Sensitivity analysis of runoff modeling to statistical downscaling models in the western Mediterranean

Benjamin Grouillet[1], Denis Ruelland[1], Pradeebane Vaittinada Ayar[2], and
Mathieu Vrac[2]

[1]CNRS, Laboratoire HydroSciences, Place Eugene Bataillon, 34095 Montpellier, France
[2]LSCE, Laboratoire des Sciences du Climat et de l'Environnement, UMR CEA-CNRS-UVSQ
1572, CE Saclay l'Orme des Merisiers, 91191 Gif-sur-Yvette, France

Correspondence to: B. Grouillet (b.grouillet@gmail.com), D. Ruelland (denis.ruelland@um2.fr)

Abstract. This paper analyzes the sensitivity of a hydrological model to different methods to statistically downscale climate precipitation and temperature over four western Mediterranean basins illustrative of different hydro-meteorological situations. The comparison was conducted over a common 20-year period (1986–2005) to capture different climatic conditions in the basins. The daily GR4j conceptual model was used to simulate streamflow that was eventually evaluated at a 10-day time step. Cross-validation showed that this model is able to correctly reproduce runoff in both dry and wet years when high-resolution observed climate forcings are used as inputs. These simulations can thus be used as a benchmark to test the ability of different statistically downscaled datasets to reproduce various aspects of the hydrograph. Three different statistical downscaling models were tested: an analog method (ANALOG), a stochastic weather generator (SWG) and the "cumulative distribution function – transform" approach (CDFt). We used the models to downscale precipitation and temperature data from NCEP/NCAR reanalyses as well as outputs from two GCMs (CNRM-CM5 and IPSL-CM5A-MR) over the reference period. We then analyzed the sensitivity of the hydrological model to the various downscaled data via five hydrological indicators representing the main features of the hydrograph. Our results confirm that using high-resolution downscaled climate values leads to a major improvement of runoff simulations in comparison to the use of low-resolution raw inputs from reanalyses or climate models. The results also demonstrate that the ANALOG and CDFt methods generally perform much better than SWG in reproducing mean seasonal streamflow, inter-annual runoff volumes as well as low/high flow distribution. More generally, our approach provides a guideline to help choose the appropriate statistical downscaling models to be used in climate change impact studies to minimize the range of uncertainty associated with such downscaling methods.

1 Introduction

Climate Change Impact Studies (CCIS) focusing on water resources have become a hot topic in the last decade. However, such studies need reliable climate simulations to drive hydrological models

25 efficiently. General circulation models (GCMs) have demonstrated significant skills in simulating climate variables at continental and hemispherical scales but are inherently incapable of representing the local sub-grid-scale features and dynamics required for regional impact analyses. For most hydrologically relevant variables (precipitation, temperature, wind speed, humidity, etc.), GCMs currently do not provide reliable information at scales that are appropriate for impact studies (e.g.

30 Maraun et al., 2010). The mismatch between the spatial resolution of the GCM outputs and that of the data required for hydrological models is a major obstacle (e.g. Fowler et al., 2007). Some post-processing is thus required to improve these large-scale models for impact studies and downscaling methods have been developed to meet this requirement.

Downscaling methods can be dynamical or statistical, both approaches being driven by GCMs

35 or reanalysis data. Dynamical downscaling methods correspond to the so-called "Regional Climate models" (RCMs), aiming at generating detailed regional and local information (from a few dozen km down to a few km) from low-resolution simulations (generally with a horizontal resolution ranging from 100 to 300 km) by simulating high-resolution physical processes consistent with the required large-scale dynamics. Easier and less costly to implement as compared to dynamical downscaling

40 techniques, statistical downscaling models (SDMs) are also used in anticipated hydrologic impact studies under climate change scenarios (for a review, see e.g. Fowler et al., 2007). SDMs rely on determining statistical relationships between large- and local-scale variables and do not try to solve the physical equations that model atmospheric dynamics. Due to their statistical formulation, they generally have a low computational cost and provide simulations relatively rapidly. SDMs are based

45 on a static relationship, i.e. the mathematical formulation of the relation between predictands (i.e. the local-scale variable to be simulated) and predictors (i.e. the large-scale information or data used as inputs in the SDMs) has to be valid not only for the current climate on which the relationship is calibrated, but also for future climates, for example. Most state-of-the-art SDMs belong to one of the four following families (Vaittinada Ayar et al., 2015): "transfer functions", "weather typing",

50 methods based on "stochastic weather generators" and "Model Output Statistics" (MOS) models, which generally work on cumulative distribution functions (CDFs). Many studies demonstrated that caution is required when interpreting the results of climate change impact studies based on only one downscaling model (e.g. Chen et al., 2011). It is thus recommended to use more than one SDM to account for the uncertainty of the downscaling (e.g. Chen et al., 2012). However, uncertainty can be

55 very high due to the inability of some SDMs to realistically reproduce the local climate, and this can be critical when the aim is to produce accurate inputs for hydrological models at the basin scale in the context of CCIS. On the other hand, a sensitivity analysis of hydrological modeling to different downscaling methods can produce an indicator to assess the quality of downscaled climate forcings via their ability to generate reasonable simulations of discharge from hydrological modeling. This

60 analysis can also help to quantify the impact of the error in a runoff simulation that stems from SDMs.

Several works have already attempted to compare climate simulations, downscaled or not, from a hydrological point of view. Although these studies revealed significant differences between SDMs on hydrological responses including seasonal variability of runoff (e.g. Dibike and Coulibaly, 2005; Prudhomme and Davies, 2009; Chen et al., 2012; Teng et al., 2012), interannual discharge dynamics (e.g. Wood et al., 2004; Salathé, 2005), or the distribution of extreme events (e.g. Diaz-Nieto and Wilby, 2005), they were not able to clearly conclude on how to choose one method over another. Difficulties in selecting among different SDMs may arise from the choice of relevant criteria. While some may be appropriate from the statistical or climatological point of view, these criteria may not adequately highlight the differences between the methods with respect to the hydrological responses. As a result, the aforementioned studies generally suggest an ensemble approach including several methods to offer a range of downscaling uncertainty when studying climate change impact on runoff. However, this uncertainty range can be reduced to a minimum if inappropriate statistical downscaling methods are excluded from the ensemble approach.

Our analysis of the literature revealed that no consensus has emerged on the best downscaling techniques among the state-of-the-art SDMs in the context of CCIS on runoff. This calls for an original protocol to assess the strengths and weaknesses of the different SDMs in providing accurate hydrological simulations according to different insights. Indeed, assessing water resource availability for different uses requires accounting for different aspects of the hydrograph including interannual runoff volumes, mean seasonal streamflow, and low/high flow distribution. First, hydrologists need to correctly reproduce the interannual water balance in order to evaluate changes in the storage capacity of the hydrosystems, for instance. Second, analysis of the interannual variability of flows makes it possible to test the ability of the climate simulations to reproduce the occurrence of dry and wet years, as well as the frequency and intensity of change. Third, surface water resources can be evaluated through a seasonal analysis so as to focus on intra-annual high and low flow events. While high flows are particularly important, e.g. when the focus is on flood risk, low flows are generally studied in connection with the water needed for agriculture and tourism, as in these cases, there is generally an increase in water demand when flows are low (see e.g. Fabre et al., 2015; Grouillet et al., 2015). Consequently, assessing water availability means focusing on low flows, which generally occur during peak water demand.

Water resource issues are particularly important in the Mediterranean region, which has been identified as a hot-spot of climate change (Giorgi, 2006). The western Mediterranean basins are of particular interest since they are characterized by complex and varying hydro-climatic conditions due to the contrasted influences of the Atlantic Ocean and the Mediterranean Sea, and of mountain ranges. These contrasted conditions offer an opportunity to account for the uncertainty linked to the differences in spatial and temporal patterns that may arise from one downscaling technique to another.

3

The aim of this study is to propose a method to analyze the sensitivity of hydrological responses to different methods used to statistically downscale climate values by means of criteria that are commonly used in CCIS to assess the impact on water resources: volume of water flow, interannual and seasonal variability of runoff, distribution of extreme events including high and low flows. We compare statistical downscaling methods via a guideline aimed at providing an overview of their capabilities to reproduce the main features of the hydrograph in view of their use in CCIS.

The rest of this article is organized as follows. In section 2 we describe the basins in the western Mediterranean and a hydro-climatic analysis based on the available data. In section 3, we provide an overview of downscaling models and of the steps involved in hydrological modeling. In section 4, we summarize the results for each hydrological indicator, and in section 5 we discuss these results and provide a short conclusion.

2   Study areas and hydro-climatic context

2.1   Four catchments in the western Mediterranean

Four catchments were chosen to account for the variety of hydro-climatic conditions in the western Mediterranean region (Fig. 1): the Herault basin at Laroque (910 km^2, France), the Segre basin at Seo de Urgel (1 265 km^2, Spain), the Irati basin at Liedena (1 588 km^2, Spain) and the Loukkos basin at Makhazine (1 808 km^2, Morocco). These basins were also chosen because they are located upstream from storage dams and in areas in which withdrawals are negligible (Ruelland et al., 2015), so their streamflow regime can be considered as natural. For brevity's sake, the basins are referred to as Herault, Segre, Irati and Loukkos.

The Herault basin, from 165 to 1 565 m asl. comprises two-thirds karstified limestone favoring delayed and sometimes sudden restitution and one third of basement rocks with low groundwater reserves favoring surface runoff. The mountainous basin of Segre, located upstream from the Ebro basin in northern Spain from 670 to 2 830 m asl., is characterized by basement rocks (granite and quartzite) and a rugged topography that favors runoff. The Irati basin, from 407 to 2 017 m asl., is located upstream from the Ebro basin. This mountainous catchment, composed mainly of limestone and conglomerate, is characterized by a high upstream-downstream topographic gradient, favoring a rapid hydrological response. The Loukkos basin, from 55 to 1 668 m asl., is characterized by sandstone and marl successions favoring surface runoff.

2.2   Hydro-climatic data

Preliminary studies (Tramblay et al., 2013; Fabre et al., 2015; Ruelland et al., 2015) provided daily hydro-climatic data (precipitation, temperature and streamflow) over a common 20-year period (1986–2005), thus making it possible to compare the basins. Climate data for the Herault basin were extracted from the SAFRAN 8×8 km meteorological analysis system (Vidal et al., 2010) and

4

observed runoff was provided by the French Ministry of ecology and sustainable development from their database *Banque Hydro* (MEDDE, 2010). As mentioned by Vidal et al. (2010), SAFRAN is a gauge-based analysis system using the Optimal Interpolation (OI) method described by Grandin (1965). This method has been found to outperform other objective techniques for precipitation notably in studied in France over the Cévennes area, a region with very high spatial and temporal variability (Creutin and Obled, 1982). Climate data for the Segre and Irati basins were obtained by interpolating daily precipitation and temperature measurements on an 8×8 km grid with the inverse distance weighted (IDW) method (Shepard, 1968). This method is particularly efficient for gauge-based analyses of global daily precipitation (Chen et al., 2008). The precipitation and temperature data were extracted based on numerous stations available at the Ebro basin scale (Dezetter et al., 2014), of which around 19 and 6 precipitation stations, and 10 and three temperature stations concern the Irati and Segre catchments respectively. Elevation effects on temperature distribution were taken into account using a digital elevation model and a lapse rate of $-6.65°C/1\,000$ m estimated from the data. Daily streamflow data were provided by the Center of studies and experiments on hydraulic systems (CEDEX, 2012). In the Loukkos basin, precipitation data were interpolated on a 8×8 km grid based on 11 stations using the IDW method. Since daily temperature data were only available from a station located at the basin outlet, a universal lapse rate of $-6.5°C/1\,000$ m was used for temperature interpolation. Hydro-climatic data including daily streamflow were provided by the Moroccan *Département de Planification des Ressources en Eau* (DPRE). Due to the lack of additional data such as wind and humidity in the Moroccan basin, a simple formula relying on solar radiation and temperature was chosen (Oudin et al., 2005) to assess daily potential evapotranspiration (PE) in each basin.

The atmospheric variables used for the calibration of the SDMs as predictors were selected from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) daily reanalysis data (Kalnay et al., 1996) with a 2.5° spatial resolution, from January 1, 1976 to December 31, 2005. The variables covered the region $[-15°E; 42.5°E] \times [27.5°N; 50°N]$ encircling the Mediterranean Sea as defined in Vrac and Yiou (2010) and corresponding to 240 grid cells. For the temperature models, five predictors were used: the temperature at 2 m (T2), the sea level pressure (SLP), as well as the geopotential height and the zonal and meridional wind components at 850 hPa (respectively Z850, U850 and V850). For precipitation models, the same five predictors were used, and the dew point temperature at 2 m (D2) was added. The predictors and the pre-processing of those predictors according to the SDM and the predictands are summarized in the table 1. Calibration was performed over the usual four seasons in the northern hemisphere. The calibrated SDMs were forced with three different datasets: NCEP reanalysis data over the 1976–2005 calibration period and with the IPSL-CM5A-MR (from the French "Institut Pierre Simon Laplace", IPSL Climate Modelling Centre, Dufresne et al., 2013) and CNRM-CM5 (from the French National Centre for Meteorological Research, CNRM, Voldoire et al., 2013) GCMs, regridded at a 2.5° spa-

5

tial resolution, over the GCMs historical (or CTRL) period (i.e. 1986–2005).The regridding was done through a bilinear interpolation in order to have the GCMs and NCEP data at the same resolution. This is a requirement in order to use GCMs as predictors in the different SDMs calibrated from NCEP at a 2.5° resolution. Over the mid-latitudes, 2.5° correspond approximately to 250km. The Herault, Segre and Loukkos basins are included in a single GCM grid cell. The Irati basin straddles two grid cells, split equally. Also, the basins are not on the edge of the GCM grid and therefore are not subject to border effects in interpolation.

The SDMs have been calibrated over a 30-year period (1976–2005) for the Herault, Irati and Segre basins and a 20-year period (1986–2005) for the Loukkos due to data availability before 1986. This choice results from the need to use the maximum available time period for the SDM calibrations to have them as robustly calibrated as possible. However, the GCM historical period was defined over 1986–2005 in order to have a 20-year common period for all the SDMs to be evaluated through their ability to provide reliable hydrological simulations.

2.3 Hydro-climatic analysis

The four basins are characterized by a more or less pronounced Mediterranean climate with low precipitation in summer and more abundant precipitation in winter (see Fig. 1). Mean annual precipitation decreases from north to south, from 1 397 mm in the Herault basin to 935 mm in the Loukkos basin. Mean annual precipitation in the Segre basin (813 mm) is low compared to neighboring basins because of the rain shadow effect of the mountains surrounding the basin, which often stops precipitation from the Atlantic (West) as well as from the Mediterranean sea (East). Summer is hot and dry, especially in the Loukkos basin, which causes severe low flows during this season. In contrast, winter is milder and wetter. In the Herault and the Irati basins, the precipitation peaks in spring and fall are produced by precipitation events whose intensity can vary greatly over short periods. The spring and fall streamflows are strongly influenced by these precipitation events as well as by snowmelt in spring in the mountainous basins (mostly in the Segre and the Irati basins).

No significant trends in interannual variations in precipitation and streamflow were observed in the four basins over the period 1986–2005. Nevertheless, mean precipitation during the first 10 years of the study period was from 4% to 19% higher than during the last 10 years, except in the Segre basin (–3%). Furthermore, the analysis of the precipitation indices (Eq. 1) showed that the wet and dry years observed in the four basins occurred at the same time in nearly half the years (grey lines in Fig. 2). Mean annual temperature remained almost constant during the 1986–2005 period and the temperature indices (Eq. 2) were the same in the four basins in two thirds of the years (Fig. 2).

$$I_P = (P_y - \overline{P_y})/\sigma_P \tag{1}$$

$$I_T = (T_y - \overline{T_y})/\sigma_T \tag{2}$$

6

where P_y is the annual precipitation for the year y, $\overline{P_y}$ is the mean of the annual precipitation, σ_P is the standard deviation of the annual precipitation. T_y is the annual temperature for the year y, $\overline{T_y}$ is the mean of the annual temperature, σ_T is the standard deviation of the annual temperature.

3 Models and evaluation procedures

3.1 Statistical downscaling models

Based on the preliminary climatological study of Vaittinada Ayar et al. (2015), three downscaling methods were retained according to their ability to reproduce commonly used climatic patterns on E-OBS (Haylock et al., 2008) grid scale (0.44° or approximatively 50 km spatial resolution). These SDMs were thus used to provide the climate data, i.e. precipitation and temperature, used as inputs for the hydrological model at the basin scale. For each variable, three models were calibrated and applied: analogs of atmospheric circulation patterns (ANA), the "cumulative distribution function – transform" approach (CDFt) and a stochastic weather generator (SWG). The analog method and the stochastic weather generator are both calibrated and run on a seasonal basis, using the usual four seasons of the northern hemisphere, whereas the CDFt approach is run on a monthly basis. For ANA-LOG and SWG, the calibration was performed on NCEP reanalysis. Conversely, for CDFt, coming from the family of the bias correction (BC) techniques, the calibration was performed directly on the GCM to downscale. Although CDFt is derived from the quantile-mapping technique, none of the three SDMs is bias corrected. Those three models (i.e., CDFt included) have all the particularity of providing high-resolution precipitation and temperature simulations (constrained by large-scale re-analysis or GCM data) and therefore belong all to the family of the statistical downscaling methods. For all the three models, calibration was done over 1976–2005 (except for Loukkos on which data availability limited the calibration to 1986–2005). Their assessment when applied to NCEP reanalysis and GCM data was performed according to a common 20-year 1986–2005 evaluation period. Sections 3.1.1 to 3.1.3 describe the different models.

3.1.1 The Analog model

The "analogs" model used here is based on the approach of Yiou et al. (2013) and applied on the fields of anomalies fields over the Mediterranean region $[-15°E; 42.5°E] \times [27.5°N; 50°N]$ as defined in section 2.2. For any given day to be downscaled in the validation period, it consists in determining the day in the calibration period with the closest large-scale atmospheric situation X_{ANA}. More precisely, for a given day, the analog is taken from the 15 days before and after this date in the calibration data set. Note that the days in the same year are excluded. Therefore, this prevents the analog day to be too close (in time) to the day to be downscaled. The closest large-scale atmospheric situation X_{ANA} is determined by minimizing a distance metric (here the Euclidian distance) between the large-scale situation (X_d) of the day to be downscaled and the large-scale situation (X_c) of all the

7

days in the calibration period. More technically, this can be written as:

$$X_{ANA} = argmin(dist(X_d, X_c))$$ (3)

where *argmin(f)* is the function returning the minimum value of a function *f*, here computed over all the X_c situations of the calibration period. The daily large-scale atmospheric situations correspond to the daily fields of anomalies of the predictors. Those anomalies were calculated with respect to the seasonal cycle, as is classically done in analog techniques, see e.g., Yiou et al. (2013) and references therein. X_d – the large-scale situation of the day to be downscaled – corresponds to the fields of anomalies of all the predictors of that day. X_c corresponds to any large-scale situation (defined in the same way) in the calibration period. Hereafter this model is referred to as ANA.

3.1.2 The CDFt model

The "cumulative distribution function – transform" (CDFt) method was originally developed by Michelangeli et al. (2009) to downscale wind velocity and was later applied to temperature and precipitation, in, for example Vrac et al. (2012) and Vigaud et al. (2013). The CDFt model is a quantile-mapping-based approach, which consists in relating the local-scale cumulative distribution function (CDF) of the variable of interest to the large-scale CDF (here from NCEP or GCMs) of the same variable. Let $F_{Gc}(x)$ and $F_{Oc}(x)$ define the CDFs of the variable of interest, respectively from a GCM (subscript G) and from local-scale observations-based dataset (subscript O) over the calibration period (subscript c), and $F_{Gv}(x)$ and $F_{Ov}(x)$ the CDFs over the validation period (subscript v). First, CDFt estimates $F_{Ov}(x)$ as:

$$F_{Ov}(x) = F_{Oc}\left(F_{Gc}^{-1}(F_{Gv})\right)$$ (4)

with *x* in the range of the physical variable of interest. Then, a quantile-mapping between F_{Gv} and F_{Ov} is performed to retrieve the physical variable of interest at the local scale. All the technical details on Eq. (4) and subsequent quantile-mapping can be found in Vrac et al. (2012). Note that for this method, only the variable of interest (i.e. precipitation or temperature) at a large scale is used as predictor. Contrary to ANALOG and SWG, the CDFt approach comes from the family of the bias correction (BC) techniques. In that sense, CDFt does not need NCEP reanalyses for its calibration but is directly calibrated to link GCM simulations and high-resolution data (through their CDF). Note that CDFt is used here as a downscaling technique and not a BC, since it is applied here to downscale (i.e., to go from large-scale to high-resolution) temperature and precipitation time series.

3.1.3 The Stochastic Weather Generator model

The stochastic weather generator (SWG) model used in this study is based on conditional probability distribution functions in a vector generalized linear model (VGLM) framework, as in Chandler and Wheater (2002). This means that the distribution family is fixed and the distribution parameters are estimated as functions of the selected predictors.

8

Modeling precipitation is usually divided into two steps: first the occurrence and second the intensity. The modeling of intensity has been introduced in previous sections. The rain occurrence at a given location is modeled as a binomial distribution $B(1,p)$ using a logistic regression (LR, e.g. Buishand et al., 2004; Fealy and Sweeney, 2007). Let p_i be the probability of rainfall on day i conditionally on an N-length predictor (or covariate) vector $X_i = (X_{i1}, \cdots, X_{iN})$ as defined in the previous section. The conditional probability of occurrence p_i is formulated through a LR as:

$$\log\left(\frac{p_i}{1-p_i}\right) = p^0 + \overbrace{\sum_{j=1}^{N} p^j X_{i,j}}^{S=} \tag{5}$$

$$p_i = \frac{\exp(S)}{1 + \exp(S)} \tag{6}$$

where $(p_0, \cdots, p_N)$ is the vector of coefficients to be estimated. The LR is only used for SWG. The analog and CDFt models directly provide zeros or positive precipitation values.

Temperature is expected to follow a Gaussian distribution and rain intensity a Gamma distribution. The mean μ and the standard deviation σ of the Gaussian distributions and the shape α and the rate β of the Gamma distributions are estimated as functions of the large-scale predictors. The parameters σ, α and β at day i are computed with a common formulation, illustrated here for the α parameter:

$$\log(\alpha_i) = \alpha^0 + \sum_{j=1}^{N} \alpha^j X_{i,j} \tag{7}$$

with $(\alpha^j)_{j=0,\cdots,N}$ the regression coefficients to be estimated, N the number of predictors, and $X_{i,j}$ the jth daily large-scale predictor for day i. Note that Eq. (7) models the logarithm of the parameter of interest to ensure that the parameter obtained (σ, α or β) is positive. The parameter μ is formulated in the same way but without the positivity (i.e. log) constraint:

$$\mu_i = \mu^0 + \sum_{j=1}^{N} \mu^j X_{i,j} \tag{8}$$

As in Vaittinada Ayar et al. (2015), the predictors used for this model are the two first principal components (PCs) calculated from a principal component analysis (PCA, Barnston and Livezey, 1987) applied separately to each variable.

3.2 Hydrological simulations

3.2.1 Hydrological model

The GR4j lumped conceptual model (Perrin et al., 2003), was chosen to simulate the seasonal and interannual variations in runoff at a daily time step (see Fig. 3). Many studies have demonstrated the ability of the model to perform well under a wide range of hydro-climatic conditions (e.g. Perrin

et al., 2003; Vaze et al., 2010; Coron et al., 2012) and notably in the Mediterranean region (e.g. Tramblay et al., 2013; Fabre et al., 2015; Ruelland et al., 2015). This model relies on precipitation (*P*) and potential evapotranspiration (*PE*) and is based on a production function that determines the effective precipitation (the fraction of the precipitation involved in runoff) that supplies the production reservoir and on a routing function based on a unit hydrograph. According to the available data (cf Section 2.2), a simple formula relying on solar radiation and temperature (cf Eq. 9) was chosen (Oudin et al., 2005) to assess daily potential evapotranspiration (*PE*).

$$PE = \frac{R_e}{\lambda \rho} \times \frac{T+5}{100} \qquad if \ (T+5) > 0 \qquad else \qquad PE = 0 \tag{9}$$

where R_e is the extraterrestrial solar radiation $(\mathrm{MJ/m^2/d})$ given by the Julian day and the latitude, λ net latent heat flux (2,45 MJ/kg), ρ water density $(\mathrm{kg/m^3})$ and T is the mean air temperature at a 2 m height (°C).

Four parameters are used in the GR4j basic version: the maximum capacity of the soil moisture accounting store *x1*, a groundwater exchange coefficient *x2*, the maximum capacity of routing storage *x3*, and a time base for unit hydrographs *x4*. A three-parameter snow module based on catchment-average areal temperature (Ruelland et al., 2011, 2014) was activated to account for the contribution of snow to runoff from the catchments. Below a temperature threshold *x5*, a fraction *x6* of precipitation is considered as snowfall; this fraction feeds the snow reservoir. Above the threshold *x5*, a fraction *x7*, weighted by the difference between the daily temperature and the threshold *x5*, is taken from the snow reservoir to represent snowmelt runoff.

3.2.2 Optimization of hydrological simulations

The model parameters were calibrated and the simulation performances were analyzed by comparing simulated and observed streamflow at a 10-day time step (averaged from daily streamflow outputs) in a multi-objective framework. This time step was retained because it constitutes an interesting compromise for CCIS on water resources, between a daily time step useful to represent small runoff effects and a monthly time step too coarse to capture hydrological variability. The following objectives were considered: (*i*) the overall agreement of the shape of the hydrograph via the Nash-Sutcliffe efficiency (NSE) metric (Nash and Sutcliffe, 1970); (*ii*) the agreement of the low flows via a modified, log version of the NSE criterion; and (*iii*) the agreement of the runoff volume via the cumulated volume error (VE_C) and the mean annual volume error (VE_M).

$$NSE = 1 - \left\{ \sum_{t=1}^{N} \left(Q_{obs}^t - Q_{sim}^t \right)^2 \Big/ \sum_{t=1}^{N} \left(Q_{obs}^t - \overline{Q_{sim}} \right)^2 \right\} \tag{10}$$

$$NSE_{log} = 1 - \left\{ \frac{\sum_{t=1}^{N} \left(\log \left(Q_{obs}^t + 0.1 \right) - \log \left(Q_{sim}^t + 0.1 \right) \right)^2}{\sum_{t=1}^{N} \left(\log \left(Q_{obs}^t + 0.1 \right) - \log \left(\overline{Q_{obs}} \right) \right)^2} \right\} \tag{11}$$

$$VE_C = \left(\sum\nolimits_{y=1}^{N_{years}} V_{obs}^y - \sum\nolimits_{y=1}^{N_{years}} V_{sim}^y \right) / \sum\nolimits_{y=1}^{N_{years}} V_{obs}^y \tag{12}$$

335

$$VE_M = \sum\nolimits_{y=1}^{N_{years}} \left(|V_{obs}^y - V_{sim}^y| / V_{obs}^y \right) / N_{years} \tag{13}$$

where Q_{obs}^t and Q_{sim}^t are, respectively, the observed and simulated discharges for the time step t, N is the number of time steps for which observations are available, Q_{obs}^y and Q_{sim}^y are the observed and simulated volumes for year y, and N_{years} is the number of years in the simulation period.

340   The NSE criterion is as well-known form of the normalized least squares objective function. Perfect agreement between the observed and simulated values yields an efficiency of 1, whilst a negative efficiency represents a lack of agreement worse than if the simulated values were replaced with the observed mean values. The optimal value of the VE_C and VE_M criteria is zero. The latter criteria express the relative difference between observed and simulated values. This multi-objective cali-

345   bration problem was transformed into a single-objective optimization problem by defining a scalar objective function F_{obj} that aggregates the different objective functions:

$$F_{obj} = (1 - NSE) + (1 - NSE_{log}) + |VE_C| + VE_M \tag{14}$$

Calibration was performed in a 7D parameter space by searching for the minimum value of F_{obj}. To achieve this high-dimensional optimization efficiently, the shuffle complex evolution (SCE) al-

350   gorithm was used (Duan et al., 1992).

3.2.3   Cross-calibration and validation of hydrological model

To test the performance of the hydrological model in contrasted conditions, the calibration-validation periods were sub-divided using a differential split-sample testing (DSST) scheme (Klemeš, 1986). Thus, two sub-periods of 10 years each divided according to the median annual precipitation for the

355   period were used either for calibration or for validation.These two sub-periods define dry and wet year periods.

For the cross calibration-validation process, three calibration-validation periods (for the whole period, for dry years, and for wet years) were used to test the performance of the hydrological model in contrasted conditions. A 2-year warm-up period was included at the beginning of each period to

360   attenuate the effect of the initialization of storage. In addition, hydrological years starting in typical low-flow period in the Mediterranean region (from September to August) were used in the modeling process to minimize the boundary limits of the model reservoir. The quality of the simulations was then assessed by comparing the "optimal" parameter set for each calibration period. For each basin, three simulations based on the three sets of parameters were compared (see Fig. 4). The four criteria

365   employed for the multi-objective function (NSE, NSE_{log}, VE_C and VE_M) were used to assess the quality of the simulations. F_{obj} is optimal at 0, and considered satisfactory below 1.

11

The hydrographs in figure 4a illustrate the ability of the model to correctly simulate runoff in the basins, according to the parameter sets used for the calibration periods: "whole period", "dry years" and "wet years". All F_{obj} values were below 1, underlining the quality of the simulations. Whatever the calibration period (whole period, dry or wet years), the objective function F_{obj} did not vary more than 0.1 over the validation period (except the Segre basin in the wet year validation period). This shows the stability of the simulations when the model is calibrated under contrasted hydro-climatic conditions.The lower quality of the simulations for the Segre basin may be attributed to: (i) complex snowmelt processes that are not well represented by the hydrological model; (ii) insufficient quality of data inputs due to the limited number of precipitation and temperature gauges (e.g. only 2 precipitation gauges on a total of 6 stations are included within the Segre basin while 10 stations for the Irati basin); (iii) the very particular hydro-climatic context characterized by a mountainous climatic barrier, which limits Atlantic influence and reduces the quantity of solid and liquid precipitation supplying the streamflow inside the basin. Although the hydrological simulations were less efficient in this basin than in the others, we found them sufficiently correct to provide an additional basin for the inter-comparison of the SDMs through a regional analysis in different hydro-climatic contexts.

Figure 4b shows that the parameter sets are quite stable whatever the calibration period used for the basins. However, the model parameters were normalized with respect to the lower and upper limits of the parameters obtained. As a result, the more the bounds are widened, the less the normalized parameters are able to account for the differences between the calibration periods. Nonetheless, the relative stability of the normalized parameters underlines the robustness of the model under contrasted climatic conditions. However in the Segre basin, differences on the GR4j native parameters reflect the difficulty to correctly simulate runoff in this basin including NSE values of around 0.7. Snow module parameters (*x5*, *x6* and *x7*) in the Herault and Loukkos basins are less stable but the contribution of snowfall in these basins is rather small.Finally, the low drift of the parameters and the relatively homogeneous simulations obtained whatever the calibration period led us to retain the parameter set from the whole period to simulate streamflow under the various climate datasets. To facilitate interpretation and to limit biases in hydrological modeling, the simulated streamflow produced with the best parameter set for the "whole period" calibration period was used as a benchmark (instead of the observed data) for the comparison between the climate datasets in the following steps.

3.3 Comparing downscaling methods from the point of view of water resources

Based on the preliminary calibration of the hydrological model, runoff simulations forced by statistically downscaled climate simulations were compared using hydrological indicators that reflect the main issues of impact studies on water resources. Figure 5 illustrates the different steps of this approach.

12

First, three low-resolution climate datasets (NCEP, CNRM and IPSL) were downscaled using three different statistical methods (ANALOG, CDFt and SWG) to produce new high-resolution hydro-climatic datasets (P and T). Daily PE time series were calculated using the same formula (Oudin et al., 2005) as that used to estimate PE from observed temperature.

After preliminary calibration over the whole reference period under observation-based climate inputs, the hydrological model was then forced with the nine sets of downscaled hydro-climatic data (high resolution) and the three raw datasets (low resolution) to produce an ensemble of 12 runoff simulations. These simulations were compared to a reference runoff simulation (REF) corresponding to the model ouputs over the whole reference period calibrated with observation-based climate inputs. This comparison relies on hydrological indicators that are relevant to the water resource challenges according to four complementary aspects of the hydrograph: volume of the water flow, interannual and seasonal variability of runoff, and streamflow distribution. The water flow volume was assessed according to the cumulated volume error (VE_C, see Eq. 12). Interannual variability was assessed according to a root mean square error applied to the sorted annual flows. This criterion was then normalized by dividing the RMSE value by the mean of annual observed discharge. Choosing a normalized root mean square error criterion (NRMSE, Eq .15) applied to this distribution gets round the non-synchronicity of the simulations. Note that applying the NRMSE criterion to sorted flows may favor high flows. Seasonal variability was assessed using a NSE criterion (Eq. 10) applied to the mean 10-day discharge series. The last comparison criterion was based on the flow duration profile, divided between high and low flows. High flows correspond to daily flows exceeding the 95th percentile (> Q95), i.e. the 5% highest daily flows or flows exceeded 5% of the time. Low flows correspond to daily flows not exceeding the 80th percentile (< Q80), i.e. the 80% lowest daily flows or flows exceeded 20% of the time. This value was deliberately chosen to cover a wide range of flows to enable a meaningful distinction between simulations while correctly representing low flows. Both high and low flows were evaluated using a NSE criterion applied to the high and low flow time series.

$$NRMSE = \frac{\sqrt{\sum_{i=1}^{N} \left(X_{obs,i} - X_{sim,i}\right)^2 / N}}{\overline{X_{obs}}} \tag{15}$$

where X_{obs} is observed values and X_{sim} is simulated values at time/place i. $\overline{X_{obs}}$ is the mean of observed values.

The 12 runoff simulations were compared via these five hydrological indicators. Finally, the downscaling methods (from the runoff simulations forced by the downscaled climate time series) were ranked using the same indicators. The median of the related criterion (VE_C, $NRMSE_{INT}$, NSE_{SEAS}, NSE_{HF} or NSE_{LF}) in the four study areas made it possible to rank the downscaling methods according to their respective performances in a given configuration "climate data – indicator". Next, the simulations were combined by computing the median of the criteria values of the four basins and the three climate datasets to make it possible to rank them. Finally, an additional criterion

(Eq. 16) was used to aggregate the different goodness-of-fit criteria to provide an overview of the performance of the different downscaling models driven by distinct climate datasets. The lower the aggregation criterion, the better the ranking.

$$I_{AGG} = |VE_C| + NRMSE_{INT} + (1 - NSE_{SEAS}) + (1 - NSE_{HF}) + (1 - NSE_{LF}) \qquad (16)$$

For the remainder of this paper, REF refers to the simulated runoff with the parameters calibrated over the whole period based on the observed climate data. RAW refers to the simulations with raw low-resolution climate data from NCEP/NCAR reanalysis or GCMs outputs over the reference period. ANA, CDFt and SWG refer to the simulations based on climate data downscaled via ANALOG, CDFt and SWG methods respectively.

4 Comparative analysis of hydrological responses to downscaled climate forcings

4.1 Water volumes

Water volumes were assessed through the cumulative volume error, i.e. the error in the percentage of the cumulated volume of water flow over the whole period (Table 2). ANALOG-based simulations generally reproduced water volumes better than the other simulations. Nevertheless, differences appeared depending on the input data used (NCEP, CNRM or IPSL) and on the basin concerned (Fig. 6). Except in the Loukkos basin and for CNRM in the Herault and Segre basin, RAW-based simulations were always improved by downscaling. CDFt-based simulations were slightly better than ANALOG-based simulations in reproducing cumulated volume of water with VE_C absolute values averaged between the four basins, with 12% for CDFt and with 14% for ANALOG. In addition, the results of ANALOG-based simulations were more constant without outlier criterion values. Criterion values can be considered as outliers when VE_C is greater than 50%, which may be seen as an unacceptable error. In the Loukkos basin, simulations provided many outliers with both SWG and CDFt. The CDFt method improved the results according to the VE_C criterion better than the other models. SWG-based simulations ranked first for both criteria with NCEP as inputs, but performed poorly with GCMs.

4.2 Interannual variability of streamflow

The ability to reproduce interannual runoff variability was assessed through a root mean square error ($NRMSE_{INT}$) criterion applied to the sorted time series of annual discharge and normalized by dividing RMSE by the mean annual discharge of the reference (see Fig. 7). In other words, for each basin, the downscaling method and input data, and the annual discharge values were sorted from the highest value to the lowest one to generate new decreasing time series on which the NRMSE criterion was calculated with respect to the sorted reference time series. The results show that the interannual variability of runoff is correctly reproduced by the simulations based on most of the downscaled cli-

mate datasets, particularly ANALOG- and CDFt-based simulations in which NRMSE values rarely reached more than 30%. On the whole, RAW-based simulations were improved by downscaling, especially when driven by NCEP and IPSL, except for SWG-based simulations driven by GCMs (Fig. 7). Indeed, when driven by NCEP, the SWG method reproduced interannual variability better than the other methods for three of the four basins, but produced poor results with GCMs, in which case ANALOG- and CDFt-based simulations generally performed better.

4.3 Seasonal variability of streamflow

Seasonal variability was assessed using an NSE criterion (Eq. 10) applied to the mean 10-day discharge series. In most cases, the downscaling methods improved the reproduction of the seasonal variability of streamflow compared to the low-resolution raw datasets (see Fig. 8). This was particularly true of NCEP reanalyses, for which downscaled inputs considerably improved the simulation of the seasonal dynamics more realistically than with RAW-based simulations. Although the ANALOG method did not systematically match the best NSE values, on the whole, the method reproduced the seasonal variability better than CDF-t and SWG. The CDFt method performed particularly well with GCMs as inputs, but proved to be unsuitable with NCEP under the particular hydro-climatic conditions that prevail in the Segre basin. Except with NCEP, SWG-based simulations reproduced poorly the seasonal variability of runoff, due notably to systematic overestimation of high-flow events.

4.4 Streamflow distribution: high and low flows

Streamflow distribution was divided between high flows, i.e. the 5% highest daily flows, and low flows, i.e. the 80% lowest daily flows. Both were evaluated using a NSE criterion applied to the high and low flow time series. On the whole, the downscaling methods improved the reproduction of the distribution of sorted high flows (Fig. 9a). However, it should be noted that the downscaled simulations with CNRM data deteriorated raw data in the Segre basin. Results showed that ANALOG generally reproduced the 5% highest flows best; the NSE values were quite stable and never below 0.47. The CDFt-based simulation results were very close to those obtained with ANALOG, with equivalent scores when NCEP or GCM data were used as inputs. Nevertheless, it should be noted that ANA and CDFt reproduced less accurately high flows in the Segre basin than in the other basins. This can be explained by a lower efficiency of the hydrological model in this area as shown in the section 3.2.3., thus leading to a reference simulated streamflow more uncertain than in the other basins. The SWG method reproduced high flows well with NCEP data as inputs, but not with GCM data.

Figure 9b shows the distribution of sorted low flows and the associated NSE criterion. Moreover, applying a NSE criterion to the sorted low flows tended to emphasize the differences between the simulations and thus made it easy to distinguish simulations that reproduced low flows poorly. The downscaling methods improved the representation of the 80% lowest flows in all basins, except

15

for the SWG method with GCM data used as inputs. In general, the best results were obtained from ANALOG-based simulations, with NSE values always above 0.81. The CDFt-based simulations performed significantly better when forced with GCMs than with NCEP. The SWG-based simulations were unable to reproduce low flows when GCMs data were used as inputs.

5 Discussion and conclusions

The aim of this study was to test the ability of different statistical downscaling climate models to provide accurate hydrological simulations for use in climate change impact studies (CCIS) on water resources. To get round the constraints represented by the inherent characteristics of each climate model, we compared three statistical downscaling methods applied on three low resolution raw datasets: NCEP/NCAR reanalysis data and two GCM data (CNRM and IPSL). The three downscaling methods were an analog method (ANALOG), a stochastic weather generator (SWG) and the "cumulative distribution function – transform" approach (CDFt). This allowed us to analyze the sensitivity of runoff modeling at the catchment scale to 12 climatic series (three raw low-resolution datasets and nine downscaled high-resolution datasets). The sensitivity analysis was based on a previously calibrated hydrological model validated with local hydro-climatic observed data over a 20-year reference period. The model simulations served as a benchmark for the comparison between the raw and downscaled datasets from NCEP reanalysis and GCM outputs over the same period. The comparison with the runoff simulations forced with raw and downscaled climate datasets was based on hydrological indicators describing the main features of the hydrograph: the ability to reproduce the cumulated volume of water flow, interannual and seasonal variability of runoff, and the distribution of streamflow events, including high and low flows. To account for uncertainty related to the spatial variability of the downscaled climate simulations, this approach was applied over four western Mediterranean basins of similar size but that represent a with a wide range of hydro-meteorological situations.

The proposed sensitivity analysis enabled us to identify the strengths and weaknesses of different statistical downscaling methods with respect to the sensitivity of runoff simulations to low-resolution and high-resolution downscaled climate datasets (see Fig. 10). Our study revealed the performances that could be expected from downscaling techniques applied to large-scale datasets to provide acceptable hydrological simulations. To complement the usual calibration/validation exercises conducted by climatologists for assessing the suitability of SDMs based on predictors and reanalyze grids (see e.g. Vaittinada Ayar et al., 2015), we focused on a validation protocol directly based on streamflow thus allowing the combined impacts of the downscaled precipitation and temperature inputs to be considered through the hydrological response.

On the whole, the ANALOG-based simulations performed well in all the situations tested, whatever the large-scale climate dataset used as inputs (NCEP or GCMs), notably in reproducing in-

terannual and seasonal runoff and low flows. ANALOG-based simulations were closely followed by CDFt-based simulations, notably when GCM outputs were used, but with a lower variability of scores than with ANALOG. To the contrary, the results clearly showed that the SWG method should not be used 'as is' in climate change impact studies on water resources. Indeed, although the SWG-based simulations were satisfactory when based on the NCEP large-scale climate dataset, they significantly underperformed when based on GCM outputs. Biases of the GCM data with respect to the NCEP/NCAR reanalyses may explain the poor performances of the SWG method. As SWG is calibrated with "perfect" predictors from reanalyses, its application to biased GCM predictors led to unsatisfactory SWG-based hydrological simulations. To make the SWG method more applicable in climate change impact studies on runoff, one solution could be correcting the GCMs predictors with respect to reanalyses, as done for example by Colette et al. (2012) before performing a dynamical downscaling.

Although the ANALOG method appeared to be the best SDM in this study, it may suffer from certain limitations when used in a climate change context, notably when downscaling GCM projections over the 21st century. One main limitation is that ANALOG is not able to provide suitable simulations for the extreme events if such events increase in intensity in the future (see e.g. Teng et al., 2012). Indeed, by construction, as ANALOG works by resampling the calibration set, it never supplies downscaled values beyond the range of the calibration reference dataset.

On the other hand, although CDFt-based simulations were less consistent than ANALOG simulations, they were more sensitive to climate forcing and also more sensitive to the chosen indicators. The CDFt method was particularly appropriate when we focused on the cumulated volume, seasonal variability and high flows. In addition, it should be noted that the CDFt method is the most parsimonious technique since it generally needs only one variable as predictor. This could obviously be considered an advantage since the complexity of CDFt is very low. However, this low level of complexity could mean that some climate information needed to drive the CDFt more efficiently will be missing. In that sense, one possible improvement could consist in incorporating additional covariates in CDFt, as done by Kallache et al. (2011). Nevertheless, the approach including those additional predictors means that this conditional CDFt has to be calibrated on reanalyses or, at a minimum, on the outputs of a climate model of which the day-to-day evolution of large-scale weather states matches that of the real world. This could be a limitation, since additional biases may appear with those constraints.

The next step will be exploring the potential impact of climate change on the runoff in the basins studied here. To this end, an ensemble approach will be proposed based on the construction of high-resolution climate scenarios using different climate models, gas emission scenarios, and downscaling techniques. In view of the acceptable hydrological simulations obtained with ANALOG and CDFt methods, it may be useful to develop high-resolution climate forcings downscaled with these two methods in order to account for the uncertainty of the downscaling, as recommended by some authors

(e.g. Chen et al., 2011, 2012) for applications in climate change impact studies. Our study also showed the benefits of evaluating the relevance of SDMs in a given hydro-climatic context using a suitable validation protocol. Indeed, selecting unsuitable downscaling methods, such as SWG with GCM outputs, can expand the range of uncertainty linked to the range of SDMs.

Furthermore, our study showed that hydrological responses were sensitive to the climate datasets used as inputs. Indeed, despite the significant contribution of the downscaling methods, hydrological simulations are better from reanalysis data than from GCM data. This demonstrates the limits of GCMs to reproduce current climatic conditions and therefore the associated hydrological responses. This point raises the question about the use of GCM, and thus about the need to correct them afterwards for the evaluation of future hydrological impact in CCIS. Finally, although it is commonly acknowledged that the uncertainty resulting from climate modeling (GCMs, gas emission scenarios and downscaling methods) is highest in a context of climate change (e.g. Wilby and Harris, 2006; Arnell, 2011; Teng et al., 2012), it should be noted that the uncertainty stemming from hydrological modeling may also be high. Several authors (e.g. Benke et al., 2008; Brigode et al., 2013; Hublart et al., 2015; Ruelland et al., 2015) showed that the choice of the hydrological model (structural uncertainty) and its parameterization (parameter uncertainty) could cause significant variability in runoff simulations. Consequently, further analyses of the applicability of the model parameters in a non-stationary context and with different calibration criteria are needed before the model is used in future climate conditions.

Similarly, the different sources of uncertainties and their propagation in the hydrological projections need to be evaluated. To this end, a standard ensemble approach based on various climatic, downscaling and hydrological models may not be sufficient, since using many models without prior validation of their efficiency can lead to very large uncertainty bounds due to the poor quality of some models in the ensemble framework. Minimizing uncertainty thus requires selecting models that perform reasonably well over the reference period in the context of current climate. Although this cannot guarantee the quality of the models for future conditions, we believe it is an essential step to provide more reliable and relevant hydrological projections.

References

615   Arnell, N. W.: Uncertainty in the relationship between climate forcing and hydrological response in UK catchments, Hydrol. Earth Syst. Sci., 15, 897–912, 2011.

Barnston, A. G. and Livezey, R. E.: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, Monthly Weather Review, 115, 1083–1126, 1987.

Benke, K. K., Lowell, K. E., and Hamilton, A. J.: Parameter uncertainty, sensitivity analysis and prediction

620   error in a water-balance hydrological model, Mathematical and Computer Modelling, 47, 1134–1149, 2008.

Brigode, P., Oudin, L., and Perrin, C.: Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?, J. Hydrol., 476, 410–425, 2013.

Buishand, T. A., Shabalova, M. V., and Brandsma, T.: On the choice of the temporal aggregation level for statistical downscaling of precipitation, J. Clim., 17, 1816–1827, 2004.

625   Chandler, R. E. and Wheater, H. S.: Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland, Water Resour. Res., 38, 1192, 2002.

Chen, H., Xu, C. Y., and Guo, S.: Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff, J. Hydrol., 434–435, 36–45, 2012.

Chen, J., Brissette, F. P., and Leconte, R.: Uncertainty of downscaling method in quantifying the impact of

630   climate change on hydrology, J. Hydrol., 401, 190–202, 2011.

Chen, M., Shi, W., Xie, P., Silva, V. B. S., Kousky, V. E., Higgins, R. W., and Janoviak, J.: Assessing objctive techniques for gauge-based analyses of global daily precipitation, J. Geophys. Res., 113, doi:"10.1029/2007JD009132", d04110, 2008.

Colette, A., Vautard, R., and Vrac, M.: Regional climate downscaling with prior statistical correction of the

635   global climate forcing, Geophysical Research Letters, 39, l13707, 2012.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water Resour. Res., 48, w05552, 2012.

Creutin, J. D. and Obled, C.: Objective analyses and mapping techniques for rainfall fields: An objective com-

640   parison, Water Resour. Res., 18, 413–431, doi:"10.1029/WR018i002p00413", 1982.

Dezetter, A., Fabre, J., Ruelland, D., and Servat, E.: Selecting an optimal climatic dataset for integrated modeling of the Ebro hydrosystem, in: Hydrology in a changing world: environmental and human dimensions (Proc. 7th FRIEND Int. Conf., Montpellier, France, 7–10 Oct. 2014). IAHS Publ., vol. 363, pp. 355–360, 2014.

645   Diaz-Nieto, J. and Wilby, R. L.: A comparison statistical downscaling and climate change factor methods: impacts on low flows in the river Thanes, United Kingdom, Climatic Change, 69, 245–268, 2005.

Dibike, Y. B. and Coulibaly, P.: Hydrologic impact of climate change in the Saguenay watershed: comparison of downscaling methods and hydrologic models, J. Hydrol., 307, 145–163, 2005.

Duan, Q. Y., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-

650   runoff models, Water Resour. Res., 28, 1015–1031, 1992.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J.-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedling-

19

stein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas,
J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F.,
Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y.,
Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier,
C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Climate Dynamics, 40, 2123–2165, doi:10.1007/s00382-012-1636-1,
http://dx.doi.org/10.1007/s00382-012-1636-1, 2013.

Fabre, J., Ruelland, D., Dezetter, A., and Grouillet, B.: Simulating past changes in the balance between water demand and availability and assessing their main drivers at the river basin scale, Hydrology and Earth
System Sciences, 19, 1263–1285, doi:10.5194/hess-19-1263-2015, http://www.hydrol-earth-syst-sci.net/19/
1263/2015/, 2015.

Fealy, R. and Sweeney, J.: Statistical downscaling of precipitation for a selection of sites in Ireland employing a generalised linear modelling approach, International Journal of Climatology, 27, 2083–2094,
doi:10.1002/joc.1506, http://dx.doi.org/10.1002/joc.1506, 2007.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recent
advances in downscaling techniques for hydrological modelling, International Journal of Climatology, 27,
1547–1578, doi:10.1002/joc.1556, http://dx.doi.org/10.1002/joc.1556, 2007.

Giorgi, F.: Climate change hot-spots, Geophysical Research Letters, 33, doi:10.1029/2006GL025734, http://dx.
doi.org/10.1029/2006GL025734, l08707, 2006.

Grandin, L. V.: Objective analysis of meteorological fields, Israel Program for Scientific Translations:
Jerusalem, 1965.

Grouillet, B., Fabre, J., Ruelland, D., and Dezetter, A.: Historical reconstruction and 2050 projections of water demand under anthropogenic and climate changes in two contrasted Mediterranean catchments, Journal
of Hydrology, 522, 684–696, doi:http://dx.doi.org/10.1016/j.jhydrol.2015.01.029, http://www.sciencedirect.
com/science/article/pii/S0022169415000463, 2015.

Haylock, M., Hofstra, N., Tank, A. K., Klok, E., Jones, P., and New, M.: A European daily high-resolution
gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res (Atmospheres),
113, doi:doi:10.1029/2008JD102019, http://dx.doi.org/10.1029/2008JD102019, d20119, 2008.

Hublart, P., Ruelland, D., Dezetter, A., and Jourde, H.: Reducing structural uncertainty in conceptual hydrological modelling in the semi-arid Andes, Hydrology and Earth System Sciences, 19, 2295–2314,
doi:10.5194/hess-19-2295-2015, http://www.hydrol-earth-syst-sci.net/19/2295/2015/, 2015.

Kallache, M., Vrac, M., Naveau, P., and Michelangeli, P.-A.: Nonstationary probabilistic downscaling of
extreme precipitation, Journal of Geophysical Research: Atmospheres, 116, doi:10.1029/2010JD014892,
http://dx.doi.org/10.1029/2010JD014892, d05113, 2011.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G.,
Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo,
K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The ncep/ncar 40-year reanalysis project, Bulletin
of the American Meteorological Society, 77, 437–471, 1996.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrol. Sci. J., 31, 13–24, 1986.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, rG3003, 2010.

Michelangeli, P.-A., Vrac, M., and Loukos, H.: Probabilistic downscaling approaches: Application to wind cumulative distribution functions, Geophysical Research Letters, 36, doi:10.1029/2009GL038401, http://dx.doi.org/10.1029/2009GL038401, 111708, 2009.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, doi:http://dx.doi.org/10.1016/0022-1694(70)90255-6, http://www.sciencedirect.com/science/article/pii/0022169470902556, 1970.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, Journal of Hydrology, 303, 290–306, doi:http://dx.doi.org/10.1016/j.jhydrol.2004.08.026, http://www.sciencedirect.com/science/article/pii/S0022169404004056, 2005.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275–289, doi:http://dx.doi.org/10.1016/S0022-1694(03)00225-7, http://www.sciencedirect.com/science/article/pii/S0022169403002257, 2003.

Prudhomme, C. and Davies, H.: Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: baseline climate, Climatic Change, 93, 177–195, doi:10.1007/s10584-008-9464-3, http://dx.doi.org/10.1007/s10584-008-9464-3, 2009.

Ruelland, D., Brisset, N., Jourde, H., and Oyarzun, R.: Modeling the impact of climatic variability on the groundwater and surface flows from a mountainous catchment in the Chilean Andes, in: Cold regions hydrology in a changing climate, edited by Publ, I., vol. 346, pp. 171–179, 2011.

Ruelland, D., Dezetter, A., and Hublart, P.: Sensitivity analysis of hydrological modeling to climate forcing in a semi-arid mountainous catchment, in: Hydrology in a changing world: environmental and human dimensions (Proc. 7th FRIEND Int. Conf., Montpellier, France, 7–10 Oct. 2014). IAHS Publ., vol. 363, pp. 145–150, 2014.

Ruelland, D., Hublart, P., and Tramblay, Y.: Assessing uncertainties in climate change impacts on runoff in Western Mediterranean basins, in: Hydrologic non-stationarity and extrapolating models to predict the future (Proc. of symp. HS02 held during IUGG2015 in Prague, Czech Republic, July 2015). IAHS Publ., vol. in press, 2015.

Salathé, E. P.: Downscaling simulations of future global climate with application to hydrologic modelling, International Journal of Climatology, 25, 419–436, doi:10.1002/joc.1125, http://dx.doi.org/10.1002/joc.1125, 2005.

Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data, Proceedings of the 1968 ACM National Conference, pp. 517–524, doi:10.1145/800186.810616, 1968.

Teng, J., Chiew, F. H. S., Timbal, B., Wang, Y., Vaze, J., and Wang, B.: Assessment of an analogue downscaling method for modelling climate change impacts on runoff, Journal of Hydrology, 472–473,

111–125, doi:http://dx.doi.org/10.1016/j.jhydrol.2012.09.024, http://www.sciencedirect.com/science/article/pii/S002216941200827X, 2012.

Tramblay, Y., Ruelland, D., Somot, S., Bouaicha, R., and Servat, E.: High-resolution Med-CORDEX regional climate model simulations for hydrological impact studies: a first evaluation of the ALADIN-Climate model in Morocco, Hydrology and Earth System Sciences, 17, 3721–3739, doi:10.5194/hess-17-3721-2013, http://www.hydrol-earth-syst-sci.net/17/3721/2013/, 2013.

Vaittinada Ayar, P., Vrac, M., Bastin, S., Carreau, J., Déqué, M., and Gallardo, C.: Intercomparison of statistical and dynamical downscaling models under the EURO- and MED-CORDEX initiative framework: present climate evaluations, Climate Dynamics, pp. 1–29, doi:10.1007/s00382-015-2647-5, http://dx.doi.org/10.1007/s00382-015-2647-5, 2015.

Vaze, J., Post, D., Chiew, F., Perraud, J.-M., Viney, N., and Teng, J.: Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies, Journal of Hydrology, 394, 447–457, doi:http://dx.doi.org/10.1016/j.jhydrol.2010.09.018, http://www.sciencedirect.com/science/article/pii/S0022169410005986, 2010.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, International Journal of Climatology, 30, 1627–1644, doi:10.1002/joc.2003, http://dx.doi.org/10.1002/joc.2003, 2010.

Vigaud, N., Vrac, M., and Caballero, Y.: Probabilistic downscaling of GCM scenarios over southern India, International Journal of Climatology, 33, 1248–1263, doi:10.1002/joc.3509, http://dx.doi.org/10.1002/joc.3509, 2013.

Voldoire, A., Sanchez-Gomez, E., y Mélia, D. S., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., Déqué, M., Deshayes, J., Douville, H., Fernandez, E., Madec, G., Maisonnave, E., Moine, M.-P., Planton, S., Saint-Martin, D., Szopa, S., Tyteca, S., Alkama, R., Belamari, S., Braun, A., Coquart, L., and Chauvin, F.: The CNRM-CM5.1 global climate model: description and basic evaluation, Climate Dynamics, 40, 2091–2121, doi:10.1007/s00382-011-1259-y, http://dx.doi.org/10.1007/s00382-011-1259-y, 2013.

Vrac, M. and Yiou, P.: Weather regimes designed for local precipitation modeling: Application to the Mediterranean basin, Journal of Geophysical Research: Atmospheres, 115, doi:10.1029/2009JD012871, http://dx.doi.org/10.1029/2009JD012871, d12103, 2010.

Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L., and Somot, S.: Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment, Natural Hazards and Earth System Science, 12, 2769–2784, doi:10.5194/nhess-12-2769-2012, http://www.nat-hazards-earth-syst-sci.net/12/2769/2012/, 2012.

Wilby, R. L. and Harris, I.: A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, Water Resources Research, 42, doi:10.1029/2005WR004065, http://dx.doi.org/10.1029/2005WR004065, w02419, 2006.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P.: Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs, Climatic Change, 62, 189–216, doi:10.1023/B:CLIM.0000013685.99609.9e, http://dx.doi.org/10.1023/B%3ACLIM.0000013685.99609.9e, 2004.

22

Table 1. Selected predictors according to the SDM and the predictand. These variables are: the dew point at 2m (D2), the temperature at 2m (T2), the sea level pressure (SLP), the relative humidity, the zonal and meridional wind components, the geopotential height at 850 hPa pressure level (R850, U850, V850 and Z850) and the large-scale precipitation (PR). The pre-processing (PC) of the predictors depends on the SDM.

SDM	Predictand	D2	SLP	T2	U850	V850	Z850	PR
ANA	PR	Field of anomalies	Field of anomalies	Field of anomalies	Field of anomalies	Field of anomalies	Field of anomalies	-
	T	-	Field of anomalies	Field of anomalies	Field of anomalies	Field of anomalies	Field of anomalies	-
CDFt	PR	-	-	-	-	-	-	Raw
	T	-	-	Raw	-	-	-	-
SWG	PR	2 first PCs	2 first PCs	2 first PCs	2 first PCs	2 first PCs	2 first PCs	-
	T	-	2 first PCs	2 first PCs	2 first PCs	2 first PCs	2 first PCs	-

Table 2. Cumulative volume error (VE_C) between hydrological simulations based on downscaled or raw climate data (ANA, CDFt, SWG, RAW) and the reference (REF). Values are expressed in % of difference in the total volume of water flowed during the period.

	NCEP				CNRM				IPSL			
	RAW	ANA	CDFt	SWG	RAW	ANA	CDFt	SWG	RAW	ANA	CDFt	SWG
Herault	-98%	-13%	18%	-13%	-12%	-17%	14%	42%	-53%	-13%	2%	57%
Segre	-77%	-15%	38%	-18%	-4%	-14%	1%	49%	-90%	-20%	12%	61%
Irati	-71%	-9%	19%	-4%	65%	6%	21%	34%	-70%	-2%	21%	54%
Loukkos	-79%	-31%	7%	-10%	-96%	-39%	-14%	124%	-100%	-20%	9%	195%

Yiou, P., Salameh, T., Drobinski, P., Menut, L., Vautard, R., and Vrac, M.: Ensemble reconstruction of the atmospheric column from surface pressure using analogues, Climate Dynamics, 41, 1333–1344, doi:10.1007/s00382-012-1626-3, http://dx.doi.org/10.1007/s00382-012-1626-3, 2013.
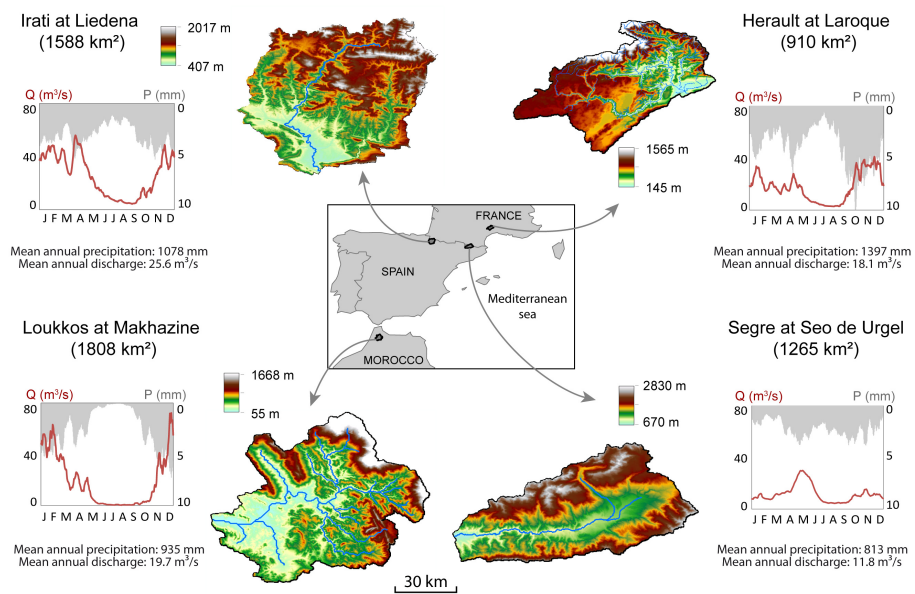
Figure 1. Study catchments (Herault, Segre, Irati and Loukkos) in the western Mediterranean region with their topography and mean seasonal variability in precipitation (P) and discharge (Q) for the period 1986–2005.
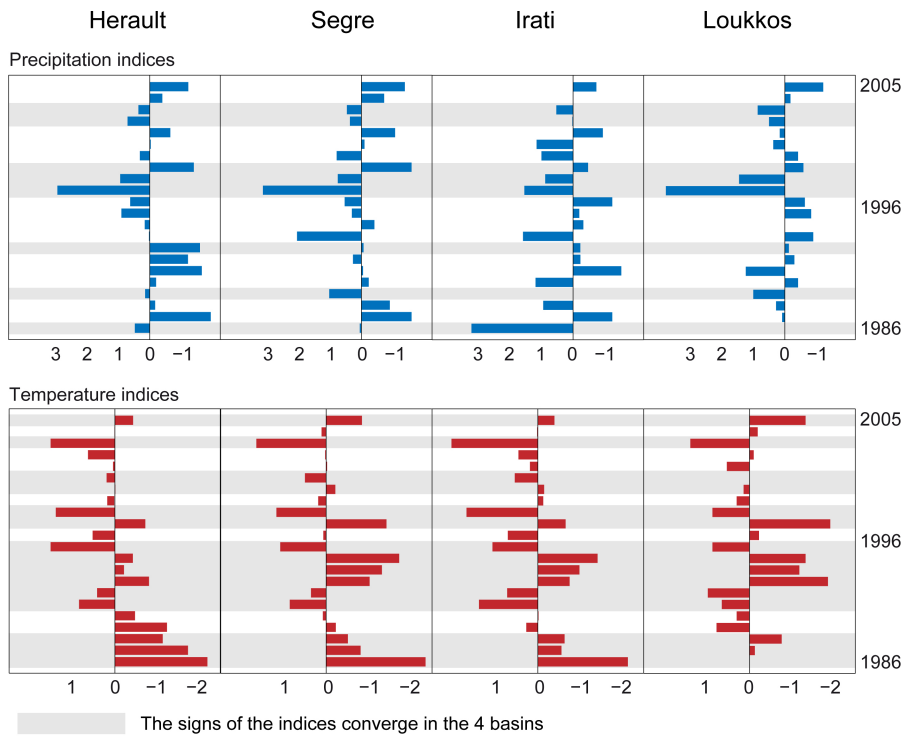
Figure 2. Precipitation ($I_P = (P_y - \overline{P_y})/\sigma_P$) and temperature indices ($I_T = (T_y - \overline{T_y})/\sigma_T$) applied on the four basins over the 1986–2005 period. The grey lines highlights years when the signs of the indices are the same for the four basins. P_y is the annual precipitation for the year y, $\overline{P_y}$ is the mean of the annual precipitation, σ_P is the standard deviation of the annual precipitation. T_y is the annual temperature for the year y, $\overline{T_y}$ is the mean of the annual temperature, σ_T is the standard deviation of the annual temperature.
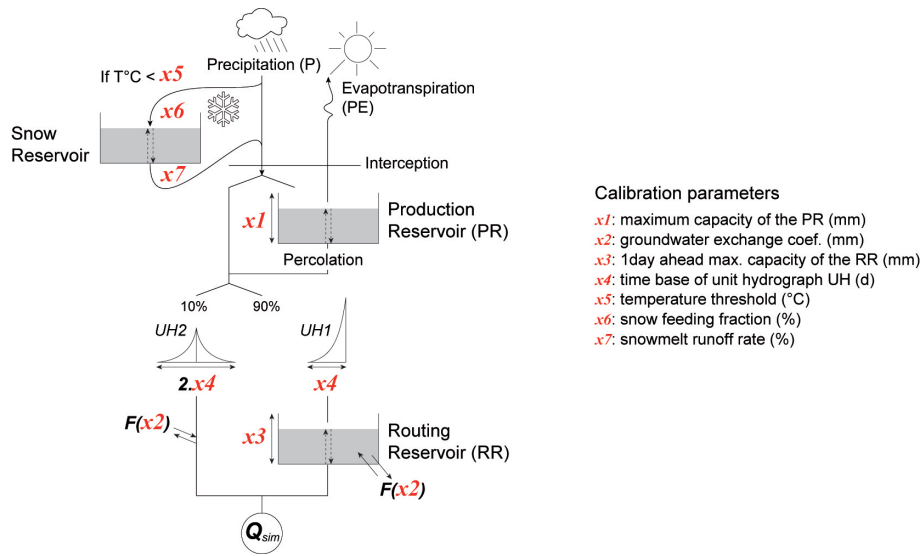
Figure 3. Schematic diagram of the hydrological model GR4J. Adapted from Perrin et al. (2003); Ruelland et al. (2011).
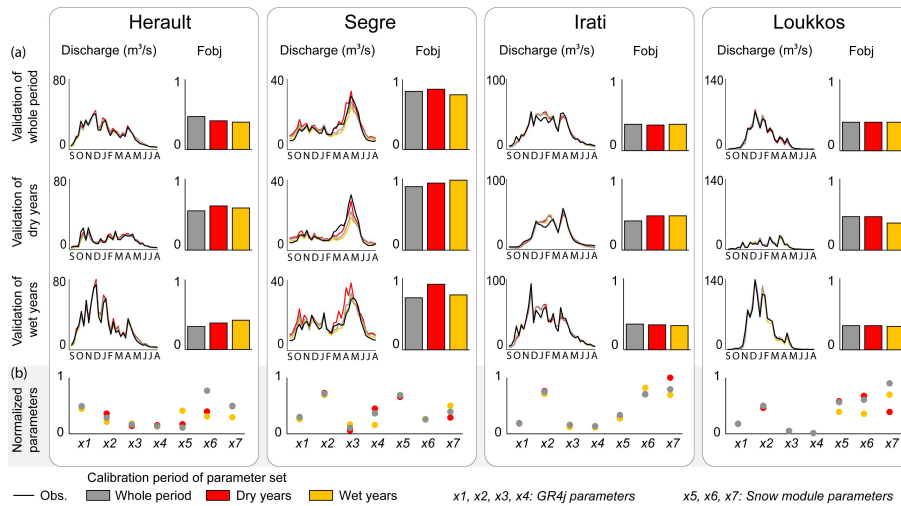
Figure 4. Cross calibration/validation of the hydrological model. (a) Seasonal representation (from September to August) of simulated and observed runoff during the whole period (WHO, first row), dry years (DRY, second row) and wet years (WET, third row) according to parameter sets optimized respectively for the whole period (in grey), dry years (red) and wet years(yellow). F_{obj} ($F_{obj} = (1 - NSE) + (1 - NSE_{log}) + |VE_C| + VE_M$) is computed on daily series. F_{obj} is optimal at 0, but considered satisfactory below 1. (b) Normalized model parameters obtained over the three calibration periods.
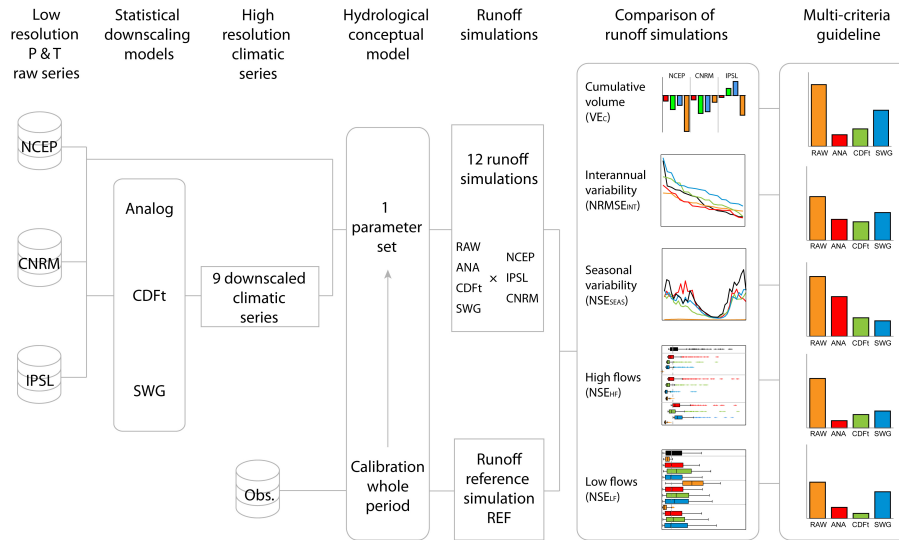
Figure 5. Flow chart illustrating the method used to compare the three downscaling methods through a hydrological sensitivity analysis.
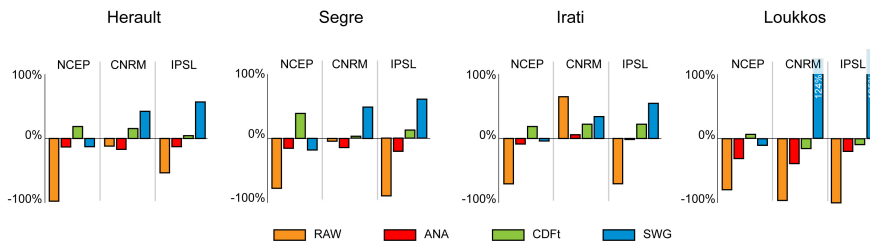

Figure 6. Comparison of the downscaling methods according to the cumulative volume error (VE_C) used as criterion to compare the downscaling methods applied to NCEP, CNRM and IPSL climate inputs in the four basins. The smaller the absolute value of the criterion, the better the simulation.
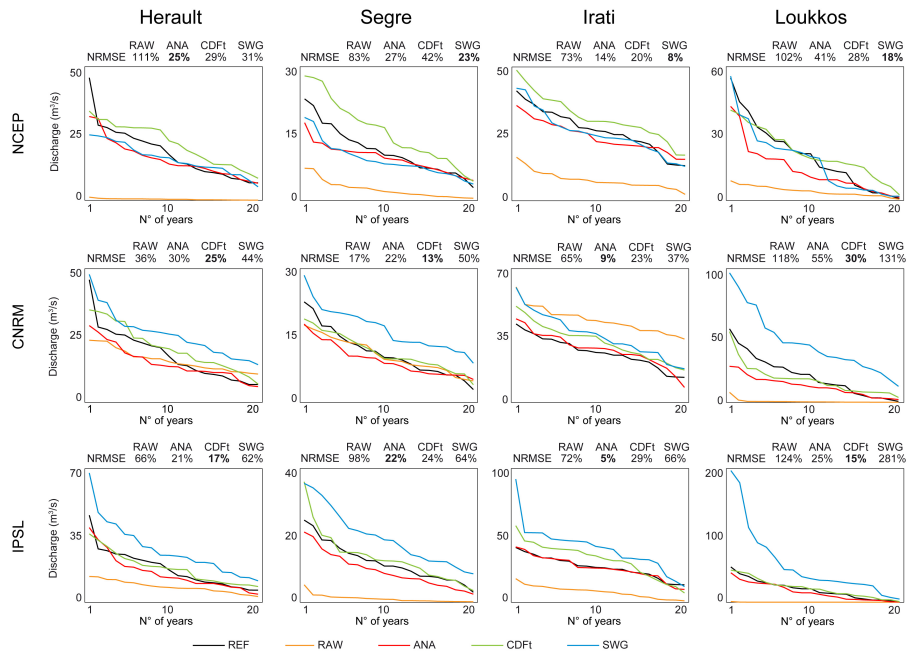
Figure 7. Comparison of the sorted annual discharge simulated using REF data, RAW (NCEP or GCM) data, and the three downscaling methods (applied to NCEP, CNRM and IPSL) for each basin. The NRMSE values above each panel represent a root mean square error applied to the sorted time series of annual discharge normalized by dividing RMSE by the mean annual discharge of the reference time series. The best values are in bold.
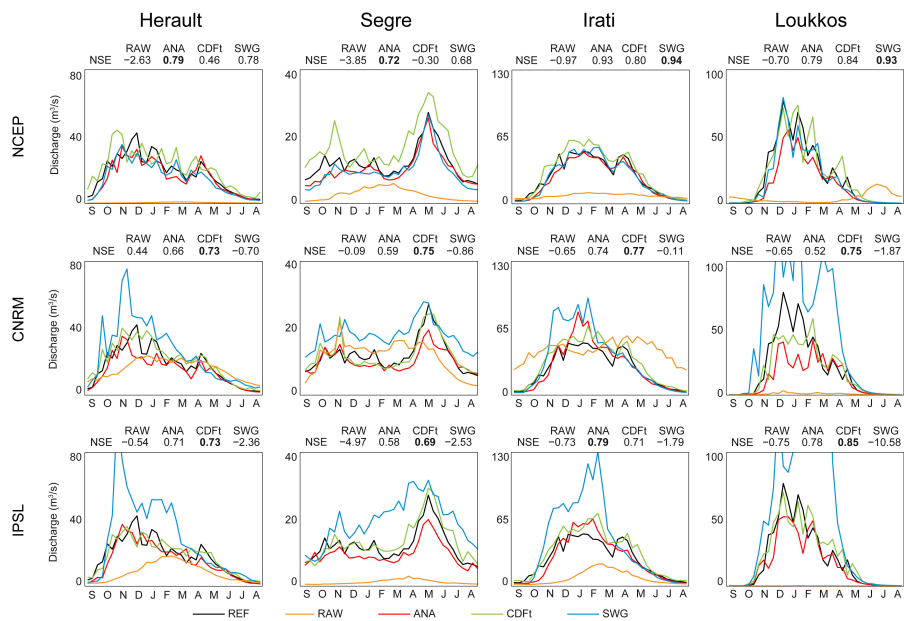
Figure 8. Comparison of seasonal variations in streamflow simulated using REF data, RAW (NCEP or GCM) data, and the three downscaling methods (applied to NCEP, CNRM and IPSL) for each basin. The NSE values for the mean 10-day discharge between REF and the simulation concerned are given above each panel. The best values are in bold.
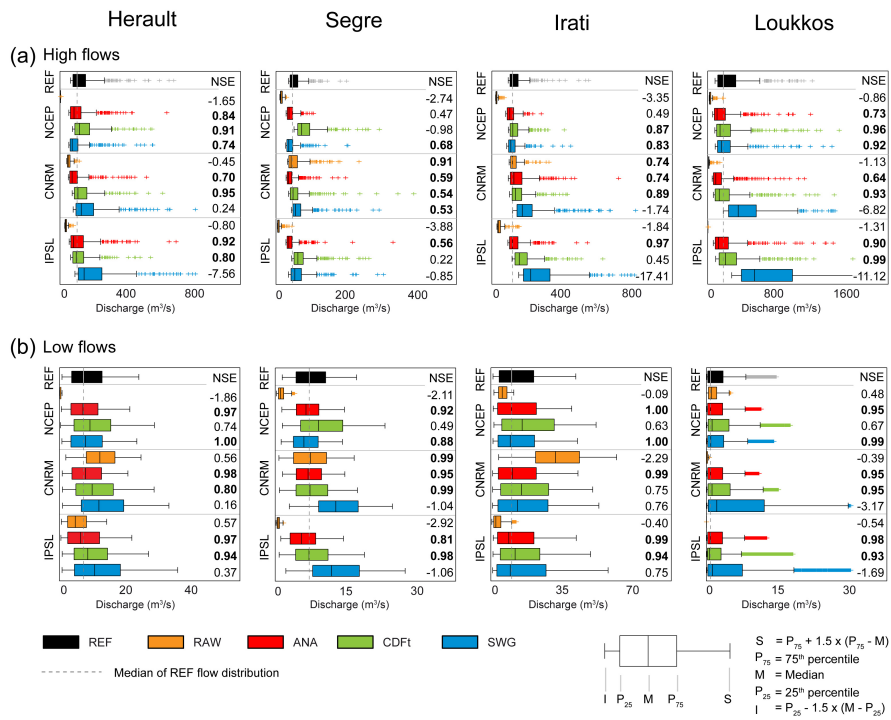
Figure 9. Comparison of (a) the 5% daily high flows and (b) the 80% daily low flows simulated with REF data, RAW (NCEP or GCM) data, and the three downscaling methods (applied to NCEP, CNRM and IPSL) for each basin. The NSE values calculated on the 5% high and the 80% low flows are indicated on the right in each panel. NSE values higher than 0.5 for high flows and 0.8 for low flows are in bold.
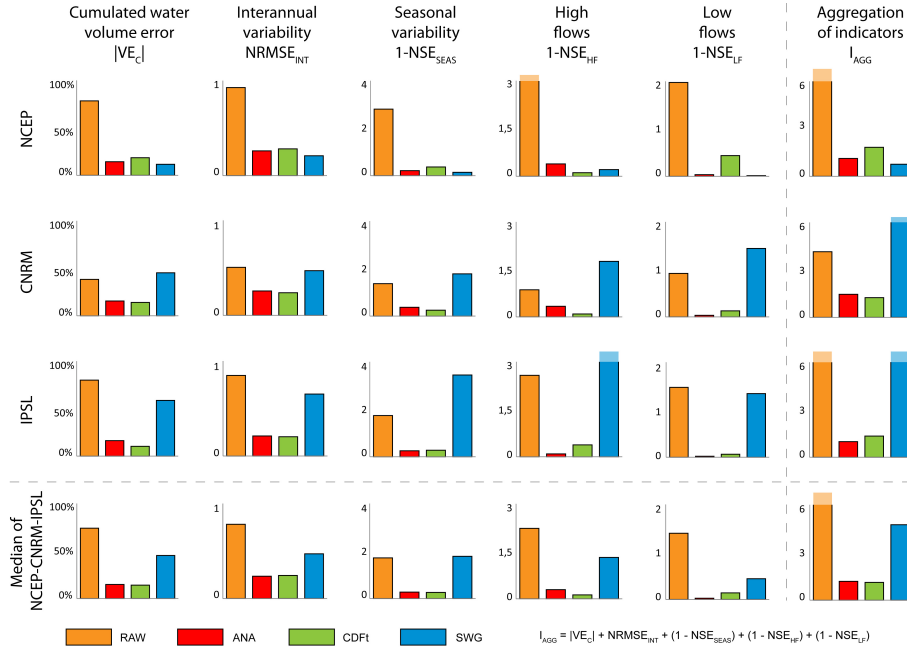
Figure 10. Efficiency of the different climatic datasets to reproduce different aspects of the hydrographs from the four basins over the period 1986–2005: comparison of low resolution datasets (RAW) and high resolution datasets downscaled using the ANALOG, CDFt or SWG methods forced by NCEP/NCAR reanalyses and outputs from the CNRM and IPSL. The bars represent the median of the indicator values of the four basins. The smaller is the bar, the better the result. The row "Median of NCEP-CNRM-IPSL" corresponds to the median of the four basins for the three large-scale climate datasets (NCEP, CNRM and IPSL). The column "Aggregation of indicators" sums the six indicators values according to the following equation: $I_{AGG} = |VE_C| + NRMSE_{INT} + (1 - NSE_{SEAS}) + (1 - NSE_{HF}) + (1 - NSE_{LF})$.