Uncertainty assessment in river flow projections for Ethiopia's Upper Awash Basin using multiple GCMs and hydrological models

Chan, W.C.H[a,1] , Thompson, J.R[a], Taylor, R.G.[a], Nay, A.E.[a], Ayenew, T.[b], MacDonald, A.M.[c], and Todd, M.C.[d]

[a]UCL Department of Geography, University College London, UK; [b]School of Earth Sciences, University of Addis Ababa, Ethiopia; [c]British Geological Survey, Edinburgh, UK; [d]Department of Geography, University of Sussex UK.

[1]Current institution: Department of Meteorology, University of Reading, UK.
Email : wilson.chan@pgr.reading.ac.uk

Uncertainty assessment in river flow projections for Ethiopia's Upper Awash Basin using multiple GCMs and hydrological models

Uncertainty in climate change impacts on river discharge in Ethiopia's Upper Awash Basin is assessed using five MIKE SHE hydrological models, six CMIP5 GCMs, and two RCP scenarios for the period 2071–2100. Hydrological models vary in their spatial distribution and process representations of unsaturated and saturated zones. Very good performance is achieved for 1975–1999 (NSE: 0.65–0.8; *r:* 0.79–0.93). GCM-related uncertainty dominates variability in projections of high and mean discharges (mean: -34% to +55% for RCP4.5, -42% to +195% for RCP8.5). Although GCMs dominate uncertainty in projected low flows, inter-hydrological model uncertainty is considerable (RCP4.5: -60% to +228%, RCP8.5: -86% to +337%). ANOVA uncertainty attribution reveals that GCM-related uncertainty occupies an average 68% of total uncertainty for median and high flows and hydrological models no more than 1%. For low flows, hydrological model uncertainty occupies an average 18% of total uncertainty; GCM-related uncertainty remains substantial (average 28%).

Keywords: Upper Awash Basin; hydrological modelling; climate change; uncertainty

1. Introduction

Climate change is expected to alter the quantity, quality and timing of river flow, groundwater recharge and other hydrological processes (Jiménez Cisneros *et al.* 2014). In turn, modifications to the distribution of freshwater resources have the potential to significantly impact global water and food security (Betts *et al.* 2018). Quantifying the hydrological impacts of climate change is critical to informing strategies designed to sustain water security and maintain aquatic environments and the ecosystem services they provide (Thompson *et al.* 2014a). The hydrological impacts of climate change are commonly evaluated by perturbing meteorological inputs to hydrological models with climate projections derived from General Circulation Model (GCM) simulations of alternative radiative forcing scenarios (e.g. Todd *et al.* 2011; Thompson *et al.* 2013). However, these climate change impact assessments are subject to multiple sources of uncertainty that cascade down through each step of the impact modelling chain (Wilby and Dessai 2010).

The cascade of uncertainty in hydrological climate change impact assessments typically consists of uncertainties related to emission scenarios, GCMs and hydrological models. The first source of uncertainty relates to diverging trajectories of greenhouse gas concentrations (and other radiative drivers) that arising from uncertainty in economic development, technological change and climate mitigation policies. Current projections are expressed through a range of Representative Concentration Pathways (RCPs) (IPCC 2014). GCM-related uncertainty is the next, and most frequently studied, source of uncertainty. It arises from differences in spatial resolution, parameterization and process descriptions between GCMs (e.g. Todd *et al.* 2011; Hattermann *et al.* 2018). Integrated multi-scale analyses have shown that at a global scale GCM-related uncertainty dominates uncertainty in hydrological projections (e.g. Todd *et al.* 2011; Vetter *et al.* 2015; Krysanova *et al.* 2017).

Uncertainty related to the structure of hydrological models is often neglected despite the wide range of hydrological model codes applied in climate change impact studies. While previous studies agree that hydrological model uncertianty is generally smaller than GCM-related uncertainty, it cannot be ignored (e.g. Thompson *et al.* 2013). Hydrological model uncertainty is epistemic and mainly arises from a lack of knowledge in the representation of natural hydrological processes within models (Beven 2016). Hydrological model structures can differ according to 1) parameterization, 2) process coupling, 3) model domain discretization, and 4) classification of spatially-distributed catchment characteristics including land use, geology, soil, and hydrological processes (Butts *et al.* 2004). Acceptable and similar performance can be achieved using different model structures (i.e. model equifinality; Beven 1993). Whilst different hydrological models may perform similarly for a baseline period, they can produce different responses when forced with the same climate change projections (e.g. Poulin *et al.* 2011; Thompson *et al.* 2013).

This study investigates the cascade of uncertainty for projections of river discharge under climate change within Ethiopia's Upper Awash Basin with a particular focus on hydrological model structure uncertainty. Uncertainty associated with model parameters during calibration can also lead to model equifinality but are not explicitly addressed in this study. Hydrological model structure uncertainty is assessed by developing five individual hydrological models of varying complexity using the same model code, MIKE SHE. While previous studies (such as QUEST-GSI; Todd *et al.* 2011 and ISI-MIP; Hattermann *et al.* 2018) have investigated hydrological model structural uncertainty using a number of different model codes, this study assesses structural uncertainty by varying the representation of spatial variability and the complexity of process descriptions within a single hydrological model code. This study is the first time the MIKE SHE modelling system has been applied to the Upper Awash Basin and is likely the first study that has undertaken sensitivity analyses for alternative combinations of process representations within MIKE SHE for an African catchment.

1.1 The Upper Awash Basin

The Awash River Basin (Fig. 1) is Ethiopia's most developed and economically important river basin and the first in which modern agriculture and large-scale irrigation were established (Berhe *et al.* 2013). It has an area of ~113,709 km^2 and is bordered to the west by the Abbay Basin (Blue Nile). The main river runs for 1200 km from the central Ethiopian highlands at ~3000 mamsl (metres above mean sea level) within the East African Rift Valley in a northeastern direction to the endorheic Lake Abe at ~250 mamsl (Fig.1b). Based on topography, climate and elevation it can be divided in to three zones: Upper Awash, Middle Awash and Lower Awash (Taddese *et al.* 2003).

The Upper Awash Basin (UAB) is home to at least 15.7 million people (~17% of Ethiopia's population) and includes Ethiopia's capital, Addis Ababa. The UAB terminates at the Koka Dam and Reservoir, 75 km southeast of Addis Ababa. It has an area of 11,500 km^2 and accounts for 7% of the Awash River Basin. The UAB has frequently been the sole focus of hydrological research given the significantly altered river regime downstream of the Koka Dam (Berhe *et al.* 2013; Müller *et al.* 2016). Precipitation within the UAB is substantially modulated by the seasonal migration of the Inter-tropical Convergence Zone. Mean annual rainfall varies from 1400 mm in the headwaters of the Ethiopian highlands to 800 mm near Koka Dam. The climate is characterized by a short rainy season in spring (February to March) and a main summer wet season (June to September). Potential evapotranspiration (PET) is inversely correlated to altitude (Berhe *et al.* 2013) and varies from 1810 mm in the humid Ethiopian highlands to over 2300 mm in the arid lower valley. Sandy clay loam (42%) and clay (39%) are the dominant soil types. The hydrostratigraphy of the UAB is dominated by three main units (Kebede, 2013; Jira 2019): an upper or shallow (80 to 150 m depth) aquifer of Quaternary alluvium as well as weathered and fractured basalt; an intermediary (~100 m thick) pyroclastic confining bed; and a confined aquifer (>300 m thick) of Tertiary volcanics.

Precipitation over the Ethiopian highlands is strongly influenced by large-scale controls on climate variability such as the El Niño Southern Oscillation (ENSO). El Niño phases are associated with below-average summer rainfall and are considered to be a major driver of past drought episodes (e.g. Seleshi and Zanke 2004; Philip *et al.* 2018). Despite its importance, previous research has shown that the GCMs of the Coupled Model Intercomparison Project Phase 5 (CMIP5) are limited in their ability to simulate ENSO behaviour in the observed record and show little consensus over how its behaviour may change in the future (Kociuba and Power 2014; Chen *et al.* 2016). Projections for East Africa over the 21st Century generally point towards increasing temperature (and hence PET), in response to enhanced global mean temperatures, and precipitation, linked to enhanced wet seasons under both RCP4.5 and RCP8.5 greenhouse-gas concentration trajectories (Niang *et al.* 2014).

The scarcity of high quality and complete hydro-meteorological data has been a particular challenge in the Awash Basin and neighbouring East African catchments (Mekonnen *et al.* 2009). Sparse hydrologgcal monitoring below Koka Dam imposes further restrictions, beyond the dam's regulation of discharge, on the spatial extent of hydrological investigations. In this study, discharge data from two gauging stations in the UAB (Fig. 1) were selected on the basis on the duration, continuity and integrity of their records. Gauging stations at Melka and Hombole have drainage areas of 4456 km^2 and 7656 km^2, respectively. These stations were used for model calibration / validation. Models were forced with daily precipitation data from 11 rain gauges (Fig. 1), which were selected based on a review of data quality including initial double mass analysis. Temperature records are limited and only available at an elevation of 2354 m. This record was adjusted for four 500 m elevation ranges between 1500 m and 3500 m to account for spatially varying PET, the other meteorological parameter forcing the models which was calculated using the adjusted temperature records and the Hargreaves method (Hargreaves and Samani 1985).

[FIGURE 1]

2. Methods

2.1 Model development, calibration and validation

Each of the five hydrological models of the UAB was developed using the MIKE SHE modelling system which is capable of simulating the major processes of the land phase of the hydrological cycle (e.g. Graham and Butts 2005). MIKE SHE is commonly described as being deterministic, fully-distributed and physically-based but its modular structure is flexible and includes process descriptions of varying levels of complexity, some of which are conceptual and semi-distributed in nature (Refsgaard *et al.* 2010). The spatial distribution of model inputs across the MIKE SHE model grid can also be readily modified and these inputs can either be uniformly or spatially distributed. By altering the complexity of process descriptions and spatial distribution, MIKE SHE can be used to explore the impacts of hydrological model structural uncertainty by developing models of alternative process descriptions and spatial distributions (e.g. Rochester 2010; Robinson 2018; Vansteenkiste *et al.* 2014a, b). The development of models of the UAB adheres to the approaches used by MIKE SHE modelling of other large river systems (e.g. Anderson *et al.* 2001; Thompson *et al.* 2013; Hudson and Thompson 2019). MIKE 11, a 1D hydraulic model, is dynamically coupled to MIKE SHE and simulates channel flow (Thompson *et al.* 2004). Table 1 summarises the data requirements and sources of data for each component of the five alternative MIKE SHE / MIKE 11 models of the UAB.

[TABLE 1]

The five models span a range of commonly used hydrological model structures. They were developed using existing process representations available within the MIKE SHE modelling system, which include relatively simple spatially uniform, conceptual approaches through to spatially-distributed, physically-based process descriptions as well as a combination of the two. Whilst each model employs the same 1 km × 1 km grid size and a maximum time step of 1 day, they vary according to the computational approaches used to represent the unsaturated (UZ) and saturated zones (SZ) (Table 2). The models also employed two alternative approaches to the spatial distribution of surface soil parameters in the UZ. The first assumes uniform distribution of soil types; the second spatially varies soil classes based on the FAO Digital Soil Map of the World (FAO 1998). Each hydrological model is given a name according to the following three criteria: 1) whether the model has a uniformly distributed (U) or spatially distributed (D) unsaturated zone, 2) whether the model has conceptual (C) or physically-based (P) representation of unsaturated and saturated flow, and 3) whether it used Gravity Flow (G) or Richards equation (R) to describe unsaturated flow.

[TABLE 2]

A split-sample calibration/validation approach (Klemeš 1986) was adopted for all five hydrological models. In each case, the periods 1975–1987 and 1988–1999, for which the most complete hydro-meteorological datasets are available, were used for calibration and validation, respectively. Calibration/validation was based on comparison of observed and simulated discharge at the two gauging stations with model performance being assessed using multiple statistical measures; Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe 1970), percentage deviation (Dv; Henriksen *et al.* 2003) and Pearson product moment correlation (r). Model performance using these statistics was assessed based on the scheme proposed by Henriksen *et al.* (2003). This scheme has previously been used in a number of modelling studies (e.g. Thompson *et al.* 2013; Ho *et al.* 2016; Hudson and Thompson 2019). Previous research has suggested that NSE tend to be more sensitive to high and extreme flows (Pushpalatha *et al.* 2012; Krause *et al.* 2005). As the hydrology at the UAB is relatively flashy and the difference between high and low flows can exceed 150 m^3s^{-1} in a single year, model performance was also assessed through the calculation of daily and monthly logNSE using the logarithmic values of river discharge to give more weight to low flows (Krause *et al.* 2005). In order to minimize the risks of over-parameterization, the number of parameters subject to calibration was kept as low as possible according to the framework outlined by Refsgaard (1997). Horizontal and vertical hydraulic conductivity was varied during calibration for all hydrological models with the exception of Model 5, which employed the conceptual linear reservoir representation of the SZ. The time constants for the interflow and baseflow reservoirs were varied in the calibration of Model 5 instead. Hydraulic conductivity within the UZ was varied in models 1 and 2, which both employed a conceptual representation of the unsaturated zone. Bypass fraction for the dominant soil type was varied in all models.

2.2 Climate change scenarios

Each of the five calibrated and validated MIKE SHE/MIKE 11 models was perturbed with projected changes in precipitation and PET from six CMIP5 GCMs (Table 3) for the RCP4.5 and RCP8.5 scenarios. Changes in mean annual precipitation and temperature over the UAB for these six GCMs broadly represent the range of change across the complete CMIP5 ensemble with the exception of a few notable outliers (Supplementary Fig.1). In this way, the range of projected climate change across the CMIP5 ensemble is represented by a similar number of GCMs to the hydrological models that were developed for the UAB. The delta-factor approach was used to establish future precipitation and PET time series. Initially monthly change factors for precipitation (%) and mean, maximum and minimum temperatures (°C) were derived by comparing basin-wide projections from each GCM for the baseline (1975-1999) and future periods (2071-2100). Daily baseline observed precipitation and temperature data were perturbed by these monthly delta factors (e.g. Anandhi *et al.* 2011). PET was then recalculated using the new temperature series and the Hargreaves method (Hargreaves and Samani 1985). The delta-factor approach is widely used in hydrological climate change impact assessments (e.g. Arnell 2003; Poulin *et al.* 2011; Ho *et al. 2016*; Hudson and Thompson 2019). It should be noted that this approach retains the climate variability of the baseline period but does not consider modifications to rainfall intensity (Fowler *et al.* 2007). Additional simulations employed perturbed precipitation whilst using baseline PET and vice versa. This enabled assessment of the relative importance of inter-GCM uncertainty in precipitation and PET to overall uncertainty (Gosling *et al.* 2011; Thompson *et al.* 2013).

[TABLE 3]

2.3 Uncertainty analysis

Systematic analysis of multiple sources of uncertainty enables quantification of the relative magnitude of each source to overall uncertainty. For example, Bosshard *et al.* (2013) used three-way ANOVA to quantify the relative dominance of different sources of uncertainty along the entire impact modelling chain. ANOVA has since been employed to quantify uncertainties in multi-catchment global investigations (Karlsson *et al.* 2016; Vetter 2015, 2017; Krysanova *et al.* 2017; Hatterman *et al.* 2018). It enables variance decomposition where overall impact assessment uncertainty is decomposed into elements (i.e. individual sources of uncertainty) and interactions among them.

Following the approach of Bosshard *et al.* (2013), overall uncertainty (Y_Q) is defined in terms of annual Q10, Q50 and Q90 river discharges (i.e. discharges equalled or exceeded for 10, 50 and 90 percent of the time in each year, respectively) for the climate change signal for each of the impact modelling chain combinations comprising five hydrological models, six GCMs, and two RCP scenarios (eq. 1). The river discharge quantiles selected match those of recent multi-site impact assessments (Vetter *et al.* 2015; Krysanova *et al.* 2017; Hattermann *et al.* 2018) and characterise high, median and low flows, respectively. Each source of uncertainty that is considered is an 'effect' that is hypothesized to influence overall climate change signal variability. ANOVA, conducted using SPSS 22, splits the total sum of squares into sum of squares for each effect and their interactions (eq. 2). The variance fraction (n^2) (between 0 and 1) is then calculated for each effect and represents the percentage contribution of each effect and interactions (eq. 3).

$$Y_Q = Q_Q^{CC} - Q_Q^{BL} \qquad (1)$$

Q = monthly river discharge quantiles (Q10, Q50, Q90)
CC = scenario
BL = baseline
Y = overall uncertainty

$$SST = SS_{HM} + SS_{GCM} + SS_{RCP} + SS_{HM*GCM} + SS_{HM*RCP} + SS_{GCM*RCP} + SS_{HM*GCM*RCP} \qquad (2)$$

SST = total sum of squares
SS_{HM}, SS_{GCM} and SS_{RCP} correspond to SST partitioned into sum of squares of the effects (hydrological model, GCM, RCP)

$$\eta^2_{GCM} = \frac{1}{I}\sum_{i=1}^{I}\frac{SS_{GCM}}{SST},$$

$$\eta^2_{HM} = \frac{1}{I}\sum_{i=1}^{I}\frac{SS_{HM}}{SST},$$

$$\eta^2_{RCP} = \frac{1}{I}\sum_{i=1}^{I}\frac{SS_{RCP}}{SST} \quad (3)$$

I is the number of subsamples and $\eta^2$is the variance fraction for each effect.

3. Results

3.1 Model calibration and performance

Table 4 summarises the final values of the main calibration parameters for each of the five hydrological models. As described above, the alternative process representations means that not all of these parameters are relevant to each model and the parameters modified during calibration varied between models. The values of horizontal and vertical hydraulic conductivity within the SZ, representing the upper aquifer unit of Quaternary alluvium and weathered/fractured volcanics, are up to an order of magnitude larger in models 3 and 4 that use physically-based conceptualizations of the UZ, compared with Model 1 with a simpler representation of the UZ. It is probable that the higher hydraulic conductivities for the former are required to compensate for the longer transit times through the UZ as simulated by the Gravity Flow and Richard’s equation approaches. Indeed, the calibrated hydraulic conductivity values are more than an order of magnitude lower than those employed by Jira (2019) in SZ models using MODFLOW. The bypass fraction, which determines the proportion of precipitation that under certain conditions is routed directly to the water table, was similar across all models. Similarly, the same values were established for the UZ hydraulic conductivity for the two models (1 and 2) employing the conceptual 2-layer water balance approach.

[TABLE 4]

Fig. 2a&b shows observed and simulated mean monthly discharge at both gauging stations for the calibration and validation periods. Monthly mean discharges are shown in the interests of clarity since simulation results are provided for each of the five hydrological models. Statistical measures of model performance and their classification using the scheme of Henriksen *et al.* (2003) using both daily and monthly mean discharges at each station for both periods and all five models are presented in Table 5. Simulated and observed low flows in the dry season (DJF) and logNSE values for both daily and monthly mean discharges are shown in Table 6.

According to the performance scheme, monthly NSE values indicate at least 'very good' model performance (NSE > 0.7) for all models at both stations for the calibration period. NSE values for this period suggest improvement in performance between models 1 and 2 that is associated with the introduction of spatial distributed soil parameters within the UZ. Model performance according to NSE is more variable in the validation period although it is at least 'fair' across all models with the exception of Model 3 at Melka. NSE values indicate weaker performance at a daily time-steps compared to monthly metrics (see also Thompson *et al.* 2014b) with Model 5 providing the best performance at this shorter time step. In most cases *r* values are above 0.9 in the calibration period and are all above 0.8 with some exceeding 0.9 in the validation period.

[TABLE 5]

According to the values of D_v, all models provide at least 'fair' performance during the calibration period and indicate an overall underestimation of river flow at Melka and an overestimation at Hombole. Dv values are poorer for the validation period with classifications ranging from 'very poor' to 'fair'. All models tend to overestimate peak discharges in the validation period, albeit to varying degrees. This is particularly notable for models 4 and 5. Given the flashy nature of river flow in the UAB, Dv values are likely skewed by the inability of all of the models to fully capture the range of flow extremes. This is supported by the daily and monthly logNSE values which are generally lower than the NSE values across both the calibration and validation periods. In common with the NSE values, logNSE values are weaker at a daily time-step compared to monthly. Monthly logNSE values range from 0.22-0.58 in the calibration period and -0.21 to 0.68

in the validation period across both stations, suggesting a large disparity among the hydrological models in their ability to simulate low flows. The logNSE values at Hombole are generally better than at Melka across all hydrological models. The ability to reproduce low flows is notably poor in the validation period for models 4 and 5 with negative logNSE values.

[TABLE 6]

Fig. 2c-f show observed and simulated river regimes (monthly mean discharge) for all hydrological models in the calibration and validation periods. Although timing of the observed seasonal peak is reproduced very well for the calibration period, all models underestimate wet season July-August-September (JAS) discharges at Melka. This underestimation ranges from, on average, 14% for Model 4 to 35% for Model 3. Underestimation of wet season discharges also dominates at Hombole (three models) with a range of 6-15% (models 2 and 3, respectively). Models 4 and 5 overestimate discharge at this time by, on average, 10% and 17%, respectively. All of the models except Model 3 overestimate the low dry season (DJF) flows. This overestimation is more pronounced at Melka. The largest overestimates are for Model 4 (375% for Melka, 210% for Hombole) whereas the smallest are for Model 2 (74% at Melka, 1% at Hombole). The average absolute differences in DJF flows across all hydrological models are, however, relatively small (4.1 m^3s^{-1} and 4.9 m^3s^{-1} at Melka and Hombole, respectively).

Inter-hydrological model ranges in simulated discharge, indicated by the river regimes, are larger for the validation period. At Melka, mean overestimation of JAS discharge ranges from 16% (Model 5) to 52% (Model 3) whereas at Hombole different models are responsible for the smallest (16%, Model 2) and largest (32%, Model 5) overestimates. Underestimates in mean JAS discharges are restricted to Model 1 and are by comparison relatively small (2% at Melka and 4% at Hombole). The representation of dry season (DJF) river discharge remains an area of significant inter-hydrological model variability. Model 3 is, as for the calibration period, the only model to underestimate low flows (12% at Melka, 36% at Hombole). The remaining models all overestimate low flows by at least 100% at both stations. Overestimates at Melka range from a mean of 280% (Model 2) to 750% (Model 4) and at Hombole from 100% (Model 2) to 430% (Model 4).

[FIGURE 2]

A comparison of annual observed and simulated Q10 discharges (i.e. discharges equalled or exceeded for 10 percent of the year) at the two gauging stations for both the calibration and validation periods and each hydrological model demonstrates relatively good replication of high flow with reasonable *r* values cross the calibration (0.49-0.79) and validation (0.77-0.91) periods. The spread of model results (lowest *r* values) is larger for Model 3, which produces a notably higher bias at Melka. The *r* values for comparisons between annual observed and simulated Q90 discharges (discharges equalled or exceeded for 90 percent of the year) are less favourable for both the calibration (0.16-0.69) and validation (0.14-0.69) periods. Models 1, 4 and 5 produce particularly large differences between observed and simulated low flows. As demonstrated by the logNSE values, Models 2 and 3 are comparatively superior in simulating annual Q90 discharges and Model 3 is the only model to underestimate low flows. Inadequate low-flow performance likely stems from a relatively small weighting given to low flows in the calibration process and the dominant reliance on NSE and *r* that favour the replication of peak flows.

3.2 Projected changes in precipitation and PET

Catchment-averaged baseline and projected mean monthly precipitation and PET under the RCP4.5 and RCP8.5 scenarios for each of the six GCMs are shown in Fig. 3a-d. Percentage changes in mean annual precipitation and PET for each GCM and both scenarios are shown in Fig. 4e-h.

The magnitude of precipitation and PET changes, as well as inter-GCM variability, is considerably larger for RCP8.5 compared to RCP4.5. All but one GCM project increases in mean annual precipitation for both RCP scenarios. The exception is CSIRO-Mk3 that projects a decline of 7% for RCP8.5. This same GCM projects only very modest (0.5%) increases under RCP4.5. In both cases, increases are concentrated in the dry season and precipitation declines in most wet season months. The remaining five GCMs project increases at this time of year and consequentially higher annual precipitation totals. The GCM responsible for the largest increase varies between RCP scenarios. MPI-ESM-MR accounts for the largest increase for RCP4.5 (+31%) and CanESM2 (58%) for RCP8.5. It is also notable that a number of GCMs project a change in the temporal distribution of rainfall with an enhanced bimodal distribution developing in the form of a second, but smaller, rainy season between January and April. This is most notable for CSIRO-Mk3.6.0, IPSL-CM5A-MR and MPI-ESM-MR (RCP4.5) and CanESM2 and IPSL-CM5A-MR (RCP8.5).

All six GCMs project increases in annual PET for both RCP scenarios although the magnitude of the changes varies. For RCP4.5 increases in annual PET range from 2% (CanESM2) to 10% (CSIRO-Mk3). In most cases the magnitude of gains in PET increase under the RCP8.5 scenario although at the lowest extreme the 2% increase for CanESM2 is repeated. The largest increase of 17% is again for CSIRO-Mk3 and there is an approximate consistency in the relative order of magnitude of change for different GCMs between the two scenarios. CanESM2 is the only GCM that projects decline in any of the mean monthly PET totals. In most cases these declines occur in months when baseline PET is relatively low.

[FIGURE 3]

3.3 Projected changes in river discharge

Fig. 4 shows percentage changes in mean discharge from the baseline at both Melka and Hombole as simulated by the five hydrological models for each GCM and the two RCP scenarios. Near-consistent increases in mean catchment precipitation for the six GCMs are not repeated for mean discharge. A larger proportion, albeit still a minority of the GCM-hydrological model results, projects declines in mean discharge. This contrast reflects consistent increases in PET. For RCP4.5 declines in mean discharge at both stations for all five hydrological models are projected for CCSM4 and CSIRO-Mk3 whereas all but one hydrological model (Model 1) projects declines for HadGEM2-ES. The largest declines are projected for CSIRO-Mk3 with a mean across the five models of 35% at Melka and 32% at Hombole. At the other extreme, consistent increases across the five hydrological models are projected for CanESM2, IPSL-CM5A-MR and MPI-ESM-MR with the last projecting the largest mean increase across the hydrological models of 55% at both Melka and Hombole.

Under RCP8.5 fewer GGM-hydrological model results are associated with declines in mean discharge. The GCM responsible for the largest increases (mean 169% and 193% for Melka and Hombole, respectively) is CanESM2. At Melka all hydrological models project declines for CCSM4 and CSIRO-Mk3 whereas at Hombole one model (Model 1) projects an increase for the first of these GCMs. As with RCP4.5, CSIRO-Mk3 projects the largest declines (on average 47% and 37% for Melka and Hombole, respectively and larger than those for RCP4.5). For CCSM4, mean discharge is projected to decline by an average of 7% at Melka and Hombole excluding Model 1. In general, where increases in discharges are projected they are larger than those for RCP4.5 with the exception of MPI-ESM-MR (a smaller increase relative to RCP4.5). HadGEM2-ES projects an average increase of 28.5% at both stations compared to the mean decline of 5% for RCP4.5.

[FIGURE 4]

Following the approach of Gosling *et al.* (2011) and Thompson *et al.* (2013), Supplementary Fig. 3 and 4 show percentage change in mean annual discharge from the alternative application of scenario precipitation and PET for RCP4.5 and RCP8.5. For each GCM and hydrological model, projected changes in mean annual discharge are obtained from the alternative application of either scenario PET and scenario precipitation whilst retaining baseline time series for the other model input. The results confirm that inter-GCM uncertainty in projected river discharge mainly arises from uncertainty in precipitation projections. Where gains in precipitation are relatively small, enhanced PET results in reductions in discharge whereas higher PET magnifies drying under a projected reduction in precipitation. Mean discharge decreases in the majority of cases regardless of hydrological model if only perturbed PET is applied. Comparing the range of projected changes in mean annual discharge if only scenario precipitation is applied with the equivalent value if only scenario PET is applied shows that projected changes in mean annual discharge due to the application of perturbed precipitation alone is on average 7 times (15 times) larger than the corresponding range due to the application of perturbed PET alone under RCP4.5 (RCP8.5).

Fig. 4 shows that the direction of change in mean discharge remains the same regardless of hydrological model used with the few exceptions involving Model 1 described above. In contrast, and as also described above, different GCMs produce both increases and decreases in mean discharge. The range of change between hydrological models for an individual GCM is indicative of inter-hydrological model uncertainty. Similarly, the range of changes in projected discharges for different GCMs simulated by a single hydrological model provides an assessment of inter-GCM uncertainty (Dams *et al.* 2015). Table 7 reports the percentage changes in projected mean discharge for inter-hydrological model and inter-GCM uncertainty. These results demonstrate that inter-GCM uncertainty is larger than the uncertainty associated with the use of different hydrological models. For example, for the RCP4.5 scenario the percentage range in mean discharge at Melka simulated for a given GCM by the different hydrological models ranges between 7% and 25% (mean: 17%). The corresponding range for Hombole is 7–26% (mean: 14%). This contrasts with the average inter-GCM percentage for mean discharge of 91% (73–119%) and 87% (70–111%) for the two gauging stations, respectively. A similar pattern is evident for RCP8.5 albeit with an increase in both sets of ranges. For example, at Melka the mean inter-hydrological model range for mean discharge is 25% (8–51%) compared to 216% (195–259%) for the inter-GCM range. At Hombole, inter-hydrological model range is larger (mean: 79%, 56–140%) than at Melka but still smaller than the inter-GCM range (mean: 221%, 176–281%).

[TABLE 7]

Baseline and projected river regimes at Hombole for each GCM and both RCP scenarios are shown in Fig. 5. Results for Hombole, the downstream station, are shown in light of the overall consensus in the direction of changes projected at the two gauging stations for the same RCP scenario / hydrological model. Changes in the regime at Melka follow those at Hombole (Supplementary Figure 2). There is considerable inter-GCM uncertainty in the seasonal distribution of river flow. Changes in river regimes are more pronounced for those hydrological models that include a spatially-distributed unsaturated zone (models 2-4) with particularly pronounced variability in peak discharge being evident for models 2 and 3. Inter-GCM variability in the regimes simulated by Model 4, which used the fully distributed physically-based Richards equation, is relatively subdued but is noticeably larger than for models 1 and 5 which employed spatially uniform and more conceptual approaches to represent key hydrological processes.

[FIGURE 5]

For RCP4.5, the largest increase in peak discharges across all hydrological models and at both gauging stations is projected by MPI-ESM-MR. On average this GCM projects increases in JAS discharges of 42.5% and 43% at Melka and Hombole, respectively. CSIRO-Mk3 projects the largest decreases (mean JAS declines of 45.1% and 46.5%, respectively). Whilst for RCP8.5 the same GCM (CSIRO-Mk3) projects the largest decreases in JAS discharge (58% and 59% for Melka and Hombole, respectively), the largest increases are projected by CanESM2 (76% and 81%, respectively). All GCMs project increases in river discharge during the dry season (DJF) when low flows (Q90) occur. For RCP4.5, these range between 9% and 10% for Melka and Hombole, respectively for HadGEM2-ES and between 252% and 285% for MPI-ESM-MR. For RCP8.5, mean changes in DJF flows range from between 52% and 66% for HadGEM2 to between 1000% and 1280% for CanESM2 for Melka and Hombole, respectively.

Inter-GCM and inter-hydrological model differences in projected changes in high and low flows are further demonstrated in Fig. 6. This figure shows percentage changes in Q10 and Q90 as simulated by each hydrological model when forced with both RCP scenarios derived from the six GCMs. Results are provided for Hombole and are broadly representative of those for Melka. The relative magnitude of change in low flows (Q90) is much larger than changes in both mean (Fig. 4) and high (Q10) flows. Projections for RCP4.5 from the different GCMs are approximately evenly split with CCSM4, HadGEM2 and CSIRO-Mk3 generally producing declines in both Q10 and Q90 for all of the hydrological models. Increases in these flows dominate results for the remaining three GCMs (CanESM2, IPSL-CM5A-MR and MPI-ESM-MR). The magnitude of these changes tends to increase in both directions for RCP8.5 although, in general, gains in the magnitude of the increases in Q10 and Q95 are larger than those where these flows decline.

Figure 6 shows substantial inter-hydrological model variability in projected high and low flows at Hombole. This variability is particularly pronounced for Q90 compared to Q10. In a number of cases different hydrological models project a different direction of change in low flows for the same GCM. For example, for RCP4.5 Model 5 projects relatively large increases for CCSM4, CSIRO-Mk3, HadGEM-ES and IPSL-CM5A-MR whereas most other hydrological models project declines or only small increases. This pattern is repeated for CCSM4 and CSIRO-Mk3 in the case of RCP8.5. The inter-hydrological model range of changes in Q90 for a given GCM varies between 147% (HadGEM-ES) and 228% (MPI-ESM-MR) under RCP4.5 and 161% (CCSM4) to 419% (CanESM2) under RCP8.5. Inter-hydrological model variability in high (Q10) flows is comparatively smaller and varies between 13% (IPSL-CM5A-MR) and 43% (MPI-ESM-MR) under RCP4.5 and 10% (HadGEM-ES) and 67% (CanESM2) under RCP8.5. There are fewer instances of the direction of change in high flows for a given GCM varying according to hydrological model. Such disagreements are limited to CCSM4 for both RCP scenarios and HadGEM-ES for RCP4.5. In each case, the single hydrological model that projects a different direction of change (Model 5 for CanESM2 and Model 1 for HadGEM-ES) projects only a very small change from the baseline.

[FIGURE 6]

3.4 Uncertainty quantification

Variance decomposition using ANOVA for individual river discharge quantiles is presented in Figure 7. Uncertainty attribution confirms that GCM-related uncertainty is the largest and most dominant source of uncertainty. The mean contribution of GCMs to overall uncertainty over Q10, Q50 and Q90 runoff quantiles is 54% across both gauging stations. The average fraction of uncertainty attributed to the different hydrological models is 7%. For low flows (Q90) the average fraction of uncertainty attributed to the different hydrological models (18%) is considerably higher than the corresponding figures for both median and high flows (Q10) (1% and <0.5%, respectively). For climate change signals of both high flows (Q10) and median flows (Q50), GCM and RCP scenarios are the largest contributors to uncertainty (>70% for both stations). However, for low flows (Q90), the contribution of hydrological model uncertainty to overall uncertainty is considerably higher. The contribution of different hydrological models to overall uncertainty in projections of Q90 is particularly pronounced at Hombole (27%) where it is comparable to GCM-related uncertainty and larger than the combined contribution of interactions between GCM and RCP-related uncertainties.

[FIGURE 7]

4. Discussion

4.1 Model performance and uncertainty sources

Model performance reflects findings from previous model inter-comparison studies. Despite being relatively less complex, models that are conceptual or have uniformly distributed parameter values (i.e. Model 1, 2 and 5 in this study) can exhibit similar, if not better, performance to physically-based or spatially distributed models (i.e. Models 3 to 4) when calibrated against observations from an historical period (Reed *et al.* 2004; Duan *et al.* 2006). It should be noted that the models were calibrated using observations from only two gauging stations and similar model performance across all models is expected. Despite this, model performance at Melka and Hombole for the different hydrological models was comparable to, if not better than, previous hydrological models of the UAB (Table 8).

[TABLE 8]

Inter-hydrological model uncertainty was considerably higher for low flows (Q90). Previous continental-scale modelling studies have similarly found greater variability in projections of low flows that were context- and catchment-specific (Vetter *et al.* 2015; Krysanova *et al.* 2017). It has also been suggested that commonly used hydrological models tend to have relatively poorer predictive ability for low flows given the focus of most models on reproducing a basin's response to precipitation (Staudinger *et al.* 2011; Trudel *et al.* 2017). Results from this study also reinforce the need to consider variables beyond mean annual and monthly discharge that may reveal critical differences between model structures that might otherwise not be identifiable (Gosling *et al.* 2011).

A number of global multi-model studies have similarly suggested higher hydrological model-related uncertainty in projections of low flows under climate change. Inter-hydrological model uncertainty in Q90 from the current study was also comparable to the absolute percentage differences found between hydrological models (>30%) by Vansteenkiste *et al.* (2014a,b). Comparing multiple distributed and semi-distributed models (including MIKE SHE) with different conceptualizations of groundwater-surface water interactions for a Belgian catchment, it was concluded that projections exhibited common impact trends for high/mean flows among the models but that results were highly variable for low flows. This variability in low flows occurred without any specific conceptualization of groundwater flow yielding superior model performance during calibration. Using four hydrological models of varying complexity, including both lumped and distributed approaches, in a climate change impact assessment for the Tualatin River Basin (Oregon, USA), Najafi *et al.* (2011) similarly concluded that choice of model exerts considerable uncertainty in discharge projections during the dry season.

Uncertainty attribution showed that the mean contribution of GCMs to overall uncertainty for the UAB was comparable to the GCM fraction of total uncertainty (57%) averaged over 12 global basins by Krysanova *et al.* (2017). The average fraction of uncertainty attributed to the different hydrological models for Q90 is comparable to the fraction calculated for four out of the 12 basins (Niger, Darling, Upper Amazon and Blue Nile) in this earlier study. In contrast, the hydrological model-related fraction of uncertainty for Q10 is substantially lower than projected for all 12 basins by Krysanova *et al.* (2017). Our results reinforce that the contributions of individual uncertainty sources vary in space and that this might warrant different modelling philosophies to reduce the relative dominance of different sources of uncertainties (Hattermann et al. 2018).

4.2 Comparison between hydrological models

Uncertainties associated with hydrological model structures can be considerable depending on the calibration strategy and the selected hydrological variables. The addition of spatially distributed soil classes from Model 1 to Model 2 while the other process representations of unsaturated and saturated flow remained the same was assumed to be advantageous. However, additional parameterization as a result of specifying individual calibration parameters for each soil type may contribute to over-parameterization and an improvement in model performance may not result across different climatic and environmental conditions (Jakeman and Hornberger 1993). Comparing MIKE SHE models with spatially uniform and distributed parameterisations of the Tern catchment (UK), Rochester (2010) found better performance from distributed models at locations underlain by impermeable geology. The domination of clay within the UAB and the representation of the saturated zone as a single layer of basaltic volcanic strata that is impermeable in certain regions (Yitbarek *et al.* 2012), may contribute to better performance of Model 2. However, although also including spatially distributed soils, the performance of Models 3-5 was only comparable and in some cases, marginally worse than models 1 and 2. This may relate to the additional parameters associated with the variation of process descriptions of unsaturated and saturated flow and the limitation of the manual calibration strategy employed in this study. Model equifinality exhibited for mean flows reflect findings of previous model intercomparison studies but considerable uncertainties exists for the representation of low flows. Even though models 1-4 all employed a finite difference method for saturated zone flow, their performance for low flows varied substantially. This suggests that parameter and structural uncertainty associated with the representation of the unsaturated zone and so groundwater recharge is particularly important for the simulation of low flows.

Given that the models used in this study span a range of commonly-used model structures, variations in process descriptions within MIKE SHE show that the evaluation of singular model components within an individual model code should be as fundamental as comparisons between hydrological model codes. Considerably different parameter values used in the alternative models demonstrate the importance of considering model equifinality. Variable model performance due to the use of different process representations of the unsaturated zone highlight the fact that considerable uncertainties still remain in the representation of the subsurface and the contributions from groundwater to low flows among hydrological models. The addition of conceptual effective parameters to represent macropore flows and spatial heterogeneity within physically-based process equations is further indicative of the need to better characterize and constrain epistemic uncertainties (Beven and Germann 2013). Comparing the simulation of root-zone dynamics using reservoir schemes (e.g. 2-layer water balance) and the Richards equation, Baroni *et al.* (2010) found comparable performance especially when no site-specific calibration was conducted. Although the Darcy-Richards equation remains the dominant description of the unsaturated zone in physically-based distributed models, non-linearity in spatially heterogeneous soils, inaccuracy at the catchment scale and the exclusion of preferential flow are concerns (Gupta *et al.* 2012; Beven and Germann 2013). The inclusion of preferential flow in the model employing the gravity flow unsaturated zone approach (Model 3) did not have a large influence on overall model performance.

4.3 Climate change implications

Projections from the six different GCMs demonstrate the dominance of projected increases in precipitation over the UAB. This reflects results of past multi-model studies and ensemble projection that have consistently projected increased precipitation over East Africa (Niang *et al.* 2014). Examining historical global precipitation data, Knoben *et al.* (2019) detected a gradual transition from bimodal to unimodal precipitation regimes latitudinally across Africa with a bimodal regime over Ethiopia. A number of GCMs used in this study project the development of a unimodal regime over the UAB under both RCP4.5 and RCP8.5. It is therefore plausible that under climate change, the latitudinal gradient in precipitation modality across Africa may shift. Projections of increased precipitation are in apparent contrast to significant drought events experienced across East Africa in recent decades, sometimes referred to as the 'East African Climate Paradox' (Souverijns *et al.* 2016; Nicholson 2017). Possible reasons for the discrepancy include the impacts of anthropogenic aerosol emissions, uncertainty in GCM representation of key processes, changes in seasonality of the rainy season and natural variability (Rowell *et al.* 2015; Wainwright *et al.* 2019).

Consistency in outcomes of the impact of climate projections from different GCMs for increases in discharge and high flows (Q10) is greater for RCP8.5 than for RCP4.5, indicating a degree of confidence over the projected direction of change under scenarios of higher greenhouse-gas concentrations (i.e. RCP8.5). Projected increases in mean flows, Q10 and flood frequency over East Africa and other monsoonal regions are well documented in global-scale modelling studies (Arnell and Gosling 2013; Koirala *et al.*

2014). Greater model agreement in projections of increased discharge under RCP8.5 in this study is also consistent with an evaluation of CMIP5 model agreement in global streamflow change (Koirala *et al.* 2014). Given that this study is only based on a subset of CMIP5 GCMs, an extension of this study would be to conduct a complete assessment using all 41 CMIP5 GCMs or genealogical-based model groups (e.g. Ho *et al.* 2016; Thompson *et al.* 2017; Hudson and Thompson 2019).

Effective decision-making in the UAB may be hampered by the large range of uncertainties revealed in this study. Projected increases in precipitation and river flow could be expected to benefit agricultural production. However, reductions in precipitation and river flow are equally plausible. Assessing the economic impacts of hydrological changes in the Awash River Basin, Borgomeo *et al.* (2018) found that a 5% reduction in precipitation or a spatial redistribution of rainfall under climate change could incur up to a 10% drop in GDP of the agricultural sector. The high sensitivity of low flows to hydrological model structural uncertainty relative to mean and high flows will have significant implications for both drought mitigation (e.g. Duan *et al.* 2014) and environmental flows (e.g. Thompson *et al.* 2014a). Diverging scenarios projected by different GCMs and hydrological models may be plausible but could easily be omitted in ensemble analyses. Recent studies have suggested employing a 'storylines' approach to navigate uncertainties incurred along the modelling chain (Clark *et al.* 2016; Shepherd 2019). Increasingly popular in climate science, storylines are suites of equally plausible, quantitative narratives that are catered towards providing regional climate change information to better enable decision-making (Shepherd 2019). Applying a storylines approach could better present information of plausible hydrological changes directed at operational decision making and stress-testing water resources systems to improve climate resilience.

5. Conclusions

Climate change impacts on river discharge in the Upper Awash Basin (UAB) of Ethiopia, assessed using an ensemble of five MIKE SHE hydrological models, six CMIP5 GCMs, and two greenhouse-gas concentration trajectories (RCP4.5, RCP8.5) reveal substantial GCM-related uncertainty in projected river discharge that determines both the direction and magnitude of change from baseline to the end of this century, 2071-2100 (RCP4.5: -34% to +55%; RCP8.5: -42% to +195%). Our application of an ensemble of five MIKE SHE hydrological models found, consistent with previous model inter-comparison studies, that models with spatially uniform parameter values exhibit similar performance to physically-based models with spatially distributed parameterizations. Model performance generally exceeded that of previous hydrological models of the UAB and demonstrated a bias to its representation of peak flows (Q10) compared to low flows (Q90). Uncertainty attribution using ANOVA shows that GCM-related uncertainty represents, on average, 68% of the total uncertainty for mean and high flows whereas hydrological model uncertainty constitutes an average 18% of total uncertainty in the low-flow projections. At the downstream gauging station in the UAB (Hombole), the contribution of uncertainty in hydrological model structure (27%) to total uncertainty was comparable to that of GCM-related uncertainty for low (Q90) flows. Of note is that uncertainties arising from different hydrological model structures are masked if only projections of mean annual discharge are considered.

Substantial uncertainties in the representation of low flows attributed to hydrological model structure have significant implications for the prediction and management of drought risks in semi-arid catchments such as the UAB. The lack of integrated monitoring infrastructure observing precipitation, surface waters and groundwater levels currently impairs the development of robust conceptual models of basin hydrology including critically seasonal interactions between groundwater and streamflow, and the observed contribution of groundwater to baseflow. On the modelling side, possible extensions to this study include consideration of additional sources of uncertainties along the impact modelling chain such as PET-related uncertainty stemming from the choice of algorithm used in its calculation (e.g. Kingston *et al.* 2009; Thompson *et al.* 2014b). Alternative approaches to bias correcting climate projections such as quantile mapping (see Rahman *et al.* 2020) could also be explored. Characterizing the propagation of impact model uncertainty in the hydrological projections in terms of environmental flows (e.g. Thompson *et al.* 2014a) could also be a next step to better understand potential hydro-ecological impacts of climate change on the Upper Awash Basin.

Acknowledgements

This study was supported by a consortium grant, *GroFutures* (grant refs: NE/M008932/1, NE/M008584/1, NE/M008207/1) funded by the UK Natural Environment Research Council (NERC) and Economic and Social Research Council (ESRC) and the UK Department for International Development (DfID): Unlocking the Potential of Groundwater for Poverty Alleviation (UpGro) consortium project. The authors would like to thank the associate editor and reviewers for their constructive comments and suggestions, which improved the clarity of presented arguments in the paper.

References

Anandhi, A., Frei, A., Pierson, D.C., Schneiderman, E.M., Zion, M.S., Lounsbury, D. and Matonse, A.H. 2011. Examination of change factor methodologies for climate change impact assessment. *Water Resources Research*, *47*(3), W03501.

Andersen, J., Refsgaard, J.C. and Jensen, K.H., 2001. Distributed hydrological modelling of the Senegal River Basin—model construction and validation. *Journal of Hydrology*, *247*(3-4), 200-214.

Allen, R.G., Luis, S., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration—Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56. FAO, Rome, Italy.

Arnell, N.W., 2003. Effects of IPCC SRES* emissions scenarios on river runoff: a global perspective. *Hydrology and Earth System Sciences Discussions*, *7*(5), 619-641.

Arnell, N.W. and Gosling, S.N., 2013. The impacts of climate change on river flow regimes at the global scale. *Journal of Hydrology*, *486*, 351-364.

Baroni, G., Facchi, A., Gandolfi, C., Ortuani, B., Horeschi, D. and Van Dam, J.C., 2010. Uncertainty in the determination of soil hydraulic parameters and its influence on the performance of two hydrological models of different complexity. *Hydrology and Earth System Sciences*, *14*(2), 251-270.

Berhe, F.T., Melesse, A.M., Hailu, D. and Sileshi, Y., 2013. MODSIM-based water allocation modeling of Awash River Basin, Ethiopia. *Catena*, *109*, 118-128.

Betts, R.A., Alfieri, L., Bradshaw, C., Caesar, J., Feyen, L., Friedlingstein, P., Gohar, L., Koutroulis, A., Lewis, K., Morfopoulos, C. and Papadimitriou, L., 2018. Changes in climate extremes, fresh water availability and vulnerability to food insecurity projected at 1.5° C and 2° C global warming with a higher-resolution global climate model. *Phil. Trans. R. Soc. A*, *376*(2119), 20160452.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in water resources*, *16*(1), 41-51.

Beven, K., 2016. Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal*, *61*(9), 1652-1665.

Beven, K. and Germann, P., 2013. Macropores and water flow in soils revisited. *Water Resources Research*, *49*(6), 3071-3092.

Borgomeo, E., Vadheim, B., Woldeyes, F.B., Alamirew, T., Tamru, S., Charles, K.J., Kebede, S. and Walker, O., 2018. The distributional and multi-sectoral impacts of rainfall shocks: Evidence from computable general equilibrium modelling for the Awash Basin, Ethiopia. *Ecological economics*, *146*, 621-632.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M. and Schär, C., 2013. Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. *Water Resources Research*, *49*(3), 1523-1536.

Butts, M.B., Payne, J.T., Kristensen, M. and Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of hydrology*, *298*(1-4), 242-266.

Chen, C., Cane, M.A., Wittenberg, A.T. and Chen, D., 2017. ENSO in the CMIP5 simulations: Life cycles, diversity, and responses to climate change. *Journal of Climate*, *30*(2), 775-801.

Clark, M.P., Wilby, R.L., Gutmann, E.D., Vano, J.A., Gangopadhyay, S., Wood, A.W., Fowler, H.J., Prudhomme, C., Arnold, J.R. and Brekke, L.D., 2016. Characterizing uncertainty of the hydrologic impacts of climate change. *Current Climate Change Reports*, *2*(2), 55-64.

Dessu, S.B., Seid, A.H., Abiy, A.Z. and Melesse, A.M., 2016. Flood forecasting and stream flow simulation of the upper Awash river basin, Ethiopia using geospatial stream flow model (GeoSFM). In *Landscape Dynamics, Soils and Hydrological Processes in Varied Climates* (pp. 367-384). Springer, Cham.

Duan, K. and Mei, Y., 2014. Comparison of meteorological, hydrological and agricultural drought responses to climate change and uncertainty assessment. *Water resources management*, *28*(14), 5039-5054.

Duan, Q., Schaake, J., Andreassian, V., Franks, S., Goteti, G., Gupta, H.V., Gusev, Y.M., Habets, F., Hall, A., Hay, L. and Hogue, T., 2006. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, *320*(1-2), 3-17.

FAO (1998): Soil map of the world: Revised legend. World Soil Resources Report 60, Food and Agriculture Organization of the United Nations, Rome.

Fowler, H.J., Blenkinsop, S. and Tebaldi, C. 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, *27*(12), 1547-1578.

Gizaw, M.S., Biftu, G.F., Gan, T.Y., Moges, S.A. and Koivusalo, H., 2017. Potential impact of climate change on streamflow of major Ethiopian rivers. *Climatic Change*, *143*(3-4), 371-383.

Gosling, S.N., Taylor, R.G., Arnell, N. and Todd, M.C., 2011. A comparative analysis of projected impacts of climate change on river runoff from global and catchment-scale hydrological models. *Hydrology and Earth System Sciences*, *15*(1), 279-294.

Graham, D.N. and Butts, M.B., 2005. Flexible, integrated watershed modelling with MIKE SHE. *Watershed models*, *849336090*, 245-272.

Gupta, H.V., Clark, M.P., Vrugt, J.A., Abramowitz, G. and Ye, M., 2012. Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*, *48*(8). W08301.

Haddeland, I., Clark, D.B., Franssen, W., Ludwig, F., Voß, F., Arnell, N.W., Bertrand, N., Best, M., Folwell, S., Gerten, D. and Gomes, S., 2011. Multimodel estimate of the global terrestrial water balance: setup and first results. *Journal of Hydrometeorology*, *12*(5), 869-884.

Hailemariam, K., 1999. Impact of climate change on the water resources of Awash River Basin, Ethiopia. *Climate Research*, *12*(2-3), 91-96.

Hargreaves, G.H. and Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. *Applied engineering in agriculture*, *1*(2), 96-99.

Hattermann, F.F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., Fekete, B., Flörke, F., Gosling, S.N., Hoffmann, P. and Liersch, S., 2018. Sources of uncertainty in hydrological climate impact assessment: a cross-scale study. *Environmental Research Letters*, *13*(1), 015006.

Henriksen, H.J., Troldborg, L., Nyegaard, P., Sonnenborg, T.O., Refsgaard, J.C. and Madsen, B., 2003. Methodology for construction, calibration and validation of a national hydrological model for Denmark. *Journal of Hydrology*, 280(1-4), 52-71.

Ho, J.T., Thompson, J.R., Brierley, C.B., 2016. Projections of hydrology in the Tocantins-Araguaia Basin, Brazil: uncertainty assessment using the CMIP5 ensemble. *Hydrological Sciences Journal*, 61(3), 551-567.

Hudson, C.E. and Thompson, J.R., 2019. Hydrological modelling of climate change impacts on river flows in Siberia's Lena River Basin and implications for the Atlantic Meridional Overturning Circulation. *Hydrology Research*. Corrected Proof Online.

IPCC, 2014. *Climate Change 2014: Synthesis Report*. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland.

Jakeman, A.J. and Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model?. *Water Resources Research*, *29*(8), 2637-2649

Jiménez Cisneros, B.E., T. Oki, N.W. Arnell, G. Benito, J.G. Cogley, P. Döll, T. Jiang, and S.S. Mwakalila, 2014: Freshwater resources. In: Climate Change 2014: Impacts,Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L.White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 229-269.

Jira, M.N. 2019. *Numerical groundwater flow modelling for planning and management of the resource in the Bacho Plain, Upper Awash Basin, Central Ethiopia*, Thesis (MSc), Addis Ababa University, Addis Ababa.

Karlsson, I.B., Sonnenborg, T.O., Refsgaard, J.C., Trolle, D., Børgesen, C.D., Olesen, J.E., Jeppesen, E. and Jensen, K.H., 2016. Combined effects of climate models, hydrological model structures and land use scenarios on hydrological impacts of climate change. *Journal of Hydrology*, *535*, 301-317.

Kebede, S., 2013. *Groundwater in Ethiopia: features, numbers and opportunities*. Springer, Heidelberg.

Kingston, D.G., Todd, M.C., Taylor, R.G., Thompson, J.R. and Arnell, N.W., 2009. Uncertainty in the estimation of potential evapotranspiration under climate change. *Geophysical Research Letters* 36, L20403.

Klemeš, V. 1986. Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, *31*(1), 13-24.

Knoben, W.J., Woods, R.A. and Freer, J.E., 2019. Global bimodal precipitation seasonality: A systematic overview. *International Journal of Climatology*, 39(1), 558-567.

Kociuba, G. and Power, S.B., 2015. Inability of CMIP5 models to simulate recent strengthening of the Walker circulation: Implications for projections. *Journal of Climate*, *28*(1), 20-35.

Koirala, S., Hirabayashi, Y., Mahendran, R. and Kanae, S., 2014. Global assessment of agreement among streamflow projections using CMIP5 model outputs. *Environmental Research Letters*, *9*(6), 064017.

Krause, P., Boyle, D.P. and Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *5*, 89-97.

Krysanova, V., Vetter, T., Eisner, S., Huang, S., Pechlivanidis, I., Strauch, M., Gelfan, A., Kumar, R., Aich, V., Arheimer, B. and Chamorro, A., 2017. Intercomparison of regional-scale hydrological models and climate change impacts projected for 12 large river basins worldwide—a synthesis. *Environmental Research Letters*, *12*(10), 105002.

Mekonnen, M.A., Wörman, A., Dargahi, B. and Gebeyehu, A., 2009. Hydrological modelling of Ethiopian catchments using limited data. *Hydrological Processes: An International Journal*, *23*(23), 3401-3408.

Müller, R., Gebretsadik, H.Y. and Schütze, N., 2016. Towards an optimal integrated reservoir system management for the Awash River Basin, Ethiopia. *Proceedings of the International Association of Hydrological Sciences*, 373, 215-219.

Najafi, M.R., Moradkhani, H. and Jung, I.W., 2011. Assessing the uncertainties of hydrologic model selection in climate change impact studies. *Hydrological processes*, *25*(18), 2814-2826.

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, 10(3), 282-290.

Niang, I., O.C. Ruppel, M.A. Abdrabo, A. Essel, C. Lennard, J. Padgham, and P. Urquhart, 2014. Africa. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L.White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1199-1265.

Nicholson, S.E., 2017. Climate and climatic variability of rainfall over eastern Africa. *Reviews of Geophysics*, 55(3), 590-635.

Philip, S., Kew, S.F., Jan van Oldenborgh, G., Otto, F., O'Keefe, S., Haustein, K., King, A., Zegeye, A., Eshetu, Z., Hailemariam, K. and Singh, R., 2018. Attribution analysis of the Ethiopian drought of 2015. *Journal of Climate*, 31(6), 2465-2486.

Poulin, A., Brissette, F., Leconte, R., Arsenault, R. and Malo, J.S., 2011. Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. *Journal of Hydrology*, 409(3-4), 626-636.

Pushpalatha, R., Perrin, C., Le Moine, N. and Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology*, 420, 171-182.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.J. and Participants, D.M.I.P., 2004. Overall distributed model intercomparison project results. *Journal of Hydrology*, *298*(1-4), 27-60.

Refsgaard, J.C., 1997. Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology*, 198(1-4), 69-97.

Refsgaard, J.C., Storm, B. and Clausen, T. 2010. Système Hydrologique Europeén (SHE): review and perspectives after 30 years development in distributed physically-based hydrological modelling. *Hydrology Research* 41, 355–377.

Robinson, A. J. 2018.Uncertainty in Hydrological Scenario Modelling: An Investigation Using the Mekong River Basin, SE Asia. Thesis (PhD), Department of Geography, University College London, London.

Rochester, R.E. 2010. *Uncertainty in hydrological modelling: A case study in the Tern Catchment, Shropshire, UK,* Thesis (PhD), Department of Geography, University College London, London.

Rowell, D.P., Booth, B.B., Nicholson, S.E. and Good, P., 2015. Reconciling past and future rainfall trends over East Africa. *Journal of Climate*, *28*(24), 9768-9788.

Seleshi, Y. and Zanke, U., 2004. Recent changes in rainfall and rainy days in Ethiopia. *International Journal of Climatology*, *24*(8), 973-983.

Shepherd, T.G., 2019. Storyline approach to the construction of regional climate change information. *Proceedings of the Royal Society A*, *475*(2225), p.20190013.

Souverijns, N., Thiery, W., Demuzere, M. and Van Lipzig, N.P., 2016. Drivers of future changes in East African precipitation. *Environmental Research Letters*, 11(11), 114011.

Staudinger, M., Stahl, K., Seibert, J., Clark, M.P. and Tallaksen, L.M., 2011. Comparison of hydrological model structures based on recession and low flow simulations. *Hydrology and Earth System Sciences*, *15*(11), 3447-3459.

Taddese, G., Sonder, K. and Peden, D., 2003. The water of the Awash River basin a future challenge to Ethiopia. *International Livestock Research Institute, Addis Ababa*.

Tadono, T., Ishida, H., Oda, F., Naito, S., Minakawa, K. and Iwamoto, H., 2014. Precise global DEM generation by ALOS PRISM. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *2*(4), 71-76.

Thompson, J.R., Crawley, A. Kingston, D.G. 2017. Future river flows and flood extent in the Upper Niger and Inner Niger Delta: GCM-related uncertainty using the CMIP5 ensemble. *Hydrological Sciences Journal*, 62(14), 2239-2265.

Thompson, J.R., Green, A.J. and Kingston, D.G., 2014b. Potential evapotranspiration-related uncertainty in climate change impacts on river flow: An assessment for the Mekong River basin. *Journal of Hydrology*, 510, 259-279.

Thompson, J.R., Green, A.J., Kingston, D.G. and Gosling, S.N. 2013. Assessment of uncertainty in river flow projections for the Mekong River using multiple GCMs and hydrological models. *Journal of Hydrology*, 486, 1-30.

Thompson, J.R., Laizé, C.L.R., Green, A.J., Acreman, M.C., Kingston, D.G. 2014a. Climate change uncertainty in environmental flows for the Mekong River. *Hydrological Sciences Journal* 59, 935-954.

Thompson, J.R., Sørenson, H.R., Gavin, H. and Refsgaard, A., 2004. Application of the coupled MIKE SHE/MIKE 11 modelling system to a lowland wet grassland in southeast England. *Journal of Hydrology*, 293(1-4), 151-179.

Todd, M., Taylor, R., Osborn, T., Kingston, D., Arnell, N. and Gosling, S. 2011. ‘Uncertainty in climate change impacts on basin-scale freshwater resources – preface to the special issue: the QUEST-GSI methodology and synthesis of results’. *Hydrology and Earth System Sciences*, 15(3), 1035-1046.

Tolera, M.B., Chung, I.M. and Chang, S.W., 2018. Evaluation of the Climate Forecast System Reanalysis Weather Data for Watershed Modeling in Upper Awash Basin, Ethiopia. *Water (20734441)*, 10(6).

Trudel, M., Doucet-Généreux, P.L. and Leconte, R., 2017. Assessing River Low-Flow Uncertainties Related to Hydrological Model Calibration and Structure under Climate Change Conditions. *Climate*, *5*(1), 19-43.

Vansteenkiste, T., Tavakoli, M., Ntegeka, V., De Smedt, F., Batelaan, O., Pereira, F. and Willems, P., 2014a. Intercomparison of hydrological model structures and calibration approaches in climate scenario impact projections. *Journal of Hydrology*, 519, 743-755.

Vansteenkiste, T., Tavakoli, M., Van Steenbergen, N., De Smedt, F., Batelaan, O., Pereira, F. and Willems, P., 2014b. Intercomparison of five lumped and distributed models for catchment runoff and extreme flow simulation. *Journal of Hydrology*, 511, 335-349.

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V. and Hattermann, F., 2015. Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents. *Earth System Dynamics*, 6(1), 17-43.

Wainwright, C.M., Marsham, J.H., Keane, R.J., Rowell, D.P., Finney, D.L., Black, E. and Allan, R.P., 2019. ‘Eastern African Paradox’rainfall decline due to shorter not less intense Long Rains. *npj Climate and Atmospheric Science*, *2*(1), 1-9.

Wilby, R.L. and Dessai, S., 2010. Robust adaptation to climate change. *Weather*, 65(7), 180-185.

Yitbarek, A., Razack, M., Ayenew, T., Zemedagegnehu, E. and Azagegn, T., 2012. Hydrogeological and hydrochemical framework of Upper Awash River basin, Ethiopia: with special emphasis on inter-basins groundwater transfer between Blue Nile and Awash Rivers. *Journal of African Earth Sciences*, 65, 46-60.

Figure titles

Figure 1. The River Awash Basin (a) location within East Africa; (b) delineation of the Upper Awash Basin within the Awash Basin; and (c) the Upper Awash Basin including elevation, main drainage network and hydro-meteorological monitoring infrastructure from which data are used in the current study. (Note: this figure is purely to illustrate the location of the UAB within Ethiopia and is not representative of an official endorsement of disputed country borders)

Figure 2. Baseline and simulated river discharge at the Melka and Hombole gauging stations: (a,b) monthly mean discharge (through the simulation period with calibration and validation periods indicated; (c-f) river regimes for calibration and validation periods; note: different y-axis ranges.

Figure 3. a-d) Percentage change in precipitation and PET from the baseline (1979-1999) for each GCM and RCP scenario (2071-2100); e-h) Baseline and projected mean monthly catchment-averaged precipitation and PET for each GCM and both RCP scenarios.

Figure 4. Percentage changes in mean discharge at Melka and Hombole relative to the baseline for each hydrological model and GCM under RCP4.5 (top row) and RCP 8.5 (bottom row); note: different y-axis ranges for the two RCP scenarios.

Figure 5. Baseline and simulated river regimes at Hombole for each GCM and hydrological model under RCP4.5 and RCP8.5.

Figure 6. Percentage changes in Q10 and Q90 at Hombole for each hydrological model and GCM under the RCP4.5 and RCP8.5 scenarios.

Figure 7. Contribution of each source of uncertainty and interactions between them to overall uncertainty in projections of Q90, mean and Q10 discharges at Melka and Hombole.

Supplementary Figure 1 Boxplots of delta factors for annual mean precipitation and annual mean temperature over the Upper Awash catchment for all CMIP5 GCMs and the six selected GCMs.

Supplementary Figure 2 Baseline and simulated river regimes at Melka for each GCM and hydrological model under RCP4.5 and RCP8.5.

Supplementary Figure 3. Percentage change in mean discharge at Melka and Hombole simulated by each hydrological model from the combined and individual application of scenario precipitation and PET for the six GCMs and RCP 4.5 scenario.

Supplementary Figure 4. Percentage change in mean discharge at Melka and Hombole simulated by each hydrological model from the combined and individual application of scenario precipitation and PET for the six GCMs and RCP 8.5 scenario.

Table 1. Summary of inputs and data sources for key components of the coupled MIKE SHE/MIKE11 model of the Upper Awash Basin.

Model component	Input	Data source/derivation
MIKE SHE		
Model Domain	Catchment extent	ESRI polygon shapefile established from ALOS digital elevation model
Topography	Digital Elevation Model (DEM)	30m × 30m resolution ALOS Digital Elevation Model resampled to 1km MIKE SHE grids (Tadono *et al.* 2014)
Land use/vegetation	Land use	Seven land use classes specified using a 2015 land cover map (rainfed agriculture, Irrigated agriculture, grassland, bushland, forest, wetland, open water and urban areas)
	LAI/Root depth	Literature (Allen *et al.* 1998)
Overland flow	Manning's M; Spatial distribution	- Uniform - Derivation: 2D finite difference method
Catchment meteorology	Precipitation	Observed daily station data from 11 meteorological stations distributed by Thiessen's polygons. A grid file defined 11 meteorological sub-catchments based on the areas of each Thiessen's polygons to account for the spatial distribution in precipitation across the entire catchment
	Evapotranspiration (PET)	- Time-varying PET derived by calculating lapse rate for four elevation ranges (1750m, 2250m, 2750m and 3250m) from daily minimum and maximum temperature at 2354m - PET derived using the Hargreaves method (Hargreaves and Samani 1985)
Unsaturated Zone	Soil classes	- Spatial distribution vary among HMs according to the FAO soil map of the world (FAO 1996)
	Soil hydraulic properties	Literature and USDA soil classes hydraulic properties
Saturated Zone	Spatial distribution	- Uniform - Derivation: Varies among HMs between finite difference or linear reservoirs
MIKE 11		
	River network	- River delineation from aerial photography (Google Earth Pro) and stream order from ALOS DEM
	Cross-sections of stream network	- Synthetic cross sections (Representative max. cross section depths and profiles estimated from stream orders derived from DEM and Google Earth Pro)
	Hydrodynamic parameters: High order; Fully dynamic	- Uniform Manning's n of 0.04 (Chow 1959) throughout river network

Table 2. Alternative specification of the unsaturated and saturated zones within each MIKE SHE model.

	Model name[1]				
	Model 1 (UCP)	**Model 2 (DCP)**	**Model 3 (DPP-G)**	**Model 4 (DPP-R)**	**Model 5 (DCC)**
Unsaturated Zone Spatial Distribution					
Uniform (U)	✓				
Distributed (D)		✓	✓	✓	✓
Unsaturated Flow Process Representation					
2-layer Water Balance (C)	✓	✓			✓
Gravity Flow (P-G)			✓		
Richards Equation (P-R)				✓	
Saturated Flow Process Representation					
Finite Difference (P)	✓	✓	✓	✓	
Linear Reservoir (C)					✓

[1]Model name refers for whether or not the hydrological model is uniformly distributed (U), spatially distributed (D) conceptual (C) or physically-based (P) in their representation of the spatial distribution of the unsaturated zone, and the process representation of the unsaturated zone and saturated zone respectively.

Table 3. Selected GCMs and their respective host institutions.

Modelling institute	**GCM**
Canadian Centre for Climate Modelling and Analysis	CanESM2 (CA)
National Centre for Atmospheric Research	CCSM4 (CC)
Commonwealth Scientific and Industrial Research (AU)	CSIRO-Mk3.6.0 (CS)
Met Office Hadley Centre	HadGEM2-ES (HA)
Institut Pierre-Simon Laplace	IPSL-CM5A-MR (IP)
Max Planck Institute for Meteorology	MPI-ESM-MR (MP)

Table 4. Final values of calibration parameters for each MIKE SHE model.

Calibration parameter	**Model 1 (UCP)**	**Model 2 (DCP)**	**Model 3 (DPP-G)**	**Model 4 (DPP-R)**	**Model 5 (DCC)**
UZ Saturated hydraulic conductivity (ms^{-1})	9.8e-009	9.8e-008	NA[1]	NA[1]	NA[1]
Horizontal hydraulic conductivity (ms^{-1})	2.1e-008	3.5e-008	3.5e-007	5e-007	NA[1]
Vertical hydraulic conductivity (ms^{-1})	2.4e-009	3e-009	6e-008	8e-007	NA[1]
Bypass fraction	0.25	0.15	0.2	0.25	0.18

[1]NA denotes that the process representation included within a specific model does not include this parameter.

Table 5. Daily and monthly model performance statistics at Melka and Hombole for each hydrological model for the calibration (cal.) and validation (val.) periods.

Station		Dv (%)	Daily NSE	Monthly NSE	Monthly *r*
Model 1 (UCP)					
Melka	cal.	-0.98 ★★★★★	0.45 ★★	0.78 ★★★★	0.90
	val.	26.9 ★★	0.45 ★★	0.82 ★★★★	0.93
Hombole	cal.	11.8 ★★★	0.45 ★★	0.82 ★★★★	0.91
	val.	14.6 ★★★	0.52 ★★★	0.83 ★★★★	0.93
Model 2 (DCP)					
Melka	cal.	-5.72 ★★★★	0.53 ★★★	0.85 ★★★★★	0.92
	val.	41.4 ★	0.28 ★★	0.53 ★★★	0.88
Hombole	cal.	3.3 ★★★★★	0.61 ★★★★	0.88 ★★★★★	0.94
	val.	28.4 ★★	0.55 ★★★	0.78 ★★★★	0.90
Model 3 (DPP-G)					
Melka	cal.	-29.2 ★★	0.42 ★★	0.79 ★★★★	0.79
	val.	58.8 ★	0.19 ★	0.33 ★★	0.91
Hombole	cal.	-10.5 ★★★	0.33 ★★	0.73 ★★★★	0.87
	val.	32.2 ★★	0.32 ★★	0.60 ★★★	0.93
Model 4 (DPP-R)					
Melka	cal.	9.79 ★★★★	0.38 ★★	0.79 ★★★★	0.91
	val.	68.2 ★	0.13 ★	0.59 ★★★	0.87
Hombole	cal.	36.1 ★★	0.54 ★★★	0.77 ★★★★	0.91
	val.	55.7 ★	0.21 ★★	0.67 ★★★★	0.89
Model 5 (DCC)					
Melka	cal.	-6.18 ★★★★	0.69 ★★★★	0.80 ★★★★	0.92
	val.	40.8 ★★	0.64 ★★★	0.65 ★★★★	0.90
Hombole	cal.	38.7 ★★	0.65 ★★★★	0.66 ★★★★	0.87
	val.	40.5 ★★	0.62 ★★★	0.55 ★★★	0.89

Performance indicator[1]	Excellent ★★★★★	Very Good ★★★★	Fair ★★★	Poor ★★	Very Poor ★
Dv (%)	<5	5-10	10-20	20-40	>40
NSE	>0.85	0.65-0.85	0.50-0.65	0.20-0.50	<0.20

[1]Model performance criteria based on Henriksen *et al.* (2003)

Table 6. Daily and monthly logNSE values and mean DJF river discharge for each hydrological model for the calibration (cal.) and validation (val.) periods

Station		Daily logNSE	Monthly logNSE	Sim. DJF flow (m^3s^{-1})	Obs. DJF flow (m^3s^{-1})
Model 1 (UCP)					
Melka	cal.	0.49	0.58	3.56	1.46
	val.	0.33	0.41	9.55	1.61
Hombole	cal.	0.43	0.55	5.83	4.23
	val.	0.39	0.49	13.28	4.33
Model 2 (DCP)					
Melka	cal.	0.33	0.47	2.48	
	val.	0.44	0.54	6.25	
Hombole	cal.	0.27	0.34	4.27	
	val.	0.39	0.51	8.91	
Model 3 (DPP-G)					
Melka	cal.	0.22	0.23	0.92	
	val.	0.54	0.63	1.40	
Hombole	cal.	0.25	0.23	2.01	
	val.	0.58	0.68	2.92	
Model 4 (DPP-R)					
Melka	cal.	0.23	0.37	6.92	
	val.	0.12	-0.21	13.76	
Hombole	cal.	0.28	0.33	13.11	
	val.	-0.04	0.11	23.14	
Model 5 (DCC)					
Melka	cal.	0.12	0.22	4.65	
	val.	-0.03	-0.21	7.20	
Hombole	cal.	0.22	0.31	7.64	
	val.	0.36	0.34	10.49	

Table 7. Inter-GCM uncertainty range (difference in maximum and minimum % change in mean discharge) between hydrological models and inter-hydrological model uncertainty range between GCMs.

		RCP 4.5		RCP8.5	
	Model	Melka	Hombole	Melka	Hombole
Inter-GCM uncertainty range	Model 1 (UCP)	75	70	207	281
	Model 2 (DCP)	102	99	237	231
	Model 3 (DPP-G)	119	111	259	239
	Model 4 (DPP-R)	85	78	194	176
	Model 5 (DCC)	73	76	184	179
Inter-HM uncertainty range within single GCM	CanESM2	17	13	51	140
	CCSM4	7	7	10	62
	CSIRO-MK3	24	19	24	56
	HadGEM2-ES	14	10	8	77
	IPSL-CM5A-MR	16	11	34	74
	MPI-ESM-MR	25	26	22	63

Table 8. Model performance of previous hydrological models of the UAB; cal. and val. refer to calibration and validation periods, respectively.

Model	**No. of stations**	**Time-step**	**Period**	**NSE**	***r***	**Reference**
WatBal	3	Monthly	cal.	NA	0.88-0.96	Hailemariam (1999)
			val.	NA	0.72-0.93	
MODSIM	2	Monthly	combined	NA	0.59-0.76	Berhe *et al.* (2013)
GeoSFM	3	Daily	cal.	0.60-0.63	0.69-0.71	Dessu *et al.* (2016)
			val.	0.60-0.61	0.64-0.66	
HSPF	4	Daily	cal.	0.53-0.88	0.60-0.90	Gizaw *et al.* (2017)
			val.	0.53-0.72	0.64-0.89	
SWAT	2	Daily	cal.	0.67-0.89	0.67-0.89	Tolera *et al.* (2018)
			val.	0.26-0.94	0.26-0.94	

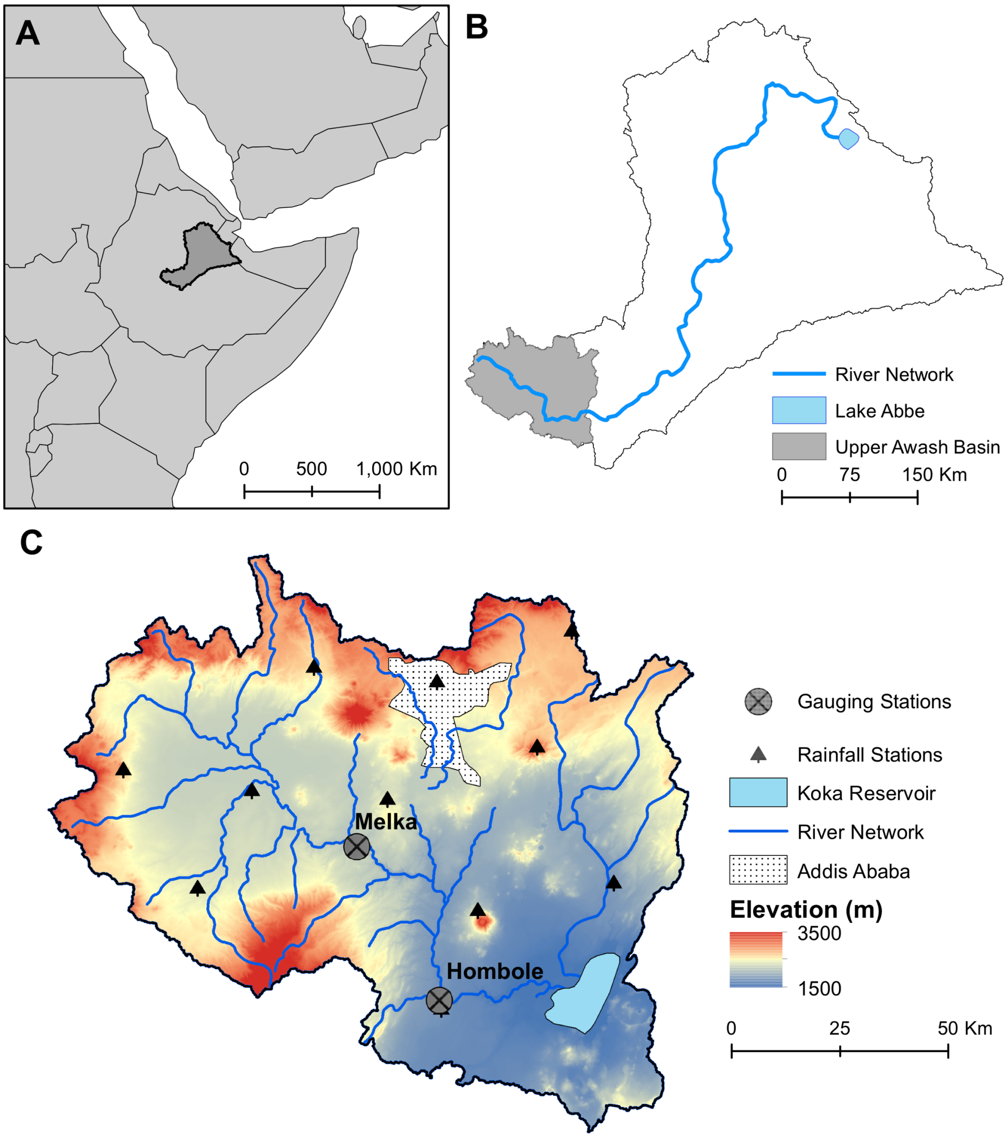

A
0
500
1,000 Km
B
River Network
Lake Abbe
Upper Awash Basin
0
75
150 Km
C
Melka
Hombole
Gauging Stations
Rainfall Stations
Koka Reservoir
River Network
Addis Ababa
Elevation (m)
3500
1500
0
25
50 Km

Figure 1

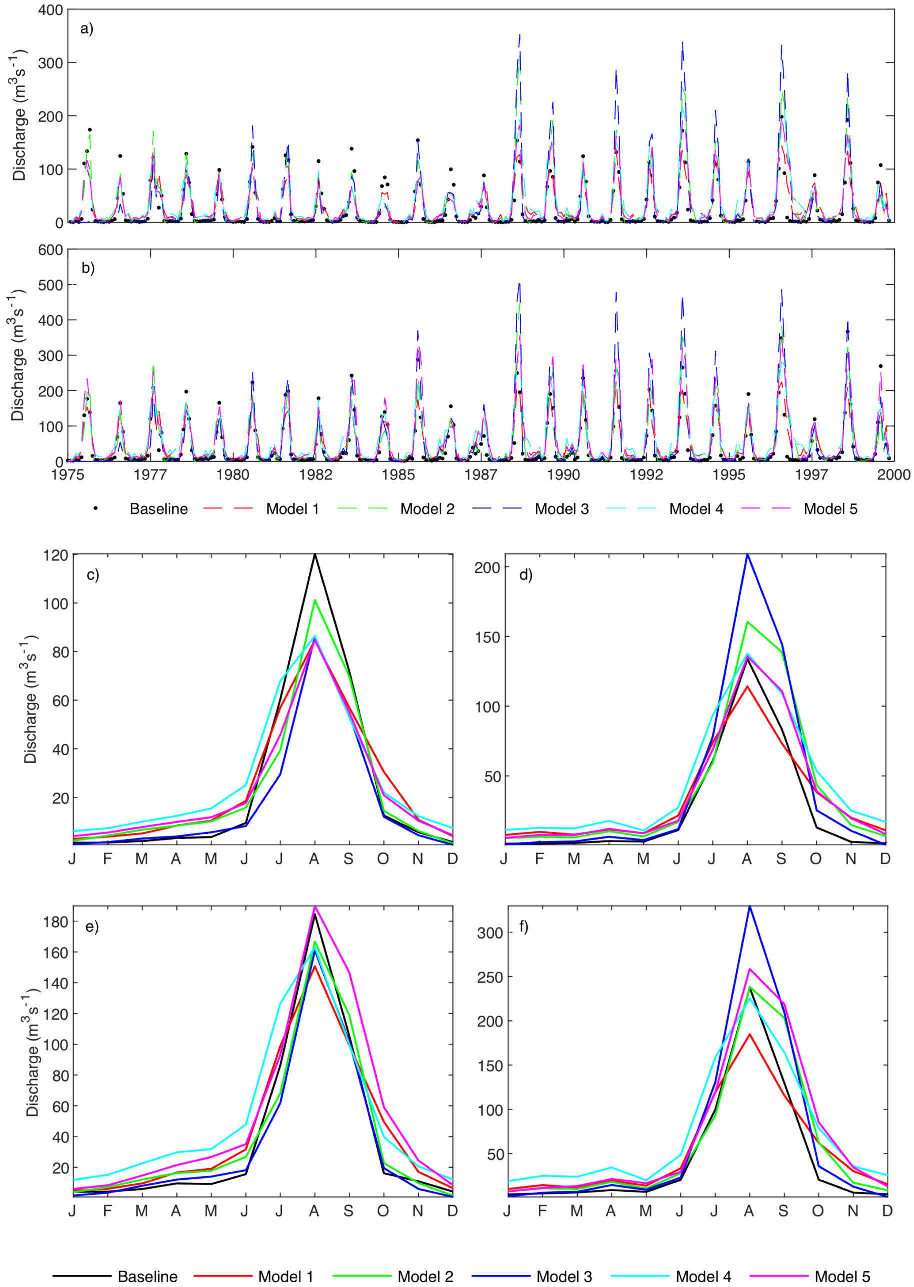
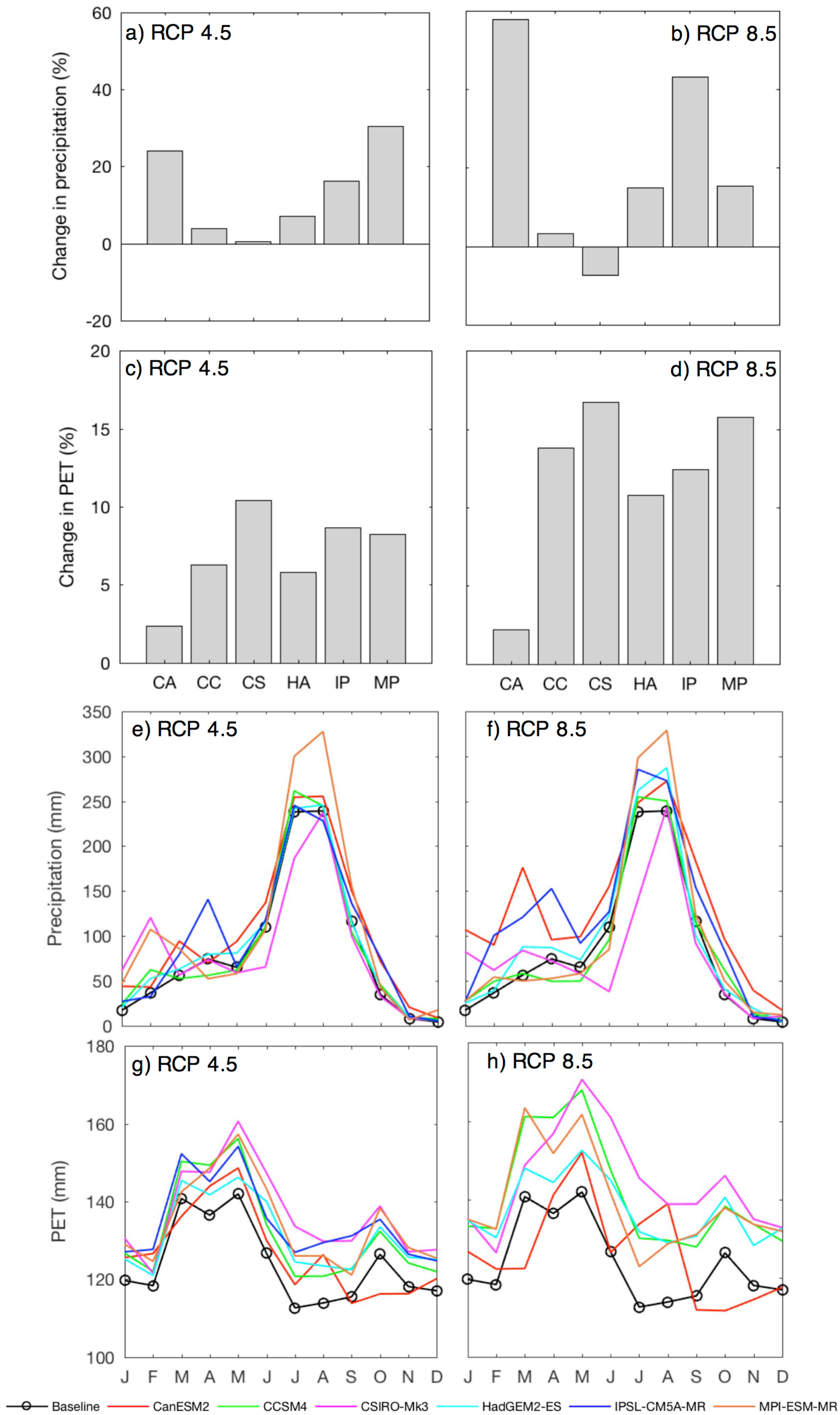

a)
b)
c)
d)
e)
f)
Discharge (m3s-1)
1975
1977
1980
1982
1985
1987
1990
1992
1995
1997
2000
J F M A M J J A S O N D
Baseline
Model 1
Model 2
Model 3
Model 4
Model 5

Figure 2



a) RCP 4.5
b) RCP 8.5
Change in precipitation (%)
60
40
20
0
-20
c) RCP 4.5
d) RCP 8.5
Change in PET (%)
20
15
10
5
0
CA
CC
CS
HA
IP
MP
e) RCP 4.5
f) RCP 8.5
Precipitation (mm)
350
300
250
200
150
100
50
0
g) RCP 4.5
h) RCP 8.5
PET (mm)
180
160
140
120
100
J F M A M J J A S O N D
Baseline
CanESM2
CCSM4
CSIRO-Mk3
HadGEM2-ES
IPSL-CM5A-MR
MPI-ESM-MR

Figure 3

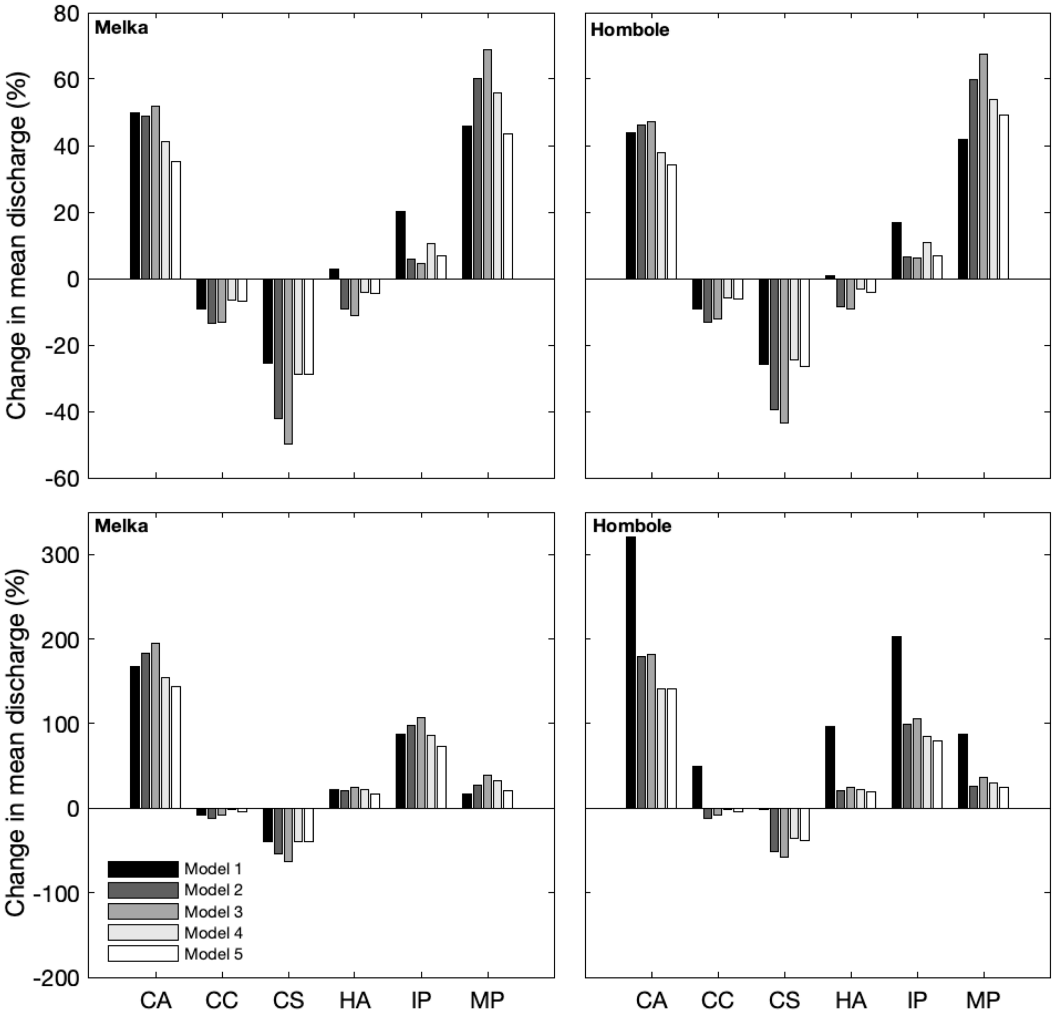

Melka
Hombole
Melka
Hombole
Change in mean discharge (%)
Change in mean discharge (%)
80
60
40
20
0
-20
-40
-60
300
200
100
0
-100
-200
Model 1
Model 2
Model 3
Model 4
Model 5
CA
CC
CS
HA
IP
MP
CA
CC
CS
HA
IP
MP

Figure 4

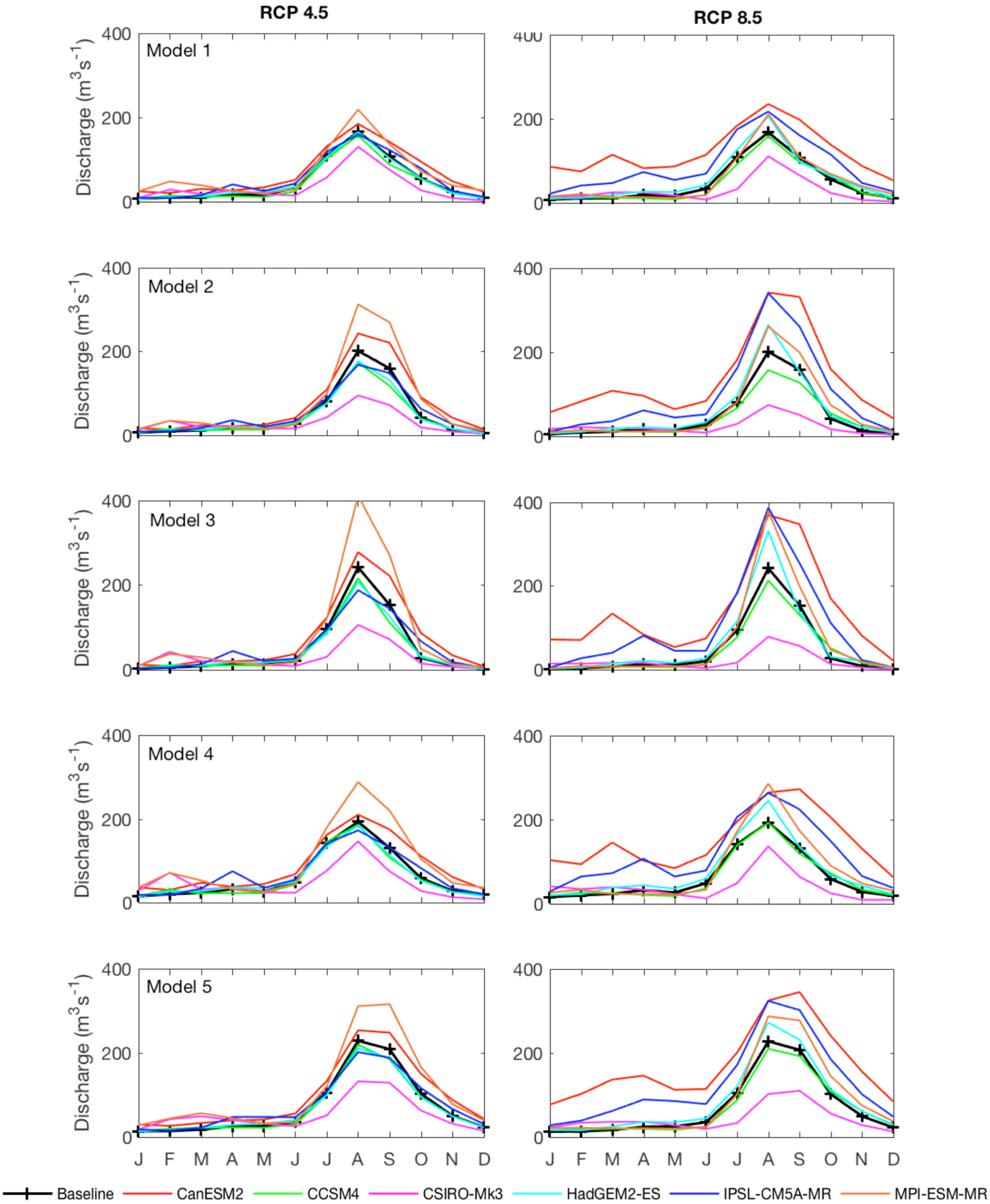

RCP 4.5
RCP 8.5
Model 1
Model 2
Model 3
Model 4
Model 5
Discharge (m3s-1)
400
200
0
J F M A M J J A S O N D
Baseline
CanESM2
CCSM4
CSIRO-Mk3
HadGEM2-ES
IPSL-CM5A-MR
MPI-ESM-MR

Figure 5

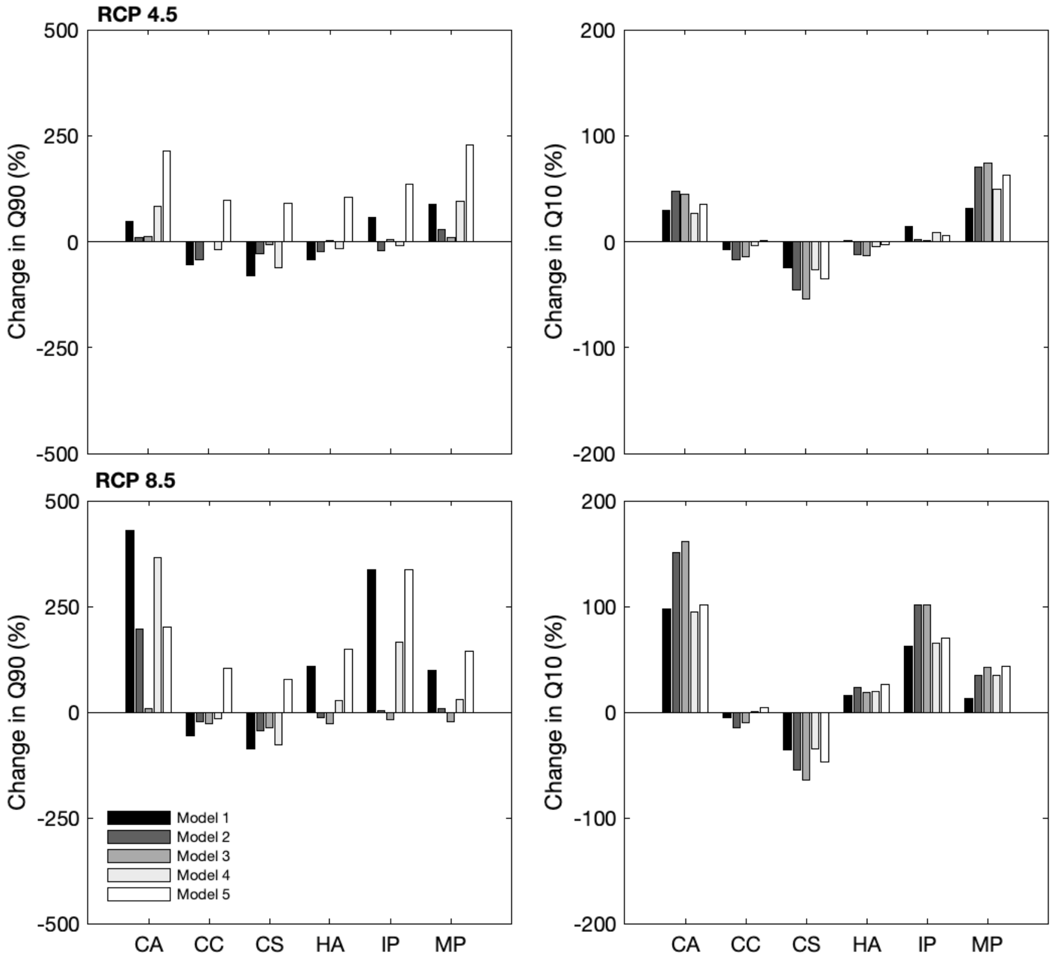

RCP 4.5
RCP 8.5
Change in Q90 (%)
Change in Q10 (%)
500
250
0
-250
-500
200
100
0
-100
-200
Model 1
Model 2
Model 3
Model 4
Model 5
CA
CC
CS
HA
IP
MP

Figure 6

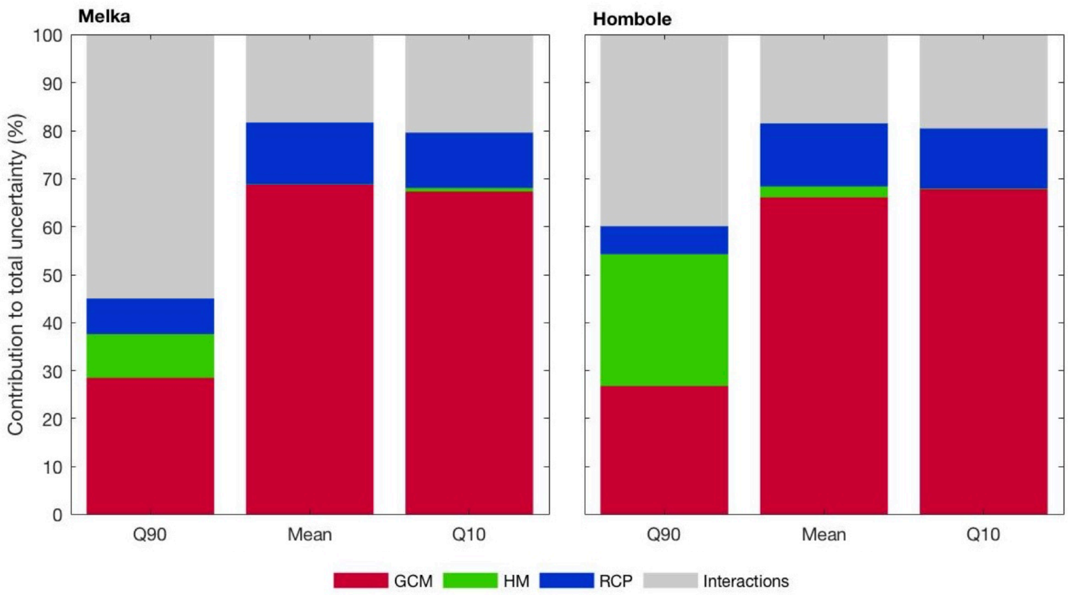

Melka
Hombole
Contribution to total uncertainty (%)
100
90
80
70
60
50
40
30
20
10
0
Q90
Mean
Q10
GCM
HM
RCP
Interactions

Figure 7