Hydrol. Earth Syst. Sci. Discuss., 9, C4703–C4711, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C4703/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Regional climate models downscaling in the Alpine area with Multimodel SuperEnsemble" *by* D. Cane et al.

D. Cane et al.

daniele.cane@arpa.piemonte.it

Received and published: 12 October 2012

First of all, let us thank the anonymous reviewer for the very accurate review: your detailed criticism will help us to improve the comprehensibility of our paper because it made us aware that some parts of the presentation of our technique were too brief and must be specified better. We also agree that we must include an evaluation of multi-model uncertainty and we thank the anonymous reviewer for signalling this lack: we can do it quite easily for both the parameters and include it in a revised version of this paper.

The main objective of this work is to obtain a reliable projection, trying to include information from many different models, which differ quite a lot in the Alpine area and do not

permit to drive conclusions about the climate projections in the XXI century (of course with the limitation of the use of RCMs on a single scenario, but we think that due to our temporal scope limited to 2050, this is quite reasonable). When we started this work we kept in mind the recommendations of the IPCC group on multi-model ensembles (Knutti et al, 2010), which includes the criticism by Weigel et al. (2008) about the risks of model weighting in Multimodel Climate Projections. "The performance metric is most powerful if it is relatively simple but statistically robust, if the results are not strongly dependent on the detailed specifications of the metric and other choices external to the model (e.g., the forcing) and if the results can be understood in terms of known processes (e.g., Frame et al., 2006)." (Knutti et al., 2010) We know that the use of a given metric implies limitations and we think we honestly stated them in our paper, but we try to state them again.

TEMPERATURES

Yes, the temperatures are a linear combination of the un-biased model values, with weights calculated with a Gauss-Jordan minimization technique in the training period. Both model statistics (biases) and weights are calculated from the corresponding time-series: ERA40-driven RCMs have specific weights calculated from the confrontation between their daily values in the training period and observations, GCM-driven RCMs have different weights calculated from their values. Fig. 1 shows an extract of a paper of our group (Cane & Milelli, 2010) with all the (simple) mathematics of the Multimodel SuperEnsemble in our implementation.

We performed a quite extensive evaluation of this technique to understand if we were able to reduce the errors of the projections, of which figures 5 and 6 in the paper are only a part. We evaluated the multimodel with weights calculated on a 20-year period (1961-1980) and applied them to a subsequent 20-year period (1981-2000). The weights and biases are calculated independently for each model and gridpoint, but do not depend on time (i.e. we applied the same bias correction and weights for any day of the year).

We know very well that the value projected by a given model in the "scenario mode" in a given day is not correlated with the value observed in that day, but what we are searching here is something very "gross": for each model we had 7305 (quite big) anomalies vs the observations, partly due to the models and partly to the uncorrelation between GCM scenarios and reality. The weights we are calculating are not trying to disentangle these contributions, but to evaluate which is the best way to average the different contributions to obtain the best approximation of observations (in the average). Please notice that the weights are not normalized and can be negative: they are simply weighting the contribution of each model anomaly to the final corrected value. While applying Multimodel SuperEnsemble, we are doing two hypotheses:

1) for any given point, the (annual averaged) biases of the models in the future scenario will remain the same they were in the past scenario (training period).

2) for any given point, the weights of the models, hence their relative contribution to the final result, will remain the same. From our experience, if you are using fixed weights for a multi-model, it is better to have the larger possible training period, while if you are interested in a better correction you must use short statistics close to the forecast period you are interested in.

Even with these limitations, the multimodel scenario is quite good in reproducing the observed data trends in the control period 1981-2000 (fig 5 in the paper), and very good in reproducing the monthly regime (fig 6 in the paper), eliminating the big problems some of the input model show in this case.

Multimodel SuperEnsemble is then able to reduce the above-cited two big issues of temperature fields in ENSEMBLES RCMs in the Alpine area: such post-processed data can be of interest for the impact application, while the original models are definitely unuseful. To evaluate the projected scenarios, we re-calculated the weights on a 40-year period (1961-2000) and applied them on a 50 year period (2001-2050). Again, the weights are computed on a time window comparable with the "forecast" window.

According to the metric we used, we know very well the limitation of the multimodel data we are producing: they can be used on a aggregated timescale (minimum time resolution: monthly data aggregated over a period of minimum 20 years), they cannot be used for assessment of temperature extremes but only of temperature averages, but they include a signal which is coming from the observations and is definitely hard to find in the input models: we can see quite well the alpine chain (even at the coarse resolution of 14 km) and its behaviour is different (more enhanced) from that of the plains like it really happened in the observed data, while in the input models there is a very light difference. We are here writing a rapid answer to your comment, but we intend to evaluate the uncertainty of the Multimodel SuperEnsemble via a combination of the input model uncertainties.

PRECIPITATION

First of all, let us repeat that the RCMs direct model output precipitation as it is is definitely not useful, because 1) it shows biases up to 100% (in winter) in the Alps and 2) the different models show a quite different behaviour in the scenario. If we use any kind of single-model correction, as for the quantile mapping or similar techniques, we will still have the big question of having conflicting answers from different corrected models.

The metric we applied in this case is quite different: we tried to correct the Probability Density Function of the observed precipitation conditioned to the RCMs forecasts, hence having a precipitation-depending correction.

In other words, we try to answer to the question: when the given model is forecasting 2 mm, how much precipitation was really measured? This can be done in an average, but we think that a PDF evaluation is more correct. We followed this algorithm:

1) we took the ERA40-driven RCMs and confronted them with the observed precipitation on all gridpoints

2) we considered all the days and points where the model forecast a given precipitation (in our example: 2 mm/day)

3) we built the distribution of the observed values of that days/points

4) we repeated for any reasonable forecast value (up to 300 mm/day)

5) we fitted the distributions with a set of functions, finding that the Weibull distribution is the best to represent all of them (not surprisingly)

6) we interpolated and extrapolated the observed distributions to obtain all possible distribution for all forecast values

7) we calculated the individual RCMs Continuous ranked Probability Scores (CRPS) from ERA40-driven models

8) we applied the model-specific PDFs and (inverse, normalized) CRPSs to the GCM-driven models to obtain weighted PDFs

Please notice that in this case (as we stated honestly in the paper) the model metrics is evaluated on ERA40-driven models and not on GCM-driven models: this is necessary, because we cannot loose the correlation between forecast and observed precipitation, but at the same time what these PDFs say is how much we must correct any single model to obtain more reasonable precipitation, while CRPSs say how much the single model is able to reproduce the statistical distribution of the data (part of the CRPS error arises from possible errors in ERA40 fields, but is common to all the models).

In this case we think we can apply this two metrics evaluated on reanalyses to scenarios, because we think that the under/overestimation of precipitation (described by PDF) will be the same for a given atmospheric pattern, both evaluated by ERA40 and by the GCM, while the CRPS is the ability of a given model to reproduce distributions and we think that we can use it to asses which model has to be given the higher importance. We think that this is quite reasonable, what is quite difficult to evaluate is the contribution of the different GCMs to the potential CRPS, but we consider here each

model as a GCM-RCM system.

Please notice that, while the weights of the standard Multimodel SuperEnsemble techniques require the contemporary evaluation of any given model weight, in this case the CRPSs are evaluated independently, and only at the end we calculate the inverse and normalize these inverses to obtain the weighting.

We evaluated again the precipitation multi-model on our control period (1981-200) after having evaluated all the metrics in the 1961-1980 training period. We were quite surprised about the results: again, we reduced the (quite enormous) biases and we reproduced quite reasonably the monthly precipitation regime. Fig. 2 shows the un-biased Walter and Lieth diagrams for Piemonte. It corresponds to fig. 7 of the paper, we did not included such a figure because it is unfair to compare un-biased models vs a bias-corrected post-processed output, but you can understand that we cannot propose the larger part of these models for any use without some kind of correction. Even without any further bias-correction, Multimodel SuperEnsemble data are quite reasonable when compared with observations.

Again, we here state the limitations of our technique: the data must be evaluated on a monthly basis, with aggregation on a long period, they cannot be used for extreme precipitation events because of the averaging effect (although reduced by the use of the PDFs), the consecutive wet days number is not conserved but the consecutive dry days number is, then these data can be reasonably used for integrating indices like fire weather indices. The multi-model precipitation provide almost bias-free data, then they can be used for long-term evaluation of droughts.

Again, we will include an analysis of uncertainty based on our PDFs: we can easily extract as many realizations we want from those PDFs and evaluated the Multimodel SuperEnsemble Dressing uncertainty.

REFERENCES

Cane, D. and Milelli, M.: Can a Multimodel SuperEnsemble technique be used for precipitation forecasts?, Adv. Geosci., 25, 17–22, doi:10.5194/adgeo-25-17-2010, 2010
Knutti, R., Abramowitz, G., Collins, M., Eyring, V., Gleckler, P. J., Hewitson, B., and Mearns, L.: Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections,in: Meeting Report of the Intergovernmental Panel on Climate Change Expert Meetingon Assessing and Combining Multi Model Climate Projections, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., and Midgley, P. M., IPCC Working Group I Technical SupportUnit, University of Bern, Bern, Switzerland, 2010.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 9425, 2012.

$$S=\overline{O}+\sum_{i=1}^{N}a_i\left(F_i-\overline{F}_i\right) \quad (3)$$

where N is the number of models, a_i are the SuperEnsemble weights, F_i is the forecast value, $\overline{F}_i$ is the mean forecast value in the training period and $\overline{O}$ is the mean observation in the training period.

The calculation of the parameters a_i is given by the minimisation of the mean square deviation

$$G=\sum_{k=1}^{T}(S_k-O_k)^2 \quad (4)$$

where T is the training period length (days). By derivation we obtain a set of N equations:

$$\begin{pmatrix} \sum_{k=1}^{T}(F_{1_k}-\overline{F}_1)^2 & \sum_{k=1}^{T}(F_{1_k}-\overline{F}_1)(F_{2_k}-\overline{F}_2) & \cdots & \sum_{k=1}^{T}(F_{1_k}-\overline{F}_1)(F_{N_k}-\overline{F}_N) \\ \sum_{k=1}^{T}(F_{2_k}-\overline{F}_2)(F_{1_k}-\overline{F}_1) & \sum_{k=1}^{T}(F_{2_k}-\overline{F}_2)^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \sum_{k=1}^{T}(F_{N_k}-\overline{F}_N)(F_{1_k}-\overline{F}_1) & \cdots & \cdots & \sum_{k=1}^{T}(F_{N_k}-\overline{F}_N)^2 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ \vdots \\ \vdots \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{T}(F_{1_k}-\overline{F}_1)(O_k-\overline{O}) \\ \vdots \\ \vdots \\ \sum_{k=1}^{T}(F_{N_k}-\overline{F}_N)(O_k-\overline{O}) \end{pmatrix} \quad (5)$$

We then solve these equations using the Gauss-Jordan method.

Fig. 1. Equations of the Multimodel SuperEnsemble (from Cane & Milelli, 2010)

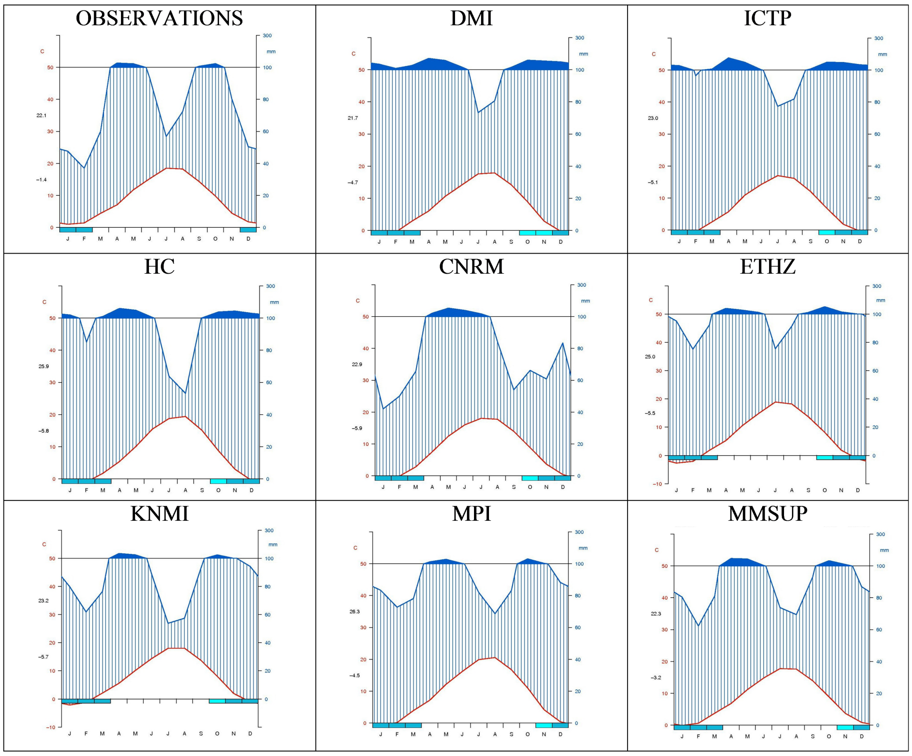

Fig. 2. Walter and Lieth diagrams of the models and Multimodel (MMSUP) for the values averaged over Piedmont OI gridpoints, period 1981–2000.