Combining climate model output via model correlations

Sain, S R ; Furrer, R

DOI: https://doi.org/10.1007/s00477-010-0380-5

Combining Climate Model Output via Model Correlations

Stephan R. Sain[1] and Reinhard Furrer[2]

September 23, 2009

SUMMARY: In climate science, collections of climate model output, usually referred to as ensembles, are commonly used devices to study uncertainty in climate model experiments. The ensemble members may reflect variation in initial conditions, different physics implementations, or even entirely different climate models. However, there is a need to deliver a unified product based on the ensemble members that reflects the information contained in whole of the ensemble. We propose a technique for creating linear combinations of ensemble members where the weights are constructed from estimates of variation and correlation both within and between ensemble members. At the heart of this approach is a Bayesian hierarchical model that allows for estimation of the correlation between ensemble members as well as the study of the impact of uncertainty in the parameter estimates of the hierarchical model on the weights. The approach is demonstrated on an ensemble of regional climate model output.

KEY WORDS: Model averaging, model correlations, total variation, regional climate models, Bayesian hierarchical model.

[1]Geophysical Statistics Project, Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, ssain@ucar.edu.

[2]Institute of Mathematics, University of Zurich, Switzerland, reinhard.furrer@math.uzh.ch

1 Introduction

Computer-based simulations of geophysical processes, also refereed to as computer models, often represent a synthesis of an entire field of knowledge and offer researchers a way of studying processes that would otherwise be too difficult or too expensive to observe directly. They also offer a framework for obtaining predictions about such processes. While these models are typically deterministic, there are a number sources of uncertainty in these simulations that range from unknown initial conditions, incomplete knowledge of the underlying physics, or even uncertainties about the implementations of these complex codes.

Ensembles or collections of model output are often used to characterize the uncertainty in computer models. Ensemble members may represent variation in initial conditions, variation across physics implementations, or even entirely different computer models. With these ensembles, there is a need to deliver a unified product that captures the common structure in such models or, possibly, spans the variation in the computer model output. These products are often constructed through weighted averages of the ensemble members. The weights control the influence of individual ensemble members, and may allow equal weights (a simple arithmetic average) or weights based on some performance measure or predictive skill.

The study of climate and climate change are scientific areas that commonly uses ensembles of computer model output for studying climate variability and prediction of future climate change. However, combining ensemble members via weights based on some measure of skill at reproducing current climate is not guaranteed to perform well for model runs aimed at a future climate. While this phenomenon is difficult to measure empirically, the fundamental processes simulated by climate models during future runs are inherently different than control runs due to changing model forcings (e.g., increasing greenhouse gasses in the atmosphere).

Recognizing that ensemble members that are aimed at modeling the same physical process are inherently positively correlated, we offer an approach to ensemble weighting

that is based on estimates of internal variation and correlations between ensemble members. Essentially, we fit a statistical model linking the ensemble members explicitly through the covariance structure. We then choose optimal model weights based on a measure of total variation of a weighted average of the ensemble members.

Combining ensemble members via some type of averaging is well established (e.g. Hoeting *et al.*, 1999; Neuman, 2003; Ye *et al.*, 2004). However, many of these approaches treat the individual ensemble members as stochastically independent. The need to link the ensemble members based on correlations between the ensemble members is not new, at least from the perspective of climate science (e.g. Furrer *et al.*, 2007; Tebaldi and Knutti, 2007), but to our knowledge, this is the first formal attempt.

In the following section, we discuss computing averages of ensemble members through an analysis of the total variation of the linear combination of the computer model output in the ensemble. Section 3 gives a brief discussion of regional climate modeling, while Section 4 demonstrates the approach on an ensemble of regional climate model output from the North American Regional Climate Change Assessment Program (NARCCAP). We wrap up the paper in Section 5 with some concluding remarks.

2 Model Averaging via Total Variation

We propose to create a weighted average of the model outputs in an ensemble based on the covariance structure between ensemble members. Denote the collection of model outputs in the ensemble by $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$, where each of the $\mathbf{Y}_i$ are vectors of length n. These multivariate outputs may be, for example, time series or spatial fields. Further, assume that this collection of model outputs has an $nN \times nN$ joint covariance given by $\mathbf{\Sigma} = \text{Block}(\mathbf{\Sigma}_{ij})$. In other words, $\mathbf{\Sigma}$ is a block matrix with N^2 elements $\mathbf{\Sigma}_{ij}$ and where each $n \times n$ matrix $\mathbf{\Sigma}_{ij} = \text{Cov}[\mathbf{Y}_i, \mathbf{Y}_j]$.

Consider a weighted average of the ensemble members of the form

$$\widetilde{\mathbf{Y}} = \sum_{i=1}^{N} w_i \mathbf{Y}_i,$$

where $\{w_i\}$ is a collection of weights. We note that a measure of the total variation (Mardia *et al.*, 1979) of $\widetilde{\mathbf{Y}}$ is given by

$$\operatorname{tr}\left[\operatorname{Var}\widetilde{\mathbf{Y}}\right] = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \operatorname{tr}\boldsymbol{\Sigma}_{ij} = \boldsymbol{w}'\mathbf{A}\boldsymbol{w}, \tag{1}$$

where tr indicates the trace of matrix (sum of diagonal elements), $\boldsymbol{w} = (w_1, \ldots, w_n)'$, and the symmetric matrix $\mathbf{A}$ has elements $a_{ij} = \operatorname{tr}\boldsymbol{\Sigma}_{ij}$.

We consider an approach for choosing weights based on the information about the variation and correlations between ensemble members contained in $\mathbf{A}$. Assuming that the $\mathrm{E}[\mathbf{Y}_i] = \boldsymbol{\mu}$ for all $i = 1, \ldots, N$, then one strategy would be to minimize (1) subject to the constraint $\boldsymbol{w}'\mathbf{1} = 1$ (i.e. the weights sum to one). The assumption of $\boldsymbol{w}'\mathbf{1} = 1$ is equivalent to the unbiasedness constraint in mathematical statistics, $\mathrm{E}[\widetilde{\mathbf{Y}}] = \boldsymbol{\mu}$, as

$$\mathrm{E}[\widetilde{\mathbf{Y}}] = \mathrm{E}\Big[\sum_{i=1}^{N} w_i \mathbf{Y}_i\Big] = \sum_{i=1}^{N} w_i \,\mathrm{E}[\mathbf{Y}_i] = \boldsymbol{\mu}\sum_{i=1}^{N} w_i.$$

However, we are using this device to assume that there is some common structure in the model outputs, $\mathbf{Y}_i$. It is still possible that $\boldsymbol{\mu}$ is not equivalent everywhere to some measure of ground truth. For example, if the ensemble represents climate model output, $\widetilde{\mathbf{Y}}$ may represent some common structure in the ensemble, but it may still include biases for the true climate of the Earth.

Using a Lagrange multiplier to handle the constraint, the minimization problem becomes

$$\min_{\boldsymbol{w}} \boldsymbol{w}'\mathbf{A}\boldsymbol{w} + 2\lambda(\boldsymbol{w}'\mathbf{1} - 1).$$

After taking partial derivatives, the problem can be written as the solution to a system of linear equations:

$$\begin{bmatrix} \mathbf{A} & \mathbf{1} \\ \mathbf{1} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{w} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}.$$

Further, the solution for $\boldsymbol{w}$ is given by

$$\boldsymbol{w} = \mathbf{A}^{-1}\mathbf{1}/(\mathbf{1}'\mathbf{A}^{-1}\mathbf{1}). \tag{2}$$

3 Regional Climate Modeling

Climate models simulate weather over long time periods to capture the long-term behavior of weather processes. Atmosphere-ocean general circulation models (AOGCMs or, more simply, GCMs) couple an atmospheric model to an ocean model and seek to simulate the Earth's global climate system. These computer models have grid boxes on the spatial scale of 100s of kilometers and are extremely useful for the study of large-scale circulation and forcings that affect the Earth's global climate.

There are, however, limitations to the use of GCMs for studying regional climate, and downscaling techniques have been receiving a great deal of attention, in particular for utilizing the output from climate models in impacts studies (public health, agriculture, energy, water, etc.). These techniques use the large-scale information in a GCM to produce representations of climate at much finer spatial scales. Statistical downscaling uses empirical relationships to connect the coarse-scale GCM output to regional or local variables (e.g, Kim *et al.*, 2007). Alternatively, dynamic downscaling uses various forms of higher-resolution climate models to obtain higher-resolution climate simulations.

Regional climate models (RCMs), which are the focus of this work, use grid boxes that are typically on the order of 50 kilometers. However, there is a price for the increased computational burden. For example, RCMs typically use a limited spatial domain and there is generally some simplification of the modeling of ocean processes. Initial conditions and time-dependent lateral boundary conditions for the domain (winds, temperature, moisture, etc.) are obtained from a GCM or a global data set. Thus, global circulation and large-scale forcings are consistent with the GCM, but the higher-resolution forcings in the regional model have the potential to improve the simulation of climate on regional and local scales.

The Intergovernmental Panel on Climate Change (IPCC, http://www.ipcc.ch) assessment reports contain excellent overviews of the issues associated with climate modeling, including downscaling and regional climate modeling. These reports also contain many excellent scientific references. In particular, see the contributions of Working Group I to the Third and Fourth Assessment Reports (Houghton *et al.*, 2001; Solomon *et al.*, 2007). Keller

(2009) also provides an excellent review of some of the more recent issues and controversies.

4 Case Study

In this work, we focus on a subset of the regional climate model output from Phase I of NARCCAP that utilizes five regional models driven by NCEP-DOE Reanalysis 2 (Kanamitsu *et al.*, 2002). (At the time of writing only five of the planned six regional models were available.) The regional models include the OURANOS Canadian Regional Climate Model (CRCM), the UC San Diego/Scripps Experimental Climate Prediction Center Regional Spectral Model (ECPC), the Iowa State University MM5PSU/NCAR mesoscale model (MM5I), the UC Santa Cruz Regional Climate Model version 3 (RCM3), and the Pacific Northwest National Laboratory Weather Research and Forecasting model (WRFP).

Output from each of these models, including the appropriate fields from the NCEP reanalysis data, was interpolated to a common 120×98 grid that covers most of North America. Note that the so-called "sponge zone" (the region on the boundary used to ameliorate the transition from the lower-resolution driving data and the higher-resolution RCM) has been removed from the model outputs. Total winter (December, January, February) precipitation was computed for each year of the twenty-year period spanning the winter of 1980–1981 through the winter of 1999–2000. As climate is often thought of as a long-term average, the twenty-year average fields for each model were computed. Precipitation is typically strongly right-skewed; a fourth-root transformation was applied. Let $\mathbf{Y}_i,\ i = 1, \ldots, 5$, denote these transformed precipitation fields that make up the ensemble; Figure 1 displays these fields. There are higher precipitation values along both coasts, although the highest values are along the northern Pacific coast of the U.S. and British Columbia, Canada.

4.1 A Hierarchical Model

A key aspect of the weighting approach outlined in Section 2 is obtaining a statistical representation for the covariance structure between the ensemble members. While there are many ways to specify such statistical models that ultimately depend on the data being

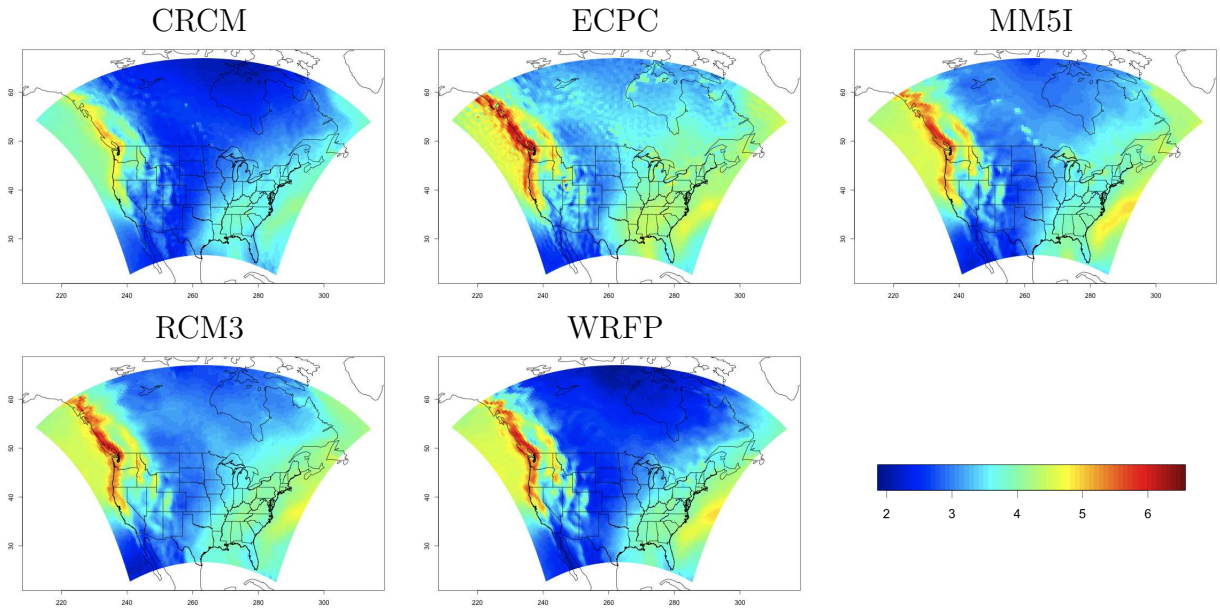

Figure 1: Transformed 20-year-average winter precipitation fields $\{\mathbf{Y}_i\}$.

studied, we focus on a simple three-level hierarchical formulation tailored to this case study. The specific elements of the statistical model have been determined through extensive initial exploratory analysis of NARCCAP's NCEP driven regional climate model output and are summarized below.

The hierarchical structure involves the specification of a data model that links the observed fields $\{\mathbf{Y}_i\}$ to an unobserved spatial process. The process model defines the spatial structure and links the spatial process to parameters. Finally, prior distributions are specified for the parameters. Schematically, this structure can be thought of as specifying three (conditional) densities:

$$\begin{aligned} \text{Data model:} &\quad [\text{data}|\text{process}] \\ \text{Process model:} &\quad [\text{process}|\text{parameters}] \\ \text{Prior model:} &\quad [\text{parameters}] \end{aligned}$$

where $[A]$ indicates the probability density for A and $[A|B]$ indicates a conditional probability density for A given B. See, for example, Berliner (2003).

Letting $\mathbf{Y} = [\mathbf{Y}_1', \ldots, \mathbf{Y}_5']'$, the data model is given by

$$\mathbf{Y}|\boldsymbol{\beta}, \mathbf{U}, \{\sigma_i^2\} \sim \mathcal{N}\left(\mathbf{1}_5 \otimes \boldsymbol{\mu}_{\text{NCEP}} + (\mathbf{I}_5 \otimes \mathbf{X})\boldsymbol{\beta} + \mathbf{U}, \boldsymbol{\sigma} \otimes \mathbf{I}_n\right), \tag{3}$$

where $\otimes$ indicates a Kronecker product, $\boldsymbol{\mu}_{\text{NCEP}}$ indicates the mean of the NCEP (transformed) winter precipitation fields, $\mathbf{X}$ denotes a $n \times q$ matrix of scaled predictors (here we use intercept, longitude, latitude, and elevation, and $q = 4$), $\boldsymbol{\beta} = [\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_5]'$ is a vector of regression parameters with each $\boldsymbol{\beta}_i$ specific to each model output, $\mathbf{U} = [\mathbf{U}_1', \ldots, \mathbf{U}_5']'$ is a vector of spatial random effects, and $\boldsymbol{\sigma} = \text{diag}[\sigma_1^2, \ldots, \sigma_5^2]'$ is a diagonal matrix of scale parameters.

This specification of the process model reflects two important points about NCEP-driven runs of the regional climate models in the NARCCAP experiment. First, the models are all simulating the same common climate and they are all being driven by the same boundary conditions. Ideally, there should be some commonalities in the output fields. However, there are different dynamical features to the individual regional climate models that may lead to subtle (and perhaps not so subtle) differences in the output fields. Hence, the process model in (3) reflects these competing ideas by including a common component relating the output to what might be expected from the driving data (i.e., $\boldsymbol{\mu}_{\text{NCEP}}$) as well as mechanisms for capturing both individual and common deviations from the driving data relating to the increased resolution through covariates such as elevation and through the spatial random effects.

The process model has two parts, given by

$$\begin{aligned} \boldsymbol{\beta}_i|\sigma_{\boldsymbol{\beta}}^2 &\sim \mathcal{N}\left(\mathbf{0}, \sigma_{\boldsymbol{\beta}}^2\mathbf{I}_q\right), \qquad i = 1, \ldots, 5, \\ \mathbf{U}|\boldsymbol{\Omega}, \phi &\sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega} \otimes \mathbf{S}_\phi\right), \end{aligned} \tag{4}$$

where the $\{\boldsymbol{\beta}_i\}$ are assumed to be independent, $\sigma_{\boldsymbol{\beta}}^2$ is a scale parameter, $\boldsymbol{\Omega}$ is a 5×5 covariance matrix and $\mathbf{S}_\phi = \mathbf{V}^{-1}(\phi)$ is an $n \times n$ spatial covariance matrix based on a Markov random field model and indexed by the spatial dependence parameter ϕ with $-1 < \phi < 1$ (see Appendix for details).

The form of the covariance structure for the spatial random effects in (4) assumes separability between the model-to-model covariance and the spatial covariance for a particular

model. Specifically, this implies that the spatial covariance structure is similar across all of the models and the model-to-model correlation does not vary spatially. While this assumption may seem to be restrictive, it provides sufficient flexibility to model the spatial dependence in the individual regional climate models while isolating the model-to-model correlation structure for use in constructing weights (see Section 5 for further discussion on this issue).

Prior distributions are generally taken to be noninformative. Specifically, the prior distributions for parameters $\{\sigma_i^2\}$ and $\sigma_{\boldsymbol{\beta}}^2$ are independent $G(1.0, 0.01)$ where $G(\alpha, \beta)$ indicates a gamma distribution with shape parameter α and scale parameter β, the prior distribution for $\boldsymbol{\Omega}^{-1}$ was taken to be a Wishart distribution with 6 degrees of freedom and a scale matrix equal to $0.01 \cdot \mathbf{I}_5$, and the prior for ϕ was taken to be uniform on $(-1, 1)$.

4.2 Bayesian Computation

Parameter estimation and inference is done in a Bayesian context by sampling from the posterior density given generally as

$$[\text{process}, \text{parameters}|\text{data}] \propto [\text{data}|\text{process}][\text{process}|\text{parameters}][\text{parameters}].$$

Of course, in this case, the posterior density $[\{\sigma_i^2\}, \sigma_{\boldsymbol{\beta}}^2, \boldsymbol{\Omega}, \phi|\mathbf{Y}]$ cannot be computed in closed form, so we use Markov chain Monte Carlo (MCMC) through the implementation of a Gibbs' sampler that utilizes full-conditional distributions or Metropolis-Hastings steps as necessary (Gilks *et al.*, 1996). A single chain of length 5000 was run, with convergence indicated after about 1000 iterations.

Figure 2 shows estimated posterior distributions for the parameters of the variance components $\sigma_{\boldsymbol{\beta}}^2$, $\{\sigma_i^2\}$, and the diagonal elements of $\boldsymbol{\Omega}$. The correlations between the elements of $\mathbf{U}$ in $\boldsymbol{\Omega}$ are shown in Table 1, which show strong positive correlations between the ensemble members in the spatial random effect. While the regression on the predictors contributed substantially to the variability, there is also a great deal of spatial variation with the mean of the posterior distribution of ϕ equal to 0.88 (posterior standard deviation of 0.0021), which is close to its maximal value of 1.0.

Table 1: Fifth and 95th percentiles of posterior density for correlations in $\mathbf{U}$.

	ECPC	MM5I	RCM3	WRFP
CRCM	(0.57,0.63)	(0.73,0.76)	(0.80,0.83)	(0.75,0.78)
ECPC		(0.59,0.64)	(0.67,0.72)	(0.51,0.57)
MM5I			(0.77,0.80)	(0.90,0.92)
RCM3				(0.76,0.79)

Assessing model fit is often difficult with such complex hierarchical models, and more traditional methods (e.g., cross-validation) do not lend themselves to the size and complexity of the regional climate model output analyzed here. However, carefully monitoring the convergence of the Gibb's sampler and carefully examining the posterior densities of the parameters can often indicate such issues as overdetermined models or parameterizations that the data cannot support. Further, examination of the posterior and residual fields can also highlight any difficulties with model fit. However, there was no evidence of such issues for the hierarchical model formulation presented here.

4.3 Results

Integrating over the process model, the first two levels of the hierarchical model outlined in Section 4.1 can be rewritten as

$$\mathbf{Y}|\sigma^2_{\boldsymbol{\beta}}, \boldsymbol{\Omega}, \phi, \{\sigma_i^2\} \sim \mathcal{N}\left(\mathbf{1}_5 \otimes \boldsymbol{\mu}_{\text{NCEP}}, \sigma^2_{\boldsymbol{\beta}}\, \mathbf{I}_5 \otimes \mathbf{X}\mathbf{X}' + \boldsymbol{\Omega} \otimes \mathbf{S}_\phi + \boldsymbol{\sigma} \otimes \mathbf{I}_n\right). \tag{5}$$

In other words, the structure of the hierarchical model is capturing the deviations in the model output from the precipitation field associated with the driving model through a random structure with three components. The first component captures the variation associated with the regression, the second captures the covariance between ensemble members and the spatial dependence, and the third captures the small-scale residual error.

The form in (5) gives us the explicit structure for the components of $\boldsymbol{\Sigma}$ discussed in Section 2:

$$\begin{aligned} \boldsymbol{\Sigma}_{ii} &= \omega_{ii}\mathbf{S}_\phi + \sigma^2_{\boldsymbol{\beta}}\mathbf{X}\mathbf{X}' + \sigma_i^2\mathbf{I}_n \\ \boldsymbol{\Sigma}_{ij} &= \omega_{ij}\mathbf{S}_\phi, \end{aligned}$$

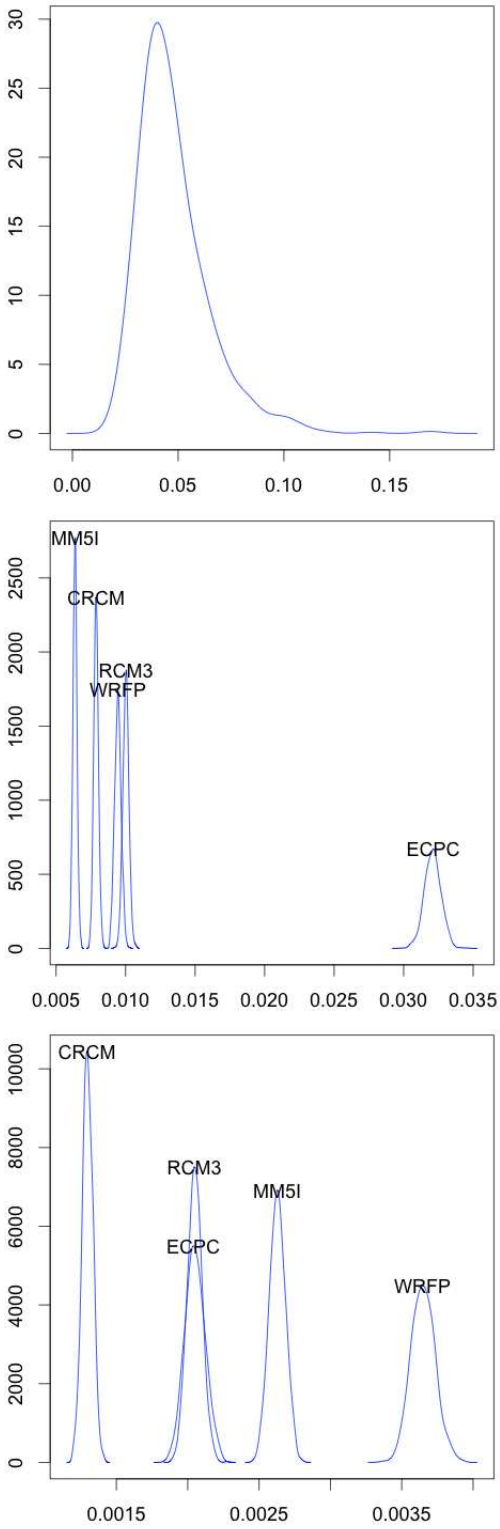

Figure 2: Top plot shows estimated posterior density for $\sigma^2_{\boldsymbol{\beta}}$, middle plot shows estimated posterior density for $\{\sigma_i^2\}$, and bottom plot shows estimated posterior density for the diagonal elements of $\boldsymbol{\Omega}$.

where ω_{ij} are the ijth elements of $\mathbf{\Omega}$. Hence we easily obtain $\mathbf{A}$ with elements given by

$$\begin{aligned} a_{ii} &= \omega_{ii}c_1 + \sigma^2_{\boldsymbol{\beta}}c_2 + n\sigma^2_i \\ a_{ij} &= \omega_{ij}c_1, \end{aligned} \tag{6}$$

where $c_1 = \text{tr}[\mathbf{S}_\phi]$ and $c_2 = \text{tr}[\mathbf{X}\mathbf{X}']$.

Samples of the parameters can be drawn from the posterior and then plugged into (6) to obtain an $\mathbf{A}$ from which weights can be constructed and $\widetilde{\mathbf{Y}}$ can be computed. To demonstrate, consider the following draw of parameters from the posterior: $\sigma^2_{\boldsymbol{\beta}} = 29.8$,

$$\boldsymbol{\sigma} \;=\; 10^{-3} \cdot \text{diag}\left[\begin{array}{ccccc} 7.70 & 31.4 & 6.22 & 10.1 & 9.34 \end{array}\right]',$$

and

$$\mathbf{\Omega} \;=\; 10^{-3} \cdot \left[\begin{array}{ccccc} 1.30 & 0.95 & 1.38 & 1.29 & 1.70 \\ 0.95 & 2.07 & 1.52 & 1.41 & 1.51 \\ 1.38 & 1.52 & 2.72 & 1.80 & 2.91 \\ 1.29 & 1.41 & 1.80 & 1.97 & 2.12 \\ 1.70 & 1.51 & 2.91 & 2.12 & 3.80 \end{array}\right].$$

With $c_1 = 227081$ ($\phi = 0.879$) and $c_2 = 47037$, then

$$\mathbf{A} = \left[\begin{array}{ccccc} 1963.8 & 215.5 & 312.6 & 293.6 & 386.5 \\ 215.5 & 2415.9 & 345.9 & 319.3 & 342.4 \\ 312.6 & 345.9 & 2268.5 & 409.2 & 660.1 \\ 293.6 & 319.3 & 409.2 & 2143.4 & 480.5 \\ 386.5 & 342.4 & 660.1 & 480.5 & 2550.8 \end{array}\right].$$

From (2), the weights are given by

$$\boldsymbol{w} = \left[\begin{array}{ccccc} 0.27 & 0.21 & 0.18 & 0.21 & 0.13 \end{array}\right]'. \tag{7}$$

Note these weights take into account both the relative variation in the random effects, but are also adjusted by the correlations in the ensemble members.

To compare different strategies for weighting ensemble members, Figure 3 shows an equally weighted average of the ensemble members as well as $\widetilde{\mathbf{Y}}$ with weights computed in (7). Of course, the plots show similar patterns, being based on constrained weights and only five models. However, there are higher precipitation values in $\widetilde{\mathbf{Y}}$ across much of the

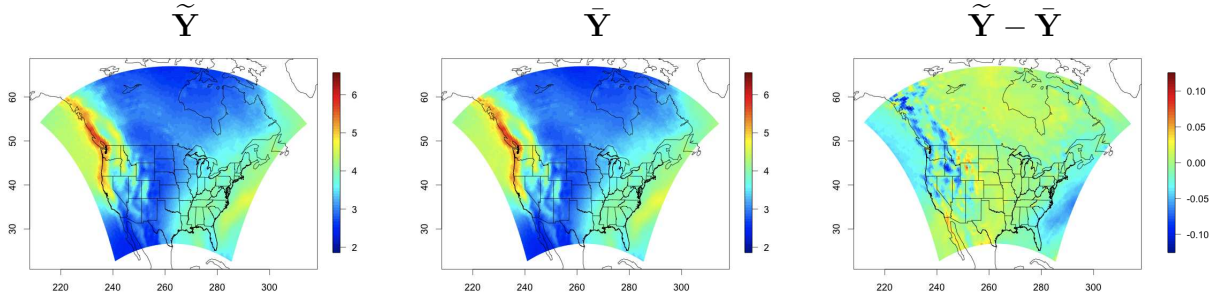

Figure 3: $\widetilde{\mathbf{Y}}$ (left frame) and $\bar{\mathbf{Y}}$ (middle frame) and their difference (right frame) based on a random draw from the posterior distribution of the statistical model parameters.

middle portion of the domain. In what follows we demonstrate that these differences are persistent and are not just due to uncertainty in the estimation of the statistical model parameters that determine the weights.

One of the benefits of the Bayesian formulation is that we can draw from the posterior distribution of the parameter estimates to understand the uncertainty in the weights and hence $\widetilde{\mathbf{Y}}$. Figure 4 shows estimated densities of $\log(w_i/0.2)$ taken from 1000 random draws from the posterior distribution of the statistical model parameters, which shows the deviations in the weights from an equally weighted scheme (i.e. a simple average). Weights for CRCM are typically larger, while weights for WRFP are typically smaller, but both show larger variation in their weights than the other three models. The weights for ECPC, MM5I, and RCM3 are typically closer to 0.2, but with ECPC and RCM3 slightly larger and MM5I slightly lower. Reassuringly, the density estimates suggest that there is little uncertainty in the estimates of the weights.

While a rigorous assessment of the general properties of $\widetilde{\mathbf{Y}}$ is beyond the scope of this paper, we can compare $\widetilde{\mathbf{Y}}$ to both the twenty-year average of the NCEP fields as well as to an equally weighted combination of the $\{\mathbf{Y}_i\}$. Figures 5 and 6 show differences in the mean $\widetilde{\mathbf{Y}}$ and the simple average of the ensemble members and the average of the NCEP precipitation fields, respectively. The mean of $\widetilde{\mathbf{Y}}$ was constructed by taking 1000 draws from the posterior distribution of the statistical model parameters, constructing the weights for each draw, and computing a $\widetilde{\mathbf{Y}}$ for each set of weights. The probabilities of $\widetilde{\mathbf{Y}}$ being

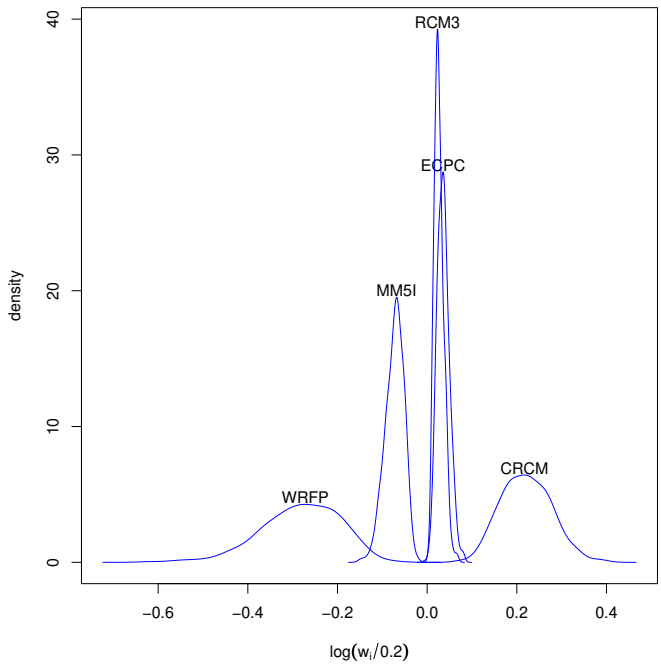

Figure 4: Estimated densities of $\log(w_i/0.2)$ based on samples from the posterior distributions of the statistical model parameters.

greater than both of these means are also shown in Figures 5 and 6. In general, the mean of $\widetilde{\mathbf{Y}}$ shows much higher precipitation values than the twenty-year average of NCEP over much of the domain. The consistency of the weights, which reflects the uncertainty, or lack thereof, in the estimation of parameters in the statistical model, induces a nearly binary map of the $P[\widetilde{\mathbf{Y}} > \boldsymbol{\mu}_{\text{NCEP}}]$ in the right frame of Figure 6. Further, $\widetilde{\mathbf{Y}}$ seems to have consistently higher precipitation values than the simple average of the ensemble members over much of the midwestern United States, as well as northern Canada.

5 Discussion

Motivated by research in quantifying uncertainty in climate model experiments, we have outlined an approach for weighting ensemble members that is fundamentally different from approaches that are based solely on some measure of predictive skill or accuracy at reproducing some ground truth. Our approach recognizes that climate models are typically

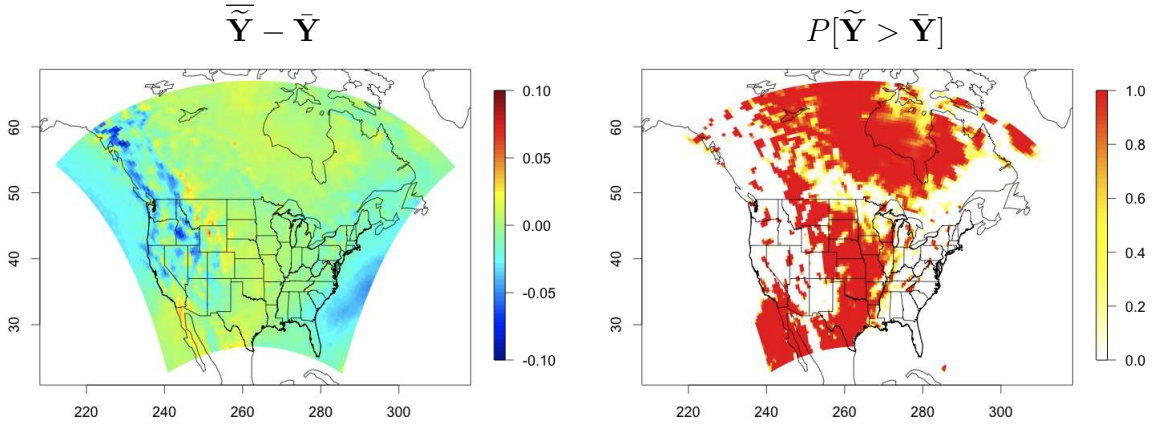

Figure 5: Difference between the mean of $\widetilde{\mathbf{Y}}$s and the simple average of the ensemble members (left frame) and the estimated pointwise probabilities of $\widetilde{\mathbf{Y}}$ being greater than the simple average (right frame). Both are computed from weights constructed from 1000 draws from the posterior distribution of the statistical model parameters.

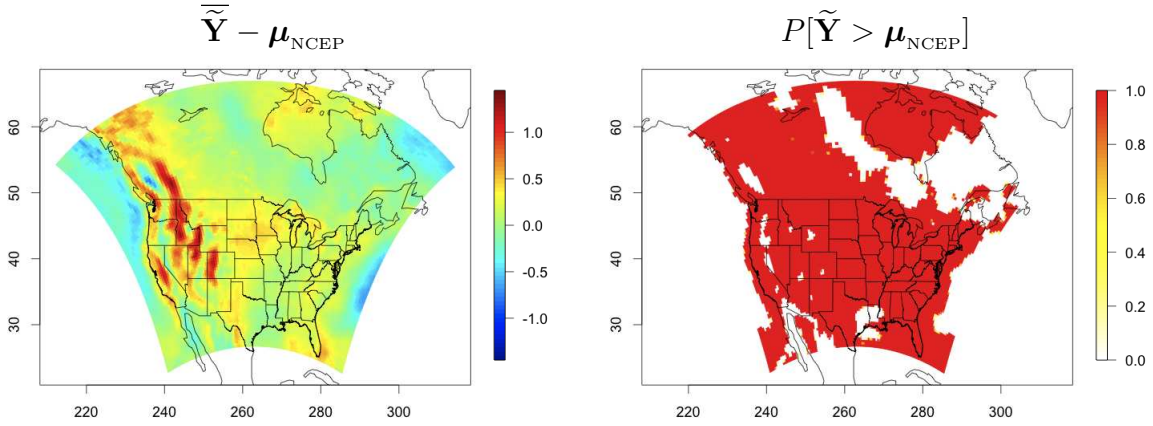

Figure 6: Difference between the mean of $\widetilde{\mathbf{Y}}$s and the twenty-year average of the NCEP precipitation fields (left frame) and the estimated pointwise probabilities of $\widetilde{\mathbf{Y}}$ being greater than the twenty-year average of NCEP (right frame). Both are computed from weights constructed from 1000 draws from the posterior distribution of the statistical model parameters.

positively correlated and utilizes this correlation structure in constructing weights for computing linear combinations of ensembles of climate model output. This is especially important when considering climate models predicting a future climate when there is no ground truth available. We are currently examining how such a weighting scheme performs with climate model ensembles that include future runs. While the methodology was demonstrated on RCMs, there is no reason not to consider correlations for weighting global models and this is also a current line of our research.

We expect that situations similar to ensembles of climate models can be found in other problems in geoscience, as well as in many other disciplines. Further, it is possible to consider ensembles that include the output of stochastic models in this framework, including predictions from various statistically-based analyses.

At the heart of this paper is a statistical representation for the ensemble members that specifically includes a correlation structure between the ensemble members. Parameters for this statistical model are estimated in a Bayesian context. This model includes some simplifications, in particular a Kronecker form that is a simplification of a more general multivariate version of a Markov random field. Specially, the covariance structure on the spatial random effects that account for individual model differences is based on the assumption of separability between the covariance structure between the models and the spatial covariance within the models. It is possible to consider more complex forms of Markov random field models (e.g. Sain and Cressie, 2007; Sain *et al.*, 2008a) or even other forms of multivariate spatial models (e.g. Cressie, 1993; Wackernagel, 2003) that relax this assumption and allow for the spatial dependence structure to vary across models or allow for the correlation structure between models to vary spatially or some combination of the two. Further, the model outlined here could be expanded to include correlations in the regression parameters in (4). Of course, computing constructing weights under these more general situations is not so clear, and these issues are currently being examined by the authors in current research projects.

We note that this approach is not necessarily aimed at prediction, at least in the sense

of constructing a predictive distribution for a new (or additional) model. Such a statistical model would require a fundamentally different structure in the process model of the hierarchy outlined in Section 4.1. This is also an area of active research for the authors.

Appendix

The form for $\mathbf{S}_\phi = \mathbf{V}^{-1}(\phi)$ in (4) was assumed to be induced by a type of Markov random field, a spatial statistical model ideal for gridded or lattice type spatial observations (Rue and Held, 2005). The essential idea behind such models is the characterization of the spatial dependence through the specification of conditional distributions that link the observation at each spatial location on the lattice to its neighbors. The collection of conditional distributions effectively leads to the specification of the joint precision matrix (inverse covariance matrix) for all of the locations on the grid or lattice.

Assuming Gaussian conditional distributions, these conditional autoregressive (CAR) models are specified through the conditional mean and variance. A very simple specification of the mean and variance are given by

$$\mathrm{E}[y_i|\mathbf{Y}_{-i}] = \mu_i + \phi \sum_{j \in N_i} (y_j - \mu_j) \qquad \mathrm{Var}[y_i|\mathbf{Y}_{-i}] = \sigma^2, \qquad i = 1, \ldots, n, \tag{8}$$

where $\mathbf{Y} = [y_1, \ldots, y_n]$, $\mathbf{Y}_{-i}$ indicates all of the elements of $\mathbf{Y}$ except the ith one, N_i denotes a collections of indices representing the neighbors of the ith location on the lattice, and ϕ is a partial or conditional correlation between two neighbors. This collection of conditional distributions can be shown to lead to joint Gaussian distribution with mean $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]$ and with covariance matrix $\boldsymbol{\Sigma} = \sigma^2(\mathbf{I}_n - \phi\mathbf{C})^{-1}$ with $\mathbf{I}_n$ the $n \times n$ identity matrix and $\mathbf{C}$ the $n \times n$ incidence matrix determined by the neighborhood structures $\{N_i\}$.

The choice of neighborhoods often plays a crucial role in the behavior of CAR models. For example, in the simple specification in (8), one might simply choose neighbors as grid boxes that share an edge. The version of a Markov random field model used in this work follows a Kronecker form between the precision matrices for two one-dimensional processes, both indexed by the dependence parameter ϕ with one process for the rows and the other for

the columns of the output grid for the climate models (see Sain *et al.*, 2008b). Specifically, the conditional means and variances for interior grid boxes are given by

$$\mathrm{E}[y_i|\mathbf{Y}_{-i}] = \mu_i + \frac{\phi}{1+\phi^2}\sum_{j\in N_{1i}}(y_j - \mu_j) - \left(\frac{\phi}{1+\phi^2}\right)^2 \sum_{j\in N_{2i}}(y_j - \mu_j)$$
$$\mathrm{Var}[y_i|\mathbf{Y}_{-i}] = \sigma^2 \frac{1}{(1+\phi^2)^2},$$

where N_{1i} and N_{2i} represents indices for neighboring grid boxes that share an edge or vertices, respectively. (Note that the specification for boundary and corner grid boxes are slightly different; see Sain *et al.*, 2008b, for details.) While this formulation is more complex, the (unconditional) spatial covariance structure is stationary (which the formulation in (8) lacks) and the additional neighbors gives more smoothness to the spatial fields.

Finally, it is important to note that the specification of these models defines the inverse of the covariance matrix, typically referred to as the precision matrix. Furthermore, the precision matrix is generally a sparse matrix. Hence, one can dramatically improve computational performance of the statistical model for large spatial grids, both in terms of the storage of the precision matrix and the linear algebra computations associated with computing the likelihood (see also Furrer and Sain, 2009).

Acknowledgements

This research of the first author was supported in part by National Science Foundation (NSF) grants ATM-0502977 and ATM-0534173. The second author was supported in part by the NSF grant DMS-0621118. In addition, much of the research in this paper was done while the second author was an assistant professor in the Department of Mathematical and Computer Sciences at the Colorado School of Mines in Golden, CO. We wish to thank the North American Regional Climate Change Assessment Program (NARCCAP) for providing the data used in this paper. NARCCAP is funded by the NSF, the U.S. Department of Energy (DoE), the National Oceanic and Atmospheric Administration (NOAA), and the U.S. Environmental Protection Agency Office of Research and Development (EPA). The

National Center for Atmospheric Research is managed by the University Corporation for Atmospheric Research under the sponsorship of the NSF.

References

Berliner, L. M. (2003). Physical-statistical modeling in geophysics. *Journal of Geophysical Research (Atmospheres)*, **108**, 8776, doi:10.1029/2002JD002865.

Cressie, N. A. C. (1993). *Statistics for Spatial Data, rev. edn.* John Wiley & Sons Inc., New York.

Furrer, R., Knutti, R., Sain, S. R., Nychka, D., and Meehl, G. A. (2007). Spatial patterns of probabilistic temperature change projections from a multivariate Bayesian analysis. *Geophysical Research Letters*, **34**, L06711, doi:10.1029/2006GL027754.

Furrer, R. and Sain, S. R. (2009). spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *Journal of Statistical Software*, to appear.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). Introducing Markov chain Monte Carlo. In *Markov Chain Monte Carlo in Practice*, 1–19. Chapman & Hall, London.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical Science*, **14**, 382–417.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C. A., editors (2001). *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S., Hnilo, J. J., Fiorino, M., and Potter, G. L. (2002). NCEP–DOE AMIP–II reanalysis (R-2). *Bulletin of the American Meteorological Society*, **83**, 1631–1643.

Keller, C. F. (2009). Global warming: A review of this mostly settled issue. *Stochastic Environmental Research and Risk Assessment*, **23**, 643–676.

Kim, B. S., Kim, H. S., Seoh, B. H., and Kim, N. W. (2007). Impact of climate change on water resources in Yongdam Dam Basin, Korea. *Stochastic Environmental Research and Risk Assessment*, **21**, 355–373.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.

Neuman, S. P. (2003). Maximum likelihood Bayesian averaging of uncertain model predictions. *Stochastic Environmental Research and Risk Assessment*, **17**, 291–305.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Application*. Chapman & Hall/CRC Press, Boca Raton, FL.

Sain, S. R. and Cressie, N. (2007). A spatial model for multivariate lattice data. *Journal of Econometrics*, **140**, 226–259.

Sain, S. R., Furrer, R., and Cressie, N. (2008a). Combining ensembles of regional climate model output via a multivariate Markov random field model. *Annals of Applied Statistics*, Submitted.

Sain, S. R., Kaufman, C., and Tebaldi, C. (2008b). A spatial analysis of a regional climate model experiment. Unpublished manuscript.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., editors (2007). *Climate Change 2007: The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC*. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Tebaldi, C. and Knutti, R. (2007). The use of the multimodel ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A*, **365**, 2053–2075.

Wackernagel, H. (2003). *Multivariate Geostatistics - An Introduction with Applications*. Springer-Verlag, Berlin, 3rd edition.

Ye, M., Neuman, S. P., and Meyer, P. D. (2004). Maximum likelihood Bayesian averaging of spatial variability models in unsaturated fractured tuff. *Water Resources Research*, **40**, W05113, doi:10.1029/2003WR002557.