

Evaluating the impact of rainfall–runoff model structural uncertainty on the hydrological rating of regional climate model simulations

Hamouda Dakhlaoui ^{a,b,*} and Khalil Djebbi^a

^a LMHE, Ecole Nationale d'Ingénieurs de Tunis, Université Tunis El Manar, BP 37, 1002 Tunis Le Belvédère, Tunisia

^b Ecole Nationale d'Architecture et d'Urbanisme, Université de Carthage, Sidi Bou Said, Tunisia

*Corresponding author. E-mail: hammouda.dakhlaoui@laposte.net

 HD, 0000-0002-5271-4350

ABSTRACT

We propose to evaluate the impact of rainfall–runoff model (RRM) structural uncertainty on climate model evaluation, performed within a process-oriented framework using the RRM. Structural uncertainty is assessed with an ensemble approach using three conceptual RRMs (HBV, IHACRES and GR4J). We evaluate daily precipitation and temperature from 11 regional climate models forced by five general circulation models (GCM–RCMs), issued from EURO-CORDEX. The assessment was performed over the reference period (1970–2000) for five catchments situated in northern Tunisia. Seventeen discharge performance indexes were used to explore the representation of hydrological processes. The three RRMs performed well over the reference period, with Nash–Sutcliffe efficiency values ranging from 0.70 to 0.90 and bias close to 0%. The ranking of GCM–RCMs according to hydrological performance indexes is more meaningful before the bias correction, which considerably reduces the differences between GCM- and RCM-driven hydrological simulations. Our results illustrate a strong similarity between the different RRMs in terms of raw GCM–RCM performances over the reference period for the majority of performance indexes, in spite of their different model structures. This proves that the structural uncertainty induced by RRMs does not affect GCM–RCM evaluation and ranking, which contributes to consolidate the RRM as a standard tool for climate model evaluation.

Key words: bias correction, climate model evaluation, EURO-CORDEX, rainfall–runoff modelling, structural uncertainty, Tunisia

HIGHLIGHTS

- Structural uncertainty induced using rainfall–runoff models does not affect climate model evaluation and ranking performed using hydrological modelling.
- Importance of considering a wide set of hydrological performance indexes to evaluate climate model performance.
- Climate models that rank well on high-flow performance indexes rank poorly on low-flow performance indexes and vice versa.

1. INTRODUCTION

The widespread interest in understanding climate change, along with the response and feedback of ecosystems, has driven the creation and application of different models and methods to investigate these dynamics (Jacob *et al.* 2007; Deidda *et al.* 2013; Hakala *et al.* 2019). Numerical simulation of palaeo and future weather (including atmospheric circulation and surface-climate variables) generally needs to be derived from climate models. These models account generally for dynamic shifts in the global atmosphere and oceans, powered by global boundary conditions such as atmospheric trace gases, aerosols, earth–sun dynamics, sea ice, sea level and continental ice sheets. Climate models offer tools for reproducing and anticipating climate. General circulation models (GCMs) are the principal tools used to create projections of future climate with a resolution of around 100–300 km (horizontal grid spacing) (Hakala *et al.* 2019). GCMs' resolution presents a problem when local-scale information is required (Piani *et al.* 2010). Dynamical downscaling (DD) (e.g. Maraun 2013; Jury *et al.* 2015) can be used to translate coarse resolution information to a finer scale. Within DD, a regional climate model (RCM) is used to simulate climate dynamics on the scale of 10–50 km (horizontal grid spacing), where GCM output is used as the boundary condition.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Climate system complexity has made it difficult to devise any model or method of downscaling that can reproduce this system perfectly. In practice, GCM–RCMs are generally selected for impact studies based on their ability to simulate current climate (Jacob *et al.* 2007; Jury *et al.* 2015; Hakala *et al.* 2018). For example, Mendlik & Gobiet (2016) propose the removal of extremely unrealistic models according to their performance under current climate and recommend that an impact study utilize more than just the ‘best-performing’ models because a good performance under current climate does not guarantee the most realistic future projections. Acheng & Zhu (2019) recommend the Bayesian averaging of GCM–RCMs streamflow projection. They suggest that individual GCM–RCMs’ performance under current climate before they are averaged should be consistent with their respective performances in the Bayesian framework.

While several performance indexes to evaluate the climate model were established, there is a lack of a standard performance index or procedure (Hakala *et al.* 2018). The evaluation of climate model performance over a historical period is generally based on climate variables and is generally performed for each climate variable separately (e.g. Gleckler *et al.* 2008; Khan & Koch 2018; Haj-Amor *et al.* 2020). However, the impact of climate change on the water cycle is the result of the interaction between climate variables, which are highly correlated with one another. For example, Dimitriadis *et al.* (2021) who evaluated the stochastic analogies in key hydroclimatic processes related to the hydrological cycle, found several stochastic similarities in the marginal structure and in the second-order dependence structure of hydroclimatic variables. They prove a specific hierarchy among their marginal and dependence structures, similar to the one in the hydrological cycle, which allows for the development of a universal stochastic view of the hydrological cycle under the Hurst–Kolmogorov dynamics (Hurst 1951; Koutsoyiannis *et al.* 2008; O’Connell *et al.* 2016; Markonis *et al.* 2018; Dimitriadis & Koutsoyiannis 2019). Indeed, the separate evaluation of variables simulated by climate models could be a limitation, and, therefore, assessing climate model inter-variable relationships is essential. Climate model inter-variable relationships can be assessed in a statistical framework (e.g. Jury *et al.* 2015). However, the statistical framework is not able to take into consideration the non-linear interaction between climate variables. In addition, with these methods (separate evaluation of climate model outputs and statistical intervariable relationship assessment), it is difficult to establish whether climate model’s performance is satisfactory in terms of predictions of practical interest, such as hydrological projection. To tackle these limitations, a few recent studies used a more process-based investigation of climate model simulations based on hydrological modelling (Foughali *et al.* 2015; Hakala *et al.* 2018; Santos *et al.* 2019).

In effect, the hydrological models are generally able to capture the interactions between temperature and precipitation, at the catchment scale, leading to discharge. In a first illustrative example, showing the interactions between temperature and precipitation into hydrological models, threshold air temperature is generally used to distinguish rainfall (liquid precipitation that flows or infiltrates into soil) from snowfall (that forms snowpack and does not generate immediate runoff). When air temperature increases above a given threshold, the snowpack melts gradually. The timing and the magnitude of discharge resulting from snowmelt is dependent on the feature of air temperature. In a second example, illustrating interactions between temperature and precipitation into the RRM, the surface runoff genesis depends mainly on both rainfall intensity and antecedent soil moisture. Antecedent soil moisture, in turn, depends on previous precipitation and air temperature conditions (for potential evapotranspiration estimation). Furthermore, RRM performances have shown to be sensitive to change in the pattern of climate forcing input (Merz *et al.* 2011; Dakhloui *et al.* 2017). Indeed, RRM parameters have proven to be climate-dependent (i.e. influenced by climate conditions), pointing out that RRM parameters could represent a whole system including both the receiver environment and the forcing climate, as revealed by several studies (Ruelland *et al.*, 2009; Deshmukh & Singh 2019; Stephens *et al.* 2019).

While RRM have been widely used to evaluate simulations from weather generators (e.g. Li *et al.* 2013; Breinl 2016), the stochastic rainfall model (e.g. McMillan & Brasington 2008; Bennett *et al.* 2019) and gridded precipitation datasets (from different sources such as radar, gauge, satellite, analysis, or reanalysis, or combinations thereof) (e.g. Beck *et al.* 2017), their use to evaluate GCM–RCM simulations is still limited. Foughali *et al.* (2015) evaluated four GCM–RCM outputs from the ENSEMBLE project (Jacob *et al.* 2007) in the Sejnane catchment located in northern Tunisia, using a bucket with a bottom hole water balance model, over a 12-year reference period (1972–1984). They found that the predicted monthly runoff regime from the water balance model forced by the bias-corrected climate series reflects more or less the observed regime, depending on the climate model. Hakala *et al.* (2018) used Hydrologiska Byråns Vattenbalansavdelning (HBV) RRM to investigate GCM–RCM simulations from EURO-CORDEX over eight Swiss catchments using seven performance indexes evaluating simulated discharge. These performance indexes serve to rate climate models and test the quantile mapping (QM) effect on the simulated discharge. Hakala *et al.* (2018) prove that the equifinality of HBV parameters (i.e.

non-uniqueness of parameters, [Beven 2006](#)) does not affect climate model ratings by the RRM. [Santos et al. \(2019\)](#) used the Soil and Water Assessment Tool (SWAT) to evaluate precipitation and surface air temperature projections from two regional climate models (Eta-HadGEM2-ES and Eta-MIROC5) over the Paraguaçu River Basin. They evaluated two bias correction methods (linear scaling and distribution mapping) using five performance indexes evaluating the simulated discharge. They found that the distribution mapping yielded better results for hydrological extreme simulations.

The use of RRM could be an additional source of uncertainty in the assessment of GCM–RCM performance. The usual sources of uncertainty in hydrological modelling are linked to the structure of the models, the calibration procedures, erroneous data used for calibration/validation and parameter instability (e.g. [Beck 1987](#); [Beven 2006](#); [Coron et al. 2012](#); [Brigode et al. 2013](#); [Poissant et al. 2017](#); [Adeyeri et al. 2020c](#); [Ben Jaafar & Bargaoui 2020](#)). However, to our knowledge, except [Hakala et al. \(2018\)](#), who investigated the effect of equifinality of RRM parameters on GCM–RCM rating, there is no study that investigated the effect of hydrological modelling uncertainties on GCM–RCM evaluation performed by the RRM. This is especially the case for structural uncertainty.

RRM structural uncertainty can be defined in terms of inadequacy and non-uniqueness ([Hublart et al. 2015](#)). Model inadequacy refers to the simplifying assumptions in a given model that fail to represent accurately the real system it is intended to characterize. Non-uniqueness arises from the existence of different model structures performing equally to a given observed data ([Hublart et al. 2015](#)). Evaluating structural RRM uncertainty is generally based on an ensemble approach, in which models of different complexities are tested to characterize their output range (e.g. [Hublart et al. 2015](#); [Seiller et al. 2017](#)). However, the few studies that proposed a process-based evaluation of GCM–RCM simulations using RRM have used only one hydrological model ([Foughali et al. 2015](#); [Hakala et al. 2018](#); [Santos et al. 2019](#)), which shows the need for more investigation of the effect of RRM structural uncertainty on GCM–RCM evaluation.

The North Africa region, including Tunisia, has been identified as a hot spot of climate change ([Fader et al. 2020](#)). Recent climate projections guess that there will be a substantial decrease in precipitation (around 20%) and an increase in mean annual temperature (around +1 °C to +3 °C) by the 2050 horizon compared with the 1971–1990 period ([Schilling et al. 2012](#); [Milano et al. 2013](#)). This could cause severe water stress in the future ([Hachani et al. 2017](#); [Dakhlaoui et al. 2019b](#)). The climate change impact studies need reliable climate models, and several studies report that the reliability of climate models could vary among regions (e.g. [Khan & Koch 2018](#)). However, the literature review of studies that investigated the hydroclimatic impacts of climate change in Tunisia shows that they either have not provided any evaluation of climate models over the reference period (e.g. [Nasr et al. 2008](#); [Allani et al. 2019](#)) or have performed an evaluation of each climate variable separately (e.g. [Deidda et al. 2013](#); [Bargaoui et al. 2014](#); [Sellami et al. 2015, 2016](#); [Dakhlaoui et al. 2019a](#); [Haj-Amor et al. 2020](#); [Adham et al. 2019](#)). The unique study that proposed a climate model evaluation in Tunisia, taking into account climate variable interactions (into a process-based framework using the RRM) ([Foughali et al. 2015](#)), was performed for the previous generation of climate models (ENSEMBLE, [Jacob et al. 2007](#)) and was based on only one single catchment and the RRM. This shows the need for evaluating recent climate model simulations in this part of the world, especially with a process-based investigation considering the interaction between climate variables.

In light of the above, the aim of the present study was to assess whether RRM structural uncertainty affects the evaluation and ranking of GCM–RCMs, performed in a process-oriented evaluation framework based on hydrological models. To achieve this overall goal, the paper focused on three specific objectives: (i) to conduct an evaluation of recent high-resolution simulations from GCM–RCMs under the present climate based on their hydrological performance using different RRM and different hydrological performance indexes, (ii) to rank GCM–RCMs based on how well they enabled us to capture hydrological variables and (iii) to check if the ranking of climate models is comparable across different conceptual structures of RRM.

These questions were investigated for the case of five catchments in northern Tunisia. The modelling chain of the process-oriented evaluation framework is composed of 11 RCMs from the European domain (EURO-CORDEX), forced by five GCMs and three conceptual RRM (HBV, GR4 J and IHACRES). An evaluation of the ability of this modelling chain to reproduce the discharge observed is performed over the 1970–2000 reference period.

2. DATA AND METHODS

2.1. Study catchments

Five catchments located in Northern Tunisia with sizes ranging from 80 to 315 km² were selected for the study. The human influence on the discharge regime can be considered negligible since these catchments are located upstream from chief

hydraulic installations (dams or water transfers). The five catchments (Figure 1) are Rhezala, Melah, Maaden, Joumine and El Abid. They are located in the main hydrographical basins (rivers) of Northern Tunisia: Extreme North, Ichkeul, High Medjerdah and Cap Bon, having a strategic role as a surface water supplier for the country (Ben Fraj *et al.* 2019). The climate of the study catchments could be categorized as semi-arid to humid Mediterranean climate with a warm season (Dakhlaoui *et al.* 2017, 2020). The catchment's hydro-climatic characteristics are reported in Figure 1.

2.2. Observational data

The precipitation data are available from 511 daily precipitation gauges situated in or around the study catchments. However, data quality is important for hydrological modelling and could be an important source of uncertainty (Adeyeri *et al.* 2019, 2020a). For this purpose, we use only the 123 precipitation gauges with less than 30% of daily gaps over the period 1970–2000 (see Figure 1). In addition, these gaps are randomly distributed over the whole time period, offering a coherent network of gauges for the spatial interpolation of precipitation in the region. More details about precipitation data quality can be found in Dakhlaoui *et al.* (2017). The monthly mean series of air temperatures, issued from eight synoptic meteorological stations, were used to estimate air temperature and potential evapotranspiration (PET). An inverse distance weighting technique, accounting for a lapse rate of $-0.65\text{ }^{\circ}\text{C}/100\text{ m}$ (Ruelland *et al.* 2014), was used to interpolate temperature. Precipitation was interpolated by the method proposed by Valéry *et al.* (2010) that accounts for altitude via a 4×10^{-4} patch factor. This interpolation method was initially developed to improve RRM performance. It proposes an objective elevation-dependent correction procedure, which helps to improve the estimation of areal precipitation and correct the apparent water balance anomalies of the RRM. The PET was estimated using the Oudin formula (Oudin *et al.* 2005) that relies on clear daily sky solar radiation and mean daily air temperature.

Mean daily discharge data are provided from the five gauge stations situated at the outlets of the study catchments. Several hydrological reports (e.g. Manai 1983) outline the good quality hydrological data for these gauges.

2.3. General circulation model–regional climate model

Daily precipitation and temperature series were extracted from 11 ensemble members from the EURO-CORDEX project (<http://www.euro-cordex.net/>; Table A1, Supplementary Material). The RCM outputs are at a considerably finer scale (0.11° , equivalent to the horizontal grid spacing of $\sim 12.5\text{ km}$) than the forcing GCMs. The period from 1970 to 2000 corresponds to the historical model's simulation. Although northern Tunisia is also located in the domain of Middle East and North Africa (MENA-CORDEX), the domain of Africa (CORDEX-AFRICA) and the domain of the Mediterranean region (MED-CORDEX), we preferred to use EURO-CORDEX because it offers a larger number of climate models and climate scenarios and finer resolution than the other CORDEX domains (www.cordex.org).

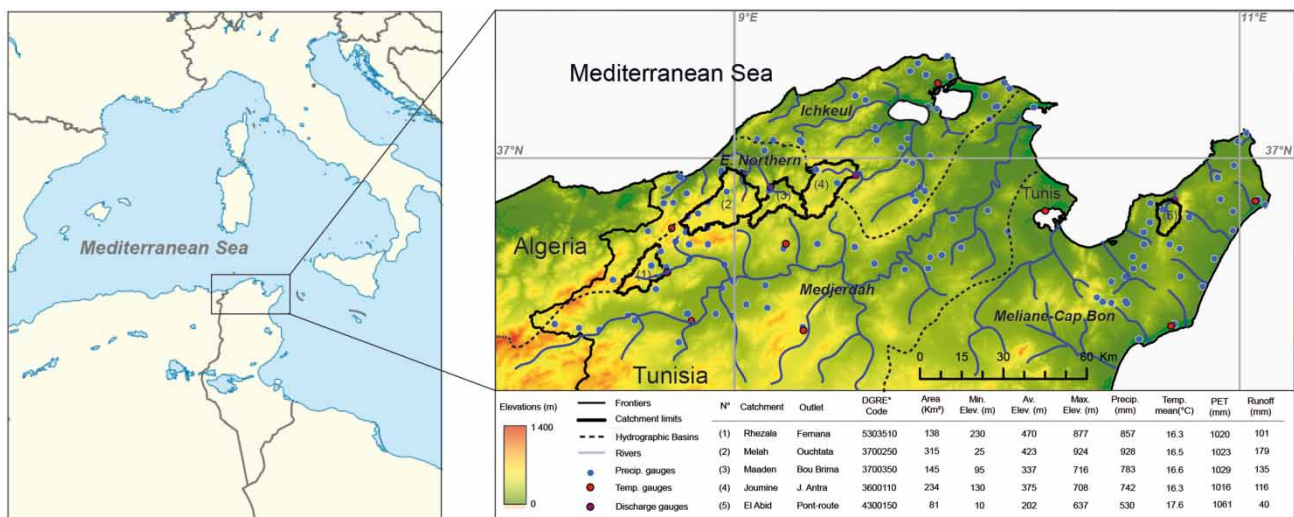


Figure 1 | Locations of study catchments are indicated by a bold black line along with the locations of the precipitation (blue circles), temperature (red circles) and stream flow gauges (purple circles). The main hydro-climatic characteristics are averaged over the period 1970–2000.

The extraction of the GCM–RCM data at the catchment scale was performed using the area-weighted mean of GCM–RCM grids corresponding to the catchment.

2.4. Bias correction

Climate models are not intended to create an isomorphism of nature (Stainforth *et al.* 2007) and are, therefore, not able to reproduce this system perfectly. This implies applying post-processing tools such as statistical adjustment and bias correction before using their outputs in impact studies (Piani *et al.* 2010; Teutschbein & Seibert 2010). The correction of climate model output biases has become a frequent procedure in most recent impact studies (Chen *et al.* 2013). Bias correction methods are used to assist in adjusting some particular aspects of climate models (e.g. spatial, multivariate, temporal, and marginal aspects) (Adeyeri *et al.* 2020b). Popular methods include statistical transformations that aim to adjust the distribution of RCMs or GCMs to observed distributions (Piani *et al.* 2010; Themeßl *et al.* 2011; Teutschbein & Seibert 2012). In the context of hydrological impact studies, previous studies have shown QM to outperform other bias correction methods that correct only the mean or mean and variance of climate series (Teutschbein & Seibert 2012; Chen *et al.* 2013).

We used in the present study QM to bias-correct the daily precipitation and temperature of the GCM–RCMs. QM aims to correct the distribution of the climate model data, so that it matches the distribution of the observational data. It consists of estimating quantiles for both observation and modelled climate variables under a control (calibration) period. A transfer function is then created by interpolation between corresponding quantile values, which is applied to climate variables over the projection or validation period. The empirical cumulative distribution functions (CDFs) of both observed and modelled climate variables were assessed using empirical percentiles for each season. Values within the observation range were approached using linear interpolation. Values outside the range of observations (such as from the projected period) were estimated using a linear regression fit that extrapolates beyond the range of observations.

For the assessment and ranking of the GCM–RCMs, the QM transfer function was inferred over the 30-year control period (1970–2000), which is the time-length recommended for climate applications (WMO 2011) that allow us to capture well the climate variability.

In the proposed methodology of this study, the QM transfer function is not used outside the reference period over which it is calibrated (e.g. future projection period). However, for sanity check of the bias correction method (e.g. Adeyeri *et al.* 2020b), we also propose calibration and validation exercises of QM over the historical period. For this purpose, the historical period was split into calibration (1970–1985) and validation (1986–2000) periods, over which the performance of QM transfer function was evaluated.

2.5. Hydrological models, calibration method and goodness-of-fit criteria

To evaluate the impact of RRM structural uncertainty on GCM–RCM evaluation and ranking, we used an ensemble approach based on three conceptual models, with different levels of complexity, and their output range was characterized. GR4 J, HBV and IHACRES (see Table A2 in Supplementary Material), a global conceptual RRM, running at the daily time step, were considered in this study. These models have different ways to conceptualize the hydrological processes and have four to eight free parameters. The successful applications of these models in northern Tunisia (Bargaoui *et al.* 2008; Dakhlaoui *et al.* 2009, 2012; Abbaris *et al.* 2014) stimulated us to consider them for this study.

The model parameters were calibrated using the Shuffled Complex Evolution (SCE) algorithm (Duan *et al.* 1992), applied to optimize Kling–Gupta efficiency criteria (*KGE*, Gupta *et al.* 2009). *KGE* is the result of the decomposition of Nash and Sutcliffe Efficiency index (*NSE*) (Nash & Sutcliffe 1970). It represents a conciliation between three performance criteria (correlation coefficient, bias error and standard deviation ratio). *KGE* is expressed as follows:

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (1)$$

where r is the linear correlation coefficient between observed and simulated discharges (respectively, q_{oi} and q_{ci}), α is a measure to compare the variability of the observed and simulated values (the standard deviation of q_{ci} divided by the standard deviation of q_{oi}) and β is the bias (the mean of q_{ci} divided by the mean of q_{oi}).

To achieve this calibration exercise efficiently (i.e. with a low processing time to find the optimum) and effectively (i.e. the ability to find the optimum), the algorithmic parameters of SCE were set to values recommended by Duan *et al.* (1994) and Kuczera (1997).

The hydrological models were run at a daily time step. However, their calibration and validation were performed at a 10-day time step.

NSE criterion and the cumulated volume error (*VE*) were also used to evaluate model performance. These criteria are computed according to the following equations:

$$VE = \sum_{i=1}^n (q_{ci} - q_{oi}) / \sum_{i=1}^n q_{oi} \quad (2)$$

$$NSE = 1 - \left(\frac{\sum_{i=1}^n (q_{ci} - q_{oi})^2}{\sum_{i=1}^n (q_{oi} - \bar{q}_o)^2} \right) \quad (3)$$

where \bar{q}_o is the mean observed runoff.

The RRM models were calibrated and validated over the periods 1970–1985 and 1986–2000, respectively, and also calibrated over the whole observed period (1970–2000).

2.6. Performance indexes for climate model performance evaluation

An evaluation of the climate models over the reference period was performed via the assessment of the hydrological simulation of the three RRM models forced by precipitation and temperature data from GCM–RCMs. Following a bibliographical review, a wide variety of hydrological performance indexes were selected for this assessment. They are based on metrics that characterize different aspects of the hydrograph, and they are sensitive to various hydroclimatic processes (Addor *et al.* 2018; Hakala *et al.* 2018; Gudmundsson *et al.* 2018). The selected metrics describe discharge distribution, dynamic, seasonality and timing. They treat water balance and high, low and medium flows. These hydrological metrics include:

- *Qmean*: mean annual flow
- *Q5*, *Q25*, *Q50*, *Q75* and *Q95*: flow exceedance percentiles from the CDF
- *Qmax*: annual maximum flow
- *Qmin*: annual minimum flow
- *STD*: standard deviation of flows
- *HFD*: half flow date, corresponding to the day of the year when half the annual discharge has been measured
- *QmaxDate*: date corresponding to annual maximum flow
- *QminDate*: date corresponding to annual minimum flow
- *SON*, *DJF*, *MAM*, *JJA*: respectively, Autumn, Winter, Spring and Summer mean flow

These hydrological metrics were calculated using 10-day discharges (36 values per year). After a metric was calculated for each individual year, the median of those values was then used. Median metrics, *Mref*, *Mraw* and *Mqm*, were calculated, respectively, for:

- *Qref*: discharge simulated by the RRM forced by observed climate data
- *Qraw*: discharge simulated by the RRM forced by raw GCM–RCM climate data
- *Qqm*: discharge simulated by the RRM forced by quantile-mapped GCM–RCM climate data

The performance hydrological indexes (*Eraw* and *Eqm*) were obtained by calculating the relative errors comparing *Mref* with *Mraw* and *Mqm*:

$$Eraw = (Mraw - Mref) / (Mref + \varepsilon) \quad (4)$$

$$Eqm = (Mqm - Mref) / (Mref + \varepsilon) \quad (5)$$

where ε was set to 0.1.

An explanation for the addition of ε is that the relative error is very sensitive to low-flow metrics where small discharge values are compared with one another. ε avoids dividing error by *Mref* close to zero, where small differences can result in large relative errors. For the *HFD*, *QmaxDate* and *QminDate* metrics, the difference was used (by calculating *Mraw* – *Mref* and *Mqm* – *Mref*). In addition, we calculated $1 - NSE$ applied to the long-term means of 10-day discharges (36 values per year), which allowed us to evaluate the seasonality reconstitution of discharge.

Both raw and quantile-mapped GCM–RCMs can then be ranked according to the performance of runoff simulations.

3. RESULTS AND DISCUSSION

3.1. Calibration and validation of RRM

The RRM forced by observed climatic data show satisfactory efficiency for the calibration period (1970–1985), with *NSE* values ranging from 0.62 to 0.87, *VE* close to 0% and less than 9%, and *KGE* between 0.69 and 0.96, for the five study catchments (Table A3 in Supplementary Material).

Th results show that the RRM were very robust. In effect, a satisfactory model efficiency was obtained for the validation period (1986–2000), with *NSE* values ranging from 0.54 to 0.83, *VE* less than 18%, and *KGE* between 0.58 and 0.90, for the five study catchments (Table A3 in Supplementary Material).

Satisfactory RRM efficiency was also obtained for calibration on the whole reference period (1970–2000), with *NSE* values ranging from 0.70 to 0.90, *VE* close to 0% and less than 2%, and *KGE* between 0.76 and 0.90, for the five study catchments (Table 1). The performance of all three models was generally similar for all catchments. Series of the mean seasonal discharge covering the whole period from the HBV model calibrated over the whole period (1970–2000) forced by observed climatic data are shown in Figure 2(c) for J. Antra and Rhezala catchments and compared with the observed discharges (the results obtained from all catchments and RRM are given in Supplementary Material, Fig A3). The spread of the hydrological simulations over the reference period shows that seasonal patterns of runoff are all well captured for all RRM over all the catchments.

Yapo *et al.* (1996) reported that 8 years of observed data are sufficient for the calibration of RRM in Southeastern USA. However, Vaze *et al.* (2010) noted that at least 20 years are needed for the calibration of RRM in South Australia. Several studies showed that a long calibration period gives a more robust parameter than a short period (Motavita *et al.* 2019; Ben Jaaffer & Bargaoui 2020). This is especially true for a semi-arid climate, characterized by a strong interannual variability, where longer periods of calibration are needed to capture the whole range of variability that can be observed (Tramblay *et al.* 2013). Therefore, to evaluate RCM simulations, RRM were run using the set of parameters obtained by calibration during the whole 30-year observed period (1970–2000).

3.2. Bias correction of GCM-RCM simulations

3.2.1. Calibration and validation of the quantile mapping method

We found a substantial improvement in the representation of climate variables after bias correction throughout the annual cycle for both calibration (1970–1985) and validation (1986–2000) periods (Figures A1 and A2 in Supplementary Material). The bias correction led to an ensemble average that was much closer to the observed reference precipitation and temperature for both calibration and validation periods, which proves the validity and robustness of the QM method.

3.2.2. Bias correction of GCM-RCMs over the whole reference period

In Figures 2(a) and 2(b), we compare raw climate model simulations with observations over the whole reference period (1970–2000) for the J. Antra and Rhezala catchments (the results obtained from all catchments are provided in Supplementary Material, Fig A3). The 11 raw GCM-RCMs underestimate rainfall during the rainy months and overestimate it during dry seasons. In addition, these models show different precipitation seasonalities than the observed one. The median of the 11 GCM-RCMs presents a long wet season (from September to May) and a short dry season (from June to August). For the temperature, the 11 GCM-RCMs reproduce the seasonality well, but there is still a certain lag compared with the observed temperature (Figure 2(b)).

Table 1 | Performance of rainfall-runoff models calibrated over the whole reference period (1970–2000) using the observed climatic data for the five study catchments

Criteria RRM/catchment	KGE			NSE			VE		
	HBV	IHACRES	GR4J	HBV	IHACRES	GR4J	HBV	IHACRES	GR4J
Abid	0.89	0.76	0.78	0.78	0.74	0.81	−0.1%	8.1%	6.7%
Rhezala	0.88	0.86	0.91	0.77	0.72	0.82	2.0%	2.0%	−1.9%
Melah	0.85	0.83	0.83	0.70	0.71	0.70	0.2%	4.6%	3.1%
J. Antra	0.89	0.88	0.88	0.78	0.77	0.79	−0.3%	−0.8%	2.0%
Maaden	0.88	0.87	0.88	0.76	0.74	0.76	−0.2%	0.9%	−0.3%

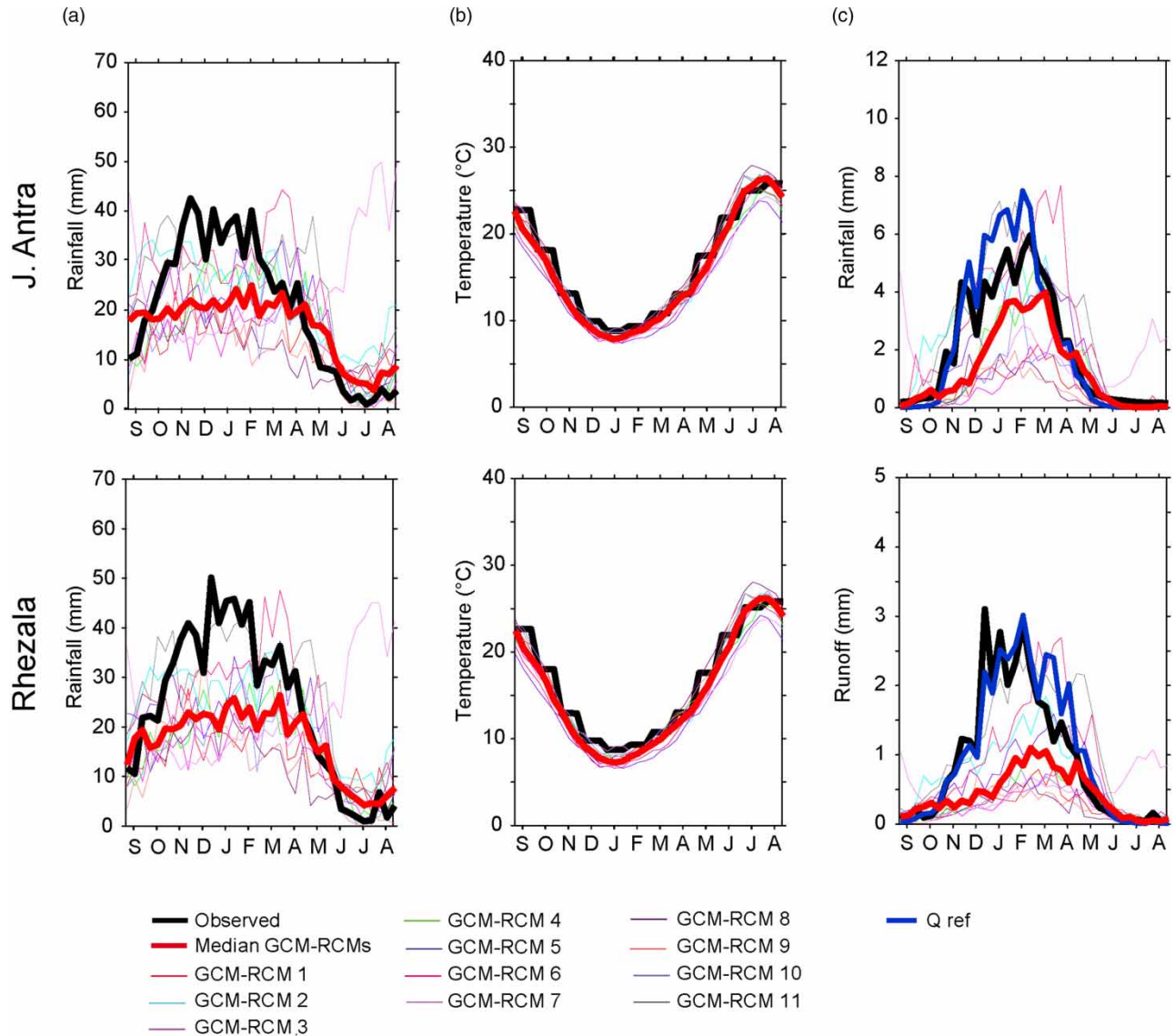


Figure 2 | The long-term mean decadal precipitation (a), temperature (b), and discharge (c) for two example catchments (J. Antra and Rhezala). The precipitation and temperature data from the 11 raw GCM-RCMs are compared with the observed ones. The simulated discharge by the HBV model forced by the 11 raw GCM-RCMs is compared with the simulated discharge by the same RRM model forced by observed climatic data (*Qref*) and with the observed discharge.

In Figures 4(a) and 4(b), we compare bias-corrected GCM-RCM precipitation and temperature with the observed ones over the reference period (1970–2000) for the J. Antra and Rhezala catchments (the results obtained from all catchments and RRM are given in Supplementary Material, Fig A4). We find an improvement in the representation of climate variables after bias correction throughout the annual cycle (Figure 4(a) and 4(b)). The bias correction led to an ensemble average that was much closer to reference precipitation and temperature.

3.3. Evaluation of raw GCM-RCMs by hydrological modelling

In Figure 2(c), we compare simulated runoff by the HBV model, forced by the raw climate model, with the reference discharge (discharge simulated by the RRM forced by observed climate data) for the J. Antra and Rhezala catchments (the results obtained from all catchments and RRM are given in Supplementary Material, Fig A3). It seems that the bias in climate variables was transmitted to runoff.

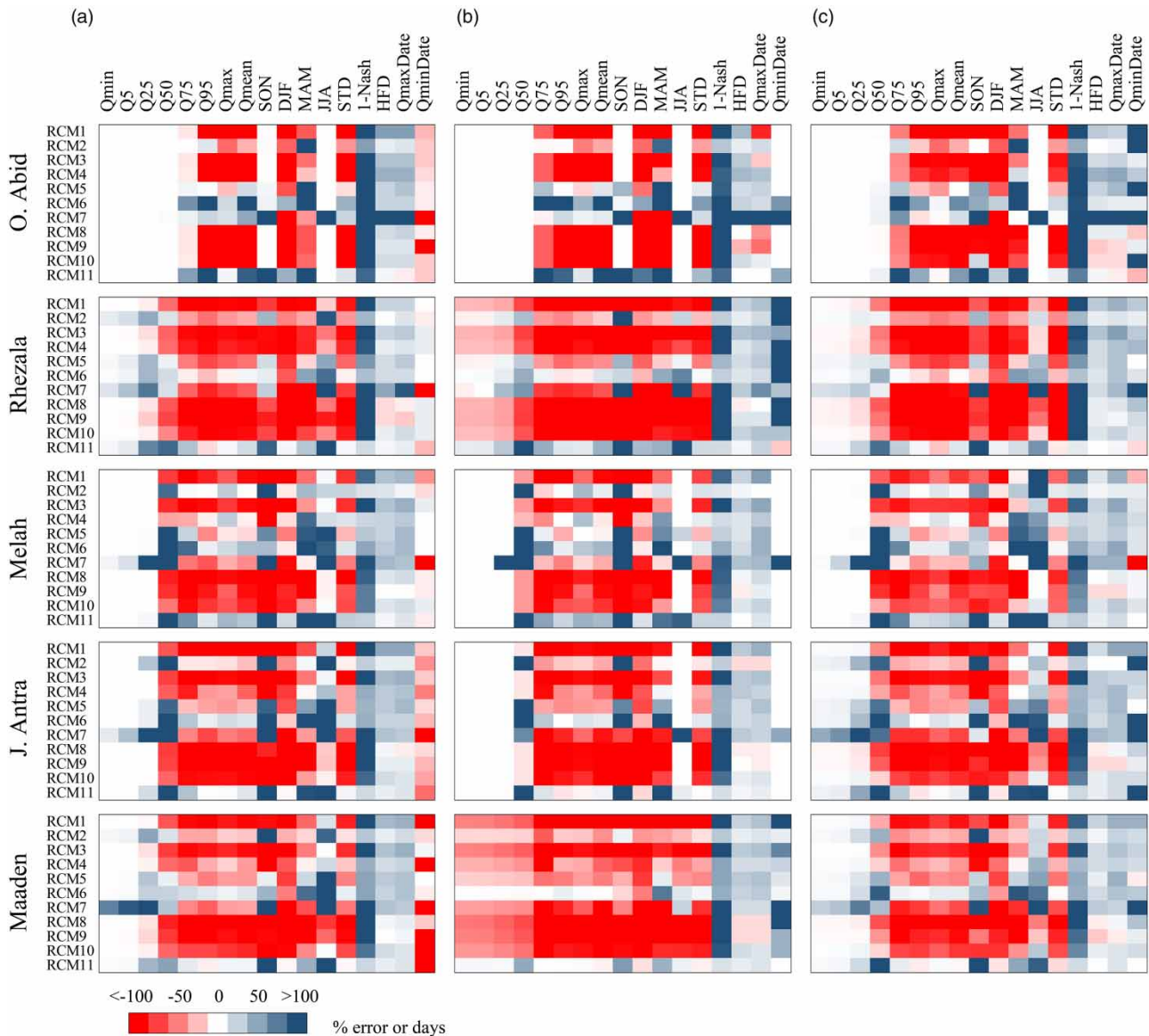


Figure 3 | The range in the performance of the raw GCM-RCMs (y-axis) according to hydrological performance indexes (x-axis) using three hydrological models: (a) HBV, (b) IHACRES and (c) GR4 for the five study catchments. The colours in the heat maps illustrate the values of error of raw GCM-RCMs with values in units of percent error, except *HFD*, *QmaxDate* and *QminDate*, which are in the units of days.

Figure 3 shows the hydrological performance indexes for climate model evaluation applied to the 11 raw climate models and three RRM. Our results show that performance indexes related to seasonality were the hardest to reconstitute. In effect, the high values of $1 - NSE$, except those of GCM-RCM2 and 11 (CNRM-SMHI and MPI-SMHI), reflect the difficulty of reconstituting seasonality with raw climate models. The climate models generally underestimate the standard deviation of 10-day discharges, which shows that it is difficult to ascertain the intra-seasonal variability of discharge with these climate models; instead, a more monotonous discharge is presented by these models (Figure 2(c)). The climate models generally present a dry bias in *Qmean*. In addition, performance indexes related to high flows were also hard to reconstitute. In effect, raw climate models generally underestimate the mean flow of the wet season (*DJF*, *Q95* and *Qmax*). However, performance indexes related to low flows (*Qmin*, *Q5* and *JJA*) were the easiest to simulate by the majority of GCM-RCMs (Figure 3). These performance indexes have generally minimum bias, but if they present bias, it was generally wet bias.

Our results also show that raw climate models generally underestimate the mean flow of autumn and spring (*MMA* and *SON*).

When comparing date performance indexes, we find an overestimation of *HFD* and *QmaxDate*, generally about one to two decades, which reflects a delay in runoff by the climate models. We can point to a great similarity between RRM for these

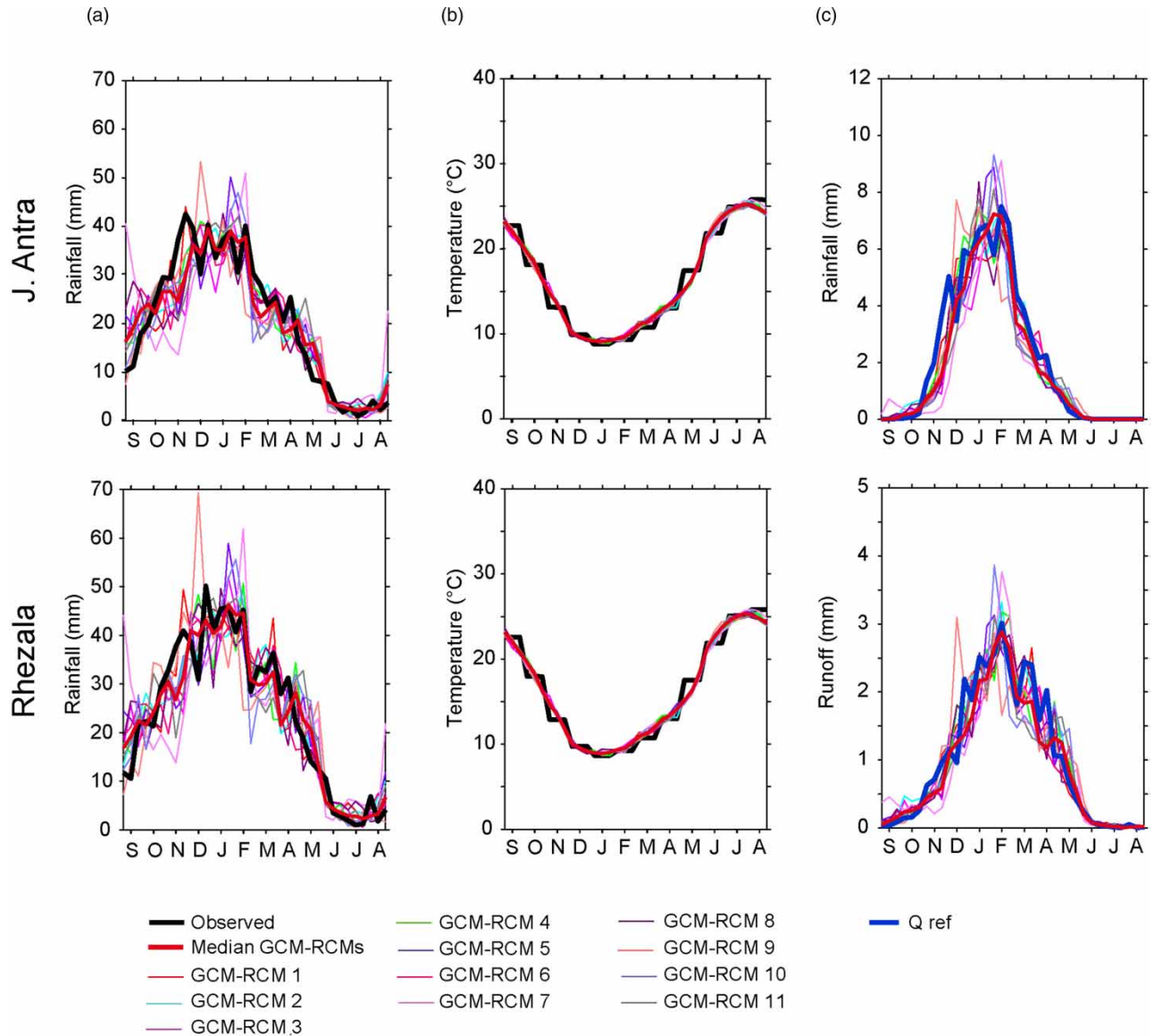


Figure 4 | The long-term mean decadal precipitation (a), temperature (b), and discharge (c) for two example catchments (J. Antra and Rhezala). The precipitation and temperature data from the 11 bias-corrected GCM-RCMs are compared with the observed ones. The simulated discharge by the HBV model forced by the 11 bias-corrected GCM-RCMs is compared with the simulated discharge by the same RRM model forced by observed climatic data (*Qref*).

performance indexes. However, for *QminDate*, HBV generally underestimates it by about one to two decades (exceptional until 31 decades), but IHACRES and GR4J overestimate this date by the same rate. It seems that the origin of these differences is that the date of minimum flow could migrate from the end of summer to the beginning of autumn.

Our results illustrate a strong similarity between the different RRM in terms of raw GCM-RCMs performance over the reference period for the majority of performance indexes, in spite of their different structures and level of complexity. In effect, the *NSE* criterion calculated for each couple of RRM series of hydrological performance index values (of all GCM-RCMs evaluation over all study catchments) (Table 2) was between 0.70 and 0.97 for high flows, medium flows and seasonality performance indexes. The *NSE* calculated for the series of hydrological performance indexes related to low flows and dates was generally lower than the *NSE* calculated for the other performance indexes. The *NSE* values are non-meaningful in these cases, because low flows are close to zero and are already well reconstituted by all RRM; and dates (especially *QminDate* and *QmaxDate*) could easily migrate from one season to another, as mentioned above.

Table 2 | NSE criteria calculated for each couple of RRM series of hydrological performance indexes (from all GCM-RCM and catchments)

	Low flows						Medium flows		High flows			Seasonality			Dates		
	Qmin	Q5	Q25	JJA	SON	MAM	Q50	Q75	Q95	DJF	Qmax	Qmean	STD	1-Nash	HFD	QmaxDate	QminDate
HBV/IHACRES	0.03	0.01	0.24	0.75	0.84	0.88	0.77	0.68	0.94	0.95	0.97	0.95	0.95	0.94	0.84	0.47	0
HBV/GR4J	0.38	0.53	0.66	0.57	0.44	0.91	0.81	0.81	0.87	0.96	0.91	0.94	0.93	0.98	0.78	0.72	0
IHACRES/GR4J	0	0.01	0.57	0.95	0.41	0.98	0.82	0.89	0.95	0.92	0.91	0.94	0.92	0.93	0.87	0.65	0.01

3.4. Evaluation of bias-corrected GCM–RCMs by hydrological modelling

In Figure 4(c), we compare the simulated discharge by the HBV model forced by bias-corrected GCM–RCMs with the reference discharge (the discharge simulated by the RRM forced by observed climate data) for the J. Antra and Rhezala catchments (the results obtained from all catchments and RRM are shown in Supplementary Material, Fig A4). Figure 5 shows the hydrological performance indexes applied to the discharge from the RRM forced by bias-corrected GCM–RCMs.

It was found that due to the improvements in the representation of climate variables after bias correction throughout the annual cycle (Figure 4(a) and 4(b)), the discharge greatly improved after bias correction. The discharge from RRM forced by bias-corrected GCM–RCMs came very close to the reference seasonal discharge values (Figure 4(c)). The improvements in discharge were substantial for all catchments (see Fig A4 in Supplementary Material). The bias correction led to an ensemble average that was much closer to the reference discharge with a significantly decreased error range for the simulated winter

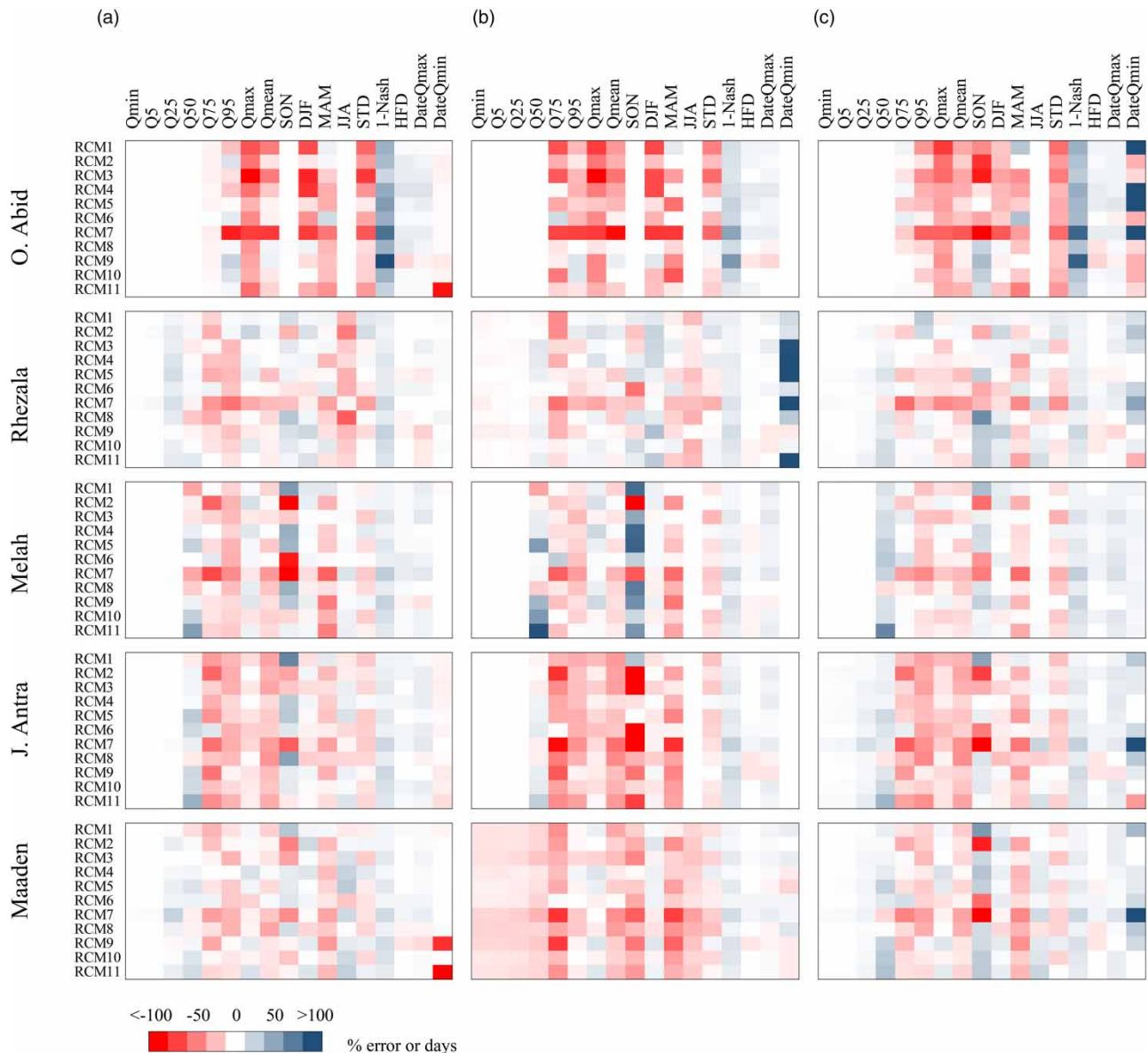


Figure 5 | The range in the performance of the bias-corrected GCM–RCMs (y-axis) according to hydrological performance indexes (x-axis) using three hydrological models: (a) HBV, (b) IHACRES and (c) GR4 for the five study catchments. The colours in the heat maps illustrate the values of error of bias-corrected GCM–RCMs with values in units of percent error, except *HFD*, *QmaxDate* and *QminDate*, which are in the units of days.

and spring flood peaks and annual complete discharge. The enhancement was not as obviously noticeable for summer discharge, which is already well reconstituted by raw climate models.

All performance indexes, except *DateQmin*, show an improvement after bias correction for all GCM-RCMs, except GCM-RCM7 (IPSL-IPSL). It seems that the origin of the deterioration in *DateQmin* is that the date of minimum flow could migrate easily from the end of summer to the beginning of autumn.

The performance of GCM-RCMs in the O. Abid catchment in terms of the used hydrological performance indexes seems to be poor. However, the results from the other catchments are close.

This study has shown that the performance of raw GCM-RCM7 (IPSL-IPSL) was poor over the reference period. Even though bias correction has improved the performance of this model, it is still seen to be poor.

3.5. Ranking climate models

For each performance index considered, the raw and bias-corrected climate models were ranked according to the measure of absolute relative error ($|E_{raw}|$ and $|E_{qm}|$) (Figure 6). To synthesize our results, we combined all of the hydrological performance indexes into a single performance index, referred to each RRM. The estimation of performance for each RRM involves considering the median across all of the hydrological performance indexes (except dates that have a different unit) and all of the catchments for a given GCM-RCM. We considered the median, because it helps to prevent a particularly poor or high performing index or catchment from affecting the ranking. To synthesize our results, we also give ranking of raw and quantile-mapped GCM-RCMs to the median performance of the three RRMs.

We point out that the ranking of raw climate models varies considerably from one performance index to another. The ranking of raw GCM-RCMs shows that generally GCM-RCMs that rank well on high-flow performance indexes rank poorly on low-flow performance indexes and vice versa.

As Figure 6 shows, we notice a switch in the placement of GCM-RCMs, except that of GCM-RCM7 (IPSL-IPSL), due to their performance improvement after quantile mapping. However, as Figure 5 shows, except GCM-RCM7 (IPSL-IPSL), all bias-corrected GCM-RCMs performed well over the reference period, and their ranking (Figure 7) became non-meaningful.

Our results illustrate a strong similarity between the different RRMs in terms of raw GCM-RCMs ranking over the reference period for the median performance indexes (Figure 6), in spite of their different model structures. In effect, the ranks of

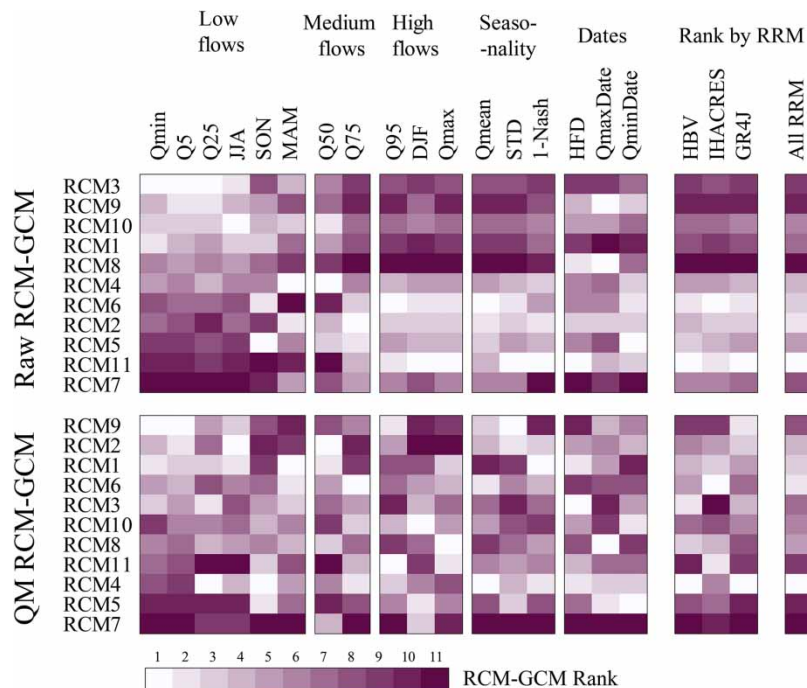


Figure 6 | The ranking of raw and bias-corrected GCM-RCMs (y-axis) according to various hydrological performance indexes (x-axis). The ranking was elaborated according to the median absolute error over the five study catchments. (The GCM-RCMs in the heat plot are sorted according to Q5 rank.)

GCM-RCMs according to RRM are similar for 13 cases (ranks according to a couple of RRM of a given GCM-RCM). The difference in ranking is equal to one position in 18 cases and two positions for only one case. However, this similarity becomes less strong for bias-corrected GCM-RCMs due to the good performance of all bias-corrected GCM-RCMs, except that of GCM-RCM7 (IPSL-IPSL).

4. DISCUSSION

The objective of this study was to examine whether the structural uncertainty induced by the use of RRM affects climate model evaluation and ranking performed by using a process-oriented framework with RRM.

An important key finding was related to the strong behavioural similarities between the different hydrological models, which give similar results in terms of GCM-RCM performance and ranking. This proves that structural uncertainty generated by the use of RRM does not have any great impact on GCM-RCM evaluation and ranking. However, we acknowledge that the other sources of uncertainty need more investigation in further studies, especially uncertainty related to parameter instability and to erroneous data used for calibration/validation. The reference and observed discharges may not be similar enough, implying that the parameter sets of RRM need further tuning, perhaps taking more into consideration the uncertainty related to observed data on parameter estimation.

The difference in the ranking of climate models between the different performance indexes proves the importance of considering a wide set of performance indexes to evaluate GCM-RCM performance.

Biases are often still present after bias correction of precipitation and temperature of raw GCM-RCMs, although they are minimized. This is not surprising, as has been reported by the previous literature. [Addor & Seibert \(2014\)](#) show that differences between the observations and the GCM-RCM simulations remain for other time scales after performing a bias correction of precipitation over a daily time stage, for example. Our research shows a similar manifestation of this principle, in that regular biased-corrected precipitation and temperature data contain monthly biases.

However, a major concern of bias correction techniques is that they only target the symptoms of model imperfections (i.e. biases in the simulations) and not the origins of these imperfections ([Maraun et al. 2017](#)). This leads to doubt about the ability of bias correction techniques to correct future simulations in an effective way. In a sense, bias correction provides the right answer (i.e. simulations looking like observations) but not necessarily for the right reasons ([Hakala et al. 2019](#)). The assessment and ranking of raw climate models could then be useful for impact studies; in fact, it provides an assessment of the climate model realism. We suppose that a bias-corrected model having the best result before bias correction, and needing less correction, could be considered more credible for impact studies than a bias-corrected model having a large error before correction, even though this error is sufficiently corrected by the bias correction method.

Some studies have proposed that focusing on climate models with good performance at simulating some aspects of historical climate may allow adequate coverage of GCM uncertainty and produce accurate projections without incurring significant computational costs. [Weiland et al. \(2012\)](#) showed that selecting climate models based on numerous criteria, including their ability to reproduce observed mean discharge, produced results that were reasonable and were comparable to results obtained by averaging climate models with weighting based on the same criteria. We suggest then to use the 10 best-performing GCM-RCMs for impact studies over the study catchments. GCM-RCM7 (IPSL-IPSL) should be excluded due to its poor performance over the study region.

1. The O. Abid catchment has given the poorest performance of GCM-RCMs according to the proposed hydrological performance indexes. This could be due to the catchment size (81 km², the smallest one in the study), which is smaller than one entire RCM grid cell of EURO-CORDEX (12.5*12.5 = 156.25 km²). This makes the use of GCM-RCM output challenging since they were not designed to represent features at the spatial scale smaller than their grid ([Hakala et al. 2018](#)).

A univariate bias correction technique has been used in this study, which implies that temperature and precipitation have been fixed separately to each other. This technique is restricted in that the intervariable dependence structure between temperature and precipitation is not specifically considered. There are more sophisticated techniques to adapt, such as the multivariate bias correction, that take into account the intervariable dependence structure (e.g. between different quantities like temperature and precipitation or between sites) ([Vrac & Friederichs 2015](#); [Cannon 2018](#); [Vrac 2018](#); [Adeyeri et al. 2020b](#)) and that are supposed to better bias-correct GCM-RCM simulation by taking into account other aspects. [Bellil & Dakhlaoui \(2019\)](#) found that there is not much difference between univariate bias correction (Quantile Delta Mapping)

and a bivariate bias correction method (MBCn; Cannon 2018) of daily temperature and precipitation to correctly reproduce the seasonality of these variables for a catchment situated in northern Tunisia (El Abid). However, other multivariate bias correction methods considering the inter-site dependence structure could be more promising, especially for this region known for its high spatial variability of precipitation (Ouachani *et al.* 2013; Dhib *et al.* 2017).

5. CONCLUSION

The main objective of the present study was to assess whether RRM's structural uncertainty affects the evaluation and ranking of GCM-RCMs, performed in a process-oriented evaluation framework based on hydrological models. For this purpose, we evaluated 11 GCM-RCM climate variables (daily precipitation and temperature) from EURO-CORDEX using three conceptual rainfall-runoff models, over a 30-year reference period (1970–2000), for five catchments in northern Tunisia. Seventeen discharge performance indexes were used to investigate the representation of the hydrological processes.

Our results illustrate a strong similarity between the different RRM's in terms of raw GCM-RCM performance over the reference period for the majority of performance indexes, in spite of their different model structures, which proves that structural uncertainty generated by the use of RRM's does not have any significant impact on GCM-RCM evaluation and ranking. The similarity is less marked for bias-corrected GCM-RCMs that performed well over the reference period. Their ranking is non-meaningful in such a situation. Such results contribute to confirm that hydrological modelling can be used in an integrated way at the catchment level to evaluate and rank climate model simulations.

In addition, we point out that the ranking of climate models varies considerably from one performance index to another, which proves the importance of considering a wide set of performance indexes to evaluate GCM-RCM performance. We also found that, generally, raw climate models that rank well on high-flow performance indexes rank poorly on low-flow performance indexes and vice versa.

This study offers a first assessment of GCM-RCMs from EURO-CORDEX over northern Tunisia at the catchment scale. We recommend the use of the 10 best-performing GCM-RCMs for impact studies over the study catchments. IPSL-IPSL GCM-RCMs should be excluded due to their poor performance.

Future work should focus on ways to directly integrate hydrological performance indexes in the bias correction method, which could allow considering intervariable dependence in a process-oriented framework using hydrological modelling. Future studies should also focus on the impact of other sources of uncertainty of hydrological modelling on the assessment and ranking of GCM-RCMs, such as uncertainty related to the calibration procedures, erroneous data used for calibration/validation and parameter instability (e.g. Coron *et al.* 2012; Brigode *et al.* 2013; Adeyeri *et al.* 2020c).

ACKNOWLEDGEMENTS

The authors are grateful to the INM (Institut National de la Météorologie) and to the DGRE (Direction Générale des Ressources en Eau) in Tunisia for providing the necessary hydro-climatic data for this study. They acknowledge the EURO-CORDEX community for the provision of climate modelled data. The authors thank Dr Urs Beyerle for his assistance with the retrieval of EURO-CORDEX data, Dr Yves Trambly and Dr Kirsti Hakala for their constructive comments and also the editor and anonymous reviewers for their interest in this work and their useful comments, which helped to improve the manuscript.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Abbaris, A., Dakhlaoui, H., Thiria, S. & Bargaoui, Z. 2014 Variational data assimilation with the YAO platform for hydrological forecasting evolving water resources systems. In: *Understanding, Predicting and Managing Water-Society Interactions (Proc. of ICWRS2014, Bologna, Italy, June 2014)*, Vol. 364. IAHS Publ., pp. 1–6.
- Achieng, K. O. & Zhu, J. 2019 Application of Bayesian framework for evaluation of streamflow simulations using multiple climate models. *J. Hydrol.* **574**, 1110–1128. <https://doi.org/10.1016/j.jhydrol.2019.05.018>.
- Addor, N. & Seibert, J. 2014 Bias correction for hydrological impact studies – beyond the daily perspective. *Hydrol. Process.* **28**, 4823–4828. doi:10.1002/hyp.10238.

- Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N. & Clark, M. P. 2018 A ranking of hydrological signatures based on their predictability in space. *Water Resour. Res.* **54** (11), 8792–8812.
- Adeyeri, O. E., Lawin, A. E., Laux, P., Ishola, K. A. & Ige, S. O. 2019 Analysis of climate extreme indices over the Komadugu-Yobe basin, Lake Chad region: past and future occurrences. *Weather Clim. Extremes.* **23**, 2019. <https://doi.org/10.1016/j.wace.2019.100194>.
- Adeyeri, O. E., Laux, P., Lawin, A. E., Ige, S. O. & Kunstmann, H. 2020a Analysis of hydrometeorological variables over the transboundary Komadugu-Yobe basin, West Africa. *J. Water Clim. Change* **11** (4), 1339–1354. <https://doi.org/10.2166/wcc.2019.283>.
- Adeyeri, O. E., Laux, P., Lawin, A. E. & Oyekan, K. S. A. 2020b Multiple bias-correction of dynamically downscaled CMIP5 climate models temperature projection: a case study of the transboundary Komadugu-Yobe river basin, Lake Chad region, West Africa. *SN Appl. Sci.* **2**, 1221. <https://doi.org/10.1007/s42452-020-3009-4>.
- Adeyeri, O. E., Laux, P., Arnault, J., Lawin, A. E. & Kunstmann, H. 2020c Conceptual hydrological model calibration using multi-objective optimization techniques over the transboundary Komadugu-Yobe basin, Lake Chad Area, West Africa. *J. Hydrol. Reg. Stud.* **27**, 100655. <https://doi.org/10.1016/j.ejrh.2019.10065>.
- Adham, A., Wesseling, J. G., Abed, R., Riksen, M., Ouessar, M. & Ritsema, C. J. 2019 Assessing the impact of climate change on rainwater harvesting in the Oum Zessar watershed in Southeastern Tunisia. *Agric. Water Manage.* **221**, 131–140.
- Allani, M., Mezzi, R., Zouabi, A., Béji, R., Joumade-Mansouri, F., Hamza, M. E. & Sahli, A. 2019 Impact of future climate change on water supply and irrigation demand in a small Mediterranean catchment. Case study: Nebhana dam system, Tunisia. *J. Water Clim. Change.* <https://doi.org/10.2166/wcc.2019.131>.
- Bargaoui, Z., Dakhlaoui, H. & Houcine, A. 2008 Modélisation pluie-débit et classification hydroclimatique. *Rev. Sci. Eau.* **21**, 233–245. doi:10.7202/018468ar.
- Bargaoui, Z., Trambly, Y., Lawin, E. A. & Servat, E. 2014 Seasonal precipitation variability in regional climate simulations over northern basins of Tunisia. *Int. J. Climatol.* doi:10.1002/joc3683.
- Beck, M. B. 1987 Water quality modelling: a review of the analysis of uncertainty. *Water Resour. Res.* **23** (8), 1393–1442.
- Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J. & Wood, E. F. 2017 Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.* **21**, 6201–6217. <https://doi.org/10.5194/hess-21-6201-2017>.
- Bellil, O. & Dakhlaoui, H. 2019 Effects of univariate and multivariate bias correction on hydrological projections in northern Tunisia. In: *13th International Geological Conference*, 03–06 March 2020, Jeddah, Saudi Arabia.
- Ben Fraj, W., Elloumi, M. & Molle, F. 2019 The politics of interbasin transfers: socio-environmental impacts and actor strategies in Tunisia. *Nat. Resour. Forum* **43**, 17–30. <https://doi.org/10.1111/1477-8947.12165>.
- Ben Jaafar, A. & Bargaoui, Z. 2020 Generalized split-sample test interpretation using rainfall runoff information gain. *J. Hydrol. Eng.* **25** (1). doi:10.1061/(ASCE)HE.1943-5584.0001868.
- Bennett, B., Thyer, M., Leonard, M., Lambert, M. & Bates, B. 2019 A virtual hydrological framework for evaluation of stochastic rainfall models. *Hydrol. Earth Syst. Sci.* **23**, 4783–4801. <https://doi.org/10.5194/hess-23-4783-2019>.
- Even, K. J. 2006 A manifesto for the equifinality thesis. *J. Hydrol.* **320**, 18–36.
- Breinl, K. 2016 Driving a lumped hydrological model with precipitation output from weather generators of different complexity. *Hydrol. Sci. J.* **61** (8), 1395–1414. doi:10.1080/02626667.2015.1036755.
- Brigode, P., Oudin, L. & Perrin, C. 2013 Hydrological model parameter instability: a source of additional uncertainty in estimating the hydrological impacts of climate change? *J. Hydrol.* **476**, 410–425. doi:10.1016/j.jhydrol.2012.11.012.
- Cannon, A. J. 2018 Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. *Clim. Dynam.* **50**, 31–49. <https://doi.org/10.1007/s00382-017-3580-6>.
- Chen, J., Brissette, F. P., Chaumont, D. & Braun, M. 2013. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resour. Res.* **49**, 4187–4205. <https://doi.org/10.1002/wrcr.20331>.
- Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M. & Hendrickx, F. 2012 Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments. *Water Resour. Res.* **48**, W05552. <https://doi.org/10.1029/2011WR011721>.
- Dakhlaoui, H., Bargaoui, Z. & Bárdossy, A. 2009 Comparaison de trois méthodes d’usage de la technique des voisins les plus proches en vue d’amélioration de la performance de l’algorithme SCE-UA appliqué pour le calage du modèle pluie-débit HBV. In: *Hydroinformatics in Hydrology, Hydrogeology and Water Resources*, Vol. 331 (I.D. Cluckie, Y. Chen, B. Babovic, L. Konikow, A. Mynett eds.). IAHS Publ., Wallingford, UK, pp. 139–153.
- Dakhlaoui, H., Bargaoui, Z. & Bárdossy, A. 2012 Toward a more efficient calibration schema for HBV rainfall-runoff model. *J. Hydrol.* **444–445**, 161–179. doi:10.1016/j.jhydrol.2012.04.015.
- Dakhlaoui, H., Ruelland, D., Trambaly, Y. & Bargaoui, Z. 2017 Evaluating robustness of conceptual rainfall-runoff models under climate variability in Northern Tunisia. *J. Hydrol.* **550** (2017), 201–217. <https://doi.org/10.1016/j.jhydrol.2017.04.032>.
- Dakhlaoui, H., Ruelland, D. & Trambaly, Y. 2019a A bootstrap-based differential split-sample test to assess the transferability of conceptual rainfall-runoff models under past and future climate variability. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2019.05.056>.
- Dakhlaoui, H., Seibert, J. & Hakala, K. 2019b Hydrological impacts of climate change in Northern Tunisia. Advances in sustainable and environmental hydrology, hydrogeology, hydrochemistry and water resources. In: *Proceedings of the 1st Springer Conference of the Arabian Journal of Geosciences (CAJG-1), Tunisia 2018. Series: Advances in Science, Technology & Innovation* (Chaminé, H. I., Barbieri, M., Kisi, O., Chen, M. & Merkel, B. J., eds.). https://doi.org/10.1007/978-3-030-01572-5_71.

- Dakhlaoui, H., Seibert, J. & Hakala, K. 2020 Sensitivity of discharge projections to potential evapotranspiration estimation in northern Tunisia. *Reg. Environ. Change*. doi:10.1007/s10113-020-01596-8.
- Deidda, R., Marrocu, M., Caroletti, G., Pusceddu, G., Langousis, A., Lucarini, V., Puliga, M. & Speranza, A. 2013 Regional climate models' performance in representing precipitation and temperature over selected Mediterranean areas. *Hydrol. Earth Syst. Sci.* **17**, 5041–5059. <http://dx.doi.org/10.5194/hess-17-5041-2013>.
- Deshmukh, A. & Singh, R. 2019 A Whittaker-Biome based framework to account for the impact of climate change on catchment behavior. *Water Resour. Res.* <https://doi.org/10.1029/2018WR023113>.
- Dhib, S., Mannaerts, C. M., Bargaoui, Z., Retsios, V. & Maathuis, B. H. P. 2017 Evaluating the MSG satellite multi-sensor precipitation estimate for extreme rainfall monitoring over northern Tunisia. *Weather Clim. Extreme*. **16**, 14–22. <https://doi.org/10.1016/j.wace.2017.03.002>.
- Dimitriadis, P. & Koutsoyiannis, D. 2019 The mode of the climacogram estimator for a Gaussian Hurst-Kolmogorov process. *J. Hydroinform.* **22**, 160–169.
- Dimitriadis, P., Koutsoyiannis, D., Iliopoulou, T. & Papanicolaou, P. A. 2021. Balasko-scale investigation of stochastic similarities in marginal distribution and dependence structure of key hydrological-cycle processes. *Hydrology* **8** (2), 59. <https://doi.org/10.3390/hydrology8020059>.
- Duan, Q. Y., Sorooshian, S. & Gupta, V. 1992 Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* **28**, 1015–1031. doi:10.1029/91WR02985.
- Duan, Q., Sorooshian, S. & Gupta, V. 1994 Optimal Use of the SCE-UA global optimisation method for calibrating watershed models. *J. Hydrol.* **158**, 265–284.
- Fader, M., Giupponi, C., Burak, S., Dakhlaoui, H., Koutroulis, A., Lange, M. A., Llasat, M. C., Pulido-Velazquez, D. & Sanz-Cobena, A. 2020 Water. In: *Climate and Environmental Change in the Mediterranean Basin – Current Situation and Risks for the Future. First Mediterranean Assessment Report* (Cramer, W., Guiot, J. & Marini, K., eds.). Union for the Mediterranean, Plan Bleu, UNEP/MAP, Marseille, France, pp. 181–236. doi:10.5281/zenodo.4768833.
- Foughali, A., Trambly, Y., Bargaoui, Z., Carreau, J. & Ruelland, D. 2015 Hydrological modeling in northern Tunisia with regional climate model outputs: performance evaluation and bias-correction in present climate conditions. *Climate* **2015** (3), 459–473. doi:10.3390/cli3030459.
- Gleckler, P. J., Taylor, K. E. & Doutriaux, C. 2008 Performance performance indexes for climate models. *J. Geophys. Res. Atmos.* **113**, 1–20. doi:10.1029/2007JD008972.
- Gudmundsson, L., Do, H. X., Leonard, M. & Westra, S. 2018 The global streamflow indices and metadata archive (GSIM) – part 2: quality control, time-series indices and homogeneity assessment. *Earth Syst. Sci. Data* **10** (2), 787–804.
- Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. 2009 Decomposition of the mean squared error and NSE performance criteria implications for improving hydrological modelling. *J. Hydrol.* **377**, 80–91.
- Hachani, A., Ouessar, M. & Zerrim, A. 2017 Chapter 1: a study of water stress on olive growing under the effect of climate change in South East of Tunisia. In: *Water and Land Security in Drylands* (M. Ouessar, D. Gabriels, A. Tsunekawa & Evett, S. eds.). doi:10.1007/978-3-319-54021-4_1.
- Haj-Amor, Z., Acharjee, T. D., Dhaouadi, L. & Bouri, S. 2020 Impacts of climate change on irrigation water requirement of date palms under future salinity trend in coastal aquifer of Tunisian oasis. *Agric. Water Manage.* <https://doi.org/10.1016/j.agwat.2019.105843>.
- Hakala, K., Addor, N. & Seibert, J. 2018 Hydrological modeling to evaluate climate model simulations and their bias correction. *J. Hydrometeor.* doi:10.1175/JHM-D-17-0189.1. In press. <https://doi.org/10.1002/joc.1779>.
- Hakala, K., Addor, N., Teutschbein, C., Vis, M., Dakhlaoui, H. & Seibert, J. 2019 Hydrological climate change impact modeling. In: *Encyclopedia of Water: Science, Technology, and Society* (Maurice, P., ed.). doi:10.1002/9781119300762.wsts0062.
- Hublart, P., Ruelland, D., Dezetter, A. & Jourde, H. 2015 Reducing structural uncertainty in conceptual hydrological modelling in the semi-arid Andes. *Hydrol. Earth Syst. Sci.* **19**, 2295–2314.
- Hurst, H. E. 1951 Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civ. Eng.* **116**, 770–799.
- Jacob, D., Bärring, L., Christensen, O. B., Christensen, J. H., de Castro, M., Déqué, M., Giorgi, F., Hagemann, S., Hirschi, M., Jones, R., Kjellström, E., Lenderink, G., Rockel, B., Sánchez, E., Schär, C., Seneviratne, S.I., Somot, S., van Ulden, A. & van den Hurk, B. 2007 An inter-comparison of regional climate models for Europe: model performance in present-day climate. *Clim. Chang.* **81**, 31–52.
- Jury, M. W., Prein, A. F., Truhetz, H. & Gobiet, A. 2015 Evaluation of CMIP5 models in the context of dynamical downscaling over Europe. *Am. Meteorol. Soc.* **28**, 5575–5582. doi:10.1175/JCLI-D-14-00430.1.
- Khan, A. J. & Koch, M. 2018 Selecting and downscaling a set of climate models for projecting climatic change for impact assessment in the upper Indus basin (UIB). *Climate* **2018** (6), 89. doi:10.3390/cli6040089.
- Koutsoyiannis, D., Yao, H. & Georgakakos, A. 2008 Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods/Prévision du débit du Nil à moyen terme: une comparaison de méthodes stochastiques et déterministes. *Hydrol. Sci. J.* **53** (1), 142–164.
- Kuczera, G. 1997 Efficient subspace probabilistic parameter optimization for catchment models. *Water Resour. Res.* **33**, 177–185.
- Li, Z., Chen, Y., Shen, Y., Liu, Y. & Zhang, S. 2013 Analysis of changing pan evaporation in the arid region of Northwest China. *Water Resour. Res.* **49**, 2205–2212.
- Manai, K. 1983 *Etude Hydrologique De L'oued Rhezala Fernana*. PFE, ESIER, Tunisie, p. 28.

- Maraun, D. 2013 Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. *J. Clim.*, 2137–2143. doi:10.1175/JCLI-D-12-00821.1.
- Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A. & Mearns, L. O. 2017 Towards process-informed bias correction of climate change simulations. *Nat. Clim. Change* 7, 764–773. <https://doi.org/10.1038/nclimate3418>.
- Markonis, Y. Y., Moustakis, C., Nasika, P., Sychova, P., Dimitriadis, M., Hanel, P. & Máca Papalexioiu, S. M. 2018 Global estimation of long-term persistence in annual river runoff. *Adv. Water Resour.* 113, 1–12.
- McMillan, H. K. & Brasington, J. 2008 End-to-end flood risk assessment: a coupled model cascade with uncertainty estimation. *Water Resour. Res.* 44, W03419. <https://doi.org/10.1029/2007WR005995>.
- Mendlik, T. & Gobiet, A. 2016 *Clim. Change* 135, 381–393. doi:10.1007/s10584-015-1582-0.
- Merz, R., Parajka, J. & Blöschl, G. 2011 Time stability of catchment model parameters: Implications for climate impact analyses. *Water Res. Research* 47, W02531. doi:10.1029/2010WR009505.
- Milano, M., Ruelland, D., Fernandez, S., Dezetter, A., Fabre, J., Servat, E., Fritsch, J.-M., Ardoin-Bardin, S. & Thivet, G. 2013 Current state of Mediterranean water resources and future trends under climatic and anthropogenic changes. *Hydrol. Sci. J.* 58, 498–518.
- Motavita, D.F., Chow, R., Guthke, A. & Nowak, W. 2019 The comprehensive differential split-sample test: A stress-test for hydrological model robustness under climate variability. *Journal of Hydrology* 573, 501-515, <https://doi.org/10.1016/j.jhydrol.2019.03.054>.
- Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models. 1. A discussion of principles. *J. Hydrol.* 10, 282–290.
- Nasr, Z., Almohammed, H., Gafrej Lahache, R., Maag, C. & King, L. 2008 Drought modelling under climate change in Tunisia during the 2020 and 2050 periods. In: *Drought Management: Scientific and Technological Innovations* (López-Francos, A., ed.). CIHEAM, Zaragoza, pp. 365–370 (Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 80).
- O'Connell, P., Koutsoyiannis, D., Lins, H. F., Markonis, Y., Montanari, A. & Cohn, T. 2016 The scientific legacy of Harold Edwin Hurst (1880–1978). *Hydrol. Sci. J.* 61, 1571–1590.
- Ouachani, R., Bargaoui, Z. & Ouarda, T. 2013 Power of teleconnection patterns on precipitation and streamflow variability of upper Medjerda Basin. *International Journal of Climatology* 33 (1), 58–76.
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F. & Loumagne, C. 2005 Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2: towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *J. Hydrol.* 303, 290–306.
- Piani, C., Haerter, J. O. & Coppola, E. 2010 Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.* 99, 187–192.
- Poissant, D., Arsenault, R. & Brissette, F. 2017 Impact of parameter set dimensionality and calibration procedures on streamflow prediction at ungauged catchments. *J. Hydrol. Reg. Stud.* 12, 220–237.
- Ruelland, D., Guinot, V., Levavasseur, F. & Cappelaere, B., 2009. Modelling the long-term impact of climate change on rainfall-runoff processes over a large Sudano-Sahelian catchment. *IAHS-AISH Publication* 333., 59–68.
- Ruelland, D., Dezetter, A. & Hublart, P. 2014 Sensitivity analysis of hydrological modelling to climate forcing in a semi-arid mountainous catchment. In: *Hydrology in a Changing World: Environmental and Human Dimensions (Proc. 7th FRIEND-Water Int. Conf., Montpellier, France, Oct. 2014)*, Vol. 363. IAHS Publ., pp. 145–150.
- Santos, C. A. S., Rocha, F. A., Ramos, T. B., Alves, L. M., Mateus, M., Oliveira, R. P. & Neves, R. 2019 Using a hydrologic model to assess the performance of regional climate models in a semi-Arid watershed in Brazil. *Water* 11, 170.
- Schilling, J., Freier, K. P., Hertig, E. & Scheffran, J. 2012 Climate change, vulnerability and adaptation in North Africa with focus on Morocco. *Agric. Ecosyst. Environ.* 156, 12–26. doi:10.1016/j.agee.2012.04.021.
- Seiller, G., Roy, R. & Anctil, F. 2017 Influence of three common calibration performance indexes on the diagnosis of climate change impacts on water resources. *J. Hydrol.* 547, 280–295.
- Sellami, H., Benabdallah, S., La Jeunesse, I. & Vanclooster, M. 2015 Quantifying hydrological responses of small Mediterranean catchments under climate change projections. *Sci. Total Environ.* 543, 924–936.
- Sellami, H., Benabdallah, S., La Jeunesse, I. & Vanclooster, M. 2016 Climate models and hydrological parameter uncertainties in climate change impacts on monthly runoff and daily flow duration curve of a Mediterranean catchment. *Hydrol. Sci. J.* 1415–1429. <https://doi.org/10.1080/02626667.2015.1040801>.
- Stainforth, D., Allen, M., Tredger, E. & Smith, L. 2007 Confidence, uncertainty and decision-support relevance in climate predictions. *Phil. Trans. R. Soc. A* 365 (1857), 2145–2161.
- Stephens, C. M., Marshall, L. A. & Johnson, F. M. 2019 Investigating strategies to improve hydrologic model performance in a changing climate. *J. Hydrol.* 579. <https://doi.org/10.1016/j.jhydrol.2019.124219>.
- Teutschbein, C. & Seibert, J. 2010 Regional climate models for hydrological impact studies at the catchment scale: a review of recent modeling strategies. *Geogr. Compass* 4 (7), 834–860.
- Teutschbein, C. & Seibert, J. 2012 Bias correction of regional climate model simulations for hydrological climate change impact studies: review and evaluation of different methods. *J. Hydrol.* 456, 12–29.
- Thiemeßl, M. J., Gobiet, A. & Leuprecht, A. 2011 Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int. J. Climatol.* 31, 1530–1544.

- Tramblay, Y., Ruelland, D., Somot, S., Bouaicha, R. & Servat, E. 2013 High-resolution Med-CORDEX regional climate model simulations for hydrological impact studies: a first evaluation of the ALADIN-Climate model in Morocco. *Hydrol. Earth Syst. Sci.* **17**, 3721–3739. <https://doi.org/10.5194/hess-17-3721-2013>.
- Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J. M., Viney, N. R. & Teng, J. 2010 Climate non-stationarity – validity of calibrated rainfall-runoff models for use in climate change studies. *Journal of Hydrology* **394**, 447–457. doi:10.1016/j.jhydrol.2010.09.018.
- Valéry, A., Andréassian, V. & Perrin, C. 2010 Regionalization of precipitation and air temperature over high-altitude catchments: learning from outliers. *Hydrol. Sci. J.* **55**, 928–940. doi:10.1080/02626667.2010.504676.
- Vrac, M. 2018 Multivariate bias adjustment of high-dimensional climate simulations: the rank resampling for distributions and dependences (r2d2) bias correction. *Hydrol. Earth Syst. Sci.* **22**, 3175–3196. <https://doi.org/10.5194/hess-22-3175-2018>.
- Vrac, M. & Friederichs, P. 2015 Multivariate – intervariable, spatial, and temporal – bias correction. *J. Climate* **28**, 218–237. <https://doi.org/10.1175/JCLI-D-14-00059.1>.
- Weiland, F. C. S., van Beek, L. P. H., Weerts, A. H. & Bierkens, M. F. P. 2012 Extracting information from an ensemble of GCMs to reliably assess future global runoff change. *J. Hydrol.* **412–413**, 66–75. <https://doi.org/10.1016/j.jhydrol.2011.03.047>.
- World Meteorological Organization. 2011 *Guide to Climatological Practices (WMO-No. 100)*. World Meteorological Organization, Geneva.
- Yapo, P. O., Gupta, H. V. & Sorooshian, S. 1996 Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology* **181**, 23–48. doi:10.1016/0022-1694(95)02918-4.

First received 20 December 2020; accepted in revised form 31 August 2021. Available online 17 September 2021