Please find in this document, the two revision notes related to C4467 and C4792 respectively.

Contenu

C4467 - Revision note in response to the anonymous review

General comment:

*The paper investigates the results with respect to decadal and monthly runoff of applying three different methods of downscaling of daily air temperature and precipitation data from GCM (2.5° spatial resolution) to watershed scale (watershed area around 1000 km2); Four watersheds belonging to the western Mediterranean region are studied. A reference period of 20 years is considered (1986-2005). The downscaling is relative to inputs from reanalysis data as well as two GCM data. Thus, the analysis deals with nine series of results (3 input data sources * three downscaling methods). A rainfall runoff model is calibrated using in situ input data and runoff observations. Two downscaling methods are found to perform better than the third one in reproducing the simulated runoff obtained when using in situ data as inputs. The illustrations are of good quality. My general comment is about the lack of explanation of downscaling methods used to build the comparison. Also, the discussion of intermediate results (downscaling comparison) is quite missing. My principal critic is about the omission of comparing runoff generated using downscaled data to observed runoff.*

Authors' response:

Thank you for your encouraging comments. We have tried to answer the questions and remarks available in the specific comment responses. This particularly concerns the lack of explanation of downscaling methods used to build the comparison, the lack of comparison between the downscaled climate outputs and the omission of comparing runoff generated using downscaled data to observed runoff.

Specific comment:

*Authors say that 818 and 264 stations rainfall and climatic stations are available in Ebro basin. What is the size of this basin? What is the link with Segre and Iraki basins (they said they are upstream; are these basins nested?) - Compared to these basins the network of the basin from the South Mediterranean sea is too sparse while authors adopted a 5*5 km2 grid in the sparse network basin and 8*8 km2 in the well observed basins. What is the idea behind that?*

Authors' response:

The Irati and Segre catchments are two sub-basins located in the left bank of the Ebro River in the Pyrenees Mountains. These two basins drain an area of 1588 km² and 1265 km² respectively while the whole Ebro catchment covers an 85.000 km² area. As shown in Figure 1, these basins are not nested. Preliminary studies over the whole Ebro catchment (Dezetter et al., 2014; Fabre et al., 2015) provided an 8 x 8 km climatic grid that was interpolated based on 818 and 264 climatic stations for precipitation and temperature. We used an extract of this grid in the studied basins.

On the same manner, preliminary studies (Tramblay et al., 2013; Ruelland et al., 2015) made it possible to provide climatic stations over the Loukkos basin. A 5 x 5 km grid was used originally in the framework of these studies. This is why the 5 x 5 km grid was mentioned erroneously in the text. In fact, by concern of consistency, we first re-built this grid based on an 8 x 8 km resolution as well.

Note however that the network of climatic stations on the Moroccan basin is of the same order of density than in the other catchments. For instance, 6 precipitation gauges (on a total of 11 stations used) are included within the Loukkos catchment while 10 stations are included within the Irati catchment.

Authors' changes in manuscript:

The error regarding the resolution grid has been updated in the manuscript.

The sentence "In the Loukkos basin, precipitation data were interpolated on a 5 x 5 km grid based on 11 stations using the IDW method" has been replaced by "In the Loukkos basin, precipitation data were interpolated on an **8 x 8 km** grid based on 11 stations using the IDW method"

Specific comment:

page 10074 line 16: The sentence "The calibrated SDMs were forced with three different datasets: NCEP reanalysis data over the 1976–2005 calibration period and with the IPSL-CM5A-MR (Dufresne et al., 2013) and CNRM-CM5 (Voldoire et al., 2013) GCMs, regridded at a 2.5° spatial resolution, over the GCMs historical (or CTRL) period (i.e. 1986–2005). This is not clear. First it was necessary for authors to state that 1986-2005 was adopted as control period in line 12 of page 10073. Secondly, authors should begin by explaining the differences between NCEP/NCAR daily reanalysis data and IPSL-CM5A-MR and CNRM-CM5 data. These kind of data are "by essence" different. So a differentiation should be adopted from the beginning. A brief description of the science behind these data should be included (domain, model, data, assumptions ...). Also authors should specify the chosen GCM control period and how it was defined. Why do they use NCEP reanalysis data over the 1976–2005 while the period of control is 1986-2005?

Authors' response:

The SDMs have been calibrated over a 30-year period (1976–2005) for the Herault, Irati and Segre catchments, but not for the Loukkos catchment that only had a 20-year period data availability: hence a 20-year calibration was performed for this catchment.

This idea was 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.

Concerning the differences between NCEP/NCAR and GCMs, the biggest one is certainly that NCEP/NCAR data are reanalyses (i.e., model simulations that are constrained/updated through data assimilation of observations) and therefore can generally be considered as "observations" at a large scale (see Kalnay et al., 1996 for details). GCMs cannot be considered as observations, at least in the sense that a GCM output for a given day has absolutely no reason to be in agreement with the observation of this given day. At best, the statistical properties of the GCM outputs are equivalent to those of the observations. This type of GCM with no "synchronicity" with observations is called a "free running" GCM. Specific details (domain, physical assumption, etc.) concerning the two GCMs involved in this study are given in Dufresne et al. (2013) and Voldoire et al. (2013).

Authors' changes in manuscript:

The following paragraph has been added at the end of the section 2.2:

"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., 165 2013) GCMs, regridded at a 2.5° spatial resolution, over the GCMs historical 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 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."

Specific comment:

Page 10074. How are positioned the studied basins in comparison to the 240 grid points of the GCM? It may be important because of frontier effects in interpolation. How many grid points in each basin? A fig. and/or a Table should be added

Authors' response:

As illustrated in the figure below, three of the four basins (Herault, Segre and Loukkos) are included in a single grid cell. The Irati basin straddles two grid cells, split equally (50/50). Also, the basins are not on the edge of the GCM grid (240 grid cells), and therefore are not subject to border effects in interpolation.

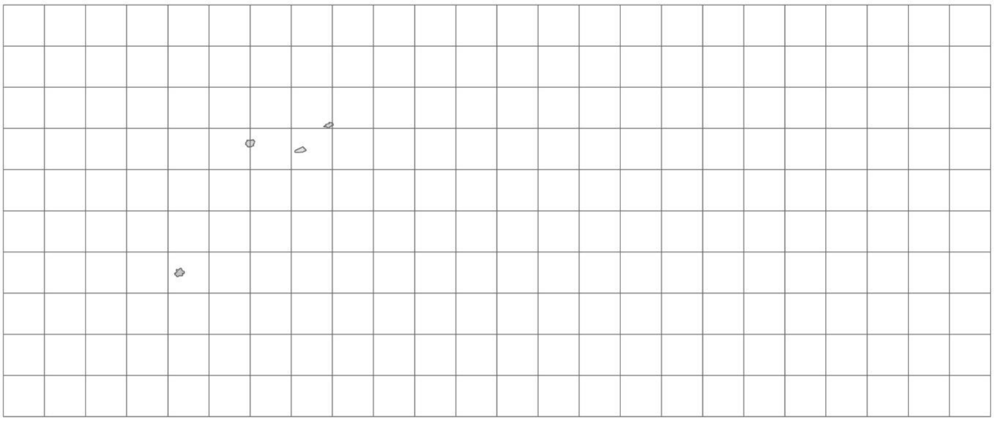


Figure 1 : Location of the four basins in the GCM grid. Loukkos, Irati, Segre, Herault from the west to the east respectively.

Authors' changes in manuscript:

This paragraph has been added at the end of the Section 2.2: "**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.**"

Specific comment:

For the ANALOG model I would like to see some intermediate results. What are the neighbor days for a given day? Are authors satisfied with this classification? Did they examine its results?

Authors' response:

For the ANALOG model, 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. For example, if the day to downscale is the 1st of July 2002, only the time period 1976–2001 U 2003–2005 is considered and only the days between June, 15th and July ,15th. Therefore, this prevents the analog day to be too close (in time) to the day to be downscaled.

Moreover, for this specific study, we did not look at "when" the selected days are for two consecutive days to be downscaled. In previous studies (e.g., Vaittinada Ayar et al., 2015), focusing on the statistical downscaling models, this has been investigated but consecutive days were not necessarily found (not shown), showing the capability of the ANALOG model to capture the specific temporality of the downscaled sequence.

Specific comment:

The validation period should be specified before presenting the three methods. The sentence "Thus, two sub-periods of 10 years each divided according 5 to the median annual precipitation for the period were used either for calibration and for validation" should be reported page 10076 otherwise the reader is not aware about the existence of a validation period

Authors' response:

There is a distinction between the calibration/validation periods regarding the hydrological model (differential split sample test (DSST) over 1986–2005 between 10 dry years vs. 10 wet years) and the preliminary calibration of the SDMs, which was over 1976–2005. The DSST applied to the hydrological model aimed at testing the model's robustness under contrasted climate conditions over 1986–2005. As explained in the manuscript in Section 3.2.3, this preliminary calibration/validation exercise enabled us to show that the hydrological model was able to reproduce the outlet streamflow with a high degree of realism whatever the calibration periods was used (dry years, wet years or whole period over 1986–2005). Consequently, the runoff simulated under the observed climate datasets with the parameters calibrated over the whole 1986–2005 period was retained as a benchmark for the comparison with the runoff simulated based on the raw and downscaled climate datasets to be compared through their ability in providing accurate hydrological simulations with the same calibration parameters. As mentioned before, the GCM historical period is 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.

Authors' changes in manuscript:

The following paragraph has been added at the end of the section 2.2 to clarify the calibration/validation periods:

"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."

Specific comment:

How do authors define large-scale atmospheric situation XANA. ? page 10076

Authors' response:

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.

Authors' changes in manuscript:

The following sentence in Section 3.1.1 "Note that this method is applied on the anomalies of the predictors with respect to the seasonal cycle (Yiou et al., 2013)." has been replaced by "**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.**"

Specific comment:

Also what do they mean by "the anomalies of the predictors with respect to the seasonal cycle"? Do they look for the most similar situation given the season? they said that ANALOG was calibrated and run on season basis.

Authors' response:

The anomalies of a large-scale variable were calculated by subtracting the seasonal cycle to the raw predictor fields. For example, for the 22th of October 2000, the mean of all the 22th October, over the whole calibration period was subtracted in order to obtain the daily anomaly. More precisely, for the ANALOG model, 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. For example, if the day to downscale is the 1st of July 2002, only the time period 1976-2001 U 2003-2005 is considered and only the days between June, 15th and July ,15th. Therefore, this prevents the analog day to be too close (in time) to the day to be downscaled.

Authors' changes in manuscript:

The first paragraph of Section 3.1.1 has been amended:

The "analogs" model used here is based on the approach of Yiou et al. (2013). 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 days in the calibration period. More technically, this can be written as:

Specific comment:

For ANALOG method it is important to describe how authors split from the identification of the closest day (from anomaly perspective) to the downscaled data.

Authors' response:

For the ANALOG method, the date of the day that has the closest large-scale (field of anomalies) situation to the day to be downscaled is obtained. Then, the local-scale data corresponding to the obtained date is assigned to the day to be downscaled. The hypothesis here is that similar large-scale situations imply similar local-scale conditions. The closest day is determined by the Euclidian distance (see section 3.1.1).

Authors' changes in manuscript:

The first paragraph of Section 3.1.1 has been amended:

The "analogs" model used here is based on the approach of Yiou et al. (2013). 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 days in the calibration period. More technically, this can be written as:

Specific comment:

In CDFT model what do authors mean by predictor? How do they use these predictors? It is important to specify this aspect and also to compare the maps of daily results obtained from the three downscaling methods. What about persistence aspects?

Authors' response:

For CDFt, the predictor and the local-scale variables correspond to the same meteorological variable. For example, to generate local-scale temperatures, the large-scale temperature (from the GCM grid-cell containing the local-scale location of interest) is taken as predictor. The difference (between the three SDMs) in terms of persistence aspects is thoroughly discussed in Vaittinada Ayar et al. (2015). The aim of this paper is to discriminate SDMs through a hydrological point of view.

Authors' changes in manuscript:

A table has been added to summarize the predictors and the pre-processing of those predictors according to the SDM and the predictands. Moreover, the following sentence has been added at the in Section 2.2: "**The predictors and the pre-processing of those predictors according to the SDM and the predictands are summarized in the table 1."**

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 (anomalies or PC=principal components) 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	-

Some elements of the following modification of the Section 3.1.2 can also clarify this point: "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."**

<u>**Specific comment:**</u>

In 3.2.3 title, authors should add "of hydrological model", because one may think that they are assessing the downscaling quality which is not performed here.

Authors' response:

Agreed.

Authors' changes in manuscript:

Done

<u>**Specific comment:**</u>

In 3.3 the sentence "the quality of runoff simulations forced by statistically downscaled climate simulations was evaluated" is not reflecting what authors are doing. In effect authors are evaluating to what extent outputs are similar to simulations of runoff forced by observed data, which is not the same thing. Can authors report the discrepancy with observations?

Authors' response:
Agreed. This sentence could lead to misunderstanding, and has been changed in the manuscript. For the remark about observations, please see the detailed answer of the Specific comment "P10083…".
Authors' changes in manuscript:
The sentence "Based on the preliminary calibration of the hydrological model, the quality of runoff simulations forced by statistically downscaled climate simulations was evaluated using hydrological indicators that reflect the main issues of impact studies on water resources" has been changed in "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**".

<u>**Specific comment:**</u>
Page 10076. What is the link with "reanalyze grid scale (0.44° spatial resolution)"? - Data in Herault basin were extracted from the SAFRAN 8 km_8 km meteorological analysis system. The key word of reanalysis should be used here in the text. Authors need to write a sentence about the method Safran is a gauge-based analysis system using the Optimal Interpolation (OI) method described by Gandin (1965). From Vidal et al. 2010 https://hal-meteofrance.archivesouvertes.fr/meteo-00420845/document " One may find in this document conclusions about validation of these data sets. - What is the link between SAFRAN et Xie et al.?; Also authors need to add the reference of Obled et Creutin which is very important point of departure of many works in the same field Â´n Creutin and Obled (1982) examined several well-known schemes and recommended the optimal interpolation (OI) of Gandin (1965). Â˙z that's what said Xie and al. Xie et al. said that "while similar performance statistics can be achieved by other inverse-distance interpolation algorithms if the anomaly, instead of the total, is interpolated". Now it is the anomaly which is interpreted in the present case?
Authors' response:
The three statistical downscaling techniques have been firstly chosen according to their ability to reproduce commonly used climatic patterns on E-OBS grid scale (0.44° spatial resolution) in preliminary studies. Based on this finding, we chose to apply these methods at finer resolutions, i.e. at the basin scale.
We agree that a few explanations about Safran features are needed. In that way, we amended the section 2.2.
Authors' changes in manuscript:
The following paragraph has been modified: "Climate data for the Herault basin were extracted from the SAFRAN 8 km x 8 km meteorological analysis system (Vidal et al., 2010) and 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 Gandin (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).**"

The following references have been added:
Creutin, J.-D., Obled, C. 1982. Objective analyses and mapping techniques for rainfall fields: An objective comparison. Water Resources Research 18 : 413-431. DOI:10.1029/WR018i002p00413
Gandin, L. V. 1965. Objective analysis of meteorological fields. Israel Program for Scientific Translations: Jerusalem.

<u>**Specific comment:**</u>
The daily climatology used in Xie et al. is not the median value "First time series of 1978–97 20-yr mean daily precipitation are calculated for the 365 calendar days for all stations with 80% or higher

reporting rates. Fourier truncation is then performed for the 365-day time series of raw mean daily precipitation, and the accumulation of the first six harmonic components is defined as the daily climatology of precipitation at the stations." What did authors do exactly in the present work? " interpolating the ratio of total rainfall to the climatology, instead of the total rainfall itself, the OI is capable of better representing the spatial distribution of precipitation, especially over regions with substantial orographic effects [Xie et al., 2007]. "in Chen et al. *ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/DOCU/Chen_et_al_2008_JGR_Gauge_Algo.pdf ; Did authors interpolated the ratio ?*

Field Code Changed

Authors' response:

We do not clearly understand why the reviewer expects so many details on the SAFRAN reanalysis. As mentioned before, we have provided additional references and information on this dataset that serves as reference for climate in the Herault basin. However, given the reference listed on SAFRAN and the additional information provided, we believe it is unnecessary and far beyond the scope of this paper to detail more deeply this dataset as compared to the other catchments. We hope the modifications brought are sufficient for the reviewer.

Specific comment:

the IDW method should be documented and reported with the key reference of Shepard. Its quality assessment (See Chen et al. 2008) should be reported

Authors' response:

Agreed. The manuscript has been modified consequently.

Authors' changes in manuscript:

"Climate data for the Segre and Irati basins were obtained by interpolating daily precipitation and temperature measurements on an 8 x 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).**"

References section has been updated:

Shepard, Donald (1968). « A two-dimensional interpolation function for irregularly-spaced data » Proceedings of the 1968 ACM National Conference: 517–524. DOI:10.1145/800186.810616

Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. Wayne Higgins, and J. E. Janowiak (2008), Assessing objective techniques for gauge-based analyses of global daily precipitation, J. Geophys. Res., 113, D04110, doi:10.1029/2007JD009132.

Specific comment:

In Eq 1 and 2 Do authors have an idea about the statistical properties of the anomaly defined in this way?

Authors' response:

After checking, it appears that a mistake was made on Equation 1 and 2.The mean should be used instead of the median that is statistically incorrect to be used with the standard deviation. However, the analysis of new precipitation and temperature indices concluded in the same way.

Authors' changes in manuscript:

Equation 1 and 2 has been corrected. The figure 2 has been updated. Equation and figure captions have been also updated consequently.

Specific comment:

In Eq 3 Xd and Xc are not specified. What is meant by "large scale situation"? Do authors map the anomalies before describing large scale situation?

Authors' response:

The daily large-scale atmospheric situations correspond to the daily fields of anomalies of the predictors with respect to the seasonal cycle. 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.

Authors' changes in manuscript:

In addition to the answer of the reviewer comment about X_{ANA} (see previous comment above) the last paragraph of Section 3.1.1 was completed by: "**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."**

Specific comment:

ANALOG method: the approach of Yiou should be briefly presented. Also, this approach has been criticized. Authors should report about these critics.

Authors' response:

The authors agree that there are many ways to formulate an analog method (e.g. Grenier et al., 2013; Radanovics et al., 2013; Yiou et al., 2013) and that the approach retained here (Yiou et al., 2013) has some particularities as compared to others. Accordingly, a new sentence has been inserted in order to provide the reader with additional details on the ANALOG method used.

However, note that the aim of this study is to set an inter-comparison framework of SDMs through a hydrological point of view. Hence, our goal is not to test and apply all possible SDMs including all their variants. Here we rather want to point out the advantage and inconvenient due to the use of different types of SDMs in hydrology.

Grenier P, Parent AC, Huard D, Anctil F, Chaumont D (2013) An assessment of six dissimilarity metrics for climate analogs. J Appl Meteorol Climatol 52(4):733–752. doi:10.1175/JAMC-D-12-0170.1

Radanovics S, Vidal JP, Sauquet E, Ben Daoud A, Bontron G (2013) Optimising predictor domains for spatially coherent precipitation downscaling. Hydrol Earth Syst Sci 17(10):4189–4208. doi:10.5194/hess-17-4189-2013. http://www.hydrol-earth-systsci.net/17/4189/2013/

Authors' changes in manuscript:

The first sentence of Section 3.1.1 has been amended: "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] x [27.5°N; 50°N] as defined in section 2.2."**

The following sentence has been added in Section 3.1.1 "**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.**"

Specific comment:

In CDF method it is improper to write that it is from "local scale observations". Because authors don't use gauging data (observations) but interpolated data. What is done in Vrac et al. 2012 should be briefly reported here. Otherwise a normal reader of the journal will spend a lot of time in reading the bibliography cited.

Authors' response:

The authors agree that the interpolated data are not the real observed data.
This remark stands also for all the SDMs not only for CDFt.

In consequence, the terms “observations” of “observed” have been replaced by “observation-based” in the appropriate places of the manuscript.
Moreover, if the reviewer wants that we include more details about the CDFt method (as described in Vrac et al., 2012), we do feel that it could increase too much the length of this paper, while we do not want to focus on any specific approach. The description provided in Section 3.1.2., indeed, does not provide all details but this is on purpose: our idea was more to present the philosophy and the key-aspects of this method, without repeating technicalities that can be found in the cited references.

Authors’ changes in manuscript:

The terms observations” of “observed” have been replaced by “**observation-based**” in the following places: P10077 L6, P10083 L27 and P10084 L4

<u>Specific comment:</u>

Authors have to simplify the reading by giving the methodology you used and not always refer to the other works. - The choice of the predictors should be explained as it is the case for example in Vrac et YIOU 2010 (paragraph [13]) http://onlinelibrary.wiley.com/wol1/doi/10.1029/2009JD012871/full

Authors’ response:

Agreed.

Authors’ changes in manuscript:

A table has been added to summarize the predictors and the pre-processing of those predictors according to the SDM and the predictands. Moreover, the following sentence has been added at the in Section 2.2: “**The predictors and the pre-processing of those predictors according to the SDM and the predictands are summarized in the table 1.”**

Table 2 : 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 (anomalies or PC=principal components) 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	-

<u>Specific comment:</u>

Page 10083 “To facilitate interpretation and to limit biases in hydrological modeling when comparing downscaled climate-based hydrological simulations, in the following, the whole period hydrological simulation is used as a reference instead of the observation time series.” This is a critical point. Why do authors do so? Bias in hydrological modeling is generally related to the difference between observations and predictions (of runoff). May authors present the results when the observations are used as reference?

Authors’ response:

We agree that this point is crucial. As mentioned at the end of Section 3.2.3 and in the discussion section, we confirmed that the simulated streamflow produced with the best parameter set for the “whole period” calibration period was used as a benchmark for the comparison between the raw and downscaled datasets from NCEP reanalysis and GCM outputs over the period 1986–2005. In Section

3.2.3, we showed that these simulations were very closed to observed streamflow in the four basins. In this way, we try to overcome uncertainties related to hydrological modeling (including parameter uncertainty, structural uncertainty, experimental uncertainty -errors on the observed streamflow data-, etc.). As the model is quite efficient with observed data as input, we find this approach relevant. Accordingly, comparing with the observed streamflow data, the ranking would be the same with a lower degree of confidence.

Specific comment:

The GR4j (six parameters) was calibrated on 10 days time step. (page 10080). It is important to say this in the abstract.

Authors' response:

Agreed.

Authors' changes in manuscript:

In the abstract, the sentence "Streamflow was simulated using the GR4j conceptual mode" has been replaced by "**The daily GR4j conceptual model was used to simulate streamflow that was eventually evaluated at a 10-day time step**."

Specific comment:

The reference Dezetter and al. 2014 is not possible to download.

Authors' response:

We do not understand why the IAHS press (http://iahs.info/) published only the abstract of the proceedings related to the 7th FRIEND-2014 International conference (Redbook # 363 "Hydrology in a Changing World: Environmental and Human Dimensions", 7–10 October 2014, Montpellier, France, IAHS publ., 363, 355–360). Usually, the entire proceedings are downloadable (see e.g. IAHS Publ. 347). We have asked the editor to update this on the website. Note however that this paper is already referenced in ISI Web of Knowledge and has already been cited three times. If needed, this paper can be easily sent by the authors on demand.

C4792 - Revision note in response to the anonymous review

General comment:

*The presented study of Gruilett et al. (2015) is focussing on the analysis of three different statistical downscaling methodologies as boundary conditions for the lumped hydrological model GR4J (Génie Rural à 4 paramètres Journalier). The presented procedure is introduced as a framework to analyse different downscaling products for climate change impact studies with a sensitivity analysis procedure. Therefore the authors used the reanalysis data set of the National Centres for Environmental Prediction/National Centre for Atmospheric Research (NCEP/NCAR) and two general circulation models (GCM's) the CNRM-CM5 from the French National Centre for Meteorological Research and the IPSL-CM5A-MR of the French IPSL Climate Modelling Centre as input data. The data sets of precipitation and temperature were downscaled with the following three statistical downscaling models (SDM): "analogs of atmospheric circulation patterns" (ANA) "cumulative *distribution function - transform" (CDFt) "stochastic weather generator" (SWG). Because of lag of meteorological observation data in the Marroquin catchment Loukkos a simple module to estimate potential evapotranspiration is implemented in the hydrological model framework. That equation is based on extraterrestrial radiation and temperature. Four Mediterranean catchments located in the western Mediterranean Sea are firstly calibrated/validated with observed station data of 20 years (1986-2005) on a daily time step based on an aggregation of different objective functions (Nash-Sutcliffe, the log version of the Nash-Sutcliffe, the cumulative volume error and the mean annual volume error) with cross calibration – validation scheme of differential split sample testing. Seven parameters were optimised with the shuffle complex evolution algorithm to the complete time series and to dry and wet years. The validated model setups were driven by the BC of the three SDM's of the two GCM's and reanalysis data set plus the pure data sets of GCM's and reanalysis data (RAW). The hydrological outputs are finally analysed based on different quality values (cumulative volume error, RMSE based on sorted data, and a seasonal, high and low flow Nash-Sutcliffe) in comparison with the simulated runoff of the reference period (1986-2005) driven by observed precipitation and temperature.*

Authors' response:

Thank you for these comments, which represent a good summary of the methodology presented in the paper.

General comment:

The manuscript needs improvement in different directions. The authors present a complex scheme, with a lot of information. Here they should reduce the presented data set to a value where the readers still can follow. The Pyrenean catchment Segre was not well calibrated and the reason therefore can be anything. What is the reason that the Pyrenean catchment Segre is responding during the winter and spring period so different from Irati and Herault? I guess it is more affected by snow processes, than the other three. Higher mountain ranges and the more linear morphology of the channel network could be a reason. That would be a hint of the low quality of the observed runoff data or less representative meteorological stations describing the input signal. Here they can start to reduce the presented material.

Authors' response:

The hydrological simulations on the Pyrenean Segre catchment showed indeed less efficiency than in the other studied catchment. This can be explained by many reasons:

- This basin is more snow-dominated than in the others, which leads to more complex hydrological functioning that are not well simulated by the hydrological model.
- There are fewer precipitation and temperature gauges in this basin than in the others. For instance, 2 precipitation gauges (on a total of 6 stations used) are included within the Segre catchment while 10 stations for the Irati catchment.

- The lower quality of the simulation may be attributed to 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.

If the hydrological simulations were less efficient in this catchment than in the others, we found them sufficiently correct to provide an additional catchment for the inter-comparison of the SDMs through a regional analysis in different hydro-climatic contexts.

Authors' changes in manuscript:

The sentence in Section 3.2.3 has been modified: "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**."

<u>**General comment:**</u>

A short description of the two GCMs (CNRM-CM5 from the French National Centre for Meteorological Research and IPSL-CM5A-MR of the French IPSL Climate Modelling Centre) is missing in the manuscript. Abbreviation should be explained.

Authors' response:

Agreed, done in section 2.2.

Authors' changes in manuscript:

The last sentence of the section 2.2 has been modified consequently.

Please find modification in bold.

"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., 165 2013) GCMs, regridded at a 2.5° spatial resolution, over the GCMs historical (or CTRL) period (i.e. 1986–2005)"

<u>**General comment:**</u>

The figures are very complex and need more explanation.

Authors' response:

Agreed, some figure comments have been modified consequently. Please find details in the following answers.

<u>**General comment:**</u>

Scientific English has to be improved and should be reviewed by a native speaker. The authors tend to use long sentences, which were hard to follow.

Authors' response:

The paper has been reviewed by a scientific native English speaker before the submission. Consequently, we believe that English is generally very acceptable. However, we have tried to cut some sentences in order to ease the reading.

General comment:

One major point is that they don't show the differences between observed reanalysis data sets and GCM's. It is important to understand the uncertainties, which arise in the meteorological drivers, before analyzing the hydrological response. They already discuss that in the manuscript at P10091 25-29.

Authors' response:

Since the purpose of the paper is to compare three different downscaling techniques in their ability to provide accurate hydrological simulations, we did not want to develop further a comparison of the raw large-scale climate datasets except through the hydrological responses they provide. In that sense, the comparison between large-scale reanalysis data sets (NCEP/NCAR) and GCM outputs is realized through our hydrological protocol.

Indeed, as explained in the introduction section, we believe it is particularly relevant to propose a selection protocol directly based on the streamflow variable since this variable integrates the combined impacts of the precipitation and temperature variables inputs through the hydrological response. Moreover, the streamflow variable is the most suitable for quantifying the impact of the bias of the downscaling techniques on key issues for water management related to surface water availability and high and low flow events. Consequently, we do think that the originality of our paper lies on the hydrological assessment of different statistically downscaled datasets that were preliminary calibrated and validated by the climatologists who co-authored the paper. Moreover, we think that presenting the ability of different statistically downscaled datasets to reproduce for instance the inter-annual and seasonal hyetograph or the distribution of precipitation extremes would significantly increase the paper length while potentially confusing the purpose.

General comment:

The other point is that it is rather unfair to compare one bias corrected SDM (CDFt) with two uncorrected ones. It is like comparing apples with oranges. For a revised manuscript all SDM should be treated equivalent.

Authors' response:

The three statistical downscaling models (SDMs) are based on different concepts:

- ANALOG is based on analog circulation determination;
- SWG is a stochastic weather generator conditional on large-scale information;
- CDFt is a quantile-mapping approach performed over the projection period (large-scale and local-scale) CDFs – and not over the calibration period CDFs as in the classical quantile-mapping (see e.g., Vrac et al., 2012).

Although CDFt is derived from the quantile-mapping technique (that is classical in bias correction methodologies), we insist on the fact that those three models (i.e., CDFt included) have all the particularity of providing high-resolution precipitation and temperature simulations (constrained by large-scale reanalysis or GCM data). Therefore they all belong to the family of the statistical downscaling methods. In any way, CDFt is NOT a "bias corrected" SDM as understood by the reviewer. Actually, none of the three SDMs is bias corrected.

Thank you for allowing us to clarify this point that was indeed potentially confusing.

Moreover, from a strictly technical point of view, it is absolutely impossible to treat the three SDMs with exactly the same information as input (i.e., predictors) since they are built from different philosophies and therefore different constraints (see Vaittinada Ayar et al., 2015).

Therefore, those three SDMs have been treated equally in the sense that we tried to calibrate them as good as possible to make their downscaled simulations representative of what they can really generate in their optimal version.

With all those points clarified (three SDMs, calibrated as good as possible, no specific bias correction to any of them, etc.), we clearly do not have the feeling to compare apples and oranges…

Authors' changes in manuscript:

This sentence has been added at the end of the section 3.1 to clarify this point: **"Although CDFt is derived from the quantile-mapping technique (that is classical in bias correction methodologies), 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 reanalysis or GCM data) and therefore belong all to the family of the statistical downscaling methods."**

General comment:

They should think about reducing the amount of study sites and maybe integrate one or two additional hydrological models to give a broader view on the uncertainties, which arise through hydrological modelling via the model framework.

Authors' response:

Regarding the reduction of the amount of study sites, we do believe that it was important to consider four different catchments with various hydro-climatic conditions even if hydrological simulations were less efficient on one of them (Segre).

Obviously a broader view on the uncertainties could be provided, notably by exploring in details the uncertainty that arises from hydrological modelling, for instance by using different hydrological models. However, as it is stated in the discussion section, we think it is far beyond the scope of this paper. While we tried to preliminary investigate issues regarding parameter identifiability under climate-contrasted conditions (see 3.2.3), we showed that the model was rather efficient either on dry or wet years. Consequently, even if the uncertainty stemming from hydrological modeling cannot be ignored (as stated in the discussion section p. 10091 l25-29), we assumed here that it was not meaningful in the framework of the comparison of downscaling techniques through the proposed protocol.

Specific comment:

P10070, 26-29: Prove English

Authors' response:

Agreed.

Authors' changes in manuscript:

The sentence "Difficulties in choosing one SDM among several may arise from the choice of criteria which may be relevant from the statistical or climatological point of view, but may not adequately highlight the differences between the methods with respect to the hydrological responses with respect to the main CCIS issues." has been replaced by "**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.**"

Specific comment:

P10073 L20 It is more important to show how many stations of the measurement network could be used for the catchment, than how many stations are available in the complete Ebro catchment.

Authors' response:

Agreed. Done.

Authors' changes in manuscript:

The sentence "The precipitation and temperature data were extracted from respectively 818 and 264 stations available at the Ebro basin scale (Dezetter et al., 2014)." has been replaced by "**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**."

Specific comment:

P10073 L22 and L27: How are the laps rates estimated or from which source are they taken?

Authors' response:

Lapse rates were estimated in the mentioned publication (Dezetter et al., 2014).

Specific comment:

P10074 L16 and P1076 L8: DJF, MAM, JJA, SON is not helpful and can be deleted

Authors' response:

Agreed.

Authors' changes in manuscript:

This has been deleted in both locations.

Specific comment:

P10074 L16-20: It is hard to follow that sentence. It needs improvement. How has the regridding been conducted to the GCM to a resolution of 2.5°? How many km are 2.5°? Explain the abbreviation CTRL

Authors' response:

The regridding was done through a "largest area fraction remapping" (consisting in taking the native grid cell value with the largest area fraction for each target grid cell) 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.

Moreover, "CTRL" is as classical term in GCM terminology. It means "control". A CTRL run is a GCM simulation run performed over a historical time period. To prevent any confusion, the abbreviation CTRL was removed from the manuscript to keep historical time period.

Authors' changes in manuscript:

This sentence has been added at the end of Section 2.2:

"**The regridding was done through a largest area fraction remapping (consisting in taking the native grid cell value with the largest area fraction for each target grid cell) 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.**"

Specific comment:

P10075, 3-4, 10-11: Check English

Authors' response:

Agreed. Done.

Authors' changes in manuscript:

The sentence "In the Herault and the Irati basins, peaks in spring and fall precipitation are produced by precipitation events whose intensity can vary greatly over short periods." has been replaced by "In the Herault and the Irati basins, **the precipitation** peaks in spring and fall are produced by events whose intensity can vary greatly over short periods."

The sentence "Furthermore, analysis of the precipitation indices (Eq. 1) showed that the wet and dry years in the four basins were the same in nearly half the years (Fig. 2)." has been replaced by "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 (Fig. 2)."

<u>Specific comment:</u>

P10075, L12: Figure 2 is hard to interpret. I cannot identify that 50 % of the catchments respond similar in time. But is that information important for the manuscript?

Authors' response:

The gray lines in Figure 2 underline the years that are equivalently dry or wet, and cold or warm for all the basins. For example, a dry year (according to the precipitation index) for the 4 basins is highlighted in gray, as well as for a cold year (according to the temperature index) for the 4 basins in the same time.

First, the analysis of precipitation and temperature indices underlines that no climate trend is observed over the study period. Secondly, it highlights a relative climate consistency between the basins, despite their different geographical characteristics in the Mediterranean.

Authors' changes in manuscript:

The sentence "Furthermore, analysis of the precipitation indices (Eq. 1) showed that the wet and dry years in the four basins were the same in nearly half the years (Fig. 2)." has been replaced by "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). This analysis shows that no climate trend is observed over the study period, and highlights a relative climate consistency between the basins, despite their different geographical characteristics in the Mediterranean region."

<u>Specific comment:</u>

P10075 P17: It is statistically not perfect to use the combination of median and standard deviation and could lead to irritations. Why do they use the median and the standard deviation and not the average with the standard deviation or median with MAD?

Authors' response:

Agreed. After checking, it appears that a mistake was made on Equation 1 and 2. The mean should be used instead of the median, which is statistically incorrect to be used with the standard deviation. However, the analysis of new precipitation and temperature indices concluded in the same way.

Authors' changes in manuscript:

Equation 1 and 2 has been corrected. The figure 2 has been updated. Equation and figure captions have been also updated consequently.

<u>Specific comment:</u>

P10076, L 1: What is 0.44° in km?

Authors' response:

0.44° = 48.926 Km

Authors' changes in manuscript:

The sentence "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 reanalyze grid scale (0.44° spatial resolution)." has been replaced by "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)."

Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones and M. New. 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J. Geophys. Res (Atmospheres), 113, D20119, doi:10.1029/2008JD10201

<u>Specific comment:</u>

3.1.4 and 3.1.3 3.1.4: is only important for the SWG SDM. For sake of simplicity I would merge the two parts and start with the modelling of the occurrence of precipitation.

Authors' response:

Very good point. This has been done.

Authors' changes in manuscript:

Section 3.1.3 and 3.1.4 have been merged.

<u>Specific comment:</u>

P10077 L15-18: I cannot follow. The SDM is calibrated with the GCM and there is a link to a bias correction? Pleas clarify for all SDM's, how they are calibrated and validated, which data was used, etc.

Authors' response:

As explained earlier in the present document, ANALOG, SWG and CDFt are three SDMs, but they differ in their philosophies and constraints: 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.

For clarification, the notion of "bias-correct" has been removed from this sentence. This was indeed somehow confusing.

To summarize how the calibrations are performed:

- For ANALOG, the calibration is performed on NCEP reanalyses;
- For SWG, the calibration is performed on NCEP reanalyses;
- For CDFt, the calibration is performed directly on the GCM to downscale.

For all the three models, calibration is done over 1976–2005 (for Herault, Irati and Segre, but for 1986–2005 for Loukkos due to data availability) and evaluation is performed with GCM data as input over the 1986–2005 time period to have a common 20-year evaluation period.

Authors' changes in manuscript:

The end of the section 3.1 has been amended with this paragraph: "**For ANALOG 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 reanalysis 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."

The end of the section 3.1.2 has been modified: "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."**

<u>Specific comment:</u>

P10080 L15: What is the reason for the average over 10 days for calibration? The model was not able to represent small runoff effects in time?

Authors' response:

The 10-day time step was retained because it constitutes an interesting compromise for Climate Change Impact Studies 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. This time step leads more easily to realistic hydrological simulations than with a daily time step, while providing a better insight on the hydrological variability than the monthly time step.

Authors' changes in manuscript:

The following sentence was added in the text to justify this time step.

"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**."

<u>Specific comment:</u>

P10082 L8: What are the criteria's of a dry and wet year?

Authors' response:

A modification of the previous sentence (P10082 L4) should clarify this point.

Authors' changes in manuscript:

The sentence P10082 L4 "Thus, two sub-periods of 10 years each divided according to the median annual precipitation for the period were used either for calibration and for validation." has been replaced by "Thus, two sub-periods of 10 years each divided according to the median annual precipitation for the period were used either for calibration **or** for validation. **These two sub-periods define dry and wet year periods."**

<u>Specific comment:</u>

P10082 L11: the hydrological year after American and British system is from the first October to the 30iest September. Just to prevent confusion, the specific system which was used (France?) should be provided or the standard should be used.

Authors' response:

Based on hydrological situations, September 1 or October 1 can be selected as the start of a hydrological year. In our case, September 1 is typically a low-flow period while October 1 can register significant precipitation because of the Mediterranean climate context. That is why September 1 was used in this study to limit memory effects from one year to another in the calibration/validation DSST process.

Authors' changes in manuscript:

The following sentence was modified to enhance this point:

"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."

<u>Specific comment:</u>

P10082 L21-23: The difference between what? Validation to calibration? In the figure 4 only calibration or validation is presented. In text and caption the information is missing what they present. I would present both calibration and validation.
Authors' response:
To be more precise: whatever the calibration period used (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.
In the figure 4, validation of hydrological modeling is presented using parameter sets provided by the calibration step.
We agreed that the caption of the figure 4 needs to be improved to precise if calibration or validation is concerned.
Authors' changes in manuscript:
The sentence "The differences between the F_{obj} of the validation simulations never exceeded 0.1 (except the Segre basin in the wet year validation period) emphasizing the stability of the simulations under different hydro-climatic conditions" has been modified by "**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**."
"**Cross calibration/validation of the hydrological model**" has been added at the beginning of the caption of the figure 4.

Specific comment:
P10083 L9-15: Prove English, split sentences. As far as I understand the authors correctly they use the simulated runoff data instead of the observed data to minimise the errors.
Authors' response:
Agreed.
Authors' changes in manuscript:
The following paragraph has been modified to clarify this point:
"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.**"

Specific comment:
P10084 L10 Equation of the NRMSE is missing.
Authors' response:
Agreed. The equation of NRMSE can help the reader.
Authors' changes in manuscript:
The following equation has been added in Section 3.3.

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

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

Specific comment:
P10085 L6: Check English
Authors' response:
Agreed.
Authors' changes in manuscript:
The following paragraph has been modified to clarify this point:
"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**."

Specific comment:
P10085 L6-L11: That block is already in the caption of the figures.
Authors' response:
OK, this information is repeated in the caption of the Figure 5 in order to ease its comprehension. It can be removed but we think it is useful both to ease the text readability and to assist the reader in interpreting the figure.
Authors' changes in manuscript:
The figure 6 has been deleted. Please find more explanations in the next Specific Comment.

Specific comment:
P10085 L15-17: That is not presented in the manuscript, but would be essential to prove the results of meteorological drivers. In figure 6 only the data of the reanalysis is shown, which gave no hint about the effect of the two GCM's.
Authors' response:
This figure underlines how the hydrological indicators have been evaluated for every downscaled or raw climate data (reanalysis and 2 GCMs) on the four basins. For instance in the figure 6, we have deliberately chosen to present the interim results of one climate dataset (NCEP/NCAR) and one of the four basins (Herault). Obviously, displaying 12 detailed graphs (3 climate datasets x 4 basins) would not have been concise and readable. This introductive section (and related graph) aimed at helping the reader to understand how the hydrological indicators had been evaluated before being aggregated in the discussion part of the paper.
However, we agree that showing a unique example in the beginning of the result section can lead to misunderstanding. So we decided to delete the Section 4.1 and the figure 6.
Authors' changes in manuscript:
We deleted Section 4.1 and figure 6. Section and figure numbers have been updated consequently.

Specific comment:
P10086 L6: Unclear, add a table.
P10086 L15 the section is hard to follow. An additional table with the specific values would be helpful to check the mean statistics of the volume performance.
Authors' response:
Agreed.
Authors' changes in manuscript:
This table has been added in Section 4.1 Water volumes.

This table has been called in the following sentence “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)**”

Table 2: Cumulative volume error (VEC) 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%

Specific comment:

P10086 L16-18: The outliers' are not clear for me, does that mean in case the simulated absolute value per time step increases 50% of the simulated runoff driven by observations is classified as an outlier and in that case not taken into account? These values need to be presented in the figure or a table. But in the presented form it is unclear.

Authors' response:

In this case, outliers are simply VE_C values exceeding 50%. This threshold of 50% is only used to help the reader to understand which criteria value is acceptable or not, in our point of view.
Nevertheless, we agree that the sentence introducing outliers was not clear.

Authors' changes in manuscript:

The sentence “In addition, the results of ANALOG-based simulations were more constant without outlier criterion values. Criterion values are considered as outliers when VE_C is greater than 50 %.” has been replaced by “In addition, the results of ANALOG-based simulations were more constant**, i.e.** 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**.”

Specific comment:

P10087, L23-25: Improve English, hard to follow. SWG is the worst of the SDM's but it outperforms still the raw data sets and it tends to overestimate the volume.

Authors' response:

Agreed. This part needed more explanation.

Authors' changes in manuscript:

The sentences “Except with NCEP, SWG-based simulations reproduced seasonal variability poorly, more in terms of intensity than occurrence: as a result, with this SDM, the shape of the streamflow seasonality was reasonably well reproduced but not the values of discharge.” have been replaced by “Except with NCEP, SWG-based simulations **reproduced poorly the seasonal variability of runoff, due notably to systematic overestimation of high-flow events.”**

Specific comment:

P10088 L13-14: Why is only CDFt affected by snow processes?

Authors' response:

In fact, the reproduction of high flows is also less efficient in the basin with the ANALOG method. This is probably due to the fact that the hydrological model is less efficient 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.

Authors' changes in manuscript:

Consequently, the sentence in Section 4.5 "Nevertheless, CDFt appeared to be less able to reproduce high flows in the Segre basin characterized by a hydrological context including snowmelt." has been modified by "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.**"

Specific comment:

P1088 l9 and L18: The explanation of the achievement of the NSE criteria is missing: 0.5 for high flows and 0.8 for low flows. Is that information important? There is no additional use of those criteria.

Authors' response:

We agree that these thresholds are not necessary for the comment.

Authors' changes in manuscript:

Consequently, the text in Section 4.5 has been modified so as to remove the two related sentences:

"Due to the nature of the "high flows" indicator and the NSE criterion used to evaluate it, the reproduction of high flows was considered to be satisfactory for NSE values greater than 0.5."

"The reproduction of low flows was considered to be satisfactory when NSE values were higher than 0.8."

Specific comment:

P10090 L16-20: It is not clear for me if the method CDFt has an automatically bias correction including that a similar procedure is not used for the other SDM's. In case of the SWG which is the weakest approach it is unfair to use not bias corrected data sets.

Authors' response:

As explained earlier in the present document, ANALOG, SWG and CDFt are three SDMs, but they differ in their philosophies and constraints: 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.

In any way, CDFt is NOT a "bias corrected" SDM. Actually, none of the three SDMs is bias corrected.

However, the question of using a bias correction step into a SDM approach is very interesting

This leads to the question of bias correcting the large-scale GCM data (with respect to NCEP) before applying a downscaling procedure. This is clearly out of the scope of this paper but this is discussed in Section 5 "Discussion and conclusion".

Authors' changes in manuscript:

The end of the section 3.1 has been amended with this paragraph: "**For ANALOG 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 reanalysis 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 for all catchments (except for the Loukkos on which calibration was limited over 1986–2005 due to data availability). Their assessment when**

applied to NCEP reanalysis and GCM data was performed according to the common 20-year 1986–2005 evaluation period. Sections 3.1.1 to 3.1.3 describe the different models."

Specific comment:

P10090 L21: Although

Authors' response:

"Although" is correctly wrote in the source file. We will check with the editor next time.

Specific comment:

P10090 L21-26: But in that study GCM-SDM tandem is not used to predict data and the Nash of 10 days does not allow such interpretations due to the smoothing. That is part of the description of the model not of the discussion conclusion.

Authors' response:

We agree that the GCM-SDM tandems were not used in this study to provide future climate projections, but were used in a sensitivity analysis over a 30-year climatic reference period. We also agree that the reproduction of daily hydrological extreme events can be smoothed by a larger time step (10-day time step in this study) in the analysis of the seasonal hydrographs.

However, this sentence (P10090 L21-26) tends to underline the fact that the ANALOG method globally better performed than the other methods over the reference period, in terms of water volumes, seasonal and interannual distributions and extreme events such as high and low flows, analyzed at the 10-day time step. Nevertheless, to provide climatic projections at a mid or long term horizon, the ANALOG method is facing some limitations. In particular, as shown by Teng et al. (2012), this method is not able to provide suitable simulations for extreme events if such events increase in intensity in the future.

This point needs to be mentioned here to underline the fact that the CDFt method, whose results are close to ANALOG ones, does not face such limitations, as stated in the discussion section. Accordingly, we assumed that these limitations have to be mentioned in the discussion section rather than in the methodology section. Indeed, this helps to qualify the quality of the results obtained with the ANALOG method while providing elements of comparison with the CDFt method.

Specific comment:

P10091 L26 I would not write gas emission scenarios, which are the old IPCC scenarios. I would keep it broad and general to all scenario types.

Authors' response:

Agreed.

Authors' changes in manuscript:

We have simplified the text by deleting the list of the sources of climate modeling uncertainty.

The sentence "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…" has been modified in "Although it is commonly acknowledged that the uncertainty resulting from climate modeling is highest in a context of climate change…"

Specific comment:

P10091 L25-29: That sentence needs simplification, modification and splitting. Here arises the question, why the uncertainty of the GCM's compared to the reanalysis data set is not presented. The uncertainty of the boundary conditions could be used to clarify the range of the uncertainty of hydrology, by expecting that GR4J is a perfect model. They could easily show the uncertainty in the drivers and the used model.

Authors' response:

The sentence has been modified (see last specific comment).
Moreover, we added a comment on the uncertainty that was highlighted regarding the GCM outputs in comparison with the use of reanalyses.
However, we do not clearly understand how we could clarify the range of hydrological uncertainty. We think that a specific study on the whole uncertainties that arise in CCIS (including the hydrological model uncertainty) is far beyond this paper and could not be easily highlighted here.

Authors' changes in manuscript:

Modification in bold have been added in the discussion section:

"**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 for the evaluation of future hydrological impact in CCIS. Finally,** although it is commonly acknowledged that the uncertainty resulting from climate modeling is highest in a context of climate change (e.g. Wilby and Harris, 2006; Arnell, 2011; Teng et al.,2012)…"

Specific comment:

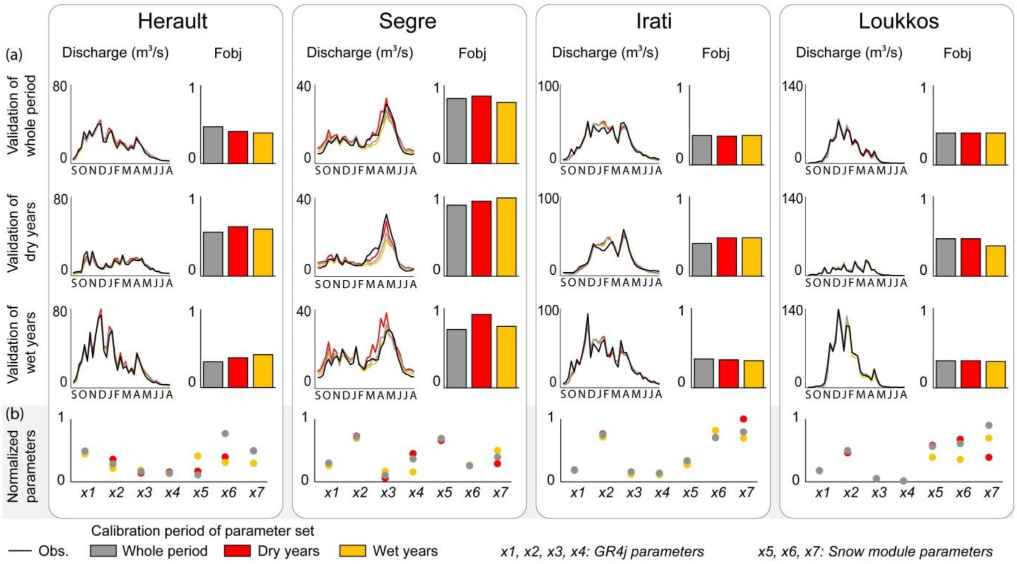
Figure 4: The differences are hard to prove especially for the low flows. A log scale here would be helpful. The line in the parameters suggested that they are related, which they are hopefully not. They should use point symbols instead of lines.

Authors' response:

The figure 4 aims at showing how robust is the hydrological model under contrasted hydro-climatic conditions. We assumed that this goal was achieved. Increase readability in low flow part of the graphs with a log scale for example was not done because this criterion was not considered as discriminant visually. Moreover, NSE_{log} criterion used in the objective function F_{OBJ} already attempts to highlight low flows. However, this figure illustrates the lower quality of the hydrological simulations in the Segre basin including low flows.
On the other hand, the reviewer is absolutely right about the choice of curves instead of points on the graphs of "Normalized parameters". This can actually suggest a correlation between them, which is obviously not the case.

Authors' changes in manuscript:

Thus the figure 4 has been changed accordingly.


Herault
Segre
Irati
Loukkos
(a)
Discharge (m³/s)
Fobj
Validation of whole period
Validation of dry years
Validation of wet years
SONDJFMAMJJA
(b)
Normalized parameters
x1 x2 x3 x4 x5 x6 x7
Calibration period of parameter set
Obs.
Whole period
Dry years
Wet years
x1, x2, x3, x4: GR4j parameters
x5, x6, x7: Snow module parameters

Manuscript prepared for Hydrol. Earth Syst. Sci.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 19 December 2015

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, interannual 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 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. 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 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 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 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”, 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 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 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.

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 $\mathrm{m\,asl.}$ 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 $\mathrm{m\,asl.}$, is characterized by basement rocks (granite and quartzite) and a rugged topography that favors runoff. The Irati basin, from 407 to 2 017 $\mathrm{m\,asl.}$, 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 $\mathrm{m\,asl.}$, 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 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] × [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° spatial 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).

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

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

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

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

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:

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

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.

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 j^{th} 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 $2\ \mathrm{m}$ 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).

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

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

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

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

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.

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

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

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) algorithm 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 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 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 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.

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.

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 95$^{\text{th}}$ 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 80$^{\text{th}}$ 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 climate 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 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 interannual 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.

Acknowledgements. This work was part of the StaRMIP project (Statistical Regionalization Models Intercomparisons and hydrological impacts Project, grant agreement ANR-12-JS06-0005-01), and the REMEMBER project (grant agreement ANR- 12-SENV-0001-01), both funded by the French National Research Agency (ANR), as well as part of the GICC REMedHE project (2012–2015) funded by the French Ministry of Ecology, Sustainable Development and Energy and the ENVI-Med *CLIHMag* (*Changement cLimatique et Impacts Hydrologiques au Maghreb*) project funded by the program INSU-MISTRALS. The authors are grateful to *Météo-France*, the AEMET (*Agencia Estatal de Meteorología*), the department of Water Research and Planning (DRPE) of Morocco and the Hydraulic Basin Agency of Loukkos-Tetouan for having provided the observed hydro-climatic data.

References

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 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.

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 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 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 comparison, 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. 7^{th} FRIEND Int. Conf., Montpellier, France, 7–10 Oct. 2014). IAHS Publ., vol. 363, pp. 355–360, 2014.

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-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., Friedlingstein, 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, l11708, 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.

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.

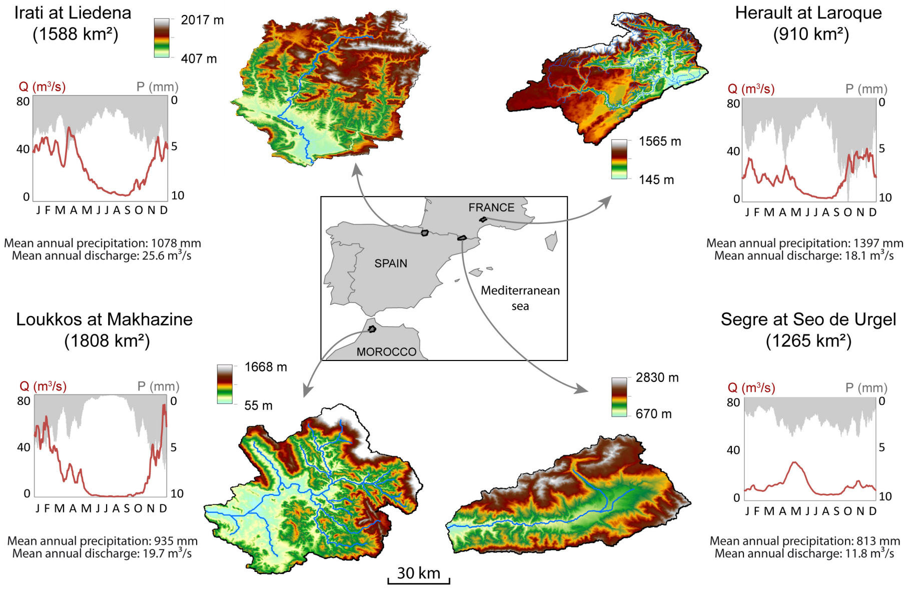


Figure 1. Study catchments (Herault, Segre, Irati and Loukkos) in the western Mediterranean region with their topography and mean seasonal variability in precipitation (P) and discharge (Q) for the period 1986–2005.

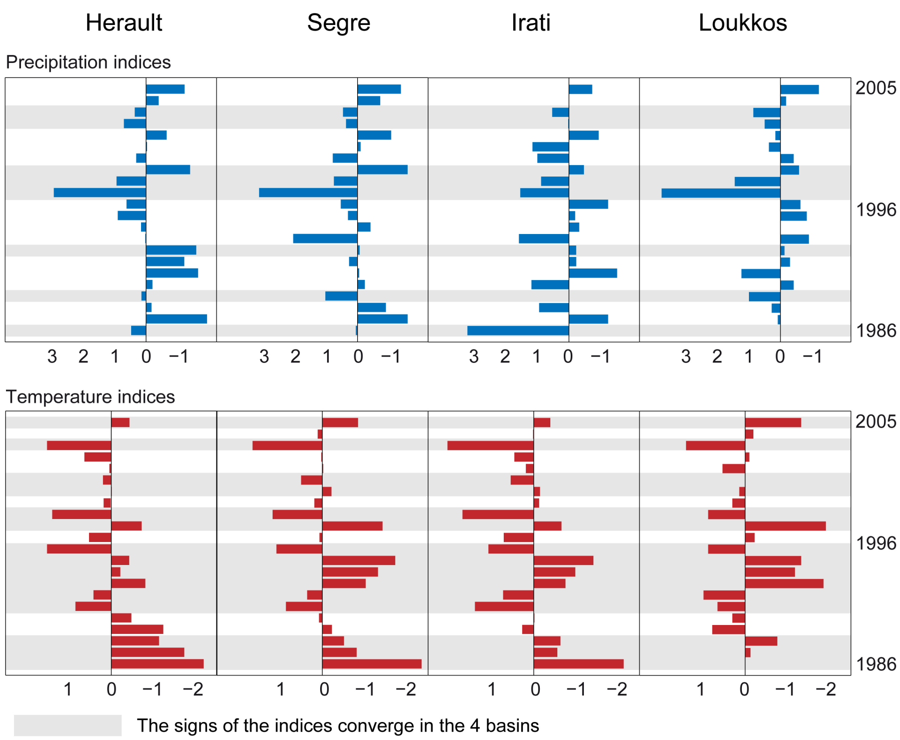


Figure 2. Precipitation ($I_P = (P_y - \overline{P_y})/\sigma_P$) and temperature indices ($I_T = (T_y - \overline{T_y})/\sigma_T$) applied on the four basins over the 1986–2005 period. The grey lines highlights years when the signs of the indices are the same for the four basins. P_y is the annual precipitation 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.

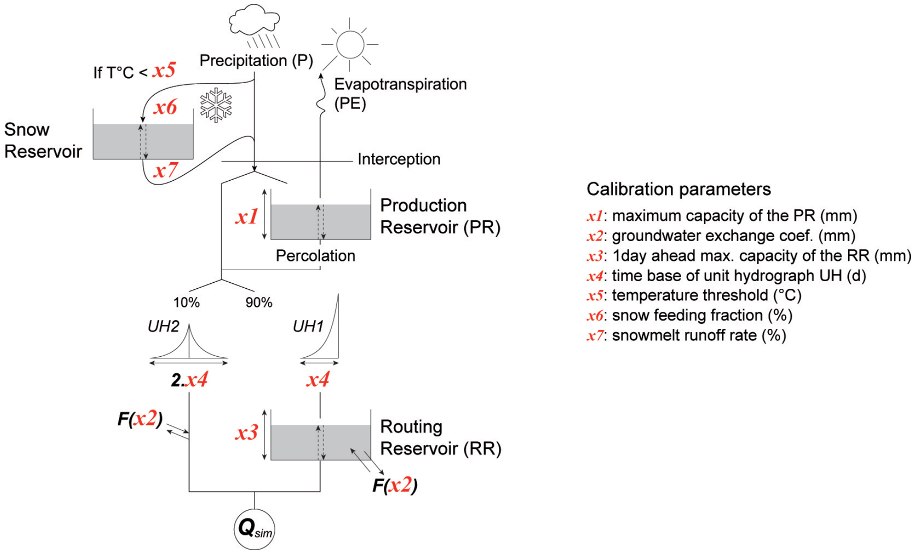


Figure 3. Schematic diagram of the hydrological model GR4J. Adapted from Perrin et al. (2003); Ruelland et al. (2011).

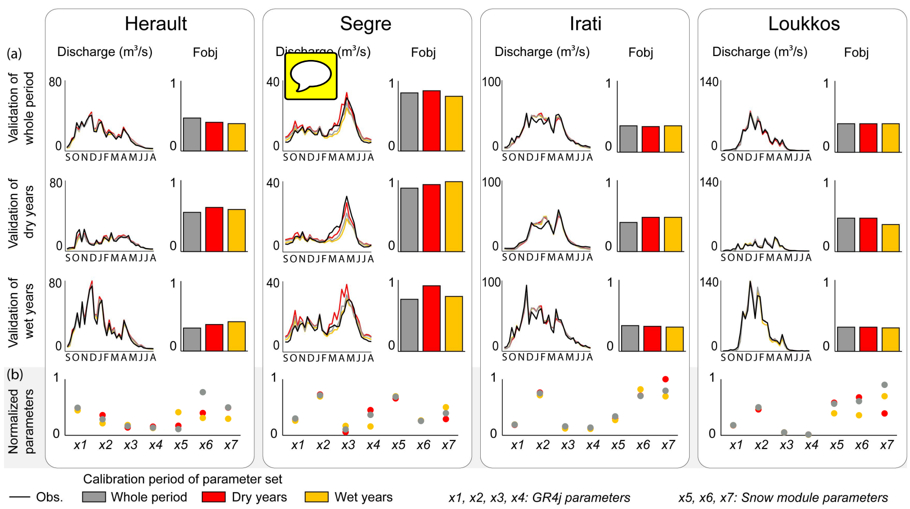


Figure 4. Cross calibration/validation of the hydrological model. (a) Seasonal representation (from September to August) of simulated and observed runoff during the whole period (WHO, first row), dry years (DRY, second row) and wet years (WET, third row) according to parameter sets optimized respectively 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.

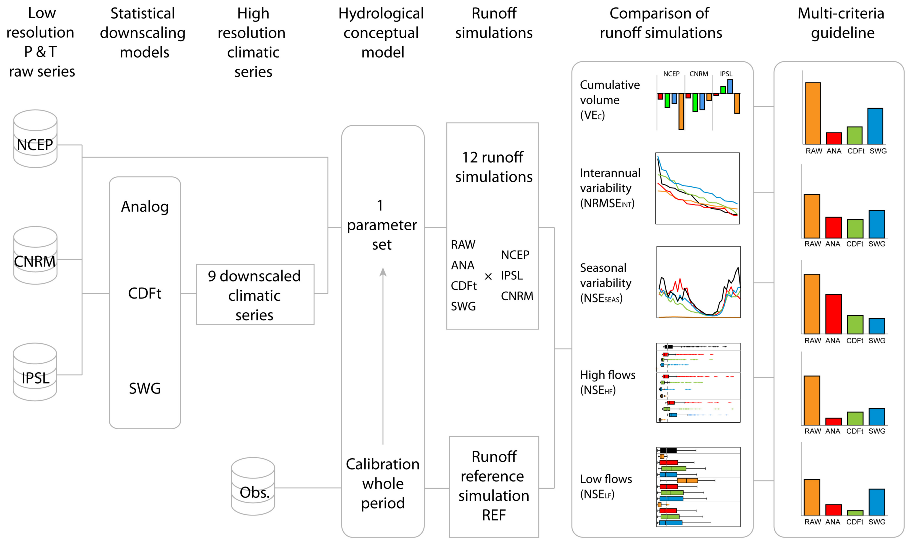


Figure 5. Flow chart illustrating the method used to compare the three downscaling methods through a hydrological sensitivity analysis.

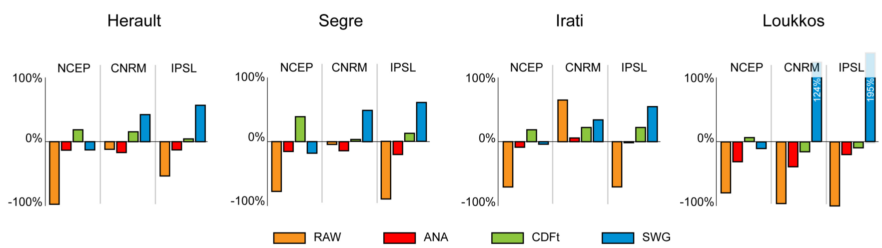


Figure 6. Comparison of the downscaling methods according to the cumulative volume error (VE_C) used as criterion to compare the downscaling methods applied to NCEP, CNRM and IPSL climate inputs in the four basins. The smaller the absolute value of the criterion, the better the simulation.

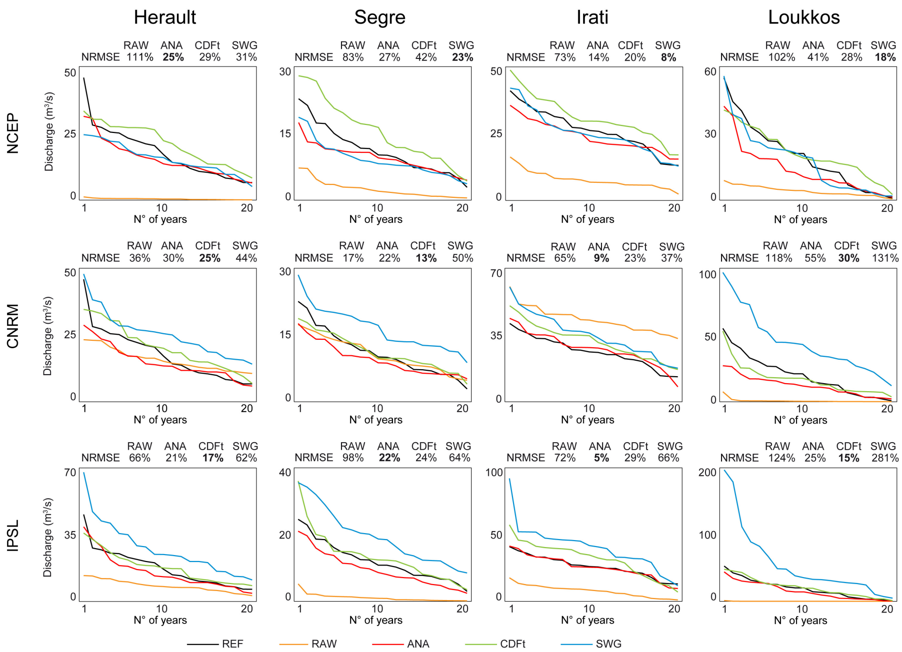


Figure 7. Comparison of the sorted annual discharge simulated using REF data, RAW (NCEP or GCM) data, and the three downscaling methods (applied to NCEP, CNRM and IPSL) for each basin. The NRMSE values above each panel represent a root mean square error applied to the sorted time series of annual discharge normalized by dividing RMSE by the mean annual discharge of the reference time series. The best values are in bold.

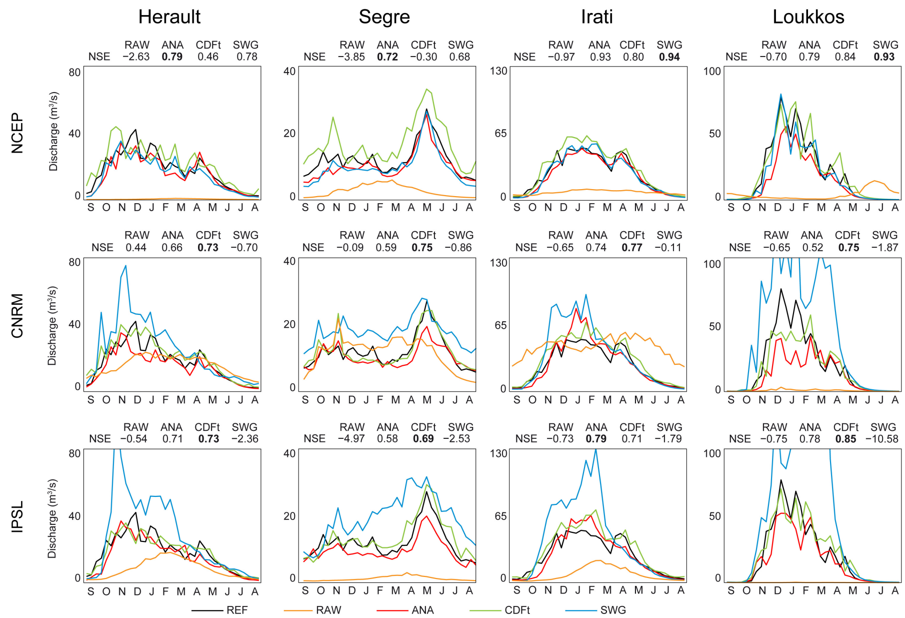


Figure 8. Comparison of seasonal variations in streamflow simulated using REF data, RAW (NCEP or GCM) data, and the three downscaling methods (applied to NCEP, CNRM and IPSL) for each basin. The NSE values for the mean 10-day discharge between REF and the simulation concerned are given above each panel. The best values are in bold.

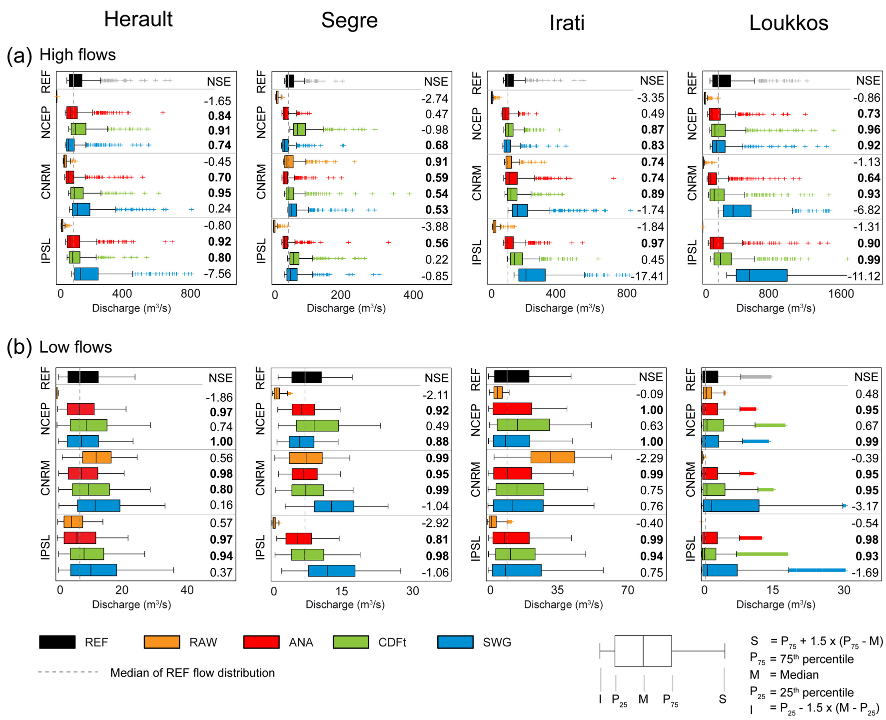


Figure 9. Comparison of (a) the 5% daily high flows and (b) the 80% daily low flows simulated with REF data, RAW (NCEP or GCM) data, and the three downscaling methods (applied to NCEP, CNRM and IPSL) for each basin. The NSE values calculated on the 5% high and the 80% low flows are indicated on the right in each panel. NSE values higher than 0.5 for high flows and 0.8 for low flows are in bold.

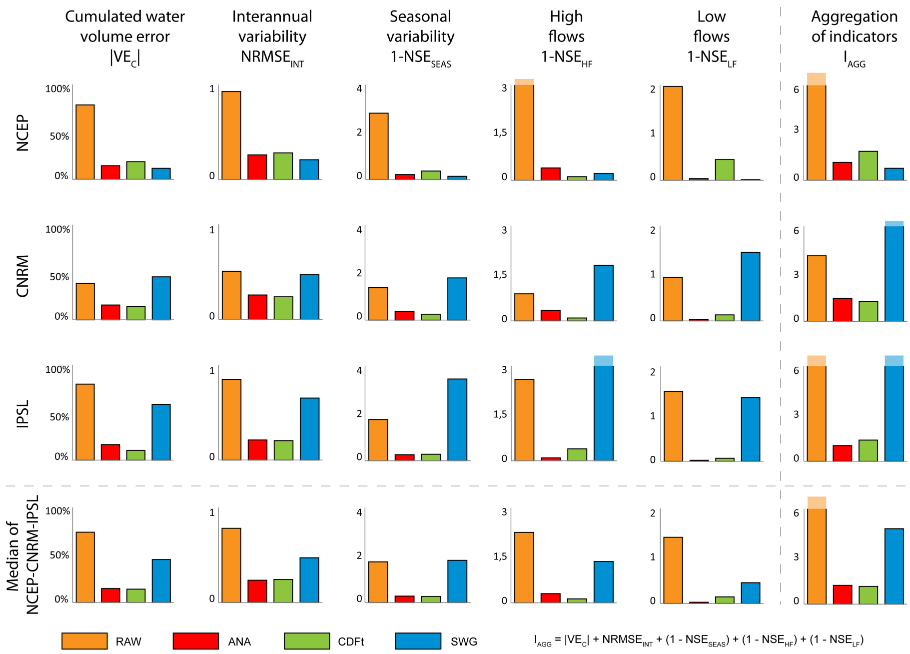


Figure 10. Efficiency of the different climatic datasets to reproduce different aspects of the hydrographs from the four basins over the period 1986–2005: comparison of low resolution datasets (RAW) and high resolution datasets downscaled using the ANALOG, CDFt or SWG methods forced by NCEP/NCAR reanalyses 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: $I_{AGG} = |VE_C| + NRMSE_{INT} + (1 - NSE_{SEAS}) + (1 - NSE_{HF}) + (1 - NSE_{LF})$.