Special Issue Paper

Received: 9 July 2010, Accepted: 5 August 2010, Published online in Wiley Online Library: 5 November 2010

(wileyonlinelibrary.com) DOI: 10.1002/env.1068

Functional ANOVA and regional climate experiments: a statistical analysis of dynamic downscaling†

Stephan R. Sain[a]*, Doug Nychka[a] and Linda Mearns[a]

Climate model experiments incorporating model runs conducted using different conditions have become popular for the study of uncertainty affecting model output and projections of climate change. Recently, such experiments have been used to also study of the uncertainties in producing high-resolution projections of climate change based on methods for dynamic downscaling of global models. In this paper, we discuss an initial analysis of a subset of the ensemble being produced by the North American Regional Climate Change Assessment Program. Using an approach based on a functional analysis of variance, we examine the differences between two different dynamic downscaling methods and demonstrate that there are significant differences between the two models and their projections of summer temperature and precipitation over North America. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: functional ANOVA; Markov random fields; regional climate models

1. INTRODUCTION

The physical processes that determine the Earth's climate exist on many different temporal and spatial scales, and it is a grand scientific challenge to model the dynamics of this system over time and in response to external changes. Moreover, the scientific consensus that human activities has led to warming of the atmosphere has added an urgency to predict changes in our climate due to our future behavior. The primary tools used to project future climate are large numerical models, based on physics, that simulate the different components of the climate system. Known as atmosphere-ocean general circulation models (AOGCMs) they include components for the atmosphere, the ocean, land processes and sea ice. These AOGCMs have proven invaluable in the scientific study of climate and climate change and a key source of information supporting the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (Solomon *et al*., 2007, http://www.ipcc.ch).

Although AOGCMs are effective at reproducing and studying circulation and forcings (factors that influence the Earth's climate, such as natural changes in solar or volcanic activity, human-induced changes to greenhouse gasses or land cover, etc.) on global scales, they have a limited use for studies focused on regional climate. Due to computational resources, they are restricted to coarse spatial scales and may not capture the spatial and temporal variability of meteorology that is observed in the real world and that is relevant to economic, social, and ecological systems. For example, climate model experiments for the IPCC Fourth Assessment Report had spatial resolutions on the order of 150–200 km and so would reduce the area of the United States to a few hundred grid points. Even the large western states are discretized to just a few points. For example, Colorado, with an area of nearly 270 000 km^2 and wildly diverse topography, including high plains in the eastern portion of the state and the Rocky Mountains in the western portion of the state, would be spanned by approximately five to six grid boxes. Part of this problem is that an AOGCM simulates climate by generating a series of weather events that are then compiled to estimate a climatological distribution. Such a model must have time steps on the order of minutes even if the final goal is estimating annual or even decadal averages of a physical variable. Typically it takes a month of dedicated time on a supercomputer for a state-of-the-art AOGCM to simulate a century of the climate system.

The approach used in climate science to achieve higher spatial resolution is to use the AOGCM output as a starting point and downscale this information using additional models. In this work, we provide a statistical analysis that evaluates two different methods of downscaling that involve geophysical models running at finer resolution. It is of scientific interest to see how these two methods differ not only in reproducing current climate but also when the downscaling is used to infer future changes.

* *Correspondence to: Stephan R. Sain, Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, CO, USA. E-mail: ssain@ucar.edu*

a *Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, CO, USA*

† *These articles are published in Environmetrics as a special issue on Handling complexity and uncertainty in environmental studies, arising from the TIES- GRASPA joint conference held in Bologna in 2009 and is edited by Daniela Cocchi, Department of Statistics University of Bologna, Italy and E. Marian Scott, School of Mathematics and Statistics, University of Glasgow, UK.*

Dynamic downscaling, which is the focus of this paper, uses climate models that take "boundary information" from the full AOGCM and produce a higher resolution simulation of the atmosphere. Perhaps the most common way to do this is to define a subdomain of the Earth's surface and use the time-varying, geophysical variables of the AOGCM on the boundary of the domain to provide the constraints for a high-resolution simulation within the domain. These simulations are carried out using a *regional climate model* (RCM). Global circulation and large-scale forcings are then consistent with the AOGCM, but the higher-resolution geophysical features included in RCMs (e.g., topography and land cover) should improve climate simulations on regional scales. Another strategy, termed a *time slice experiment* uses observed sea surface temperatures as well as perturbations of the surface conditions of the ocean simulated in the AOGCM and couples these to a higher-resolution version of the atmospheric component of the AOGCM. A time slice experiment has the advantage that it is not restricted to a subdomain, but it may involve more computation. In either case, these higher-resolution models depend on the AOGCM to provide information either at the domain boundary or about the state of the ocean. However, there is also the potential that these two modeling strategies downscale the AOGCM simulation in very different ways.

The climate model experiments analyzed in this work are a subset of a larger suite of experiments associated with the North American Regional Climate Change Assessment Program (NARCCAP; see Mearns *et al*., 2009). NARCCAP is an ambitious program involving climate research groups from around the world that is focused on examining the uncertainties in RCMs and producing high-resolution projections of climate change over North America. The primary experimental design is to use a collection of four different AOGCMs to drive six different RCMs, yielding a multi-model ensemble of RCM output. In addition, two global atmospheric models are included in the basic experimental design as time slice experiments. Although there are many specific questions about physical processes that can be studied within this experiment, our perspective is that a functional analysis of variance (FANOVA) approach can be useful in summarizing the overall results of NARCCAP with respect to different AOGCM/RCM and time slice combinations. The methods that we present in this work can extend to the full NARCCAP experiments when they are completed.

In this paper, we consider two NARCCAP runs that downscale a climate simulation run by Geophysical Fluid Dynamics Laboratory (GFDL). These are (1) time slice experiments using a high-resolution version of the atmospheric component of the GFDL climate model (AM2.1; see The GFDL Global Atmospheric Model Development Team, 2004) and (2) the regional climate model RegCM3 (see, for example, Pal *et al*., 2007) supported primarily by scientific groups in the Abdus Salam International Centre for Theoretical Physics and the University of California, Santa Cruz. One expectation is that the downscaled results would be similar since they share the same parent AOGCM. However, the two incorporate information from the GFDL global climate run in different ways. The time slice is constrained by perturbations in sea-surface temperatures and sea ice while the regional model is constrained by the time-dependent atmospheric fields used as lateral boundary conditions.

The downscaling experiments are done for these two models, (1) and (2), under current climate conditions and for a future scenario, yielding a classical 2^2 factorial experiment. Some basic scientific questions are whether the change in temperature or precipitation is influenced by the choice of downscaling method and if there is a significant interaction between the downscaling method and the future projections. An interaction in this case is important because it indicates that the downscaling method is not an additive effect and makes the experiments more difficult to interpret. Downscaling methods, and climate models in general, must be validated against past or present climate and their reliability for future projections can only be inferred indirectly. The presence of interactions makes it more difficult to connect the validation of the model under current climate with quantifying the uncertainty of future projections. From a scientific perspective, interactions are also interesting because they may indicate a nonlinear response among the physical processes between the global climate model and the downscaling model.

What is novel in the statistical problem is that the responses for each of these four runs are two-dimensional spatial fields of average surface temperature and average precipitation over the NARCCAP domain. We have found it useful to formulate this structure as a Bayesian hierarchical model where the main effects and interactions have priors expressed as spatial processes. The Bayesian context also facilitates inferences for these fields. A computational contribution in this work are some useful approximate models to make the statistical calculations feasible. Specifically, we use Markov random fields (MRFs) and Kronecker products of covariance matrices to improve the efficiency of the computations needed to evaluate multivariate Gaussian likelihoods for covariance parameters.

In the following section, we describe the model output studied in this paper. Section 3 describes our statistical approach which is based on a two-way FANOVA model, while Section 4 will describe some results and insights discovered when applying the FANOVA model in this setting. Finally, we discuss some issues and future work in Section 5.

2. MODEL OUTPUT

The NARCCAP protocol specifies two runs of the regional models, a current run from 1971 to 2000 that uses the "Climate of the 20th Century" runs produced by the AOGCMs and a future run from 2041 to 2070 that uses the A2 scenario runs of the AOGCMs. The A2 scenario (Nakicenovic *et al*., 2000) represents a relative high emissions scenario, increasing CO_2 concentration levels from the current values of around 380 ppm to about 870 ppm by the end of the 21st century.

The output from both the time slice and the RCM is archived in 3-hour increments on their native spatial grid. Seasonal summaries for average daily temperature and precipitation are computed and interpolated to a common 134×83 grid, yielding 30 annual fields for both the current and future runs. The focus in this paper will be limited to the summer season (June, July, and August), although the methods outlined here could be used to examine the other seasons, as well as other output fields of interest.

Figures 1–4 display selected temperature and precipitation fields from the timeslice and from the RCM. A fourth-root transformation was applied to the precipitation fields to account for the skewness in the distribution. (The fourth-root was chosen after an exploratory analysis with a selection of different years and locations and examining a sequence of power transformations.) The yearly labels in the figures provide a temporal index but do not reflect the observed weather for that particular year. Rather, due to the particular initial conditions, boundary

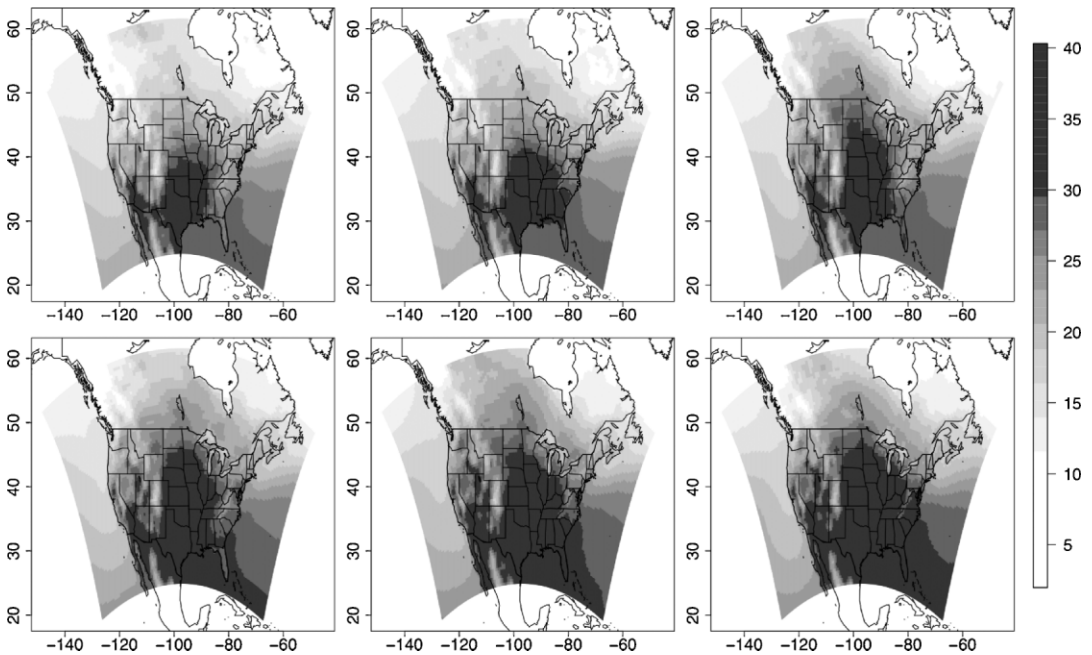

Figure 1. Selected average daily temperature (°*C*) fields for the GFDL timeslice. Top row includes 1980, 1990, and 2000 (left to right) while the bottom row includes 2050, 2060, and 2070 (left to right)

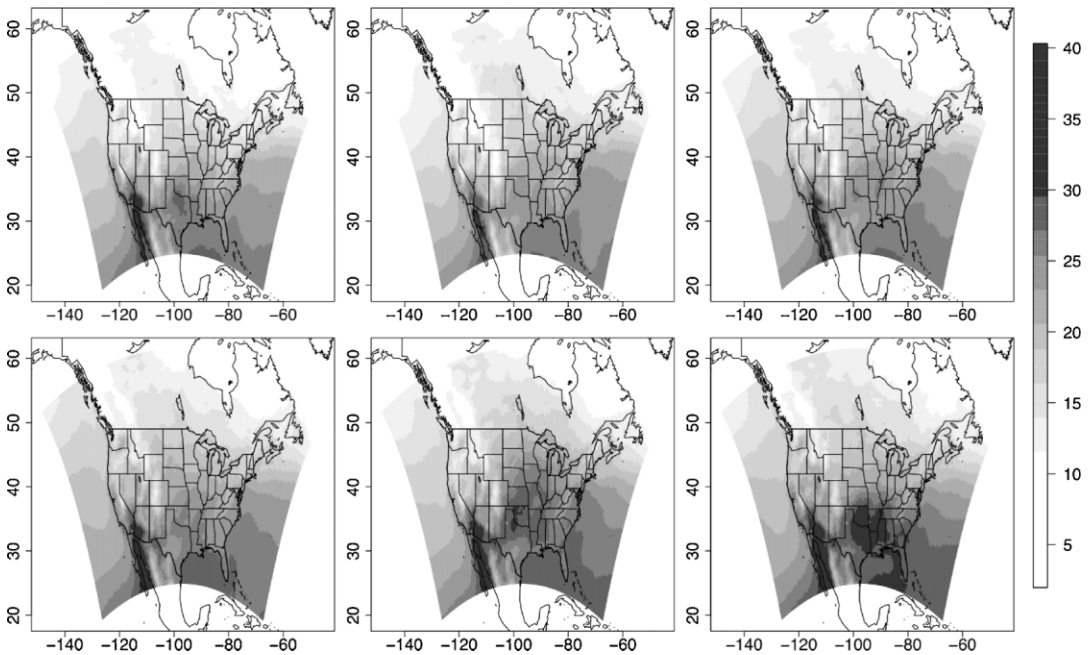

Figure 2. Selected average daily temperature (°*C*) fields for the GFDL-driven regional model. Top row includes 1980, 1990, and 2000 (left to right) while the bottom row includes 2050, 2060, and 2070 (left to right)

conditions, and forcings, the output from climate models more typically reflects the distribution of the weather for a particular set of conditions and time periods. (It should be noted that the time slice uses observed oceans and, hence, the output may have more in common with observed weather for a specific time period than the RCM, which only uses information from the forcing AOGCMs. However, neither is likely to reproduce specific weather events.)

There are some observations of note that can be made from Figures 1–4. First the GFDL time slice typically exhibits higher temperatures than the GFDL-driven RCM, and both show evidence of wide-spread warming over the entire domain when comparing years from the current run to the future run. The RCM appears to have greater summer precipitation, although the patterns of change between current and future runs seem less consistent between the two models.

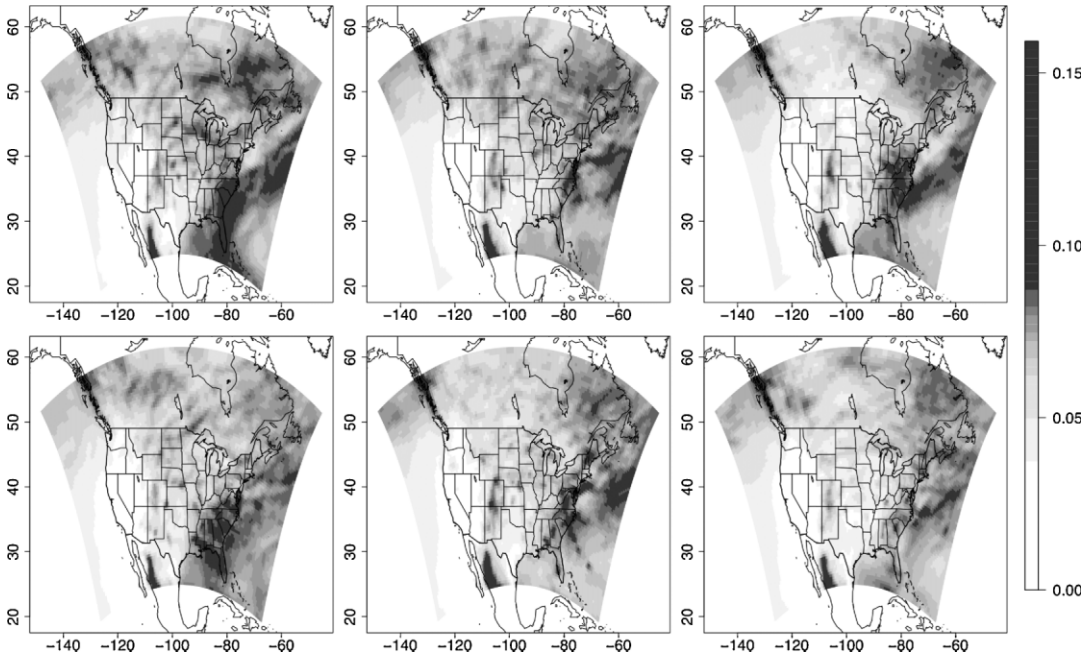

Figure 3. Selected average daily precipitation fields for the GFDL timeslice. Top row includes 1980, 1990, and 2000 (left to right) while the bottom row includes 2050, 2060, and 2070 (left to right)

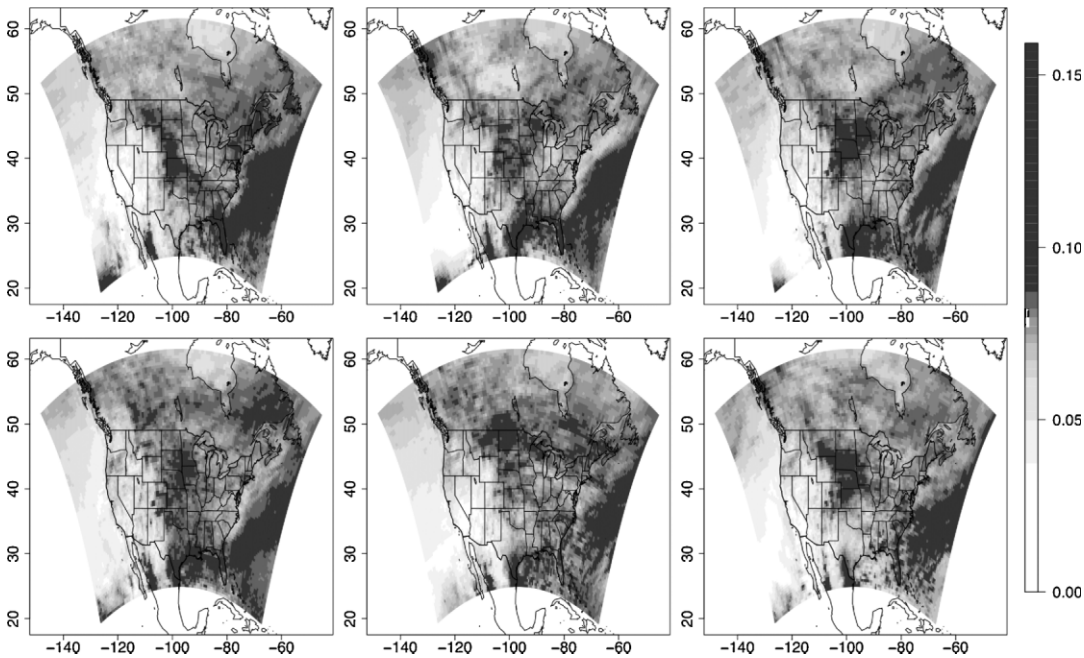

Figure 4. Selected average daily precipitation fields for the GFDL-driven regional model. Top row includes 1980, 1990, and 2000 (left to right) while the bottom row includes 2050, 2060, and 2070 (left to right)

3. FUNCTIONAL ANALYSIS OF VARIANCE

Climate models are deterministic, yet there is great flexibility in these computer models to produce a wide variety of output. And, while the size and quantity of output at times seems overwhelmingly large, the number of runs is still limited, making statistical methods necessary to describe the distribution of the model output. To accomplish the comparison between the time slice and the RCM, we expand on the FANOVA technology developed in Kaufman and Sain (2010); see also the related work of Zhang *et al*. (2009) on utilizing an analysis of variance approach in a spatial setting.

Let $\mathbf{Y}_{ijt}$ denote the output field from the ith model ($i = 0$ indicates the time slice and $i = 1$ indicates the RCM), the jth run ($j = 0$ indicates the current run and $j = 1$ indicates the future run), and the tth year ($t = 1, \ldots, 30$). The primary structure of the model expands $\mathbf{Y}_{ijt}$ into

factors that capture the variation across the two downscaling procedures and across the current and future runs; that is, expand $\mathbf{Y}_{ijt}$ as

$$\mathbf{Y}_{ijt} = \boldsymbol{\alpha}_0 + i\boldsymbol{\alpha}_1 + j\boldsymbol{\alpha}_2 + ij\boldsymbol{\alpha}_3 + \boldsymbol{\epsilon}_{ijt} \quad (1)$$

where $\boldsymbol{\alpha}_0$ represents a baseline corresponding to the current run of the time slice, $\boldsymbol{\alpha}_1$ is an adjustment representing the difference between the time slice and the RCM for the currrent run, $\boldsymbol{\alpha}_2$ is an adjustment representing the difference between the current and future runs (i.e., a common estimate of climate change across both time slice and RCM), and $\boldsymbol{\alpha}_3$ represents an interaction that allows for differences between the two models to vary between the current and future run. The error term, $\boldsymbol{\epsilon}_{ijt}$, captures the year-to-year variation.

While Equation (1) is reminiscent of a standard two-way analysis of variance model, it is important to remember that the $\mathbf{Y}$'s and each of the $\boldsymbol{\alpha}$ components represent surfaces. Kaufman and Sain (2010) placed Gaussian process distributions on such components. In this work, we do something similar, although with two important differences. First, noting that the two downscaling models share the same parent AOGCM, we incorporate information from the GFDL AOGCM into these distributions. Second, the size of the spatial fields (over 11 000 grid boxes) make computing with traditional approaches to spatial covariance specifications problematic. However, the regular lattice structure of the models' spatial grid is an ideal case for spatial covariance specifications based on MRF models (e.g., Rue and Held, 2005), and we incorporate these into the specification of the distributions for the $\boldsymbol{\alpha}$ components. Hence, we assume that each of the $\boldsymbol{\alpha}$ components are realizations from a Gaussian MRF:

$$\boldsymbol{\alpha}_0 \sim \mathcal{N}\big(\boldsymbol{\mu}_{\text{current}}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\big) \quad \boldsymbol{\alpha}_1 \sim \mathcal{N}\big(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_1)\big)$$
$$\boldsymbol{\alpha}_2 \sim \mathcal{N}\big(\boldsymbol{\mu}_{\text{difference}}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_2)\big) \quad \boldsymbol{\alpha}_3 \sim \mathcal{N}\big(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_3)\big)$$

where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ represents the spatial covariance matrix based on a Gaussian MRF with parameter vector $\boldsymbol{\theta}$ (see Appendix), $\boldsymbol{\mu}_{\text{current}}$ is the 30-year average temperature constructed from the current run of the GFDL AOGCM, and $\boldsymbol{\mu}_{\text{difference}}$ is the 30-year average change in temperature constructed from the difference between the future and current runs from the GFDL AOGCM.

It is expected that temperatures will rise in response to the greenhouse gas forcings associated with the A2 scenario. After some initial exploratory analysis, it was found that a linear term, held constant over each grid box, captured much of the variation associated with the increase within the current or future runs. This suggest a form for the error term that has two components; that is,

$$\boldsymbol{\epsilon}_{ijt} = \gamma_j(t - 15.5) + \boldsymbol{\eta}_t$$

where $\gamma_j \sim \mathcal{N}(\gamma_j^*, \sigma_\gamma^2)$ and $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_t))$. It should be noted that the γ_j^* are slopes constructed from the control and future runs of the GFDL AOGCM (again assuming a constant slope across each spatial location).

Finally, to complete the FANOVA model specification, prior distributions are needed for the MRF parameters $\boldsymbol{\theta}$. The parameter vector $\boldsymbol{\theta} = (\sigma^2, \phi)$ consists of a scale parameter and a dependence parameter, respectively; see Appendix for details. Prior distributions for both are non-informative, with the dependence parameter having a uniform prior distribution (i.e., $\pi(\phi) = U(-1, 1)$) and the prior distribution for the scale parameter taken to be $\pi(\sigma^2) \propto 1/\sigma^2 \cdot A$ (i.e., $\pi(\sigma) = U(0, A)$) where $A = 100$.

Our choices for the distributions for the $\boldsymbol{\alpha}$s and the prior distributions on the $\boldsymbol{\theta}$ parameters requires some comment. Our working assumption is that both the timeslice and the RCM will bear some similarity to the parent AOGCM. Hence, the mean of the distributions for the $\boldsymbol{\alpha}$s that represent the current climate ($\boldsymbol{\alpha}_0$) and the common estimate of climate change ($\boldsymbol{\alpha}_2$) are informed by the 30-year averages from the parent AOGCM. Correspondingly, the terms that represent differences in the two downscaling procedures ($\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_3$) have zero mean. However, we want to allow sufficient flexibility for the ANOVA model to discern discrepancies between the two downscaling procedures, so the prior distributions on the parameters controlling the variability of those distributions are sufficiently vague.

4. RESULTS

Direct computation of the posterior distribution is not possible in this setting and we use Markov chain Monte Carlo (MCMC) (e.g., Gilks *et al*., 1996) to generate samples from the posterior. A Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990; Gelfand *et al*., 1990) was constructed with full conditional distributions derived for all but the spatial dependence parameters; Metropolis-Hastings steps (Metropolis *et al*., 1953; Hastings, 1970) with Gaussian proposal distributions were used for these. A single Markov chain was run with 5000 iterations. The proposal variance on the Metropolis-Hastings steps was updated every 100 iterations for the first 2000 iterations. Convergence was monitored graphically, and samples were taken from the final 3000 iterations for posterior inference.

4.1. Temperature

Figure 5 shows posterior means for the $\boldsymbol{\alpha}$ components for summer temperature. From the plots of the posterior means, it is clear that (1) the size of the difference between the models captured in $\boldsymbol{\alpha}_1$ is generally larger than the common climate change captured in $\boldsymbol{\alpha}_2$, (2) the common climate change suggests widespread warming throughout the extent of the domain, and (3) the size of the interaction captured in $\boldsymbol{\alpha}_3$ is very small.

Of course, the strength of such models is not only in computing averages but also in being able to capitalize on the ability to draw samples from the distribution of the $\boldsymbol{\alpha}$ components. For example, computing the posterior probability that the interaction term $\boldsymbol{\alpha}_3$ is different from zero for each gridbox, Figure 6 shows grid boxes where the interaction is credibly negative (white) or credibly positive (dark grey)

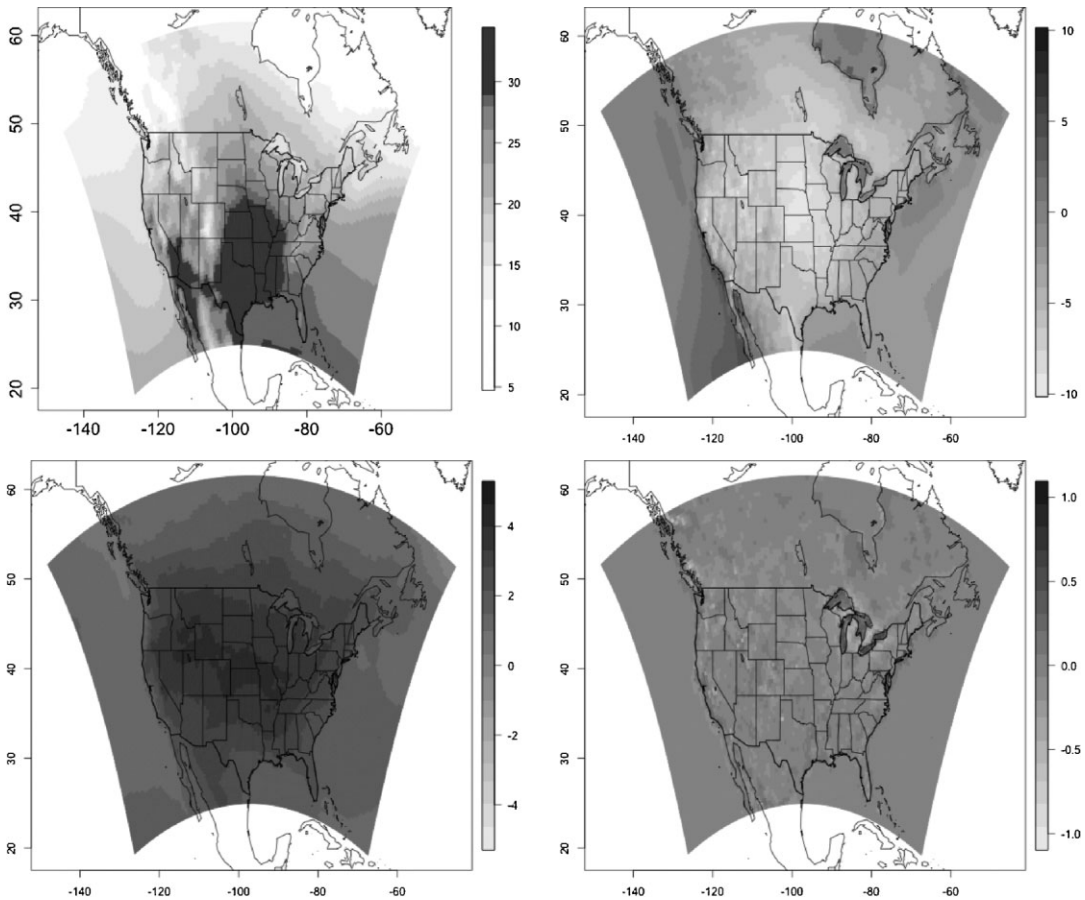

Figure 5. Posterior means of the $\boldsymbol{\alpha}$ components for summer temperature: $\boldsymbol{\alpha}_0$ (top left), $\boldsymbol{\alpha}_1$ (top right), $\boldsymbol{\alpha}_2$ (bottom left), and $\boldsymbol{\alpha}_3$ (bottom right). Units are in $^\circ C$

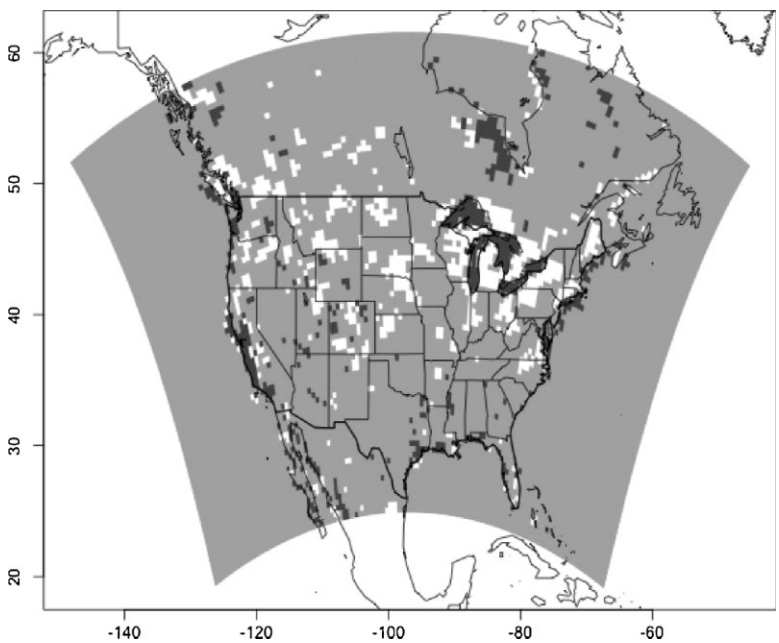

Figure 6. Grid boxes where the interaction term $\boldsymbol{\alpha}_3$ is significantly different from zero for summer temperature. White regions suggest a negative value for $\boldsymbol{\alpha}_3$ while dark grey regions suggest a positive value for $\boldsymbol{\alpha}_3$

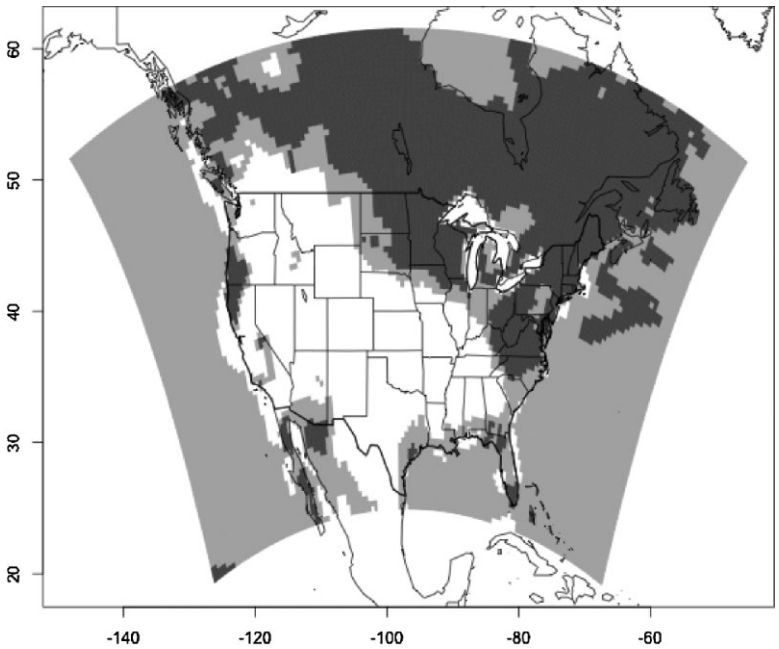

Figure 7. Grid boxes where $\boldsymbol{\alpha}_2$ is significantly different from $\mu_{\text{difference}}$ for summer temperature. Dark grey regions suggest $\boldsymbol{\alpha}_2 > \mu_{\text{difference}}$, while white regions suggest $\boldsymbol{\alpha}_2 < \mu_{\text{difference}}$

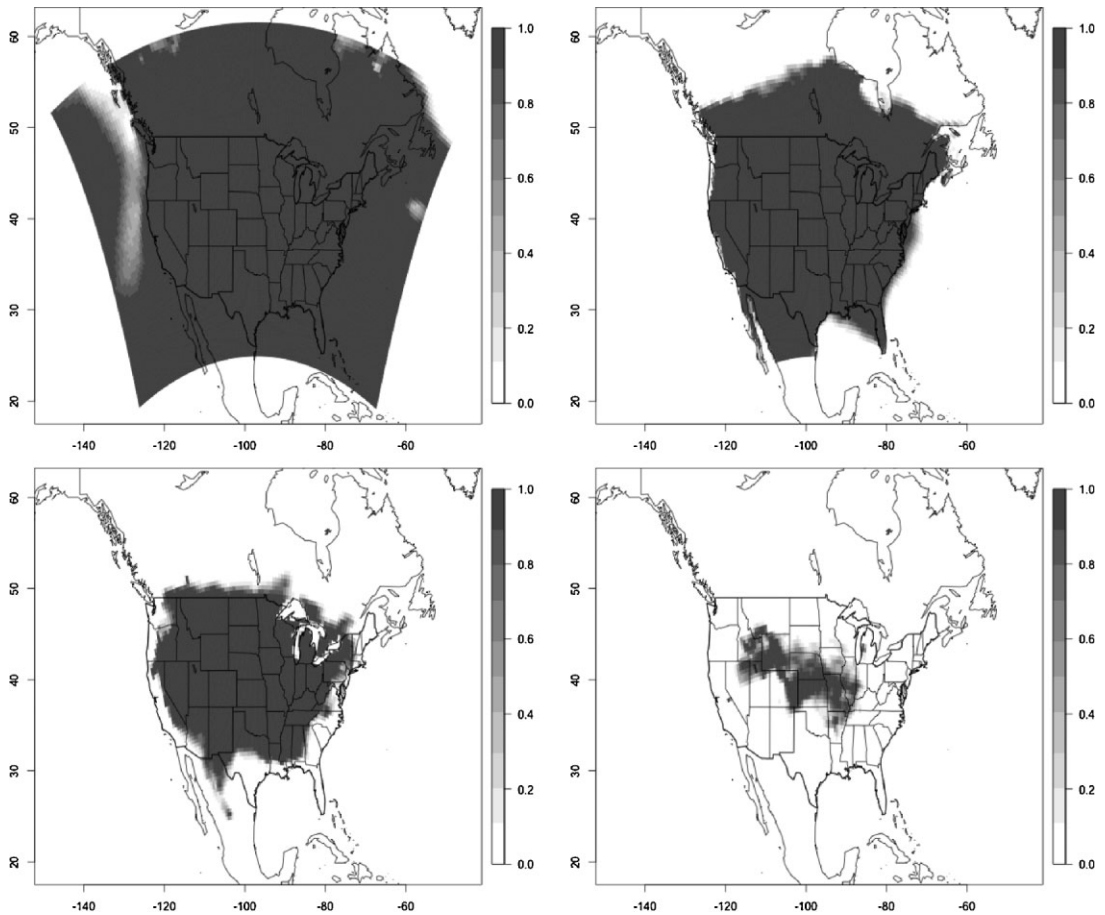

Figure 8. Estimated (pointwise) probabilities of increases in temperature of more than $1^\circ C$ (top left), $2^\circ C$ (top right), $3^\circ C$ (bottom left), and $4^\circ C$ (bottom right)

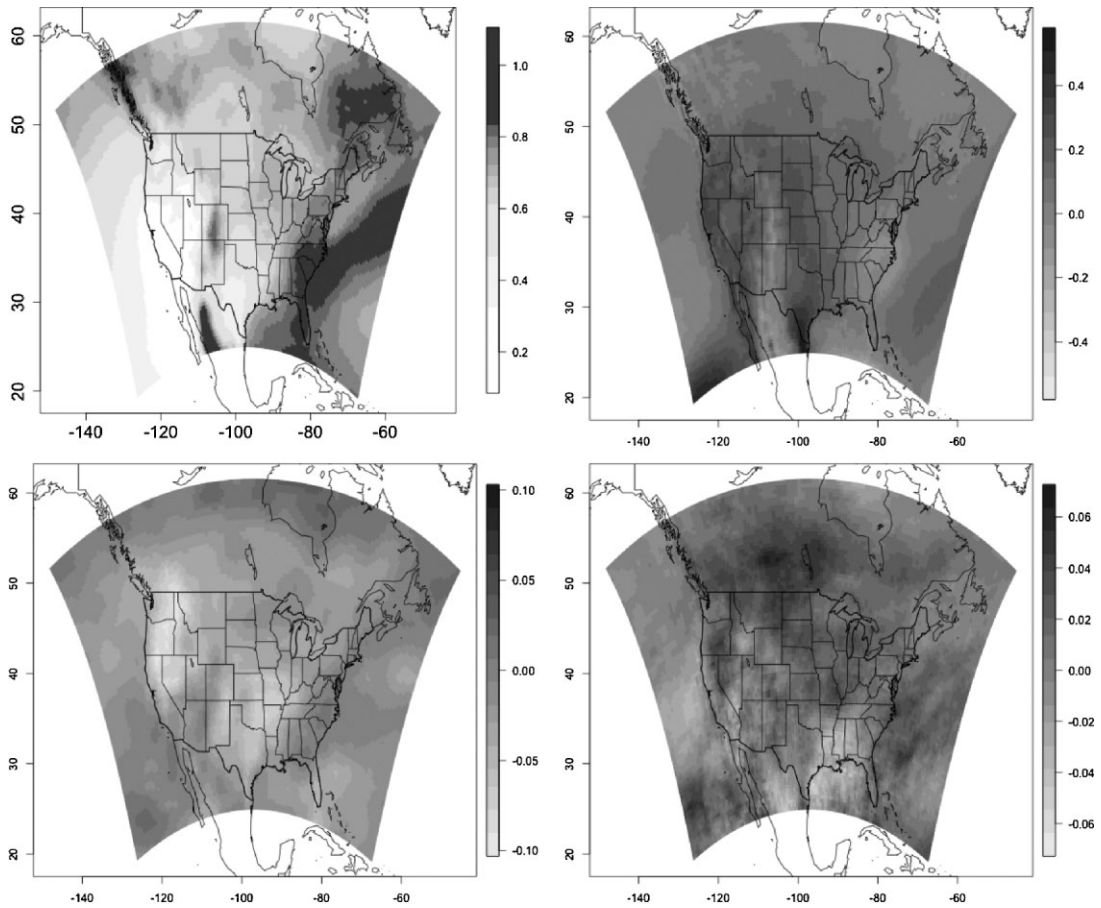

Figure 9. Posterior means of the $\boldsymbol{\alpha}$ components for summer precipitation: $\boldsymbol{\alpha}_0$ (top left), $\boldsymbol{\alpha}_1$ (top right), $\boldsymbol{\alpha}_2$ (bottom left), and $\boldsymbol{\alpha}_3$ (bottom right)

(i.e., pointwise posterior probabilities were greater than 0.95). Recall that the interaction suggests regions where the difference in the two models disagrees between the current and future run. The size of the interaction suggests that it does not significantly affect our inference for the main effects and, in particular, the common temperature change. Yet, the grid boxes where the interaction is different from zero are centered around coastlines, lakes, and mountain areas. It is speculated that this is, to some degree, related to the different ways in which the models incorporate information about the orography (e.g., mountains) across the domain. However, errors from the interpolation of the model output to a common grid could also play a role, although it is expected that these errors are small. This is currently still under investigation.

Given that the size of the interaction term $\boldsymbol{\alpha}_3$ is small, we can turn our attention to the main effects, and specifically, $\boldsymbol{\alpha}_2$, which reflects the common estimate of temperature change between the current and future runs. Figure 7 shows regions where $\boldsymbol{\alpha}_2$ is different from $\boldsymbol{\mu}_{\text{difference}}$. The figure suggests that the posterior for $\boldsymbol{\alpha}_2$ moves significantly away from the prior distribution based on the GFDL AOGCM and suggesting increased warming across the northern portion of the domain and along some of the coastlines and generally less warming across much of the western and southern United States.

The posterior mean of $\boldsymbol{\alpha}_2$ in Figure 5 suggests warming across the entire domain, but we can also say something about the uncertainty associated with that statement. For example, Figure 8 shows estimated probabilities of increases of greater than $1^\circ C$ (top left), $2^\circ C$ (top right), $3^\circ C$ (bottom left), and $4^\circ C$ (bottom right) computed for each gridbox. We see more serious warming stretching across the interior of the United States, in a band from eastern Idaho across Wyoming, northeastern Colorado, Nebraska, Kansas, and into Missouri, southern Illinois, and northern Arkansas.

4.2. Precipitation

The same model structure and MCMC approach were also used to examine summer precipitation. Figure 9 shows the posterior mean fields for the $\boldsymbol{\alpha}$ components. A similar situation arises in that the size of the model difference captured in $\boldsymbol{\alpha}_1$ is larger than either the difference between current and future runs or the interaction. However, the size of the interaction in $\boldsymbol{\alpha}_3$ is much larger than the interaction for temperature (relative to the other effects for precipitation). This suggests significant differences in the two models and that the nature of those differences change between current and future runs; see Figure 10.

Of course, this makes interpreting $\boldsymbol{\alpha}_2$ as a common measure of precipitation change problematic. Figure 11 shows a comparison of posterior means of $\boldsymbol{\alpha}_2$ (representing the change associated with the time slice) and $\boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_3$ (representing the change associated with the RCM). There

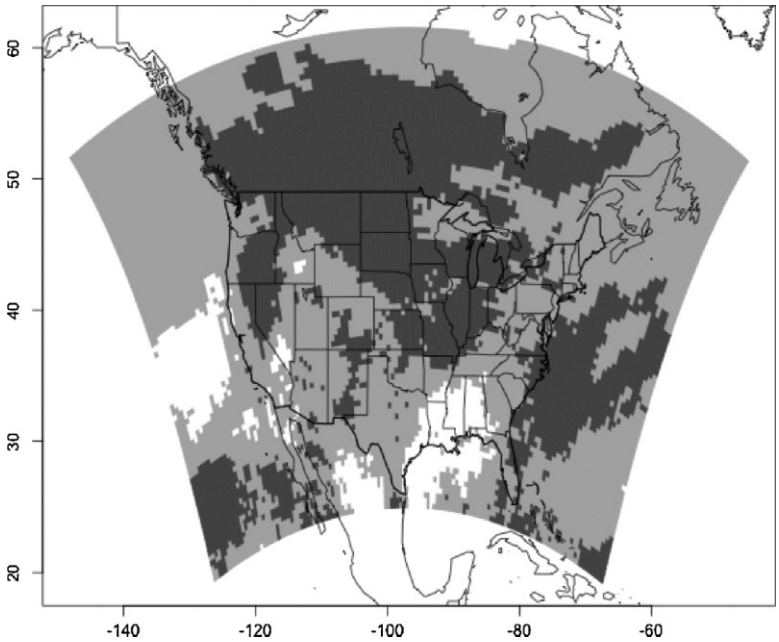

Figure 10. Grid boxes where the interaction term $\boldsymbol{\alpha}_3$ is significantly different from zero for summer precipitation. White regions suggest a negative value for $\boldsymbol{\alpha}_3$ while dark grey regions suggest a positive value for $\boldsymbol{\alpha}_3$

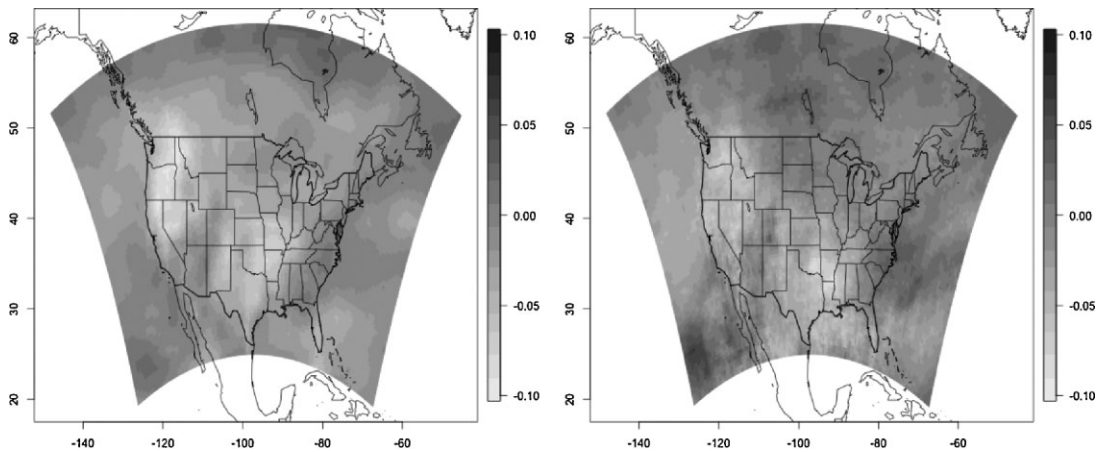

Figure 11. Posterior means for $\boldsymbol{\alpha}_2$ (left) and $\boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_3$ (right) for summer precipitation

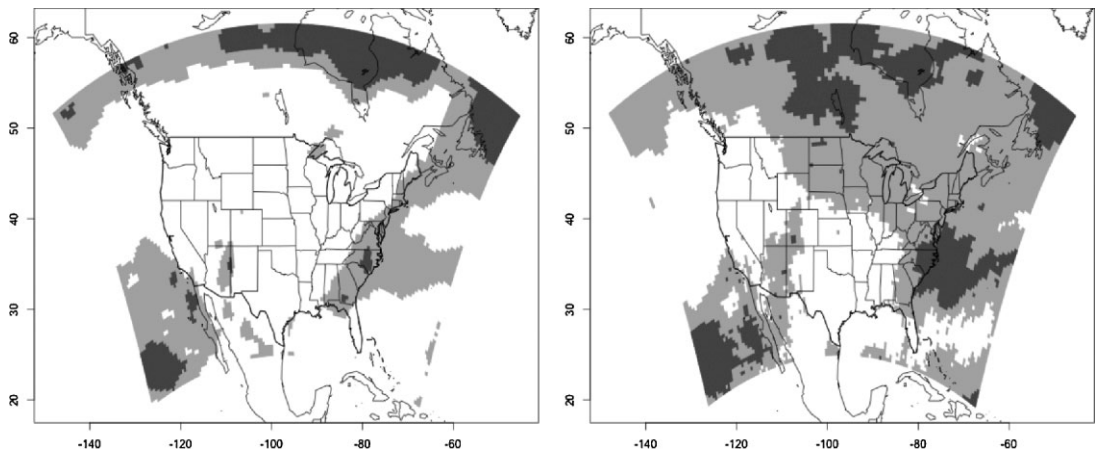

Figure 12. Significant probabilities of a change in precipitation based on $\boldsymbol{\alpha}_2$ (left) and $\boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_3$ (right) for summer precipitation. White regions suggest a significant decline while dark grey regions suggest a significant increase

are similarities in their spatial patterns, but the time slice tends to have higher decreases in precipitation in the northern portion of the domain. Figure 12 shows regions of significant probabilities of a decrease (white) or an increase (dark grey) in precipitation.

5. CONCLUDING REMARKS

This work has combined a FANOVA approach with spatial models to provide a rigorous statistical analysis of a subset of a large regional climate study. Besides providing specific results for downscaling the GFDL global climate simulations, we also consider this an example for summarizing other parts of the NARCCAP experiment.

We have found that overall summer temperatures are well described by the main effects of the downscaling method (time slice/RCM) and time period (current/future) while exhibiting a small interaction effect. The largest warming in this experiment that is distinct from the parent global climate model is in the central and Rocky Mountain region of the United States. The analysis also points to a more complex pattern of temperature response around the Great Lakes region, and the possible explanation is the artifact of how the lakes are discretized to the different downscaling grids. Summer precipitation, however, includes a large interaction between the two factors and the complex response is more difficult to interpret. These results are in line with the qualitative expectation that simulated climate changes with respect to temperature are more consistent than with respect to precipitation. For example, the IPCC Fourth Assessment Report finds that global models tend to produce more agreement among spatial patterns in warming and tend to have more disagreement among changing patterns of precipitation.

Given the large number of spatial locations in the (regridded) model output, being able to complete a Bayesian analysis using even an approximate statistical model is an accomplishment. The use of MRF and Kronecker structures in the spatial process models has made a statistical analysis feasible. Some extensions of this are to consider patterns of neighborhood dependence that produce a more isotropic covariance shape and to add an additional hidden layer that allows for the analysis to be done on the different, native grids of the model outputs. Including additional levels of the factors, such as additional downscaling methods or additional models, is a natural extension that requires some thought as to how these factors are handled. Kaufman and Sain (2010) utilize a class of prior distributions that impose an identifiability constraint, although a random effects assumption that does not need such constraints can also be imposed. The role of prior information, particularly prior distributions that are informed from the GCMs driving the regional models, will also become more important to these analysis with more model runs. Determining how exactly the information from the GCM will be incorporated is a subject of ongoing study, although recent research including Berliner *et al*. (2000) and Leith and Chandler (2010) may help determine a way forward. These issues, including improved handling of trends over time and the application of high-performance and massively parallel computing platforms, will be considered as we expand these analyses to a larger part of the NARCCAP experiment.

An important extension to these models is to consider a multivariate response where variables such as precipitation and temperature are considered jointly. This can done formally by augmenting the lattice to have links between different variables at coincident and neighboring locations (Sain and Cressie, 2007; Sain *et al*., 2010). Often a multivariate inference that includes both precipitation and temperature simultaneously is important for some agricultural and economic impacts of a changed climate.

Finally, an important qualification of this analysis is that it considers variation among model experiments but not between model results and actual climate. It is necessary to incorporate observational data to quantify model biases. The NARCCAP design also includes a part where the RCMs are forced by boundary conditions constructed from available data. These runs should reproduce the observed weather patterns and so can provide a measure of the RCMs' biases. Incorporating this observational information is important in the process of weighting different models and quantifying the value added by downscaling over the global climate simulations. Given the complexity of efforts such as NARCCAP we believe statistics will play an increasing role in highlighting key effects across numerous factors and also in designing future multi-model experiments.

ACKNOWLEDGEMENTS

This research was supported, in part, by National Science Foundation (NSF) grants ATM-0502977, ATM-0534173, and DMS-0707069. The authors thank the North American Regional Climate Change Assessment Program (NARCCAP) for providing the data used in this paper, and, in particular, Isaac Held, GFDL, for the time slice runs and Lisa Sloan, University of California, Santa Cruz, for the RCM runs. 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 National Science Foundation.

REFERENCES

Banerjee S, Carlin BP, Gelfand AE. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC Press: Bacon Raton, FL.

Berliner LM, Levine RA, Shea DJ. 2000. Bayesian climate change assessment. *Journal of Climate* **13**(21): 3805–3820. DOI:10.1175/1520-0442(2000)013<3805:BCCA>2.0.CO;2

Besag JE. 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B* **35**: 192–236.

Cressie NAC. 1993. *Statistics for Spatial Data* (revised edition). John Wiley & Sons Inc: New York.

Furrer R, Sain SR. 2010. spam: a sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *Journal of Statistical Software* **36**: 1–25.

Gelfand AE, Hills SE, Racine-Poon A, Smith AFM. 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* **85**: 972–985.
Gelfand AE, Smith AFM. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**: 398–409.
Geman S, Geman D. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.
Gilks WR, Richardson S, Spiegelhalter DJ. 1996. Introducing Markov chain Monte Carlo. In *Markov Chain Monte Carlo in Practice*. Chapman & Hall: London; 1–19.
Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97–109.
Kaufman CG, Sain SR. 2010. Bayesian functional ANOVA modeling using Gaussian process prior distributions. *Bayesian Analysis* **5**: 123–150.
Leith NA, Chandler RE. 2010. A framework for interpreting climate model outputs. *Journal of the Royal Statistical Society: Series C* **59**: 279–296. DOI:10.1111/j.1467-9876.2009.00694.x
Mardia KV, Kent JT, Bibby JM. 1979. *Multivariate Analysis*. Academic Press: London.
Mearns LO, Gutowski W, Jones R, Leung R, McGinnis S, Nunes A, Qian Y. 2009. A regional climate change assessment program for North America. *EOS Transactions* **90**: 311.
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**: 1087–1091.
Nakicenoniv N., Swart R. (eds). 2000. *Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge, UK.
Pal JS, Giorgi F, Bi X, Elguindi N, Solmon F, Gao X, Rauscher SA, Francisco R, Zakey A, Winter J, Ashfaq M, Syed FS, Bell JL, Diffenbaugh NS, Karmacharya J, Konar A, Martinez D, da Rocha RP, Sloan LC, Steiner AL. 2007. Regional climate modeling for the developing world: The ICTP RegCM3 and RegCNET. *Bulletin of the American Meteorological Society* **88**: 1395–1409.
Rue H, Held L. 2005. *Gaussian Markov Random Fields: Theory and Application*. Chapman & Hall/CRC Press: Boca Raton, FL.
Sain SR, Cressie N. 2007. A spatial model for multivariate lattice data. *Journal of Econometrics* **140**: 226–259.
Sain SR, Furrer R, Cressie N. 2010. Combining ensembles of regional climate model output via a multivariate Markov random field model. *Annals of Applied Statistics* (in press).
Sain SR, Kaufman C, Tebaldi C. 2009. A spatial analysis of a regional climate model experiment. Unpublished Manuscript. `//www.image.ucar.edu/~ssain/Steve_Sain_Webpage/Publications.html`
Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds). 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.
The GFDL Global Atmospheric Model Development Team. 2004. The new GFDL global atmosphere and land model AM2-LM2: evaluation with prescribed SST simulations. *Journal of Climate* **17**: 4641–4673.
Zhang Y, Hodges JS, Banerjee S. 2009. Smoothed ANOVA with spatial effects as a competitor to MCAR in multivariate spatial smoothing. *Annals of Applied Statistics* **3**: 1805–1830.

APPENDIX

In this Appendix, we give an overview of the Gaussian MRF specification used in this work. For more details, see Sain *et al*. (2009). MRF models are ideal for situations where the data lie on a spatial lattice, such as the regular grids found in climate models or remote sensing or the irregular grids associated with governmental or administrative boundaries. Conditional distributions based on neighborhoods give rise to the spatial dependence as opposed to more traditional geostatistical methods where spatial dependence is captured through functions of distance and possibly direction. The more influence neighboring values have on a particular location on the lattice, the stronger the spatial dependence is across the lattice. Rue and Held (2005) is an excellent general resource for MRF models as well as the overviews in Cressie (1993, Part II) or Banerjee *et al*. (2004, Chapter 3).

Besag (1974) laid out the basic structure, showing that, for random variables $Y_1, \ldots, Y_n$ at n locations on a spatial lattice, the conditional distributions $f(Y_i|Y_{-i}), i = 1, \ldots, n$, (where Y_{-i} refers to all random variables except the ith one) can be combined to form the join distribution $f(Y_1, \ldots, Y_n)$ as long as certain regularity conditions hold. In the Gaussian case, also referred to as the conditional autoregressive (CAR) model, the conditional distributions are specified through the mean and variance:

$$\mathrm{E}[Y_i|Y_{-i}] = \mu_i + \sum_{j=1}^{n} b_{ij}(Y_j - \mu_j) \qquad \text{and} \qquad \mathrm{var}[Y_i|Y_{-i}] = \tau_i^2$$

for $i = 1, \ldots, n$. These, in turn, lead to the joint distribution $\mathcal{N}(\boldsymbol{\mu}, (\mathbf{I} - \mathbf{B})^{-1}\mathbf{T})$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$, $\mathbf{I}$ is an $n \times n$ identity matrix, $\mathbf{B}$ is the $n \times n$ matrix with the i, jth element b_{ij}, and $\mathbf{T} = \mathrm{diag}(\tau_1^2, \ldots, \tau_n^2)'$. Of course, the spatial dependence parameters, $\{b_{ij}\}$, must be chosen to ensure that the resulting matrix, $(\mathbf{I} - \mathbf{B})^{-1}\mathbf{T}$, is a valid covariance matrix.

One of the drawbacks of MRF models is that, in general, the resulting covariance matrix is not stationary. However, following Sain *et al*. (2009) we use a stationary form that follows from the following specification for interior grid boxes:

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

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

where N_{1k} denotes the indices of the first-order neighbors (two grid boxes share an edge) and N_{2k} denotes the indices of the second-order neighbors (two grid boxes share a corner). Adjustments are made for grid boxes on the boundary, and the resulting covariance matrix is given by $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2\mathbf{R}(\phi)^{-1} \otimes \mathbf{C}(\phi)^{-1}$, where $\mathbf{R}(\phi)$ and $\mathbf{C}(\phi)$ are precision matrices specific to the rows and columns, respectively, of the grid. The

matrix $\mathbf{R}(\phi)$ is a symmetric, positive-definite matrix with the same number of rows and columns as the number of rows in the regular grid. Its form is given by

$$\mathbf{R}(\phi) = \begin{bmatrix} 1 & -\phi & & & & \\ -\phi & 1+\phi^2 & -\phi & & & \\ & \ddots & \ddots & \ddots & & \\ & & -\phi & 1+\phi^2 & -\phi & \\ & & & -\phi & 1 & \end{bmatrix} \tag{2}$$

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

There are a number of advantages to CAR models in general and specifically to this construction. First, the specification of CAR models defines the inverse of the covariance matrix so no matrix inversions are necessary in likelihood computations. Further, these precision matrices are sparse and significant computational gains are possible using sparse matrix methods (Furrer and Sain, 2010). Second, the Kronecker form of the covariance matrix also allows for significant computational savings by simply exploiting the properties of Kronecker products (see, for example, Mardia *et al*., 1979, Appendix A).

Of course, we should note that the resulting covariance does not exhibit the traditional spherical contours and is simply an approximation to more traditional forms. However, our focus is not on estimating covariance functions but on examining the differences in the distribution of the climate model output. This covariance allows us to balance computational considerations while still allowing for sharing across space. See Sain *et al*. (2009) for further discussion.

