Hydrol. Earth Syst. Sci. Discuss., 9, C5760–C5770, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C5760/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Comparing dynamical, stochastic and combined downscaling approaches – lessons from a case study in the Mediterranean region" *by* N. Guyennon et al.

N. Guyennon et al.

ivan.portoghese@ba.irsa.cnr.it

Received and published: 11 December 2012

Response to Referee #2 comments.

With regard to the manuscript submitted for publication on Hydrology and Earth System Sciences Discussions, entitled "Comparing dynamical, stochastic and combined downscaling approaches – lessons from a case study in the Mediterranean region", we have revised it following point by point both the reviewer's suggestions that allowed to significantly improve its quality and readability.

Changes and modifications are described in the following pages providing a detailed

description of the modifications introduced in the same order proposed in the review. Each specific comment (RC) is reported and followed by the corresponding reply and appropriate manuscript integration (AC).

Sincerely, Nicolas Guyennon and co-authors.

RC: This paper compares three different methods to downscale GCM data over the southern part of Italia. The first method is a dynamical downscaling method based on a regional model. The second method is based on statistical downscaling method and the last one is a combination of the dynamical and the statistical methods. This study sheds light on the comparison between different downscaling techniques but I am not convince by the interest since there is no comparison with other new methods that can be better than the methods used here. Also, I do not understand how the authors compare the dynamical method, which is independent to the observations, to the statistical method that is corrected by the observations. I consider that the following comments addressed need to be clarify before publishing this paper. Major comments: 1-I am not convinced whether it is valuable to compare these three methods. In fact, there are so many different methods to downscale that one can be easily overwhelmed by the sheer range of possibilities. It would be nice to provide some contextual information indicating where do your methods fit within this large number of existing methods. Related with the point raised above, the authors must shown objectively without ambiguities the interest of this study. There is no information in the current version of the manuscript that allows a fair and objective comparison of this approach with other techniques. A comparison with other downscaling approaches such as stochastic weather generation or the CDF-transform method, that are freely available in the net, should be done.

AC: The reviewer arises here two different issues. a) Concerning the lack of contextual information about the different downscaling methods, the authors reformulated the introduction in order to express without ambiguities the interest of the study: the main goal is to further investigate the advantages and limitations of using a DD as a pre-processing before SD to obtain monthly scenarios of temperature and precipitation.

b) Concerning the use of a single statistical downscaling, the authors agree that choosing one among a large number of existing model is a limit of the study, but they consider this limit similar to the choice of a unique GCM, a unique DD and a unique case study. This is an important limitation and for this reason a large part of the discussion is dedicated to this issue. The referee proposes to add other statistical downscaling models. While comparing a regression model with weather typing schemes or weather generators (Fowler et al., 2007) would be of particular interest at shorter time scale, the authors think that it would not be useful for the overall objective of this study (although adopting other regression models may lead to more robust results), because of the monthly time scale and the long term indicator of performance adopted. Further investigations on the advantages and limitations of using a DD as a pre-processing before more SD (as WG) at daily time scale require additional indicator of performances (e.g. to monitor extreme events) and can be the focus of a future work. As an answer to this issue, we tried to better explain within the manuscript that the overall goal of this study is to further investigate the advantages and limitations of using a DD as a pre-processing before SD to obtain monthly scenarios of temperature and precipitation. Following this line, the authors have revised the manuscript: -Reformulating the title, part of the abstract and of the introduction to better focus the overall objective; -Moving the discussion of the SD choice from the Discussion section to Data and Methods section to better individuate this limitation before presenting the results.

RC: 2-The authors apply the quantile matching method directly. It should be done in two steps : a learning period to calibrate then a projection or validation period. Moreover, the quantile matching method used here is not really a downscaling method but more a correction method. So the values from DD are corrected by the observations for the same period, whereas the DD method is independent from the observations. So this seems difficult to compare these two products.

AC: The reviewer arises here three different issues. a) Concerning splitting the dataset into calibration and validation period for the data processing involving the SD: we pro-

pose to present the results of a k-fold cross validation rather than dividing an already limited data set into calibration and validation period. In fact: -Reducing calibration period will reduce the data samples available for the monthly quantile mapping that already result statistically limited (47), as the method is applied with a monthly aggregation. -Reducing validation period will reduce the period over which the non stationarity and the trend analysis are applied, that will result meaningless. Our basic assumption is that the quantile mapping performance is not significantly affected by the overlapping of calibration/validation periods. To asses this hypothesis we present the results of a k-fold cross validation (k=4) (Kohavi 1995) over 4 sub-sets randomly selected from the overall dataset of each land control station. For each step (k), 3 sub-sets are used for calibration, while validation is performed on the remaining sub-set. The root mean squared error (RMSE) between results of the SD and the reference is used as indicator of the quantile mapping performances and applied to the four validation sets (resulting from k-fold permutation). This is done for each land control station. We also consider the RMSE resulting from the overall dataset as in the first submitted manuscript thus obtaining five groups of RMSE. An ANOVA is then performed for the resulting five groups of RMSE, in order to test the null hypothesis of no significant difference between the spatial distribution of the RMSE resulting from each k-fold permutation and the complete dataset. The p-values are reported in the next table for precipitation, minimum and maximum temperature, for both data processing (3) and (4). The null hypothesis is rejected when the p-value is less than the significance level (usual significance levels are 0.05 or 0.01). None of these tests indicate significant differences of the SD performances among the calibration/validation periods selected. Precipitation Minimum Temperature Maximum Temperature GCM-SD (3) / (ref) 0.85 0.69 0.88 GCM-DD-SD (4) / (ref) 0.65 0.54 0.89

Thus we propose to insert a short comment in the reviewed manuscript as a response to the issue risen by the referee together with the conclusions from the ANOVA test.

b) Concerning the fact that the quantile mapping is not really a downscaling method but more a correction method: Giorgi and Hewitson (2001) use the term ‘transfer function’ to describe methods that directly quantify a relationship between predictand and predictor. In that context, the quantile mapping can be classified as a regression model within the statistical downscaling family (Fowler et al. 2007) as recently done in Hayhoe (2010). We propose to keep the current terminology and to reformulate both the introduction to better contextualize the application of quantile mapping as SD at monthly time scale, and the section data and methods, in order to better explain why we considered the quantile mapping a statistical downscaling when applied at monthly scale.

c) Concerning the direct comparison between the data processing independent from the observations ((1) and (2)) and data processing involving the SD ((3) and (4)), we agree with the reviewer on the fact that any indicator based on time correlation is biased as the GCM and DD are not initialized. For this reason a particular attention was paid to the indicators of performances used to compare these products: -The mean bias analysis involves the spatial distribution of the mean values, and does not depend on the time correlation. -The non stationarity analysis is based on quantile distribution over decades, and thus is related to the time scale of changes in the radiative balance of the earth system which is more driven by the observed concentration of CO_2 than by the initial state. -The trend analysis based on the Sen’s slope spatial distribution over the whole period does not imply time correlation between data processing. Moreover, the authors wish to highlight that the scope of the study is to further investigate the advantages and limitations of using a DD as a pre-processing before SD rather than compare the absolute performances of each data processing independently.

RC: Minor comments: Abstract p9848, l27, these sentences should be in the introduction section

AC: There should be some confusion in the line number as the current abstract ends at line 22. Some changes have been implemented in the abstract. We hope they will answer to this comment.

RC: P9850, l7: “In this context...”. The originality of this study is not clear in relation to previous studies. Is it only the study area ?

AC: The main goal of the present study is to further investigate the pros from using a DD as a pre-processing before SD, in particular in terms of non stationarity. The authors tried to better present the context by rewriting part of the abstract and introduction. We hope that the originality of this study is clear now.

RC: p9852, l10. You mean 6h accumulated precipitation and temperature every 6h ?

AC: Exactly. The sentence has been reformulated.

RC: p9852, l24. Following the constraints associated with the method, a discussion on the ability to study extreme events using the Quantile matching method is necessary here.

AC: This section aims to shortly present the adopted methods. Following the answer of the first major comment, the selection of a statistical downscaling to match particular features suitable for impact studies is not the point of interest of the present work. In particular, in the present case study, the monthly time scale is unsuitable for extreme events analysis up the selection of the statistical downscaling.

RC: P9854, l12 “The GCM ...”. Please clarify, do you mean that it is reanalysis?

AC: The GCM (Global Circulation Model) is presented in section 2.1.1 and does not refer to a reanalysis.

RC: P9858, l3: The temperature evolution of the DD (2) seems to be strongly correlated to the temperature evolution provided by the GCM (1). The effect of the downscaling seems to be only in term of bias. Can you comment this result. Do you consider the regional domain large enough to evolve independently of the large scale forcing ?

AC: At the annual scale, the temperature evolution of the DD is strongly correlated to the temperature evolution of the GCM. In fact, the model domain adopted in this study is comparable to the scale of baroclinic synoptic systems affecting climate variability in this area. Therefore, over time scales larger than the characteristic baroclinic scale, the system is expected to preserve the properties of the large scale driver (Déqué et al., 2007).

RC: P9858, l12: For the section 3.1. Do you apply the quantile matching method for the entire period, before performing these scores? In that case, it is not correct to compare a method corrected from the observation to the DD without correction. To make this kind of comparison, the authors should distinguish two periods for the SD method. The first period is the learning period and the second one, the validation period (without information about the observations). In this section, I consider feasible only the comparison between (1) and (2) or between (3) and (4).

AC: Our basic assumption is that the quantile mapping performance is not significantly affected by the overlapping of calibration/validation periods. To asses this hypothesis, we performed a k-fold cross validation (k=4) over four sub-sets randomly selected from the overall dataset of each land control station (for the details please refer to the previous authors answer on this issue). None of the ANOVA tests over the k-fold cross validation indicate significant differences of the SD performances among the calibration/validation periods selected. For this reason the authors consider feasible to compare data processing with and without the SD step, to highlight the benefits of SD in term of mean bias, even without distinguishing between learning and validation period.

RC: P9860, The results for the 4 methods are strongly correlated. I understand for the SD but I am surprised for the DD. Moreover, from these results, it seems that the tendency over this area is only explained by the large scale forcing and is not offset by the regional modelling. Can you comment these two points?

AC: The average tendency over a given area is dominated by the large scale (see answer to comment P9858, l3). Where the regional downscaling can add information is in improving the description of spatial patterns. We cited results already reported in

the literature (Déqué, 2007).

RC: P9863: l 3 How to interpret the tendency observed in Fig. 6 for precipitation in the GCM and the Sen's slope closed to 0 in Fig. 8 second line ?

AC: In Fig. 8 l 2, the GCM presents a homogeneous dark green Sen's slope color code, indicating value lying between 0 and -1.0 mm/yr. Actual Sen's slope is about -0.6 mm/yr, in agreement with the tendency observed in Fig. 6. The confusion may have resulted from the pseudo color code resolution, which has been optimized to highlight the impact of downscaling processing's rather than absolute values. We choose to use a linear scale of the color coding rather than any other classification method as we think it is the clearest way to present these results.

RC: P9863: l17 How do you explain the impact of the DD in precipitation without significant modulation regarding temperature?

AC: The different behavior of the DD temperature and rainfall in terms of their spatial heterogeneity is linked to two important aspects of the dynamical downscaling process. First, in DD the local topography can generate rainfall patterns even in the absence of significant temperature inhomogeneities at the surface. Second, local feedbacks affecting rainfall patterns and involving convective instability depend on soil moisture and therefore on the land-use characteristics and land-sea contrast, which can amplify irregular rainfall patterns even in case of weak temperature gradients (SchÓŞr et al. 1999).

RC: P9864: The third paragraph at the beginning of the discussion section should be moved to the introduction section. This discussion section is too long given its interest. The authors should be more concise.

AC: The third paragraph at the beginning of the discussion has been moved to the introduction. Discussion has been rewritten to be more concise.

RC: Figures: Fig 1: The indications on the spatial scale are oversimplified. The SD is based on local observations, how do you estimate the associated spatial scale ?

AC: We do agree on the fact that the SD spatial scale depends on the local observation density, and may thus vary over a large range of value depending on the case study. Similarly, scale associated with GCM and DD may vary significantly depending on the adopted models. The spatial scale reported in figure 1 are the typical global, regional and local scale found in literature: - Concerning the GCM, model resolution varies from a typical coarse horizontal grid of $3.8^\circ \times 3.8^\circ$ (T31) to a fine grid of $0.75^\circ \times 0.75^\circ$ (T159) ((Werner, Langebroek et al. 2011), meaning typical scale roughly ranging from 5*105 to 104 km2. We report 105 km2, which is also representative of the implemented GCM. - Concerning the DD, Fowler el al. (2007) state that Regional climate model are typically resolved at the 0.5° latitude and longitude, meaning a typically scale of 2*103 km2. We report 103 km2, which is also representative of the implemented DD. - Concerning the SD, we refer to the minimum densities recommended of precipitation stations by the WMO (1994): 1 station for 2.5*102 km2 for the mountainous area, 1 for 9*102 km2 for the coastal area and 1 for 10 km2 for urban areas. We report 102 km2, which is also representative of the implemented land control network.

RC: Fig 6 and Fig 7: It is difficult to distinguish the lines, please use different colours.

AC: We add colors to figures 6 and 7 to improve the readability

References: Déqué M., D. P. Rowell, D. Lüthi, F. Giorgi, J. H. Christensen, B. Rockel, D. Jacob, E. Kjellström, M. de Castro, B. van den Hurk: An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections, Climatic Change, 81, 53-70, 2007.

Fowler, H. J., Blenkinsop, S. and Tebaldi C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology 27(12): 1547-1578, 2007.

Giorgi F., Hewitson BC.: Regional climate information – evaluation and projections.

In Climate Change 2001: The Scientific Basis. C , Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dia X, Maskell K, Johnson CA (eds). Cambridge University Press: Cambridge, 2001

Hayhoe K. A.: A standardized framework for evaluating the skill of regional climate downscaling techniques. Thesis (Ph.D.), University of Illinois at Urbana-Champaign, 2010.; Publication Number: AAT 3430975; ISBN: 9781124315010; Source: Dissertation Abstracts International, Volume: 71-12, Section: B, 158 p, 2010.

Ines, A. V. M. and Hansen, J. W.: Bias correction of daily GCM rainfall for crop simulation studies. Agricultural and Forest Meteorology 138(1–4): 44-53, 2006. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection, in Proc. 14th Int. Joint Conf. Artificial Intelligence, pp. 338–345, 1995.

Schär, C., Lüthi, D., Beyerle, U. and Heise, E.: The Soil–Precipitation Feedback: A Process Study with a Regional Climate Model. J. Climate, 12, 722–741, 1999.

Werner, M., Langebroek, P. M., Carlsen, T., Herold, M. and Lohmann, G.: Stable water isotopes in the ECHAM5 general circulation model: Toward high-resolution isotope modeling on a global scale. J. Geophys. Res. 116(D15): D15109, 2011.

World Meteorological Organization.: Guide to hydrological practices: data acquisition and processing, analysis, forecasting and other applications. 5th edition. WMO-No 168, Geneva, 1994.

Kohavi, R.: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection.in C. S. Mellish, ed., Proceedings of the 14th International Joint Conference on Artificial Intelligence, Morgan Kaufmann Publishers, Inc., pp. 1137-1143, 1995.

Werner, M., Langebroek, P. M., Carlsen, T., Herold, M. and Lohmann G.: Stable water isotopes in the ECHAM5 general circulation model: Toward high-resolution isotope modeling on a global scale. J. Geophys. Res. 116(D15): D15109, 2011.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 9847, 2012.