Review by Ezio Todini of

Regional Climate Models Downscaling in the Alpine Area with Multimodel Super Ensemble by D. Cane, S. Barbarino, L. A. Renier, C. Ronchi

The manuscript discusses the use of two multi-model post-processing techniques. The first one applied to temperature refers to the technique introduced by Krishnamurti et al. in two papers dated 1999 and 2000. The second one, used to merge precipitation conditional density functions.

The manuscript is interesting and well written, but I do not think it is acceptable in its present form.

The main problem lies in the fact that the authors seem not aware of important literature in the field of multi-model uncertainty post-processors which has led to the Bayesian Model Averaging (BMA) introduced by Raftery et al., (2003; 2005) and the Model Conditional Processor (MCP) developed by Todini, (2008), to quote two of the most recent ones.

A multi-model approach obviously aims at obtaining a better estimate of the real unknown value by eliminating the bias and most of all by reducing predictive uncertainty, which in the case of the proposed approach, for the temperatures is obtained by minimizing the variance of the observed minus predicted values. In practice this corresponds to a multi-regression approach, which implicitly assumes an underlying Normal heteroscedastic conditional probability density. This underlying hypothesis is neither formally acknowledged nor proven by the authors. A way of overcoming this problem is for instance given by MCP, in which the development of a multi-dimensional regression in order to derive the conditional distribution of the predictand (the actual temperature in this case) given the predictors (the model forecasts), is preceded by the conversion of predictand and predictors into a Normal space via the Normal Quantile Transform (which in practice corresponds to a Normal Copula approach) (Van der Waerden B.L., 1952; 1953a; 1953b; Krzysztofowicz, 1997). In this space it is possible to identify a multi-Normal joint probability density from which one can analytically derive the conditional probability density (Todini, 2008). This again corresponds to a multi-regression type of approach, which is justified by the fact that the joint density is now multi-Normal. Although in practice the estimation of the weights is very close to the one proposed, the transformation into the multi-Normal space has the advantage of allowing for the correct estimation of the weights. Moreover, as opposed to proposed approach, which implies symmetry and is limited to the estimation of the first and second order moments, the derivation of the conditional density in the multi-Normal space allows for the estimation the full “predictive uncertainty” probability density function, which can be, and generally is, a-symmetrical. The transformation into the multi-Normal space and the re-projection of results into the real space marginally increases the complexity of the proposed algorithm, but offers instead improved performances.

Alternatively, one can also use BMA, which builds the probability density of the observations conditional to the different models as a Bayesian mixture of densities.

$$g\left(y\middle|\hat{y}^{(1)},\hat{y}^{(2)},\dots,\hat{y}^{(m)}\right)=E\left\{f_1\left(y\ \middle|\hat{y}^{(1)}\right),f_2\left(y_t\middle|\hat{y}^{(2)}\right),\dots,f_m\left(y_t\middle|\hat{y}^{(m)}\right)\right\}$$

Also in this case it is appropriate to convert all the data into the Normal space prior to forming the BMA likelihood function (Vrugt and Robinson, 2007; Todini, 2008), which again implies a multi-Normal probability density function. Results of BMA can be proven to be similar but less performing than results from MCP, because the resulting conditional density is a generic conditional density obtained as the weighted average of the single model conditional densities, which differs form the one descending from the Kolmogorov definition as the joint divided by the marginal density.

$$f\left(y\ \middle|\hat{y}^{(1)},\hat{y}^{(2)},\dots,\hat{y}^{(m)}\right)=\frac{f\left(y\ ,\hat{y}^{(1)},\hat{y}^{(2)},\dots,\hat{y}^{(m)}\right)}{\int_y f\left(y\ ,\hat{y}^{(1)},\hat{y}^{(2)},\dots,\hat{y}^{(m)}\right)dy}=\frac{f\left(y\ ,\hat{y}^{(1)},\hat{y}^{(2)},\dots,\hat{y}^{(m)}\right)}{f\left(\hat{y}^{(1)},\hat{y}^{(2)},\dots,\hat{y}^{(m)}\right)}$$

Where y is the predictand, and $\hat{y}^{(1)},\hat{y}^{(2)},\dots,\hat{y}^{(m)}$ the m models predictors.

One thing which is not clearly specified in the manuscript. In Appendix A of Krishnamurti et al (1999) it is clearly stated that "*The weights a_i are computed at each grid point by minimizing the following function…*". In other words the weights are estimated at each single pixel "independently" from neighboring pixels. Apart from the resulting large number of weights, it seems to me that either one of the models is consistently well performing in a region (the aggregation of a number of pixels or a super-pixel) or it is not. Therefore the weights should be, in my view, computed per regions (which is what BMA does for instance), not per pixel. Otherwise a model could be the best in one pixel and the worst in the adjacent one, which is not totally reasonable. In any case the authors are invited to clarify this point.

As far as the approach used to merge precipitation forecasts, referred to as "Probabilistic Multimodel SuperEnsemble Dressing", this is an alternative way to BMA for building a conditional predictive density, which again does not match the "conditional density" descending from Kolmogorov definition.

There are three points to be addressed by the authors.

The first one is to describe the approach within the frame of alternative used approaches and in particular to BMA, to which the approach is closer.

The second one is to avoid using in their graphs of a log-probability scale, which tends to hide the fact that over 90% of observations and of at least of two models forecasts are zeroes (while it is very strange that two of the used models do not show this characteristics). This is a very important point because one of the major problems for effectively estimating predictive precipitation conditional densities lies in the fact that “misses” and “false alarm” rates can be quite large. Given that the rate of zeroes accounts for over 90% of the cases the effect is enhanced when no precipitation occurs or when the models predict no precipitation.

The third one is that also in the case of precipitation it is not clear if the averaged conditional densities are derived for each pixel. This implies the use of a large number of estimated parameters (number of merged models times number of pixels).

References

- Krishnamurti T. N., Kishtawal C. M., LaRow T.E., Bachiochi D. R, Zhang Z., Williford C.E., Gadgil S., and Surendran S., (1999) Improved weather and seasonal climate forecasts from Multimodel SuperEnsemble, *Science* 285, 1548-1550
- Krishnamurti, T.N., C.M. Kishtawal, T. LaRow, D. Bachiochi, Z. Zhang, C.E. Williford, S. Gadgil, S. Surendran, (2000). Multi-model superensemble forecasts for weather and seasonal climate. *J. Climate* 13, 4196-4216.
- Krzysztofowicz, R., (1997). Transformation and normalization of variates with specified distributions, *J. Hydrol.*, 197, 286–292.
- Raftery, A. E., F. Balabdaoui, T. Gneiting, and M. Polakowski, (2003). *Using Bayesian averaging to calibrate forecast ensembles*, Tech. Rep. 440, Dep. of Stat., Univ. of Wash., Seattle.
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, (2005). Using Bayesian model averaging to calibrate forecast ensembles, *Mon. Weather Rev.*, 133, 1155-1174.
- Todini, E., (2008). A model conditional processor to assess predictive uncertainty in flood forecasting, *Intl. J. River Basin Management*, 6 (2), 123-137.
- Van der Waerden B.L. (1952). Order tests for two-sample problem and their power I. *Indagationes Mathematicae*, 14, 453-458.
- Van der Waerden B.L. (1953a). Order tests for two-sample problem and their power II. *Indagationes Mathematicae*, 15, 303-310.
- Van der Waerden B.L. (1953b). Order tests for two-sample problem and their power III. *Indagationes Mathematicae*, 15, 311-316.
- Vrugt, J.A. and Robinson, B.A. (2007). Treatment of Uncertainty Using Ensemble Methods: Comparison of Sequential Data Assimilation and Bayesian Model Averaging, *Water Resour. Res.,* 43, W01411, doi:10.1029/ 2005WR004838.