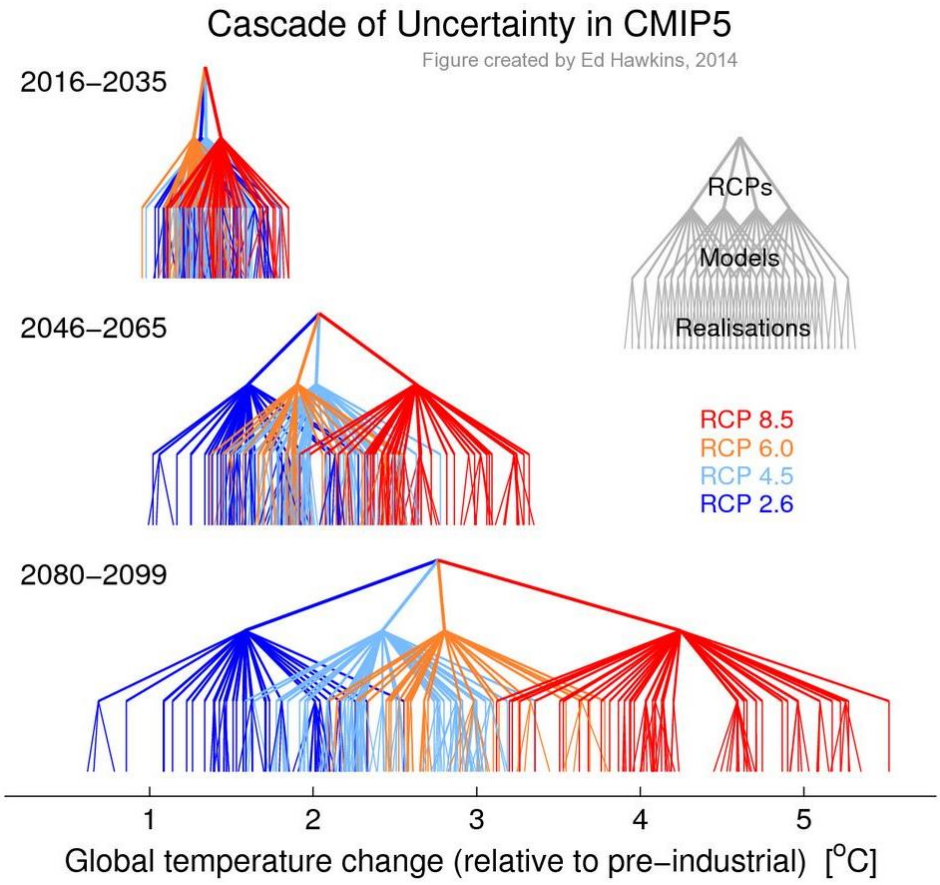


Figure 7.5 CMIP5 Cascade of Uncertainty for global mean surface temperature over different time period (Hawkins, 2014). "*The three levels of the pyramid highlight the uncertainty due to choice of RCP, GCMs and realisation of climate variability. Unfortunately not all the simulations have multiple realisations, resulting in a vertical line in the lowest layer. The intersection on the top row for each time period is the multi-scenario, multi-realisation mean" (Hawkins, 2014).*

Decision makers prefer to receive a single estimate such as a multi-model mean, which places the risk with the scientist. Scientists however, prefer to present their results with a range of uncertainty and allow the decision makers to make their decisions with as much information as possible, therefore placing the risk with the decision maker. So what is the best compromise?

There are many blogs dedicated to visualising uncertainty (e.g. www.visualisingdata.com, www.understandinguncertainty.org/visualising-uncertainty) as well as several blog posts on wider data science blogs. Probabilistic

representations are commonly in use now, as they present the full range (or as far as can be estimated) of uncertainty, whilst maintaining a best estimate for the decision maker. Spiegelhalter et al. (2011) examine the success of graphic visualizations for communicating probabilities to a wider public. This research made use of graphs that displayed the best estimate, bounded by shaded areas that gave a range of uncertainty from a suite of top performing models. Due to the size of the ensemble developed in this study, there was difficulty in employing methods that displayed the full ensemble, especially as even the matrices of data for 21 catchments were too large to be stored in the memory of a desktop computer for analysis and plotting.

If this were not the case, fan graphs would have been a good option, which show the mean with shading indicating ranges of uncertainty. Fan graphs have originated from economic forecasts, an example of which is shown in Figure 7.6. This type of graph is an improvement on the 'spaghetti' graph that just plotted individual lines for each model forecast. In a blog post about visualising data uncertainty, Krusz (2013) present a shaded alternative to error bars, which provides more information about the uncertainty distribution about the point. Examples of such plots are given in Figure 7.8. This idea could be extended to line graphs, with some interpolation to present uncertainty that may not be evenly distributed for all data points.

One recent advance in uncertainty presentation in mapping was the progression from stippling areas where models agree, to increasing the saturation of colour where there is more confidence. This technique was introduced in 2014 as part of the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP), and this method was published just after the IPCC 5th Assessment Report. Figure 7.9 shows the difference in clarity between the IPCC and the ISI-MIP techniques. Whilst the stippling and hatching in the IPCC map in Figure 7.9a gives additional information to the ISI-MIP saturation map, the hatching and the stippling

indicate degrees of uncertainty and confidence respectively (by assessing the multi-model mean change in relation to internal variability). The saturation map is much easier to read and the results are apparent even without the need to consult the caption. In the saturation map, the areas of colour with the deepest saturation stand out, such as the increase in runoff that can be seen in arctic regions and the decrease in runoff in the Mediterranean, whilst the areas that are less certain are shown in paler colours, which give the immediate impression of uncertainty. Another advantage of the saturation technique is that small localised areas of certainty can still be displayed, where a single dot using the stippling approach might be missed by the reader.

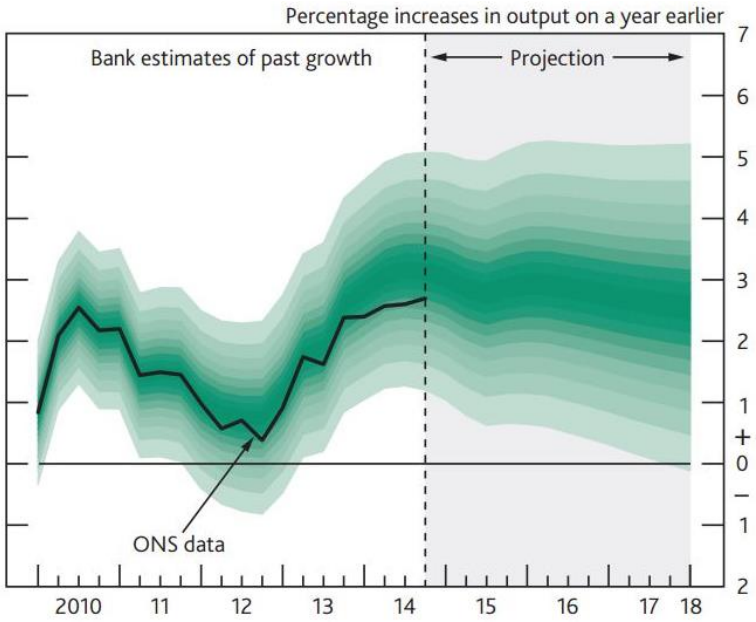


Figure 7.6 Fan Chart of GDP projections from the Bank of England (taken from Bank of England, 2015). *The distribution to the left of the vertical dashed line reflects the likelihood of revisions to the data over the past. Over the forecast period, the distribution reflects the uncertainty over the evolution of GDP growth. If economic circumstances identical to today's were to prevail on 100 occasions, the MPC's best collective judgement is that the GDP growth would lie within the darkest central band on only 10 of those occasions. In any particular quarter of the forecast period, GDP growth are expected to lie somewhere within the fan on 90 out of 100 occasions. And on the remaining 10 out of 100 occasions they can fall anywhere outside the coloured area of the fan chart. This has been depicted by the light grey background.*

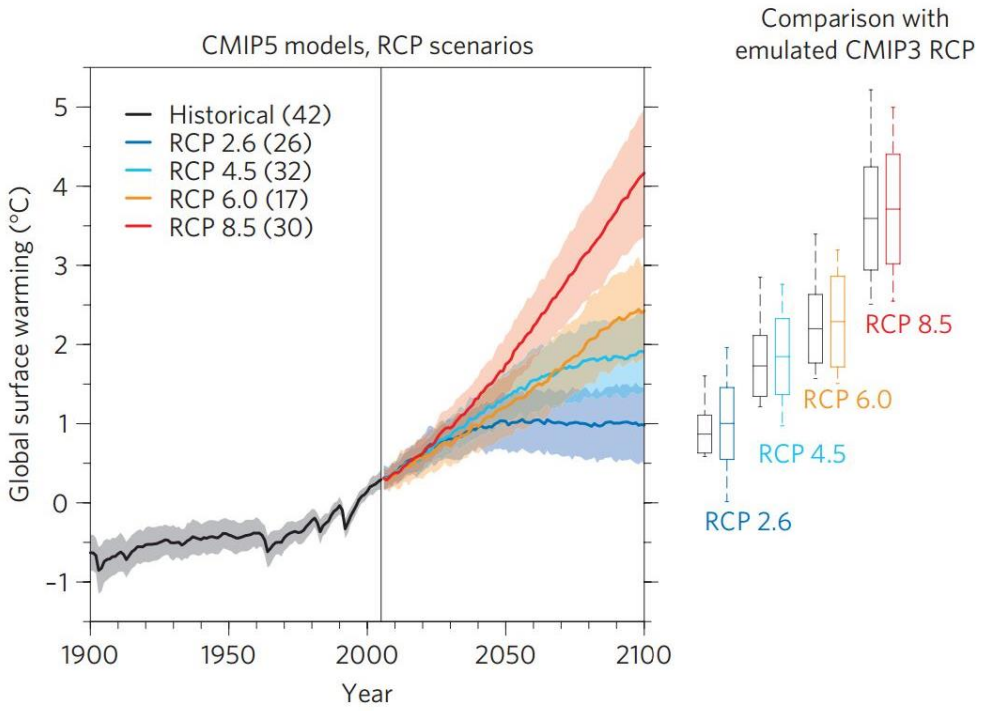


Figure 7.7 Fan and bar charts combined: Global temperature changes and uncertainty (taken from Knutti and Sedlacek, 2013). *Global temperature change (mean and one standard deviation as shading) relative to 1986-2005 for the RCP scenarios run by CMIP5. The number of models is given in bracket. The box plots (mean, one standard deviation, and minimum to maximum range) are given for 2080-2099 for CMIP5 (colours) and for the MAGICC model calibrated to 19 CMIP3 models (black), both running the RCP scenarios. Copyright licence granted.*

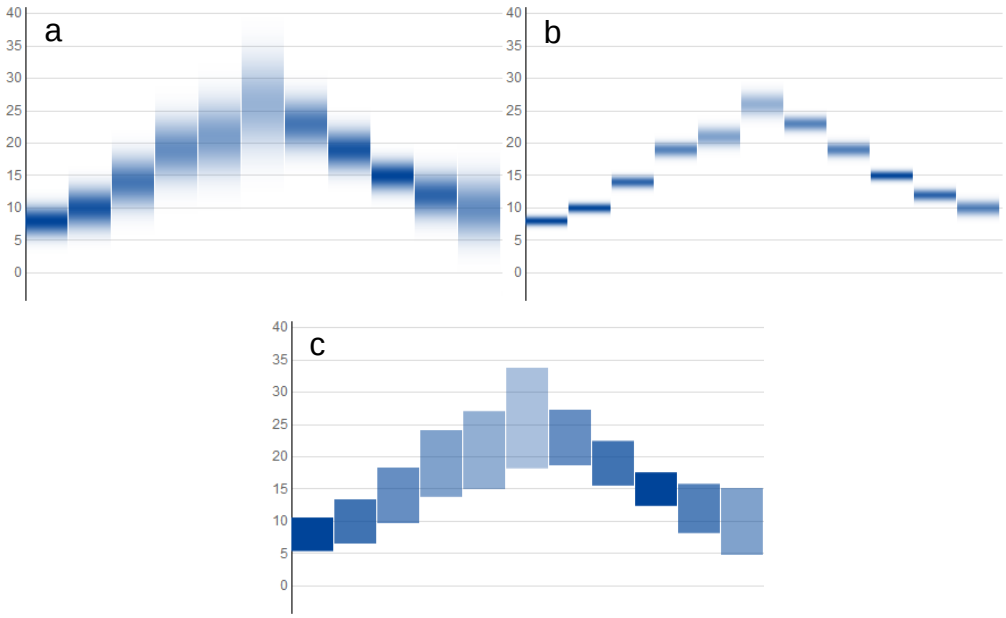


Figure 7.8 An alternative to error bars (Krusz, 2013). a) *Normally distributed uncertainty with high variance. b) Normally distributed uncertainty with lower variance. c) Uniformly distributed uncertainty.*

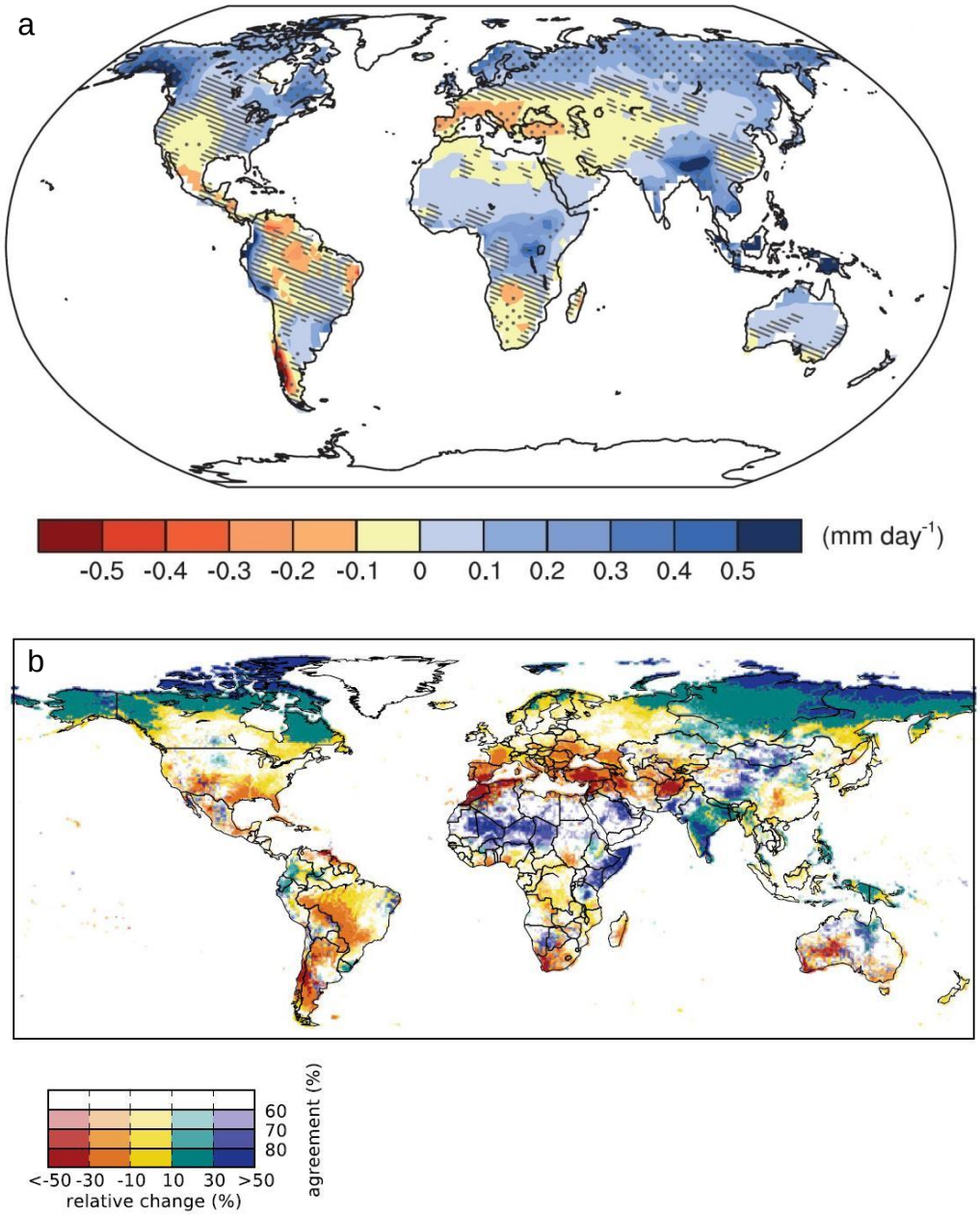


Figure 7.9 Presenting uncertainty in maps: stippling to saturation. *a) Change in annual mean runoff relative to the reference period 1986-2005 projected for 2081-2100 from the CMIP5 ensemble for RCP8.5. Hatching indicated regions where the multi-model mean change is less than one standard deviation of internal variability. Stippling indicates regions where the multi-model mean change is greater than two standard deviations of internal variability and where at least 90% of the models agree on the sign of change. (Collins et al., 2013) b) Relative change in annual discharge at 2°C compared with present day (1980-2010 average), under RCP8.5. Colour hues show the multi-model mean change, and saturation shows the agreement on the sign of change across all GHM-GCM combinations (percentage of model runs agreeing on the sign of change) (Schewe et al., 2014).*

One final consideration in the presentation of scientific data that is re-entering discussions is that of colour itself. Colour blindness affects between 5 and 10% of men and 0.5% of women, and red-green colour blindness (deuteranopia) is the most common (Hawkins, 2015) (see Figure 7.10). Discussions of colour blindness in science have been ongoing for more than a decade, and Light and Bartlein (2004) highlighted the perception of rainbow scales by viewers with protanopic vision (see Figure 7.11).

Presenting uncertainty is both an important challenge, and an opportunity for scientists. It is essential to be able to present information on the uncertainties inherent in modelling studies, whilst avoiding undermining the message of the research results. Producing a quickly interpretable graphic is often required in science communication with both peers and public, and it can be incredibly rewarding. Yet the complexity of many scientific studies make producing such graphics a significant challenge.

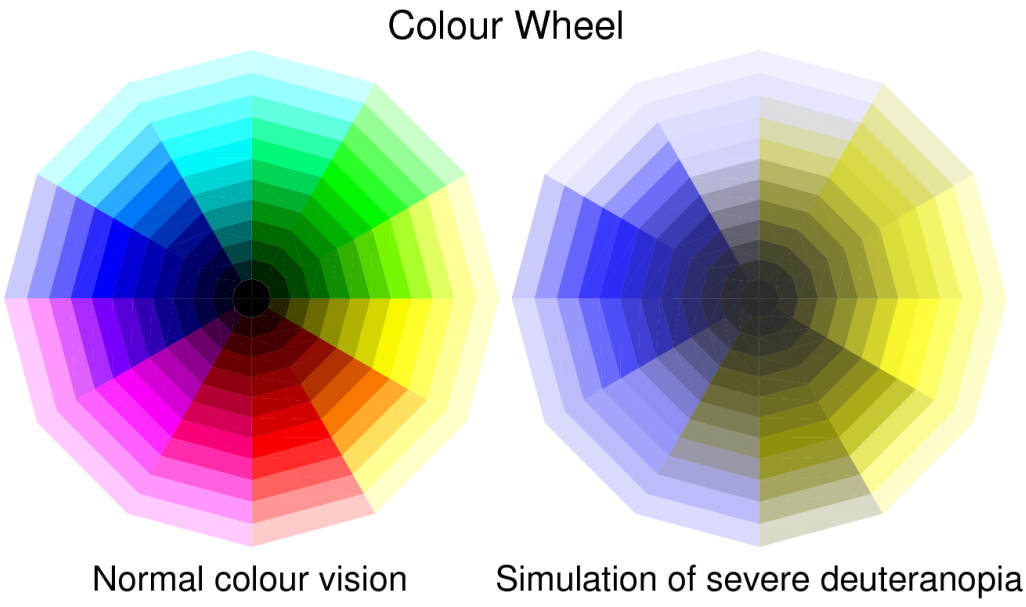


Figure 7.10 Simulation of deuteranopic vision of a colour wheel (taken from Hawkins, 2015).

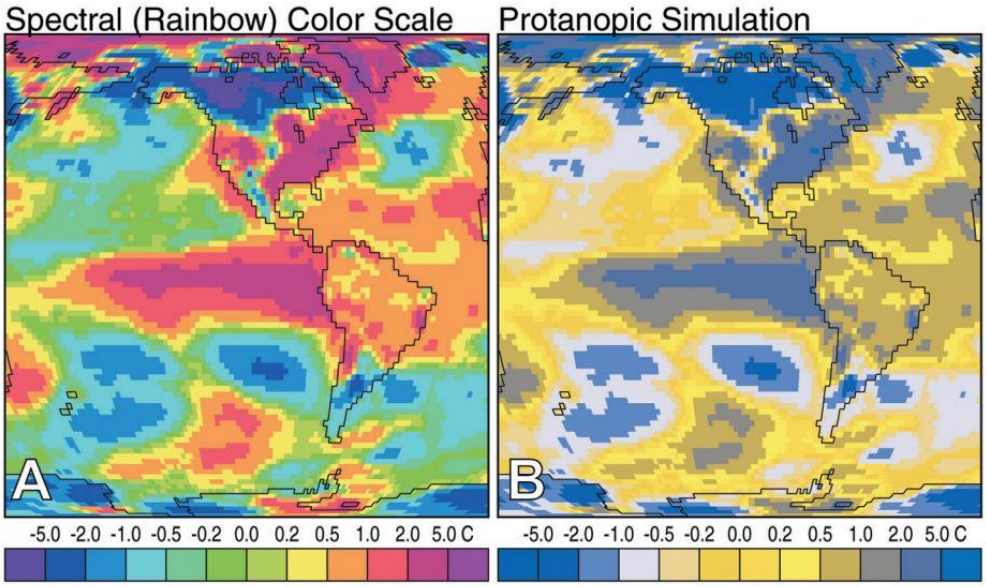


Figure 7.11 Simulation of protanopic vision of a rainbow scaled map. *Air temperature anomalies (1971-2000 mean) for January 1998, during an El Niño event (taken from Light and Bartlein, 2004).*

7.10 Deeper into the Roots of Uncertainty

Presenting uncertainty becomes more complex when different sources of uncertainty are taken into account. The cascade of uncertainties in the CMIP5 project presentation (see Figure 7.5) is a good start towards making the full range of model uncertainty clear, and helping the reader determine what levels of uncertainty are associated with each part of the modelling process. There are many more aspects of uncertainty that need to be addressed in modelling studies though, before it can be said that we have a full understanding of uncertainty in hydrological modelling. A few examples of these include the choice of climate model to use as climate input, the choice of potential evapotranspiration equation used, climate modelling techniques (e.g. downscaling methods, boundary conditions etc.), the choice and integrity of abstraction data if it is used, and routing methods that may be employed to simulate the translation of runoff to river discharge.

To return to the analogy of the tree of uncertainty, as presented in Chapter 1.5, these options and decisions along the path of a modelling

study can be seen as an incredibly dense network of roots at the base of a tree, where each choice and subsequent combination leads to a different leaf at the top of the tree. The huge variety of models available for both climate input and hydrology modelling, alongside the plethora of techniques of analysis and evaluation, provide hydrological scientists with a veritable forest of possibilities. Whilst all combinations cannot possibly be explored, comparative studies such as parameter experiments and multi-model ensembles help us understand the impacts of the choices that are made in scientific investigations.

Chapter Eight:

Conclusions

8.1 Introduction

Global hydrology modelling has advanced rapidly over the last two decades since its inception. Emerging into a research field of their own, global hydrology models offer large scale simulations of water resources that have not previously been possible. Many global hydrology models are now in widespread use, and whilst a few studies have investigated aspects of model uncertainty by comparing these models, the uncertainties within global hydrology models have remained un-investigated. This study aimed to "*address the issue of uncertainties within a global hydrology model by analysing parameter uncertainties"*. This was carried out using the Mac-PDM global hydrology model for a set of 21 large river catchments across the globe. Mac-PDM.09 was updated using recent land cover mapping products, and a more comprehensive classification of soil texture. The model update was renamed Mac-PDM.14, and this model was taken forward for use in addressing the aim of this thesis.

8.2 Research Questions

The three main research questions posed are reviewed here.

Research Question 1: How can uncertainties within global hydrology models be assessed and quantified?

Methods of investigating parameter uncertainty within global hydrology models were reviewed (Chapter 3). Simple techniques such as One-At-a-Time (OAT) sensitivity analysis were considered alongside more rigorous methods such as Generalised Likelihood Uncertainty Estimation (GLUE), and Bayesian Recursive Estimation (BaRE) which vary several parameters simultaneously. There are numerous techniques available, but all centre on the same principal steps:

1. choosing parameters to investigate,
2. sampling defined ranges of the parameters,

3. running the model with the sampled parameters,
4. comparing the model output(s) with the observed record, and
5. determining whether the set(s) of parameters is/are either acceptable or optimal

Some of the uncertainty analysis techniques are recursive, using 'hill climbing' methods to seek a better set of parameters using the knowledge gained from the previous model run(s) (e.g. Shuffled Complex Evolution Metropolis Algorithm). Others, such as GLUE, use random or near-random sampling techniques to run samples of the entire parameter space simultaneously, assess the model performance once all simulations have been completed, and then determine which, if any, of the parameter sets are acceptable. There remains debate in the literature between these techniques, as GLUE theorists suggest that there is unlikely to be only one optimal solution, and therefore hill climbing techniques may lead to a perceived optimal parameter set that may actually be matched, or indeed exceeded, elsewhere in the parameter space.

Three uncertainty techniques were employed in this study using the Mac-PDM model: OAT sensitivity analysis, GLUE, and Approximate Bayesian Rejection. The One-At-a-Time sensitivity analysis cannot quantify the parameter uncertainty of the model as a whole, as it cannot account for parameter interactions from non-independent parameters. However, sensitivity analysis can inform the modeller about the relative importance of a parameter within a model. In the Mac-PDM model, it was found that the soil parameters were the most sensitive, whilst the routing parameters were the least sensitive. This does provide some information on model uncertainty though, as it shows that in this case the soil parameters are the most uncertain, and therefore need to be carefully considered in the model calibration process.

The Generalised Likelihood Uncertainty Estimation experiment carried out on Mac-PDM.14, used a few alternative methods to the original

GLUE technique. 100,000 parameter perturbations were sampled using a Latin Hypercube Sampling technique. This sampling technique is more efficient than the Monte-Carlo random sampling that is traditionally used in GLUE analysis, and so allowed for fewer simulations to be run. This was important, as all 123 of Mac-PDM.14's parameters were included in this GLUE experiment. The parameters were sampled from prior distributions which were determined from the literature using the @RISK software package. The Mean Absolute Relative Error (MARE) measure was used to assess the models performance with each parameter set.

From the 100,000 model runs, the best parameter set gave a MARE of 0.9, or an average error of 90% over all 21 catchments, but when the Murray Darling and Nile catchment were excluded, where the model performed particularly badly, the best model parameterisation had a MARE of 0.47 or 47%. The previous version of Mac-PDM, before the GLUE experiment, and before the update of the land cover and soil maps (Mac-PDM.09) scored a MARE of 1.04 (104%) over the 19 study catchments excluding the Murray Darling and the Nile. This shows that not only can a GLUE experiment assess and quantify the uncertainty of a model, it can also significantly improve the models performance.

Of the 100,000 model runs, 2 parameterisations scored a MARE <0.5, 1,238 scored a MARE <0.75, and 25,532 scored a MARE <1. The ranges of the model outputs from these parameterisations were plotted against the model results from 9 Global Hydrology Models (GHMs) from a multi-model ensemble (MME) - the EU-WATCH project. This demonstrated that the range of model outputs with a MARE <0.75 was comparable to the range from the choice of GHM, demonstrating that parameter uncertainty can easily be as large as structural uncertainty, but that it depends on the limits of acceptability adopted by the modeller (if only models with a MARE<0.5 were determined as acceptable, then the range of model output is much smaller than that of the MME). An

additional important result here, was that the MARE of the top performing Mac-PDM.14 model parameterisation was lower than that of all of the GHMs that participated in the EU-WATCH project, whilst the previous version of the model (Mac-PDM.09) ranked 6th out of 9 in the MARE scores of the GHMs.

The Approximate Bayesian Rejection method was used to determine whether any trends could be found in the parameter values of the top ranking model runs. The top 1,000 of the 100,000 model parameterisations (the top 1%) were used in this investigation. The prior distribution of the parameters were compared to the posterior distribution of the 1,000 top model parameterisations.

This technique was applied to determine whether any trends could be seen in the catchment specific results, and to ascertain whether there were groups of climatically similar catchments showing trends towards similar parameter values. Some differences between catchments were identified for a few of the model parameters, but the results could not easily be attributed to climatic regime, latitude or hydrological characteristics. Many of the parameters appeared to be insensitive to perturbations, which suggested that the Mac-PDM model is over-parameterised.

Mac-PDM.14 was then validated using an alternative input dataset from another MME experiment (ISI-MIP). This was an important exercise, because it would conclude whether the extensive uncertainty estimation experiment, and subsequent calibration of the model, would remain an improvement on the model performance when applied to a different modelling exercise. The results showed that the top performing model parameterisation from the 100,000 GLUE simulations scored better when run with the alternative (ISI-MIP) input data than the calibration (WATCH) input data, with MARE scores of 0.45 and 0.47 respectively. Again, Mac-PDM.14 scored better than all of the other models in the

ISI-MIP ensemble, when Mac-PDM.09 came 6th out of the 10 MMEs in a ranking of MARE scores.

These findings showed that alongside uncertainty assessment and quantification, significant improvements in model performance can be made. However, there remains more work to be done on understanding how and why these sets of identified parameter values improve the models performance so significantly.

The potential improvement in a models outputs from an exercise such as the one carried out in this thesis provides a great incentive for modellers to consider this type of assessment when employing their models.

Research Question 2: What is the feasibility of including rigorous uncertainty estimation experiments in the global hydrology model calibration process?

This research question was addressed (in Chapter 4.6) using a set of GLUE experiments of different sizes. Due to the large number (123) of model parameters being considered in this study, 100,000 model parameterisations were run in the main GLUE experiment that was used in the various analyses for the main aim of this thesis. However, this required significant computational resources. GLUE ensembles of 10,000, 5,000, and 1,000 model parameterisations were also undertaken. The same sampling method was employed for each, and MARE was used to assess the model outputs over the 19 study catchments.

The distribution of the MARE scores across the range of parameterisations was very similar between the different sample sized experiments, which demonstrated that with a smaller sample size, the modeller is not less likely to obtain a ‘good’ model parameterisation, but he may obtain fewer ‘good’ model parameterisations. For example, in the 100,000 sample experiment, 1,238 model parameterisations scored

a MARE <=0.7, whilst for the 1,000 sample experiment, only 6 parameterisations scored <=0.7. This indicates that if a modeller were happy with a model with a MARE of 0.7, then perhaps only 1,000 model parameter perturbations might suffice to achieve a good model. However, for a MARE <=0.5, 2 models parameterisations from the 100,000 sample ensemble achieved this score, whilst no parameterisations from the 10,000, 5,000 or 1,000 sample experiments could. Furthermore, if a modeller was trying to consider the uncertainty in their model parameters, the ranges of model outputs from the models scoring a MARE less than a certain threshold is significantly affected by the experiment sample size. Whilst the top ranking model output is very similar, the range of outputs from models scoring <0.75 is much larger from a sample size of 100,000, than it is for a sample size of 1,000. This might give modellers that use a small sample sized experiment false hope that the range of their uncertainty is smaller than a modelling experiment that used a larger sample size.

The implication of this is that a modeller should carry out as many parameter perturbations as they can afford, in order to get as good a grasp on the true range of the parameter uncertainty of their model. Valuable further research could be carried out (ideally using a model with fewer parameters), to determine whether the range of model outputs converges with increasing sample size.

Access to high performance computing is rapidly increasing, so investigations such as this, should be accessible to most modellers globally.

Research Question 3: To what extent are "global" hydrology models fit for purpose?

Global hydrology models have been available since 1989, with an increase in focus and complexity during the 21st century. Model

evaluations that have been carried out on catchment models have not yet been widely applied to global models.

How can models be evaluated and validated?

Evaluation of the uncertainty in global hydrology models was considered above. In order to be calibrated, observed data is usually required, and in this thesis observed discharge records for the chosen study sights were obtained from the Global Runoff Data Centre, the Bangladesh Water Development Board, and the United States Geological Survey. This data allowed the model to be calibrated 'globally' using an average of the Mean Absolute Error of the models outputs across these catchments. However, data is not available for all catchments worldwide, so the model's performance in other catchments may not be optimal. A global calibration of the model (see chapter 5) also requires a trade-off in the results of each catchment, and it was found in this model that the Amazon catchment performed badly as a result of the model calibration. Model validation can be carried out by examining the models performance over a different time period, or, as here, by applying the model using a different input climate dataset. In this study, Mac-PDM.14 was found to perform very well with an alternate input dataset, thus endorsing the results of the model calibration.

How do global hydrology models perform in a catchment context?

Global calibrations of Mac-PDM.14 have been evaluated in a catchment context (chapter 4), whilst the potential application of catchment specific calibrations of the model have been considered (chapters 5-6). The top 20 global model calibrations from the 100,000 model parameterisation experiment were compared to the top 20 catchment specific calibrations. The catchment specific calibrations were shown to provide significant improvement in the model performance in all study catchments. In several catchments where the global model calibrations

performed well, the range of the top 20 calibrations was significantly reduced from the catchment calibrations compared to the range of the global calibrations, which shows a reduction in uncertainty in catchment calibrated model simulations. The natural variability of the catchments was taken into account using the observed record, which further confirmed the merits of employing catchment calibrations of the model.

This is a novel concept in global hydrology modelling, as the aim of a global model is to enable the model to be employed worldwide without the need to recalibrate to each catchment. However, global models are regularly applied to catchment scale research problems, particularly in large 'global scale' catchments, and the significant improvement in the model performance with catchment specific calibration demonstrates the potential merit of this approach. In this study, just 21 catchments were considered, however the 100,000 model runs used in this experiment were run at the global scale and the catchment average runoff was extracted. Therefore, with the retention of the model outputs from the 100,000 parameterisation ensemble, any catchment where observed data can be obtained, can be calibrated very quickly indeed. This would be significantly faster than calibrating a catchment model, and the results would be better than the globally calibrated model realisation. Where observed data cannot be found however, the global calibration can be applied, and from this experiment, the global calibration of Mac-PDM.14 is known to perform significantly better than the previous version of the model (Mac-PDM.09), which is due to the rigorous calibration procedure employed. This global calibration could also be iteratively updated as more catchments are added to the repertoire of the models applications, covering more areas of the globe. Where truly global scale applications of the model are required, there could be the potential to 'stitch together' the catchment calibrations of the model, where they are available, and the global calibration of the model where it is not. This would alleviate the issue of the reduced model performance in catchments such as the Amazon under the global

calibration of the model due to the trade-off involved between catchments in the global model calibration process.

Are the uncertainties in global hydrology models acceptable?

Whether the uncertainties in global hydrology models are acceptable depends upon how the models are used. Currently, the results of global hydrology models are mostly used within the research community, investigating global water resources and potential implications on society. GHMs are featured in the significant governmental reports of the IPCC, and uncertainty is carefully considered in these reports. However, the uncertainty that is presented to date is established from multi-model ensembles, and does not consider the uncertainty within a GHM, as has been addressed in this thesis. This study has shown that parameter uncertainty can be as significant as model uncertainty, so this is an important issue in GHM presentation. There is no definition of what level of uncertainty is acceptable, though having several different GHMs agree on a model result increases confidence. Some definition of level of confidence within each model included in a multi-model ensemble needs to be considered. The graphs in this thesis, that compared a parameter ensemble with the models from the WATCH and the ISI-MIP MMEs, is a step towards this. They highlight that following the rigorous uncertainty experiment and subsequent calibration of the Mac-PDM model, the model performs significantly better than any other model in each of the ensembles. If each of the participating GHMs carried out this sort of experiment, the results of the MME as a whole could be dramatically improved.

8.3 Further Research

The findings of this study propose that global hydrology models undergo much more rigorous calibration and uncertainty estimation before they are employed in multi-model ensembles. It is important to first understand the uncertainties within a model before the

uncertainties between models are evaluated. However, even this level of uncertainty analysis could not address all of the issues that arose in the study. The cause of the improvements in model performance with certain model parameterisations was not clear, which could have been the result of over-parameterisation of the model. Further research in this area in the form of model emulation would be valuable. Employing a catchment calibration of a global hydrology model has been suggested, however it would be an important investigation to compare the performance of a catchment model with a catchment calibrated global model. Presenting uncertainty is also an important issue in all scientific fields, and discussions with the end users of hydrological models over the understanding and application of uncertainties in decision making exercises would be extremely valuable.

This study has demonstrated the benefits of carrying out an uncertainty experiment and calibration of global hydrology models to a level that has not been previously considered. This process has improved the Mac-PDM model, from performing centrally within the range of current GHMs, to performing significantly better than the others. The cascade of uncertainty (see chapter 7) is currently missing one of the levels in the cascade in its application to global hydrological science. Research methods surrounding the progression of uncertainty analyses in catchment hydrological research could be explored to bridge this gap in the cascade. The uncertainties in global hydrological models need to be better understood, and the methods of assessing uncertainties applied here have potential to improve the integrity of global hydrological models.

Appendix

Tables A.1-A.9 Mac-PDM.14 Parameter values for the top model calibration for each catchment, including a global model calibration.
Catchment abbreviations are detailed in figure 3.3, land cover classification codes are detailed under table 4.5. Soil classifications read as follows: Sa=Sand, LoSa=Loamy Sand, SaLo=Sandy Loam, Lo=Loam, SiLo=Silt Loam, Si=Silt, ClLo=Clay Loam, SaClLo=Sandy Clay Loam, SiClLo=Silty Clay Loam, SaCl=Sandy Clay, SiCl=Silty Clay, Cl=Clay, Li=Lithosols, and Hi=Histosols (Organic).

T A.1	*b*	*delta*	*grout*	*srout*	*xmelt*
Glob	0.14016	0.45757	0.44591	0.78633	2.90955
Ama	1.68783	0.33573	63.48982	0.01869	3.21013
AmuD	0.52899	0.01592	203.38471	0.01941	0.06829
Brah	0.10751	0.00414	0.04708	2.53721	2.74085
Con	0.90732	0.84094	82.71066	0.00022	4.61507
Dan	0.11255	0.28031	9.62159	2.86746	2.16181
Euph	1.80536	0.47914	136.14668	0.00608	3.24252
Gan	0.11821	0.36105	7.72376	0.12504	0.70960
Koly	2.27161	0.02000	5.59315	2.31960	1.90494
LaPla	0.53914	0.42641	185.79728	0.01029	5.01949
Lena	0.16351	0.01748	5.19359	2.00554	0.50530
Mek	0.14863	0.72693	0.17829	1.89923	2.81414
Miss	0.24322	0.134522	110.3278	0.04895	5.96656
MurD	0.15243	0.53826	42.48102	0.00012	2.65001
Nig	0.11975	0.49909	0.94046	0.79454	4.44280
Nile	0.10190	0.42713	45.78040	2.70914	1.01223
Ob	0.36829	0.59693	107.19559	0.00907	0.13351
Oka	0.13597	0.38871	13.53619	2.02728	3.68800
Yang	0.14146	0.07034	179.85346	0.03053	3.09886
Yell	0.11964	1.15314	0.99580	2.40614	3.69460
Yen	1.69586	0.05561	18.46418	0.00270	3.00990
Yuk	3.58583	0.01590	180.39943	0.03065	1.16364

T A.2	Fcpc													
	Sa	*LoSa*	*SaLo*	*Lo*	*SiLo*	*Si*	*ClLo*	*SaClLo*	*SiClLo*	*SaCl*	*SiCl*	*Cl*	*Li*	*Hi*
Glob	18.2848	11.0300	13.2277	29.4653	33.8388	19.3153	15.9231	20.2731	38.3983	21.9400	36.1485	38.6093	39.8282	26.3327
Ama	13.1976	18.3189	23.3683	32.6310	32.1874	20.1431	41.5249	36.2733	42.7508	45.3571	41.8309	35.8731	19.5188	46.9997
AmuD	11.2804	13.3341	34.1583	21.7604	34.7608	32.6394	32.4004	39.3057	38.8676	40.7585	34.8443	40.0980	36.1650	73.5274
Brah	12.1799	14.9873	15.2278	31.4647	27.1187	19.6841	31.3475	36.6085	31.0728	26.3367	32.4336	40.5218	28.5339	57.0493
Con	13.4908	17.0278	21.8362	34.2173	44.2073	20.6245	33.2451	30.7913	43.6748	35.8128	33.7699	41.4331	40.0027	61.3527
Dan	8.14447	16.7105	29.2522	24.0931	30.9954	19.5184	31.0618	36.5397	32.4805	38.2985	37.1930	43.9906	18.3281	53.4246
Euph	15.7382	14.2435	40.6456	31.5399	26.0100	26.5710	26.5352	34.0090	30.8135	43.9576	43.2694	43.8992	39.3002	43.3628
Gan	20.8271	31.9378	35.7097	20.0298	39.3647	43.0371	24.6629	39.5770	36.1532	25.0799	39.7758	43.6599	37.8180	34.2312
Koly	9.21702	17.1926	17.1926	29.2190	35.5197	35.2436	33.5793	26.0785	34.0678	25.7133	41.8737	43.0930	14.1598	49.2532
LaPla	18.0046	34.6554	34.6554	29.8006	32.3544	32.9967	41.5659	18.2992	32.8408	21.6556	37.3355	39.4482	26.2078	44.8120
Lena	17.1983	14.5968	14.5968	24.9577	32.0765	23.3245	33.7794	17.7819	42.1042	47.6479	46.1268	38.8238	33.2339	67.1560
Mek	10.5617	35.1261	35.1261	22.2643	28.8271	47.3211	33.9687	21.8402	37.5772	21.9005	38.1765	38.4602	24.0528	65.9827
Miss	10.4195	27.5593	27.5593	25.5101	35.9142	41.0750	36.2234	25.9313	35.4712	40.9392	41.6386	40.4233	24.0294	39.3151
MurD	14.3607	25.4499	25.4499	34.8221	31.9935	32.9901	28.6028	22.0482	38.5910	47.2851	34.1149	38.5734	38.6844	30.3520
Nig	10.5953	21.1717	21.1717	27.7658	29.8351	37.6374	15.0518	17.3238	35.4117	43.3063	39.2083	36.3057	39.4246	54.5969
Nile	10.5913	16.6296	35.8280	20.8973	18.9318	24.5727	47.2233	30.4328	38.7359	32.5504	39.2083	39.9764	35.2957	60.9101
Ob	21.4051	26.7898	49.4849	28.7514	18.4890	30.6513	32.4229	31.0243	40.6110	19.8608	38.2862	33.3810	24.5507	67.4393
Oka	29.3227	12.3034	16.8742	27.0943	22.3155	25.9762	35.7322	20.6331	40.8458	28.1239	41.1621	41.8307	20.6072	63.8681
Yang	10.6221	11.9407	15.9144	30.9871	37.7603	41.1435	38.3510	32.5847	36.6825	20.4647	38.841	38.4758	39.0619	45.8757
Yell	16.3452	14.3118	42.5013	29.8811	37.1567	42.3272	19.9992	32.9246	38.0053	39.7968	44.0565	39.9061	39.0446	52.4096
Yen	19.6774	14.9544	21.8602	21.5346	33.1655	32.7835	20.3323	31.7908	28.4155	41.4656	38.5210	38.1925	37.5055	53.7492
Yuk	8.47127	11.0604	28.2366	30.3072	31.2743	24.5263	39.6326	37.1143	42.0492	38.1658	41.4541	39.6831	33.4604	35.8577

T A.3	Satpc													
	Sa	***LoSa***	***SaLo***	***Lo***	***SiLo***	***Si***	***ClLo***	***SaClLo***	***SiClLo***	***SaCl***	***SiCl***	***Cl***	***Li***	***Hi***
Glob	30.8078	46.4010	43.8847	56.5867	45.3690	30.0916	51.5986	41.4113	47.8537	36.5115	67.3125	43.5797	54.8484	121.303
Ama	41.3190	57.3957	43.1130	51.1734	49.5754	33.2810	46.9290	30.3976	41.6034	47.6088	66.4606	34.9440	27.2931	124.520
AmuD	42.0365	32.4506	46.0865	41.4076	48.1316	42.6990	44.7789	38.3580	44.8004	40.6269	42.7383	47.3574	35.5033	130.857
Brah	37.1628	35.2855	37.4285	50.6478	46.4054	70.5123	40.5445	41.3185	46.3810	41.7433	67.2927	51.6410	34.1465	78.4272
Con	37.0939	56.7155	39.5956	60.2964	45.8157	37.8081	54.0535	31.3611	34.8583	45.9597	65.6791	48.6937	40.4467	116.934
Dan	37.0939	37.5626	47.5702	49.0109	48.4941	56.4896	58.6702	49.1485	35.9507	43.4778	51.4550	50.4314	34.9719	99.4160
Euph	41.6127	46.2527	40.9469	37.2146	42.2221	34.5224	47.5838	31.5403	45.1392	55.6350	34.0675	47.3805	25.3341	77.9416
Gan	43.2970	44.8208	39.7323	31.2423	47.8833	37.4442	46.4632	37.4402	47.6527	47.3395	45.6370	48.2573	52.6260	115.362
Koly	37.0841	53.3901	44.2648	59.4634	46.6477	32.4034	48.9896	42.1597	39.6882	33.7179	36.6033	47.5129	43.1776	129.285
LaPla	45.5406	48.5104	39.6481	53.5363	42.0846	38.3594	47.6213	52.7775	64.4306	50.8255	49.6631	50.5180	70.5421	129.381
Lena	31.4052	42.7484	44.3356	39.6551	46.2019	52.4489	44.8781	47.3781	41.9407	57.1244	66.7533	49.8808	65.9546	77.4533
Mek	39.7362	42.7991	49.4682	42.7121	47.6990	31.5688	54.6272	42.3781	57.9588	31.8639	47.2622	49.8198	49.8802	76.5826
Miss	41.9445	55.9185	38.3083	34.6251	46.4518	36.7548	48.0693	36.4739	48.4530	28.4904	38.9019	47.9917	45.8396	127.854
MurD	28.2990	48.7164	43.7503	50.4890	44.1933	69.0266	49.4697	50.8133	55.7250	34.6603	47.5454	59.5810	36.8040	84.6401
Nig	48.5643	58.2269	33.5932	39.7012	50.1818	40.7153	60.0028	41.5298	55.0098	29.3041	56.3971	49.5431	73.9453	121.026
Nile	32.2276	49.3954	44.2095	58.4566	47.3409	45.0366	47.2401	35.4987	53.3391	35.4115	49.6436	52.1348	35.3024	106.202
Ob	48.8578	56.1590	41.2041	43.7957	43.1753	38.9612	53.1681	36.5062	51.9099	33.1395	59.4153	49.7527	53.8925	127.957
Oka	48.6732	35.5098	45.3807	36.1762	44.6442	53.7429	41.0472	41.2922	42.3044	44.9216	66.3340	43.1208	46.6668	131.731
Yang	40.7190	55.3034	44.2817	46.1839	49.2617	39.9576	49.5611	40.5669	33.7321	35.1653	69.9045	39.4644	33.4859	120.585
Yell	48.0206	37.7934	39.6196	36.6239	41.3589	62.6916	44.5424	43.8100	49.8787	32.3101	33.0007	48.5980	64.7707	113.318
Yen	50.6794	38.8799	44.9628	50.4807	47.6833	59.0434	51.2754	31.6590	69.9661	41.4242	40.2439	50.3110	66.0592	119.946
Yuk	39.0677	45.4898	42.7048	49.9499	47.9463	64.6466	49.0911	43.2973	51.2135	29.5123	48.3817	46.4825	60.5717	62.9116

T A.4	Rootg														
	EN	*EB*	*DN*	*DB*	*MF*	*Mtv*	*Mcv*	*Sh*	*Gr*	*Cr*	*SpV*	*Ba*	*Bfl*	*Vfl*	*AA*
Glob	3.5030	3.6055	2.5047	2.5047	6.1802	3.5468	2.6765	1.4931	2.3239	1.0836	3.2876	1.4177	3.3504	4.2825	0.8347
Ama	0.9565	1.4154	2.9038	2.6907	0.8747	1.1308	0.8486	0.9052	1.3160	2.8012	0.0344	1.7625	7.6495	1.1791	0.6039
AmuD	1.3294	4.1870	2.1755	1.3531	11.161	1.6323	2.918	1.4249	1.5168	1.0591	3.4173	1.0888	5.5314	4.0402	1.1817
Brah	2.1015	2.1009	2.0182	0.7716	12.591	7.2320	1.4629	0.7222	0.4867	1.0636	0.9586	0.5334	4.8450	4.6370	.8597
Con	1.5151	2.1088	1.4199	2.0954	1.4486	3.3301	1.1444	2.1358	0.3186	0.7109	1.3256	1.4835	5.5873	3.3058	0.9010
Dan	2.4928	5.1687	1.7430	1.7328	4.3316	6.7266	2.2679	4.2474	1.1372	1.6229	1.5044	0.5274	14.098	4.7065	0.7559
Euph	2.2353	1.4525	0.9747	0.6075	1.8323	2.6518	4.3004	4.9025	1.0501	2.0329	1.9988	1.1697	6.7405	3.4591	0.4003
Gan	2.5122	3.7016	4.0450	1.1825	9.9085	2.6812	0.8093	3.6597	0.7798	1.4548	1.0325	1.6125	1.5203	3.6816	1.0952
Koly	0.9286	1.4354	1.2425	0.8570	5.5942	3.7961	1.1294	0.3447	1.6705	1.2017	0.4532	1.9438	5.3877	0.8807	0.4664
LaPla	2.7763	1.2561	1.6738	2.0058	2.0645	6.8530	1.0218	4.3720	0.8272	1.7996	3.2342	1.8086	11.122	3.5953	0.5678
Lena	2.0569	1.2895	1.1518	2.6358	8.9105	0.9733	4.2206	3.7823	0.6242	2.3539	3.6171	0.5478	11.948	3.7760	0.6886
Mek	1.5766	3.0663	0.4953	5.6954	7.1038	1.9292	4.1386	2.7837	1.9580	1.1246	3.0774	0.7377	7.7235	0.8284	0.9758
Miss	1.3787	4.5999	1.4384	3.0221	5.6386	8.1223	3.2248	0.4055	1.5593	1.4003	2.0203	1.0785	1.5784	2.5807	1.1359
MurD	1.3959	1.9291	0.5021	1.2162	2.0092	2.2713	1.5465	1.1399	1.5987	1.0779	3.7800	0.7414	5.0184	3.8478	0.4034
Nig	4.4790	3.1347	1.5971	1.8060	7.0143	2.9760	4.5568	0.3594	1.7023	0.8182	1.8330	0.9911	7.0803	4.7830	1.1467
Nile	1.0977	1.3368	1.8584	2.6170	3.7600	5.4507	2.6969	3.0752	1.8100	1.7903	3.1394	1.8850	6.0212	0.3131	0.4793
Ob	1.8819	3.8107	0.5871	1.4077	6.6253	5.5980	2.3112	2.8050	1.2790	1.1979	1.4760	0.5575	1.0375	4.3232	1.0480
Oka	1.5138	1.3067	0.9979	1.7372	8.3150	1.0299	0.4561	4.8110	2.8022	2.3764	1.6287	1.2634	7.1072	0.8220	0.9232
Yang	0.6523	2.5669	1.4054	2.2726	1.9824	5.8608	2.1534	1.8557	1.1620	1.1469	0.7508	1.0186	6.8251	2.1692	1.0346
Yell	1.1788	4.2883	1.6360	1.2576	8.0657	6.6916	2.2106	1.0121	0.4820	2.4848	1.6407	0.2513	6.9570	3.0131	0.4288
Yen	1.8146	2.7111	0.5876	2.0540	2.2304	2.0140	2.8058	1.3233	1.3009	0.9093	1.9034	1.8400	6.2262	1.3055	1.0710
Yuk	1.8533	0.3511	5.4990	3.0768	3.0491	7.5502	4.8982	4.5990	1.3591	1.5402	1.7787	0.3053	7.6052	2.0009	0.5206

T A.5

	Rsc														
	EN	***EB***	***DN***	***DB***	***MF***	***Mtv***	***Mcv***	***Sh***	***Gr***	***Cr***	***SpV***	***Ba***	***Bfl***	***Vfl***	***AA***
Glob	5070.3	1890.1	135.31	177.51	2628.2	3388.4	1520.4	482.37	167.34	191.81	4866.8	611.67	856.88	1635.5	2124.2
Ama	462.09	591.31	313.30	148.98	3163.7	2208.6	4349.0	173.78	327.24	415.46	2383.1	4178.1	1816.9	639.33	2572.2
AmuD	1158.7	752.92	962.46	179.49	1016.3	1101.1	1052.1	296.53	391.27	323.02	1192.0	1558.7	3763.9	2739.6	1831.1
Brah	576.76	3461.0	332.59	335.17	2874.0	4258.3	1830.7	1934.3	235.6	75.563	447.10	3377.3	1645.5	1881.0	1625.8
Con	504.31	652.91	94.440	315.82	1551.8	541.55	582.25	93.593	222.82	80.778	1057.3	847.37	2180.1	1946.0	1132.3
Dan	211.69	736.82	526.48	139.54	3066.6	1797.1	422.27	459.04	203.75	55.365	4650.5	2307.8	978.03	2487.5	3019.5
Euph	342.79	436.63	207.83	263.66	271.67	1448.8	4054.7	630.61	86.417	200.31	1787.3	3643.0	2238.9	2476.7	861.03
Gan	476.02	383.67	267.45	182.27	4341.2	208.74	311.83	1575.5	110.05	49.045	4679.5	3865.2	995.57	865.48	802.70
Koly	1560.5	1857.1	191.24	400.37	782.75	2900.6	1229.0	35.939	219.22	139.37	2661.5	4733.4	876.20	3842.1	293.92
LaPla	296.95	381.85	461.73	197.99	3335.0	3167.5	1903.5	683.53	107.43	275.94	2267.3	1539.9	496.22	174.42	1595.0
Lena	644.89	812.73	275.72	259.86	1206.1	2056.1	518.79	251.30	263.70	293.42	2071.4	1610.4	1769.6	3294.8	1202.4
Mek	850.24	331.75	428.89	779.08	2944.4	2203.7	3251.2	84.639	36.116	170.01	2861.2	912.28	1054.9	2524.9	444.48
Miss	831.59	270.40	194.60	784.80	2610.9	671.80	3308.7	45.458	39.502	313.26	1232.4	1659.4	5041.8	2513.2	818.81
MurD	659.57	159.72	90.153	538.12	1981.7	2107.5	1845.5	685.43	125.83	526.12	415.96	3232.6	2763.5	877.19	1811.3
Nig	655.41	620.15	386.15	330.28	1877.1	3116.9	775.96	360.74	133.28	153.37	2295.6	3317.1	1163.2	1230.5	755.82
Nile	477.40	642.70	359.54	117.70	2024.3	2247.5	2863.3	579.74	64.825	212.49	499.01	4179.3	576.8	3982.0	1887.6
Ob	1366.8	394.46	583.59	808.13	2419.1	3231.2	1435.3	285.16	624.64	297.39	945.22	3352.2	437.67	2112.0	2095.7
Oka	346.11	138.26	306.46	256.71	821.03	1009.2	1724.9	319.34	25.974	674.78	3814.6	2648.2	1867.2	1003.7	349.36
Yang	74.279	657.63	244.94	244.23	3600.2	3273.8	1316.7	54.311	375.39	121.48	580.44	2085.4	507.43	1298.5	3817.5
Yell	787.02	568.37	267.25	94.558	1761.7	3389.0	3843.8	94.787	36.156	346.33	3656.6	740.88	3378.9	757.95	246.82
Yen	169.31	1951.0	626.50	202.01	4324.1	2883.0	2415.5	202.35	167.19	328.74	804.31	631.82	441.75	2249.6	589.88
Yuk	1157.6	258.76	713.13	604.57	628.42	1944.5	2523.3	305.93	1290.1	89.611	4606.2	2333.8	3040.5	1270.8	1353.5

T A.6

	Capg														
	EN	***EB***	***DN***	***DB***	***MF***	***Mtv***	***Mcv***	***Sh***	***Gr***	***Cr***	***SpV***	***Ba***	***Bfl***	***Vfl***	***AA***
Glob	5.8757	6.2889	5.3934	4.5561	2.2306	8.1428	6.3831	10.981	2.6516	5.3171	0.4706	3.8177	4.4509	4.3082	1.9173
Ama	4.3461	4.6089	2.6387	5.1364	3.6957	7.9350	1.0034	9.6929	1.6562	6.6657	4.7048	2.5002	4.2865	4.9489	4.5562
AmuD	4.2364	5.2987	5.5012	6.5664	2.7338	5.8442	1.2529	0.4358	0.7849	3.6292	5.3222	2.396	5.3405	4.9966	1.5390
Brah	3.8302	3.8145	6.4002	5.2219	8.8118	5.4386	0.8483	10.410	0.6330	2.9508	0.6268	1.2197	7.4220	7.8714	4.6392
Con	2.9472	6.0006	6.7851	6.4288	5.8986	6.9402	5.4161	8.0627	1.9741	4.4930	0.9330	0.6115	5.2492	1.0718	1.0331
Dan	3.4074	5.7846	6.9659	6.7744	5.9586	6.5031	1.8050	10.944	0.0113	3.3723	1.6288	1.9339	2.4112	6.1516	2.6116
Euph	6.5419	6.3006	7.7890	6.5853	2.7543	9.9879	6.8752	6.3501	2.2858	0.6053	0.5294	3.2203	2.3669	4.3096	1.5415
Gan	7.6996	7.3371	4.6788	4.4239	3.7890	5.7605	7.9210	5.2351	3.0829	7.5186	0.7817	4.1276	7.1014	2.3825	4.6200
Koly	5.8905	5.9027	4.4631	5.5909	7.9315	8.4808	1.2321	10.896	4.9572	5.4789	5.5623	0.2393	2.5908	4.5651	0.8324
LaPla	4.3926	8.8054	5.4534	4.3929	6.5475	1.0834	5.1477	3.7402	3.0370	6.3662	5.5551	3.1276	7.7906	5.1322	2.4649
Lena	4.4681	7.4699	7.1729	7.2477	7.5152	0.2373	4.9992	8.0651	0.8041	4.4226	5.0358	0.0696	0.7805	0.5023	0.6788
Mek	2.0342	6.2223	6.7607	5.4424	8.4145	0.7012	4.8459	2.9031	1.8932	0.4940	0.7950	4.5983	4.7234	2.1839	4.4388
Miss	6.1203	5.9218	6.0669	7.5680	5.0390	9.5475	2.9020	3.4414	3.1556	0.9906	4.3067	0.4529	7.1493	3.7244	0.6132
MurD	7.7271	6.2435	4.4806	7.3847	7.3380	3.4176	4.3944	4.4585	1.8466	1.1693	4.2870	4.3297	6.2873	0.6008	4.1418
Nig	6.9881	6.2002	7.4725	6.4727	1.0265	7.8670	7.4143	8.8256	3.6067	3.3178	5.9099	4.1285	7.1010	4.8332	1.3322
Nile	6.4639	7.9514	7.9865	4.7652	2.9902	7.2214	7.7403	5.8993	4.0461	1.4836	5.6837	1.8552	2.2564	3.0782	2.1106
Ob	6.6056	6.4986	7.0371	4.1614	9.0043	4.3280	1.2772	1.9267	3.4047	4.2521	0.6235	1.5404	7.2026	0.9858	2.1947
Oka	2.4510	5.1199	6.6927	5.3082	4.7694	5.3762	7.8669	5.6312	2.3745	3.4629	1.1323	0.3353	7.4777	1.4058	2.4373
Yang	7.9821	6.2516	4.7771	7.6716	6.7053	4.3032	5.1743	5.7194	2.6448	2.6242	2.6519	2.3536	2.0170	6.4098	1.7247
Yell	7.1587	3.9743	5.0690	6.5968	1.2658	7.9597	4.2859	9.9575	1.5364	7.2409	0.3054	3.2600	4.2486	6.5273	4.0255
Yen	6.2423	4.7065	7.7891	7.5950	6.5654	3.3398	0.6267	2.9088	3.5478	6.6352	1.6216	3.7788	6.4398	5.8477	3.8317
Yuk	6.7129	2.4535	4.0124	5.5225	4.0008	5.0325	4.7560	7.4078	1.7642	5.9166	0.2427	4.2994	5.9629	2.2630	1.8273

T A.7	*Rlai*														
	EN	***EB***	***DN***	***DB***	***MF***	***Mtv***	***Mcv***	***Sh***	***Gr***	***Cr***	***SpV***	***Ba***	***Bfl***	***Vfl***	***AA***
Glob	3.5313	1.1273	5.1762	8.3513	8.6323	3.7752	2.1127	3.2300	10.860	3.3034	0.7065	1.9094	19.529	9.1165	2.8683
Ama	2.0688	11.717	5.2575	6.1506	3.9600	3.4112	12.387	1.7113	0.6410	3.4352	3.4771	0.6990	5.3863	3.2279	4.5138
AmuD	7.7757	7.6830	4.2454	2.2756	5.3528	4.1988	5.7971	3.0413	8.1334	3.6156	3.5168	0.1269	5.5374	10.508	2.3616
Brah	17.659	2.1717	3.7834	3.6337	5.1902	4.2339	10.690	2.3421	2.0401	3.3440	4.3428	0.1525	14.128	9.1393	5.2500
Con	8.0820	5.8618	5.3467	4.3459	3.9416	3.8521	4.9658	1.4578	4.0033	7.6707	6.4110	0.1073	10.704	1.5006	4.8170
Dan	3.3943	4.1698	4.4346	3.5359	4.7332	5.6419	9.2421	2.8883	1.6910	6.4057	6.0701	0.0298	4.5936	4.4283	2.2319
Euph	3.1353	10.602	5.2977	8.2674	6.9553	3.6553	1.4079	1.8779	2.2050	3.4069	5.2268	1.1801	15.961	9.6621	4.6004
Gan	12.194	7.1580	5.2534	6.1694	6.0275	5.0902	5.3296	2.2730	3.4106	4.6298	2.4416	2.2487	11.307	3.9653	5.8049
Koly	6.6733	9.6205	4.4513	4.9821	6.1457	5.6311	4.3085	5.3983	7.9110	0.9962	3.1102	1.4409	3.3813	1.5867	1.9284
LaPla	4.8960	2.6815	7.6055	5.2595	6.4380	4.8168	1.5750	0.7316	7.2511	3.0781	1.5429	0.4012	19.728	9.4571	0.7090
Lena	7.3501	6.8058	4.3848	5.9634	8.0120	4.7242	3.9862	0.1296	15.975	3.7417	4.8475	1.6117	21.946	10.093	4.1989
Mek	2.7117	4.8428	8.3190	7.4879	3.3592	5.1983	14.033	5.4142	1.3658	6.2351	6.0863	1.4642	4.9329	6.3110	0.0366
Miss	2.1598	6.7657	7.2149	2.1385	4.8416	5.0804	6.0272	1.7889	4.5382	7.6334	3.1156	3.4540	17.310	2.1473	2.0130
MurD	3.4787	9.8238	7.9646	4.6012	4.9803	4.9983	3.7676	1.1094	4.5192	4.8807	4.7426	2.6536	2.3595	12.103	5.1640
Nig	6.5056	5.0468	4.7191	4.6796	3.7251	5.1214	8.0491	2.8533	4.5875	2.2538	0.7841	1.2051	5.1611	3.5646	4.5007
Nile	0.2393	1.2608	4.5865	6.5999	2.3892	4.9351	1.6712	4.2003	6.6134	3.4812	1.1956	0.6237	18.200	13.259	2.7823
Ob	5.8307	7.0602	5.8921	4.6747	3.9268	3.3118	2.1271	0.2590	5.0777	2.9121	6.7332	0.8088	6.8271	2.2888	5.4908
Oka	3.2894	4.6440	8.4198	5.9158	6.3343	3.5652	4.3896	4.5286	3.3066	1.6546	3.0808	0.4707	8.3323	7.3986	0.4958
Yang	5.2115	3.6080	7.2770	5.7491	4.9040	4.4081	3.5604	4.8124	6.9376	3.4151	5.7763	2.9351	10.414	5.7578	1.2735
Yell	2.7477	6.3530	5.4892	7.3332	7.4433	4.2017	2.7705	3.2000	6.2668	2.7958	2.5722	0.6468	3.4789	10.555	4.7288
Yen	6.6217	4.4489	5.8640	6.7322	2.0485	6.0493	3.7420	1.5650	6.4340	1.4484	1.4928	1.0114	20.811	11.034	3.2999
Yuk	6.8047	6.5024	6.6513	7.6974	5.7130	5.0292	4.9197	4.2680	16.563	2.8577	2.8238	3.6953	19.937	4.5325	2.6679

T A.8	*Hc*														
	EN	***EB***	***DN***	***DB***	***MF***	***Mtv***	***Mcv***	***Sh***	***Gr***	***Cr***	***SpV***	***Ba***	***Bfl***	***Vfl***	***AA***
Glob	14.529	1.7672	21.793	6.5219	27.043	26.935	8.9393	12.093	0.4839	6.3315	0.6247	0.9616	13.047	2.3514	0.7728
Ama	3.9307	18.022	4.9165	35.827	36.826	16.155	15.412	11.394	0.9639	0.9292	0.3285	0.3360	39.917	7.7699	0.4031
AmuD	3.6590	7.9057	11.488	5.9617	5.9617	39.997	18.268	12.178	1.7590	0.1931	3.8479	1.0200	40.067	4.2593	0.6094
Brah	11.517	22.338	26.683	37.100	37.100	67.635	7.6656	0.9944	0.1769	0.4617	3.3778	0.3003	12.033	7.2219	0.6410
Con	22.305	30.508	6.3804	25.258	25.258	34.933	7.8465	5.4580	0.5716	1.2999	1.6026	0.9423	12.083	11.864	0.8774
Dan	10.966	7.1340	2.5585	2.0325	2.0325	5.3272	7.3726	4.3838	1.1417	0.8197	0.7683	0.3892	46.475	6.8641	0.3027
Euph	17.534	14.859	14.296	26.889	26.889	27.573	10.868	14.997	1.2076	1.0704	2.3211	0.6516	41.197	3.4565	0.5402
Gan	12.074	24.730	8.2371	8.0728	8.0728	20.597	8.1009	2.2477	2.0559	0.5658	0.9415	0.1570	63.307	3.4561	0.3091
Koly	29.829	20.094	5.7813	46.321	46.321	39.851	8.4714	6.7063	0.8772	0.3886	1.113	0.7925	57.463	3.3160	0.5010
LaPla	22.663	8.7987	12.870	13.345	13.345	35.248	20.596	1.5993	3.6401	1.5191	0.2779	0.7550	20.220	10.340	0.1665
Lena	10.882	23.883	39.137	6.4066	6.4066	12.574	6.9642	3.4552	1.3684	2.3905	0.4365	0.1483	20.351	12.203	0.7137
Mek	31.091	16.487	12.839	2.4053	2.4053	16.794	9.6240	8.4367	0.7290	1.2958	1.7311	1.1342	22.018	5.7805	1.0370
Miss	9.1801	38.839	10.614	9.4199	9.4199	14.780	17.379	1.8370	0.7786	0.5320	0.5471	0.7518	38.852	3.4547	0.3294
MurD	32.297	15.223	0.8860	5.4681	5.4681	15.762	13.204	5.8328	1.0735	0.9349	3.4927	1.1693	25.707	8.1699	0.6867
Nig	14.891	28.711	53.933	3.7829	3.7829	60.700	3.3108	10.254	2.7091	1.0169	2.5102	0.5357	52.086	4.3095	1.1183
Nile	8.8834	6.3681	1.1518	15.870	15.870	3.0503	7.8658	2.0228	1.7444	0.7039	4.2197	0.3079	23.561	11.915	1.1290
Ob	25.158	7.1700	6.4562	36.622	36.622	12.535	7.1155	5.8214	1.2305	3.0563	1.8238	0.4830	38.033	6.3547	0.3465
Oka	23.189	4.6119	9.9868	25.402	25.402	45.074	19.096	5.3182	1.8929	0.0822	0.5829	0.7325	18.763	7.1938	1.1453
Yang	41.490	2.5870	18.872	10.840	10.840	19.554	4.6532	0.8143	2.8505	1.0883	3.0668	0.6371	49.145	16.668	0.7268
Yell	11.884	5.1982	14.902	16.500	16.500	41.696	11.215	2.2857	1.1457	1.1694	0.3137	0.7912	57.422	7.1789	0.0952
Yen	20.823	21.538	18.130	1.0419	1.0419	18.397	2.9554	1.9294	1.4967	0.4170	0.3466	0.2325	40.006	3.6920	0.0111
Yuk	22.466	13.298	4.5968	0.5984	0.5984	8.6392	10.208	15.440	0.9512	0.6167	1.7938	0.6142	38.322	7.4535	0.9681

T A.9 *Percov*

	EN	*EB*	*DN*	*DB*	*MF*	*Mtv*	*Mcv*	*Sh*	*Gr*	*Cr*	*SpV*	*Ba*	*Bfl*	*Vfl*	*AA*
Glob	80.590	97.598	55.851	86.476	78.065	64.388	81.963	22.854	0.0000	47.443	59.895	76.912	93.543	68.722	96.194
Ama	79.000	92.173	86.364	81.958	81.498	53.759	66.846	57.035	0.0000	24.983	51.423	91.891	96.649	63.672	97.273
AmuD	81.372	95.509	78.156	85.763	77.803	34.631	85.157	45.719	0.0000	24.508	70.234	80.779	94.112	76.206	87.646
Brah	79.200	86.977	82.510	81.993	80.641	58.119	79.893	58.030	0.0000	55.764	65.762	90.269	87.697	62.153	82.363
Con	88.174	94.369	62.794	86.216	82.729	44.050	56.924	22.584	0.0000	37.171	61.504	81.270	96.112	66.547	77.860
Dan	82.098	92.292	61.631	93.994	80.037	81.185	43.584	56.810	0.0000	17.025	52.870	76.930	96.384	85.951	92.887
Euph	83.490	93.111	56.262	89.708	95.942	21.981	85.414	38.491	0.0000	77.042	69.174	81.064	94.538	88.377	96.729
Gan	87.481	86.185	59.952	81.166	81.514	94.157	89.937	37.018	0.0000	65.807	80.470	86.878	96.455	63.077	82.455
Koly	87.667	94.547	60.979	82.044	94.044	73.477	53.037	47.955	0.0000	69.366	56.929	92.553	71.383	89.476	80.287
LaPla	77.802	92.794	59.744	93.850	75.394	51.098	88.948	35.200	0.0000	25.370	57.70	79.020	77.051	62.732	91.333
Lena	83.170	90.968	85.215	75.048	84.100	67.712	27.509	24.164	0.0000	37.577	93.127	78.231	70.217	66.960	90.823
Mek	91.213	96.435	81.934	82.856	84.716	54.489	64.332	70.869	0.0000	60.287	53.632	80.170	84.902	89.388	88.561
Miss	94.038	97.666	81.934	76.224	89.371	77.067	26.931	55.167	0.0000	45.649	69.881	94.536	89.413	77.124	90.485
MurD	75.450	93.029	76.898	93.519	95.188	34.247	68.189	65.329	0.0000	69.428	62.318	81.495	70.149	62.304	87.361
Nig	85.956	86.091	85.483	91.421	77.276	52.725	60.206	70.236	0.0000	51.954	91.416	77.727	91.324	84.241	92.907
Nile	76.702	87.453	73.896	92.485	87.342	42.120	79.475	30.430	0.0000	56.237	71.915	75.963	80.122	70.196	75.139
Ob	76.252	94.262	82.170	91.906	88.254	45.218	42.718	17.115	0.0000	25.545	60.180	94.033	81.656	88.044	96.212
Oka	87.467	91.617	76.710	86.250	94.777	44.336	18.298	18.293	0.0000	59.267	91.910	83.793	81.359	75.112	82.805
Yang	87.563	97.825	86.644	76.931	81.927	77.633	78.946	36.405	0.0000	10.254	52.446	81.109	70.061	77.549	84.059
Yell	90.652	97.046	60.038	86.748	77.804	44.670	53.129	20.395	0.0000	46.713	63.387	77.373	91.342	63.207	85.126
Yen	91.706	86.553	79.530	87.341	93.639	70.908	33.528	51.592	0.0000	67.588	59.793	78.363	88.427	88.581	76.299
Yuk	82.094	90.818	69.606	84.613	94.089	47.782	81.786	41.364	0.0000	22.287	59.771	89.537	85.778	83.610	85.325

Bibliography

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'connell, P. E. & Rasmussen, J. (1986). An Introduction to the European Hydrological System—Systeme Hydrologique Europeen,"She", 1: History and Philosophy of a Physically-Based, Distributed Modelling System. *Journal of hydrology,* 87**,** 45-59.

Abdulla, F. A., Lettenmaier, D. P., Wood, E. F. & Smith, J. A. (1996). Application of a Macroscale Hydrologic Model to Estimate the Water Balance of the Arkansas-Red River Basin. *Journal of Geophysical Research: Atmospheres (1984–2012),* 101**,** 7449-7459.

Abe, J., Wellens-Mensah, J., Diallo, O. S. & Mbuyil, C. (2004). Guinea Current, Giwa Regional Assessment 42. *In:* UNEP (ed.) *Global International Waters Assessment.* University of Kalmar, Sweden: United Nations Environment Programme.

Akhtar, M., Ahmad, N. & Booij, M. J. (2009). Use of Regional Climate Model Simulations as Input for Hydrological Models for the Hindukush-Karakorum-Himalaya Region. *Hydrol. Earth Syst. Sci.,* 13**,** 1075-1089.

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rosch, T. & Siebert, S. (2003). Development and Testing of the Watergap 2 Global Model of Water Use and Availability. *Hydrological Sciences Journal,* 48**,** 317-337.

Alcamo, J., Doll, P., Kaspar, F. & Siebert, S. (1997). Global Change and Global Scenarios of Water Use and Availability: An Application of Watergap 1.0. University of Kassel, Germany: Center for Environmental Systems Research (CESR).

Alcamo, J., Flörke, M. & Märker, M. (2007). Future Long-Term Changes in Global Water Resources Driven by Socio-Economic and Climatic Changes. *Hydrological Sciences Journal,* 52**,** 247-275.

Alcamo, J., Henrichs, T. & Rösch, T. (2000). World Water in 2025 - Global Modeling Scenario Analysis for the World Commission of Water for the 21st Century. Kassel, Germany: Center for Environmental Systems Research, University of Kassel.

Allen, R. (2005). Penman-Monteith Equation. *In:* HILLEL, D., ROSENZWEIG, C., POWLSON, D., SCOW, K., SINGER, M. & SPARKS, D. (eds.) *Encyclopedia of Soils in the Environment.* Academic Press.

Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. (1998). Crop Evapotranspiration-Guidelines for Computing Crop Water Requirements-Fao Irrigation and Drainage Paper 56. *FAO, Rome,* 300**,** 6541.

Alpas, H., Smith, M. & Kulmyrzaev, A. (2011). *Strategies for Achieving Food Security in Central Asia*, Springer.

Alvisi, S., Mascellani, G., Franchini, M. & Bardossy, A. (2006). Water Level Forecasting through Fuzzy Logic and Artificial Neural Network Approaches. *Hydrology and Earth System Sciences,* 10**,** 1-17.

Amerman, C. (1965). The Use of Unit-Source Watershed Data for Runoff Prediction. *Water Resources Research,* 1**,** 499-507.

Andréassian, V., Lerat, J., Loumagne, C., Mathevet, T., Michel, C., Oudin, L. & Perrin, C. (2007). What Is Really Undermining Hydrologic Science Today? *Hydrological Processes,* 21**,** 2819-2822.

Arino, O., Ramos, J., Kalogirou, V., Defourny, P. & Achard, F. (Year) Published. Globcover2009. ESA Living Planet Symposium, 2010 Bergen, Norway.

Arnell, N. (2002). *Hydrology and Global Environmental Change*, Prentice Hall.

Arnell, N. & Gosling, S. (2014). The Impacts of Climate Change on River Flood Risk at the Global Scale. *Climatic Change***,** 1-15.

Arnell, N. W. (1999). A Simple Water Balance Model for the Simulation of Streamflow over a Large Geographic Domain. *Journal of Hydrology,* 217**,** 314-335.

Arnell, N. W. (2003). Effects of Ipcc Sres Emissions Scenarios on River Runoff: A Global Perspective. *Hydrology and Earth System Sciences,* 7**,** 619-641.

Arnell, N. W., Wheeler, T., Osborne, T., Rose, G., Gosling, S. N., Dawson, T., Penn, A. S. & Perryman, A. H. (2010). The Implications of Climate Policy for Avoided Impacts on Water and Food Security. *Work stream 2, Report 6 of the AVOID programme: Avoiding dangerous climate change.* Walker Institute, University of Reading.

Aus Der Beek, T., Voß, F. & Flörke, M. (2011). Modelling the Impact of Global Change on the Hydrological System of the Aral Sea Basin. *Physics and Chemistry of the Earth, Parts A/B/C,* 36**,** 684-695.

Avoid. (2014). *Avoid: Can We Avoid Dangerous Climate Change?* [Online]. UK Met Office. Available: www.avoid.uk.net [Accessed 13/05/2014.

Balsamo, G., Viterbo, P., Beljaars, A., Van Den Hurk, B., Hirschi, M., Betts, A. K. & Scipal, K. (2009). A Revised Hydrology for the Ecmwf Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. *Journal of Hydrometeorology,* 10.

Bank of England (2015). Inflation Report Fan Charts February 2015 (http://www.bankofengland.co.uk/publications/Documents/inflationreport/2015/feb5.pdf).

Barthem, R. B., Charvet-Almeida, P., Montag, L. F. A. & Lanna, A. E. (2004). Giwa Regional Assessment 40b. *In:* UNEP (ed.) *Global International Waters Assessment.* University of Kalmar, Sweden: United Nations Environment Programme.

Bartholomé, E. & Belward, A. (2005). Glc2000: A New Approach to Global Land Cover Mapping from Earth Observation Data. *International Journal of Remote Sensing,* 26**,** 1959-1977.

Bennett, V., Kershaw, P., Pritchard, M., Churchill, J., Del Cano Novales, C., Juckes, M., Pascoe, S., S., P., Stephens, A., Lawrence, B., Muller, J., Kharbouche, S., Latter, B. & Styles, J. (Year) Published. Eo Science from Big Eo Data on the Jasmin-Cems Infrastructure. Big Data from Space (BiDS' 14), 12-14 November 2014 2014 European Space Agency-ESRIN, Frascati, Italy.

Bergström, S. (1995). The Hbv Model. *In:* SINGH, V. (ed.) *Computer Models of Watershed Hydrology.*

Bergström, S., Lindström, G. & Pettersson, A. (2002). Multi-Variable Parameter Estimation to Increase Confidence in Hydrological Modelling. *Hydrological Processes,* 16**,** 413-421.

Beven, K. (2006a). A Manifesto for the Equifinality Thesis. *Journal of Hydrology,* 320**,** 18-36.

Beven, K. (2006b). On Undermining the Science? *Hydrological Processes,* 20**,** 3141-3146.

Beven, K. (2009). *Environmental Modelling: An Uncertain Future?,* Abingdon, Oxon, Routledge.

Beven, K. (2012). *Rainfall-Runoff Modelling: The Primer (Second Edition),* Chichester, Wiley-Blackwell.

Beven, K. & Binley, A. (2013). Glue: 20 Years On. *Hydrological Processes***,** n/a-n/a.

Beven, K. & Freer, J. (2001). Equifinality, Data Assimilation, and Uncertainty Estimation in Mechanistic Modelling of Complex Environmental Systems Using the Glue Methodology. *Journal of Hydrology,* 249**,** 11-29.

Beven, K. & Kirkby, M. J. (1979). A Physically Based, Variable Contributing Area Model of Basin Hydrology. *Hydrological Sciences Journal,* 24**,** 43-69.

Beven, K., Smith, P. & Freer, J. (2007). Comment on "Hydrological Forecasting Uncertainty Assessment: Incoherence of the Glue Methodology" by Pietro Mantovan and Ezio Todini. *Journal of Hydrology,* 338**,** 315-318.

Beven, K., Smith, P., Westerberg, I. & Freer, J. (2012). Comment on "Pursuing the Method of Multiple Working Hypotheses for Hydrological Modeling" by P. Clark Et Al. *Water Resources Research,* 48**,** W11801.

Beven, K. J. & Binley, A. M. (1992). The Future of Distributed Models: Model Calibration and Uncertainty Prediction. *Hydrological Processes,* 6**,** 279-298.

Beven, K. J., Smith, P. J. & Freer, J. E. (2008). So Just Why Would a Modeller Choose to Be Incoherent? *Journal of Hydrology,* 354**,** 15-32.

Blyth, E. (2009). Processes That Impact Runoff Generation in Northern Latitudes. *In:* WATCH (ed.). CEH.

Bowden, G. J., Maier, H. R. & Dandy, G. C. (2002). Optimal Division of Data for Neural Network Models in Water Resources Applications. *Water Resources Research,* 38**,** 2-1-2-11.

Bowling, L., Pomeroy, J. & Lettenmaier, D. (2004). Parameterization of Blowing-Snow Sublimation in a Macroscale Hydrology Model. *Journal of Hydrometeorology,* 5.

Bowling, L. C. & Lettenmaier, D. P. (2010). Modeling the Effects of Lakes and Wetlands on the Water Balance of Arctic Environments. *Journal of Hydrometeorology,* 11.

Bradford, R. B. & Marsh, T. J. (2003). Defining a Network of Benchmark Catchments for the Uk. *Proceedings of the ICE - Water and Maritime Engineering,* 156**,** 109-116.

Breuer, L., Eckhardt, K. & Frede, H.-G. (2003). Plant Parameter Values for Models in Temperate Climates. *Ecological Modelling,* 169**,** 237-293.

Breuer, L. & Frede, H. G. (2003). Plapada - an Online Plant Parameter Data Drill for Eco-Hydrological Modelling Approaches. Institute for Landscape Ecology and Resources Management.

Brilly, M. (2010). *Hydrological Processes of the Danube River Basin: Perspectives from 10 Danubian Countries*, Springer.

Budescu, D. V., Broomell, S. & Por, H.-H. (2009). Improving Communication of Uncertainty in the Reports of the Intergovernmental Panel on Climate Change. *Psychological Science,* 20**,** 299-308.

Burek, P., Alfieri, L., Thielen-Del Pozo, J., Muraro, D., Pappenberger, F. & Krzeminsk, B. (Year) Published. Looking at the Big Scale-Global Flood Forecasting. EGU General Assembly Conference Abstracts, 2012. 3993.

Calder, I. R. (1986). A Stochastic Model of Rainfall Interception. *Journal of Hydrology,* 89**,** 65-71.

Calder, I. R. (1990). *Evaporation in the Uplands*, Wiley.

Calder, I. R. (1998). Water Use by Forests, Limits and Controls. *Tree Physiology,* 18**,** 625-631.

Calder, I. R. (1999). *The Blue Revolution: Land Use and Integrated Water Resources Management*, Earthscan.

Calver, A., Wood, W. & Singh, V. (1995). *The Institute of Hydrology Distributed Model*, Water Resources Publications.

Cameron, D. S., Beven, K. J., Tawn, J., Blazkova, S. & Naden, P. (1999). Flood Frequency Estimation by Continuous Simulation for a Gauged Upland Catchment (with Uncertainty). *Journal of Hydrology,* 219**,** 169-187.

Cavendish, M. (2006). *World and Its Peoples*, Marshall Cavendish Corporation.

Ceh. (2014). *Gwava: Global Water Availibility Assessment Tool* [Online]. Available: http://www.ceh.ac.uk/sci_programmes/water/gwava.html.

Ceh & Bgs (2012). Water Security in the Uk: A Pilot Mode-Based Study of Current and Future Water Security in the Uk. Natural Environment Research Council.

Census Bureau (2009). *Statistical Abstract of the United States 2010 (Paperback)*, United States Government Printing Office.

Chang, F.-J., Hu, H.-F. & Chen, Y.-C. (2001). Counterpropagation Fuzzy–Neural Network for Streamflow Reconstruction. *Hydrological Processes,* 15**,** 219-232.

Chang, L.-C., Chang, F.-J. & Chiang, Y.-M. (2004). A Two-Step-Ahead Recurrent Neural Network for Stream-Flow Forecasting. *Hydrological Processes,* 18**,** 81-92.

Channan, S., Collins, K. & Emanuel, W. R. (2014). Global Mosaics of the Standard Modis Land Cover Type Data. College Park, Maryland, USA: University of Maryland and the Pacific Northwest National Laboratory.

Chapman, S. (2013). *High Performance Computing at the University of Bath: Aquila* [Online]. Available: https://wiki.bath.ac.uk/display/HPC/Aquila [Accessed 16/03/2015.

Chaubey, I., Cotter, A., Costello, T. & Soerens, T. (2005). Effect of Dem Data Resolution on Swat Output Uncertainty. *Hydrological Processes,* 19**,** 621-628.

Chauhan, Y., Das, D., Gupta, K. & Young, G. (2014a). Okavango River. *Encyclopaedia Britannica.* [Online].

Chauhan, Y., Lotha, G., Luebering, J. E., Mckenna, A., Parrott-Sheffer, C., Sinha, S. & Young, G. (2014b). Aswan High Dam. *Encyclopedia Britannica.* [Online].

Chen, C., Hagemann, S., Clarke, D., Folwell, S., Gosling, S., Haddeland, I., Hanasaki, N., Heinke, J., Ludwig, F., Voβ, F. & Wiltshire, A. (2011). Watch Technical Report No. 45 - Evaluation of Projected Hydrological Changes in the 21st Century Obtained from a Muli-Model Ensemble. Water and Global Change.

Cheng, J. & Druzdzel, M. J. (2000). Latin Hypercube Sampling in Bayesian Networks. *Uncertain Reasoning.* Pittsburgh: American Association for Artificial Intelligence.

Cherkauer, K. A., Bowling, L. C. & Lettenmaier, D. P. (2003). Variable Infiltration Capacity Cold Land Process Model Updates. *Global and Planetary Change,* 38**,** 151-159.

Christiaens, K. & Feyen, J. (2002). Constraining Soil Hydraulic Parameter and Output Uncertainty of the Distributed Hydrological Mike She Model Using the Glue Framework. *Hydrological Processes,* 16**,** 373-391.

Clapp, R. B. & Hornberger, G. M. (1978). Empirical Equations for Some Soil Hydraulic Properties. *Water Resources Research,* 14**,** 601-604.

Clark, M. P., Kavetski, D. & Fenicia, F. (2011). Pursuing the Method of Multiple Working Hypotheses for Hydrological Modeling. *Water Resources Research,* 47**,** W09301.

Clark, M. P., Kavetski, D. & Fenicia, F. (2012). Reply to Comment by K. Beven Et Al. On “Pursuing the Method of Multiple Working Hypotheses for Hydrological Modeling”. *Water Resources Research,* 48**,** W11802.

Cloke, H. L. & Hannah, D. M. (2011). Large-Scale Hydrology: Advances in Understanding Processes, Dynamics and Models from Beyond River Basin to Global Scale. *Hydrological Processes,* 25**,** 991-995.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A. J. & Wehner, M. (2013). Long-Term Climate Change: Projections, Commitments and Irriversibility. *In:* STOCKER, T. F., QIN, D., PLATTNER, G. K., TIGNOR, M., ALLEN, S. K., BOSCHUNG, J., NAUELS, A., XIA, Y., BEX, V. & MIDGLEY, P. M. (eds.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, United Kingdom: Cambridge University Press.

Cornell University. (2010). *Northeast Regoin Certified Crop Advisor (Nrcca) Study Resources: Soil Hydrology Aem* [Online]. Available: http://nrcca.cals.cornell.edu/soil/CA2/CA0212.1-3.php [Accessed 15/06/2014.

Coulibaly, P. & Baldwin, C. K. (2005). Nonstationary Hydrological Time Series Forecasting Using Nonlinear Dynamic Methods. *Journal of Hydrology,* 307**,** 164-174.

Cox, P., Betts, R., Bunton, C., Essery, R., Rowntree, P. & Smith, J. (1999). The Impact of New Land Surface Physics on the Gcm Simulation of Climate and Climate Sensitivity. *Climate Dynamics,* 15**,** 183-203.

Crawford, N. H. & Linsley, R. K. (1966). Digital Simulation in Hydrology'stanford Watershed Model 4.

Creutin, J. D., Andrieu, H. & Faure, D. (1997). Use of a Weather Radar for the Hydrology of a Mountainous Area. Part Ii: Radar Measurement Validation. *Journal of Hydrology,* 193**,** 26-44.

Criss, R. E. & Winston, W. E. (2008). Do Nash Values Have Value? Discussion and Alternate Proposals. *Hydrological Processes,* 22**,** 2723.

Crist, R. E., Parsons, J. J. & Schultz, A. R. (2014). Amazon River. *Encyclopaedia Britannica.* [Online].

Cukier, R., Fortuin, C., Shuler, K. E., Petschek, A. & Schaibly, J. (1973). Study of the Sensitivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients. I Theory *The Journal of Chemical Physics* 59**,** 3873-3878.

Cukier, R., Levine, H. & Shuler, K. (1978). Nonlinear Sensitivity Analysis of Multiparameter Model Systems. *Journal of computational physics,* 26**,** 1-42.

D'orgeval, T., Polcher, J. & De Rosnay, P. (2008). Sensitivity of the West African Hydrological Cycle in Orchidee to Infiltration Processes. *Hydrology & Earth System Sciences Discussions,* 5.

Daniel, C. (1958). On Varying One Factor at a Time. *Biometrics,* 14**,** 430-431.

Daniel, C. (1973). One-at-a-Time Plans. *Journal of the American statistical association,* 68**,** 353-360.

Davie, J., Falloon, P., Kahana, R., Dankers, R., Betts, R., Portmann, F., Wisser, D., Clark, D., Ito, A. & Masaki, Y. (2013). Comparing Projections of Future Changes in Runoff from Hydrological and Biome Models in Isi-Mip. *Earth System Dynamics,* 4**,** 359-374.

Dawdy, D. R. & O'donnell, T. (1965). Mathematical Models of Catchment Behavior. *J. Hydraul. Div. Am. Soc. Civ. Eng,* 91**,** 113-137.

Dawson, C. W., Abrahart, R. J. & See, L. M. (2007). Hydrotest: A Web-Based Toolbox of Evaluation Metrics for the Standardised Assessment of Hydrological Forecasts. *Environmental Modelling & Software,* 22**,** 1034-1052.

Dawson, C. W., Abrahart, R. J., Shamseldin, A. Y. & Wilby, R. L. (2006). Flood Estimation at Ungauged Sites Using Artificial Neural Networks. *Journal of Hydrology,* 319**,** 391-409.

De Rosnay, P. & Polcher, J. (1998). Modelling Root Water Uptake in a Complex Land Surface Scheme Coupled to a Gcm. *Hydrology and Earth System Sciences,* 2**,** 239-255.

Defries, R. S., Hansen, M., Townshend, J. R. G. & Sholberg, R. (1998). Global Land Cover Classification at 8km Spatial Resolution: The Use of Training Data Derived from Landsat Imagery in Decision Tree Classifiers. *International Journal of Remote Sensing,* 19**,** 3141-3168.

Dehghani, M., Saghafian, B., Nasiri Saleh, F., Farokhnia, A. & Noori, R. (2014). Uncertainty Analysis of Streamflow Drought Forecast Using Artificial Neural Networks and Monte-Carlo Simulation. *International Journal of Climatology,* 34**,** 1169-1180.

Dingman, S. L. (2002). *Physical Hydrology,* Long Grove, IL, USA, Waveland Press Inc.

Doherty, J. (2010). Pest: Model-Independent Parameter Estimation User Manual: 5th Edition. Brisbane, Australia: Watermark Numerical Computing.

Döll, P., Fiedler, K. & Zhang, J. (2009). Global-Scale Analysis of River Flow Alterations Due to Water Withdrawals and Reservoirs. *Hydrology & Earth System Sciences,* 13**,** 2413-2432.

Döll, P., Kaspar, F. & Alcamo, J. (1999). Computation of Global Water Availability and Water Use at the Scale of Large Drainage Basins. *Mathematische Geologie,* 4**,** 111-118.

Döll, P., Kaspar, F. & Lehner, B. (2003). A Global Hydrolgical Model for Deriving Water Availability Indicators: Model Tuning and Validation. *Journal of Hydrology,* 270**,** 105-134.

Dooge, J. C. (1959). A General Theory of the Unit Hydrograph. *Journal of geophysical research,* 64**,** 241-256.

Dooge, J. C. I. (1986). Looking for Hydrologic Laws. *Water Resources Research,* 9**,** 46S-58S.

Duan, Q., Sorooshian, S. & Gupta, V. (1992). Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models. *Water resources research,* 28**,** 1015-1031.

Dumont, E., Williams, R., Keller, V. & Folwell, S. (2010). Modelling Water Scarcity across Europe in Terms of Water Quantity and Quality. *In:* SERVAT, E., DEMUTH, S., DEZETTER, A. & DANIELL, T. (eds.) *Global Change: Facing Risks and Threats to Water Resources.* Wallingford: Int Assoc Hydrological Sciences.

Dumont, E., Williams, R., Keller, V., Voß, A. & Tattari, S. (2012). Modelling Indicators of Water Security, Water Pollution and Aquatic Biodiversity in Europe. *Hydrological Sciences Journal,* 57**,** 1378-1403.

Ellis, J. B., Geodesy, I. U. O. & Assembly, G. G. (1999). *Impacts of Urban Growth on Surface Water and Groundwater Quality: Proceedings of an International Symposium Held During Iugg 99, the Xxii General Assembly of the International Union of Geodesy and Geophysics, at Birmingham, Uk 18-30 July 1999*, IAHS.

Engelhardt, I., De Aguinaga, J., Mikat, H., Schüth, C., Lenz, O. & Liedl, R. (2012). Complexity Versus Simplicity: An Example of Groundwater Model Ranking with the Akaike Information Criterion. *Hydrology and Earth System Sciences Discussions,* 9**,** 9687-9714.

Essery, R., Best, M., Betts, R., Cox, P. M. & Taylor, C. M. (2003). Explicit Representation of Subgrid Heterogeneity in a Gcm Land Surface Scheme. *Journal of Hydrometeorology,* 4.

Fair, G. M. & Hatch, L. P. (1933). Fundamental Factors Governing the Streamline Flow of Water through Sand *Journal (American Water Works Association),* 25**,** 1551-1565.

Fao. (1995). *Digital Soil Map of the World* [Online]. Available: http://www.fao.org/geonetwork/srv/en/metadata.show?id=14116 [Accessed 18/03/2013.

Fao (2014a). Aquastat Database.

Fao. (2014b). *Soil Texture Web Page* [Online]. Available: ftp://ftp.fao.org/fi/CDrom/FAO_training/FAO_training/general/x6706e/x6706e06.htm [Accessed 26/05/2014.

Federer, C. A., Vörösmarty, C. & Fekete, B. (2003). Sensitivity of Annual Evaporation to Soil and Root Properties in Two Models of Contrasting Complexity. *Journal of Hydrometeorology,* 4**,** 1276-1290.

Fekete, B. & Vorosmarty, C. J. (2011). Islscp Ii Unh/Grdc Composite Daily Runoff. *In:* HALL, F. G., COLLATZ, G. J., MEESON, D., LOS, S., BROWN DE COLSTOUN, E. & LANDIS, D. (eds.) *ISLSCP Initiative II Collection.*

Fekete, B. M., Vörösmarty, C. J. & Grabs, W. (1999). Global, Composite Runoff Fields Based on Observed River Discharge and Simulated Water Balances. Fed. Inst. of Hydrology.

Fekete, B. M., Vörösmarty, C. J., Roads, J. O. & Willmott, C. J. (2004). Uncertainties in Precipitation and Their Impacts on Runoff Estimates. *Journal of Climate,* 17**,** 294-304.

Fernando, D. & Jayawardena, A. (1998). Runoff Forecasting Using Rbf Networks with Ols Algorithm. *Journal of Hydrologic Engineering,* 3**,** 203-209.

Folwell, S. & Farquharson, F. (2006). The Impacts of Climate Change on Water Resources in the Okavango Basin. *Climate Variability and Change - Hydrological Impacts (Proceedings of the Fifth Friend World Conference Held at Havana, Cuba, November 2006).* IAHS.

Frame, D. J., Aina, T., Christensen, C. M., Faull, N. E., Knight, S. H. E., Piani, C., Rosier, S. M., Yamazaki, K., Yamazaki, Y. & Allen, M. R. (2009). The Climateprediction.Net Bbc Climate Change Experiment: Design of the Coupled Model Ensemble. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 367**,** 855-870.

Freer, J., Beven, K. & Ambroise, B. (1996). Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the Glue Approach. *Water Resour. Res.,* 32**,** 2161-2173.

Freeze, R. A. & Harlan, R. (1969). Blueprint for a Physically-Based, Digitally-Simulated Hydrologic Response Model. *Journal of Hydrology,* 9**,** 237-258.

Fung, C. F., Farquharson, F. & Chowdhury, J. (2006). Exploring the Impacts of Climate Change on Water Resources-Regional Impacts at a Regional Scale: Bangladesh. *Climate variability and change: hydrological impacts***,** 389-393.

Fung, F., Lopez, A. & New, M. (2011). Water Availability in +2°C and +4°C Worlds. *Philosophical Transactions of the Royal Society A.,* 369**,** 99-116.

Gain, A. K. & Wada, Y. (2014). Assessment of Future Water Scarcity at Different Spatial and Temporal Scales of the Brahmaputra River Basin. *Water resources management,* 28**,** 999-1012.

Gao, H., Tang, Q., Shi, X., Zhu, C., Bohn, T., Su, F., Sheffield, J., Pan, M., Lettenmaier, D. P. & Wood, E. F. (2010). Water Budget Record from Variable Infiltration Capacity (Vic) Model. *Algorithm Theoretical Basis Document for Terrestrial Water Cycle Data Records.*

Gebhart, S., Radoglou, K., Chalivopoulos, G. & Matzarakis, A. (2013). Evaluation of Potential Evapotranspiration in Central Macedonia by Empest. *Advances in Meteorology, Climatology and Atmospheric Physics.* Springer.

Giustolisi, O. & Laucelli, D. (2005). Improving Generalization of Artificial Neural Networks in Rainfall–Runoff Modelling / Amélioration De La Généralisation De Réseaux De Neurones Artificiels Pour La Modélisation Pluie-Débit. *Hydrological Sciences Journal,* 50**,** null-457.

Gosling, S., Bretherton, D. & Arnell, N. W. (2010). Global Hydrology Modelling and Uncertainty: Running Multiple Ensembles with a Campus Grid. *Philosophical Transactions of the Royal Society A.,* 388**,** 4005-4021.

Gosling, S. N. & Arnell, N. W. (2011). Simulating Current Global River Runoff with a Global Hydrological Model: Model Revisions, Validation and Sensitivity Analysis. *Hydrological Processes,* 25**,** 1129-1145.

Gosling, S. N., Taylor, R. G., Arnell, N. W. & Todd, M. C. (2011). A Comparative Analysis of Projected Impacts of Climate Change on River Runoff from Global and Catchment-Scale Hydrological Models. *Hydrol. Earth Syst. Sci.,* 15**,** 279-294.

Grant, D. (1975). Comparison of Evaporation from Barley with Penman Estimates. *Agricultural Meteorology,* 15**,** 49-60.

Grdc (2014). The Global Runoff Data Base. *In:* THE GLOBAL RUNOFF DATA CENTRE (ed.). http://www.bafg.de/GRDC/EN/Home/homepage_node.html D-56002 Koblenz, Germany.

Green, A., Thompson, J., Kingston, D. & Gosling, S. (Year) Published. Quantifying the Relative Magnitude of Sources of Uncertainty in River Flow Projections under Climate Change: An Assessment for the Mekong River. EGU General Assembly Conference Abstracts, 2014. 15019.

Green, W. H. & Ampt, G. A. (1911). Studies on Soil Physics Part I - the Flow of Air and Water through Soils. *Journal of Agricultural Science,* 4**,** 1-24.

Gruber, T., Bamber, J. L., Bierkens, M. F. P., Dobslaw, H., Murbock, M., Thomas, M., Van Beek, L. P. H., Van Dam, T., Vermeersen, L. L. A. & Visser, P. N. a. M. (2011). Simulation of the Time-Variable Gravity Field by Means of Coupled Geophysical Models. *Earth Systems Science Data Discussion,* 4**,** 27-40.

Guimberteau, M., Drapeau, G., Ronchail, J., Sultan, B., Polcher, J., Martinez, J.-M., Prigent, C., Guyot, J.-L., Cochonneau, G. & Espinoza, J. (2012). Discharge Simulation in the Sub-Basins of the Amazon Using Orchidee Forced by New Datasets. *Hydrology & Earth System Sciences,* 16.

Gupta, A. (2007). *Large Rivers: Geomorphology and Management*, John Wiley.

Gupta, H. V., Sorooshian, S. & Yapo, P. O. (1998). Toward Improved Calibration of Hydrologic Models: Multiple and Noncommensurable Measures of Information. *Water Resources Research,* 34**,** 751-763.

Gupta, S. K. (2010). *Modern Hydrology and Sustainable Water Development*, John Wiley & Sons.

Guse, B., Reusser, D. E. & Fohrer, N. (2014). How to Improve the Representation of Hydrological Processes in Swat for a Lowland

Catchment – Temporal Analysis of Parameter Sensitivity and Model Performance. *Hydrological Processes,* 28**,** 2651-2670.

Haddeland, I., Clarke, D., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P. & Yeh, P. (2011). Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results. *Journal of Hydrometeorology,* 12**,** 869-884.

Haddeland, I., Lettenmaier, D. P. & Skaugen, T. (2006a). Effects of Irrigation on the Water and Energy Balances of the Colorado and Mekong River Basins. *Journal of Hydrology,* 324**,** 210-223.

Haddeland, I., Skaugen, T. & Lettenmaier, D. (2007). Hydrologic Effects of Land and Water Management in North America and Asia: 1700? 1992. *Hydrology and Earth System Sciences Discussions,* 11**,** 1035-1045.

Haddeland, I., Skaugen, T. & Lettenmaier, D. P. (2006b). Anthropogenic Impacts on Continental Surface Water Fluxes. *Geophysical Research Letters,* 33.

Hagemann, S. (2002). *An Improved Land Surface Parameter Dataset for Global and Regional Climate Models*, Max-Planck-Institut für Meteorologie.

Hagemann, S., Chen, C., Clark, D., Folwell, S., Gosling, S. N., Haddeland, I., Hannasaki, N., Heinke, J., Ludwig, F. & Voss, F. (2013). Climate Change Impact on Available Water Resources Obtained Using Multiple Global Climate and Hydrology Models. *Earth System Dynamics,* 4**,** 129-144.

Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D. & Piani, C. (2011). Impact of a Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three Gcms and Two Hydrology Models. *Journal of Hydrometeorology,* 12**,** 556-578.

Hagemann, S. & Dümenil, L. (1997). A Parametrization of the Lateral Waterflow for the Global Scale. *Climate Dynamics,* 14**,** 17-31.

Hagemann, S. & Gates, L. D. (2001). Validation of the Hydrological Cycle of Ecmwf and Ncep Reanalyses Using the Mpi Hydrological Discharge Model. *J. Geophys. Res.,* 106**,** 1503-1510.

Hagemann, S. & Gates, L. D. (2003). Improving a Subgrid Runoff Parameterization Scheme for Climate Models by the Use of High Resolution Data Derived from Satellite Observations. *Climate Dynamics,* 21**,** 349-359.

Hall, J., O'connell, E. & Ewen, J. (2007). On Not Undermining the Science: Coherence, Validation and Expertise. Discussion of Invited Commentary by Keith Beven Hydrological Processes, 20, 3141–3146 (2006). *Hydrological Processes,* 21**,** 985-988.

Hanasaki, N., Inuzuka, T., Kanae, S. & Oki, T. (2010). An Estimation of Global Virtual Water Flow and Sources of Water Withdrawal

for Major Crops and Livestock Products Using a Global Hydrological Model. *Journal of Hydrology,* 384**,** 232-244.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y. & Tanaka, K. (2008a). An Integrated Model for the Assessment of Global Water Resources - Part 1: Model Description and Input Meteorological Forcing. *Hydrology and Earth System Sciences,* 12**,** 1007-1025.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y. & Tanaka, K. (2008b). An Integrated Model for the Assessment of Global Water Resources - Part 2: Applications and Assessments. *Hydrology and Earth System Sciences,* 12**,** 1027-1037.

Hannaford, J. & Marsh, T. (2006). An Assessment of Trends in Uk Runoff and Low Flows Using a Network of Undisturbed Catchments. *International Journal of Climatology,* 26**,** 1237-1253.

Hannaford, J. & Marsh, T. J. (2008). High-Flow and Flood Trends in a Network of Undisturbed Catchments in the Uk. *International Journal of Climatology,* 28**,** 1325-1338.

Harding, R. J. & Warnaas, T. A. (2011). Water and Global Change: The Watch Projection Outreach Report. *In:* CENTRE FOR ECOLOGY AND HYDROLOGY (ed.). Wallingford: NERC Centre for Ecology and Hydrology.

Hauduc, H., Neumann, M., Muschalla, D., Gamerith, V., Gillot, S. & Vanrolleghem, P. (2011). Towards Quantitative Quality Criteria to Evaluate Simulation Results in Wastewater Treatment–a Critical Review. *Watermatex,* 11**,** 46.

Hawkins, E. 6/2/2014 The Cascade of Uncertainty in Climate Projections. *Climate Lab Book* [Online]. Available from: http://www.climate-lab-book.ac.uk/2014/cascade-of-uncertainty/ [Accessed 11/12/2014.

Hawkins, E. Reinventing the Colour Wheel. *Climate Lab Book* [Online]. Available from: http://www.climate-lab-book.ac.uk/2015/reinventing-colour-wheel/.

Henderson-Sellers, A., Mcguffie, K. & Gross, C. (1995). Sensitivity of Global Climate Model Simulations to Increased Stomatal Resistance and C02 Increases*. *Journal of Climate,* 8**,** 1738-1756.

Hipel, K. W. (1981). Geophysical Model Discrimination Using the Akaike Information Criterion. *Automatic Control, IEEE Transactions on,* 26**,** 358-378.

Hirabayashi, Y., Kanae, S., Emori, S., Oki, T. & Kimoto, M. (2008). Global Projections of Changing Risks of Floods and Droughts in a Changing Climate. *Hydrological Sciences,* 53**,** 754-772.

Hornberger, G. M. & Spear, R. (1981). Approach to the Preliminary Analysis of Environmental Systems. *J. Environ. Manage.;(United States),* 12.

Horton, R. E. (1933). The Role of Infiltration in the Hydrologic Cycle. *Transactions-American Geophysical Union,* 14**,** 446-460.

Horton, R. E. (1945). Erosional Development of Streams and Their Drainage Basins; Hydrophysical Approach to Quantitative Morphology. *Geological society of America bulletin,* 56**,** 275-370.

Hossain, F., Anagnostou, E. N., Dinku, T. & Borga, M. (2004). Hydrological Model Sensitivity to Parameter and Radar Rainfall Estimation Uncertainty. *Hydrological Processes,* 18**,** 3277-3291.

Hsu, K.-L., Gupta, H. V. & Sorooshian, S. (1995). Artificial Neural Network Modeling of the Rainfall-Runoff Process. *Water Resources Research,* 31**,** 2517-2530.

Hu, T. S., Lam, K. C. & Ng, S. T. (2001). River Flow Time Series Prediction with a Range-Dependent Neural Network. *Hydrological Sciences Journal,* 46**,** 729-745.

Huang, H.-J., Cheng, S.-J., Wen, J.-C. & Lee, J.-H. (2008). Effect of Growing Watershed Imperviousness on Hydrograph Parameters and Peak Discharge. *Hydrological Processes,* 22**,** 2075-2085.

Hyndman, D. W. (2008). *Natural Hazards and Disasters*, Brooks/Cole.

Imbeaux, É. (1892). *La Durance: Régime, Crues Et Inundations*, Ch. Dunod.

Ipcc (2007). Fourth Assessment Report - Climate Change 2007: Wg1 the Physical Science Basis. Cambridge: Cambridge University Press.

Ipcc (2012). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: A Special Report of Working Groups I and Ii of the Intergovernmental Panel on Climate Change (Srex). *In:* FIELD, C. B., BARROS, V., STOCKER, T. F., QIN, D., DOKKEN, D. J., EBI, K. L., MASTRANDREA, M. D., MACH, K. J., G.-K., P., ALLEN, S. K., TIGNOR, M. & MIDGLEY, P. M. (eds.). Cambridge: Intergovernmental Panel on Climate Change.

Ipcc (2014). Fifth Assessment Report, Working Group 2, Chapter 3: Freshwater Resources. IPCC.

Isi-Mip. (2015). *Isi-Mip Output Data* [Online]. Available: https://www.pik-potsdam.de/research/climate-impacts-and-vulnerabilities/research/rd2-cross-cutting-activities/isi-mip/for-modellers/isi-mip-fast-track/output-data [Accessed 10/05/2015.

Jain, A. & Srinivasulu, S. (2006). Integrated Approach to Model Decomposed Flow Hydrograph Using Artificial Neural Network and Conceptual Techniques. *Journal of Hydrology,* 317**,** 291-306.

Jain, S. & Sudheer, K. (2008). Fitting of Hydrologic Models: A Close Look at the Nash–Sutcliffe Index. *Journal of Hydrologic Engineering,* 13**,** 981-986.

Jain, S. K., Agarwal, P. K. & Singh, V. P. (2007). *Hydrology and Water Resources of India*, Springer.

Jones, R. N. (2000). Managing Uncertainty in Climate Change Projections–Issues for Impact Assessment. *Climatic change,* 45**,** 403-419.

Juston, J. M., Kauffeldt, A., Montano, B. Q., Seibert, J., Beven, K. J. & Westerberg, I. K. (2013). Smiling in the Rain: Seven Reasons to Be Positive About Uncertainty in Hydrological Modelling. *Hydrological Processes,* 27**,** 1117-1122.

Karunasinghe, D. S. K. & Liong, S.-Y. (2006). Chaotic Time Series Prediction with a Global Model: Artificial Neural Network. *Journal of Hydrology,* 323**,** 92-105.

Kasenow, M. (2001). *Applied Ground-Water Hydrology and Well Hydraulics*, Water Resources Publications.

Kavetski, D., Kuczera, G. & Franks, S. W. (2006). Bayesian Analysis of Input Uncertainty in Hydrological Modeling: 2. Application. *Water Resources Research,* 42**,** W03408.

Kaye, N. R., Hartley, A. & Hemming, D. (2012). Mapping the Climate: Guidance on Appropriate Techniques to Map Climate Variables and Their Uncertainty. *Geosci. Model Dev.,* 5**,** 245-256.

Kiczko, A., Romanowicz, R. J. & Napiórkowski, J. J. (Year) Published. A Study of Flow Conditions Aimed at Preserving Valuable Wetland Areas in the Upper Narew Valley Using Gsa-Glue Methodology. Proceedings 21st International Conference on Informatics for Environmental Protection, 2007. 175-183.

Kitanidis, P. K. & Bras, R. L. (1980). Real-Time Forecasting with a Conceptual Hydrologic Model: 2. Applications and Results. *Water Resources Research,* 16**,** 1034-1044.

Kliot, N. (1993). *Water Resources and Conflict in the Middle East*, Taylor & Francis Group.

Knutti, R. & Sedlacek, J. (2013). Robustness and Uncertainties in the New Cmip5 Climate Model Projections. *Nature Clim. Change,* 3**,** 369-373.

Koirala, S., Yeh, P., Oki, T. & Kanae, S. (2010). Fully Dynamic Groundwater Representation in the Matsiro Land Surface Model. *Annual Journal of Hydraulic Engineering, Japan Society of Civil Engineers,* 54**,** 37-42.

Krause, P., Boyle, D. P. & Bäse, F. (2005). Comparison of Differenct Efficiency Criteria for Hydrological Model Assessment. *Advances in Geosciences,* 5**,** 89-97.

Krusz, A. Visualizing Data Uncertainty: An Experiment with D3.Js. *Velir Blog* [Online]. Available from: http://www.velir.com/blog/index.php/2013/07/11/visualizing-data-uncertainty-an-experiment-with-d3-js/.

Kundzewicz, Z. W. (1995). *New Uncertainty Concepts in Hydrology and Water Resources,* Cambridge, Cambridge University Press.

Laio, F., Di Baldassarre, G. & Montanari, A. (2009). Model Selection Techniques for the Frequency Analysis of Hydrological Extremes. *Water Resources Research,* 45**,** W07416.

Lambers, H., Chapin, F. S. & Pons, T. L. (2008). *Plant Physiological Ecology*, Springer New York.

Landau, D. P. & Binder, K. (2005). *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press.

Lane, S., Brookes, C., Kirkby, M. & Holden, J. (2004). A Network-Index-Based Version of Topmodel for Use with High-Resolution Digital Topographic Data. *Hydrological processes,* 18**,** 191-201.

Latham, J., Cumani, R., Rosati, I. & Bloise, M. (2014). Fao Global Land Cover (Glc-Share) Beta-Release 1.0. FAO, Land and Water Division.

Lawson, W. (1865). *The Geography of River Systems*.

Ldas (1999). Mapped Static Vegetation Data. NASA.

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W. & Spracklen, D. V. (2011). Emulation of a Complex Global Aerosol Model to Quantify Sensitivity to Uncertain Parameters. *Atmos. Chem. Phys.,* 11**,** 12253-12273.

Lee, L. A., Pringle, K. J., Reddington, C. L., Mann, G. W., Stier, P., Spracklen, D. V., Pierce, J. R. & Carslaw, K. S. (2013). The Magnitude and Causes of Uncertainty in Global Model Simulations of Cloud Condensation Nuclei. *Atmos. Chem. Phys.,* 13**,** 8879-8914.

Legates, D. R. & Mccabe, G. J. (1999). Evaluating the Use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation. *Water Resources Research,* 35**,** 233-241.

Liang, X., Lettenmaier, D. P., Wood, E. F. & Stephen, J. B. (1994). A Simple Hydrologically Based Model of Land Surface Water and Energy Fluxes for General Circulation Models. *Journal of Geophysical Research,* 99**,** 415-428.

Light, A. & Bartlein, P. J. (2004). The End of the Rainbow? Color Schemes for Improved Data Graphics. *Eos, Transactions American Geophysical Union,* 85**,** 385-391.

Lighthill, M. & Whitham, G. (1955). On Kinematic Waves. I. Flood Movement in Long Rivers. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences,* 229**,** 281-316.

Lin, G.-F. & Chen, L.-H. (2004). A Non-Linear Rainfall-Runoff Model Using Radial Basis Function Network. *Journal of Hydrology,* 289**,** 1-8.

Lin, G.-F. & Chen, L.-H. (2005). Time Series Forecasting by Combining the Radial Basis Function Network and the Self-Organizing Map. *Hydrological Processes,* 19**,** 1925-1937.

Liu, Y. & Gupta, H. V. (2007). Uncertainty in Hydrologic Modeling: Toward an Integrated Data Assimilation Framework. *Water Resources Research,* 43**,** W07401.

Loos, S., Middelkoop, H., Van Der Perk, M. & Van Beek, R. (2009). Large Scale Nutrient Modelling Using Globally Available Datasets: A Test for the Rhine Basin. *Journal of hydrology,* 369**,** 403-415.

Loucks, D. P., Van Beek, E., Stedinger, J. R., Dijkman, J. P. M. & Villars, M. T. (2005). Chapter 9: Model Senstivity and Uncertainty Analysis. *In:* UNESCO (ed.) *Water Resources*

Systems Planning and Management. Paris: United Nations Educational Scientific and Cultural Organization.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, J., Yang, L. & Merchant, J. W. (2000). Development of a Global Land Cover Characteristics Database and Igbp Discover from 1-Km Avhrr Data. *International Journal of Remote Sensing,* 21**,** 1303-1330.

Lowi, M. R. (1995). *Water and Power: The Politics of a Scarce Resource in the Jordan River Basin*, Cambridge University Press.

Ludwig, F. & Voss, F. (2009). Watch Techical Report 12: Description of the Watch Modelling Framework. Wageningen University, CESR.

Mantovan, P. & Todini, E. (2006). Hydrological Forecasting Uncertainty Assessment: Incoherence of the Glue Methodology. *Journal of Hydrology,* 330**,** 368-381.

Mantovan, P., Todini, E. & Martina, M. L. V. (2007). Reply to Comment by Keith Beven, Paul Smith and Jim Freer on "Hydrological Forecasting Uncertainty Assessment: Incoherence of the Glue Methodology". *Journal of Hydrology,* 338**,** 319-324.

Marshall, L., Nott, D. & Sharma, A. (2005). Hydrological Model Selection: A Bayesian Alternative. *Water Resources Research,* 41**,** W10422.

Martinec, J., Rango, A. & Robert, R. (1998). Snowmelt Runoff Model (Srm) User's Manual. Available: http://hydrolab.arsusda.gov/cgi-bin/srmhome [Accessed 19/6/2013].

Maupin, M. A. & Barber, N. L. (2005). *Estimated Withdrawals from Principal Aquifers in the United States, 2000*, US Department of the Interior, US Geological Survey.

Mccarthy, J. J. (2001). *Climate Change 2001: Impacts, Adaptation, and Vulnerability: Contribution of Working Group Ii to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press.

Mccuen, R., Knight, Z. & Cutter, A. (2006). Evaluation of the Nash–Sutcliffe Efficiency Index. *Journal of Hydrologic Engineering,* 11**,** 597-602.

Mckay, M. D., Beckman, R. J. & Conover, W. J. (1979). Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics,* 21**,** 239-245.

Mcleod, A. I., Noakes, D. J., Hipel, K. W. & Thompstone, R. M. (1987). Combining Hydrologic Forecasts. *Journal of Water Resources Planning and Management,* 113**,** 29-41.

Mcmichael, C. E., Hope, A. S. & Loaiciga, H. A. (2006). Distributed Hydrological Modelling in California Semi-Arid Shrublands: Mike She Model Calibration and Uncertainty Estimation. *Journal of Hydrology,* 317**,** 307-324.

Meehl, G. A., Boer, G. J., Covey, C., Latif, M. & Stouffer, R. J. (2000). The Coupled Model Intercomparison Project (Cmip). *Bulletin of the American Meteorological Society,* 81**,** 313-318.

Meigh, J. R., Mckenzie, A. A. & Sene, K. J. (1999). A Grid-Based Approach to Water Scarcity Estimates for Eastern and Southern Africa. *Water Resources Management,* 13**,** 85-115.

Mekonnen, M. M. & Hoekstra, A. Y. (2011). National Water Footprint Accounts: The Green,Blue and Grey Water Footprint of Production and Consumption. *Value of Water Research Report Series.* Delft: The Netherlands: UNESCO-IHE.

Mendelsohn, J. & Obeid, S. E. (2004). *Okavango River: The Flow of a Lifeline*, Struik.

Met Office. (2015). *Uk Climate Averages* [Online]. Available: http://www.metoffice.gov.uk/public/weather/climate/ [Accessed 03/08/2015.

Mishra, V., Cherkauer, K. A. & Shukla, S. (2010). Assessment of Drought Due to Historic Climate Variability and Projected Future Climate Change in the Midwestern United States. *Journal of Hydrometeorology,* 11**,** 46-68.

Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M. & New, M. (2004). A Comprehensive Set of High-Resolution Grids of Monthly Climate for Europe and the Globe: The Observed Record (1901–2000) and 16 Scenarios (2001–2100). *Tyndall Centre for Climate Change Research Working Paper,* 55**,** 25.

Modarres, R. & Ouarda, T. B. M. J. (2013). Generalized Autoregressive Conditional Heteroscedasticity Modelling of Hydrologic Time Series. *Hydrological Processes,* 27**,** 3174-3191.

Monteith, J. L. (1965). Evaporation and Environment. *Symposium of the Society for Experimental Biology,* 19**,** 205-234.

Moore, R. (1985). The Probability-Distributed Principle and Runoff Production at Point and Basin Scales. *Hydrological Sciences Journal,* 30**,** 273-297.

Moore, R. J. (2007). The Pdm Rainfall-Runoff Model. *Hydrology and Earth System Sciences,* 11**,** 483-499.

Moors, E. J., Groot, A., Biemans, H., Van Scheltinga, C. T., Siderius, C., Stoffel, M., Huggel, C., Wiltshire, A., Mathison, C. & Ridley, J. (2011). Adaptation to Changing Water Resources in the Ganges Basin, Northern India. *Environmental Science & Policy,* 14**,** 758-769.

Moradkhani, H., Hsu, K. L., Gupta, H. & Sorooshian, S. (2005a). Uncertainty Assessment of Hydrologic Model States and Parameters: Sequential Data Assimilation Using the Particle Filter. *Water Resources Research,* 41.

Moradkhani, H., Sorooshian, S., Gupta, H. V. & Houser, P. R. (2005b). Dual State–Parameter Estimation of Hydrological Models Using Ensemble Kalman Filter. *Advances in Water Resources,* 28**,** 135-147.

Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics,* 33**,** 191-174.

Mugetti, A., Brieva, C., Giangiobbe, S., Gallicchio, E., Pacheco, F., Pagani, A., Calcano, A., González, S., Natale, O., Faure, M., Rafaelli, S., Mangnani, C., Moyano, M. C., Seoane, R. & Enriquez, I. (2004). Patagonian Shelf, Giwa Regional Assessment 38. *In:* UNEP (ed.) *Global International Waters Assessment.* University of Kalmar, Sweden: United Nations Environment Programme.

Muleta, M. K. & Nicklow, J. W. (2005). Sensitivity and Uncertainty Analysis Coupled with Automatic Calibration for a Distributed Watershed Model. *Journal of Hydrology,* 306**,** 127-145.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M. & Döll, P. (2014). Sensitivity of Simulated Global-Scale Freshwater Fluxes and Storages to Input Data, Hydrological Model Structure, Human Water Use and Calibration. *Hydrol. Earth Syst. Sci. Discuss.,* 11**,** 1583-1649.

Mulvany, T. (1851). On the Use of Self-Registering Rain and Flood Gauges in Making Observations of the Relations of Rainfall and Flood Discharges in a Given Catchment. *Proceedings of the institution of Civil Engineers of Ireland,* 4**,** 18-33.

Mutua, F. M. (1994). The Use of the Akaike Information Criterion in the Identification of an Optimum Flood Frequency Model. *Hydrological Sciences Journal,* 39**,** 235-244.

Najafi, M. R., Moradkhani, H. & Jung, I. W. (2011). Assessing the Uncertainties of Hydrologic Model Selection in Climate Change Impact Studies. *Hydrological Processes,* 25**,** 2814-2826.

Nash, J. (1957). The Form of the Instantaneous Unit Hydrograph. *IAHS Publ,* 45**,** 114-121.

Nash, J. E. & Sutcliffe, J. V. (1970). River Flow Forecasting through Conceptual Models Part I -- a Discussion of Principles. *Journal of Hydrology,* 10**,** 282-290.

Nayak, P., Sudheer, K., Rangan, D. & Ramasastri, K. (2004). A Neuro-Fuzzy Computing Technique for Modeling Hydrological Time Series. *Journal of Hydrology,* 291**,** 52-66.

Neuman, S. (2003). Maximum Likelihood Bayesian Averaging of Uncertain Model Predictions. *Stochastic Environmental Research and Risk Assessment,* 17**,** 291-305.

New, M., Anderson, K., Fung, F. & Thornton, P. (2011). Sr08: The Possible Impacts of High Levels of Climate Change in 2060 and Implications for Migration. *UK Government Foresight Project, Migration and Global Change.* UK: Government Office for Science.

Ng, T. L., Eheart, J. W. & Cai, X. M. (2010). Comparative Calibration of a Complex Hydrologic Model by Stochastic Methods Glue and Pest. *Transactions of the Asabe,* 53**,** 1773-1786.

Ngo-Duc, T., Polcher, J. & Laval, K. (2005). A 53-Year Forcing Data Set for Land Surface Models. *Journal of Geophysical Research: Atmospheres (1984–2012),* 110.

Nijssen, B., O'donnell, G. M., Hamlet, A. F. & Lettenmaier, D. P. (2001a). Hydrologic Sensitvity of Global Rivers to Climate Change. *Climatic Change,* 50**,** 143-175.

Nijssen, B., O'donnell, G. M., Lettenmaier, D. P., Lohmann, D. & Wood, E. F. (2001b). Predicting the Discharge of Global Rivers. *Journal of Climate,* 14**,** 3307-3323.

Noaa. (2015). *Climate Normals* [Online]. Available: https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals [Accessed 03/08/2015.

O'hagan, A. (2006). Bayesian Analysis of Computer Code Outputs: A Tutorial. *Reliability Engineering & System Safety,* 91**,** 1290-1300.

Oehler, S., Hartl, D., Lopez, R., Malak, R. & Lagoudas, D. (2012). Design Optimization and Uncertainty Analysis of Sma Morphing Structures. *Smart Materials and Structures,* 21**,** 094016.

Ogata, T., Valeriano, O., Yoshimura, C., Liengcharernsit, W. & Hirabayashi, Y. (2012). Past and Future Hydrological Simulations of Chao Phraya River Basin. *Journal of Japan Society of Civil Engineers Ser. B,* 1**,** 68.

Oki, T., Agata, Y., Kanae, S., Saruhashi, T., Yang, D. & Musiake, K. (2001). Global Assessment of Current Water Resources Using Total Runoff Integrating Pathways. *Hydrological Sciences Journal,* 46**,** 983-995.

Osborne, T. (2009). A User Guide for Climgen: A Flexible Tool for Generating Montly Climate Data Sets and Scenarios. University of East Anglua, Norwich: Climate Research Unit.

Palisade Corporation (2010). Guide to Using @Risk: Risk Analysis and Simulation Add-in for Microsoft Excel. Ithaca, NY, USA.

Pappenberger, F. & Beven, K. J. (2006). Ignorance Is Bliss: Or Seven Reasons Not to Use Uncertainty Analysis. *Water Resources Research,* 42**,** W05302.

Parajka, J., Naeimi, V., Blöschl, G., Wagner, W., Merz, R. & Scipal, K. (2006). Assimilating Scatterometer Soil Moisture Data into Conceptual Hydrologic Models at the Regional Scale. *Hydrology and Earth System Sciences Discussions,* 10**,** 353-368.

Park, N. & Ip, W. H. (2010). *Advances in Geosciences: Hydrological Science*, World Scientific Publishing Company, Incorporated.

Pebesma, E. J., Switzer, P. & Loague, K. (2005). Error Analysis for the Evaluation of Model Performance: Rainfall–Runoff Event Time Series Data. *Hydrological Processes,* 19**,** 1529-1548.

Peel, M. C., Finlayson, B. L. & Mcmahon, T. A. (2007). Updated World Map of the Köppen-Geiger Climate Classification. *Hydrol. Earth Syst. Sci.,* 11**,** 1633-1644.

Penman, H. (1956). Evaporation: An Introductory Survey. *Neth. J. Agric. Sci,* 4**,** 9-29.

Penman, H. L. (1948). Natural Evaporation from Open Water, Bare Soil and Grass. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences,* 193**,** 120-145.

Pest. (2014). *Pest: Uncertainty Analysis* [Online]. Available: http://www.pesthomepage.org/Uncertainty_Analysis.php [Accessed 15/06/2014.

Pinder, G. F. & Jones, J. F. (1969). Determination of the Ground-Water Component of Peak Discharge from the Chemistry of Total Runoff. *Water Resources Research,* 5**,** 438-445.

Piniewski, M., Laizé, C. L., Acreman, M. C., Okruszko, T. & Schneider, C. (2014). Effect of Climate Change on Environmental Flow Indicators in the Narew Basin, Poland. *Journal of Environmental Quality,* 43**,** 155-167.

Postel, S. (1999). *Pillar of Sand: Can the Irrigation Miracle Last?*, W.W. Norton.

Prideaux, B. & Cooper, M. (2009). *River Tourism*, CABI.

Priestley, C. H. B. & Taylor, R. J. (1972). On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. *Monthly Weather Review,* 100**,** 81-92.

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y. & Wisser, D. (2014). Hydrological Droughts in the 21st Century, Hotspots and Uncertainties from a Global Multimodel Ensemble Experiment. *Proceedings of the National Academy of Sciences,* 111**,** 3262-3267.

Qi, W., Zhang, C., Chu, J. & Zhou, H. (2013). Sobol' Sensitivity Analysis for Topmodel Hydrological Model: A Case Study for the Biliu River Basin, China. *Journal of Hydrology and Environment Research, Technical Paper,* 1.

Qu, J., Xu, Z., Long, Q., Wang, L., Shen, X., Zhang, J. & Cai, Y. (2004). East China Sea, Giwa Regional Assessment 36. *In:* UNEP (ed.) *Global International Waters Assessment.* University of Kalmar, Sweden: United Nations Environment Programme.

Ratto, M., Tarantola, S. & Saltelli, A. (2001). Sensitivity Analysis in Model Calibration: Gsa-Glue Approach. *Computer Physics Communications,* 136**,** 212-224.

Rawlins, M. A., Lammers, R. B., Frolking, S., Fekete, B. & Vorosmarty, C. (2003). Simulating Pan-Arctic Runoff with a Macro-Scale Terrestrial Water Balance Model. *Hydrological Processes,* 17**,** 3632-2539.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J. & Dmip Participants, A. (2004). Overall Distributed Model Intercomparison Project Results. *Journal of Hydrology,* 298**,** 27-60.

Refsgaard, J. C. & Knudsen, J. (1996). Operational Validation and Intercomparison of Different Types of Hydrological Models. *Water Resources Research,* 32**,** 2189-2202.

Reusser, D. E., Blume, T., Schaefli, B. & Zehe, E. (2009). Analysing the Temporal Dynamics of Model Performance for Hydrological Models. *Hydrology And Earth System Sciences,* 13**,** 999-1018.

Riad, S., Mania, J., Bouchaou, L. & Najjar, Y. (2004). Predicting Catchment Flow in a Semi-Arid Region Via an Artificial Neural Network Technique. *Hydrological Processes,* 18**,** 2387-2393.

Ryan, S. (2009). Murray-Darling Basin – Integrated Management in a Large, Dry and Thirsty Basin. *Handbook of Catchment Management.* Wiley-Blackwell.

Sadegh, M. & Vrugt, J. A. (2013). Bridging the Gap between Glue and Formal Statistical Approaches: Approximate Bayesian Computation. *Hydrol. Earth Syst. Sci.,* 17**,** 4831-4850.

Sage: Center for Sustainability and the Global Environment (2014). Global River Discharge Database (Rivdis). *In:* SAGE: CENTER FOR SUSTAINABILITY AND THE GLOBAL ENVIRONMENT (ed.). http://www.sage.wisc.edu/riverdata/.

Saleh, D. K. (2010). *Stream Gage Descriptions and Streamflow Statistics for Sites in the Tigris River and Euphrates River Basins, Iraq*, US Department of the Interior, US Geological Survey.

Saltelli, A., Chan, K. & Marian Scott, E. (eds.) 2000. *Sensitivity Analysis,* Chichester, UK: John Wiley & Sons Ltd.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. & Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer,* Chichester, England, John Wiley & Sons Ltd.

Salter, P. J. & Williams, J. B. (1969). The Influence of Texture on the Moisture Characteristics of Soil. *Journal of Soil Science,* 20**,** 126-131.

Savoskul, O. S. & Smakhtin, V. (2013). *Glacier Systems and Seasonal Snow Cover in Six Major Asian River Basins: Hydrological Role under Changing Climate,* Colombo, Sri Lanka, International Water Management Institute (IMWI).

Saxton, K. & Rawls, W. (2006). Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society of America Journal,* 70**,** 1569-1578.

Schaake, J. C., Hamill, T. M., Buizza, R. & Clark, M. (2007). Hepex: The Hydrological Ensemble Prediction Experiment. *Bulletin of the American Meteorological Society,* 88**,** 1541-1547.

Schaefli, B. & Gupta, H. V. (2007). Do Nash Values Have Value? *Hydrological Processes,* 21**,** 2075-2080.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y.,

Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L. & Kabat, P. (2014). Multimodel Assessment of Water Scarcity under Climate Change. *Proceedings of the National Academy of Sciences,* 111**,** 3245-3250.

Schneider, S. H. (1983). Co2, Climate and Society: A Brief Overview. *Social Science Research and Climate Change.* Springer.

Schneider, S. H. & Kuntz-Duriseti, K. (2002). Uncertainty and Climate Change Policy. *Climate change policy: a survey***,** 53-87.

Scurlock, J. M. O., Asner, G. P. & Gower, S. T. (2001). Worldwide Historical Estimates of Leaf Area Index, 1932-2000. Tennessee: NASA.

Sellers, P. J., Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D. A., Zhang, C., Collelo, G. D. & Bounoua, L. (1996). A Revised Land Surface Parameterization (Sib2) for Atmospheric Gcms. Part I: Model Formulation. *Journal of Climate,* 9**,** 676-705.

Sene, K. J., Tate, E. L. & Farquharson, F. a. K. (2001). Sensitivity Studies of the Impacts of Climate Change on White Nile Flows. *Climatic Change,* 50**,** 177-208.

Shaw, E. M., Beven, K. J., Chappell, N. A. & Lamb, R. (2011). *Hydrology in Practice,* Abingdon, Oxon, Spon Press.

Shen, Y., Oki, T., Utsumi, N., Kanae, S. & Hanasaki, N. (2008). Projection of Future World Water Resources under Sres Scenarios: Water Withdrawal/Projection Des Ressources En Eau Mondiales Futures Selon Les Scénarios Du Rsse: Prélèvement D'eau. *Hydrological Sciences Journal,* 53**,** 11-33.

Sherman, L. K. (1932). Streamflow from Rainfall by the Unit-Graph Method. *Eng. News Record,* 108**,** 501-505.

Sherman, L. K. (1949). The Unit Hydrograph Method. *Physics of the Earth. OE Menizer Ed. Dover Publications, Inc. New York, NY***,** 514-525.

Shiklomanov, I. A. & Rodda, J. C. (2003). *World Water Resources at the Beginning of the Twenty-First Century*, Cambridge University Press Cambridge.

Shrestha, R. R., Peters, D. L. & Schnorbus, M. A. (2013). Evaluating the Ability of a Hydrologic Model to Replicate Hydro-Ecologically Relevant Indicators. *Hydrological Processes***,** n/a-n/a.

Shukla, S. & Lettenmaier, D. P. (2011). Seasonal Hydrologic Prediction in the United States: Understanding the Role of Initial Hydrologic Conditions and Seasonal Climate Forecast Skill. *Hydrol. Earth Syst. Sci. Discuss.,* 8**,** 6565-6592.

Siderius, C., Biemans, H., Wiltshire, A., Rao, S., Franssen, W., Kumar, P., Gosain, A., Van Vliet, M. & Collins, D. (2013). Snowmelt Contributions to Discharge of the Ganges. *Science of The Total Environment,* 468**,** S93-S101.

Singh, V. P., Sharma, N. & Ojha, C. S. P. (2004). *The Brahmaputra Basin Water Resources*, Kluwer Academic Publishers.

Singh, V. P. & Woolhiser, D. A. (2002). Mathematical Modeling of Watershed Hydrology. *Journal of Hydrological Engineering,* 7**,** 270-293.

Smith, K. A. (2011). *Evaluation of Land Management Impacts on Low Flows in Northern England.* MSc by Research, Durham University.

Smith, L. A. & Stern, N. (2011). Uncertainty in Science and Its Role in Climate Policy. *Philosophical Transactions of the Royal Society A.,* 369**,** 1-24.

Smith, M. B., Koren, V., Reed, S., Zhang, Z., Zhang, Y., Moreda, F., Cui, Z., Mizukami, N., Anderson, E. A. & Cosgrove, B. A. (2012a). The Distributed Model Intercomparison Project – Phase 2: Motivation and Design of the Oklahoma Experiments. *Journal of Hydrology,* 418–419**,** 3-16.

Smith, M. B., Koren, V., Zhang, Z., Zhang, Y., Reed, S. M., Cui, Z., Moreda, F., Cosgrove, B. A., Mizukami, N. & Anderson, E. A. (2012b). Results of the Dmip 2 Oklahoma Experiments. *Journal of Hydrology,* 418–419**,** 17-48.

Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F. & Cong, S. (2004). The Distributed Model Intercomparison Project (Dmip): Motivation and Experiment Design. *Journal of Hydrology,* 298**,** 4-26.

Snidvongs, A. & Teng, S.-K. (2006). Mekong River, Giwa Regional Assessment 55. *In:* UNEP (ed.) *Global International Waters Assessment.* University of Kalmar, Sweden: United Nations Environment Programme.

Sobol', I. M. (1993). Sensitivity Estimates for Non-Linear Mathematical Models. *Mathematical Modelling and Computational Experiment,* 1**,** 407-414.

Soil Information for Environmental Modeling and Ecosystem Management. (2014). *Soil Texture Classification Triangle* [Online]. Available: http://www.soilinfo.psu.edu/index.cgi?soil_data&conus&data_cov&fract&methods [Accessed 24/05/2014.

Sperna Weiland, F. C. (2011). Hydrological Impacts of Climate Change: Interpretation of Uncertainties Introduced by Global Models of Climate and Hydrology. *Utrecht Studies in Earth Sciences,* 6.

Sperna Weiland, F. C., Van Beek, L. P. H., Kwadijk, J. C. J. & Bierkens, M. F. P. (2010). The Ability of a Gcm-Forced Hydrological Model to Reproduce Global Discharge Variability. *Hydrol. Earth Syst. Sci.,* 14**,** 1595-1621.

Sperna Weiland, F. C., Van Beek, L. P. H., Kwadijk, J. C. J. & Bierkens, M. F. P. (2012). Global Patterns of Change in Discharge Regimes for 2100. *Hydrol. Earth Syst. Sci.,* 16**,** 1047-1062.

Spiegelhalter, D., Pearson, M. & Short, I. (2011). Visualizing Uncertainty About the Future. *Science,* 333**,** 1393-1400.

Spruill, C. A., Workman, S. R. & Taraba, J. L. (2000). Simulation of Daily and Monthly Stream Discharge from Small Watersheds Using the Swat Model. *Transactions of the American Society of Agricultural Engineers,* 43**,** 1431-1439.

Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L., Van Lanen, H., Sauquet, E., Demuth, S., Fendekova, M. & Jordar, J. (2010). Streamflow Trends in Europe: Evidence from a Dataset of near-Natural Catchments. *Hydrology and Earth System Sciences,* 14**,** p. 2367-p. 2382.

Sugawara, M. (1967). The Flood Forecasting by a Series Storage Type Model. *Int. Symposium Floods and their Computation.* Association of Hydrologic Sciences.

Swart, N. C., Fyfe, J. C., Hawkins, E., Kay, J. E. & Jahn, A. (2015). Influence of Internal Variability on Arctic Sea-Ice Trends. *Nature Clim. Change,* 5**,** 86-89.

Takata, K., Emori, S. & Watanabe, T. (2003). Development of the Minimal Advanced Treatments of Surface Interaction and Runoff. *Global and Planetary Change,* 38**,** 209-222.

Tang, Q. (2006). *A Distributed Biosphere Hydrological Model for Continental Scale River Basins.* Ph.D., University of Tokyo.

Tang, Q., Oki, T., Kanae, S. & Hu, H. (2007). Hydrological Cycles Change in the Yellow River Basin During the Last Half of the Twentieth Century. *Journal of Climate,* 21**,** 1790-1806.

Tao, H. (2008). *Calibration, Sensitivity and Uncertainty Analysis in Surface Water Quality Modeling.* PhD, Tufts University.

Tateishi, R., Bayaer, Ghar, M. A., Al-Bilbisi, H., Tsendayush, J., Shalaby, A., Kasimu, A., Hoan, N. T., Kobayashi, T., Alsaaideh, B., Rahman, M. M., Tsvengee, E., Yamada, Y. & Kajikawa, S. (2008). A New Global Land Cover Map, Glcnmo. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* Beijing.

Taylor, K. E., Stouffer, R. J. & Meehl, G. A. (2011). An Overview of Cmip5 and the Experiment Design. *Bulletin of the American Meteorological Society,* 93**,** 485-498.

Tebaldi, C. & Knutti, R. (2007). The Use of the Multi-Model Ensemble in Probabilistic Climate Projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 365**,** 2053-2075.

Terragis. (2007). *Soil Moisture Classification* [Online]. UNSW. Available: http://www.terragis.bees.unsw.edu.au/terraGIS_soil/sp_water-soil_moisture_classification.html [Accessed 15/06/2014.

Theis, C. V. (1935). The Relation between the Lowering of the Piezometric Surface and the Rate and Duration of Discharge of a Well Using Ground Water Storage. *Transactions-American Geophysical Union,* 16**,** 519-524.

Thiemann, M., Trosset, M., Gupta, H. & Sorooshian, S. (2001). Bayesian Recursive Parameter Estimation for Hydrologic Models. *Water Resources Research,* 37**,** 2521-2535.

Thompson, J., Green, A., Kingston, D. & Gosling, S. (2013). Assessment of Uncertainty in River Flow Projections for the Mekong River Using Multiple Gcms and Hydrological Models. *Journal of hydrology,* 486**,** 1-30.

Thoms, M. C., Rayburg, S. C. & Neave, M. R. (2007). Physocal Diversity and Assessment of a Large River System: The Murrat-Darling Basin, Australia. *In:* GUPTA, A. (ed.) *Large Rivers: Geomorphology and Management.* Chichester, England: John Wiley & Sons.

Thornthwaite, C. W. (1948). An Approach toward a Rational Classification of Climate. *Geographical review,* 38**,** 55-94.

Times Books (2009). *The Times Concise Atlas of the World,* London, HarperCollins Publishers.

Todini, E. (1996). The Arno Rainfall—Runoff Model. *Journal of Hydrology,* 175**,** 339-382.

Todini, E. (2007). Hydrological Catchment Modelling: Past, Present and Future. *Hydrology and Earth System Sciences,* 11**,** 468-482.

Todini, E. & Mantovan, P. (2007). Comment On: 'On Undermining the Science?' by Keith Beven. *Hydrological Processes,* 21**,** 1633-1638.

Trambauer, P., Maskey, S., Winsemius, H., Werner, M. & Uhlenbrook, S. (2013). A Review of Continental Scale Hydrological Models and Their Suitability for Drought Forecasting in (Sub-Saharan) Africa. *Physics and Chemistry of the Earth, Parts A/B/C,* 66**,** 16-26.

Tsyban, A. V., Titova, G. D., Schuka, S. A., Ranenko, V. V. & A., I. Y. (2005). Russian Arctic, Giwa Regional Assessment 1a. *In:* UNEP (ed.) *Global International Waters Assessment.* University of Kalmar, Sweden: United Nations Environment Programme.

University of Cambridge. (2009). *The Darwin Cluster: Introduction and Brief History* [Online]. Available: http://www.hpc.cam.ac.uk/services/darwin.html [Accessed 16/03/2015.

University of Leicester. (2015). *Alice Architecture* [Online]. Available: http://www2.le.ac.uk/offices/itservices/ithelp/services/hpc/alice/architecture [Accessed 16/03/2015.

Usda Soil Conservation Service (1987). Soil Mechanics Level I Module 3 - Usda Textural Soil Classification Study Guide. United States Department of Agriculture.

Vajda, S., Valko, P. & Turanyi, T. (1985). Principal Component Analysis of Kinetic Models. *International Journal of Chemical Kinetics,* 17**,** 55-81.

Van Beek, L. P. H. & Bierkens, M. F. P. (2008). The Global Hydrological Model Pcr-Globwb: Conceptualization,

Parameterization and Verification. Utrecht, The Netherlands: Department of Physical Geography, Utrecht University.

Van Beek, L. P. H., Wada, Y. & Bierkens, M. F. P. (2011). Global Monthly Water Stress: 1. Water Balance and Water Availability. *Water Resour. Res.,* 47**,** W07517.

Vázquez, R. F., Feyen, L., Feyen, J. & Refsgaard, J. C. (2002). Effect of Grid Size on Effective Parameters and Model Performance of the Mike-She Code. *Hydrological Processes,* 16**,** 355-372.

Velázquez, J. A., Schmid, J., Ricard, S., Muerth, M. J., Gauvin St-Denis, B., Minville, M., Chaumont, D., Caya, D., Ludwig, R. & Turcotte, R. (2013). An Ensemble Approach to Assess Hydrological Models' Contribution to Uncertainties in the Analysis of Climate Change Impact on Water Resources. *Hydrol. Earth Syst. Sci.,* 17**,** 565-578.

Verdhen, A. & Chahar, B. R. (2012). Efficiency of Snowmelt Modelling Approaches in Watershed Models. *Internatinoal Conference on SWAT-2012.* Dehli India.

Verzano, K., Bärlund, I., Flörke, M., Lehner, B., Kynast, E., Voß, F. & Alcamo, J. (2012). Modeling Variable River Flow Velocity on Continental Scale: Current Situation and Climate Change Impacts in Europe. *Journal of Hydrology,* 424**,** 238-251.

Villarini, G., Serinaldi, F. & Krajewski, W. F. (2008). Modeling Radar-Rainfall Estimation Uncertainties Using Parametric and Non-Parametric Approaches. *Advances in Water Resources,* 31**,** 1674-1686.

Vörösmarty, C. J., Federer, C. A. & Schloss, A. L. (1998). Potential Evaporation Functions Compared on Us Watersheds: Possible Implications for Global-Scale Water Balance and Terrestrial Ecosystem Modeling. *Journal of Hydrology,* 207**,** 147-169.

Vörösmarty, C. J., Moore, B., Grace, A. L., Gildea, M. P., Melillo, J. M., Peterson, B. J., Rastetter, E. B. & Steudler, P. A. (1989). Continental Scale Models of Water Balance and Fluvial Transport: An Application to South America. *Global biogeochemical cycles,* 3**,** 241-265.

Voß, F. & Flörke, M. (2010). Watch Technical Report No 23. Spatially Explicit Estimates of Past and Present Manufacturing and Energy Water Use. *In:* WATCH (ed.). CESR.

Vrugt, J., Gupta, H., Bouten, W. & Sorooshian, S. (Year) Published. A Shuffled Complex Evolution Metropolis Algorithm for Optimization and Uncertainty Assessment of Hydrological Model Parameters. AGU Fall Meeting Abstracts, 2002. 03.

Vrugt, J. A., Diks, C. G., Gupta, H. V., Bouten, W. & Verstraten, J. M. (2005). Improved Treatment of Uncertainty in Hydrologic Modeling: Combining the Strengths of Global Optimization and Data Assimilation. *Water Resources Research,* 41.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W. & Sorooshian, S. (2003a). Effective and Efficient Algorithm for Multiobjective

Optimization of Hydrologic Models. *Water Resources Research,* 39.

Vrugt, J. A., Gupta, H. V., Bouten, W. & Sorooshian, S. (2003b). A Shuffled Complex Evolution Metropolis Algorithm for Optimization and Uncertainty Assessment of Hydrologic Model Parameters. *Water Resources Research,* 39**,** 1201.

Vrugt, J. A. & Sadegh, M. (2013). Toward Diagnostic Model Calibration and Evaluation: Approximate Bayesian Computation. *Water Resources Research,* 49**,** 4335-4345.

Vrugt, J. A., Ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M. & Higdon, D. (2009a). 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.

Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M. & Robinson, B. A. (2008). Treatment of Input Uncertainty in Hydrologic Modeling: Doing Hydrology Backward with Markov Chain Monte Carlo Simulation. *Water Resources Research,* 44.

Vrugt, J. A., Ter Braak, C. J. F., Gupta, H. V. & Robinson, B. A. (2009b). Equifinality of Formal (Dream) and Informal (Glue) Bayesian Approaches in Hydrologic Modeling? *Stochastic Environmental Research and Risk Assessment,* 23**,** 1011-1026.

Wada, Y., Van Beek, L. P. H., Van Kempen, C. M., Reckman, J. W. T. M., Vasak, S. & Bierkens, M. F. P. (2010). Global Depletion of Groundwater Resources. *Geophys. Res. Lett.,* 37.

Wada, Y., Van Beek, L. P. H., Viviroli, D., Dürr, H. H., Weingartner, R. & Bierkens, M. F. P. (2011). Global Monthly Water Stress: 2. Water Demand and Severity of Water Stress. *Water Resour. Res.,* 47**,** W07518.

Wagener, T., Mcintyre, N., Lees, M., Wheater, H. & Gupta, H. (2003). Towards Reduced Uncertainty in Conceptual Rainfall-Runoff Modelling: Dynamic Identifiability Analysis. *Hydrological Processes,* 17**,** 455-476.

Wagener, T. & Wheater, H. S. (2006). Parameter Estimation and Regionalization for Continuous Rainfall-Runoff Models Including Uncertainty. *Journal of Hydrology,* 320**,** 132-154.

Wainwright, H. M., Finsterle, S., Jung, Y., Zhou, Q. & Birkholzer, J. T. (2014). Making Sense of Global Sensitivity Analyses. *Computers & Geosciences,* 65**,** 84-94.

Wallace, J. S. & Gregory, P. J. (2002). Water Resources and Their Use in Food Production Systems. *Aquatic Sciences - Research Across Boundaries,* 64**,** 363-375.

Wang, L., Koike, T., Yang, K., Jackson, T. J., Bindlish, R. & Yang, D. (2009a). Development of a Distributed Biosphere Hydrological Model and Its Evaluation with the Southern Great Plains Experiments (Sgp97 and Sgp99). *Journal of Geophysical Research: Atmospheres,* 114**,** D08107.

Wang, L., Koike, T., Yang, K. & Yeh, P. J.-F. (2009b). Assessment of a Distributed Biosphere Hydrological Model against Streamflow and Modis Land Surface Temperature in the Upper Tone River Basin. *Journal of Hydrology,* 377**,** 21-34.

Ward, R. C., Ward, R. C. & Robinson, M. (2000). *Principles of Hydrology*, McGraw-Hill.

Warren, R., De La Nava Santos, S., Arnell, N. W., Bane, M., Barker, T., Barton, C., Ford, R., Füssel, H. M., Hankin, R. K. S., Klein, R., Linstead, C., Kohler, J., Mitchell, T. D., Osborn, T. J., Pan, H., Raper, S. C. B., Riley, G., Schellnhüber, H. J., Winne, S. & Anderson, D. (2008). Development and Illustrative Outputs of the Community Integrated Assessment System (Cias), a Multi-Institutional Modular Integrated Assessment Approach for Modelling Climate Change. *Environmental Modelling & Software,* 23**,** 592-610.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O. & Schewe, J. (2014). The Inter-Sectoral Impact Model Intercomparison Project (Isi–Mip): Project Framework. *Proceedings of the National Academy of Sciences,* 111**,** 3228-3232.

Weedon, G. P., Gomes, S., Viterbo, P., Osterle, H., Adam, J. C., Bellouin, N., Boucher, O. & Best, M. (2010). Technical Report No 22. The Watch Forcing Data 1958-2001: A Meteorological Forcing Dataset for Land Surface - and Hydrological - Models. European Union Water and Global Change Project.

Weedon, G. P., Prudhomme, C., Crooks, S., Ellis, R. J., Folwell, S. S. & Best, M. J. (2014). Evaluating the Performance of Hydrological Models Via Cross-Spectral Analysis: Case Study of the Thames Basin, Uk. *Journal of Hydrometeorology*.

Wei, S., Song, J. & Khan, N. I. (2012). Simulating and Predicting River Discharge Time Series Using a Wavelet-Neural Network Hybrid Modelling Approach. *Hydrological Processes,* 26**,** 281-296.

Werth, S., Guntner, A., Petrovic, S. & Schmidt, R. (2009). Integration of Grace Mass Variations into a Global Hydrological Model. *Earth and Planetary Science Letters,* 277**,** 166-173.

Wesseling, C. G., Karssenberg, D. J., Burrough, P. A. & Van Deursen, W. P. A. (1996). Integrating Dynamic Environmental Models in Gis: The Development of a Dynamic Modelling Language. *Transactions in GIS,* 1**,** 40-48.

Wilby, R. L. (2005). Uncertainty in Water Resource Model Parameters Used for Climate Change Impact Assessment. *Hydrological Processes,* 19**,** 3201-3219.

Wilby, R. L. & Dessai, S. (2010). Robust Adaptation to Climate Change. *Weather,* 65**,** 180-185.

Wilby, R. L. & Harris, I. (2006). A Framework for Assessing Uncertainties in Climate Change Impacts: Low-Flow Scenarios for the River Thames, Uk. *Water Resour. Res.,* 42**,** W02419.

Wilkinson, R. D. (2013). Approximate Bayesian Computation (Abc) Gives Exact Results under the Assumption of Model Error. *Statistical Applications in Genetics and Molecular Biology.*

Willmott, C. J. (1981). On the Validation of Models. *Physical Geography,* 2**,** 184-194.

Willmott, C. J. & Matsuura, K. (2005). Advantages of the Mean Absolute Error (Mae) over the Root Mean Square Error (Rmse) in Assessing Average Model Performance. *Climate Research,* 30**,** 79-82.

Wilson, M. & Henderson-Sellers, A. (1985). A Global Archive of Land Cover and Soils Data for Use in General Circulation Climate Models. *Journal of Climatology,* 5**,** 119-143.

Wisser, D., Fekete, B. M., Vörosmarty, C. J. & Schumann, A. H. (2010). Reconstructing 20th Century Global Hydrography: A Contribution to the Global Terrestrial Network-Hydrology (Gtn-H). *Hydrology and Earth System Sciences,* 14**,** 1-24.

Wood, E. F., Lettenmaier, D. P. & Zartarian, V. G. (1992). A Land-Surface Hydrology Parameterization with Subgrid Variability for General Circulation Models. *Journal of Geophysical Research: Atmospheres (1984–2012),* 97**,** 2717-2728.

Woodward, J. C., Macklin, M. G., Krom, M. D. & Williams, M. a. J. (2007). The Nile: Evolution, Quaternary River Environments and Material Fluxes. *In:* GUPTA, A. (ed.) *Large Rivers: Geomorphology and Management.* Chichester, England: John Wiley & Sons Ltd.

Zeng, X., Dickinson, R. E., Walker, A., Shaikh, M., Defries, R. S. & Qi, J. (2000). Derivation and Evaluation of Global 1-Km Fractional Vegetation Cover Data for Land Modeling. *Journal of Applied Meteorology,* 39**,** 826-839.

Zhao, Q., Ye, B., Ding, Y., Zhang, S., Yi, S., Wang, J., Shangguan, D., Zhao, C. & Han, H. (2013). Coupling a Glacier Melt Model to the Variable Infiltration Capacity (Vic) Model for Hydrological Modeling in North-Western China. *Environmental Earth Sciences,* 68**,** 87-101.

Zhao, R.-J. & Liu, H. (1992). The Xinanjiang Model Applied in China. *Journal of Hydrology,* 135**,** 371-381.