Magali Troin
CEREGE
UMR 6635 CNRS-IRD-CDF
Europôle Méditerranéen de l'Arbois BP80
13545 Aix-en-Provence cedex 4
France
troin@cerege.fr
tel. +33 (0) 4 42 97 15 89

Aix-en-Provence, 4th March 2011

Bart van den Hurk, Editor
Hydrological and Earth System Sciences

Dear Sir,

Please, find attached the manuscript entitled: "Coupling statistically downscaled GCM outputs with a basin-lake hydrological model in subtropical South America: evaluation of the influence of large-scale precipitation changes on regional hydroclimate variability" by Magali Troin, Mathieu Vrac, Myriam Khodri, Christine Vallet-Coulomb, Eduardo Piovano, and Florence Sylvestre that we would like to re-submitted to *Hydrological and Earth System Sciences*. Corresponding author: Magali Troin.

The initial version of the manuscript has been reviewed by two reviewers. We addressed each comment and question asked by each reviewer. We appreciate the reviewers' comments and suggestions, which helped us improving the quality of the manuscript and strengthened the justification of our approach and scientific results.

We have attached below our detailed response to each major and minor criticism or recommendation.

We hope this paper is now suitable to *Hydrological and Earth System Sciences*. We thank you in advance for your time and for considering our manuscript.

Yours sincerely,

Magali Troin

Detailed responses to reviewer comments #1

EVALUATION OF THE RESULTS AND DISCUSSIONS

The reviewer disagrees on the optimistic way of presenting the results and the discussion. This should be entirely rewritten and maybe some new tests are necessary.
As recommended by the reviewer #1, we have re-written some sentences in paragraphs corresponding to the results and the discussion (§ 5 and 6) to give a more critical evaluation of our results.

The proposed downscaling method (CDF-t) probably greatly improves the characteristics at station level. This is however not evaluated.
In order to evaluate the downscaling method and put more emphasis on its interest in our study, we have decided to present the results of the hydrological simulations forced directly by GCM without dowsncaling. The results are presented in paragraphs 5.2 and 5.3.

It seems that the downscaled variables are aggregated (probably arithmetic mean) before evaluating. Explain why and how this is done.
The aggregated downscaled variables however explain less than 50% of the variance of, which is very low. And this is only for the calibration period. For the validation different metrics are used than for the calibration. Why? For comparisons reasons it would be better to also present the explained variance. In stead, some metrics like the average and standard deviation are compared. The metrics in tables 5 and 6 are evaluated very briefly. However, large discrepancies exist between downscaled and observed variables. These discrepancies need to be interpreted and the application of these "biased" data in the hydrological model needs to be justified.
In the new version of the manuscript, we have used the same metrics for the evaluation of the downscaling method during the calibration and validation (Tables 3 to 8 and appendixes A to C). Additionaly, we have improved the analyse and the interpretation of the results of the downscaling method (§ 5.1). We have also explained in more details the propogation of errors in the hydrological SWAT model by comparing the different rainfall-runoff simulations (§ 5.2).

For example, the daily standard deviation (std) of the downscaled precipitation is twice as high as the std of the observed precipitation. This is probably a side-effect of the proposed method; the six meteorological stations are all dry or all wet, the highest rainfall events occur by definition on the same day for all six stations, etc. So, the spatial disaggregation is too common and only works for average conditions. To incorporate the varying spatial pattern, dynamical downscaling by a RCM could be considered. But, as RCM simulations seems not to be generally available and are too time-consuming to be performed in this study a stochastic approach, as multi-site weather generators (Fowler et al., 2005; Wilks, 1998) or conditional resampling (Brandsma and Buishand,1998; Wilby et al, 2003) could be considered.
Indeed, other statistical downscaling methods (SDMs) could give different results. This has been highlighted at the end of the paragraph 6. However, it is not the purpose of this article to compare neither the sensitivity of the climate simulations to the employed SDMs, nor the sensitivity of the SWAT model to those same SDMs. In this study, one of the main objectives is to see wether a given statistical downscaling model is able to improve the large-scale reanalysis and GCM data when "feeding" the SWAT model. Within that context, we decided to work with the CDF-t approach (Michelangeli et al., 2009). Note that, following the reviewers' comments, a comparison is now made between the SWAT results driven by the downscaled climate data, and the SWAT results driven by the "raw" data from NCEP reanalyses or LMDz outputs.
The choice of CDF-t as the SDM has been made for a particular reason: Usually, when working with GCM data, "classical" SDMs need, first, to calibrate a statistical model linking observations (e.g., a stations) and large-scale reanalysis data. Then, GCM data for the same calibration period are used as inputs in the statistical model to assess if the latter can be satisfactorily driven by GCM data instead of reanalyses. As a third step, GCM data over an evaluation period are inserted in the SDM to check the ability of the couple GCM/SDM to be applicable out of the calibration period. Those different steps that are needed to apply properly a SDM to GCM outputs are bypassed by the CDF-t approach. Indeed, as CDF-t provides and works on cumulative distribution functions (CDFs), the calibration step can be directly performed between observations and GCM simulations. This is clear advantage that we make profit of in the calibration/evaluation procedure.

Also the discrepancy between observed and downscaled temperature is quite large (0.5 – 1.0 degrees Celsius).
We have re-calculated the standard deviation of maximum and minimum temperatures and precipitation after removing the mean seasonal cycle. The results are presented in appendixes A, B, and C for the PDM calibration and validation. The discrepancies between mean observed and downscaled precipitation and temperatures are within the standard deviation range which allows us to consider these differences as non significant.

Again, interpret this bias and speculate about how this will affect the derivation of evapotranspiration and the hydrological modelling.
We have previously evidenced that the hydrological SWAT model is more sensitive to a change in precipitation than in temperature (Troin et al., 2011). Additionaly, Legesse et al. (2010) have evidenced that a 1.5°C increase in temperature should induce a 13% decrease in surface runoff and a 6% increase in potential evapotranspiration for similar climatic conditions. This study can give an idea of the amplitude of the climatic variation on a catchment.

The hydrological modelling is said to give "satisfying" results. Despite the monthly time step, the NSE is less than 50% for the calibration period, which is poor. This is likely caused by the fact that the watershed is not homogeneous at all and therefore not suitable for such evaluations.
The results in term of NSE are perhaps relatively "low" and this fact is mainly attributed to the long processing sequence used between large-scale meteorological data, the downscaling method, and the hydrological model, and thus the associated uncertainties. Additionaly, the suitability of the Sali-Dulce Basin has been tested during the SWAT calibration. We have obtained a NSE of 0.81 (Troin et al., 2011). This result allowed us to confirm the relevance to use this hydrological model for the Sali-Dulce Basin.

It seems that the construction of the Rio Hondo reservoir has greatly affected the local hydrology, at least for the first 5-10 years after construction. This should be accounted for in the hydrological modelling. If that is not possible the hydrological model output, based on the station observations could be used as a reference alternatively.
The role of the Rio Hondo reservoir on local hydrology has been previously tested during the calibration of SWAT by comparing two simulations performed with and without the reservoir (Troin et al., 2011). We have determined a weak influence of the Rio Hondo reservoir on streamflows after construction and the reservoir reduced peak flows and sustained baseflows by, on average 40 and 35%, respectively.

ABSTRACT, INTRODUCTION AND GENERAL DESCRIPTION OF THE STUDY AREA #
At page 9527 it is stated that the present GCM's are too coarse for hydrological modelling (line 4 to 8). References of 5-10 year old papers are not sufficient too state something about the presently typical GCM resolution. Is the coarse spatial resolution really a problem in this particular study where discharges and lake levels are evaluated at a monthly time step? Explain.
We have added for demonstration the performance of the discharge and lake level simulation when using the global-scale variables before downscaling. We show indeed that the problems associated to the GCM are due to both its coarse spatial resolution and the weak representation of interannual variability in precipitation and temperatures. We show that the applied downscaling method helps significantly improving the results and discuss that point in the text (§ 5.2 and 5.3).

Also the cited article of Wood et al. (2004) is somewhat dated. In the mean time lots of papers (.e.g. Leander and Buishand, 2007; Terink et al, 2010; Piani et al., 2010) have been published about bias correcting RCM output. Some of these bias corrections are by the way very similar to statistical downscaling methods. Yet, I can imagine that not many RCM simulations are available for South America. This should be more thoroughly explained.
Indeed, very few RCM simulations are available for South America and most of them cover only the 1989-2008 time period (Sanchez et al., 2010; Samuelsson et al., in prep; Solman et al., in prep.). Additionally, the comparison of the performance between statistical and dynamical downscaling methods is an interesting approach but this not the main objectif of this paper.

DATA
It is not clear how missing values are treated (section 3.1)
The hydrological SWAT model includes a weather generator (Sharpley and Williams, 1990) in order to fill in missing meteorological data. We have clarified this point in the description of SWAT (§ 4.2; lines 280-284).

The description of the NCEP/NCAR reanalysis data needs some reference (section 3.2)
We have added some reference in paragraph 3.2.

Some figure of the regions and boxes with respect to the research area is needed.
We have added the regions of NCEP/NCAR reanalyses and the boxes of LMDZ on figure 1.

Region A is much smaller than the grid cell size of the reanalysis data. How are the regional values determined? - Explicitly state that region B and box C match. Comparison of the reanalysis and GCM data is only possible if exactly the same area is used.
Geographical latitudes and longitudes were not the only criteria to guide our choice since the LMDz model has biases. Our region selection was also guided by the climatologic structure which is somewhat « contracted » in the latitude as compared to the observed climatology. For example, the latitudinal extension of the Hadley Cell is a little bit smaller than in the reanalyzes, so that all the structure have a small equator ward bias. In order to obtain comparable quantities, we choose regions in LMDz that are the most physically and climatologically consistent with the observed climatologic structure.

METHODOLOGY
At page 9533, lines 18-19 it is stated that the CDF-t method can be seen as an extension of the quantile matching method. It is not entirely clear what is the extension?
The quantile-matching approach looks for the observed quantile (i.e., it is an observed value) that has the same CDF value (in the observations "world") as the CDF value (in the GCM "world") of the large-scale data to be downscaled. In other words, the quantile-matching (QM) approach assumes that, in climate change context, the future large-scale CDF is the same as the large-scale CDF of the calibration period. As explained in Michelangeli et al. (2009), although QM and CDF-t have the same philosophy of working on CDFs, the CDF-t approach does not make this "stationnary" assumption and will derive local-scale CDFs (i.e., at the stations) based on the evolutions of the large-scale CDFs between calibration and taget (i.e., future or evaluation) period. In that sense, CDF-t can be seen as an extension of QM. We have clarified this point in the methodology (§ 4.1).

In line 20 it is stated that CDF-t assumes that a mathematical transform can be used to match the CDF of the predictors with the CDF of the predictands. What transformation is used and what do you exactly mean? ïA˛o˝ Is the same transformation (scaling, shifting, power law, etc) used for all quantiles? If so, it is no quantile matching. ïA˛o˝ Is for every quantile separately a shift or scaling factor derived?
The transformation T allowing to go from the present large-scale CDF G_p to the present local-scale one F_p is fully described in Michelangeli et al. (2009). Mathematically, this transformation can be written as verifying $T(G_p(x))=F_p(x)$ for all x. By replacing x by $G_p^{-1}(u)$, where u belongs to $[0,1]$, T can be formulated as $T(u)=Fp(G_p^{-1}(u))$. Then, assuming that this transformation remains valid in the future, i.e., with the future CDFs G_f and F_f, the researched CDF F_f is provided by $F_f(x) = T(G_f(x))$, which is equivalent to $F_f(x) = F_p(G_p^{-1}(G_f(x)))$.

In section 4.2 the water balance is very well explained, but the routing phase gets very little attention.
We have explained in more details the routing phase in the hydrological SWAT model (Lines 285-290).

Motivate why daily input is that important for the evaluation of monthly discharge and lake level? Does the CDF-t downscaling really add accuracy? It would be interesting to directly use the GCM and reanalysis output in the hydrological model and compare it to the downscaled runs.
As suggested by the reviewer #1, we have used directly the LMDZ and NCEP/NCAR reanalysis outputs in the hydrological model and compared the results with the downscaled runs. The results are presented in paragraphs 5.2 and 5.3.

Explain how the hydrological model is calibrated. Only on the monthly discharges?
The hydrological SWAT model was calibrated during the 20-years period (1973-1994) that included 10 wet and 10 dry years based on monthly discharges. The discharge measurements were available only at the monthly time step. More details of SWAT calibration can be found in Troin et al. 2011.

The wrong subscripts are used in the explanation of equation 2.
We have corrected the explanation of equation 2.

The symbol gamma in equation 3 is not explained.
We have defined the symbol gamma in equation 3.

TABLES AND FIGURES
For the evaluation of the calibration and validation of CDFt two different types of metrics are used (table 3, 4, 5 and 6). Both types should be presented.
As recommended by the reviewer #1, we have homogenized the evaluation of the calibration and validation of CDF-t in Tables 3, 4, 5, and 6. In the new version of the manuscript, we have presented the proportion of precipitation occurence for dry and wet days during the calibration and validation.

Figures 3 and 4 are not referred to in the text.
We have included in the text the references to the figures 3 and 4.

Figures 5 and 6 are too small for proper reading.
We have increased the size of figures 5 and 6.