Assessing future streamflow in Cyprus through hydrological model calibration under non-stationary climate and regional climate Model ensemble selection

Ioannis Sofokleous[1], George Zittis[2], Gerald Dörflinger[3], and Adriana Bruggeman[1]

[1] Energy, Environment & Water Research Center (EEWRC), The Cyprus Institute, Nicosia, Cyprus.

[2] Climate and Atmosphere Research Centre (CARE-C), The Cyprus Institute, Nicosia, Cyprus.

[3] Water Development Department, Nicosia, Cyprus.

Correspondence to: Ioannis Sofokleous (i.sofokleous@cyi.ac.cy)

Abstract.

The use of conceptual hydrological models for projections of future freshwater resources is challenged by non-stationary climate conditions, as these conditions may affect whether models calibrated under historical climates are suitable for future scenarios. This study aims to (i) develop a framework for the parameterization of a conceptual hydrological model under non-stationary climate conditions and (ii) bias-correct, downscale and evaluate the performance of an 18-member ensemble of Regional Climate Models (RCMs) for simulating future streamflow. The framework was applied to generate streamflow projections for 38 mountain watersheds in the eastern Mediterranean island of Cyprus over the next decades (2030–2060) with the GR4J model. Six Nash-Sutcliffe Efficiency (NSE)- and Kling-Gupta Efficiency (KGE)-based functions and a composite scaled score were used for model calibration and validation across multiple 5-year calibration and 5-year validation periods (1980–2015). Climate non-stationarity was represented by differences in total precipitation between calibration and validation periods ($\Delta P = Pval - Pcal$) using the differential split-sample test approach. The best-performing parameterization during drier periods ($\Delta P \leq -5\%$) was obtained with an NSE objective function applied to square-root transformed streamflow, achieving NSE of 0.48, KGE of 0.54 and total streamflow bias of 9% during validation. This optimized model was selected for future streamflow simulations. Nine RCMs were excluded from the impact assessment because they underestimated the fraction of wet period precipitation (60-73% instead of 82%), resulting in streamflow biases up to 40% in the 1980-2010 reference period. The median of future projections for 2030–2060 shows a 6% reduction in precipitation and a 17% reduction in streamflow. In the worst case, reductions could reach 16% and 39%, respectively. Notably, during the driest years, streamflow reductions could reach 70% relative to historical dry years. Our findings suggest that terrestrial water resources in Cyprus may decrease significantly in the coming decades.

1 Introduction

Historical records show that renewable freshwater resources in the Mediterranean region are decreasing (Gudmundsson et al., 2021). The second half of the 20th century has been characterized by decreasing trends in rainfall and river flow and increasing frequency of droughts across catchments in the Mediterranean (Vicente-Serrano et al., 2014; Marques da Silva et al., 2015; Myronides et al., 2018; Masseroni et al., 2021). The increasing frequency and intensity of water scarcity episodes in the region require the design of adaptation plans, which must take into account climate-driven projections of future water resources (Tramblay et al., 2020). Conceptual hydrological models are most often used for climate change impact studies, however, their capacity to simulate streamflow under changing climate conditions is being questioned (Refsgaard et al., 2014; Ji et al., 2023).

For the calibration of hydrological models to be used in impact assessments, the aspect of transferability for future climate conditions of model parameters that were calibrated for past conditions is under discussion. The Differential Split-Sample Test (DSST), described by Klemeš (1986), is commonly used to test how well a hydrological model performs for different climate conditions than the conditions for which the model was calibrated (Refsgaard et al., 2014). This approach is used considering changes in temperature, precipitation or both (e.g., Vaze et al., 2010; Thirel et al., 2015; Dakhlaoui et al., 2017). However, a number of studies in different climate zones found that model parameter transferability under DSST is less promising for transition towards drier conditions than for transition towards wetter conditions (e.g., Broderick et al., 2016; Le Coz et al., 2016; Yang et al., 2020). The review study of Ji et al. (2023) reported that, for credible model transferability between calibration and validation, the reported range for precipitation change towards drier conditions, is narrower (i.e., from -10% to -30%) than the corresponding range towards wetter conditions (i.e., from +10% to +80%). The model robustness for the transition to drier conditions has also been found to be lower in catchments with high runoff skewness and aridity, commonly found in the Mediterranean and dry climates (Munoz – Castro et al., 2023; Guo et al., 2020). Based on this finding, climate change impact studies recommend calibrating models using sub-periods that have similar annual mean precipitation and temperature to the future periods, rather than calibrating over the entire past period (Dakhlaoui et al., 2019). A sensitivity analysis of the model response to multiple inputs may also provide useful insight into model behaviour under changing conditions, as recommended by Wagener et al. (2022).

The selection of the evaluation measure to be used as an objective function for calibrating a hydrological model could affect the assessment of climate impact. A composite objective function, comprised of a linear combination of different functions accounting for different aspects of flow regimes (Zhang et al. 2008), was tested against commonly used objective functions by both Fowler et al. (2018) and Munoz-Castro et al. (2023). Fowler et al. (2018) found that models calibrated by the multi-objective function were outperformed by the models calibrated by simpler formulations, such as the Refined Index of Agreement (Wilmott et al., 2011) and Kling Gupta Efficiency (KGE; Kling et al., 2012) computed individually per year. According to these authors, the latter two approaches resulted in improved model performance compared to models calibrated using standard squared-error based measures, including KGE computed over the entire time series. The same study suggests that streamflow data for the computation of NSE (Nash-Sutcliffe Efficiency; Nash & Sutcliffe, 1970) or KGE values should be transformed when simulations are intended for drier

future conditions. The square root transformation of streamflow, compared to the inverse transformation, was found to be the most suitable approach for a balanced optimization of both high and low flows, with a small loss of confidence for extremely high or low flow conditions (Seiller et al., 2017). Munoz-Castro et al. (2023) found that the choice of objective function had a larger effect on model performance for high-aridity and low runoff coefficients than on model performance for wetter climates, from a comparison of 12 functions.

The impact of climate change on terrestrial water resources has been studied on global and regional levels using hydrological models forced by different global and regional climate models (GCM/RCM). Asadieh and Krakauer (2017), using bias-corrected meteorological outputs of five GCMs from CMIP5 to project streamflow changes, showed that southern Europe, the Middle East, southern North America and the Southern Hemisphere, will experience a strong decrease in all percentiles of streamflow, highlighting the drought hazard risks under RCP8.5 by the end of the century. Roudier et al. (2016) found that model projections of increasing drought magnitude and duration were more robust, i.e., with lower model spread, for southern Europe than for the rest of the continent for a global warming level of 2°C since pre-industrial. Similarly, Marx et al. (2018) found that, in a 3°C global warming scenario, low flows, defined by the flow threshold exceeded 90% of the time, in the Mediterranean region will become even lower, with up to -35% reduction.

The changes in the terrestrial water resources are driven by changes in temperature and precipitation. For the eastern Mediterranean, Zittis et al. (2022) reported that the current regional warming rate is 0.45°C per decade, nearly two times higher than the global average trend (0.27°C per decade). Cos et al. (2022) found that simulations from the two phases of the Coupled Model Intercomparison Project (CMIP5 and CMIP6, at 1° spatial resolution) project a stronger warming in the Mediterranean, relative to the global mean change, particularly in the summer, which could range from 1.8°C to an alarming 8.5°C by the end of the century. The range of these projections corresponds to the Radiative Concentration Pathways (RCPs) examined at different levels. For example, CMIP5 projections for the entire Mediterranean show temperature increases of 1.1, 2.2 and 4.4 °C under RCP2.6, RCP4.5 and RCP8.5, respectively. Precipitation changes are projected to be +1%, –6.6% and –18.8% for these scenarios over the long-term period (2081–2100), relative to 1986–2005 (Gutiérrez et al. 2021). With high-resolution (0.11° ~ 12km) RCM simulations for the Mediterranean under RCP8.5, Zittis et al. (2021a) highlighted a mean annual precipitation reduction of up to 10% for the first half of the 21st century and reductions up to 20%-40% for the second half, particularly for the southern and eastern areas. In addition to the reduction in average annual amounts, an increase in the number of consecutive days without rain in the region is expected (Reymond et al., 2019). Two major drivers for the drying trend in the Mediterranean are a robust change in the upper-tropospheric large-scale circulation and the reduction of the gradient in land-sea temperatures (Tuel & Eltahir, 2020). Due to the reduction in precipitation and streamflow volumes, the aquatic state of streams could be altered, affecting aquatic habitats and riparian ecosystems (Gallart et al., 2012; Martínez-Fernández et al., 2018).

Previous studies analyzed the performance of certain objective functions in calibrating hydrological models under a drying climate, while other studies investigated the model performance under a non-stationary climate using a single objective function. The current study investigates how multiple, commonly used objective functions and their

transformed versions affect the performance of a conceptual hydrological model under both drying and wetting conditions at different thresholds of average precipitation changes. The specific objectives are: (i) to develop a framework for the parameterization of conceptual hydrologic models for robust streamflow simulations under non-stationary climate conditions, (ii) to bias-correct, downscale and evaluate the performance of an 18-member RCMs ensemble from CMIP5 models for streamflow simulations, and (iii) to apply the optimized hydrological model parameterization with selected RCMs for 38 mountain watersheds in Cyprus, and assess the mid-term future (2030-2060) impact on the island's water resources.

2 Data and Methods

2.1 Model calibration for non-stationary climate conditions

2.1.1 Comparison of objective functions

Six objective functions were used independently for six separate hydrological model calibrations. These functions are based on the NSE and KGE criteria, each computed with the original formula of the criteria and with two types of transformation of the streamflow values: 1) no transformation, 2) square root and 3) natural logarithm (NSE, NSEsqrt, NSElog, KGE, KGEsqrt and KGElog). NSE and KGE were selected as the main objective functions due to their widespread use in hydrological modelling in both original and modified forms (Fowler et al. 2018; Guo et al. 2020). The use of the original metrics and the square-root and logarithmic transformations ensures representation of high-, medium-, and low-flow conditions (Moriasi et al. 2007; Seiller et al. 2017). To understand how the calibrated models, optimized on these objective functions, reproduced the different attributes of the observed streamflow, the six functions were also used as evaluation measures. Percent bias (PBIAS) was used as a seventh evaluation measure to account for total volume error (Moriasi et al. 2007; Coron et al. 2012). Finally, a Composite Scaled Score (CSS; eq. 1), which is used for a relative performance comparison of the model calibrated by the objective functions, was computed using all seven evaluation measures. From these, a table of six objective functions and eight evaluation measures was created, for a simultaneous evaluation of the calibration performance with different objective functions per watershed. To evaluate the overall performance of the model calibrated with alternative objective functions for all watersheds, a summarizing 6x8 table, comprised of the median values of the 38 watersheds for each measure and each objective function was used.

The CSS is computed by averaging the normalized values of the seven evaluation measures as follows (Sofokleous et al., 2021):

$$CSS_i = \frac{1}{N_s}\sum_{s=1}^{N_s}\left(\frac{Y_{s,i}-Y_{s,worst}}{Y_{s,best}-Y_{s,worst}}\right) \quad (1)$$

where i is the index identifying the objective function, s is the index of the evaluation measure out of a number of N_s (seven) measures, $Y_{s,i}$ is the value of measure s obtained by objective function i and $Y_{s,worst}$ and $Y_{s,best}$ are the worst and the best values for measure s obtained by the six objective functions. The score ranges between 0 and 1, with 1

corresponding to the best performance. The CSS has been successfully applied in performance inter-comparisons of different model configurations and methods (Sofokleous et al., 2024; Citrini et al., 2024).

2.1.2 Selection of objective function for non-stationary climate conditions

Calibration was performed for all moving 5-year windows, within the calibration period (1980 – 1998), applying all six objective functions. Observational data and periods are given in Sect. 2.2. Validation was conducted for each calibration window and each corresponding calibrated model based on the six objective functions across 5-year windows within an independent validation period (1998 – 2015). The hydrologic year preceding each 5-year window was used as a warm-up year. In total, 182 (14 × 13) study model experiments resulting from 14 5-year calibration windows within 1980 and 1998 and 13 5-year validation windows within 1998 and 2015 were performed in this study. The objective functions were evaluated for, firstly, all 5-year calibration and 5-year validation pairs (182), and secondly, pairs of 5-year calibration and 5-year validation periods corresponding to different changes in climate conditions, following the differential split-sample test approach (Klemeš 1986). The effect of climate conditions change from the calibration to the validation periods was based on the relative change of total precipitation and average temperature from the calibration to the validation. The temperature minimum increase to select a set of calibration and validation runs was 0.7 °C. This temperature threshold was based on the observed past changes in the study area and the expectation of warmer future conditions (validation) relative to the past climate (calibration). This temperature threshold value did not exclude any five-year period from being included in the study, because every five-year window within the calibration period (1980–1997) could be matched with at least one five-year window in the validation period (1998–2015) that was warmer by 0.7°C or more. The calibration-validation pairs of runs satisfying this temperature criterion were grouped in four precipitation change (ΔP) classes, representing conditions of increased wetness and increased dryness at different thresholds from the calibration to the validation period (ΔP=Pval-Pcal) , i.e., ΔP>15%, 5% <ΔP≤15%, -5%<ΔP≤5%, and ΔP≤-5%.

The evaluation of the model calibrated by six objective functions was performed for all 5-year calibration and all 5-year validation periods and for the 5-year validation periods matching each of the four precipitation change classes. The objective function that achieved the highest score (CSS), among the six functions, on average for both the calibration and the validation periods, was selected and used for the model optimization for future streamflow simulations.

Previous studies used the differential split-sample test, in which periods with distinct climate conditions were comprised of discontinuous sub-periods, i.e., not consecutive hydrological years (Dakhlaoui et al., 2017), or in which multiple consecutive sub-periods were separated by a 3-month interval (Guo et al., 2020). Although longer calibration periods are often recommended, this study adopted a 5-year calibration and validation window as a compromise between capturing interannual climate and hydrological variability typical of Mediterranean conditions and maximizing the number of sub-periods for comparison. A comparison of the 5-year against an 18-year calibration was made showing very similar median performance across 38 watersheds during calibration (NSE: 0.82 vs. 0.83; KGE: 0.89 vs. 0.88) and during an independent 17-year validation (NSE: 0.62 vs. 0.65; KGE: 0.57 vs. 0.59). During

validation,32 out of 38 watersheds exhibited NSE differences <0.1 and 28 out of 38 exhibited KGE differences <0.1. These results indicate that longer calibration periods do not provide a significant improvement in model performance that could affect future streamflow projections.

2.1.3 Selection of model parameterization for future simulations

For the selection of the optimal model parameterization, model runs obtained using the selected objective function were evaluated. For each watershed, the parameter set that achieved the highest value of the selected objective function during the validation period, corresponding to the projected precipitation change class (ΔP) for the study area, was used in the hydrological model for future simulations. Unlike the standard split-sample calibration–validation method, where the parameter set is selected based solely on calibration performance, this approach selects the parameterization based on model performance in both the calibration and validation periods, and in relation to observed changes in temperature and precipitation between the two periods.

2.2 Study area and observational data

The streamflow simulations for the past and the future are conducted in 38 watersheds in Cyprus. Cyprus is an island in the eastern part of the Mediterranean Sea, at latitude 35°N and longitude 33°E. The climate of the island is characterized by the typical variability in precipitation in terms of a clear seasonality pattern, as well as, the year-to-year variability in the distribution of precipitation in the wet period found in Mediterranean climates (Hoerling et al., 2012). December and January are the wettest months; about 80% of total annual precipitation occurs between November and April. The Troodos massif is the water tower of Cyprus and it covers more than 25% of the island's total area (9251 km^2). The 38 watersheds of the study form a radial drainage system around Troodos (Fig. 1). The 38 streamflow gauges defining the areas of the studied watersheds were selected such that any large dams and major waterworks are located downstream of the gauges. Common land use changes occurring in the studied area include agricultural land abandonment and fires over shrubland and forests. Subsequently, a conversion of the abandoned and burned area occurs, through shrub expansion and a slow (re-)growth of, predominantly, pine trees in areas where soil and rain conditions favour their development. Geologically, Troodos is constituted by an ophiolite complex with faulted and highly fractured rocks, especially the gabbro on the upper hillslope, forming fractured aquifers or aquifer systems favoring infiltration (Udluft et al., 2006). Soils in Troodos have a stony gravelly texture and a high variability in soil depth from very shallow (0–10 cm) up to about 100 cm (Camera et al., 2017). Christofi et al. (2020) used isotope and hydrogeochemical sampling and modelling to identify regional groundwater flow through the diabase and basal-group units of the Troodos Fractured Aquifer, providing evidence of groundwater flow across surface watershed boundaries in the Troodos Mountains.

A summary of the long-term average hydroclimatic characteristics of the studied watersheds is shown in Fig. 1(a). The map of Cyprus with the boundaries of the 38 watersheds and the trends in precipitation and streamflow in the 1980-2015 period are also shown in Fig. 1(b). For precipitation, Sen's slopes ranged from -2.2 to 3.2 $mm \cdot y^{-1}$, with no statistically significant ($p < 0.05$) trends (Mann-Kendall test) among the 38 watersheds. For streamflow, Sen's slopes

ranged from -0.22 to 0.01 $Mm^3 \cdot y^{-1}$ with statistically significant negative trends for two watersheds. The reference evapotranspiration (ET) trend (not shown) was statistically significant for all watersheds and Sen's slopes ranged from 1.3 to 3.8 $mm \cdot y^{-1}$.

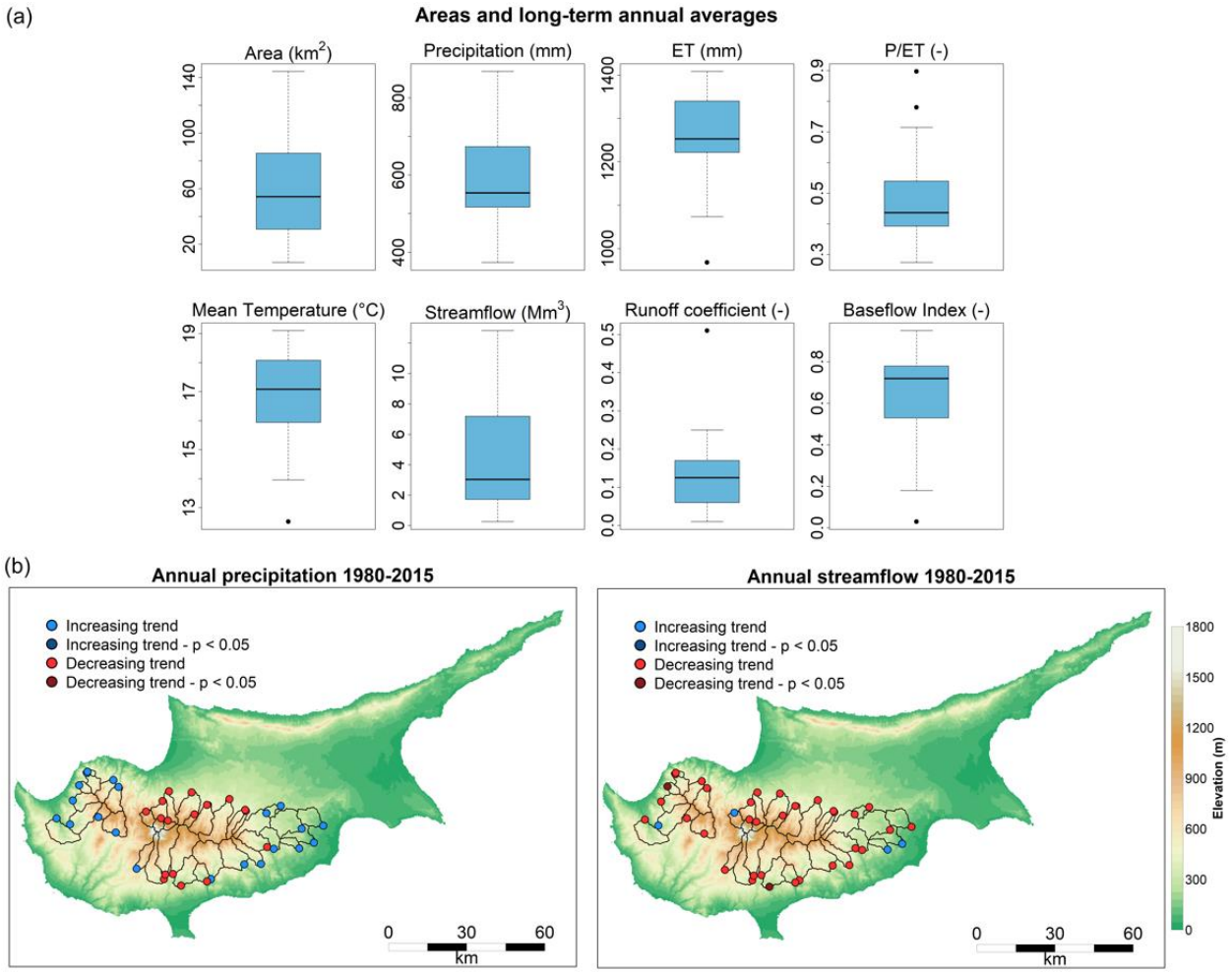


Figure 1. (a) Boxplots of the surface areas and the long-term hydrologic-year averages (1980-2015) of hydrological and climate variables of the 38 study watersheds; (b) Elevation map (m) of the island of Cyprus with the boundaries of the 38 watersheds and the trend and significance of trend, based on Mann-Kendall and Sen's slope statistics, for annual precipitation (left) and streamflow (right), for the 1980-2015 period, shown as circles at the outlets of the watersheds.

Daily precipitation and temperature observations for the area of Cyprus (CY-OBS) for 1980-2015 in gridded datasets at 1 km spatial resolution were used for forcing the hydrological model (Sect. 2.3) and the evaluation of RCM output (Sect. 2.4). The periods 1980–1998 and 1998–2015 were used for calibration and validation, respectively. These gridded observations were generated with the spatial interpolation methods described by Camera et al. (2014) and Sofokleous et al. (2021). The temperature data include daily minimum and maximum values, from which the daily

average temperature was computed. The daily reference evapotranspiration, required as input for the hydrological model was computed from the daily minimum and maximum temperature with the Hargreaves equation (Hargreaves & Samani, 1985). The daily data on watershed-level were extracted from the gridded 1-km data, by averaging the gridded data over the watershed areas. For the hydrologic model calibration and evaluation, daily streamflow observations at the outlets of the 38 watersheds from the monitoring network of the Water Development Department of Cyprus (Fig. 1), in the period 1980-2015, were used. Missing streamflow data (six watersheds, 4-14 years) were filled, by fitting linear or exponential relationships with downstream or upstream stations in the same watershed. These data were used for the calibration and evaluation of the hydrological model. Streamflow observations from Kryos watershed in the southern slopes of the Troodos mountains (r9-6-2-90, see Supplemental Material) were also corrected by taking into account monthly water diversions from Arminou Dam located on the Diarizos river (not included in the study area) to the Kryos watershed, which started in the hydrological year 1998 – 1999. The data correction periods spans outside the calibration period.

2.3 The GR4J conceptual hydrological model

The conceptual, four-parameter, daily GR4J model was used for the streamflow simulations. A detailed description of the model is given by Perrin et al. (2003). The GR4J model has shown good performance for different environments (e.g., Broderick et al., 2016; Le Coz et al., 2016; Guo et al., 2020). The four model parameters represent the watershed's soil water or production storage (X1; mm), groundwater exchange acting on streamflow (X2; mm), streamflow storage (X3; mm), and a time parameter for the unit hydrograph (X4; days). The groundwater exchange coefficient can add a positive (inflow to stream) or negative (recharge losses) contribution to the water balance. The parameters are embedded in exponential equations, which are used to represent the hydrological processes. All input data, i.e., the forcing data of daily precipitation and reference evapotranspiration are expressed in mm over the watershed area. The GR4J model implementation in R (Coron et al., 2023) was used. The parameter optimization in the model code uses the steepest descent local search algorithm. The GR4J model was applied on the 38 watersheds of the study area with streamflow records for the hydrologic years 1980-81 to 2014-15 for model calibration and validation.

2.4 Selection of RCMs for the forcing of future streamflow simulations

Daily precipitation, minimum and maximum temperatures were extracted for Cyprus from 18 CMIP5-driven RCMs from EURO-CORDEX (European hub of the international Coordinated Regional Climate Downscaling Experiment; Jacob et al., 2020), covering a historical (1980-2010) and a future period (2030-2060). This ensemble is based on a combination of six driving GCMs and seven RCMs used for dynamical downscaling. The historical period is used as the reference, to which future precipitation changes are computed. The RCM horizontal spatial resolution is 0.11° (~12 km), and the temporal resolution is at the daily time step. The RCP8.5 scenario, corresponding to the upper range of the projected global-mean surface temperature increases (Meinshausen et al., 2011) was selected to quantify the effects of the least favorable among the plausible climate change scenarios (Pedersen et al., 2021) on surface water resources.

The EURO-CORDEX data were bias-corrected and downscaled using the quantile delta mapping (QDM) algorithm (Cannon et al., 2015). The QDM method preserves the relative (precipitation) or absolute (temperature) changes in quantiles simulated by the RCM. The R MBC Package of Cannon (2024) was used for the QDM application. The CY-OBS area was covered by 55 land-based RCM grid cells. For the downscaling of the climate data from 12-km to 1-km resolution, the RCM values were re-gridded to the 1-km grid of the CY-OBS data, and the bias correction algorithm was applied with the 1-km CY-OBS and the re-gridded RCM data. To ensure that bias-corrected values of daily minimum temperature (T_{min}) do not exceed those of daily maximum temperature (T_{max}), the diurnal temperature range (DTR = T_{max} –T_{min}) and the T_{max} were bias-corrected, following the suggestion by Thrasher et al. (2012), who found that bias-correcting T_{max} and DTR resulted in smaller errors.

The 18 bias-corrected and downscaled RCMs were evaluated for their skill in reproducing observed precipitation indices and for their skill in reproducing observed streamflow, when these RCMs are used as forcing for the calibrated GR4J model in the 1980-2010 reference period. The precipitation indices were computed as long-term averages of annual index values. These are (i) the average and (ii) the standard deviation of annual precipitation, (iii) the ratio of the precipitation of the five wettest months to the annual precipitation (W5R; from November to March), (iv) the Simple Daily Intensity Index (SDII), which is the average precipitation rate on wet days (≥1mm) and (v) the number of days with daily precipitation exceeding 10 mm (R10mm). For streamflow, the relative error of the simulated to observed annual streamflow in each watershed was computed. Based on this evaluation, a subset of RCMs best simulating the reference period precipitation and streamflow was selected to be used for modeling the future streamflow with the calibrated GR4J parameterization.

2.5 Future projections of water resources

The changes in future precipitation and streamflow, relative to the reference period, were evaluated for the total period changes and the changes in the driest and wettest years, i.e., in the two consecutive driest and wettest years, as well as the five driest and five wettest years overall. The changes in precipitation, reference evapotranspiration (ET) and streamflow were also analyzed for groups of the 38 watersheds, defined by four flow regime types. The four flow regime types are: Permanent flow (P), Intermittent pools flow (I-p), Intermittent harsh flow (I-h), and Ephemeral flow (E). These types of flow regimes are derived from two indices of monthly streamflow data, including the flow permanence (Mf), i.e., the long-term mean annual relative number of months with flow, and the six-month seasonal predictability of dry periods (SD6), based on the ratio of multi-annual frequencies of the zero-flow months for the contiguous six wetter months of the year to the frequencies of the zero-flow months for the contiguous six drier months. A detailed description of the derivation of the two indices and the flow regime type is given in Gallart et al. (2012). The flow regime type for the reference period was assigned by the Water Development Department of Cyprus, and the future flow regime type was computed from the simulated flow, modelled with the climate projections of the different RCMs.

3 Results

3.1 Selection of objective function for non-stationary climate conditions

The comparison of the model performances achieved with the six objective functions based on the median values of the 38 watersheds (averaged for all 5-year periods) for each performance measure is shown in Table 1. In the calibration runs, the best evaluation measure values were achieved when the evaluation measure was used as the objective function, shown in bold in the table. These results corroborate that the optimization routine in the GR4J model implementation in R performs reasonably well.

The CSS, which normalizes the evaluation measure values and combines them into one relative performance score, shows that KGE results in the best model performance in the calibration period for 31 out of 38 watersheds, followed by NSEsqrt. Under the conventional split-sample method, the model parameterization obtained with the KGE (0.89) in the calibration would be selected and the validation, based on the full set of 182 5-year periods (14 5-year calibrations x 13 5-year validations), would confirm an acceptable performance (KGE 0.53). However, the evaluation matrices show that the parameterization obtained with the NSEsqrt in the calibration received higher scores for the validation (NSEsqrt 0.74, KGE 0.59) than the KGE-based parameterization (NSEsqrt 0.70, KGE 0.53). In the validation, the NSEsqrt-calibrated model also outperformed the calibrated models in 14 watersheds, whereas the KGE-calibrated model outperformed the other calibrated models in three watersheds only, which is the second worst performance out of the six optimizations. Interestingly, the NSE-based functions led to better model performance, based on CSS, for a larger number of watersheds in the validation compared to the KGE-based functions. The NSE-based functions also achieved a much smaller PBIAS than the KGE-based functions in the validation period, opposite to the lower PBIAS with KGE-based functions in the calibration.

The comparison of the objective functions, based on the CSS and across different long-term average precipitation of the 38 watersheds in Fig. 2, shows that the optimization using NSEsqrt is, on average, the best for all precipitation regimes in the validation for the 182 5-year validation periods, whereas the CSS based on KGE has no important difference from the CSS based on NSE, NSElog and KGElog functions. The patterns in Fig. 2 may be interpreted in relation to both the streamflow transformation and the formulation of the objective functions. The relatively stable or improving performance of the model calibrated with NSElog and KGElog with increasing watershed wetness suggests that logarithmic transformation reduces the influence of the larger and more frequent peak flows, characteristic of wetter catchments. This tendency is most apparent under wetter validation conditions ($\Delta P > 15\%$). However, this advantage is not observed during drying validation periods ($\Delta P \leq -5\%$), possibly because reduced-flow conditions increase the relative importance of bias and variability errors between simulations and observations. Since KGE explicitly incorporates correlation, bias, and variability components, KGE-based calibration may be particularly sensitive to these changes under drying conditions, even when logarithmic transformation is applied. In contrast, NSEsqrt appears more robust across precipitation regimes. The square-root transformation has a weaker effect in attenuation of peak flows, compared to the logarithmic, and thus it may retain a more balanced representation between moderate and high flows during evaluation across both wetter and drier conditions. In contrast, models calibrated with KGEsqrt show consistently weaker performance across precipitation conditions. These findings suggest that the

interaction between transformation type and objective-function structure differs between NSE and KGE formulations, with NSEsqrt providing a more stable compromise across contrasting hydroclimatic conditions.

Subselecting calibration-validation periods with specific changes in climate conditions from the calibration to the validation reveals some model performance losses or gains (Table 1). Simulations in both wetter validation periods ($\Delta P>15\%$ and $5\%<\Delta P\leq 15\%$) achieved comparable performance with simulations in the validation period with small changes in precipitation ($-5\%<\Delta P\leq 5\%$), whereas the evaluation for the drier validation ($\Delta P\leq -5\%$) showed the worst performance compared to the other three classes. This is evident in the reduction of the values of the six goodness-of-fit measures and the increase in the PBIAS values, indicating that the observed hydrographs and the total streamflow volumes are simulated with less accuracy in drier validation periods than in wetter validation periods. The NSEsqrt objective function led to the lowest increase, on average, in PBIAS in the driest validation periods, i.e., with a PBIAS of 9%, compared to the other objective functions, where PBIAS ranged from 17% to 39%. The comparison of the objective functions based on the model performance with the CSS for the wettest validation periods (Fig. 2) shows the close distance of the objective functions across all precipitation regimes of the 38 watersheds. For the drying condition in validation, NSEsqrt-calibrated model stands out as the best-performing across all precipitation regimes, with also increasing distance from the calibrated model with the other objective functions for the wettest watersheds.

In the overall comparison of the performance of the model calibrated with six objective functions with the use of the unitless and objective CSS, the model runs calibrated with KGE rank first (0.74) in the calibration, followed by those calibrated with NSEsqrt (0.65). However, in the full set of 182 validation runs, the model calibrated with NSEsqrt ranks first (0.66), followed by those calibrated with NSElog (0.62), KGE (0.60) and NSE (0.59). The model calibrated with NSEsqrt ranks also first in all four precipitation change classes (CSS from 0.66 to 0.70), whereas models calibrated with KGE show more variable ranking. Due to its consistent performance for all climate change conditions, the objective function of NSEsqrt is selected for the model optimization for the future streamflow simulations with the GR4J model.

Table 1. Median values (across 38 watersheds) of the seven evaluation measures and the composite scaled score: CSS (columns; averaged over all 5-year calibration and validation periods per watershed), and the number of watersheds for which the model parameterization obtained using a given objective function achieved the highest CSS relative to the other functions. Results are shown for the simulated streamflow in the calibration and four validation periods, based on the model parameterizations obtained with each of the six objective functions (rows). Best value for each evaluation measure (per column) is shown in bold.

Evaluation measures	NSE (-)	NSE_{log} (-)	NSE_{sqrt} (-)	KGE (-)	KGE_{log} (-)	KGE_{sqrt} (-)	PBIAS (%)	CSS (-)	#watersheds with max CSS (0-38)
Objective function	CALIBRATION (14 5-year periods)								
NSE	**0.82**	0.45	0.79	0.74	0.67	0.64	-6	0.54	1
NSElog	0.63	**0.75**	0.78	0.68	0.82	0.65	-6	0.60	1

NSEsqrt	0.74	0.64	**0.84**	0.78	0.75	0.81	-9	0.65	3
KGE	0.78	0.58	0.81	**0.89**	0.76	0.75	**0**	**0.74**	**31**
KGElog	0.55	0.72	0.78	0.71	**0.86**	0.70	6	0.60	1
KGEsqrt	0.59	0.47	0.81	0.69	0.69	**0.90**	-2	0.31	1
VALIDATION (182 5-year periods)									
NSE	**0.51**	0.35	0.71	0.53	0.58	0.58	**-1**	0.59	6
NSElog	0.45	**0.62**	0.67	0.51	**0.69**	0.49	8	0.62	9
NSEsqrt	0.50	0.49	**0.74**	**0.59**	0.68	**0.70**	4	**0.66**	**14**
KGE	0.39	0.45	0.70	0.53	0.67	0.59	15	0.60	3
KGElog	0.19	0.60	0.63	0.47	0.69	0.55	30	0.56	6
KGEsqrt	0.6	0.35	0.67	0.41	0.64	0.69	13	0.39	0
VALIDATION (18 five-year periods with ΔP>15%)									
NSE	**0.53**	0.39	0.72	0.55	0.58	0.63	-6	0.54	4
NSElog	0.52	**0.68**	0.70	0.59	**0.72**	0.51	**1**	0.62	9
NSEsqrt	**0.53**	0.58	**0.76**	**0.61**	0.68	**0.73**	-2	**0.66**	**11**
KGE	0.38	0.55	0.72	0.55	0.68	0.69	14	0.62	5
KGElog	0.44	0.64	0.69	0.54	**0.72**	0.57	18	0.61	8
KGEsqrt	0.07	0.44	0.69	0.42	0.63	0.73	16	0.42	1
VALIDATION (44 five-year periods with 5%<ΔP≤15%)									
NSE	0.51	0.48	0.74	0.58	0.61	0.64	-3	0.60	5
NSElog	0.50	**0.67**	0.72	0.58	**0.73**	0.54	5	0.62	9
NSEsqrt	**0.52**	0.57	**0.77**	**0.61**	0.69	**0.73**	**1**	**0.67**	**16**
KGE	0.43	0.55	0.73	0.58	0.67	0.67	11	0.62	3
KGElog	0.36	0.66	0.70	0.52	**0.73**	0.60	18	0.55	4
KGEsqrt	-0.05	0.45	0.69	0.45	0.64	0.74	8	0.38	1
VALIDATION (35 five-year periods with -5%<ΔP≤5%)									
NSE	**0.55**	0.40	0.71	0.55	0.62	0.57	**1**	0.61	7
NSElog	0.47	**0.64**	0.67	0.51	**0.70**	0.50	11	0.60	6
NSEsqrt	0.51	0.50	**0.75**	**0.59**	0.69	0.69	5	**0.68**	**14**
KGE	0.41	0.47	0.71	0.57	0.67	0.58	20	0.60	5
KGElog	0.23	0.60	0.64	0.46	**0.70**	0.56	33	0.54	6
KGEsqrt	-0.05	0.41	0.69	0.41	0.65	**0.69**	15	0.40	0
VALIDATION (22 five-year periods with ΔP≤-5%)									
NSE	0.46	0.19	0.61	0.48	0.56	0.48	20	0.59	5
NSElog	0.30	**0.55**	0.56	0.43	**0.68**	0.45	26	0.61	7
NSEsqrt	**0.48**	0.40	**0.68**	**0.54**	0.64	0.60	**9**	**0.70**	**21**
KGE	0.32	0.35	0.63	0.48	0.64	0.47	20	0.57	1
KGElog	0.06	0.54	0.54	0.32	0.67	0.49	39	0.55	4
KGEsqrt	-0.29	0.29	0.61	0.38	0.59	**0.62**	17	0.41	0

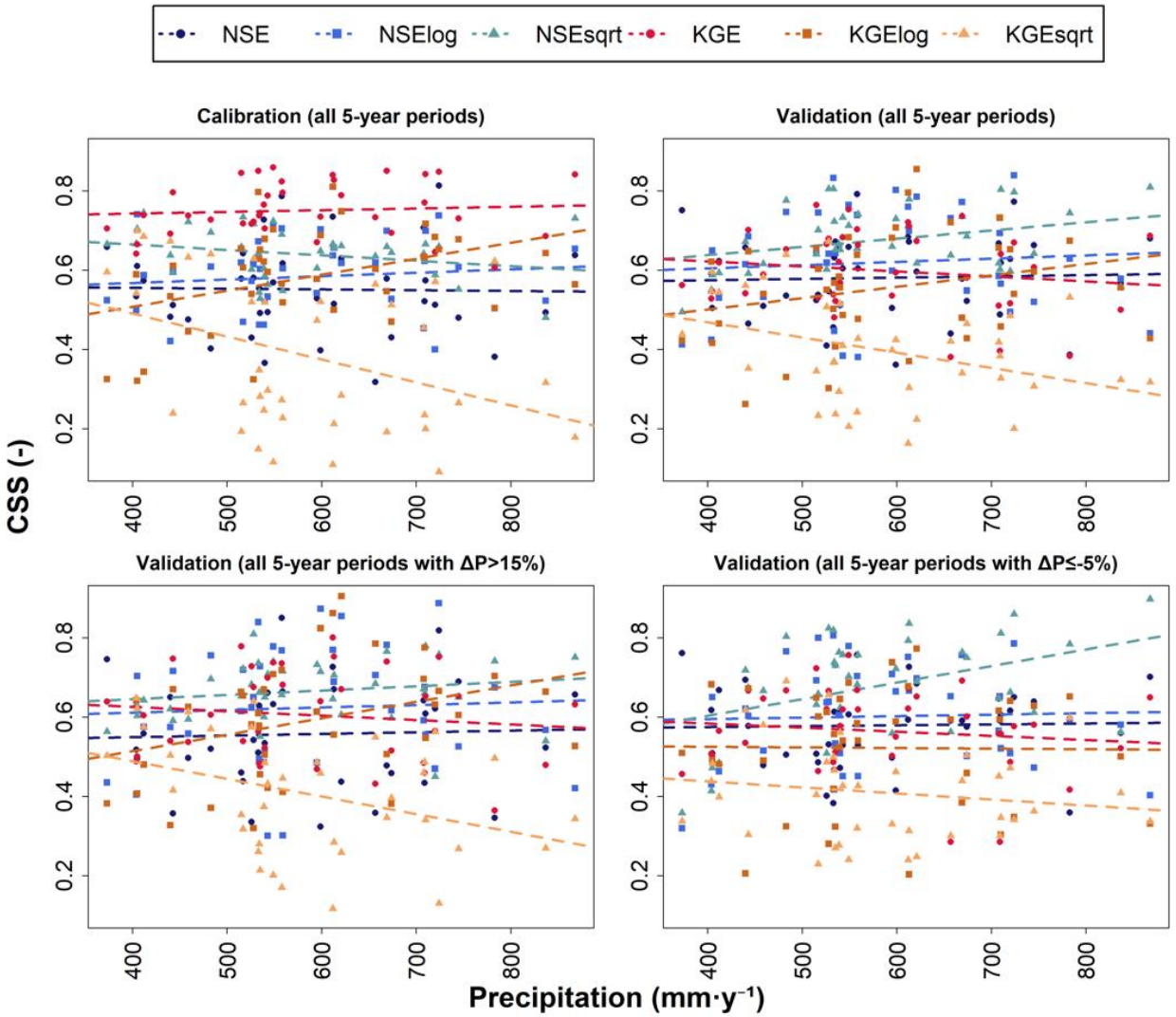


Figure 2. Scatterplots with the composite scaled score (CSS) of the parameterizations obtained with each of the six objective functions for 38 watersheds (y-axis), plotted against the long-term average annual precipitation of the watersheds (x-axis). The dashed lines represent the linear trend between CSS and precipitation per objective function.

The optimized model parameters derived using different objective functions and calibration periods were plotted against the average precipitation of the respective calibration periods (Fig. A1). Although the model parameters do not directly correspond to measurable watershed properties, their calibrated values can provide insight into integrated watershed behavior. In particular, calibrated values for parameter X2 (groundwater exchange) ranged for all watersheds from negative to near zero. In GR4J, negative X2 values imply a loss from the water balance of a watershed. This could be losses to groundwater below the level of the stream (not returning as baseflow) or diversions from streamflow for irrigation, which are processes that are not explicitly represented by the model. The three watersheds with the most negative X2 values have a larger fraction of agricultural land cover (43%-48%), whereas

the five watersheds with near-zero X2 have a larger fraction of forest cover (74%-99%) (Sofokleous et al. 2023). This pattern suggests that the calibrated model correctly accounts for water losses. In addition, the results in Fig. A1 also show that, for a given watershed, parameter values generally vary more between different objective functions than between different calibration periods. Therefore, understanding how different objective functions lead to different model outputs is important for hydrological model applications in areas with diverse watersheds conditions and high climate variability.

3.2 Selection of model parameterization for future simulations

To select the optimal GR4J parameter set for the simulation of streamflow in the future period, the NSEsqrt values of the 14 5-year calibration windows and their respective NSEsqrt values in the five validation periods were extracted per watershed. Fig. 3 shows the best NSEsqrt value of the 14 5-year calibration runs and the best average NSEsqrt value of the validation of each of these 14 model parameterizations over the 13 validation runs (all), and similarly over the validation runs for each of the four precipitation change classes. The best NSEsqrt values identify the best performing model parameterization (out of the 14) for different climate conditions, for each watershed. The median and interquartile range of the NSEsqrt values are also shown. For the future simulation experiments, the parameter set from the 5-year calibration run, with the highest validation performance in drying conditions, i.e., $-5\% < \Delta P \leq 5\%$ and $\Delta P \leq -5\%$, in line with the climate model projections for the study area, was selected for each watershed.

Interestingly, nearly all watersheds were simulated with acceptable performance by all calibration sets, with the best and median NSEsqrt of the 14 parameterizations on average equal to 0.88 and 0.83 for the calibration and 0.78 and 0.69 for all validation runs, respectively. The average interquartile range of NSEsqrt obtained with the 14 parameterizations, equal to 0.09 for the calibration and 0.15 for any validation experiment for all watersheds, shows the small variability in the performance of GR4J with most of the 14 parameterizations based on the 5-year calibration periods.

Examining the same results, with the watersheds compared by their long-term average precipitation, it is seen that lower NSEsqrt values, by about 0.1 and greater variability, up to 0.2, among the 14 parameterizations are obtained for the ten driest watersheds (average precipitation 373 to 517 $mm \cdot y^{-1}$) compared to the ten wettest watersheds (average precipitation 674 to 868 $mm \cdot y^{-1}$). The 12^{th} watershed (Gialia near Pano Gialia, 528 $mm \cdot y^{-1}$ precipitation) stands out as the worst performing out of the 38 watersheds in all validation sets with a median NSEsqrt below 0.2. This particular watershed is only 16 km^2 in size and has a small dam upstream the streamflow gauge of the watershed. The best performing watershed in all validation sets (Stavros tis Psokas, 621 $mm \cdot y^{-1}$), with median NSEsqrt above 0.89 has a size of 86 km^2, which classifies the watershed as medium to large-sized (Fig. 1).

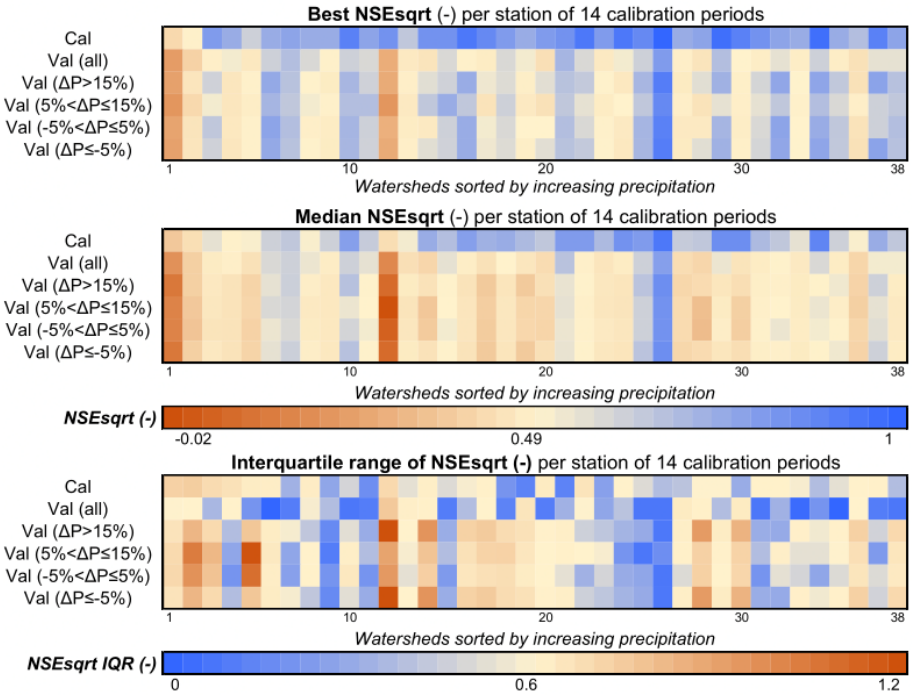


Figure 3 Heatmap with the best (top), median (middle) and interquartile range (IQR) (bottom) of NSEsqrt (-) for each watershed, sorted by increasing long-term annual average precipitation (columns) and each calibration-validation set of runs (rows). The values of NSEsqrt or the IQR for each cell in the heatmap are computed as averages of the values from the 14 5-year calibration runs, the 13 validation runs for each of the 14 calibration runs. i.e., Val (all), and the validation runs corresponding to each of the four precipitation change classes, as specified in Table 1.

3.3 Bias correction and selection of RCMs

The bias correction of the RCMs with the QDM method yielded model indices very close to the observed values for total precipitation, with a maximum difference of 3% and for the SDII and R10mm with minimal difference from the observed values for all 18 RCMS (Table 2). The first nine models listed in Table 2 comprise the subset of RCMs used to quantify future changes in streamflow. These nine models simultaneously achieved the lowest error in the simulation of the ratio of the precipitation in the five wettest months to the annual total (W5R) and in the simulation of annual average streamflow. W5R ranged between 0.79 and 0.84 for the selected RCMs and between 0.60 and 0.73 for the other nine models, compared to the observed 0.82 W5R value. The relative error in the simulation of streamflow ranged between 1% and 11% in absolute values for the selected RCMs and between 5% and 40% for the remaining models. The W5R, as an evaluator of bias-corrected precipitation, has the strongest correlation (0.86) with the error in streamflow simulations compared to the correlation of streamflow error with any of the other four precipitation indices (0.13-0.25).

Table 2. Indices of the observed (CY-OBS) and RCM-simulated precipitation for the 1980-2010 reference period for 18 RCMs before and after bias correction for the full area of the Republic of Cyprus (Sect. 2.2); and streamflow bias of the GR4J simulations with the RCM forcing relative to observed streamflow for the area of the 38 watersheds.

Pearson's correlation coefficient, denoting the correlation between each precipitation index and the streamflow bias for the 18 RCMs, is shown on the last line. In bold font, the nine models with W5R less than 10% difference from the observed W5R and streamflow relative error less than 15%.

	Precipitation					Precipitation					Streamflow
Dataset	Average annual (mm/y)	St. dev.[1] (mm/y)	W5R[1]	SDII[1] (mm/d)	R10mm[1] (days/y)	Average annual (mm/y)	St. dev. (mm/y)	W5R	SDII (mm/d)	R10mm (days/y)	Average annual ($10^6 m^3$)
CY-OBS	467	93	0.82	6.1	14	467	93	0.82	6.1	14	173
GCM/RCM pair	*Raw RCM data*					*Bias corrected RCM data (QDM)*					Relative error[2]
EC-EARTH/HIRHAM	452	90	0.84	5.1	13	477	96	**0.84**	6.2	14	**-0.01**
NorESM/RCA	609	128	0.75	4.9	19	460	122	**0.80**	6.0	14	**-0.02**
NorESM/RACMO	455	110	0.77	3.6	13	466	132	**0.80**	6.1	14	**0.04**
HADGEM/RACMO	562	112	0.76	3.8	16	474	107	**0.80**	6.2	14	**0.05**
HADGEM/HIRHAM	422	118	0.82	4.1	12	474	134	**0.83**	6.1	14	**0.07**
MPI/RACMO	509	100	0.79	3.8	14	452	103	**0.83**	6.0	14	**-0.08**
EC-EARTH/RACMO	459	75	0.79	3.8	13	472	86	**0.82**	6.2	14	**-0.09**
EC-EARTH/RCA	596	85	0.76	5.2	18	474	78	**0.80**	6.2	14	**-0.09**
MPI/RCA	617	155	0.75	5.0	19	461	141	**0.79**	6.0	14	**-0.11**
HADGEM/RCA	687	158	0.68	5.1	20	479	136	0.73	6.2	15	-0.05
CNRM/RACMO	517	111	0.67	3.5	14	470	127	0.71	6.1	14	-0.11
CNRM/ALADIN	840	149	0.67	6.0	27	476	112	0.70	6.2	14	-0.17
EC-EARTH/CLM	320	87	0.71	4.0	9	465	124	0.71	6.1	14	-0.19
MPI/REGCM	615	86	0.68	3.4	14	468	92	0.71	6.1	14	-0.31
MPI/REMO	476	98	0.75	4.2	12	454	88	0.70	6.0	14	-0.38
HADGEM/REGCM	561	78	0.64	2.9	11	479	96	0.67	6.2	14	-0.39
NorESM/REMO	386	101	0.66	4.1	10	463	117	0.60	6.0	14	-0.39
IPSL/REMO	275	83	0.69	3.5	7	474	129	0.61	6.1	14	-0.40
Pearson's r	0.21	0.39	0.67	0.32	0.39	0.13	0.25	0.86	0.23	0.15	

[1] Standard deviation (St.dev); Ratio of precipitation of the five wettest months to the annual precipitation (W5R); Simple daily intensity index (SDII); Number of days with daily precipitation exceeding 10 mm (R10mm)

[2] Relative error of modeled streamflow with RCM forcing relative to observed streamflow: (Qrcm.past-Qobs)/Qobs

3.4 Future projections of water resources

Precipitation in 2030-2060 is projected to decrease by 16% according to the driest model and by 6% according to the median model, relative to the 562 $mm \cdot y^{-1}$ over the 38 watersheds for the reference period (Table 3). Total streamflow of the 38 watersheds is projected to remain the same in the best model case or decrease up to 39%, in the worst case, relative to the 1980-2010 reference period value of 173 $Mm^3 \cdot y^{-1}$. The median projected change in streamflow is a reduction of 17%, and it is given by the NorESM-RACMO RCM. The projected change in streamflow magnitude

(14%-39%) was nearly double or more than double the change in precipitation (6%-16%) for six out of the nine models that provide the largest changes.

Table 3 shows the analysis of the 30-year annual series of precipitation for the driest years in the reference period and the driest years projected in the future period. According to these results, the driest 2-year period in the future could be from 3% up to 38% drier that the historical reference, with a median change of 11%. Streamflow in the two driest years could be 9% to 70% less in the future, with a median value of 36%. The top-5 driest years could have, on average, a median reduction of 17% in precipitation and 35% in streamflow. These reductions during these driest future periods are significantly greater than the projected 30-year long-term reductions, indicating that dry years in future will cause even more significant drought conditions than in the past.

A similar analysis for the wettest 2-year period and the top-5 wettest years shows a small increase in precipitation, with a median value of 15% and 4%, respectively, and nearly no change in streamflow relative to the streamflow of the wettest years of the past. These results indicate that, while very wet conditions are also expected in the future in some years, as in the past, the magnitude of reduction of water resources in the driest years in the future is disproportionally larger than the positive change in the wettest years.

Table 3. Annual averages of observed precipitation (CY-OBS) and streamflow of the 38 watersheds for 1980-2010, driest and wettest 2-year periods, and top-5 driest and wettest years, and relative changes for 2030-2060 versus 1980-2010, for nine RCMs.

		30-year average annual total	Driest 2-year period	Top 5 driest years	Wettest 2-year period	Top 5 wettest years	# watersheds with change in flow regime type
CY-OBS	**Precipitation**[1] ($mm \cdot y^{-1}$)	562	383	393	720	730	-
MPI/RACMO	Relative precipitation change	-0.02	-0.03	-0.07	0.18	0.22	
EC-EARTH/RACMO		-0.01	-0.10	-0.08	0.07	0.06	
MPI/RCA		0.00	-0.34	-0.20	0.04	0.15	
EC-EARTH/RCA		-0.07	-0.11	-0.09	0.20	0.03	
NorESM/RACMO		-0.06	-0.38	-0.22	0.28	0.05	-
HADGEM/RACMO		-0.09	-0.05	-0.11	0.15	0.00	
NorESM/RCA		-0.06	-0.09	-0.17	0.01	0.04	
EC-EARTH/HIRHAM		-0.16	-0.22	-0.24	0.22	-0.06	
HADGEM/HIRHAM		-0.16	-0.15	-0.21	0.10	-0.01	
Median		-0.06	-0.11	-0.17	0.15	0.04	-
OBS	**Streamflow** ($Mm^3 \cdot y^{-1}$)	173	53	51	348	345	
MPI/RACMO		0.06	-0.09	-0.15	0.30	0.52	5
EC-EARTH/RACMO		0.03	-0.36	-0.08	0.08	0.06	3

MPI/RCA	Relative streamflow change	0.00	-0.61	-0.32	-0.01	0.21	4
EC-EARTH/RCA		-0.14	-0.28	-0.32	0.02	0.02	12
NorESM/RACMO		-0.17	-0.34	-0.50	0.09	-0.03	8
HADGEM/RACMO		-0.25	-0.43	-0.35	0.00	-0.05	6
NorESM/RCA		-0.28	-0.36	-0.46	-0.24	-0.14	7
EC-EARTH/HIRHAM		-0.36	-0.70	-0.55	-0.05	-0.22	10
HADGEM/HIRHAM		-0.39	-0.70	-0.66	0.05	-0.12	9
Median		-0.17	-0.36	-0.35	0.02	-0.03	7

[1] Precipitation from the CY-OBS dataset refers to the precipitation over the 38 watershed areas, which is different from the CY-OBS precipitation over the area of the island in Table 2.

The future change in precipitation and reference evapotranspiration and the change in streamflow for the four flow type classes of the 38 watersheds are shown in Fig. 4(a). Results are presented for the three models with the minimum, median and maximum streamflow projected change among the nine RCMs. The HADGEM/HIRHAM model simulates the maximum streamflow change in the future, from -12% to -47%, for the median changes of each of the four flow types (as shown by the medians in the boxplot in Fig. 4(a)). The same model also projects the highest precipitation reduction (11%-17%) and ET increase (7%-9%) among the three models. For the median model, the NorESM/RACMO, with streamflow change up to 20% in the future for the four flow types, the ET increase is the least among the three models and equal to about 4%. Overall, the streamflow changes in the future from different RCMs have a variability attributed to the variability of both precipitation and ET.

The watersheds with permanent flow (P) and intermittent pool (I-p) flow regime, which contribute 71% in the total volume of streamflow of the 38 watersheds, are projected to exhibit the highest future reduction in streamflow, up to -47% and -42% respectively, relative to the reduction up to 20% for intermittent harsh (I-h) and up to 12% for ephemeral streams (E) according to HADGEM/HIRHAM RCM in Fig. 4(a). The median model, NorESM/RACMO and the model with the least projected streamflow change, MPI/RACMO, also show that P and I-p flow types will be subjected to the highest reductions relative to I-h and E flow regimes. The flow regime type will also become subject to change, implying a negative impact on the ecological conditions in the area of the 38 watersheds. Figure 4(b) shows that two watersheds with P, three with I-p and one with I-h flow regime have the highest probability (five or more models out of the nine) for a shift in their flow regime, while another one watershed with P, eight watersheds with I-p are and three with I-h are also projected to exhibit a change in flow regime, according to fewer than five out of the nine models. Overall, 18 out of the 38 watersheds are indicated to exhibit a change in their flow regime according to at least one of the nine RCMs.

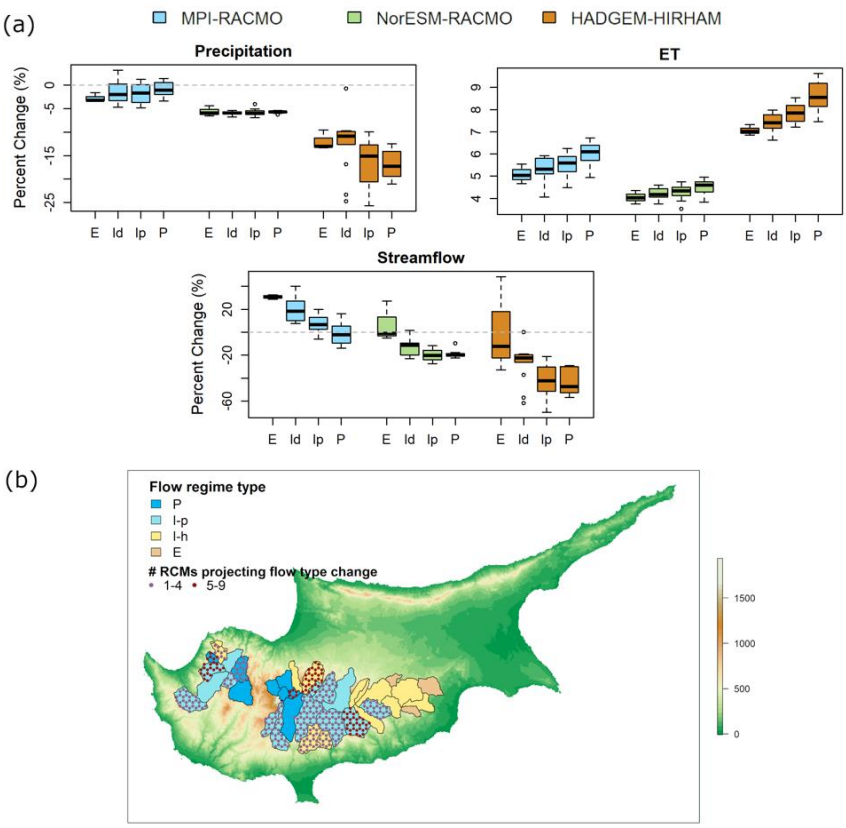


Figure 4. (a) Boxplots for the percent changes in annual precipitation, reference evapotranspiration (ET) and streamflow in the 2030-2060 period relative to the 1980-2010 period simulated by three RCMs (corresponding to minimum, median and maximum streamflow change among nine RCMs) for 38 watersheds grouped by their flow regime type (E-Ephemeral, 3 watersheds, Ih – Intermittent harsh, 13 watersheds, Ip – Intermittent pools, 14 watersheds, P – Permanent, 8 watersheds). (b) Elevation map of Cyprus (m) with the flow regime type of the 38 watersheds in the reference period (1980-2010) and the number of models out of the nine RCMs (i.e. 1 up to 4 (1-4) and 5 up to 9 (5-9)) projecting a change in the flow type in the future period (2030-2060).

4 Discussion

4.1 Selection of objective function and model parameterization for non-stationary climate conditions

This study examined the changes in the hydrological model performance when the model is used at contrasting climate periods with multiple performance measures used for the calibration and validation. Although the Differential Split Sample Test approach has been used by several studies in the context of model transferability, with different lengths in the examined sub-periods, the transferability performance was evaluated with different or fewer number of measures than those used for the model calibration. For instance, Dakhlaoui et al. (2017; 2019) used 7-8 year long sub-periods or sub-periods comprised of discontinuous years with KGE as the calibration measure and NSE and Volume error as the transferability evaluators. Guo et al. (2020) used the KGE for calibration of sub-periods separated by 3-month windows and validated the model using KGE and the three components of KGE. The evaluation of the

model calibrated with six objective functions in the current study provides a comprehensive evaluation of how KGE and NSE, with no transformation, the logarithmic and the square transformation of streamflow values affect model performance during calibration and validation under contrasting climate conditions. KGE was found to be the optimal measure for model optimization, for the calibration period only. The model calibrated with it ranked second in the validation periods with increased wetness ($\Delta P \geq 15\%$) and third or fourth in the full validation set and the validations with a small change in precipitation or increasing dryness ($-5\% < \Delta P \leq 5\%$ and $\Delta P \leq -5\%$). This outcome for the 5-year calibration runs of the study with KGE, a widely used objective function and evaluation measure, could be further investigated by assessing the model calibration and validation performance using the three components of KGE as objective functions, separately. Fowler et al. (2018) found that models calibrated with a split-KGE, i.e., computed individually per year and averaged, performed better than when KGE was computed for the full period.

The value of the composite score CSS, which normalizes and combines the values of seven evaluation measures, was highest for the NSEsqrt objective function. This finding agrees with Seiller et al. (2017), who found that the square root transformation with NSE is better, on average, for all flow regimes than calibration on NSE with inverse transformation or no transformation when the calibrated model is transferred to warmer conditions for catchments in southeastern Canada. These authors also showed that the square root transformation is optimal at the expense of some loss of confidence for very high flows. In addition to the best relative model performance with NSEsqrt as objective function, in warmer conditions, according to these authors, the current study showed that NSEsqrt was the optimal measure for warmer and wetter, as well as for warmer and drier conditions in the validation. The model runs calibrated with NSEsqrt ranked also first with increasing distance from the second highest ranked objective function with increasing precipitation reduction (ΔP) in the validation period. This result suggests that NSEsqrt becomes more robust for increasing precipitation reduction in the validation period, relative to the calibration period. The comparison of the hydrological model performance for 14 GR4J parameterizations with NSEsqrt for the different validation experiments showed that drier watersheds, i.e., with lower precipitation than their wetter counterparts in the same Mediterranean study area, are simulated with lower values and higher variability of the performance measure, among the different parameterizations. These findings are in line with the study of Munoz-Castro et al. (2023), who found that the choice of the objective function was more important for high-aridity and lower runoff coefficients than for wetter climates and that a higher model performance and higher parameter agreement was achieved for wetter basins in their study of 92 basins in continental Chile.

Dakhlaoui et al. (2017) found that a 25% difference in precipitation and 1.75 °C in temperature were the thresholds for acceptable model transferability for streamflow simulations with changing climate conditions. The difference in climate conditions of the calibration-validation periods, used for the comparison of the objective functions in the present study, had a maximum temperature change of 1.5°C and precipitation change range from -14% to +25%, except one calibration-validation set with precipitation change of 32%.

4.2 Bias correction and selection of RCMs

The long-term average precipitation data from the EURO-CORDEX ensemble, evaluated before and after bias correction (Table 2), shows that this was an essential step for obtaining values close to the observations. The good performance of the Quantile Delta Mapping (QDM) method used here was highlighted in other studies (e.g., Meyer et al., 2019). However, 9 out of 18 RCMs showed poor reproduction of the seasonal cycle of precipitation. This leads to the conclusion that if an RCM fails to reasonably represent the seasonality of the reference period, it may also not be able to model future shifts. Mascaro et al. (2018) found that an ensemble of raw EURO-CORDEX RCMs underestimated winter precipitation in Sardinia, which has a similar climate and geographical extent as Cyprus. These authors' finding is in line with lower simulated precipitation in the five wettest months (WR5) by the RCMs in the current study, which was found to be correlated with the errors in simulated streamflow.

Comparison of CMIP5 and CMIP6 models in previous global and regional studies showed that uncertainty in precipitation is high in both generations of models (Cos et al. 2022; Wu et al. 2024). For the Mediterranean, the IPCC Working Group I Interactive Atlas (Gutiérrez et al. 2021) reports a median precipitation change ranging from -7.9% to -18.3% for CMIP5 and a change from -7.1% to -17.8% for CMIP6 models for different global warming levels and for pathways RCP8.5/SSP5-8.5 (relative to the pre-industrial period). For the Eastern Mediterranean and specifically Turkey, Bagçaci et al. (2021) found that the precipitation decline is 2.5% smaller and the temperature increase is up to 0.35°C higher in CMIP6 compared to CMIP5. Despite the differences reported for the latest CMIP6- relative to CMIP5-GCMs, regionally downscaled CMIP6 models are not produced at the time of preparation of this manuscript, which does not allow a direct comparison with CMIP5 RCMs (EURO-CORDEX) used here. This work showed that bias correcting the RCM of CMIP models is necessary for the quantification of regional impacts of climate change, as shown by the errors with and without bias correction of RCM outputs for annual totals and seasonal distribution of precipitation, described in Sect. 3.3.

4.3 Future projections of water resources

The projected median change in future precipitation for 2030-2060 under RCP8.5, relative to 1980-2010, of the nine models (-6%, with a range from 2% to -16%), is in line with the overall observed drying of the island. These changes are within the range reported by other studies for the Mediterranean as a whole, about 10% in the first half and up to 20-40% in the second half of the 21st century (Zittis et al., 2021a). The projected median change of -17% for streamflow, with a range from 6% to -39% from the nine RCMs, indicates a pathway of decreasing freshwater resources by the mid-21st century for Cyprus. Projected streamflow changes for two catchments on the southern slopes of the Troodos mountains by 2050 (up to 24% and 17%), as presented by Ragab et al. (2010), are within the projections of the current study.

Dakhlaoui et al. (2019) reported for 2040-2070, relative to 1970-2000, median runoff changes up to -6.2% under RCP4.5 (scenario not examined here) and from -13% to -31% under RCP8.5 for different RCMs and basins in Tunisia. For other Mediterranean locations and the end-of-the-century horizon, which was not analyzed here, runoff reductions seem to be close to 20% under RCP4.5, i.e. 19% in southern Italy (Senatore et al., 2022; reference period 1975-2005)

and 16%-19% in Tunisia (Dakhlaoui et al., 2019). For the same horizon, relative to various 30-year periods in the second half of the 20th century, under the less optimistic, business-as-usual scenarios, i.e., B2 or RCP8.5, streamflow reductions will be exacerbated, i.e., 26%-54% in southern France (Lespinas et al., 2014), 50%-60% in central Spain (Sánchez-Gómez et al., 2023), and 37%-57% in Tunisia (Dakhlaoui et al., 2019).

The differences in the magnitude of precipitation and streamflow projected changes are comparable to the differences in the magnitudes of changes computed from two different periods in the past (1916-1969 and 1970-2000), as found in a study for the islands' water resources trends (Water Development Department, 2002). Part of this difference can be attributed to the increased reference ET in the watersheds in the future (4%-8%), driven by the robust warming (Zittis et al., 2021b; Lazoglou et al., 2024). The combined effect of increased ET and reduced streamflow suggests potential reductions in surface water availability, which may exceed the current natural replenishment of dams and groundwater. However, rainfall-runoff relations are non-linear because less runoff tends to be generated under drier watershed conditions. Additional pressures on streamflow include declines in groundwater levels, exacerbated by higher extractions for agricultural production in dry years (Zoumides et al., 2013; Leduc et al., 2017). Gutierrez et al. (2019) showed how the initial head in groundwater levels influences the onset of streamflow in Mediterranean climates. Previous modelling experience in Cyprus showed the sensitivity of simulated streamflow to the groundwater processes in hydrological modeling (Ragab et al., 2010; Camera et al., 2020; Sofokleous et al., 2023).

The change of flow regime type is most likely to occur for the wettest watersheds (Fig. 6). This was also found by Reymond et al. (2019) who related the shift from a continuous flow regime to intermittent to the lower precipitation and increased number of consecutive days without rain. Pascual et al. (2015) showed that from the comparison of flow changes in three watersheds in northeastern Spain, the two wettest watersheds were expected to experience larger reductions in streamflow (34%) than the drier watershed (25%). Schneider et al (2012) found that the most remarkable flow regime alterations in Europe are expected in the Mediterranean. The results of this climate impact assessment on the fresh water resources in Cyprus contribute to the broadening of the knowledge of changes in streamflow at watershed level in the climate hotspot region of the eastern Mediterranean.

As shown with the range of changes for precipitation and streamflow presented above, uncertainty in future projections, particularly in climate change impacts on streamflow, is large. Given that the impacts of climate change are derived from a chain of modelling steps, the uncertainty in the future projections for streamflow could be identified for each step. In this study, multiple RCMs were used to drive the hydrological simulations. The emphasis was placed on multiple RCMs rather than multiple hydrological models, as previous research has shown that for mid- to high-flow and mean annual flow conditions, variability in rainfall–runoff model outputs is greater when runoff projections are based on a single rainfall–runoff model combined with multiple climate models (Teng et al. 2012; Petheram et al. 2012). Other studies also showed that a significant source of uncertainty comes from the RCMs for studying climate impacts on streamflow (Teutschbein and Seibert 2012). RCMs exhibit systematic biases in both temperature and precipitation, influenced by the individual structures of the GCMs and RCMs. For this reason, a bias correction step is necessary (Christensen et al., 2008). In this study, 18 different RCMs were selected, representing combinations of six GCMs downscaled by seven RCMs.

A single hydrological model, GR4J, was used in this study, as the particular model is a well-established conceptual model structure for streamflow simulations across a wide range of hydroclimatic conditions. Its performance and robustness have been found to be comparable to those of a multimodel approach (Seiller et al., 2017). However, uncertainties also arise from the hydrological model structure used in impact studies. A particular limitation of "bucket-type" rainfall–runoff models, such as GR4J, is their tendency to underestimate multi-year drought conditions due to finite storage, which limits their ability to represent sustained groundwater decline (Fowler et al., 2020). The use of model parameters calibrated during observed dry conditions with trends similar to the projected climate, as suggested in this study, and structural modification of these models to account for catchment memory (e.g. Grigg and Hughes, 2018; Hughes et al., 2021) could help reduce this underestimation.

5 Conclusions

This study presented a framework of objective function selection for hydrological model calibration for use in climate impact assessments. A comparative method based on six objective functions and seven evaluation measures and a composite scaled score (CSS), which normalizes and combines multiple evaluation measures into one score, was used to evaluate the hydrological model performance optimized with the six different objective functions over multiple calibration and validation sets for non-stationary climate conditions.

The comparative method was applied to the GR4J hydrological model for 38 Mediterranean mountain watersheds, with average annual rainfall ranging between 373 and 868 mm and runoff coefficients between 0.01 and 0.51. The model parameterization obtained using the objective function of NSE with square root-transformed streamflow values (NSEsqrt) outperformed those obtained using NSE, NSElog, KGE, KGEsqrt, and KGElog during the validation period. The CSS showed that the model calibrated with KGE outperformed the model runs calibrated with the other functions only during the calibration period. The parameterization optimized with NSEsqrt also resulted in the best model performance under four different types of changing climate conditions. Specifically, with increasing dryness in the climate conditions, NSEsqrt stood out as the best objective function, showing greater performance differences from the others than under wetter or unchanged conditions. NSEsqrt can therefore be used as an objective function for streamflow simulations in Mediterranean watersheds experiencing drying trends. This method could also be applied to identify the most suitable objective functions under changing climate conditions in other environments. A comparison of the optimized parameter values obtained in different calibration periods and by different objective functions showed that the variability in optimized parameter values is greater when considering multiple objective functions, than when considering multiple calibration periods with varying climate conditions.

Comparison of validation periods with different precipitation changes relative to the calibration period showed higher total streamflow errors and degraded hydrological model performance, as described by NSE, KGE and different transformations of the streamflow with the two measures, for transitions to dry periods compared to the results for transitions to wet periods. Likewise, the watersheds with lower annual precipitation showed lower and more variable performance measures compared to the wettest watersheds of the study area.

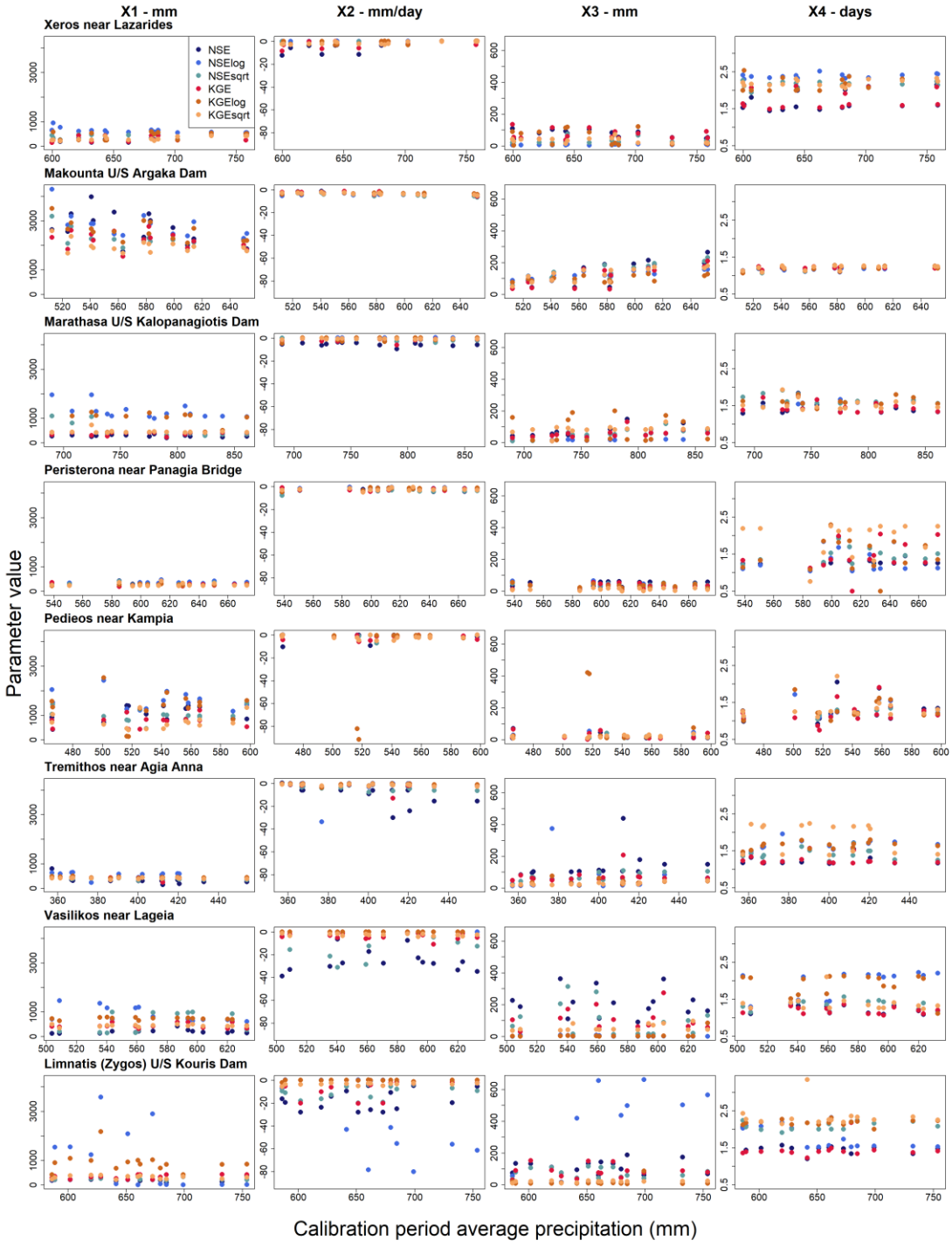
The evaluation of a large RCM ensemble showed that these models need to be bias-corrected, downscaled and evaluated for precipitation indicators and hydrological model performance when used for the assessment of climate impacts on freshwater resources. In particular, the comparison of the output of 18 RCMs and of the output of the hydrological models forced by these RCMs revealed that the correct simulation of the seasonal distribution of precipitation is correlated with the hydrological model's ability to achieve low streamflow errors. This criterion may help increase confidence in an RCM's ability to simulate future changes in both seasonal precipitation patterns and streamflow.

For Cyprus, precipitation projections under RCP8.5 scenario showed a median precipitation change of -6%, among nine RCMs, and a -16% change in the worst case for 2030-2060 relative to the 1980-2010 reference period. The streamflow reductions are amplified relative to the projected precipitation reductions. The projected median streamflow change is -17% and -39% in the worst case. The five driest years in the future are expected to become even drier than the past, from 9% up to 70% drier in terms of total streamflow for 38 watersheds of the island. This outcome highlights the importance of examining projected changes in specific future years, as they may reveal acute water shortages that must be considered in adaptation measures, which long-term average analysis alone might overlook.

Permanently flowing rivers were projected to be the most affected by reductions in streamflow, with a high risk of shifting from permanent to intermittent flow regimes. Such changes can result in reduced or even absent water flow during certain periods of the hydrological year, leading to declines in ecosystems that depend on the continuous flow of stream water. The pessimistic projections for the terrestrial water resources in 2030-2060 found in this study under the RCP8.5 scenario (business-as-usual scenario), are consistent with findings from other Mediterranean regions.

The study also highlighted the importance of maintaining reliable long-term streamflow observations for watersheds with different flow regime types. These observations are essential for parameterizing hydrological models and improving the land surface models embedded in climate models. Such models are needed to predict and project seasonal and future water resources at both regional and watershed scales and to develop climate adaptation plans and water management policies.

Appendix A



675 Figure A1: Values of the four parameters of the GR4J model optimized with six objective functions for the 14 five-year calibration periods for eight watersheds. The mean precipitation of each five-year calibration period is shown on the horizontal axis.

Data and code availability

The meteorological data used in this study are available at https://doi.org/10.5281/zenodo.17272790. Streamflow data of the reference period can be requested from the Water Development of Cyprus. EURO-CORDEX data are publicly available through the Earth System Grid Federation (https://esg-dn1.nsc.liu.se/) and the Copernicus Climate Change Services Data Store (https://cds.climate.copernicus.eu). The climate data bias correction method is available as an R package in the following link: https://cran.r-project.org/web/packages/MBC/index.html) and the hydrological model is available as an R package in: https://hydrogr.github.io/airGR/

Author contribution

IS and AB designed the simulation experiments; GZ and GD collected the data; IS analyzed the data, performed the simulations and wrote the manuscript; GZ, GD and AB reviewed and edited the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

For the meteorological data of the reference period used in the study, we would like to thank our colleagues from the Department of Meteorology of Cyprus. For the streamflow observations of the reference period, we would like to thank our colleagues from the Water Development Department of Cyprus. For the RCM data, we would like to acknowledge the CORDEX initiative. This research has received funding from the 3PRO-TROODOS Project (INTEGRATED/0609/061), co-financed by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation. This research was also supported by the PREVENT project, that has received funding from the European Union's Horizon Europe Research and Innovation Program under Grant Agreement No. 101081276.

References

Asadieh, B. and Krakauer, N. Y.: Global change in streamflow extremes under climate change over the 21st century, Hydrol. Earth Syst. Sci., 21, 5863–5874, https://doi.org/10.5194/hess-21-5863-2017, 2017.

Bağçaci, S. Ç., Yucel, I., Duzenli, E., and Yilmaz, M. T.: Intercomparison of the expected change in the temperature and the precipitation retrieved from CMIP6 and CMIP5 climate projections: A Mediterranean hot spot case, Turkey, Atmos. Res., 256, 105576, https://doi.org/10.1016/j.atmosres.2021.105576, 2021.

Broderick, C., Matthews, T., Wilby, R. L., Bastola, S., and Murphy, C.: Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods, Water Resour. Res., 52, 8343–8373, https://doi.org/10.1002/2016WR018850, 2016.

Camera, C., Bruggeman, A., Hadjinicolaou, P., Pashiardis, S., and Lange, M. A.: Evaluation of interpolation techniques for the creation of gridded daily precipitation (1× 1 km2); Cyprus, 1980–2010, J. Geophys. Res.-Atmos., 119, 693–712, https://doi.org/10.1002/2013JD020611, 2014.

Camera, C., Bruggeman, A., Zittis, G., Sofokleous, I., and Arnault, J.: Simulation of extreme rainfall and streamflow events in small Mediterranean watersheds with a one-way-coupled atmospheric–hydrologic modelling system, Nat. Hazards Earth Syst. Sci., 20, 2791–2810, https://doi.org/10.5194/nhess-20-2791-2020, 2020.

Camera, C., Zomeni, Z., Noller, J. S., Zissimos, A. M., Christoforou, I. C., and Bruggeman, A.: A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization, Geoderma, 285, 35–49, https://doi.org/10.1016/j.geoderma.2016.09.019, 2017.

Cannon, A. J.: MBC: Multivariate Bias Correction of Climate Model Outputs: November 12, 2024 Release (Version 0.10-7) [code], CRAN, available at: https://CRAN.R-project.org/package=MBC, 2024.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes?, J. Climate, 28, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1, 2015.

Christofi, C., Bruggeman, A., Kuells, C., and Constantinou, C.: Hydrochemical evolution of groundwater in gabbro of the Troodos Fractured Aquifer. A comprehensive approach, Appl. Geochem., 114, 104524, https://doi.org/10.1016/j.apgeochem.2020.104524, 2020.

Citrini, A., Lazoglou, G., Bruggeman, A., Zittis, G., Beretta, G. P., and Camera, C. A.: Advancing Hydrologic Modelling Through Bias Correcting Weather Radar Data: The Valgrosina Case Study in the Italian Alps, Hydrol. Process., 38, e15339, https://doi.org/10.1002/hyp.15339, 2024.

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., and Michel, C.: airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling (Version 1.7.6) [code], GitHub, available at: https://github.com/cran/airGR, 2023.

Cos, J., Doblas-Reyes, F., Jury, M., Marcos, R., Bretonnière, P. A., and Samsó, M.: The Mediterranean climate change hotspot in the CMIP5 and CMIP6 projections, Earth Syst. Dynam., 13, 321–340, https://doi.org/10.5194/esd-13-321-2022, 2022.

Da Silva, R. M., Santos, C. A., Moreira, M., Corte-Real, J., Silva, V. C., and Medeiros, I. C.: Rainfall and river flow trends using Mann–Kendall and Sen's slope estimator statistical tests in the Cobres River basin, Nat. Hazards, 77, 1205–1221, https://doi.org/10.1007/s11069-015-1644-7, 2015.

Dakhlaoui, H., Ruelland, D., and Tramblay, Y.: A bootstrap-based differential split-sample test to assess the transferability of conceptual rainfall-runoff models under past and future climate variability, J. Hydrol., 575, 470–486, https://doi.org/10.1016/j.jhydrol.2019.05.056, 2019.

Dakhlaoui, H., Ruelland, D., Tramblay, Y., and Bargaoui, Z.: Evaluating the robustness of conceptual rainfall-runoff models under climate variability in northern Tunisia, J. Hydrol., 550, 201–217, https://doi.org/10.1016/j.jhydrol.2017.04.032, 2017.

Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., Seo, K. W., and Western, A.: Many commonly used rainfall-runoff models lack long, slow dynamics: Implications for runoff projections, Water Resour. Res., 56, e2019WR025286, https://doi.org/10.1029/2019WR025286, 2020.

Fowler, K., Peel, M., Western, A., and Zhang, L.: Improved rainfall-runoff calibration for drying climate: Choice of objective function, Water Resour. Res., 54, 3392–3408, https://doi.org/10.1029/2017WR022466, 2018.

Gallart, F., Prat, N., García-Roger, E. M., Latron, J., Rieradevall, M., Llorens, P., Barberá, G. G., Brito, D., De Girolamo, A. M., Lo Porto, A., Buffagni, A., Erba, S., Neves, R., Nikolaidis, N. P., Perrin, J. L., Querner, E. P., Quiñonero, J. M., Tournoud, M. G., Tzoraki, O., Skoulikidis, N., Gómez, R., Sánchez-Montoya, M. M., and Froebrich, J.: A novel approach to analysing the regimes of temporary streams in relation to their controls on the composition and structure of aquatic biota, Hydrol. Earth Syst. Sci., 16, 3165–3182, https://doi.org/10.5194/hess-16-3165-2012, 2012.

Grigg, A. H. and Hughes, J. D.: Nonstationarity driven by multidecadal change in catchment groundwater storage: A test of modifications to a common rainfall–run-off model, Hydrol. Process., 32, 3675–3688, https://doi.org/10.1002/hyp.13282, 2018.

Gudmundsson, L., Boulange, J., Do, H. X., Gosling, S. N., Grillakis, M. G., Koutroulis, A. G., Leonard, M., Liu, J., Müller Schmied, H., Papadimitriou, L., Pokhrel, Y., Seneviratne, S. I., Satoh, Y., Thiery, W., Westra, S., Zhang, X., and Zhao, F.: Globally observed trends in mean and extreme river flow attributed to climate change, Science, 371, 1159–1162, https://doi.org/10.1126/science.aba3996, 2021.

Guo, D., Zheng, F., Gupta, H., and Maier, H. R.: On the robustness of conceptual rainfall-runoff models to calibration and evaluation data set splits selection: A large sample investigation, Water Resour. Res., 56, e2019WR026752, https://doi.org/10.1029/2019WR026752, 2020.

Gutiérrez, J. M., Jones, R. G., Narisma, G. T., Alves, L. M., Amjad, M., Gorodetskaya, I. V., Grose, M., Klutse, N. A. B., Krakovska, S., Li, J., Martínez-Castro, D., Mearns, L. O., Mernild, S. H., Ngo-Duc, T., van den Hurk, B., and Yoon, J.-H.: Atlas, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1927–2058, https://doi.org/10.1017/9781009157896.021, 2021.

Gutiérrez-Jurado, K. Y., Partington, D., Batelaan, O., Cook, P., and Shanafield, M.: What triggers streamflow for intermittent rivers and ephemeral streams in low-gradient catchments in Mediterranean climates, Water Resour. Res., 55, 9926–9946, https://doi.org/10.1029/2019WR025041, 2019.

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Appl. Eng. Agric., 1, 96–99, https://doi.org/10.13031/2013.26773, 1985.

Hoerling, M., Eischeid, J., Perlwitz, J., Quan, X., Zhang, T., and Pegion, P.: On the increased frequency of Mediterranean drought, J. Climate, 25, 2146–2161, https://doi.org/10.1175/JCLI-D-11-00296.1, 2012.

Hughes, J., Potter, N., Zhang, L., and Bridgart, R.: Conceptual Model Modification and the Millennium Drought of Southeastern Australia, Water, 13, 669, https://doi.org/10.3390/w13050669, 2021.

Jacob, D., Teichmann, C., Sobolowski, S., Katragkou, E., Anders, I., Belda, M., Benestad, R., Boberg, F., Buonomo, E., Cardoso, R. M., Casanueva, A., Christensen, O. B., Christensen, J. H., Coppola, E., De Cruz, L., Davin, E. L., Dobler, A., Domínguez, M., Fealy, R., Fernandez, J., Gaertner, M. A., García-Díez, M., Giorgi, F., Gobiet, A., Goergen, K., Gómez-Navarro, J. J., González Alemán, J. J., Gutiérrez, C., Gutiérrez, J. M., Güttler, I., Haensler, A., Halenka, T., Jerez, S., Jiménez-Guerrero, P., Jones, R. G., Keuler, K., Kjellström, E., Knist, S., Kotlarski, S., Maraun, D., van Meijgaard, E., Mercogliano, P., Montávez, J. P., Navarra, A., Nikulin, G., de Noblet-Ducoudré, N., Panitz, H.-J., Pfeifer, S., Piazza, M., Pichelli, E., Pietikäinen, J.-P., Prein, A. F., Preuschmann, S., Rechid, D., Rockel, B., Romera, R., Sánchez, E., Sieck, K., Soares, P. M. M., Somot, S., Srnec, L., Sørland, S. L., Termonia, P., Truhetz, H., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate downscaling over Europe: perspectives from the EURO-CORDEX community, Reg. Environ. Change, 20, 51, https://doi.org/10.1007/s10113-020-01606-9, 2020.

Ji, H. K., Mirzaei, M., Lai, S. H., Dehghani, A., and Dehghani, A.: The robustness of conceptual rainfall-runoff modelling under climate variability–A review, J. Hydrol., 621, 129666, https://doi.org/10.1016/j.jhydrol.2023.129666, 2023.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrol. Sci. J., 31, 13–24, https://doi.org/10.1080/02626668609491024, 1986.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, J. Hydrol., 424, 264–277, https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012.

Lazoglou, G., Hadjinicolaou, P., Sofokleous, I., Bruggeman, A., and Zittis, G.: Climate change and extremes in the mediterranean island of cyprus: from historical trends to future projections, Environ. Res. Commun., 6, 095020, https://doi.org/10.1088/2515-7620/ad7927, 2024.

Le Coz, M., Bruggeman, A., Camera, C., and Lange, M. A.: Impact of precipitation variability on the performance of a rainfall–runoff model in Mediterranean mountain catchments, Hydrol. Sci. J., 61, 507–518, https://doi.org/10.1080/02626667.2015.1051983, 2016.

Lespinas, F., Ludwig, W., and Heussner, S.: Hydrological and climatic uncertainties associated with modeling the impact of climate change on water resources of small Mediterranean coastal rivers, J. Hydrol., 511, 403–422, https://doi.org/10.1016/j.jhydrol.2014.01.033, 2014.

Martínez-Fernández, V., Van Oorschot, M., De Smit, J., González del Tánago, M., and Buijse, A. D.: Modelling feedbacks between geomorphological and riparian vegetation responses under climate change in a Mediterranean context, Earth Surf. Proc. Land., 43, 1825–1835, https://doi.org/10.1002/esp.4356, 2018.

Marx, A., Kumar, R., Thober, S., Rakovec, O., Wanders, N., Zink, M., Wood, E. F., Pan, M., Sheffield, J., and Samaniego, L.: Climate change alters low flows in Europe under global warming of 1.5, 2, and 3 °C, Hydrol. Earth Syst. Sci., 22, 1017–1032, https://doi.org/10.5194/hess-22-1017-2018, 2018.

Mascaro, G., Viola, F., and Deidda, R.: Evaluation of Precipitation From EURO-CORDEX Regional Climate Simulations in a Small-Scale Mediterranean Site, J. Geophys. Res.-Atmos., 123, 1604–1625, https://doi.org/10.1002/2017JD027463, 2018.

Masseroni, D., Camici, S., Cislaghi, A., Vacchiano, G., Massari, C., and Brocca, L.: The 63-year changes in annual streamflow volumes across Europe with a focus on the Mediterranean basin, Hydrol. Earth Syst. Sci., 25, 5589–5601, https://doi.org/10.5194/hess-25-5589-2021, 2021.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P. P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Clim. Change, 109, 213–241, https://doi.org/10.1007/s10584-011-0156-z, 2011.

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., and Cannon, A. J.: Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments, Hydrol. Earth Syst. Sci., 23, 1339–1354, https://doi.org/10.5194/hess-23-1339-2019, 2019.

Muñoz-Castro, E., Mendoza, P. A., Vásquez, N., and Vargas, X.: Exploring parameter (dis) agreement due to calibration metric selection in conceptual rainfall–runoff models, Hydrol. Sci. J., 68, 1754–1768, https://doi.org/10.1080/02626667.2023.2231434, 2023.

Myronidis, D., Ioannou, K., Fotakis, D., and Dörflinger, G.: Streamflow and hydrological drought trend analysis and forecasting in Cyprus, Water Resour. Manage., 32, 1759–1776, https://doi.org/10.1007/s11269-018-1902-z, 2018.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles, J. Hydrol., 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Pascual, D., Pla, E., Lopez-Bustins, J. A., Retana, J., and Terradas, J.: Impacts of climate change on water resources in the Mediterranean Basin: A case study in Catalonia, Spain, Hydrol. Sci. J., 60, 2132–2147, https://doi.org/10.1080/02626667.2014.947290, 2015.

Pedersen, J. S. T., Santos, F. D., van Vuuren, D., Gupta, J., Coelho, R. E., Aparício, B. A., and Swart, R.: An assessment of the performance of scenarios against historical global emissions for IPCC reports, Global Environ. Chang., 66, 102199, https://doi.org/10.1016/j.gloenvcha.2020.102199, 2021.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, J. Hydrol., 279, 275–289, https://doi.org/10.1016/S0022-1694(03)00225-7, 2003.

Petheram, C., Rustomji, P., McVicar, T. R., Cai, W., Chiew, F. H. S., Vleeshouwer, J., Van Niel, T. G., Li, L., Cresswell, R. G., Donohue, R. J., Teng, J., and Perraud, J.-M.: Estimating the impact of projected climate change on runoff across the tropical savannas and semiarid rangelands of northern Australia, J. Hydrometeorol., 13, 483–503, https://doi.org/10.1175/JHM-D-11-062.1, 2012.

Ragab, R., Bromley, J., Dörflinger, G., and Katsikides, S.: IHMS—Integrated Hydrological Modelling System. Part 2. Application of linked unsaturated, DiCaSM and saturated zone, MODFLOW models on Kouris and Akrotiri catchments in Cyprus, Hydrol. Process., 24, 2681–2692, https://doi.org/10.1002/hyp.7682, 2010.

Raymond, F., Ullmann, A., Tramblay, Y., Drobinski, P., and Camberlin, P.: Evolution of Mediterranean extreme dry spells during the wet season under climate change, Reg. Environ. Change, 19, 2339–2351, https://doi.org/10.1007/s10113-019-01526-3, 2019.

Refsgaard, J. C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T. A., Drews, M., Hamilton, D. P., Jeppesen, E., Kjellström, E., Olesen, J. E., Sonnenborg, T. O., Drote, D., Willems, P., and Christensen, J. H.: A framework for testing the ability of models to project climate change and its impacts, Clim. Change, 122, 271–282, https://doi.org/10.1007/s10584-013-0990-2, 2014.

Roudier, P., Andersson, J., Donnelly, C., Feyen, L., Greuell, W., and Ludwig, F.: Projections of future floods and hydrological droughts in Europe under a+ 2 C global warming, Clim. Change, 135, 341–355, https://doi.org/10.1007/s10584-015-1570-4, 2016.

Sánchez-Gómez, A., Martínez-Pérez, S., Leduc, S., Sastre-Merlín, A., and Molina-Navarro, E.: Streamflow components and climate change: Lessons learnt and energy implications after hydrological modeling experiences in catchments with a Mediterranean climate, Energy Rep., 9, 277–291, https://doi.org/10.1016/j.egyr.2022.11.191, 2023.

Schneider, C., Laizé, C. L. R., Acreman, M. C., and Flörke, M.: How will climate change modify river flow regimes in Europe?, Hydrol. Earth Syst. Sci., 17, 325–339, https://doi.org/10.5194/hess-17-325-2013, 2013.

Seiller, G., Roy, R., and Anctil, F.: Influence of three common calibration metrics on the diagnosis of climate change impacts on water resources, J. Hydrol., 547, 280–295, https://doi.org/10.1016/j.jhydrol.2017.02.004, 2017.

Senatore, A., Fuoco, D., Maiolo, M., Mendicino, G., Smiatek, G., and Kunstmann, H.: Evaluating the uncertainty of climate model structure and bias correction on the hydrological impact of projected climate change in a Mediterranean catchment, J. Hydrol. Reg. Stud., 42, 101120, https://doi.org/10.1016/j.ejrh.2022.101120, 2022.

Sofokleous, I., Bruggeman, A., and Camera, C.: The role of parameterizations and model coupling on simulations of energy and water balances–Investigations with the atmospheric model WRF and the hydrologic model WRF-hydro, J. Geophys. Res.-Atmos., 129, e2023JD040335, https://doi.org/10.1029/2023JD040335, 2024.

Sofokleous, I., Bruggeman, A., Camera, C., and Eliades, M.: Grid-based calibration of the WRF-Hydro with Noah-MP model with improved groundwater and transpiration process equations, J. Hydrol., 617, 128991, https://doi.org/10.1016/j.jhydrol.2022.128991, 2023.

Sofokleous, I., Bruggeman, A., Michaelides, S., Hadjinicolaou, P., Zittis, G., and Camera, C.: Comprehensive methodology for the evaluation of high-resolution wrf multiphysics precipitation simulations for small, topographically complex domains, J. Hydrometeorol., 22, 1169–1186, https://doi.org/10.1175/JHM-D-20-0110.1, 2021.

Teng, J., Vaze, J., Chiew, F. H., Wang, B., and Perraud, J. M.: Estimating the relative uncertainties sourced from GCMs and hydrological models in modeling climate change impact on runoff, J. Hydrometeorol., 13, 122–139, https://doi.org/10.1175/JHM-D-11-058.1, 2012.

Thirel, G., Andréassian, V., Perrin, C., Audouy, J.-N., Berthet, L., Edwards, P., Folton, N., Furusho, C., Kuentz, A., Lerat, J., Lindström, G., Martin, E., Mathevet, T., Merz, R., Parajka, J., Ruelland, D., and Vaze, J.: Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments, Hydrol. Sci. J., 60, 1184–1199, https://doi.org/10.1080/02626667.2014.967248, 2015.

Thrasher, B., Maurer, E. P., McKellar, C., and Duffy, P. B.: Bias correcting climate model simulated daily temperature extremes with quantile mapping, Hydrol. Earth Syst. Sci., 16, 3309–3314, https://doi.org/10.5194/hess-16-3309-2012, 2012.

Tramblay, Y., Koutroulis, A., Samaniego, L., Vicente-Serrano, S. M., Volaire, F., Boone, A., Le Page, M., Llasat, M. C., Albergel, C., Burak, S., Cailleret, M., Kalin, K. C., Davi, H., Dupuy, J. L., Greve, P., Grillakis, M., Hanich, L., Jarlan, L., Martin-StPaul, N., Martínez-Vilalta, J., Mouillot, F., Pulido-Velazquez, D., Quintana-Seguí, P., Renard, D., Turco, M., Türkeş, M., Trigo, R., Vidal, J. P., Vilagrosa, A., Zribi, M., and Polcher, J.: Challenges for drought assessment in the Mediterranean region under future climate scenarios, Earth-Sci. Rev., 210, 103348, https://doi.org/10.1016/j.earscirev.2020.103348, 2020.

Tuel, A., and Eltahir, E. A.: Why is the Mediterranean a climate change hot spot?, J. Climate, 33, 5829–5843, https://doi.org/10.1175/JCLI-D-19-0910.1, 2020.

Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J. M., Viney, N. R., and Teng, J.: Climate non-stationarity–validity of calibrated rainfall–runoff models for use in climate change studies, J. Hydrol., 394, 447–457, https://doi.org/10.1016/j.jhydrol.2010.09.018, 2010.

Vicente-Serrano, S. M., Lopez-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo, A., García-Ruiz, J. M., Azorin-Molina, C., Morán-Tejeda, E., Revuelto, J., Trigo, R., Coelho, F., and Espejo, F.: Evidence of increasing drought severity caused by temperature rise in southern Europe, Environ. Res. Lett., 9, 044001, https://doi.org/10.1088/1748-9326/9/4/044001, 2014.

Wagener, T., Reinecke, R., and Pianosi, F.: On the evaluation of climate change impact models, Wiley Interdiscip. Rev.: Clim. Change, 13, e772, https://doi.org/10.1002/wcc.772, 2022.

Udluft, P., Dünkeloh, A., Mederer, J., Kulls, C., and Schaller, J.: Re-evaluation of the groundwater resources of Cyprus - GRC project report, Geological Survey Department of Cyprus, Nicosia, Cyprus, 2006.

Water Development Department: Reassessment of the island's water resources and demands (TCP/CYP/8921), Ministry of Agriculture, Natural Resources and Environment of the Republic of Cyprus, Nicosia, Cyprus, available at: https://www.moa.gov.cy/moa/wdd/wdd.nsf/All/5F171454D0B5FC45C22583D20022586B/$file/1_4_1.pdf?OpenElement, 2002.

Wu, Y., Miao, C., Slater, L., Fan, X., Chai, Y., and Sorooshian, S.: Hydrological projections under CMIP5 and CMIP6: Sources and magnitudes of uncertainty, Bull. Amer. Meteor. Soc., 105, E59–E74, https://doi.org/10.1175/BAMS-D-23-0104.1, 2024.

Yang, W., Chen, H., Xu, C. Y., Huo, R., Chen, J., and Guo, S.: Temporal and spatial transferabilities of hydrological models under different climates and underlying surface conditions, J. Hydrol., 591, 125276, https://doi.org/10.1016/j.jhydrol.2020.125276, 2020.

Zhang, L., Potter, N., Hickel, K., Zhang, Y., and Shao, Q.: Water balance modeling over variable time scales based on the Budyko framework–Model development and testing, J. Hydrol., 360, 117–131, https://doi.org/10.1016/j.jhydrol.2008.07.021, 2008.

Zittis, G., Almazroui, M., Alpert, P., Ciais, P., Cramer, W., Dahdal, Y., Fnais, M., Francis, D., Hadjinicolaou, P., Howari, F., Jrrar, A., Kaskaoutis, D. G., Kulmala, M., Lazoglou, G., Mihalopoulos, N., Lin, X., Rudich, Y., Sciare, J., Stenchikov, G., Xoplaki, E., and Lelieveld, J.: Climate change and weather extremes in the Eastern Mediterranean and Middle East, Rev. Geophys., 60, e2021RG000762, https://doi.org/10.1029/2021RG000762, 2022.

Zittis, G., Bruggeman, A., and Lelieveld, J.: Revisiting future extreme precipitation trends in the Mediterranean, Weather Clim. Extremes, 34, 100380, https://doi.org/10.1016/j.wace.2021.100380, 2021.

Zittis, G., Hadjinicolaou, P., Almazroui, M., Bucchignani, E., Driouech, F., El Rhaz, K., Kurnaz, M. L., Nikulin, G., Hochman, A., Saaroni, H., Alpert, P., Önol, B., Batisati, M., and Lelieveld, J.: Business-as-usual will lead to super and ultra-extreme heatwaves in the Middle East and North Africa, npj Clim. Atmos. Sci., 4, 20, https://doi.org/10.1038/s41612-021-00178-7, 2021.