



Mass balance calibration and reservoir representations for large-scale hydrological impact studies using SWAT+

Celray James Chawanda¹ · Jeffrey Arnold² · Wim Thiery^{1,3} · Ann van Griensven^{1,4}

Received: 20 March 2020 / Accepted: 29 October 2020 / Published online: 4 December 2020

© The Author(s) 2020

Abstract

Climate change (CC) has a high impact on hydrological processes which calls for reliable projections of CC hydrological impacts at large scales. However, there are several challenges in hydrological modelling at large scales. Large-scale models are often not adapted and evaluated at regional scale due to high computation time requirements or lack of information on human interactions, such as dam operations and irrigation practices at local scale. In this study, we present a regionalised methodology that uses a hydrological mass balance calibration (HMBC) and global datasets to represent reservoir and irrigation practices and apply these to a SWAT+ model for Southern Africa. We evaluate the influence of HMBC and the representation on irrigation and reservoirs on model performance and climate projections. We propose a generalised implementation of reservoirs using global datasets and decision tables to represent irrigation and reservoir management. Results show that inclusion of irrigation, reservoirs and HMBC leads to improved simulation of discharge and evapotranspiration with fewer iterations than a full parameter calibration. There is a substantial difference between projections made by the regionalised model and default model when looking at local impacts. We conclude that large-scale hydrological studies that involve local analysis and spatial mapping of results benefit from HMBC and representation of management practices. The proposed methodology can be scaled up and improve overall projections made by global models.

Keywords SWAT+ · Hydrological modelling · Calibration · Climate change

This article is part of a Special Issue on “How evaluation of hydrological models influences results of climate impact assessment”, edited by Valentina Krysanova, Fred Hattermann and Zbigniew Kundzewicz

✉ Celray James Chawanda
celray.chawanda@vub.be

Extended author information available on the last page of the article

1 Introduction

Climate and hydrological models have been used to assess seasonal and geographic patterns of water resources at the catchment-scale (Kim et al. 2008), regional-scale (Goderniaux et al. 2009) and global-scale (e.g. Yang et al. 2010; Vörösmarty et al. 2000). Although climate change is a global phenomenon, climate change impacts are primarily manifested at the regional scale (IPCC 2014) where most adaptation measures are planned and implemented (Hattermann et al. 2017). As such, developing high-quality climate impact projections at the regional scale is of key importance.

Assessment of continental- and global-scale hydrological models often centres on overall model performance at the cost of model performance in individual catchments (Gosling and Arnell 2011). The models are usually applied at low spatial and temporal resolutions (see Schuol et al. (2008) and Sood and Smakhtin (2015)), and most global hydrological models are not calibrated (Zaherpour et al. 2018; Hattermann et al. 2017; Sood and Smakhtin 2015). In contrast, catchment-scale models focus on details and use more localised and high-resolution input data and are usually calibrated and validated. Koch et al. (2013) emphasise the importance of including water management practices in hydrological modelling, yet data on these practices is rough at the continental and global scales. A review of catchment-scale hydrological models such as MIKE SHE (Zhang et al. 2008), HBV (Lindström et al. 1997), SWAT (Krysanova and Arnold 2008; Van Griensven et al. 2012), HSPF and DWSM (Borah and Bera 2004) shows overall good performance, especially for more spatially detailed model setups. On the other hand, continental- and global-scale models like WaterGAP (Alcamo et al. 2003; Kaspar et al. 2003), Water Balance Model (Vörösmarty et al. 2000), Macro-PDM (Arnell 1999) and LPJ (Gerten et al. 2004) have generally lower performance compared to catchment-scale models for river discharge (Hattermann et al. 2017; Krysanova et al. 2018).

Trambauer et al. (2013) evaluated 16 model applications for suitability for drought studies in Africa. Only five were positively evaluated, four of which were calibrated, catchment-based hydrological models. In accord, Krysanova et al. (2018) concluded that model performance varies systematically between the calibrated regional hydrological models and non-calibrated global hydrological models in favour of the regional models. While there are techniques to deal with uncertain projections, such as process-based emergent constraints (Eyring et al. 2019) and model weighing (Lorenz et al. 2018; Brunner et al. 2019), these findings highlight the importance of model calibration in impact studies. It is clear that large-scale models need to be calibrated and validated for better performance and reliable projections. Recently, Krysanova et al. (2018) suggested five main requirements for comprehensive calibration/validation of the regional-scale hydrological models (RHMs) intended for impact studies, including (a) an observational data quality check, (b) calibration/validation using the differential split-sample (DSS) test, (c) checking model performance at multiple sites within the catchment and for several variables, (d) an assessment of the model's ability to reproduce the hydrological indicators of interest and (e) validation for observed trends and/or doing a proxy climate test. Nevertheless, calibration remains a challenge for many large-scale models due to the computational and time cost for calibration at a large scale.

Most hydrological models such as SWAT (Arnold et al. 2012) are calibrated using a time-series of discharge at the outlet of the catchment. However, calibration with *soft data* may enhance the calibration process, especially for large-scale applications. Soft data is the information on individual processes, e.g. long-term annual average estimates (Arnold et al. 2015). We consider hard data as measured time series of hydrological quantities at a point in the watershed. In their study, Arnold

et al. (2015) recommended the use of soft data to set constraints during calibration to improve the representation of hydrological processes. Still, calibration performed entirely using soft data can potentially improve the model performance at lower computational and time costs. Furthermore, soft-data calibration can improve the process representation in impact studies by more realistically simulating the components of the hydrological mass balance.

The objectives of the study are to (a) present a mass balance calibration approach to calibrate large-scale models based on soft data to reduce computation costs that come with conventional calibration methods, (b) present a method that uses global datasets to represent reservoir and irrigation practices in SWAT+ (Arnold et al. 2018), and (c) evaluate the influence of the mass balance calibration approach and management practices on model performance and climate projections.

In this study, we propose hydrological mass balance calibration (HMBC) methodology for calibrating regional-scale models. HMBC can easily be scaled to continental and global scales. We base the HMBC on soft data derived from global datasets. We also implement so-called decision tables (Arnold et al. 2018) to represent reservoir operations and irrigation practices. We investigate the influence of HMBC and implementation of management practices on model performance and future projections over Southern Africa by comparing the default model setup and the model calibrated using HMBC with implementation of human interactions (henceforth referred to as *regionalised* model setup).

2 Material and methods

2.1 SWAT+ model

The SWAT model is a time-continuous semi-distributed hydrological model. SWAT+ is a completely restructured version of SWAT (Bieger et al. 2017), and it is usually applied at catchment-scale to simulate surface water and groundwater quality and quantity. SWAT uses the concept of hydrologic response units (HRUs) which are areas of unique properties of land use, soil and slope class within each subbasin. Calculations are done at the HRU-level. Water, nutrients and sediment fluxes are aggregated for each subbasin and then routed through the channel network towards the main outlet of the catchment.

SWAT has been applied in studies across the world to project the environmental impact of land use (e.g. Shen et al. 2015), land management and climate change (e.g. Krysanova and Srinivasan 2014; Li et al. 2015; Kingston and Taylor 2010). Schuol et al. (2008) used SWAT to model blue and green water availability in Africa. In contrast, SWAT+ has not been widely used since it is relatively new.

While SWAT+ uses the same equations as SWAT to represent hydrological processes, it offers more flexibility to users in configuring the model. Connect files improve flexibility in connecting different elements in the model. SWAT+ also introduces landscape units (LSUs) which are finer divisions of subbasins to allow separation of upland processes from wetlands (Bieger et al. 2017). While subbasins are delineated based on the threshold for streams, LSUs are delineated based on channel threshold, the value of which must be smaller than or equal to the stream threshold.

Newly introduced decision tables allow users to specify conditions for various management activities within the model. The use of decision tables in SWAT+ is extensively discussed in Arnold et al. (2018) and summarised hereafter. Arnold et al. (2018) describe decision tables as a precise yet compact way to model complex rule sets and their corresponding actions. Decision tables list

conditions which must be fulfilled for an action to be performed within the model. Arnold et al. (2018) tested automated irrigation and reservoir release with decision tables and concluded that decision tables are a robust and efficient method to simulate complex, rule-based management. Management schedules which allow timing of events in the model at specific times are also available in SWAT+. The availability of decision tables and management schedules makes the implementation of river basin management practices in SWAT+ more flexible. In this study, we used revision 55 of SWAT+.

2.2 Methodological advances

2.2.1 Representation of human interactions

In this study, we devised a methodology to identify irrigated areas in the model and implement irrigation. We achieve this by intersecting global irrigation maps (shapefile) and the HRU shapefile to identify HRUs that have agricultural land use and have irrigation. Then, the HRUs are split into irrigated and rainfed HRUs based on areas acquired from the intersection. Auto irrigation, a SWAT+ feature for irrigating crops when there is water stress, is then specified in the irrigated HRUs using decision tables and management schedules (Arnold et al. 2018).

Global reservoir data such as the Global Reservoir and Dam (GRanD) database (Lehner et al. 2011) can be used to set up reservoirs in SWAT+. The setup is achieved by automatically populating reservoir properties such as depths and volumes extracted from the attribute table of the shapefile. The reservoirs are then classified based on purpose, and decision tables for the operation of the dam are generalised based on the purpose of the reservoir (water supply, hydro-electric power generation, flood control, irrigation). The implementation of both irrigation and reservoirs is done using Python scripts.

2.2.2 Hydrological mass balance calibration

Catchment-scale hydrological models require calibration and a discharge time-series at the catchment outlet is often used as a reference. Objective functions such as Nash-Sutcliffe efficiency (NSE), coefficient of determination (R^2) and percent bias (PBIAS) are used to decide whether a model has been calibrated successfully (Moriassi et al. 2007). However, models that are calibrated as such may have unrealistic proportions of the hydrological mass balance components and, thus, may not be suitable for long-term projections or case studies. Besides, calibration of only one variable, such as runoff, may not be sufficient, and additional calibration for other water balance components, such as ET, using soft data could be beneficial. The aim of the hydrological mass balance calibration (HMBC) is to adjust model parameters in order to match simulated long-term annual average water balance components (Eq. 1) expressed against the observed long-term annual average water balance components.

$$P_{\text{avg}} = ET_{\text{avg}} + SR_{\text{avg}} + GW_{\text{avg}} \quad (1)$$

where:

P_{avg}	long-term annual average precipitation
ET_{avg}	long-term annual average evapotranspiration
SR_{avg}	long-term annual average surface runoff

GW_{avg} long-term annual average of the amount of water that infiltrates and becomes subsurface flow

Calibration of the water balance using HMBC requires a far smaller number of runs and hence less time cost compared to traditional calibration on observed discharge time series. This makes it easily applicable in hydrological model applications at continental and global scales. The data requirement for HMBC is the long-term annual averages of the water balance components expressed in Eq. 1 which can be derived from observed discharge data, rainfall data and evapotranspiration data obtained from remote sensing.

The first step in performing HMBC includes subdividing the study area into calibration zones based on subcatchments or expert judgement. Secondly, long-term annual averages for each of the hydrological mass components such as ET, precipitation and surface runoff are derived for each calibration zone from global datasets. The long-term averages are expressed as ratios of the water balance components to precipitation; e.g. 400 mm annual average ET in a 1000 mm precipitation zone will have an ET ratio of 0.4. Finally, parameters for each region are automatically adjusted by an algorithm written in Python using the ratios derived from soft data. If the model hydrologic balance ratio is different from the one derived from soft data, each ratio is adjusted by changing parameters that directly affect the corresponding water balance component. Examples of parameters that can be adjusted are listed in Table S1. The HMBC method implements the following procedure:

1. Run the model without changing default model parameter values. Default parameters are assigned by SWAT+ interfaces when building the model.
2. Evaluate the results against soft data.
3. Estimate new parameter values based on difference between simulation and soft data.
4. Re-run the model with the new parameters.
5. Re-evaluate the results against soft data.
6. Repeat steps 3 to 5 until the specified threshold (default value is 2 mm) or the limit of the current parameters (default value is $\pm 20\%$) is reached.

Usually, a hydrological mass balance component for a subregion is calibrated within five iterations before moving on to the next component.

2.3 Application to Southern Africa

The method links to the five main requirements for an appropriate calibration/validation of RHMs intended for impact studies (Krysanova et al. 2018) during the setup of the SWAT+ model for the Southern African region (Fig. 1a). We use HMBC (Section 2.3.5) instead of the normal calibration suggested by Krysanova et al. (2018) but still perform (a) observational data quality check and (b) model performance check at multiple sites within the catchment and for several variables.

2.3.1 Study area

Our study area covers about 2,179,400 km² in the southern part of Africa (Fig. 1a). This area covers several countries including South Africa, Lesotho, Swaziland, and parts of Botswana,

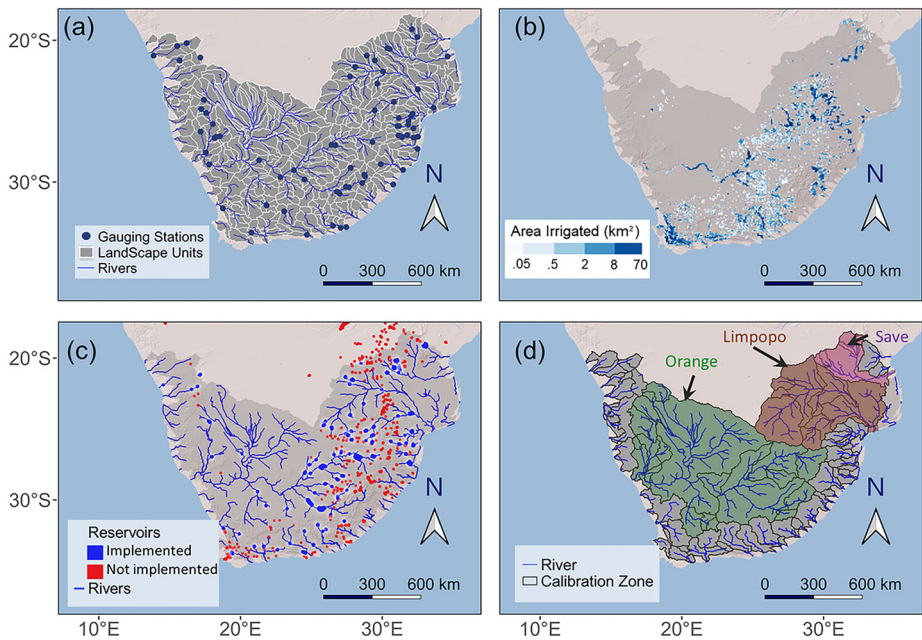


Fig. 1 Maps of the study area: the Southern Africa region including the Orange, Limpopo and Save river basins showing (a) river network generated during catchment delineation, delineated landscape units (LSUs) and points where gauging stations are located. (b) Irrigation map with irrigated area in km² in each pixel at a resolution of $0.083^\circ \times 0.083^\circ$. (c) Reservoir map showing which reservoirs were implemented and which were not. (Note that only reservoirs that are on a stream at the model's chosen level of detail were implemented in the model. The sizes of the reservoirs have been exaggerated so that they are large enough to be visible on the figure). (d) Subdivision of the study area into calibration zones

Namibia, Zimbabwe and Mozambique. Major river basins in the study area include Orange, Limpopo, Maputo, Buzi, Save, Olifants and Groot.

2.3.2 Global data sets for SWAT+ model setup and HMBC

Modelling was done using freely available data as follows:

- The digital elevation model (DEM) data was taken from the Shuttle Radar Topography Mission (Farr et al. 2007, downloadable from <http://dwtkns.com/srtm>).
- The land use map was prepared from the Land-Use Harmonization Project phase 2 (LUH2; Hurtt et al. 2011).
- Soil data with $250\text{m} \times 250\text{m}$ resolution was obtained from the Africa Soil Information Service (AfSIS) (Hengl et al. 2015).
- Monthly discharge observation data for 69 stations was obtained from Global Runoff Data Centre (GRDC, n.d.).
- Irrigated areas were obtained from Food and Agriculture Organisation at $0.083^\circ \times 0.083^\circ$ resolution (FAO; Siebert and Frenken 2014).
- Reservoir data was obtained from the Global Reservoir and Dam (GRanD) database (Beames et al. 2019).

- (g) Yearly ET data was obtained from CSIRO's Moderate resolution imaging spectroradiometer Reflectance Scaling Evapotranspiration (CMRSET; Guerschman et al. 2009) and Water Productivity through Open access of Remotely sensed derived data (WaPOR; FAO 2018) at the resolution of $0.0022^\circ \times 0.0022^\circ$.
- (h) For weather forcing, Earth2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI; Lange 2016) dataset was obtained through the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) at daily timestep. We also obtained daily weather forcing through ISIMIP for 4 bias-adjusted global climate models (GCMs) under the Representative Concentration Pathway (RCP) 6.0: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC5.

For reservoirs, the attributes extracted from the GRanD database include the name of the reservoir, its location, elevation, principal volume and corresponding surface area, emergency volume and its corresponding surface area, and year when the dam/reservoir started operation.

Detailed soil properties for SWAT modelling are derived from the AfSIS data using the Saxton and Rawls (2006) pedo-transfer functions and were acquired from Ayana et al. (2019).

EWEMBI Weather data includes records of precipitation ($\text{kg m}^{-2} \text{s}^{-1}$), minimum and maximum temperatures (K), solar radiation (W m^{-2}), wind speed and relative humidity (%) which were available from the year 1979 to 2016. EWEMBI has a spatial resolution of 0.5° . For this reason, weather variables such as precipitation may not be adequately represented, especially where the topography is very variable. However, this resolution is sufficient for large-scale modelling.

The temporal coverage of the GRDC discharge data varies from station to station. Some stations had data before the year 1970 and some beyond 1990, but the records in many stations ended in the 1980s. We eliminated stations based on temporal coverage. Stations that had less than 3 years of discharge data within the range 1979 to 1986 were removed. We also manually reviewed the discharge values to check for inconsistencies. In some stations, it was not clear if some days had zero river discharge values or missing data, e.g. GRDC station 1196400 (Fig. S8). These were also removed.

Irrigation data was provided formatted as a shapefile (Fig. 1b) with an attribute table containing area fraction equipped for irrigation and area fraction of the area equipped for irrigation that is actually irrigated every year.

Weerasinghe et al. (2020) compared actual ET products over Africa and found that CMRSET and WaPOR ranked highest in terms of bias in long-term annual mean ET estimates. The CMRSET product is derived from reflectance data from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite. The WaPOR ET dataset is derived using the ETLook-WaPOR model described in Bastiaanssen et al. (2012) which uses the Penman-Monteith (P-M) equation adapted to remote sensing input data (Blatchford et al. 2020). CMRSET and WaPOR products differ in terms of spatial ET patterns, especially in arid areas where WaPOR has more realistic patterns. There was an overlap of 7 years of WaPOR data with weather forcing and 13 years for CMRSET.

2.3.3 Setup of the default model

The QGIS SWAT+ (QSWAT+) interface was used to set up the model and run it at the daily timestep. QSWAT+ is very similar to QSWAT (Dile et al. 2017) with additional features for

creation of LSUs. Based on the DEM from SRTM, the study area was discretised into 599 LSUs using a threshold drainage area of 3500 km² (Fig. 1a).

It was not possible to directly use the LUH2 product with QSWAT+ because it was formatted as NetCDF with each layer containing a percentage of given land use (Fig. S9) instead of a single-layer raster file. The QSWAT+ code was therefore adapted to include routines that read the LUH2 NetCDF data. A ‘placeholder’ raster that was equally dimensioned as the LUH2 map was used for creating HRUs. All the values in the placeholder raster were unique. As such, initial HRUs were equal to or less than the pixel in size of the LUH2 map. The initial HRUs were later split based on the corresponding land use percentages for each pixel in all the layers of the NetCDF file; i.e. the original HRU was split by area based on the land use proportions in the LUH2 dataset.

2.3.4 Representation of management practices

Irrigation was implemented as described in Section 2.2.1 using FAO irrigation map. In total, there were 39,365 HRUs. Routing, ET estimation and surface runoff calculation methods were left as default. Reservoirs were implemented using reservoir properties acquired from the GRanD database. Only reservoirs that fell on a stream within the catchment at the threshold used for delineation were implemented (Fig. 1c). Decision tables for the operation of the dams were generalised based on the purpose of the reservoir (water supply, hydro-electric power generation, flood control, irrigation), as shown in Table S2.

2.3.5 Application of hydrological mass balance calibration

We divided subcatchments into calibration zones based on expert judgement (Fig. 1d). Long-term averages for Eq. 1 for each calibration zone were derived from global datasets as follows: ET_{avg} from WaPOR product, P_{avg} from EWEMBI (Lange 2016) and SR_{avg} from baseflow separation using Water Engineering Time Series PROCESSING tool (WETSPTO, Willems 2009) applied to GRDC discharge records. GW_{avg} was determined by subtracting ET_{avg} and SR_{avg} from P_{avg} . Thus, lateral flow from SWAT+ was lumped into GW_{avg} . Hydrological mass balance components from SWAT+ are obtained as follows: ET_{avg} is from ‘et’ column, P_{avg} is from ‘precip’, SR_{avg} is from the ‘surq_gen’ and GW_{avg} is from summation of ‘latq’ and ‘perc’ columns of SWAT+ water balance results. If the annual ET_{avg} value from WaPOR was found to be higher than the annual P_{avg} amount, the ET_{avg} for that zone was not used in HMBC.

We expressed the input data for HMBC as ratios of the water balance components to precipitation (Table S2). All the parameters listed in Table S1 (cn2, lat_len, esco, epco, alpha) were adjusted for each region automatically using an algorithm written in Python using ratios presented in Table S2.

2.3.6 Evaluation

Our study focused on river discharge and ET. We assessed the performance of the model for stream discharge at monthly timestep at multiple sites within the catchment. We did the evaluation using GRDC discharge records between 1979 and 1986 with 90% of the stations having more than 5 years discharge records data. We also assessed yearly and annual average ET fluxes using CMRSET and WaPOR over the reference period. We then apply climate scenarios to project the impact of future climate on ET and river discharge.

2.3.7 Climate change scenarios

The focus of the study is not on climate change for our study area, but to illustrate a concept. We applied climate change scenarios using the 4 GCMs (GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC5) under RCP 6.0. We selected RCP 6.0 because we expect a higher signal compared to RCP 2.6.

For the present-day reference, we ran SWAT+ from 1970 to 2000 and for the future scenario from 2060 to 2090. The scenarios were run for both the initial model setup and the regionalised model, and each has been assigned a name (Table 1). The first year of each simulation was considered spin-up and excluded from the analysis. In total, we ran 16 simulations with GCM forcing (i.e. 2 historical simulations \times 2 future simulations \times 4 GCMs). Our analysis is based on the average of four SWAT+ integrations forced with each of the 4 GCMs.

Table 1 All simulations that were run and their descriptions

Simulation name	Observed climate	GCM historical	GCM Future	Irrigation	Reservoirs	Calibrated with HMBC	Description
OBS_D	✓						Model with default parameters forced with observed climate
OBS_DI	✓			✓			Model with default parameters, implemented irrigation forced with observed climate
OBS_DRI	✓			✓	✓		Model with default parameters, implemented reservoirs and irrigation forced with observed climate
OBS_C	✓					✓	Model with parameters calibrated by HMBC forced with observed climate
OBS_CRI	✓			✓	✓	✓	Model with calibrated parameters, implemented reservoirs and irrigation forced with observed climate
HIS_D		✓					Model with default parameters forced with historical climate
HIS_CRI		✓		✓	✓	✓	Model with calibrated parameters, implemented reservoirs and irrigation forced with historical climate
FUT_D			✓				Model with default parameters forced with future climate
FUT_CRI			✓	✓	✓	✓	Model with calibrated parameters, implemented reservoirs and irrigation forced with future climate

3 Results

3.1 Importance of considering water management

The observation-driven, uncalibrated simulation (OBS_D) underestimates ET compared to the reference products WaPOR and CMRSET (Fig. 2a, c and d). WaPOR reflects a comparable spatial pattern to OBS_D ET (Fig. 2c), but that is not the case for CMRSET (Fig. 2d). The Kalahari Desert area has high ET values in CMRSET which are not reflected in the simulated ET. Such high ET values are not expected in the Kalahari, considering that there is very low annual precipitation, and there is not much irrigation according to the FAO irrigation dataset (Fig. 1b). The observation-driven calibrated model with irrigation and reservoirs (OBS_CRI) better reflects ET by simulating higher ET than OBS_D in areas where irrigation is taking place according to Fig. 1b (Fig. 2a, b).

The OBS_D simulation demonstrates poor performance in simulating river discharge when compared with GRDC data, both in terms of NSE and PBIAS. Only 8 out of 69 stations had an NSE greater than 0.5 (Fig. 3a).

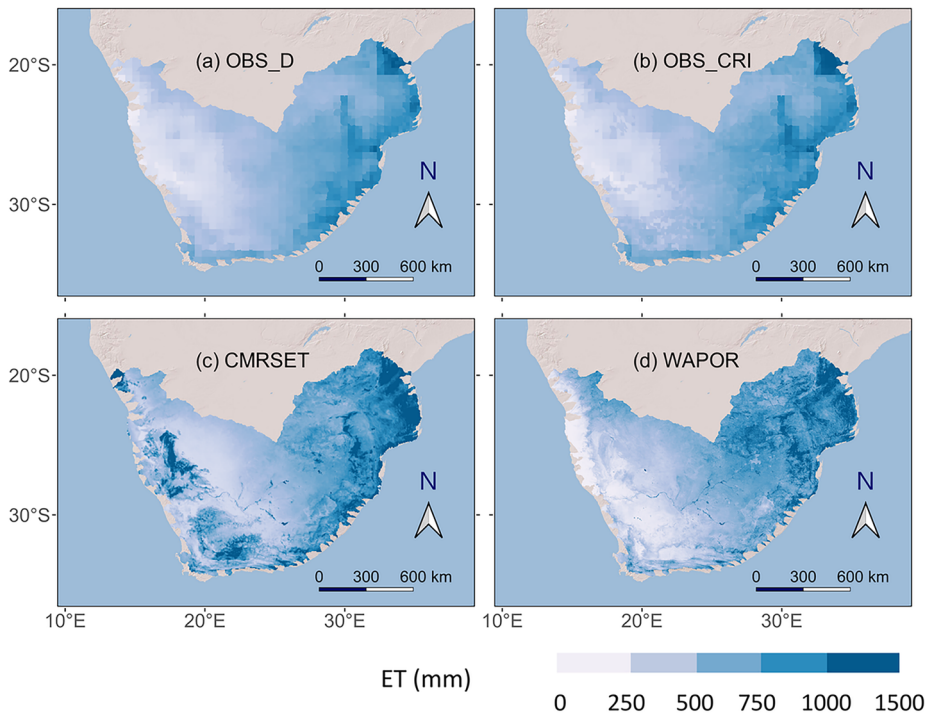


Fig. 2 Annual averages (mm) of (a) simulated ET (2001–2016) from model with default parameters forced with observed climate (OBS_D) presented at 1 km resolution. (b) Simulated ET (2001–2016) from model with calibrated parameters, implemented reservoirs and irrigation forced with observed climate (OBS_CRI) presented at 1 km resolution. (c) Observed ET from CMRSET presented at 2.77 km resolution. (d) Observed ET from WaPOR presented at 250 m resolution. Average ET values for major subbasins for (i) OBS_D: Orange: 281 mm; Limpopo: 469 mm; Save: 554 mm; (ii) OBS_CRI: Orange: 305 mm; Limpopo: 506 mm; Save: 556 mm; (iii) CMRSET: Orange: 336 mm; Limpopo: 615 mm; Save: 746 mm; (iv) WAPOR: Orange: 301 mm; Limpopo: 704 mm; Save: 696 mm

Implementation of irrigation (OBS_DI) did not result in much improvements in the simulation of both ET and discharge. However, irrigation in the model increased ET over irrigated areas (Fig. 4a). The role of reservoirs could be seen particularly when the discharge is controlled at a dam as shown in Fig. 4b (e.g. discharge is controlled at Vanderkloof (GRanD ID 4254) for station Vlyutjeskraal Station (GRDC 1159303)). After implementing reservoirs alongside irrigation (OBS_DRI), the simulation of discharge downstream of reservoirs improved. The overall distribution of performance indicators also slightly improved (Fig. 4c (ii)). The improvement in discharge simulation downstream of reservoirs can also be seen in the hydrographs (e.g. Fig. 4 (iii)).

3.2 Importance of HMBC

There was an improvement in the simulation of ET magnitude and spatial patterns after HMBC (Fig. 2b). Note the higher ET in areas where irrigation is present with reference to irrigated areas. There was also an improvement in the simulation of discharge for many stations after HMBC (Fig. 5). The number of stations with NSE > 0 increased from 41% in OBS_D simulation to 72% in the OBS_CRI simulation.

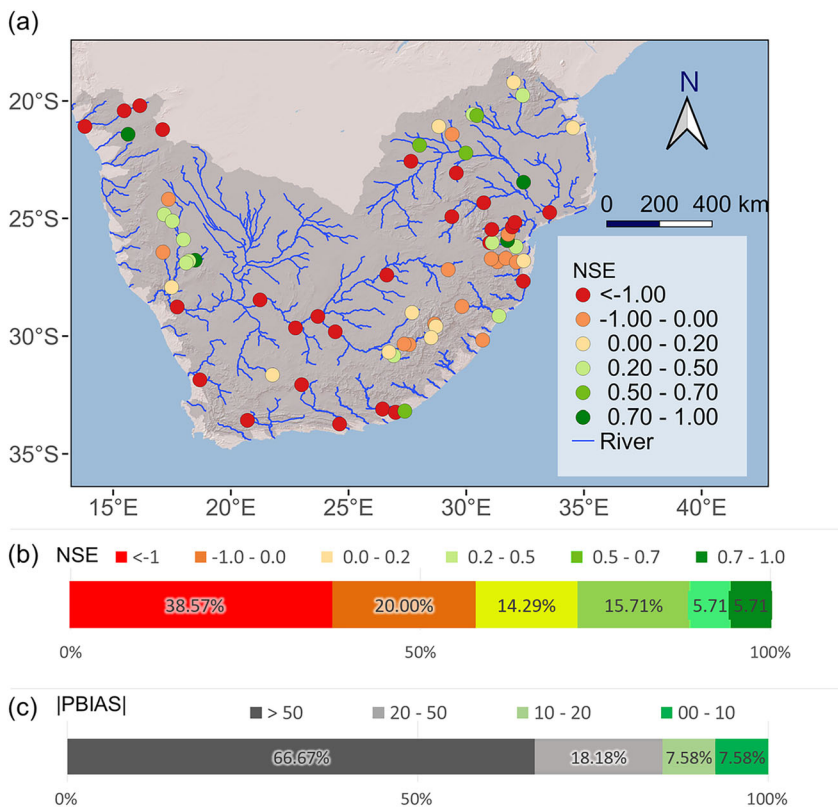


Fig. 3 Model performance for the model with default parameters forced with observed climate (OBS_D) **a** model performance in simulating discharge at monthly timestep. **b** Distribution of values of NSE for OBS_D. **c** Distribution of absolute values of PBIAS for OBS_D

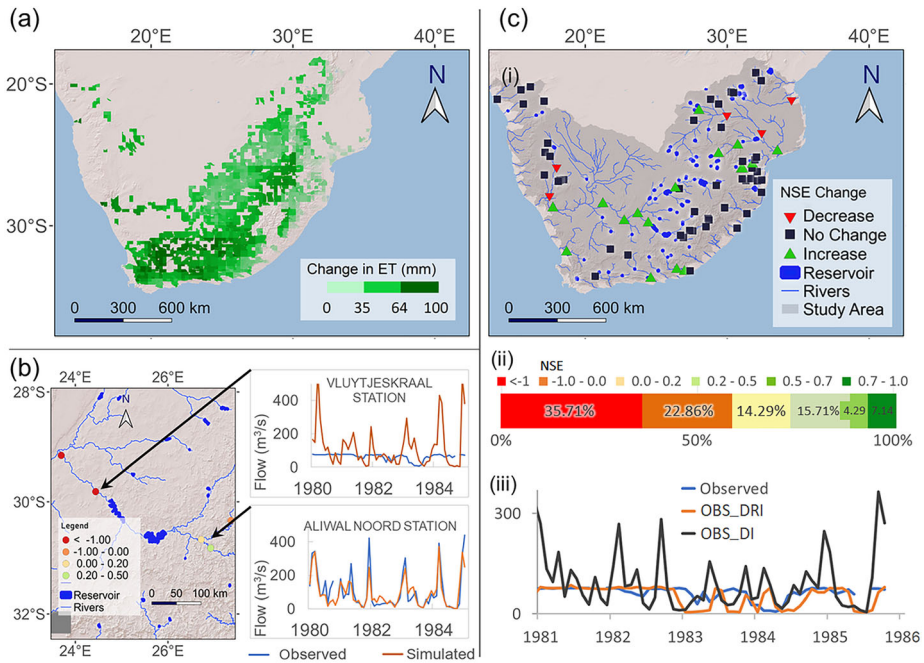


Fig. 4 Influence of implemented irrigation and reservoirs on ET and simulated discharge. **(a)** Annual average ET difference between the model with default parameters forced with observed climate (OBS_D) and OBS_DI presented at 1 km resolution. **(b)** An example of the impact of reservoirs on model performance at Vluutjeskraal Station. **(c)** (i) Impact of reservoirs on the discharge simulation performance. (No change refers to a difference of zero between NSE values OBS_DI and model with default parameters, implemented reservoirs and irrigation forced with observed climate (OBS_DRI)). (ii) Distribution of values of NSE for OBS_DRI. (iii) Time series at Vluutjeskraal Station (longitude: 24.4, latitude: -29.8) before (OBS_DI) and after implementing reservoirs (OBS_DRI) using decision tables

PBIAS distribution (Fig. 5c) improved in OBS_CRI as compared to OBS_D (Fig. 3c). Overall, there were improvements in the model after adding reservoirs and again after calibration. Implementing irrigation (OBS_DI) did not have a significant impact on model performance in terms of discharge. Addition of reservoirs to OBS_DI (OBS_DRI) slightly improves the low performing gauges while adding HMBC to OBS_DRI (OBS_CRI) improves performance in many stations considerably. Implementing the reservoirs resulted in an overall better calibrated model (Fig. 5d).

We evaluated the OBS_D, OBS_DRI and OBS_CRI model results for the period that overlaps with CMRSET and WaPOR. ET values increase from OBS_D to OBS_DRI to OBS_CRI. The modelled ET is more dynamic from year to year than CMRSET (Fig. 5e). However, the modelled ET is strongly correlated with the CMRSET data (e.g. $R^2 = 0.88$ for OBS_CRI).

3.3 Climate change scenarios

Annual average ET was higher in the calibrated, GCM-driven historical ensemble (HIS_CRI) than in the future ensemble (FUT_CRI). A similar signal was found for the uncalibrated historical (HIS_D) and future simulations (FUT_D; Fig. 6). The climate change signal varied in space (Fig. 7d). As expected, areas that were irrigated had higher

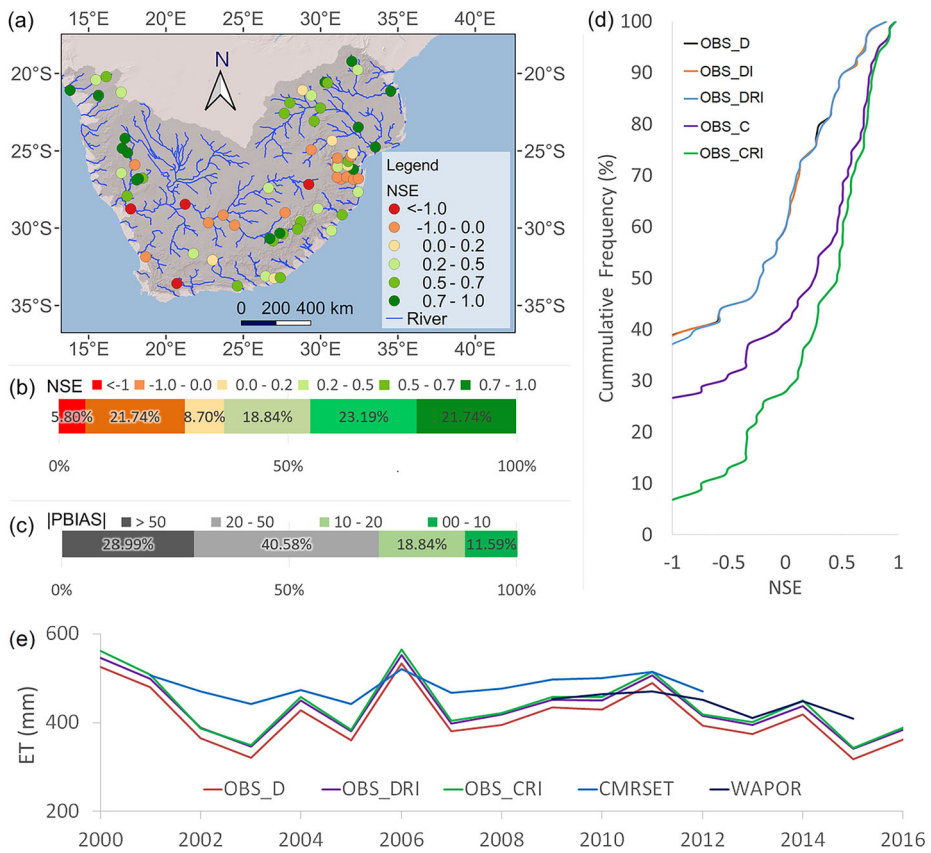


Fig. 5 Model performance after implementing HMBC. **(a)** Model performance in simulating discharge (monthly) for model with calibrated parameters, implemented reservoirs and irrigation forced with observed climate (OBS_CRI). **(b)** Distribution of NSE values for OBS_CRI. **(c)** Distribution of absolute values of PBIAS for OBS_CRI. **(d)** Overall improvements as management and HMBC are added to the model. (The model with default parameters and irrigation forced with observed climate (OBS_DI) has a small improvement in NSE values for a few stations compared to the model with default parameters forced with observed climate (OBS_D)). **(e)** Yearly ET for OBS_D, model with default parameters, implemented reservoirs and irrigation forced with observed climate (OBS_DRI) and OBS_CRI compared to CMRSET and WaPOR

simulated ET values than those which were not. A visual artefact can be seen at 34° S, 25° E in Fig. 7d. A similar pattern is visible in the same area and in many other areas when a difference between future and present precipitation of any of the GCM product is mapped (Fig. 7e). There is not much difference in surface runoff in default (HIS_D, FUT_D) and regionalised (HIS_CRI, FUT_CRI) models (Fig. 6d, e, f). Just like for ET, the relative changes in average discharge in future climate were comparable between the default and the regionalised models (Fig. 7a and b).

Still, there were some stations from the regionalised model that displayed a reverse change in the default model. There was also a substantial difference between the default and regionalised models in projected river discharge (Fig. 7c). FUT_D projected *higher* average discharge in the range 10 to 90% than FUT_CRI in 44 gauging stations and 6% to above 100% *lower* average discharge than FUT_CRI in 25 gauging stations.

4 Discussion

In this study, we modelled the Southern Africa region with SWAT+ using global datasets and implemented HMBC and reservoirs and irrigation in the model. In the analysis, we considered observational data quality check and model performance check at multiple sites within the catchment out of the requirements proposed by Krysanova et al. (2018).

4.1 Data quality and model performance

Input data quality influences model performance and evaluation. There are not many global products for the required inputs with a spatial resolution higher than 0.5°. In this study, we used daily, bias-adjusted reanalysis data at a spatial resolution of 0.5°. This resolution is too coarse to capture the fine-scale spatial variability of weather, especially for rainfall which is an important input to SWAT+. Therefore, the performance of the model in simulating different variables at higher resolutions than 0.5 is expected to be low. Many large-scale models face this limitation, and improvements in weather input resolution are necessary for better performance in large-scale hydrological modelling.

Besides model input data, observed hydrological data plays an essential role in model performance assessment. Records of most stations in Southern Africa end around 1986. Considering reanalysis weather data from EWEMBI starts only in 1979, the overlap between forcing and validation data is not long enough to enable a month-by-month evaluation. Furthermore, data from some gauging stations was removed while evaluating the quality of observational data, as recommended by Krysanova et al. (2018). A good example is Botswana Oxenham Ranch station (latitude:

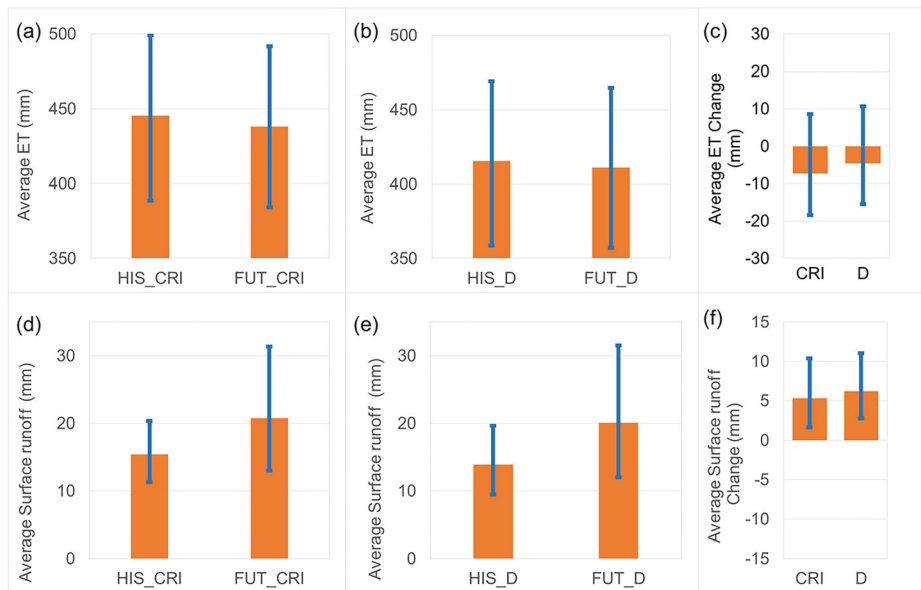


Fig. 6 Comparison of simulated ET (a, b) and surface runoff (d, e) between the historical (HIS, 1970–2000) and future (FUT, 2060–2090) periods for the HMBC calibrated model with implemented irrigation and reservoirs (CRI) and model with default parameters (D) driven by four GCM scenarios, and changes in ET (c) and surface runoff (f) between the same periods based on the same two model versions

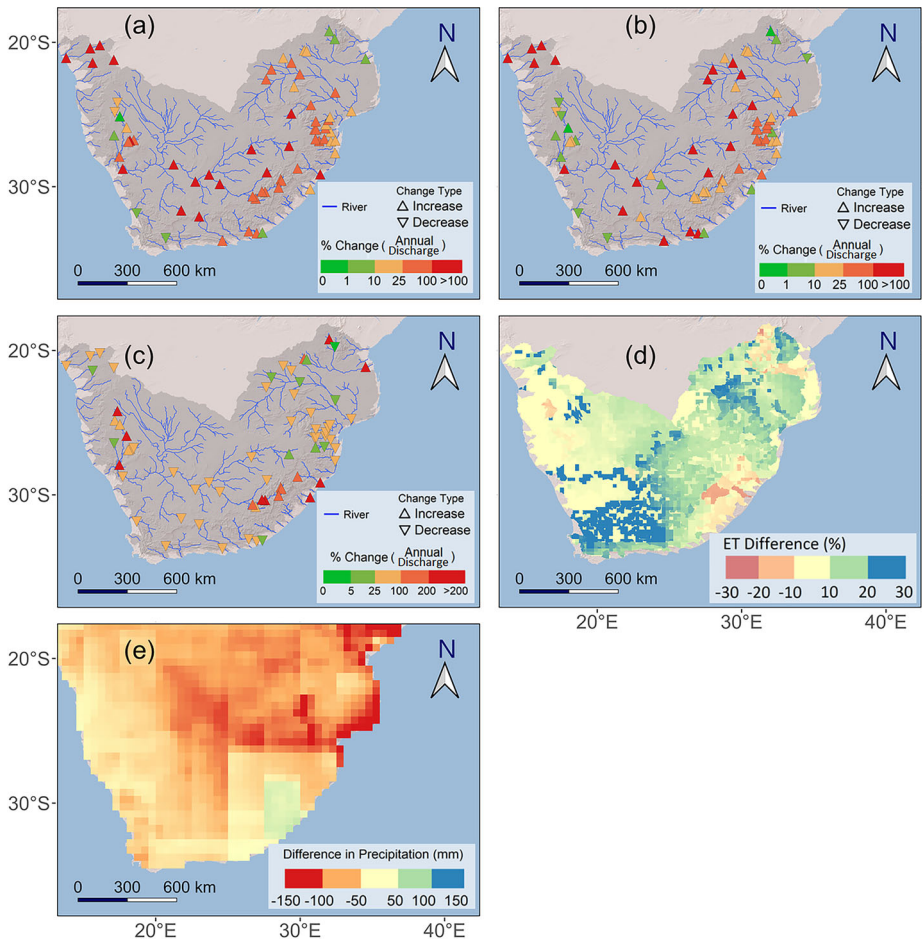


Fig. 7 Comparison of spatial variation of projected climate signal for river discharge between (a) model with default parameters (D) and (b) HMBC calibrated model with implemented irrigation and reservoirs (CRI) presented at 1 km resolution, and spatial differences in future projections (FUT, 2060–2090) of discharge (c) and ET (d) between simulations by model with calibrated parameters, implemented reservoirs and irrigation (FUT_CRI) and the model with default parameters (FUT_D) expressed as a percentage of projections from FUT_CRI, i.e. $(FUT_CRI - FUT_D) / FUT_CRI * 100$. (e) An example of visual artefacts detected when future and present average precipitation maps from any GCM (GFDL-ESM2M in this example) are subtracted. (This map is presented at 0.5° resolution)

– 22.0°; longitude: 28.0°) where discharge is zero when satellite images show flowing water. Using such data may give a picture of a bad model when the issue is poor input data quality.

While CMRSET generally shows realistic spatial patterns, there are anomalies in some areas that show high ET yet are supposed to have low ET. This is true for the Kalahari Desert area in the Western part of the study area. Comparing the ET product with others for the same area confirms the anomaly in the CMRSET product. The map of long-term average precipitation also shows low precipitation values in the same area. Without much irrigation activity (Fig. 1b), we can conclude that the spatial representation of ET for the Kalahari desert in CMRSET data is poor. This could also give an impression of poor performance in the model.

Since different ET products have different spatial patterns, temporal trends and long-term magnitudes, it is vital to check ET products used in the HMBC against long-term precipitation (P_{long}) subtracted by long-term discharge (Q_{long}), i.e. $P_{\text{long}} - Q_{\text{long}}$ (Weerasinghe et al. 2020).

Simulated discharge improved (Fig. 5) after implementing irrigation and reservoirs with generic decision tables listed in Table S2. Thus, accounting for management practices is beneficial to model performance improvement. However, insufficient data on the management of the reservoirs and irrigation and all the reservoirs that were not implemented (Fig. 1c) limit possible improvements in the model performance for gauging stations downstream of reservoirs.

4.2 Hydrological mass balance calibration

Krysanova et al. (2018) suggest DSS calibration as a second requirement aimed at conducting an intensive gauge-based calibration/validation. Gauge-based automatic calibration is time-consuming, data-intensive and computationally demanding even for small watersheds. For a detailed regional- or continental-scale hydrological model, this kind of calibration is often not feasible. As a result, almost all global-scale hydrological models are not calibrated (Zaherpour et al. 2018). This results in bad model performance, and hence potentially unreliable projections as pointed out by Krysanova et al. (2018).

In this study, we propose hydrological mass balance calibration (HMBC) which is less computationally demanding than usual streamflow-based calibration. In our case, HMBC required 4 to 5 iterations per region to calibrate one water balance component such as ET (Fig. S10).

The HMBC method also is applicable in data-scarce regions. In this study, there were subcatchments with streams having no gauging stations that otherwise could not be calibrated. With HMBC, only a long-term average for each water balance component is needed to do the calibration. While we used the WETSPRO tool to estimate long-term surface runoff and mainly satellite data to estimate long-term ET for each region, these estimates are often available in the literature from previous studies, and the water balance calibration method makes it possible to incorporate this data in the calibration procedure.

Moreover, in gauge-based automatic calibration, a high performance indicator value can be obtained even with unrealistic water balance components and without internal consistency of the simulated processes. Such models are often overfitted (Viney et al. 2009) and may not represent the dynamics of the watershed correctly. With HMBC, the hydrological mass balance changed towards the objective ratios derived from long-term data (Table S3), thus, increasing the potential for a better representation of water balance components. In this study, it was found that by calibrating the mass balance, performance indicators for discharge also improved. The HMBC further enhanced model performance for discharge after irrigation and reservoirs were implemented. Thus, representing anthropogenic activities in hydrological models can improve the performance of models calibrated by hydrological mass balance. This is significant because the performance of the model in the historical period plays a significant role in the credibility of projections. In their study, Krysanova et al. (2018) found that the results of impact assessment for calibrated and uncalibrated models were not comparable, and they argue that the change pattern suggested by the calibrated models is, probably, more trustworthy.

In addition to the improvement of model performance, the low computational cost, and better representation of hydrological processes in the calibrated model make the hydrological mass balance approach suitable for regional, continental and global hydrological models.

4.3 Historical performance for ET and discharge

Krysanova et al. (2018) have proposed that multiple variables of models should be evaluated for the historical period in multiple sites to improve the applicability of the model for impact studies and increase credibility and acceptance of projection results. We evaluated the ET for the historical period with EWEMBI as weather forcing. OBS_CRI is underestimating ET when compared to CMRSET and has similar magnitudes to WaPOR ET which was used in HMBC. Although magnitudes are not very comparable between SWAT+ simulations and CMRSET, all SWAT+ ET simulations follow the historical annual trends of CMRSET for the period with available evaluation data (Fig. 5e). Nevertheless, the modelled ET was more dynamic in magnitudes than both CMRSET and WaPOR.

The increase in ET values from OBS_D to OBS_DRI can be attributed to the irrigation-induced increase in evaporative fraction (Hirsch et al. 2017; Thiery et al. 2017, 2020) while the increase of ET from OBS_DRI to OBS_CRI is attributed to HMBC.

Model performance in simulating discharge varies in space. Streams which have controlled reservoirs have lower performance than those that do not have (Fig. 3). However, we see an improvement in performance after performing the HMBC and implementing reservoirs and irrigation with 72% of all stations having $NSE > 0$ and 45% of the stations with $NSE > 0.5$. PBIAS for discharge also improved with the HMBC (Fig. 5).

4.4 Applicability of the model

Each model should be evaluated in terms of how it can reproduce variables on which the case study focuses. In this study, we evaluated the model for discharge at monthly timestep between 1979 and 1986 and found satisfactory results. The simulation of ET was better in terms of long-term average magnitudes than yearly variations despite having a strong correlation with observations. Therefore, the model can be used in studies to assess the impacts of climate change on water resources on decadal time scales. For studies that want to look at extremes, a combination of HMBC and a more detailed calibration of high and low discharge might be needed. This combination needs to be explored in future research.

4.5 Influence of HMBC on climate change projections

Both the default (FUT_D) and the regionalised (FUT_CRI) models project a decrease of ET in the future climate. However, the default model has lower baseline ET values compared to the regionalised model. The difference indicates that calibrated large-scale models will affect the magnitudes of results even if the climate change signal remains the same. While the climate change signal for the study area as a whole remained consistent for both the default and the regionalised models, the signals were different in space (Fig. 7d). Thus, it makes a difference to calibrate models for impact studies if results are interpreted and used at the local scale. We also discovered anomalies in the map of ET differences between FUT_CRI and FUT_D. These differences are linked to anomalies in the projected changes in precipitation from the GCMs, which can, in turn, be attributed to the methods used to downscale and bias correct the forcing products (Lange 2016).

In the case of surface runoff, there were no apparent differences between the average future climate projections from the uncalibrated and calibrated simulations. However, a look at river discharges revealed that projections made by the calibrated model differ in space from those made by the uncalibrated model (Fig. 7a and b). Although climate change signals are similar in many stations for both models, the magnitude of the differences presented in Fig. 7c illustrates that it matters whether large-scale models are calibrated especially if they are used to make decisions at smaller catchment level or are used to map the results spatially.

4.6 Limitations

The limitations of this modelling study are mostly related to data: (a) non-overlapping spans of available data from different gauging stations, some with poor quality and substantial periods with missing data; (b) unavailability of reservoir management data; (c) poor resolution of input data including land use, soil and weather data; (d) limited information on agricultural management practices. There were also limitations in the model configuration. For instance, revision 55 of SWAT+ does not allow the user to specify an irrigation source. This assumes that irrigation water comes from a deep aquifer in all irrigation sites, which is unrealistic in the study area.

5 Conclusions

In this study, we built a regional hydrological model using SWAT+ for climate projections where we implemented irrigation, reservoirs and hydrological mass balance calibration (HMBC). We have demonstrated that HMBC can be applied to large-scale hydrological models without a significant simulation time cost as it generally requires less than ten iterations per region. Model performance was improved by implementing irrigation, reservoirs and calibration using an HMBC algorithm written in Python. We thus find that HMBC can remedy some barriers to the calibration of hydrological models at the large regional scale. Furthermore, the HMBC method ensures a better representation of processes in the model compared to standard flow-based calibration techniques.

This study also provides practical steps in setting up, calibrating, representing human interactions and evaluating a large-scale hydrological model using SWAT+, which can be applied at the continental and global scale. In our study, the HMBC had more effect on model performance than the implementation of human interactions. This is probably related to the fact that the terrestrial water cycle in Southern Africa remains relatively undisturbed.

End-of-century projections with the regionalised model under Representative Concentration Pathway 6.0 showed higher absolute ET values on average compared to the default model for future climate. Overall change in ET due to climate change (CC) was similar for both models, but the signal of the change in ET was different in some areas. For surface runoff, overall change due to CC was also similar for both models but there were substantial differences in annual average river discharge projections for CC between the regionalised and the default model. Thus, HMBC greatly influences model projections at local scale but has less effect on the overall long-term average magnitude of changes due to CC for the region. HMBC can be considered a valuable tool for impact studies and can be applied in large-scale hydrological models where complete automatic calibration is not feasible.

Funding The study was financially supported by the VLIR-UOS JOINT project “Global Water Academic Network” (TZ2019JOI022A105).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alcamo J et al (2003) Development and testing of the WaterGAP 2 global model of water use and availability. *Hydrol Sci J* 48(3):317–338. <https://doi.org/10.1623/hysj.48.3.317.45290>
- Arnell NW (1999) Climate change and global water resources. *Glob Environ Chang* 9:S31–S49. [https://doi.org/10.1016/S0959-3780\(99\)00017-5](https://doi.org/10.1016/S0959-3780(99)00017-5)
- Arnold JG et al (2012) Swat: model use, calibration, and validation. *Asabe* 55(4):1491–1508 Doi: ISSN 2151-0032
- Arnold JG et al (2015) Hydrological processes and model representation: impact of soft data on calibration. *Trans ASABE* 58(6):1637–1660. <https://doi.org/10.13031/trans.58.10726>
- Arnold JG et al (2018) Use of decision tables to simulate management in SWAT+. *Water (Switzerland)* 10(6):1–10. <https://doi.org/10.3390/w10060713>
- Ayana EK et al (2019) Dividends in flow prediction improvement using high-resolution soil database. *J Hydrol: Reg Stud. Elsevier* 21(2018):159–175. <https://doi.org/10.1016/j.ejrh.2019.01.003>
- Bastiaanssen WGM et al (2012) Surface energy balance and actual evapotranspiration of the transboundary Indus Basin estimated from satellite measurements and the ETLook model. *Water Resour Res* 48(11):1–16. <https://doi.org/10.1029/2011WR010482>
- Bieger K et al (2017) Introduction to SWAT+, a completely restructured version of the Soil and Water Assessment Tool. *J Am Water Resour Assoc* 53(1):115–130. <https://doi.org/10.1111/1752-1688.12482>
- Blatchford M et al (2020) Evaluation of WaPOR V2 evapotranspiration products across Africa. *Hydrol Process.* <https://doi.org/10.1002/hyp.13791>
- Borah DK, Bera M (2004) Watershed-scale. *Hydrol Nonpoint-Source Pollut* 47(3):789–804
- Brunner L et al (2019) Quantifying uncertainty in European climate projections using combined performance-independence weighting. *Environ Res Lett. IOP publishing* 14(12):124010. <https://doi.org/10.1088/1748-9326/ab492f>
- Dile Y, Srinivasan R, & George C (2017) QGIS Interface for SWAT (QSWAT). February(v 1.4), 77. <https://doi.org/10.13140/RG.2.1.1060.7201>
- Eyring V et al (2019) Taking climate model evaluation to the next level. *Nat Clim Chang* 9(2):102–110. <https://doi.org/10.1038/s41558-018-0355-y>
- FAO (2018) WaPOR database methodology: level 1. Remote sensing for water productivity technical report. Rome, Italy. Retrieved June 20, 2019. Available at: <http://www.fao.org/3/i7315en/i7315en.pdf>
- Farr TG et al (2007) The shuttle radar topography mission. *Rev Geophys* 45(2):RG2004. <https://doi.org/10.1029/2005RG000183>
- Gerten D et al (2004) Terrestrial vegetation and water balance—hydrological evaluation of a dynamic global vegetation model. *J Hydrol* 286(1–4):249–270. <https://doi.org/10.1016/j.jhydrol.2003.09.029>
- Goderniaux P et al (2009) Large scale surface-subsurface hydrological model to assess climate change impacts on groundwater reserves. *J Hydrol. Elsevier B.V.* 373(1–2):122–138. <https://doi.org/10.1016/j.jhydrol.2009.04.017>
- Gosling SN, Arnell NW (2011) Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis. *Hydrol Process* 25(7):1129–1145. <https://doi.org/10.1002/hyp.7727>
- GRDC. (n.d.). BFG—the GRDC. Retrieved June 7, 2019, from <http://grdc.bafg.de>
- Van Griensven A et al (2012) Critical review of SWAT applications in the upper Nile basin countries. *Hydrol Earth Syst Sci* 16(9):3371–3381. <https://doi.org/10.5194/hess-16-3371-2012>
- Guerschman JP et al (2009) Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia. *J Hydrol. Elsevier B.V.* 369(1–2):107–119. <https://doi.org/10.1016/j.jhydrol.2009.02.013>

- Hattermann FF et al (2017) Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins. *Clim Chang* 141(3):561–576. <https://doi.org/10.1007/s10584-016-1829-4>
- Hengl T et al (2015) Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLoS One* 10(6):1–26. <https://doi.org/10.1371/journal.pone.0125814>
- Hirsch AL et al (2017) Can climate-effective land management reduce regional warming? *J Geophys Res* 122(4): 2269–2288. <https://doi.org/10.1002/2016JD026125>
- Hurtt GC et al (2011) Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Clim Chang* 109(1):117–161. <https://doi.org/10.1007/s10584-011-0153-2>
- IPCC (2014) *Climate Change 2013 - The Physical Science Basis* (Intergovernmental Panel on Climate Change (ed)) Cambridge University Press, Cambridge
- Kaspar F, Lehner B, Döll P (2003) A global hydrological model for deriving water availability indicators: model tuning and validation. *J Hydrol* 270(1–2):105–134. [https://doi.org/10.1016/S0022-1694\(02\)00283-4](https://doi.org/10.1016/S0022-1694(02)00283-4)
- Kim U, Kaluarachchi JJ, Smakhtin VU (2008) Climate change impacts on hydrology and water resources of the upper Blue Nile River basin, Ethiopia. *Water Manag* 45(6):27
- Kingston DG, Taylor RG (2010) Sources of uncertainty in climate change impacts on river discharge and groundwater in a headwater catchment of the upper Nile Basin, Uganda. *Hydrol Earth Syst Sci* 14(7):1297–1308. <https://doi.org/10.5194/hess-14-1297-2010>
- Koch H, Liersch S, Hattermann FF (2013) Integrating water resources management in eco-hydrological modelling. *Water Sci Technol*. IWA Publishing 67(7):1525–1533
- Krysanova V et al (2018) How the performance of hydrological models relates to credibility of projections under climate change. *Hydrol Sci J*. Taylor & Francis 63(5):696–720. <https://doi.org/10.1080/02626667.2018.1446214>
- Krysanova V, Arnold JG (2008) Advances in ecohydrological modelling with SWAT—a review. *Hydrol Sci J* 53(5):939–947. <https://doi.org/10.1623/hysj.53.5.939>
- Krysanova V, Srinivasan R (2014) Assessment of climate and land use change impacts with SWAT. *Reg Environ Chang* 15(3):431–434. <https://doi.org/10.1007/s10113-014-0742-5>
- Lange S (2016) Earth2Observe, WFDEI and ERA-interim data Merged and Bias-corrected for ISIMIP (EWEMBI). GFZ Data Services. <https://doi.org/10.5880/PIK.2016.004>
- Lehner B et al (2011) High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Front Ecol Environ* 9 (9):494–502
- Lindström G et al (1997) Development and test of the distributed HBV-96 hydrological model. *J Hydrol* 201(1): 272–288. [https://doi.org/10.1016/S0022-1694\(97\)00041-3](https://doi.org/10.1016/S0022-1694(97)00041-3)
- Lorenz R et al (2018) Prospects and caveats of weighting climate models for summer maximum temperature projections over North America. *J Geophys Res-Atmos* 123(9):4509–4526. <https://doi.org/10.1029/2017JD027992>
- Moriasi ND et al (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transact ASABE St Joseph, MI: ASABE* 50(3):885–900. <https://doi.org/10.13031/2013.23153>
- Saxton KE, Rawls WJ (2006) Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Sci Soc Am J* 70(5):1569–1578. <https://doi.org/10.2136/sssaj2005.0117>
- Schuol J et al (2008) Modeling blue and green water availability in Africa. *Water Resour Res* 44(7):1–18. <https://doi.org/10.1029/2007WR006609>
- Shen Z et al (2015) Identifying non-point source priority management areas in watersheds with multiple functional zones. *Water Res*. Elsevier Ltd 68:563–571. <https://doi.org/10.1016/j.watres.2014.10.034>
- Siebert, S. and Frenken, K. (2014) *Irrigated areas Atlas of African agriculture research & development*, pp. 18–19. Available at: <http://www.fao.org/3/I9258EN/i9258en.pdf>
- Sood A, Smakhtin V (2015) Global hydrological models: a review. *Hydrol Sci J*. Taylor & Francis 60(4):549–565. <https://doi.org/10.1080/02626667.2014.950580>
- Thiery W et al. (2017) Present-day irrigation mitigates heat extremes *J Geophys Res: Atmos*. <https://doi.org/10.1002/2016JD025740>
- Thiery W et al (2020) Warming of hot extremes alleviated by expanding irrigation. *Nat Commun*. Springer US 11(1):1–7. <https://doi.org/10.1038/s41467-019-14075-4>
- Trambauer P et al (2013) A review of continental scale hydrological models and their suitability for drought forecasting in (sub-Saharan) Africa. *Phys Chem Earth*. Elsevier Ltd 66:16–26. <https://doi.org/10.1016/j.pce.2013.07.003>
- Viney, N. R. et al. (2009) The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments, 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation: interfacing modelling and simulation with mathematical and computational sciences, proceedings, (July), pp. 3421–3427

- Vörösmarty CJ et al (2000) Global water resources: vulnerability from climate change and population growth. *Science* 289(5477):284–288. <https://doi.org/10.1126/science.289.5477.284>
- Weerasinghe I et al (2020) Can we trust remote sensing evapotranspiration products over Africa. *Hydrol Earth Syst Sci* 24(3):1565–1586. <https://doi.org/10.5194/hess-24-1565-2020>
- Willems P (2009) A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. *Environ Model Softw*. Elsevier Ltd 24(3):311–321. <https://doi.org/10.1016/j.envsoft.2008.09.005>
- Yang D et al (2010) Global assessment of current water resources using total runoff integrating pathways. *Hydrol Sci J* 46(6):983–995. <https://doi.org/10.1080/02626660109492890>
- Zaherpour J et al (2018) Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. *Environ Res Lett* 13(6). <https://doi.org/10.1088/1748-9326/aac547>
- Zhang Z et al (2008) Evaluation of the MIKE SHE model for application in the Loess Plateau, China. *J Am Water Resour Assoc* 44(5):1108–1120. <https://doi.org/10.1111/j.1752-1688.2008.00244.x>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Celray James Chawanda¹ • Jeffrey Arnold² • Wim Thiery^{1,3} • Ann van Griensven^{1,4}

¹ Vrije Universiteit Brussel, Ixelles, Belgium

² USDA-ARS, Temple, Texas, USA

³ ETH Zurich, Zürich, Switzerland

⁴ IHE Delft, Delft, Netherlands