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.

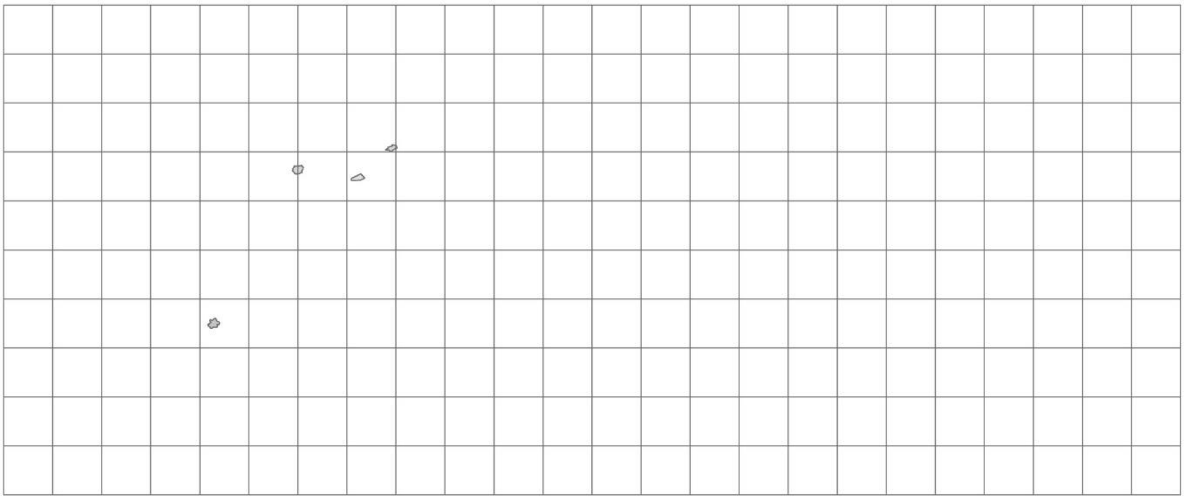

Figure 1 : Location of the four basins in the GCM grid. Loukkos, Irati, Segre, Herault from the west to the east respectively.

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."

<u>Specific comment:</u>

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.**"

<u>Specific comment:</u>

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 1^{st} of July 2002, only the time period 1976-2001 U 2003-2005 is considered and only the days between June, 15^{th} and July ,15^{th}. 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 $\mathbf{X_{ANA}}$** 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:

<u>Specific comment:</u>

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 $\mathbf{X_{ANA}}$** 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:

<u>Specific comment:</u>

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."**

Specific comment:

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

Specific comment:

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**".

Specific comment:

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.

Specific comment:

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 ?

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.**"

<u>Specific comment:</u>

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.**"

<u>Specific comment:</u>

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

Specific comment:

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	-

Specific comment:

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.