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**."

General comment:

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)"

General comment:

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.

General comment:

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)."

__Specific comment:__

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."

__Specific comment:__

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.

__Specific comment:__

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

Specific comment:

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.

Specific comment:

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.

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

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.**"

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

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.**"

<u>Specific comment:</u>
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."

<u>Specific comment:</u>
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:

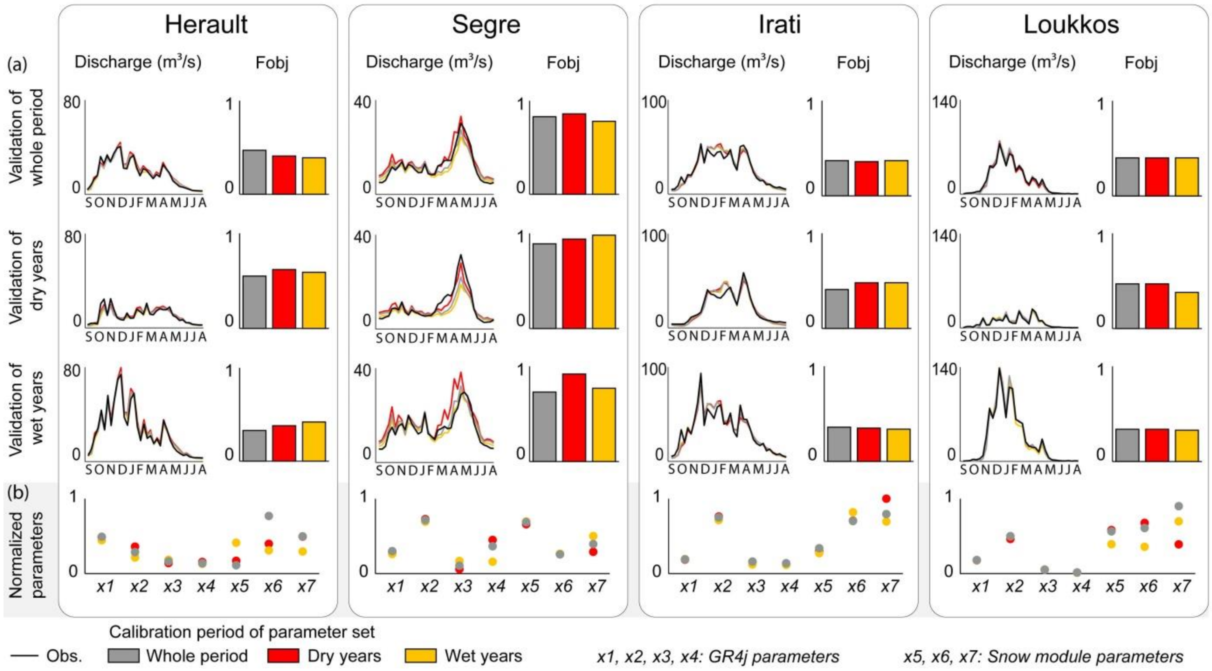
Figure 4: The differences are hard to prove especially for the low flows. A log scale here would be helpful. The line in the parameters suggested that they are related, which they are hopefully not. They should use point symbols instead of lines.

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