Hierarchical Modeling of Variability in Regional Climate Models Using Markov Random Fields

Stephan R. Sain\textsuperscript{1} and Cari Kaufman\textsuperscript{2}

July 7, 2011

SUMMARY: Collections of climate model output, called ensembles, are used to explore the uncertainties associated with projections of climate change. These uncertainties can come from many sources, and we outline a framework based on a hierarchical model for exploring and quantifying these uncertainties. A key feature of this hierarchical model is that it can be expressed as an analysis of variance (ANOVA) where the effects are spatial surfaces. These surfaces are modeled via a novel form of a Markov random field that exploits the regular grid structure to produce a computationally efficient representation. This approach is demonstrated on the first phase of the regional climate model ensemble being generated by the North American Regional Climate Change Assessment Program. This particular ensemble includes six regional models with common boundary conditions provided by reanalysis data, and the results of the statistical analysis on total winter precipitation demonstrates that there are significant differences in the models relative to the interannual variability.

KEY WORDS: Functional analysis of variance (ANOVA), climate model ensembles, uncertainty quantification.

\textsuperscript{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.

\textsuperscript{2}Department of Statistics, 315 Evans Hall, University of California, Berkeley, CA 94720, cgk@stat.berkeley.edu
1 Introduction

Climate models are useful tools for studying the Earth’s climate and producing projections of climate change. They encapsulate an entire field of scientific knowledge about the physical processes comprising the Earth’s climate system and represent the combined efforts of hundreds of scientists and researchers. However, there is still a great deal of uncertainty inherent in climate modeling, and different models can produce very different projections under the same driving conditions. The sources of variability in climate models can be studied and characterized through a collection of model output arrived at by varying the conditions under which a single model is run, by considering a collection of different models, or both. The output from such experiments is often referred to as an ensemble. Uncertainty in climate model projections stems from multiple sources, including unknown initial climate states and future driving forces (i.e. greenhouse gases), lack of knowledge of tunable model parameters, and inaccurate or incomplete representation of physical processes in the model. The variability in climate model ensembles, both over time within a model and between different models, can be used to characterize some but not all of this uncertainty. With much current interest focused on regional and even local projections of climate change, an additional uncertainty arises from the resolution at which climate projections are made and how low-resolution model output is “downscaled” to produce localized, higher-resolution projections. The North American Regional Climate Change Assessment Program (NARCCAP; Mears et al., 2009, 2011) is a large-scale, international program that is seeking to explore the uncertainties in future climate projections on a regional scale through the use of a multi-model ensemble of regional climate model (RCM) output. We focus in this paper on the development of statistical methodology to better understand and quantify the sources of variability in climate model ensembles, and, in particular, the ensemble being generated through NARCCAP.
1.1 Regional Climate Models and NARCCAP

Atmosphere-ocean general circulation models (often referred to as global climate models; denoted here as GCMs), simulate weather over long periods of time through a coupling of an atmosphere model with an ocean model. As their name suggests, with grid boxes on the order of 200 km, GCMs are excellent tools for studying climate and climate change on large, even global, scales. However, many of the smaller-scale processes important for the study of climate change and impacts of climate change on regional scales are not well-resolved. RCMs are of similar construction to their global counterparts, although with some simplification of the representation of ocean processes in the models. They are usually run over a limited spatial domain and have boundary conditions supplied by GCMs. Hence, RCMs represent a type of dynamical downscaling of the large-scale information in the GCM. RCMs typically have grid boxes on the order of 50 km.

NARCCAP is an experiment in two phases. In the first phase, boundary conditions for six RCMs were supplied by a global data set, the NCEP-DOE Reanalysis 2 (Kanamitsu et al., 2002). The regional models in NARCCAP include the OURANOS Canadian Regional Climate Model (CRCM), the UC San Diego/Scripps Experimental Climate Prediction Center Regional Spectral Model (ECPC), the Hadley Centre Hadley Regional Model 3 (HRM3), the Iowa State University MM5–PSU/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). In the second phase (still to be completed), four GCMs are supplying the boundary conditions for the same six RCMs.

The NCEP dataset used in the first phase is a gridded data product obtained by assimilating meteorological observations into a numerical model similar to those used for weather forecasts. Many types of observations, from weather stations, ships, aircrafts, radiosondes, and satellites, are conditioned upon to produce a current best estimate of climate variables over the recent past. This estimate is defined over a space-time grid, spanning the entire globe and from 1948 to the present. With boundary conditions provided by NCEP, we can evaluate the behavior of the RCMs under conditions that mimic, to the best of our
knowledge, current physical conditions on Earth.

Our focus in this paper is on this first phase, specifically to assess the within-model and between-model variability and how this differs across spatial locations. We will describe our methodology as applied to a particular outcome variable, average winter precipitation, although one of the goals of the statistical development is a method that can be applied easily and rapidly to many different variables. In assessing the differences between RCMs, we will model their output relative to the NCEP reanalysis, that is, conditioning on it. The model allows us to make inference about deviations: 1) from each RCM to the average response across RCMs, and 2) from this average to the values from NCEP. By assessing the magnitude of these differences in a probabilistic framework, we can identify variables and regions with systematic deviations, targeting areas for climate modelers to investigate.

The concepts and methods outlined here are being developed with an eye on the second phase of the NARCCAP experiment and the analysis of the full ensemble. This will include an assessment of different sources of uncertainty inherent in an ensemble of RCMs, as well as the development of probabilistic projections of climate change. See Mearns et al. (2009, 2011) for an overview of NARCCAP; see also the NARCCAP website (www.narccap.ucar.edu) for further details and references on the regional models.

1.2 Model Output

For each RCM, as well as for the NCEP reanalysis data, total winter (December, January and February) precipitation was computed for each year of the twenty-year period from the winter of 1980-1981 through the winter of 1999-2000. These seasonal summaries were interpolated from the individual model grids (each of which is at a 50 km resolution) to a common 50 km resolution grid (120 × 98) that covers most of North America. Let \( Y_{ij}, i = 1, \ldots, 6, j = 1, \ldots, 20 \), denote the precipitation field for the \( i \)th model and the \( j \)th year. A fourth-root transformation was applied to correct a strong right skew. The twenty-year averages of the transformed precipitation fields are shown in Figures 1 and 2. All the models as well as the NCEP reanalysis data show higher precipitation values
along the coasts, with the highest values along the Pacific Northwest in the United States and British Columbia in Canada. However, while there are large-scale similarities seen in Figures 1 and 2, there are a number of differences in how the models as a whole compare to NCEP and how they differ from each other. The methodology and analysis presented here will provide a probabilistic description of where the models differ from the driving model and where the model-to-model variation is large, taking into account the year-to-year variability within models.

1.3 Computation and Spatial Modeling

One of the major issues in carrying out a spatial analysis of the output from RCMs is the size of the individual fields. In our motivating example, these are made up of more than 10,000 spatial grid boxes. Spatial sample sizes for similar experiments will grow considerably with the next generation of petascale computing environments. Unfortunately, traditional geostatistical approaches to spatial modeling and interpolation involve the storage and manipulation of spatial covariance matrices that are not feasible for large datasets, as the necessary matrix computations for parameter estimation and spatial interpolation grow at a cubic rate with the number of observations.
Figure 2: Average (transformed) precipitation fields for CRCM and ECPC (top row), HRM3 and MM5I (middle row), RCM3 and WRFP (bottom row).
A number of approaches have been suggested to aid in the computation associated with large spatial problems. For example, Furrer et al. (2006) and Kaufman et al. (2008) propose covariance tapering, which introduces sparsity into spatial covariance matrices. Tapering can reduce the storage and computational resources needed to manipulate these covariance matrices while retaining sound theoretical properties. In another approach, Cressie and Johannesson (2008) discuss reduced-rank kriging, which reduces the dimensionality of the matrices that need to be manipulated through the use of a well-chosen set of basis functions. This is one of many models that exploit a low-rank representation of the spatial field; see e.g. Wikle (2010) for a recent review of such models. Finally, algorithms for the implementation of spatial interpolation on high-performance, distributed computing environments have been suggested (e.g., Kerry and Hawick, 1998), but these approaches are not practical for many users and also often ignore issues with parameter estimation.

Markov random field (MRF) models (e.g., Rue and Held, 2005) offer a different approach, tailored more specifically for spatial lattice data. The computational benefit of MRFs stems from specifying the covariance matrix through its inverse, the precision matrix. This avoids the need to invert the covariance in the Gaussian likelihood. Furthermore, this inverse is typically quite sparse, yielding improved performance through reduced storage and computational requirements.

1.4 Outline

One of the statistical goals in NARCCAP is to build an analytical framework for spatial comparisons across the multi-model ensembles being generated in the program. The approach discussed in this paper borrows elements from Tebaldi et al. (2005) and Kaufman and Sain (2010) and incorporates a hierarchical construction that allows for comparisons with both the driving model and across the multi-model ensemble. A key element in the construction is a spatial covariance structure that allows for efficient computation. Other studies that discuss combining climate model output include Furrer et al. (2007), Berliner and Kim (2008), Tebaldi and Sansó (2009), Smith et al. (2009), Buser et al. (2010), and
In the following section, we describe our hierarchical model. Section 3 discusses a Markov random field parameterization of the the spatial covariance structure in the model outlined in Section 2. Section 4 describes the results of fitting our proposed model to data from Phase I of NARCCAP. We make some concluding remarks in Section 5.

2 A Hierarchical Model

The goal of our analysis is to make probabilistic inference about sources of variability in the climate model output, conditionally on the observed output fields. We now describe the hierarchical model that relates each set of observations to these sources of variability. Start with the variability from year to year: for a particular model $i$, the observations for a given year $j$ may be modeled as a sample from the climatology of that particular model, with

$$Y_{ij}|\mu_i, \sigma^2_i, \phi_i \sim N(\mu_i, \sigma^2_i V(\phi_i)), \quad i = 1, \ldots, 6, \quad j = 1, \ldots, 20.$$  \hfill (1)

The $\{\mu_i\}$ represent the long-run average winter precipitation fields for each model. Note that the model does not include temporal dependence between the observations, as there is very little correlation in the seasonal totals from year to year. The variance structure reflects the fact that observations for a given year will tend to differ from $\mu_i$ in a spatially coherent way. That is, we could rewrite (1) as $Y_{ij} = \mu_i + \epsilon_{ij}$, where $\epsilon_{ij}$ represents a yearly anomaly that we a priori expect to have mean zero and a spatial structure determined by $\sigma^2_i V(\phi_i)$. The $\{\sigma^2_i\}$ are scale parameters and each $V(\phi_i)$ is a spatial covariance matrix indexed by a spatial dependence parameter $\phi_i$. The form of these matrices is discussed in the following section.

Now consider the variability amongst the RCMs in the long-run average winter precipitation fields. We model these as coming from a common distribution

$$\mu_i|\mu, \sigma^2, \phi \sim N(\mu, \sigma^2 V(\phi)), \quad i = 1, \ldots, 6.$$ \hfill (2)

Here $\mu$ represents the average across the RCMs, and again we have deviations from the...
overall mean that are also spatially coherent for each model. Indeed, we could likewise rewrite (2) as $\mu_i = \mu + \alpha_i$, where the $\{\alpha_i\}$ represent deviations for each model from the consensus between models.

Finally, we specify a distribution for $\mu$:

$$\mu \mid \mu_{NCEP}, \sigma^2_\mu, \phi_\mu \sim \mathcal{N}(\mu_{NCEP}, \sigma^2_\mu V(\phi_\mu)), \quad (3)$$

where $\mu_{NCEP}$ is simply the average of the (transformed) precipitation fields from the driving NCEP reanalysis data. This implies that a priori we expect the average response of the RCMs to be similar to NCEP, which is not unreasonable considering that the NCEP data provides the boundary conditions for each RCM. However, we do not exclude the possibility of differences. Indeed, as in the reparameterizations above, we can likewise rewrite (3) as $\mu = \mu_{NCEP} + \eta$, with $\eta$ representing the difference or bias compared to observations.

The model specified in (1)-(3) is fully centered in the sense of Gelfand et al. (1995) and is efficiently fit using Markov chain Monte Carlo (MCMC). However, a completely equivalent parameterization is clearly to write

$$Y_{ij} = \mu_{NCEP} + \eta + \alpha_i + \epsilon_{ij}, \quad i = 1, \ldots, 6, \quad j = 1, \ldots, 20.$$  

This representation makes the partitioning into various sources of variability clear. It is analogous to a simple analysis of variance (ANOVA) model, but the effects are functions rather than scalars, so that spatial correlation is taken into account. See Kaufman and Sain (2010) for further details regarding this interpretation. As we discuss in our results, it allows for comparisons between the models and NCEP by examining the posterior distribution of $\eta$, and amongst the models themselves by examining the joint posterior distribution of the $\{\alpha_i\}$. The magnitude of the various effects may also be summarized using derived model quantities called finite population variances, as we will discuss.

The following section describes the particular form of the spatial covariance model we adopt, after which we complete the hierarchical model description by specifying prior distributions for the hyper-parameters appearing in (1)-(3).
3 A Stationary Markov Random Field Model

Spatial data laid out on regular grids, such as remote sensing data or climate model output, or irregular lattices, such as many government or administrative boundaries, can require alternative approaches to traditional geostatistical models. One such approach is based on a Markov random field. Essentially, spatial dependence is induced through the specification of conditional distributions that are based on neighborhoods. The stronger the dependence on values at neighboring locations, the stronger the resulting spatial dependence.

More formally, Besag (1974) showed that, for random variables \(Y_1, \ldots, Y_n\) observed at \(n\) locations on a spatial lattice structure, the collection of conditional distributions \(f(Y_i|Y_{-i})\), \(i = 1, \ldots, n\), (where \(Y_{-i}\) refers to all random variables except the \(i\)th one) can be combined under certain regularity conditions to form a joint distribution \(f(Y_1, \ldots, Y_n)\). Assuming Gaussian conditional distributions for \(f(Y_i|Y_{-i})\) gives rise to an important special case referred to as the conditional autoregressive (CAR) model. The conditional mean and variance of a CAR model can be parameterized as

\[
E[Y_i|Y_{-i}] = \mu_i + \sum_{j=1}^{n} b_{ij}(Y_j - \mu_j) \quad \text{and} \quad Var[Y_i|Y_{-i}] = \tau^2_i,
\]

for \(i = 1, \ldots, n\). This collection of conditional distributions gives rise to a joint Gaussian distribution,

\[
\mathcal{N}(\mu, (I - B)^{-1}T),
\]

where \(\mu = (\mu_1, \ldots, \mu_n)'\), \(I\) is an \(n \times n\) identity matrix, \(B\) is the \(n \times n\) matrix with entries \(b_{ij}\), and \(T = \text{diag}(\tau_1^2, \ldots, \tau_n^2)'\). Of course, the spatial dependence parameters, \(\{b_{ij}\}\), must be chosen to ensure that the resulting matrix, \((I - B)^{-1}T\), is a valid covariance matrix; that is, \((I - B)^{-1}T\) is symmetric and positive-definite. For more details, see Rue and Held (2005) or the discussions in Cressie (1993, Part II) or Banerjee et al. (2004, Chapter 3).

Our approach involves a very specific form for the conditional means and variances and ultimately the spatial covariance matrices, \(V\), outlined in the previous section. This form will yield a computationally efficient and yet still stationary covariance structure. Denote
\( Y_k \) as the measurement associated with \( s_k \), the \( k \)th position in the lattice. For example, on the regular grids associated with the regional climate model output, \( s_k = (i, j) \) where \( i \) indicates the row and \( j \) the column. For points on the interior of the regular grid, the conditional mean and variance are given by

\[
E[Y_k | Y_{-k}] = \mu_k + \frac{\phi}{1 + \phi^2} \sum_{\ell \in N_{1k}} (Y_\ell - \mu_\ell) - \left( \frac{\phi}{1 + \phi^2} \right)^2 \sum_{\ell \in N_{2k}} (Y_\ell - \mu_\ell)
\]

\[
Var[Y_k | Y_{-k}] = \sigma^2 \frac{1}{(1 + \phi^2)^2},
\]

where \( N_{1k} \) denotes the indices of the first-order neighbors of \( s_k \), \( \{ \ell : ||s_k - s_\ell|| = 1 \} \), and \( N_{2k} \) denotes the indices of the second-order neighbors of \( s_k \), \( \{ \ell : ||s_k - s_\ell|| = \sqrt{2} \} \). See Figure 3. The parameter \( \phi \) controls the extent of the spatial dependence and can take on values between -1 and 1. A value of \( \phi = 0 \) implies that \( Y_k \) is conditionally independent of its neighbors and, as seen below, that the collection of \( \{Y_k\} \) are mutually independent.

For points on the corners (Figure 4), the conditional mean and variance are given by

\[
E[Y_k | Y_{-k}] = \mu_k + \phi \sum_{\ell \in N_{1k}} (Y_\ell - \mu_\ell) - \phi^2 \sum_{\ell \in N_{2k}} (Y_\ell - \mu_\ell)
\]

\[
Var[Y_k | Y_{-k}] = \sigma^2.
\]

The boundary points are more complicated (Figure 5) with

\[
E[Y_k | Y_{-k}] = \mu_k + \frac{\phi}{1 + \phi^2} \sum_{\ell \in N_{1k}} (Y_\ell - \mu_\ell) + \phi \sum_{\ell \in N_{2k}} (Y_\ell - \mu_\ell) - \left( \frac{\phi^2}{1 + \phi^2} \right) \sum_{\ell \in N_{3k}} (Y_\ell - \mu_\ell)
\]

\[
Var[Y_k | Y_{-k}] = \sigma^2 \frac{1}{1 + \phi^2},
\]

where, in this case, \( N_{1k} \) contains first-order neighbors on the interior, \( N_{2k} \) contains first-order neighbors on the edge, and \( N_{3k} \) contains second-order neighbors.

Putting these conditional means and variances together into a joint distribution as in (4) gives rise to spatial precision matrix of the form

\[
\frac{1}{\sigma^2} V(\phi)^{-1} = \frac{1}{\sigma^2} R(\phi)^{-1} \otimes C(\phi)^{-1},
\]

where \( R^{-1} \) and \( C^{-1} \) are precision matrices specific to the rows and columns, respectively, of the grid. The matrix \( R^{-1} \) is a symmetric, positive-definite matrix with the same number
Figure 3: First-order (left) and second-order (right) neighborhoods for interior points.

Figure 4: First-order (left) and second-order (right) neighborhoods for corner points.

Figure 5: First-order neighbors on the interior (left), first-order neighbors on the edge (center), and second-order neighbors (right) for edge points.
of rows and columns as the number of rows in the regular grid. Its form is given by

$$R(\phi)^{-1} = \begin{bmatrix} 1 & -\phi \\ -\phi & 1 + \phi^2 & -\phi \\ & \ddots & \ddots \\ & & -\phi & 1 + \phi^2 & -\phi \\ & & & -\phi & 1 \end{bmatrix}, \quad (6)$$

where the entries outside the diagonal and first off-diagonal are zero. $C^{-1}$ is defined similarly, although with the same number of rows and columns as the number of columns in the regular grid.

We note that (6) represents the precision matrix associated with the undirected or full conditional distributions of a univariate time-series autoregressive model (Rue and Held, 2005, Section 1.1). This form does not depend on the temporal direction, making it an appropriate model for a univariate spatial process. Further, the idea of incorporating time-series models in a spatial setting has precedent in the literature on analyzing field experiments; see, for example, Martin (1990) and Cullis and Gleeson (1991).

While Markov random field models offer an approach specifically for lattice data, the computational benefits stem from the specification of the precision matrix and the fact that this matrix is sparse (see, for example, Furrer and Sain, 2010, for more details and a discussion of these computational advantages). In this case, however, the Kronecker form in (5) adds significant additional benefits in terms of computation and storage. In particular, since we are only working with smaller matrices $R$ and $C$, the computation of the determinant and quadratic form in the likelihood can be dramatically reduced by noting that

$$|R \otimes C| = |R|^c |C|^r,$$

where $r$ and $c$ are the number of rows and columns of the regular grid, respectively, and

$$Y'(R \otimes C)Y = Y' \text{vec}(CY^*R),$$

where the vec operator stacks the columns of a matrix into a vector, and $Y^*$ is the “unstacked” matrix formed from $Y$; that is, vec($Y^*$) = $Y$. 

12
Finally, there are certain issues that arise with many common forms for MRFs, including a lack of stationarity in the implied spatial covariance structure (see, for example, Wall, 2004). The forms of $R^{-1}$ and $C^{-1}$, and ultimately $V^{-1}$, are motivated by an attempt to incorporate the computational benefits of Markov random field models but with a stationary form. The implied covariance for $V^{-1}$ can be calculated explicitly. In fact, the elements of $R$ are given by 

$$r_{ii'} = \frac{\phi |i-i'|}{1 - \phi^2}$$

for any two rows $i$ and $i'$. The elements of $C$ are defined similarly. Hence, the covariance between any two points $s_k = (i, j)$ and $s_{k'} = (i', j')$ on the grid is given by 

$$\sigma^2 \frac{\phi |i-i'| + |j-j'|}{(1 - \phi^2)^2},$$

which is does not depend on any particular location on the grid. Under this parameterization, the marginal variance for any particular element of the random vector is therefore $\sigma^2/(1 - \phi^2)^2$.

4 Results for NARCCAP Winter Precipitation

To complete the specification of hierarchical model outlined in Section 2, we set prior distributions for the scale parameters ($\{\sigma_i^2\}, \sigma^2, \sigma_\mu^2$) and the spatial dependence parameters ($\{\phi_i^2\}, \phi^2, \phi_\mu^2$). Our choice for prior distributions for the spatial dependence parameters is uniform over their possible range, i.e. $\pi(\phi) = U(-1, 1)$. Prior distributions for the scale parameters were taken to be $\pi(\sigma^2) \propto \sigma^{-2} I\{0 < \sigma^2 < A^2\}$ (i.e., $\pi(\sigma) = U(0, A)$) where $A$ is taken to be some large value. In this case, $A$ was set to be 100, which is an order of magnitude greater than any value that we might expect from the climate model output.

We simulated from the posterior distribution using the Gibbs sampling algorithm, using Metropolis Hastings steps to sample spatial dependence parameters. Convergence and mixing were monitored graphically. Twenty samples of parameters were drawn from an initial chain of 5000 iterations, and these values were used as starting values for 20 more chains with 3000 iterations each.

Figure 6 shows the posterior mean for $\mu$, the common mean for the RCMs, and $\eta$, the difference between the model ensemble taken as a whole and the NCEP precipitation field. The mean field for $\eta$ suggests that the RCMs exhibit higher winter precipitation values over the mountains in the western United States and Canada and slightly higher precipitation
Figure 6: Posterior mean for $\mu$ (top left), $\eta$ (top right), and pointwise posterior probabilities that $\eta > 0$ (bottom).
values over the coastlines in western Canada, the middle of the United States, and north of Hudson Bay. Posterior probabilities that $\eta > 0$, calculated separately for each location, are shown in the bottom frame of Figure 6 and reinforce the higher precipitation values exhibited by the RCMs over the mountain regions.

A similar approach can be used to compare the responses of different RCMs. Figure 7 displays maps of posterior probabilities that $\alpha_i(s) > 0$ for each model $i$ and location $s$. The ECPC model has precipitation values that are consistently above the common level, while the WRFP model is consistently drier than the the common level. Overall, however, the maps show considerable variability. Figure 8 shows pairwise comparisons between the models, plotting posterior probabilities that $\alpha_i(s) > \alpha_j(s)$ for $i < j$. A crude ordering of the models can be inferred from a visual inspection of these maps based on the amount of precipitation (above or below the mean). Again, WRFP appears to have the lowest precipitation overall, followed by CRCM and HRM3, then MM5I and RCM3, and then ECPC with the highest amount. Of course, these maps also allow for comparisons on specific regions in the domains as well.

While the model-to-model variability is an important component in the uncertainty
Figure 8: Pairwise comparisons between the models where the maps indicate pointwise posterior probabilities that $\alpha_i > \alpha_j$, with $i$ indicating the row and $j$ indicating the column. The scale ranges from 0 (dark blue) to 1 (dark red), so that red areas show locations for which the model indicated by row generates higher precipitation than the model indicated by column, and blue areas show locations for which it generates less precipitation. For example, the top left plot indicates that in the majority of locations, CRCM generates less precipitation than ECPC with high posterior probability, although the opposite is true in western mountain and southern ocean regions.
associated with climate model ensembles, the interannual variability within each model can also be quite large, and it is important to understand how the interannual variability compares to the model-to-model variability. To that end, Figure 9 shows regions where the model-to-model variability is larger than the residual year-to-year variability. Specifically, samples of \( \{\alpha_i\} \) were drawn from the posterior and, for each grid box location \( s \), the finite sample variance \( s^2_\alpha(s) \) and residual variance \( s^2_\epsilon(s) \) were computed. These are derived model quantities, defined as 
\[
 s^2_\alpha(s) = \alpha(s)^T [I - J/6] \alpha(s)/(6 - 1),\]
where \( I \) is the identity matrix and \( J \) is a matrix of ones, and 
\[
 s^2_\epsilon = \sum_{i=1}^6 \sum_{j=1}^{20} \epsilon_{ij}(s)^2/(6 \times 20).\]
They are spatial analogues of the between-group and within-group variances as defined in a traditional ANOVA model (Kaufman and Sain, 2010; see also Gelman, 2005). Over much of the domain the interannual variability is larger than the model-to-model variability with high probability, although there are some regions where the opposite is true. Further, it seems clear that many of these regions are driven by topographical features of the domain such as coastlines, lakes, and mountain ranges. The increased resolution of underlying topography is an important difference between RCMs and lower-resolution GCMs. The phenomenon
observed in Figure 9, that model-to-model variability is larger than interannual variability with high probability for relatively few locations, and in ways that follow topography, is consistent with results reported in Kaufman and Sain (2010), who analyzed a different collection of RCMs that were run over a different domain.

Finally, as a check on the structure of this statistical model, two reduced models were also fit to these data. The first assumes that $\alpha_i = 0$ for all $i$. In other words, the RCMs are allowed to deviate from the driving NCEP model, but their climatologies are all assumed to be the same. The second model assumes that both $\alpha_i = 0$ and $\eta = 0$, effectively treating the RCMs as if they have the same mean structure as the NCEP fields. Deviance information criterion (DIC; Spiegelhalter et al., 2002) values for three models are: -11575106 (full model), -10047211 ($\alpha_i = 0$), -9605455 ($\alpha_i = 0, \eta_0$), suggesting that there is evidence to support the full model and the results and the interpretations that the full model implies.

5 Concluding Remarks

In this work, we have presented a statistical approach to understanding and quantifying the uncertainty in climate model ensembles through a hierarchical model. This construction allows for a deliberate treatment of the sources of uncertainty through the different levels of hierarchy and has an interpretation that is analogous to a simple ANOVA. However, in contrast to traditional forms of ANOVA, the effects in this case are spatial functions. To model the spatial correlation on the regular grids associated with climate model output, a Markov random field is used. This spatial model allows for us to account for spatial dependence while also taking care with computational efficiency since the size of the data sets is larger than what can be used effectively with more traditional (i.e., geostatistical) methods.

Climate model experiments typically generate ensembles that span one or more sources of uncertainty. While the approach outlined here is entirely general and can be expanded to many different climate modeling experiments as well as into other settings, we highlight
a regional climate model experiment where the goal is to better understand how regional climate models perform when driven by near-ideal conditions. In this particular case, we have found that the regional climate models differ significantly (when compared to a measure of interannual variability) near complex topographic features such as coastlines and mountain ranges, a result that is consistent with other analysis focused on a climate model experiment using different models and over a different (and substantially smaller) domain. Finally, it should be noted that these results have been presented to the climate modelers associated with NARCCAP, and this work has spawned further collaboration to better understand the physical processes that determine the Earth’s climate and how these processes are manifest in climate models.

Acknowledgements

This research was supported, in part, by National Science Foundation (NSF) grants ATM-0502977, ATM-0534173, and DMS-0707069. 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


Mearns, L. O., Gutowski, W., Jones, R., Leung, R., McGinnis, S., Nunes, A., and Qian,


