Improving global and catchment estimates of runoff through computationally-intelligent ensemble approaches

Applications of intelligent multi-model combination, cross-scale model comparisons, ensemble analyses, and new model parameterisations

Jamal Zaherpour BSc, MSc
E-mail: zaherpour@gmail.com

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

March 2018

Host institute address

Sir Clive Granger Building, University Park Campus, University of Nottingham, Nottingham, NG7 2RD
Tel. +44 (0)115 951 5428

Home institute address

Masir Gostar Jonoub Eng. Co. (Masirgostaran), Narvan 1 Alley, Narvan Street, Eram Street, Shiraz, Iran, Postal Code: 7143714327

Tel. + 98(0)7132298593

E-mail: masirgostaran@yahoo.com

Extended thesis abstract

Water related problems (scarcity, availability and hazards) together form one of the three major crises (the two other are food and energy) for today and the future across the globe (World Economic Forum, 2016, Schewe et al., 2014a, Hanasaki et al., 2013, Rockstrom et al., 2009). Water crises are widespread and heterogeneous around the world and climate change and socioeconomic drivers are expected to accelerate these problems (Veldkamp, 2017). To deal with the above concerns, mitigation and adaptation strategies are developed at different scales (global, regional and local). Developing these strategies, as well as selecting the most appropriate one to the problem of interest, should ideally benefit from the highest possible accuracy in estimates of the hydrological cycle and water resources. More reliable decisions can in turn be made by applying tools and techniques that enhance decision makers' perception of the hydrological cycle, particularly extreme events i.e. droughts and floods. These tools should also facilitate insights into the cycle: ideally by reference to hydrological indicators, spatially (globally and locally) and temporally (present day and future). Global and catchment scale hydrological models (GHMs and CHMs) have been used as such tools that along with advances in data acquisition, analytical techniques and computation power offer powerful tools for modelling natural processes and provide useful insights into the hydrological cycle.

GHMs have a shorter history of emergence and application than CHMs. GHMs have been developed and applied from 1986 in recognition of the fact that hydrological processes and water resources are global phenomena and should be treated at global scale (Bierkens, 2015). A GHM is a pragmatic trade-off between a faithful representation of the diversity of hydrological processes found across the world's catchments, and a generalised and simplified representation of hydrological processes that can support multi-decadal, generalised hydrological simulations at global scales. Compared to hydrological models designed for catchment-scale simulations (Arnold et al., 1993; Krysanova et al., 1998; Lindstrom et al., 2010), GHMs employ coarser spatial discretisation and model the global land surface in a single instantiation.

The global scope of GHMs, limited availability and quality of observed discharge data across the global domain and their use of spatially generalised parameters make them more difficult to calibrate than catchment hydrological models. Whilst examples of calibrated GHMs do exist (Müller Schmied et al., 2016), the majority of GHMs are uncalibrated (Gosling et al., 2016; Hattermann et al., 2017). This lack of calibration, coupled with the diversity of simplifications employed in the hydrological process representations, means that there can be large inconsistency in the skill, bias and uncertainty of an individual GHM at different locations, as well as large inconsistencies between different GHMs at any given location (van Huijgevoort et al., 2013). This spatial inconsistency means that GHMs risk becoming a "jungle of models" (Kundzewicz, 1986) in which it can be difficult to determine where a particular GHM output is likely to be capable of delivering optimal hydrological simulations. It also makes it dangerous to assume that any individual GHM will be an adequate basis for making projections at any given location, even if the model's ability to replicate observed data in particular catchments is enhanced through the acquisition of higher quality input data or efforts to improve process representations (Liu et al., 2007). To an extent, these arguments are also applicable to CHMs because whilst they have been shown to generally perform better than GHMs in model evaluation studies, ensembles of such models still result in an uncertainty range when the models are run with identical inputs (Hattermann et al., 2017; Hattermann et al., 2018).

To minimise the challenge of varying outputs from different models, several model inter-comparison projects (MIPs) have been undertaken around the world (Henderson-Sellers et al., 1995, Entin et al., 1999, Guo and Dirmeyer, 2006, Koster et al., 2006, Harding et al., 2011). These projects usually use standard modelling baselines to deal with discrepancies between model outputs. This results in higher consistency in the climate forcings input to the models (where applicable), their process representations (e.g. the simulation of human impacts such as water abstractions), and the temporal and spatial resolutions of their simulations. This way, model outputs are directly comparable to each other, which supports diagnostic inter-comparisons between them (Bierkens, 2015). One of the largest, ongoing MIPs (whose data are used in this thesis) is the Inter-Sectoral Impact Model Inter-comparison Project (ISIMIP) (Schellnhuber et al., 2014, Warszawski et al., 2014). ISIMIP is a community-driven effort by more than 130 modelling groups, that covers different sectors including water (both global and catchment hydrological modelling communities). Outputs from ISIMIP are widely used in different projects, such as reports of the International Panel on Climate Change (IPCC) (http://www.ipcc.ch/).

MIPs including ISIMIP provide a unique opportunity to access data from different models and to assess their relative performance. It also facilitates continuous model improvement via the inclusion of new schemes (e.g. human impacts such as dams, reservoirs and water abstractions) accompanied by dozens of models, as well as communication between modelling groups working in the same or different sectors. Nonetheless, they do not fully address the challenge of spatial inconsistencies between models, as well as the question of what ensemble representative to select for use when trying to improve the reliability of decision-making. There remain other shortcomings or unexplored aspects within MIPs (particularly ISIMIP as this research's focus MIP), hence areas of further research and potential improvement in model evaluation and application which will be addressed later in this introduction.

The question of how to address the challenges of spatial inconsistency in hydrological models has been a feature of catchment-scale model research for several decades. In answering it, catchment modellers have recognised that reliance on a single, inconsistent model is inherently risky and should be avoided (Marshall et al., 2006; Shamseldin et al., 1997). Instead, they have developed ways to take advantage of the diversity of outputs (Clemen, 1989) generated by different models by using optimised mathematical combination methods to deliver a combined output that performs better than the individual models from which it was created (Hagedorn et al., 2005). This general approach—known as multi-model combination (MMC)—has been an important focus of catchment hydrological modelling studies, especially over the last two decades (Abrahart and See, 2002; Ajami et al., 2006; Arsenault et al., 2015; Azmi et al., 2010; de Menezes et al., 2000; Fernando et al., 2012; Jeong and Kim, 2009; Marshall et al., 2007; Marshall et al., 2006; Moges et al., 2016; Nasseri et al., 2014; Sanderson and Knutti, 2012; Shamseldin et al., 1997). Given its demonstrable potential in catchment studies, it is perhaps surprising that the potential of applying MMC to GHMs has yet to be explored.

A wide range of techniques can be used to generate an MMC solution. The simplest example is the calculation of the arithmetic mean of the input models (commonly referred to as an Ensemble Mean (EM)). More sophisticated techniques employ weighted schemes (Arsenault et al., 2015), with the differential weightings applied to each input model reflecting their relative strengths or weaknesses. The mathematical approach taken to determining the weights depends on the objective of the MMC. Where the primary objective is to minimise the difference between the MMC solution and observed data (i.e. maximise the predictive performance), without explicitly accounting for model or parameter uncertainty, the use of multiple linear regression (Doblas-Reyes et al., 2005) or machine learning algorithms to 'learn' the optimal set weights to apply to each MMC input model is a popular approach (Marshall et al., 2007). The use of algorithms such as artificial neural networks (ANNs) (Shamseldin et al., 1997; Xiong et al., 2001) or gene expression programming (GEP) (Barbulescu and Bautu, 2010; Bărbulescu and Băutu, 2009; Fernando et al., 2012) to define non-linear weighting schemes have proven to be particularly effective. This is down to their ability to generate optimised, non-linear schemes rapidly, without the need for any prior knowledge of the model parameters.

Where there is a desire to account for and minimise model and parameter uncertainty in the weighting scheme, Bayesian averaging methods are required (Ajami et al., 2007; Hoeting et al., 1999). These optimise the weights according to the posterior performance of the MMC solution under the prior probabilities of model parameter values (Duan et al., 2007; Vrugt and Robinson, 2007; Ye et al., 2004). However, these methods require knowledge of the probability density functions (PDFs) for each of the MMC's input model parameters (or at least their maximum likelihood estimates (Ye et al., 2004)). This makes their use in the MMC of GHMs problematic because the number of parameters used in GHMs is particularly high, the parameters vary considerably between models, and the PDFs of the parameters in a GHM can be extremely difficult to specify over a global domain. Consequently, the PDFs for GHM parameters are seldom specified and, in many cases, remain unknown.

On this basis, this research applies the first MMC approach to tackle the challenge of inconsistency between multitude of models and using them optimally. To this end, itexplores the potential of MMC for addressing the challenge of spatial inconsistency in simulations from several GHMs (for the first time) and CHMs, by combining outputs from diverse sets of GHMs and CHMs using the machine learning tool GEP (Ferreira, 2001; Ferreira, 2006) across up to forty major catchments (Papers II and IV of the thesis). In each catchment, the MMC's ability to replicate the observed monthly runoff is compared against that of the EM, each of the individual models from which the MMC is derived, and the best-performing individual model from the ensemble.

In addition, this thesis addresses four other areas of further research spotted within the framework of the ISIMIP as follow with more detailed rationale behind performing each part of the research and their contribution to knowledge presented in Section 1.3. It should, however, be emphasised that these explorations are applicable to any MIPs and not limited to the ISIMIP.

First, despite the development of modelling protocols as common modelling baselines in different phases of the ISIMIP, its framework lacks an evaluation protocol. This is important towards model appraisal and, in turn, model application and improvement in a comparative manner. This research tackles this problem by setting a more comprehensive evaluation framework to an ensemble of GHMs and their ensemble mean (EM) that informs, via a multi-dimensional assessment, which models are best, where, and according to which hydrological indicator.

Second, as discussed above both global and catchment models provide valuable information for decision-making and scientific understanding, but while GHMs provide output data for almost all parts of the world (applicable for aggregated assessment of e.g. climate change impacts), the applicability of outputs generated by a CHM is usually limited to the catchment for which the model has been calibrated. Given the huge resources required for developing each model type, it is of great importance to understand their relative performance and whether they can be used interchangeably both for current and future climates. Nevertheless, knowledge of the relative spreads in simulations/projections from GHM and CHM ensembles (within and outside ISIMIP) is limited to cross-scale inter-comparison of small ensembles, precluding a robust ensemble comparison. This research uses, for the first time, large multi-model ensembles of GHMs and CHMs to explore whether there are systematic differences between projections of runoff change from two ensembles comprising different types of hydrological model. It also investigates the effect of different degrees of global-mean warming on runoff for each catchment as indicated by each model type to the end of 21st century.

Third, while cross-scale inter-comparison of large GHM and CHM ensembles is more informative than their small ensembles, application of more robust representatives of each model ensemble (other than EM) was found an area of further study as the EM showed to not be necessarily the best ensemble representative. Given the potential found in the intelligent MMC approach that provided runoff simulations overall better than the EM, its capacity in filling the gap between two large ensembles of

GHMs and CHMs is sought. In addition, this study examines the performance of MMC when a group of GHMs and a group of CHMs are pooled together as a 'super-ensemble'. The performance of MMC from the super-ensemble is then compared with the performance of the EM of the super-ensemble and MMCs from GHM and CHM ensembles to unravel potential benefits of the approach.

Fourth and finally, GHM outputs available and applied at the time of this research were from two different phases of ISIMIP i.e. Fast Track and ISIMIP2a of which in the latter GHMs were run under two different conditions of with and without the inclusion of anthropogenic impacts in the GHMs' parametrisation. Human impact parameterisations (HIP) has been an important step forwards to make modelling conditions as close as possible to the real world. Nonetheless, no comprehensive evaluation and comparison of GHMs under the two abovementioned conditions had been carried out, precluding test of models' capacity to represent human activities and their impacts on assessment of freshwater resources and hydrological extremes. The final part of this research (Paper V) explores change in the performance of five GHMs that participated in ISIMIP2a and had human impacts in their parameterisation.