Report on the paper
A Bayesian posterior predictive framework for weighting ensemble regional climate models
by Fan, Olson and Evans, submitted to Geoscientific Model Development

Summary of the paper

The paper considers multimodel ensembles with two time slices and proposes weights for models based on the agreement between the model output and the observation during the control period. These weights are then used to average the posterior distributions of the unknown parameters given each model, and this averaged posterior is used for prediction of future observations. The method is then applied to a multimodel ensemble with 12 RCM-GCM chains (3 RCMs driven by 4 GCMs each).

A different presentation of the proposed method

Here is my interpretation of the idea which shows the common points and the differences to other approaches. The observations in control and future period are denoted by $y = (y_0, ..., y_T)$ and $y^f = (y_0^f, \ldots, y_T^f)$, and the output of model m in control and future period by $x^m = (x_0^m, ..., x_T^m)$ and $x^{m,f} = (x_0^{m,f}, \ldots, x_T^{m,f})$. The distribution of these time series depends on parameters θ_p, θ_p^f, θ_m and θ_m^f which are vectors in general. I denote by "data" all the known time series, i.e. data $= (x^{1,f}, \ldots, x^{M,f}, x^1, \ldots, x^M, y)$ and by η all parameters that appear in the likelihood of data, i.e. $\eta = (\theta_1^f, \ldots, \theta_M^f, \theta_1, \ldots, \theta_M, \theta_p)$. The goal is to obtain a predictive distribution

$$p(y^f|\text{data}) = \int p(y^f|\theta_p^f)p(\theta_p^f|\text{data})d\theta_p^f$$

The first factor in the integral is the likelihood of future observations which is given, and the second factor can be decomposed as follows:

$$p(\theta_p^f|\text{data}) = \int p(\theta_p^f|\eta)p(\eta|\text{data})d\eta$$

The second factor here is the posterior of η given the data and follows from Bayes formula in the usual way. The first factor quantifies the prior information about how observations in the future are related to observations in the control and model output in both periods. Such prior information is necessary as the likelihood of the data is independent of θ_p^f. Usually it is derived from assumptions about model biases and bias changes from the control to the future period. The paper here uses implicitly the following joint prior density for θ_p^f and η:

$$\frac{1}{M}\sum_{m=1}^{M}\delta(\theta_p-\theta_m)\delta(\theta_p^f-\theta_m^f)\prod_{k=1}^{M}p(\theta_k)p(\theta_k^f)$$

where δ is the Dirac delta function. This says that a priori the parameters for different models and periods are independent, there is one perfect model for both control and future and each model is equally likely to be the perfect model. It implies that

$$p(\theta_p^f|\text{data})=\sum_{m=1}^{M}w(x^m,y)\int\delta(\theta_p^f-\theta_m^f)p(\theta_m^f|x^{f,m})d\theta_m^f$$

where

$$w(x^m,y)\propto\int p(y|\theta_m)p(\theta_m|x^m)d\theta_m$$

is the posterior probability that m is the perfect model, compare equation (4).

General evaluation

As I have shown above, the method proposed here is not basically different from what is used e.g. in Buser et al. (2009), Climate Dynamics 33, 849-868. The main difference comes from the prior distribution. For me, the assumption that one model is perfect is not natural. I prefer the idea that all models have strengths and weaknesses and therefore deviations are rather on a continuum. But since it is unknown which model is perfect, in the end the analysis still uses all models and thus the results are presumably not that different. The second assumption, namely that the quality of a model can be judged on its behavior during the control alone, is harder to accept. I am not a climate scientist, but a model that agrees well with the observations in the control, but has a much slower or a much faster warming than all the other models seems doubtful to me. On the other hand, a model can be consistently too warm over the whole period from control until the end of the future period, but still give a good estimate of climate change. The authors point out that agreement between models can be due to common model errors. On the other hand, a good agreement between models and observations in the control can also be due to too much tuning, or it can be just a coincidence in case there are many models.

A different criticism concerns the fact that the dependence between different model chains is not taken into account. In my experience, there is non-negligible dependence between RCMs driven by the same GCM and this should be reflected in the likelihood. However, I guess that this would lead to complications.

Detailed comments

- p. 2, equation (1): I would use the parametrization

  $$y_t = a_p + b_p(t - t_1) + \epsilon_t, \text{ where } t_1 = t_0 + \frac{T}{2}.$$

  That is, the slope term is the same, but the intercept is the value in the middle instead of the beginning of the period. Keeping the intercept fixed as in equation (9) on p. 5 makes then much more sense to me.

- p. 3, l. 4: In my experience with multimodel ensembles for Europe (PRUDENCE, ENSEMBLES and CORDEX) it is not true that models vary less than the observations from one year to the next. On the contrary, models often overestimate the variability by a factor up to 2. Also additive corrections of standard errors are strange. Instrumental and gridding errors should be independent of natural variability which would lead to $\sigma_p = \sqrt{\sigma_m^2 + \delta^2}$.

- p.3, l. 21: the weights w^m must be normalized to sum to 1, as stated on p. 4, l. 15.

- Fig. 1: I don't understand what is shown here: The weight w for simulated x and y values (as suggested by the caption), or the likelihood for simulated y values (as suggested by the y-axis). In the latter case, it would be more interesting to show the bivariate likelihood (a function of μ and σ) with contour lines. But isn't the likelihood a well-known concept that doesn't need illustration?

- p. 4, l. 17: To sample from a mixture distribution, you cannot take the weighted average of draws from the mixture components. You have to select first randomly a component and then draw from that component, as in the procedure described at the bottom of p. 6.

- Section 2.2: I miss the information about the chosen prior distributions.

- p. 6, algorithm at the bottom: This can be simplified because the conditional distribution of $(T+1)^{-1}\sum_{t=0}^{T} y_t^f$ given $(a_b^f, b_b^f, \sigma_p^f)$ is normal with mean $a_p^f + b_p\frac{T}{2}$ and standard deviation $\sigma_p^f/sqrtT+1$. Hence one can directly simulate $(T+1)^{-1}\sum_{t=1}^{T} y_t^f$, there is no need to simulate first the y_t^f. One can even use that the conditional distribution of $T^{-1}\sum_{t=1}^{T} y_t^f$ given that m is the perfect model is a Gaussian mixture with means $a_{m,i}^f + b_{m,i}\frac{T}{2}$, standard deviation $\sigma_{m,i}^f/sqrtT$ and equal weights $1/N$. So we can directly compute its density or the quantiles.