Response to interactive comment 1

Red is reviewer comments, Black response.

The authors report on a statistical method loosely connected to bayesian analysis to post-process ensemble simulations of regional climate models. they apply the method and report results for seasonal temperatures in australia over two 20 year periods. The authors claim that their approach is entirely novel, which I do not see but which is a matter of definition of "new".

We have removed the word "entirely" from the conclusions.

More important i found a significant lack of theoretical background, rather the model is set up in a rather adhoc fashion: we need a weighted combination of regional climate models $->$ so why not taking bayesian model averaging with some weights; where to get the weights from $->$ why not taking the likelihood with some "uninformed prior"; how to calculate the weights $->$ why not taking mcmc; where to get the ensemble from $->$ why not taking the regional climate simulations.

The content of the paper is certainly worth to be published but not in this very way as it is presented currently. Papers in GMD should not only report on the technical aspects but also on the theoretical background because this allows to draw conclusions about the assumptions made for the specific implementations.

By theoretical background it appears the reviewer is asking for more justification of the choices made in establishing the framework. First we add "in order to create probabilistic projections" to the end of the first sentence in the abstract to emphasise that the calculation of model weights is not an end in itself. The introduction then discusses past attempts to do this, their limitations, and that our proposed framework overcomes some of these limitations. A sentence is added on line 2 of page 3 to justify the use of uninformed priors "The use of non-informative allows the data to discriminate amongst models, whereas informative prior reflect the scientist's personal knowledge, and can lead to more subjective analyses. Non-Informative priors are sometimes considered preferable when data contains sufficient information". Further justification for the use of MCMC is added as the first sentence of section 2.1: "The procedure for the calculation of weights is designed to be applicable regardless of the distributional forms chosen to model the data." and after the first sentence of section 2 "The framework we describe below is not limited to any particular distributional form, although the analysis presented is based on the Univariate Normal distribution. We have also implemented the same procedure using the asymmetric Laplace distribution to obtain robust estimators for our analyses, but we have to excluded them from presentation as the procedure produced similar results to that of the Normal error assumption (indicating no major violations from Normality)." The choice of regional climate projection ensemble is arbitrary but was made here due to ease of access and familiarity of the authors.

Furthermore even the technical aspect is only mildly covered because at no point except in the very last sentence it is said that the current implementation relies on (univariate! not mentioned!) normal distributed random variables. Additionally the use of a Bayesian approach is only marginally. Firstly, the modelling of uncertainty is rather adhoc (see my remark above) in the sense that the model parameters especially the precision values of the residuals are treated in a non common fashion, standard approaches eg in described in Gilks et al 1996 at least consider the normal-invers gamma model with a wide prior on the hyperparameter of the invers gamma component , secondly the treatment of observations in the likelihood and the treatment of simulations in the prior does not consider the dependency between the residual components.

Text is added after the first sentence of section 2 ”The framework we describe below is not limited to any particular distributional form, although the analysis presented is based on the Univariate Normal distribution. We have also implemented the same procedure using the asymmetric Laplace distribution to obtain robust estimators for our analyses, but we have to excluded them from presentation as the procedure produced similar results to that of the Normal error assumption (indicating no major violations from Normality).”

It is not correct to say that we treat residual variance as non-random. We model the residual variance of model output in the standard way, with a standard prior. We account for the discrepancy between model output and observation with an additional term for the variance, this discrepancy comes from the fact that observations are subject to an additional source of measurement error.

We agree that if the errors are dependent, then one should use an appropriate model to account for correlated errors.

This holds for the actual analysis where the epsilons could be modeled as an AR process in time giving rise to a multivariate normal for the x_m which is then transformed by a linear projection into a bivariate normal for the trend and offset. The likelihood can be written as a function of the same projection jointly upon observations and simulations such that the negative loglikelihood looks like this

$$nll \sim (\theta_m - \theta_o)^t P^t (\Sigma_m + \Sigma_o)^{(-1)} P(\theta_m - \theta_o) + \log(\det(\Sigma_m + \Sigma_o))$$

where $\theta_{m,o}$ are bivariate vectors containing the offset and slope of the linear fitted function (or any amplitude of a generalized additive model) and P is a matrix containing in its columns the 1′s for the offset and the $(t - t_o)$for the slope (or any function g_k in case of the generalized additive models). note that in this approach the uncertainty in the covariance matrices $\Sigma_{m,o}$ is not yet treated, but this is possible if eg $\Sigma_o = \sigma^2 I$ (I identity matrix) and setting an inverss gamma prior for σ_o^2.

If we are not mistaken, the equation above is not a likelihood. But in general, the independent univariate normal distributional assumption on x_m and y is easily extended to the multivariate case by introducing a prior on a general covariance matrix, a standard prior would be the inverse-Wishart prior.

The inclusion of the P matrix shows that the currently chosen way of trend fitting introduces a correlation between the error in offset and trend because the function $(t - t_o)$ does not sum to zero over the full time interval and therefore any error in the slope will produce an error in the average which has to be compensated by a negative deviation in the error of the mean.

We have now reparameterised the models to centre at the middle of the time series.

This shows that the analytical analysis of the approach using normal distributions firstly does not need a mcmc numerical solution and allows to draw important conclusions about the characteristics of the method.

MCMC is typically needed when the Bayesian analysis does not use specific forms of priors, leaving the posterior distributions intractable. We have left the discussion with the use of MCMC to allow for incorporations of general forms of the prior distribution, because conjugate priors can often be too restrictive. As noted above text has been added to the start of section 2 and 2.1 to clarify this, we add "Conjugate analyses for certain classes of models, including Gaussian error models are often possible, leading to analytical forms for posterior distributions. In this work, we choose to present the results with non standard priors, and use MCMC for computation. The later approach is much easier when extending to more complex modelling scenarios".

Similar remarks can be made for the predictive probability densities, also here a lot can be learned from the (multivariate) normal densities. Note that the multivariate densities always include the univariate case but not the other way around.

The predictive density for observation should take the same form as model output.

Summary The paper is worth to be published in GMD but the authors should comment/ add modifications according to my general remarks. Detailed comments are found in the annotated pdf document. Please also note the supplement to this comment: http://www.geosci-model-dev-discuss.net/gmd-2016-291/gmd-2016-291-RC1- supplement.pdf

- p2, l 4, we have changed the text to "However, their approach still suffers from shortcomings."
- p2,, l 23, see responses above.
- p3, l 8, we do not treat δ as an unknown parameters, therefore we do not place a prior on δ.
- p3, l10, the term here is the standard deviation, not the precision as is used in many Bayesian text books. We do not in fact place a prior on this term, this is a likelihood term which determines the relative weights for us.
- p3, l 20, the weights are predefined, and following the Bayesian model averaging framework, the BMA model is the weighted sum of the competing models.

- p3, l 25, the use of the word centering refers to the location, i.e., about the mean or median of the distribution, it does not imply whether the distribution is skewed or not.
- p3, l 27, we have added some text here to make it clearer. New text now appear just before Sec 2.1 on page 4.
- p4, l1, by most cases, we mean that if we deviate away from the Normal or multivariate Normal distributions, tractable solutions do not exist. See our responses above for further discussions.
- p4, l 13, our weight calculations involve a likelihood term, and a mixture of posteriors term. If the likelihood is Gaussian and the mixture of posteriors are mixtures of Gaussians, then one may expect to obtain analytic solutions. However, we do not have Gaussian posteriors unless conjugate priors were used.
- p4, l 17, this is now removed, as we do not require this calculation throughout the paper.
- p4, l 20, this takes the same form as Equation 5, only that the parameters are based on future model output, as this is the data that is used to make predictions for future observation.
- p5, l 30, we have updated our results by centering the regression on the bias parameter.
- p6, l 24, calculations with training sample do not involve mixtures, the weighted mixture model is used for prediction.
- p7, l 6, the 95% credibility internal, where the location of the interval is indicated by black lines. These are computed from 2.5 and 97.5 quantiles.
- p8, l 29, we have used a simple linear model in the demonstrations in this paper, but the approach can be used for non-linear models, or any generalised models.