Comparing and Blending Regional Climate Model Predictions for the American Southwest

Esther Salazar, Bruno Sansó, Andrew O. Finley, Dorit Hammerling, Ingelin Steinsland, Xia Wang and Paul Delamater *

Abstract

We consider the problem of forecasting future regional climate. Our method is based on blending different members of an ensemble of regional climate model (RCM) simulations while accounting for the discrepancies between these simulations, under present day conditions, and observational records for the recent past. To this end, we develop Bayesian space-time models that assess the discrepancies between climate model simulations and observational records. Those discrepancies are then propagated into the future to obtain blended forecasts of 21st century climate. The model allows for location-dependent spatial heterogeneities, providing local comparisons between the different simulations. Additionally, we estimate the different modes of spatial variability, and use the climate model-specific coefficients of the spatial factors for comparisons. We focus on regional climate model simulations performed in the context of the North American Regional Climate Change Assessment Program (NARCCAP). We consider, in particular, simulations from RegCM3 using three different forcings: NCEP, GFDL and CGCM3. We use simulations for two time periods: current climate conditions, covering 1971 to 2000, and future climate conditions under the SRES A2 emissions scenario, covering 2041 to 2070. We investigate yearly mean summer temperature for a domain in the South West of the United States. The results indicated the RCM simulations underestimate the mean summer temperature increase for most of the domain compared to our model.

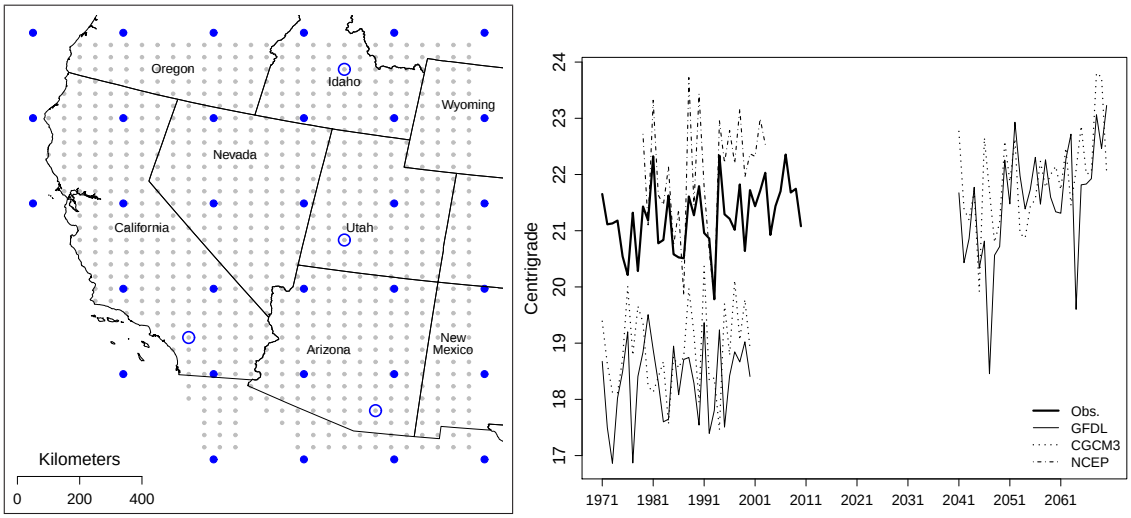
*Esther Salazar is Postdoctoral Associate, Department of Electrical and Computer Engineering, Duke University (esther.salazar@duke.edu); Bruno Sansó is Professor and Chair, Department of Applied Mathematics and Statistics, University of California Santa Cruz (bruno@ams.ucsc.edu); Andrew O. Finley is Assistant Professor, Departments of Forestry and Geography, Michigan State University (finleya@msu.edu); Dorit Hammerling is a PhD-Candidate, Department of Environmental Engineering, University of Michigan (doritmh@umich.edu); Ingelin Steinsland is Associate professor at Department of Mathematical Sciences, Norwegian University of Science and Technology, (ingelins@math.ntnu.no); Xia Wang is Postdoctoral Fellow, National Institute of Statistical Sciences, (xiawang@niss.org); Paul Delamater is a PhD candidate, Department of Geography, Michigan State University (delamate@msu.edu)

1 Introduction

Climate Models use the equations of motion to simulate changes in the global phenomena of the atmosphere. General Circulation Models (GCM) are used to describe large scale global motions and have the goal of capturing long term changes in the atmosphere. GCMs are fundamental tools to predict the evolution of climate in the future. They provide low resolution information, usually on scales of about 300 km. This is useful to quantify changes on global scales, but has limited use when the focus is on assessing the impact of future climate at regional scales. Regional Climate Models (RCM) (see, for example, Giorgi and Mearns, 1999) produce a "dynamic downscaling" of the output of GCMs. They use large scale GCM information as lateral and surface boundary conditions to simulate relatively short-term atmospheric and land-surface processes and the interactions between the two, at a spatial resolution of about 50 km. As with GCMs, RCMs simulations are subject to a number of uncertainties. Climate models use "parameterizations" to describe the physical processes that occur at resolutions that are below that of the discretization grid. Such parameterizations are one source of uncertainty. Variability in the initial conditions is also a source of uncertainty, as is variability between different GCM boundary conditions. Finally, there are structural differences in the models, due to different approaches to modeling relevant physical processes and implementing numerical solutions for the corresponding mathematical models. Rougier (2007) gives a clear account of the issues related to producing probabilistic quantifications of prediction uncertainties using climate model simulations. In this paper we focus on the variability created by running a given RCM with different GCM boundary conditions.

The idea of assessing the uncertainty in future climate predictions using multi-model ensembles is at the core of the World Climate Research Program (WCRP) Coupled Model Intercomparison Project Phase 3 (CMIP3) (Meehl et al., 2007). These simulations consist of output from 23 state of the art coupled Atmosphere-Ocean General Circulation Models (AOGCM) and were used as the base for the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC, 2007). Knutti et al. (2010) provides a good source of recent references on the challenges involved in combining different model simulations. Modern statistical methods, like the ones in Tebaldi and Sansó (2008); Smith et al. (2009); Rougier et al. (2010), use hierarchical models to obtain a distribution for the variable of interest using historical records as well as multi-model simulation ensembles. The approach considered in this paper follows along the same lines. The fundamental idea is that a climatological variable of interest, mean summer temperature in this paper, is simulated using different climate models. Such simulations are assumed to correspond to an unobserved underlying state plus a model dependent discrepancy. In correspondence to the simulations we have observational records. These are assumed to provide information about the same underlying state, but are subject to observational error. The combination of all the sources of information using a Bayesian hierarchical model produces a posterior distribution for the variable of interest. In addition, we obtain the distributions of the model discrepancies. These can be used for climate model assessment and comparison.

The North American Regional Climate Change Assessment Program (NARCCAP) (Mearns et al., 2009) is an international program to produce high resolution climate simulations to



(a) Domain of our data (b) Summer mean

Figure 1: (a) The fine grid corresponds to the resolution of the RCM output. The coarse grid has a resolution of 290 km and is used to build the predictive process. The circles correspond to locations used as examples in the results. (b) Time series of spatially averaged data.

assess climate change on a regional level. NARCCAP covers all of North America and part of the Caribbean with a resolution of 50km. It involves a number of combinations of six different RCMs and four different AOGCMs, in a way that each RCM is run with boundary conditions given by two different AOGCMs. The AOGCMs have been forced with the SRES A2 emissions scenario for the 21st century, thereby eliminating any differences in assumed emissions among the model runs. Simulations with these models were also produced for the current (historical) period. In addition, for the 20th Century, each RCM was forced with the NCEP reanalysis, which is a data assimilation product (NCEP reanalysis derived data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from `http://www.esrl.noaa.gov/psd/`). Further information about the project is available from `http://www.narccap.ucar.edu/index.html`.

The application presented in this paper focuses on mean summer temperature for an area in the Southwest of the United States. The domain is shown in Figure 1(a). We consider output simulated using RegCM3 (Pal et al., 2007), implemented by the Climate Change Research Group of the University of California Santa Cruz. In addition to the simulation obtained using NCEP forcings, we considered the ones obtained using the AOGCMs from the Geophysical Fluid Dynamics Laboratory (GFDL) (`http://nomads.gfdl.noaa.gov/CM2.X/`), and from the Canadian Center for Climate Modelling and Analysis (CGCM3) (`http://www.cccma.ec.gc.ca/data/cgcm3/cgcm3.shtml`).

Observational data for the current period, 1971–2010 Table 1, were derived from the National Oceanic and Atmospheric Administration (NOAA) daily weather station measurements and available via the National Climatic Data Center (NCDC). Given these data, prediction of daily mean temperature at RCM pixel centroids (fine gray grid in Figure 1(a))

Table 1: Data sources and available periods. Number of years in parenthesis

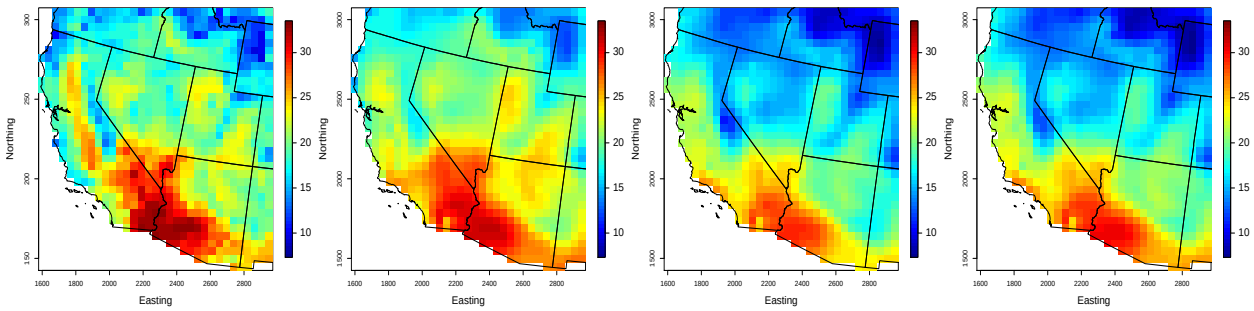
Source	Current scenario	Future scenario
Observations	1971–2010 (40)	–
GFDL	1971–2000 (30)	2041–2070 (30)
CGCM3	1971–2000 (30)	2041–2070 (30)
NCEP	1979–2003 (25)	–

was made using a spatial regression model that included an intercept, elevation as a covariate, spatial random effects, and a nugget. The spatial random effects followed a zero centered Gaussian process with an exponential spatial covariance function. Model parameters were estimated simultaneously using maximum likelihood. These daily means were then averaged to create estimates of *observed* mean summer temperature. As a result, our analysis uses four sources of information. These are summarized in Table 1.

Figure 1(b) shows the time series of the spatial averages of mean summer temperature over the region under study. Under present day conditions, we observe that the observational records and the NCEP based simulations are in reasonable agreement, while simulations based on GFDL and CGCM3 forcings show obvious discrepancies of about 2.5 °C. Additionally, we observe that RCM simulations show an increase of about 4 °C between the beginning of the 1970s and the end of the 2070s. Figure 2 shows the spatial distribution of the observations and the simulations averaged over the period 1979–2000, for which we have availability of all four sources of information. We observe that all panels have similar spatial patterns. Panels (c) and (d) are almost equivalent and they both show temperatures that are colder than those in panels (a) and (b) over most of the region. From the comparison of panel (a) with the other three panels, we observe that RCM simulations underestimate the temperature along the border of California and Arizona. They also underestimate the temperature in the California Central Valley, while they tend to overestimate the temperature along the Pacific coast.

The outputs from the RCM simulations forced by GFDL and CGCM3 are derived solely from the climatology of the models. NCEP, which is not a GCM, is "a reanalysis, a retrospective model of the atmosphere based on the observed data" (http://www.narccap.ucar.edu). Variables obtained from the NCEP driven runs are defined by observations to different degrees (Kalnay et al., 1996). The surface temperature, which is the variable we use in this study, is partially defined by the observations, but also strongly influenced by the climate model parameters. Thus, although the NCEP driven runs are close to the observations as shown in Figures 1(b) and 2, we see systematic deviations from the observations, just as in the other two RCM runs.

A fundamental problem regarding the comparison of observational records and climate model simulations is that the latter, even if indexed by a given year and location, are not meant to reproduce that year's observations. They are samples from the climate that is typical of that year, around the area of interest, as estimated by the climate model. To tackle this problem we propose statistical models for the discrepancies between observational



(a) Observations (b) NCEP (c) GFDL (d) CGCM3

Figure 2: (a) Temporal average of the observations; (b) Temporal average of the RCM simulations under NCEP forcings; (c) Temporal average of the RCM simulations under GFDL forcings; (d) Temporal average of the RCM simulations under CGM3 forcings. Period considered: 1979–2000.

records and simulations that are constant in time or vary smoothly in space and time. Thus, intuitively, our model provides the averaging that makes the comparison possible. Our statistical models and their implementation are described in detail in the next section. Section 3 presents the results of our analysis. We conclude with a discussion, presented in Section 4.

The present paper is driven by a climatological study. Nevertheless, in tackling the application, we have developed statistical methodology that is of general interest. In fact, we propose a space-time model that uses predictive processes as an effective device for dimension reduction, allows for space and time heterogeneities and provides inferences on the main modes of spatial variability of the random field under study. The decompositions developed in this paper provide effective ANOVA-like comparisons between different space-time random fields. The result is a probabilistic description of the quantities of interest obtained through a procedure that can handle several sources of information for large spatial domains.

2 Model, inference and evaluation

2.1 Proposed model

Let $y_t(\boldsymbol{s})$ be the observed temperatures at time t and site $\boldsymbol{s}$ and $y_{tj}^{CM}(\boldsymbol{s})$ be the RegCM3 output using GFDL ($j = 1$), CGCM3 ($j = 2$) and NCEP ($j = 3$) forcings. We propose a space-time model that assumes that temperature can be expressed as the sum of a baseline, a constant trend in time, a process explaining small scale spatial variability, and an observational error. The baseline is, possibly, a function of time and location dependent covariates. Climate model output follows a similar model, with the addition of a model and time dependent discrepancy term. The full specification of the model is given by the

equations:

$$\begin{aligned} y_t(\boldsymbol{s}) &= \boldsymbol{x}_t^T(\boldsymbol{s})\boldsymbol{\eta} + \xi t + \omega_t(\boldsymbol{s}) + \epsilon_t(\boldsymbol{s}), \quad t = 1, \ldots, t_0, \\ y_{jt}^{CM}(\boldsymbol{s}) &= \boldsymbol{x}_t^T(\boldsymbol{s})\boldsymbol{\eta} + \xi t + \omega_t(\boldsymbol{s}) + d_{jt}(\boldsymbol{s}) + \epsilon_{jt}(\boldsymbol{s}), \quad t = 1, \ldots, t_0, \ldots, T, \end{aligned}$$

where $\epsilon_t(\boldsymbol{s}) \sim N(0, \sigma^2)$, $\epsilon_{jt}(\boldsymbol{s}) \sim N(0, \sigma_j^2)$, for $j = 1, 2, 3$, correspond to observational errors. $\boldsymbol{x}_t(\boldsymbol{s})$ are the k-dimensional vectors of covariates, and $\boldsymbol{\eta}$ the corresponding coefficients. ξ is the slope of the long-term temperature trend. $\omega_t(\boldsymbol{s})$ is a Gaussian process that captures spatially correlated variability. $d_{jt}(\boldsymbol{s})$ are also Gaussian processes, and are associated with the discrepancy terms.

To specify $\omega_t(\boldsymbol{s})$ we use a modified predictive process as proposed in Finley et al. (2009a,b). We let $\omega_t(\boldsymbol{s}) = \tilde{\omega}_t(\boldsymbol{s}) + \tilde{\epsilon}_t(\boldsymbol{s})$, where, following the notation in Lemos and Sansó (2011),

$$\tilde{\omega}_t(\boldsymbol{s}) = \sum_{m=1}^{M} B_m(\boldsymbol{s})\gamma_{m,t} = \boldsymbol{B}(\boldsymbol{s})^T\boldsymbol{\gamma}_t, \quad \text{and} \quad \tilde{\epsilon}_t(\boldsymbol{s}) \overset{ind}{\sim} N(0, \tau^2 - \boldsymbol{v}(\boldsymbol{s})^T\boldsymbol{H}^{-1}\boldsymbol{v}(\boldsymbol{s})).$$

Thus, $\tilde{\omega}_t(\boldsymbol{s})$ is represented on a set of basis functions and $\tilde{\varepsilon}_t(\boldsymbol{s})$ are non-homogeneous independent random shocks. The basis functions are given as $B_m(\boldsymbol{s}) = [\boldsymbol{v}(\boldsymbol{s})^T\boldsymbol{H}^{-1}]_m$ where $\boldsymbol{v}(\boldsymbol{s}) = \tau^2(\rho(\boldsymbol{s}, \boldsymbol{s}_1^*; \boldsymbol{\phi}), \ldots, \rho(\boldsymbol{s}, \boldsymbol{s}_M^*; \boldsymbol{\phi}))$ and $H_{lk} = \tau^2\rho(\boldsymbol{s}_l^*, \boldsymbol{s}_k^*; \boldsymbol{\phi})$ for $l, k = 1, \ldots, M$. $\{\boldsymbol{s}_1^*, \ldots, \boldsymbol{s}_M^*\}$ represents a set of selected knots along the study area and $\rho(\cdot, \cdot)$ is a valid correlation function. In this paper we use the Matèrn with parameters $\boldsymbol{\phi} = (\phi_1, \phi_2)$ (see, for example, Banerjee et al., 2004). The evolution of $\boldsymbol{\gamma}_t$ is specified as $\boldsymbol{\gamma}_t \sim N(\varphi\boldsymbol{\gamma}_{t-1}, \boldsymbol{H})$, with $\varphi \in (-1, 1)$ and $\boldsymbol{\gamma}_0 \sim N(0, C^\gamma)$. Consider the eigenvalue/eigenvector decomposition of $\boldsymbol{H}$, then $\boldsymbol{H} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$, where $\boldsymbol{P}$ is an orthogonal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{H}$ in the diagonal. Let $\boldsymbol{\gamma}_t = \boldsymbol{P}\boldsymbol{\alpha}_t$, $\forall t$, then

$$\tilde{\omega}_t(\boldsymbol{s}) = \boldsymbol{B}(\boldsymbol{s})^T\boldsymbol{P}\boldsymbol{\alpha}_t = \boldsymbol{\psi}(\boldsymbol{s})^T\boldsymbol{\alpha}_t \text{ and } \boldsymbol{\alpha}_t \sim N(\varphi\boldsymbol{\alpha}_{t-1}, \boldsymbol{\Lambda}).$$

Given that $\boldsymbol{\Lambda}$ is a diagonal matrix, the elements of $\boldsymbol{\alpha}_t$ are independent. Assuming a decreasing order of the diagonal elements of $\boldsymbol{\Lambda}$, we have that $\psi_1(\boldsymbol{s})$ describes the main mode of spatial variability, $\psi_2(\boldsymbol{s})$ the second, and so on. Correspondingly, α_{1t} corresponds to the main factor coefficient, α_{2t} to the second, and so on. Notice that the decomposition $\tilde{\omega}_t(\boldsymbol{s}) = \boldsymbol{\psi}(\boldsymbol{s})^T\boldsymbol{\alpha}_t$ does not result in an orthogonal decomposition as the fields $\psi_j(\boldsymbol{s})$ $(j = 1, \ldots M)$ are not orthogonal.

Our spatial model for the discrepancies is also based on a predictive Gaussian process. In this case we do not add the additional random shock $\tilde{\varepsilon}_t(\boldsymbol{s})$, as we want to impose a high level of smoothness. For the j-th discrepancy we have

$$d_{jt}(\boldsymbol{s}) = \sum_{m=1}^{M} B_{jm}(\boldsymbol{s})\gamma_{jm,t} = \boldsymbol{B}(\boldsymbol{s})_j^T\boldsymbol{\gamma}_{jt} \ ,$$

where $B_{jm}(\boldsymbol{s}) = [\boldsymbol{v}_j(\boldsymbol{s})^T\boldsymbol{H}_j^{-1}]_m$, $\boldsymbol{\gamma}_{jt} \sim N(\varphi_j\boldsymbol{\gamma}_{j,t-1}, \boldsymbol{H}_j)$, $\boldsymbol{v}_j(\boldsymbol{s}) = \tau_j^2(\rho(\boldsymbol{s}, \boldsymbol{s}_1^*; \boldsymbol{\phi}), \ldots, \rho(\boldsymbol{s}, \boldsymbol{s}_M^*; \boldsymbol{\phi}))$ and $H_{j,lk} = \tau_j^2\rho(\boldsymbol{s}_l^*, \boldsymbol{s}_k^*; \boldsymbol{\phi})$. Note that $\boldsymbol{B}(\boldsymbol{s})_j = \boldsymbol{B}(\boldsymbol{s})$ and $\boldsymbol{H}_j = \boldsymbol{P}\boldsymbol{\Lambda}_j\boldsymbol{P}$. Then, we have that

$$\tilde{\omega}_t(\boldsymbol{s}) + d_{jt}(\boldsymbol{s}) = \boldsymbol{B}(\boldsymbol{s})^T(\boldsymbol{\gamma}_t + \boldsymbol{\gamma}_{jt}) = \boldsymbol{\psi}(\boldsymbol{s})^T(\boldsymbol{\alpha}_t + \boldsymbol{\alpha}_{jt}), \tag{1}$$

where $\boldsymbol{\alpha}_{jt} \sim N(\varphi_j \boldsymbol{\alpha}_{j,t-1}, \boldsymbol{\Lambda}_j)$. Equation (1) provides the basis for global comparisons of the different RCM simulations. By using a common set of bases $\boldsymbol{\psi}$ we can identify the differences between the model simulations by estimating $\boldsymbol{\alpha}_{jt}$. In our case this corresponds to a spatial ANOVA where the factors are the different forcings.

For computational purposes it is convenient to write the proposed model in vector notation as

$$\begin{aligned} \boldsymbol{Y}_t &= \boldsymbol{X}_t^* \eta^* + \boldsymbol{\Psi}\boldsymbol{\alpha}_t + \tilde{\boldsymbol{\epsilon}}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t \sim N(0, \sigma^2 I_N), \quad t = 1, \ldots, t_0, \qquad (2) \\ \boldsymbol{Y}_{jt}^{CM} &= \boldsymbol{X}_t^* \eta^* + \boldsymbol{\Psi}(\boldsymbol{\alpha}_t + \boldsymbol{\alpha}_{jt}) + \tilde{\boldsymbol{\epsilon}}_t + \boldsymbol{\epsilon}_{jt}, & \boldsymbol{\epsilon}_{jt} \sim N(0, \sigma_j^2 I_N), \quad t = 1, \ldots, T, \qquad (3) \end{aligned}$$

where $\boldsymbol{Y}_t = (y_t(\boldsymbol{s}_1), \ldots, y_t(\boldsymbol{s}_N))^T$ is a $N \times 1$ vector, $\boldsymbol{X}_t^*$ is the $N \times (k+1)$ matrix of covariates with last column given by $t\mathbf{1}_N$, $\eta^* = (\eta^T, \xi)^T$ is a $(k+1) \times 1$ vector and $\boldsymbol{\Psi} = (\boldsymbol{\psi}(\boldsymbol{s}_1), \ldots, \boldsymbol{\psi}(\boldsymbol{s}_N))^T$ is a $N \times M$ matrix. The time-varying coefficients evolve as $\boldsymbol{\alpha}_t \sim N(\varphi \boldsymbol{\alpha}_{t-1}, \boldsymbol{\Lambda})$ and $\boldsymbol{\alpha}_{jt} \sim N(\varphi_j \boldsymbol{\alpha}_{j,t-1}, \boldsymbol{\Lambda_j})$, with $\boldsymbol{\alpha}_0 \sim N(m_0, C_0)$ and $\boldsymbol{\alpha}_{j0} \sim N(m_0, C_{0j})$. Also $\eta^* \sim N(\mu_\eta, \Sigma_\eta)$ where Σ_η is a diagonal matrix so independence between covariates is assumed a priori, $\varphi, \varphi_j \sim Ntr_{(-1,1)}(\mu_\varphi, \sigma_\varphi)$, $\sigma^2 \sim IG(a, b)$, $\sigma_j^2 \sim IG(a_j, b_j)$, $\tau^2 \sim IG(c, d)$, $\tau_j^2 \sim IG(c_j, d_j)$, $\phi_1 \sim IG(2, h)$ where $h = \max(dist)/(-2\log(0.05))$. Here, the smoothness parameter ϕ_2, of the Matèrn correlation function, is fixed in at 1 as suggested by Whittle (1954).

A simplification of the model given by the Equations (2) and (3) is achieved by assuming that the discrepancies are constant in time. Thus $\boldsymbol{\alpha}_{jt} = \boldsymbol{\alpha}_j$ with the prior $\boldsymbol{\alpha}_j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Lambda_j})$. Consequently, the constant discrepancy $d_j(\boldsymbol{s}) = \boldsymbol{\Psi}(\boldsymbol{s})^T \boldsymbol{\alpha}_j$ is also based on the predictive Gaussian process. Equation (3) is then substituted by

$$\boldsymbol{Y}_{jt}^{CM} = \boldsymbol{X}_t^* \eta^* + \boldsymbol{\Psi}(\boldsymbol{\alpha}_t + \boldsymbol{\alpha}_j) + \tilde{\boldsymbol{\epsilon}}_t + \boldsymbol{\epsilon}_{jt}, \quad \boldsymbol{\epsilon}_{jt} \sim N(0, \sigma_j^2 I_N), \quad t = 1, \ldots, T. \qquad (4)$$

As commented in the introduction, climate model simulations and observational records are hard to compare on a year by year basis, thus, assuming that the discrepancies are not time-dependent can be an appealing possibility. In Section 3 we present the results of fitting and comparing the models obtained from Equations (3) and Equation (4) respectively.

2.2 Posterior inference and computational issues

To estimate the parameters in the model, specified by Equations (2) and (3), we use a purposely designed Markov chain Monte Carlo (MCMC) (see, for example, Gamerman and Lopes, 2006). First, we consider the marginalized likelihood by integrating over $\tilde{\boldsymbol{\epsilon}}_t$. Consequently, we have that $\boldsymbol{Y}_t \sim N(\boldsymbol{X}_t^* \eta^* + \boldsymbol{\Psi}\boldsymbol{\alpha}_t, \Sigma)$ and $\boldsymbol{Y}_{jt}^{CM} \sim N(\boldsymbol{X}_t^* \eta^* + \boldsymbol{\Psi}(\boldsymbol{\alpha}_t + \boldsymbol{\alpha}_{jt}), \Sigma_j)$ where $\Sigma = \text{diag}(\sigma^2 + \tau^2 - \boldsymbol{v}(\boldsymbol{s_i})^T \boldsymbol{H}^{-1} \boldsymbol{v}(\boldsymbol{s_i}))_{i=1}^N$ and $\Sigma_j = \text{diag}(\sigma_j^2 + \tau^2 - \boldsymbol{v}(\boldsymbol{s_i})^T \boldsymbol{H}^{-1} \boldsymbol{v}(\boldsymbol{s_i}))_{i=1}^N$. The parameters to be estimated are: η^*, σ^2, σ_j^2, the spatial parameters τ^2, τ_j^2, ϕ_1, the latent factors $\boldsymbol{\alpha}_t$, $\boldsymbol{\alpha}_{jt}$ and $\tilde{\boldsymbol{\epsilon}}_t$. Our MCMC scheme starts by first updating $\alpha_{m,t}$ and $\alpha_{jm,t}$ ($m = 1, \ldots, M$, $j = 1, 2, 3$, $t = 1, \ldots, T$) separately (Carlin et al., 1992). To avoid high autocorrelations within chains, we run the MCMC algorithm with a large number of iterations and thin out the chains. In addition, we also investigated estimating $\boldsymbol{\alpha}_{jt}$ and $\boldsymbol{\alpha}_t$ in blocks by using *forward filtering backward smoothing* (FFBS) (Frühwirth-Schnatter, 1994;

Carter and Kohn, 1994). That scheme requires multiplications and inversions of $4N \times 4N$ matrices. Inversions can be done efficiently by using the well known Sherman-Morrison-Woodbury matrix identity that only involves inversions of $M \times M$ matrices (see more details in Appendix). However, our experiments indicated that FFBS was still very time-consuming so we opted for the first scheme. We sample the other parameters using Gibbs sampling and Metropolis-Hastings steps. The derivation of the full conditional distributions is described in the Appendix. Finally, as discussed in Section 1, we have information available only for selected time periods. In particular, for GFDL and CGCM3 we do not have values for the period 2001–2040 and for NCEP we do not have values for the period 1971–1978. These values are considered missing values and, within the Bayesian setting they are handled as unobserved parameters. Samples of these values can be easily obtained from the predictive distribution $p(\boldsymbol{Y}_{jt}^{CM}|\boldsymbol{Y})$ (for $j = 1, 2,\ t = 31, \ldots, 70$ and $j = 3,\ t = 1, \ldots, 8$) using the MCMC algorithm.

2.3 Evaluation

The objective of our model evaluation is to rank all the available forecasters, including the three climate models and the two models we proposed, and decide which one most closely represents the predictive distribution of the true climate. Gneiting et al. (2007, 2008) propose a model diagnostic approach to evaluate predictive performance. The approach yields clear-cut ranking scoring functions for each of a number of competing forecast procedures. The method is based on empirical representations of the scoring rules, which can be easily calculated when using sample based inference (Gschlößl and Czado, 2005). Once the scores are estimated for each forecaster, they can be compared to perform model selection.

To evaluate locations individually (univariate forecasts) we use the *continuous rank probability score* (CRPS) defined as

$$crps(F, y) = \int_{-\infty}^{\infty} (F(\xi) - \mathtt{1}\{y \leq \xi\})^2 d\xi.$$

where F is the cumulative distribution function (cdf) of the predictive distribution for a given forecast model, and y the observed output.

CRPS is negative oriented, i.e. a lower values indicate better scores, and it is measured on the same scale as the observations (°C in our case). CRPS is also a valid measure for evaluating the performance of deterministic predictions. In this case it reduces to the absolute error between observations and predictions. Gneiting and Raftery (2007) show, for a given cdf F with finite first moment, $crps(F, y) = E_F(|Y - y|) - 0.5E_F|Y - Y'|$ where Y and Y' are independent random variables with common distribution F. Thus, CRPS can be conveniently calculated from ensembles (e.g. climate model output) or MCMC-samples using

$$\widehat{crps}(F_{sample}, y) = \frac{1}{m}\sum_{i=1}^{m} |y^{(i)} - y| - \frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} |y^{(i)} - y^{(j)}|,$$

where $y^{(1)}, \ldots, y^{(m)}$ are independent samples from the predictive cdf F_{sample}.

To evaluate how a model preforms for multiple locations simultaneously we follow Gneiting et al. (2008) and use the energy score, defined as

$$es(F, \boldsymbol{y}) = E_F\|\boldsymbol{Y} - \boldsymbol{y}\| - \frac{1}{2}E_F\|\boldsymbol{Y} - \boldsymbol{Y}'\| \ ,$$

where $\|\cdot\|$ denotes the Euclidean norm, and $\boldsymbol{Y}$ and $\boldsymbol{Y}'$ are independent random vectors from the multivariate predictive cdf F. The energy score is a generalization of the CRPS for multivariate forecasts, and has many of the same properties as CRPS. It is negative orientated, it reduces to the absolute error for deterministic forecasts and it can be calculated from ensembles or MCMC output using

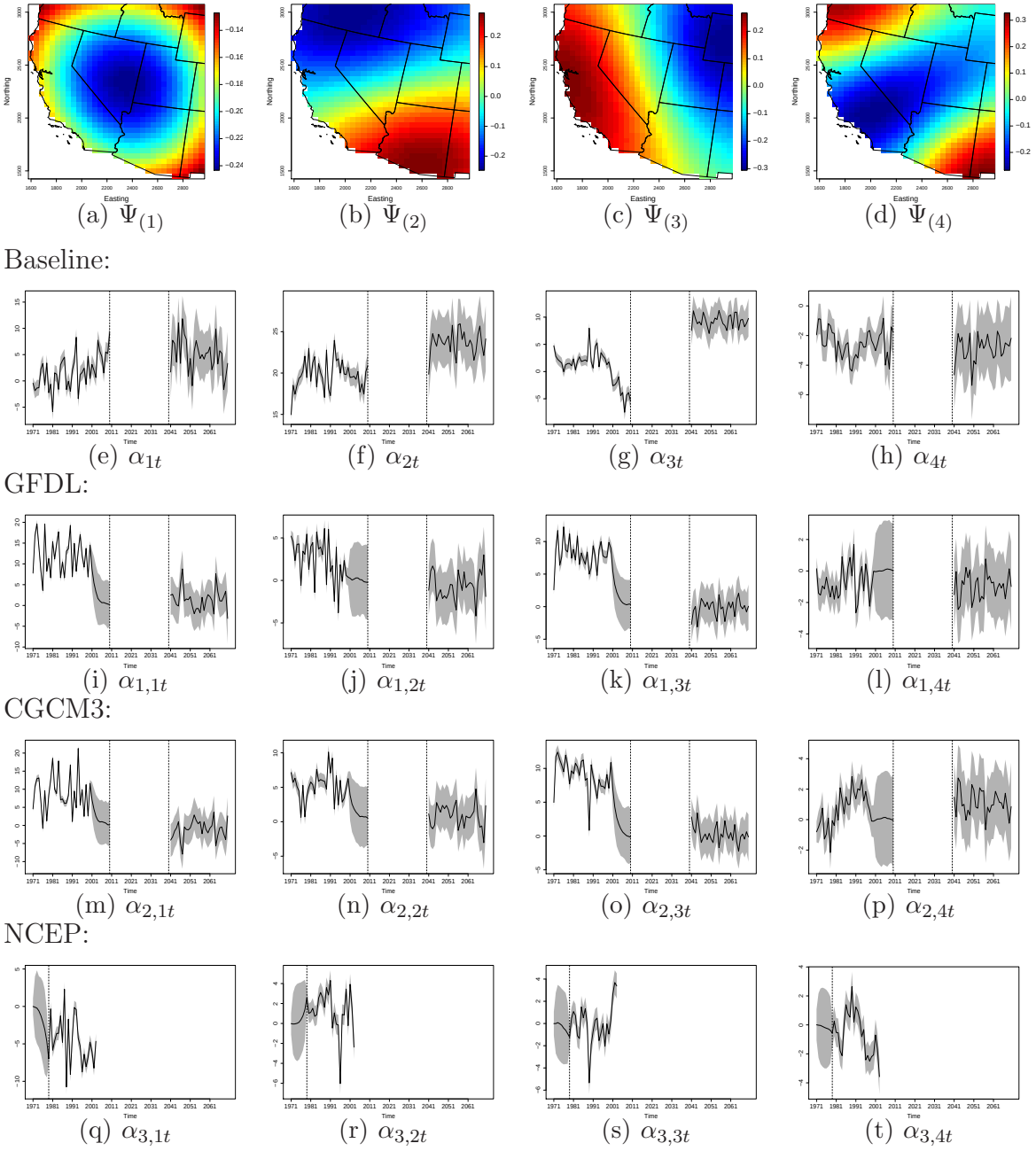
$$\widehat{es}(F_{sample}, \boldsymbol{y}) = \frac{1}{m}\sum_{i=1}^{m}\|\boldsymbol{y}^{(i)} - \boldsymbol{y}\| - \frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\|\boldsymbol{y}^{(i)} - \boldsymbol{y}^{(j)}\|$$

where $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(m)}$ are independent vectors sampled from the multivariate predictive cdf F_{sample}.

3 Results

Our results are based on keeping one sample for every five of a total of 50,000 iterations of the MCMC described in Section 2.2, after a burn-in of 10,000 iterations. We ran two parallel chains starting from different initial values of the parametric space. The convergence was checked via the Gelman-Rubin diagnostic. The predictive process uses the grid shown in Figure 1(a) with $M = 32$ knots and a resolution of 290 km. We consider the $N \times 3$ matrix of covariates $\boldsymbol{X}_t^* = (\boldsymbol{1}_N, elev, t\boldsymbol{1}_N)$ where $elev$ is the N-dimensional vector of elevations (in km) and $t = 1, \ldots, 100$. In addition, we consider the following priors for both models: $\eta^* \sim N((20, -5, 0.01)^T, \text{diag}(25, 25, 0.005))$; $\sigma^2, \sigma_j^2 \sim IG(2, 0.1)$; $\varphi, \varphi_j \sim Ntr_{(-1,1)}(0.5, 0.035)$; $\tau^2, \tau_j^2 \sim IG(2, 0.1)$; and $\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_{j0} \sim N(0\,\boldsymbol{1}_M, \boldsymbol{I}_M)$ $(j = 1, 2, 3)$.

We fitted two different models, a model with time-varying discrepancies in Equation (3), (Model 1) and a simplified model with constant discrepancies in Equation (4) (Model 2). Figure 3 presents some results for Model 1. Panels (a)-(d), shows the spatial factors corresponding to the four largest coefficients. As mentioned in Section 2 these factors are not orthogonal. Nevertheless we observe that they capture very distinct spatial features. The coefficients corresponding to Factor 1 are shown in panels (e), (i), (m) and (q). We observe that the width of the 95% probability intervals is a function of data availability. The baseline α_{1t} in (e) is strongly driven by the observational record. So predictions corresponding to 2041–2070 have pretty wide intervals. $\alpha_{3,1t}$, the coefficient of NCEP shown in (q), is not used for future climate projections. Also, NCEP forced simulations are available only for 1979–2003, so the first part of the time series has large probability bands. This is similar to what happens to $\alpha_{1,1t}$ and $\alpha_{2,1t}$, the coefficients of GFDL and CGCM3 respectively, in 2000–2010, as shown in (i) and (m). Panels (i), (m) and (q) provide a global comparison of the RegCM3 simulations under the different forcings, as a function of time. We observe



(a) $\Psi_{(1)}$ (b) $\Psi_{(2)}$ (c) $\Psi_{(3)}$ (d) $\Psi_{(4)}$

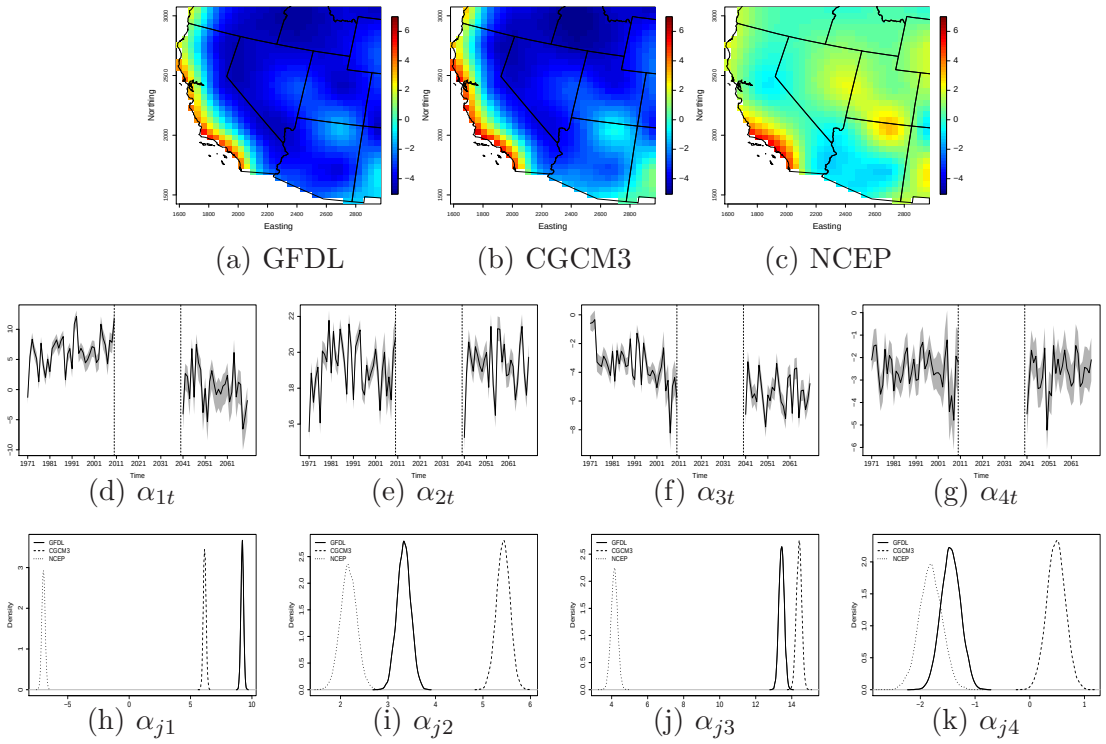
(e) α_{1t} (f) α_{2t} (g) α_{3t} (h) α_{4t}

(i) $\alpha_{1,1t}$ (j) $\alpha_{1,2t}$ (k) $\alpha_{1,3t}$ (l) $\alpha_{1,4t}$

(m) $\alpha_{2,1t}$ (n) $\alpha_{2,2t}$ (o) $\alpha_{2,3t}$ (p) $\alpha_{2,4t}$

(q) $\alpha_{3,1t}$ (r) $\alpha_{3,2t}$ (s) $\alpha_{3,3t}$ (t) $\alpha_{3,4t}$

Figure 3: Results for Model 1. (a)-(d) First four spatial factors. Time-varying coefficients corresponding to the first four factors for the (e)-(h) baseline, (i)-(l) GFDL, (m)-(p) CGCM3 and (q)-(t) NCEP. The gray shadows correspond to 95% probability intervals.



(a) GFDL (b) CGCM3 (c) NCEP

(d) α_{1t} (e) α_{2t} (f) α_{3t} (g) α_{4t}

(h) α_{j1} (i) α_{j2} (j) α_{j3} (k) α_{j4}

Figure 4: Results for Model 2. Constant discrepancies corresponding to (a) GFDL, (b) CGCM3 and (c) NCEP. (d)-(g) Time-varying coefficients corresponding to the first four factors for the baseline. The gray shadows correspond to 95% probability intervals. (h)-(k) Posterior densities for the constant coefficients α_{ji}, $j = 1, 2, 3$, $i = 1, \ldots, 4$ for the first four factors.

that NCEP based simulations have a very different behavior to the ones for the other two forcings. These, in turn, have some similarities.

Figure 4 presents some results from Model 2. The first four spatial factors are not shown, but they resemble closely the factors shown in Figure 3. Panels (a), (b) and (c) show the spatial fields of the discrepancies for the three different RCMs. Panels (a) and (b) have similar features and similar scales, while, not surprisingly, Panel (c), that corresponds to the NCEP forced RCM, shows a spatial distribution and a range of values that are very different to the ones in the panels corresponding to the other two forcings. The variability of the baseline is comparable to the one observed for Model 1, but the credibility intervals for the future years are much narrower than ones obtained for Model 1. This confirms the intuition that constant discrepancies produce less uncertain forecasts than time-varying discrepancies. Panels (h) through (k) show that, if the three different RCMs are assessed globally, they have significant differences. In fact, with the exception of some overlap between the posterior densities of α_{j4}, all the discrepancy coefficients are clearly separated. Further analysis of the results from Model 2 will be presented in Section 3.2.

3.1 Comparisons and assessments of models

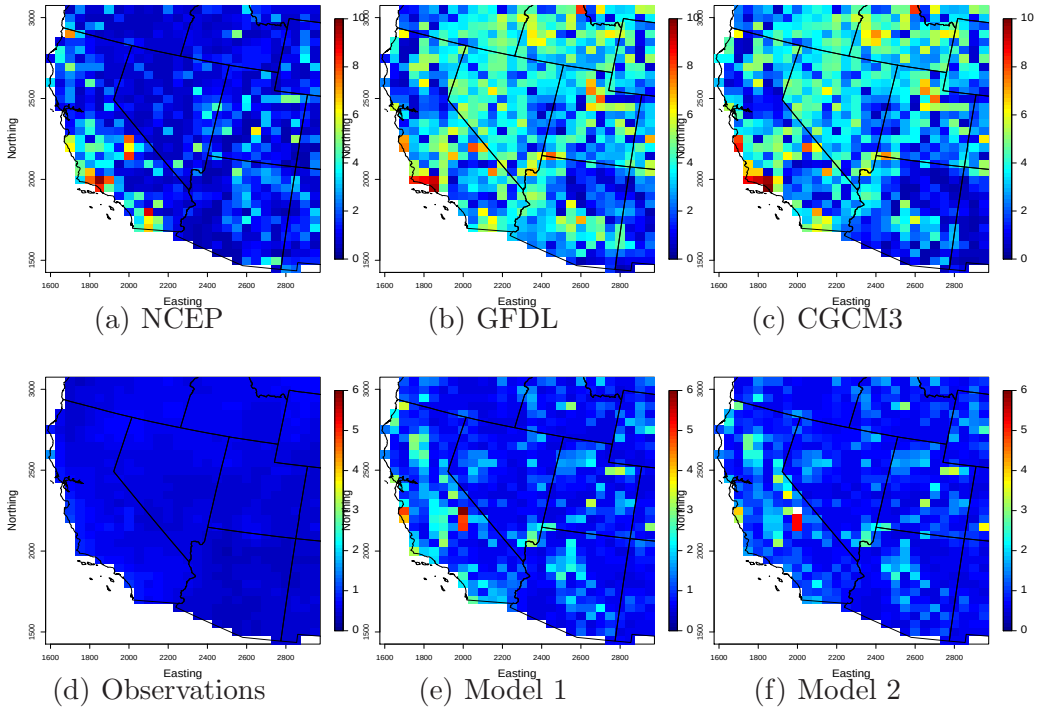
We use the scores presented in Section 2.3 to rank the three climate models as well as our two models. As commented in Section 2.3, the proposed scores will provide information on which forecast most closely represents the predictive distribution of the true climate. The observational records correspond to realizations of the climate. The empirical distribution based on the observations should therefore be a good approximation to the true climate. To establish a baseline of performance we calculated scores using the observed empirical pdf as predictive distribution. To evaluate the i-th year we used the empirical pdf based on all the other years.

Table 2 reports the CRPS, es, RMSE and MAE based on observations from 1971–2000. For the three climate models the empirical pdfs based on the output during year 1971-2000 are used as predictive distributions. For Model 1 and Model 2 their respective predictive posterior distributions are used. From Table 2 we find that all four evaluation criteria give the same rankings. The climate models with AOGCM forcings (CGCM3 and GFDL) have the poorest scores, while the climate model with NCEP forcing preforms better. This agrees with the exploratory analysis presented in the Figure 1, as considerable biases give poor predictive distributions. Our two models preform much better than any of the climate models alone, with Model 2 having the lowest scores of the two. As expected using empirical pdf based on observations gives the best scores.

In Figure 5 we present maps of CRPS for years 1971–2000 for each model. As expected, Panels (b) and (c), that correspond to the AOGCM forcings, are very similar. Similarly, Panels (e) and (f), corresponding to Models 1 and 2, are very similar between them, and they have some similarities with Panel (a), that corresponds to NCEP forcings.

Table 2: Mean continuous rank probability score (CRPS), mean energy score (es), root mean square error (RMSE) and mean absolute error (MAE) over 1971–2000 (for NCEP 1979–2000) . The es is evaluated for all locations simultaneously. For the other measures the means are taken over all locations and years. RMSE and MAE are based on ensemble mean for CGCM3, GFDL, NCEP and observations, and on posterior mean for Model 1 and Model 2.

Forecast	**CRPS**	**es**	**RMSE**	**MAE**
CGCM3	3.12	100.37	4.09	3.59
GFDL	2.85	95.31	3.91	3.31
NCEP	1.54	59.74	2.67	1.98
Observations	0.53	18.71	0.98	0.72
Model 1	1.01	31.13	1.79	1.36
Model 2	0.98	31.05	1.75	1.34



(a) NCEP (b) GFDL (c) CGCM3

(d) Observations (e) Model 1 (f) Model 2

Figure 5: Maps of mean CRPS for each locations over years 1971–2000 (for NCEP 1979–2000).

3.2 Future predictions from the best model

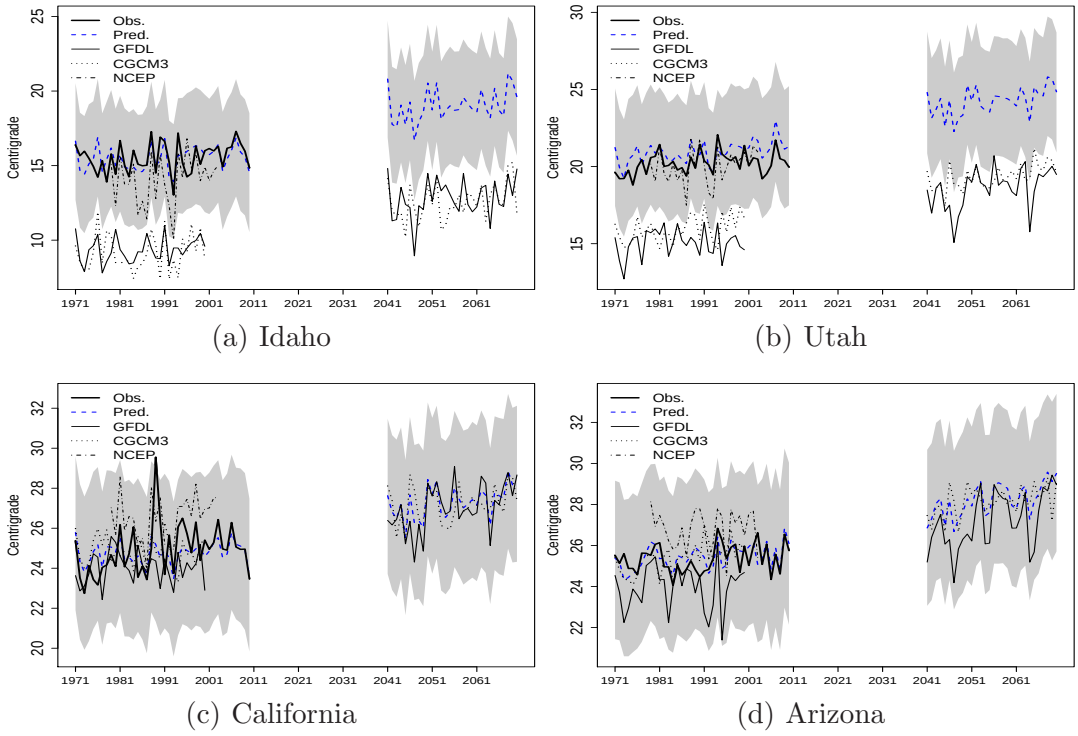
The results and the analysis presented in this section are based on the model with constant discrepancies. Table 3 provides a summary of the posterior distributions of the parameters in Model 2 that do not vary in time or space. A few of the more easily interpretable and physically relevant parameters are discussed in the following lines and their relevance is assessed using their posterior means. ξ is the coefficient associated with a constant trend in time describing the annual temperature increase in °C. So, for the 100-year time period (1971–2070) considered in this study, there is an overall predicted summer mean temperature increase of about 3 °C. η_2 is the coefficient of the elevation covariate in kilometers. Not surprisingly, this coefficient is negative indicating a temperature decrease of 0.23 °C for every hundred meters of elevation gain. ϕ_1 is the correlation length parameter used in the basis function representation of the spatial process. Given that the estimate of ϕ_1 is 378km, the correlation of two components of γ_t, which are 300km apart (about the resolution of the coarse grid used to built the predictive process), is 0.692. φ describes the year-to-year temporal autocorrelation of the spatial factor coefficient for the baseline, α_t. φ having posterior estimates close to 1 shows high temporal correlation and indicates that the process explaining small scale spatial variability is fairly consistent between years.

A detailed analysis of the resulting predictions across time can be obtained by focusing on specific locations in the modeled domain. The four locations selected as examples are indicated by circles in Figure 1(a). Figure 6 presents the time series of the predicted mean

Table 3: Posterior summaries for Model 2

Parameter	Mean	S.D.	2.5%Q	50%Q	97.5%Q
η_1	23.916	0.058	23.695	23.921	24.106
η_2	-2.302	0.006	-2.321	-2.302	-2.283
ξ	0.029	0.001	0.023	0.029	0.033
σ^2	3.251	0.009	3.190	3.252	3.311
σ_1^2	1.411	0.004	1.386	1.411	1.437
σ_2^2	1.440	0.004	1.415	1.439	1.466
σ_3^2	2.026	0.008	1.978	2.026	2.075
τ	4.930	0.032	4.724	4.928	5.136
τ_1	28.999	2.316	17.534	27.997	46.994
τ_2	34.247	2.716	20.887	33.027	54.463
τ_3	16.702	1.408	9.859	16.016	27.594
ϕ_1	377.84	4.644	377.38	377.70	379.15
φ	0.897	0.007	0.884	0.897	0.911

summer temperatures along with the corresponding time series of the RCM simulations and the observations for these four locations. The shadowed bands around the predictions correspond to 95% probability intervals. Their widths indicate the uncertainties around the predictions. On the whole, they span about 7 °C, which is rather large compared to the interannual variability of the mean summer temperatures. These four locations exemplify that the differences between our model and the RCM simulations are strongly dependent on location. For the locations in Idaho and Utah, the RCM simulations driven by GFDL and CGCM3 fall outside of the 95% probability intervals of our model for the past and future time periods. This is due to the large discrepancies between these RCM simulations and the observations at these locations. In our model the discrepancies are assumed constant in time leading to correspondingly large differences in future predictions. The top panel in Figure 4 shows the constant discrepancies for each RCM simulation and can be used as an indicator of where in the domain future predictions from our model will likely deviate from the RCM simulations. The location in Arizona serves as an example of one of the few areas where the GFDL and CGCM3 driven RCM simulations feature somewhat different discrepancies. On close observation of Figure 4(a) and (b), it can be noted that the discrepancy associated with the GFDL model is larger than the discrepancy associated with the CGCM3 model for the Arizona location. This becomes evident in Figure 6(d), where the GFDL driven simulations are on the lower side of model predictions while the CGCM3 driven simulations follow the model predictions more closely. The location in California, on the other hand, lies in area where the GFDL and CGCM3 driven RCM simulations both feature almost no discrepancies. This is visualized by the light blue band paralleling the California coastline in Figure 4(a) and (b). This lack of discrepancies can also be observed in the affinity of the RCM simulations to the future model prediction in Figure 6(c).



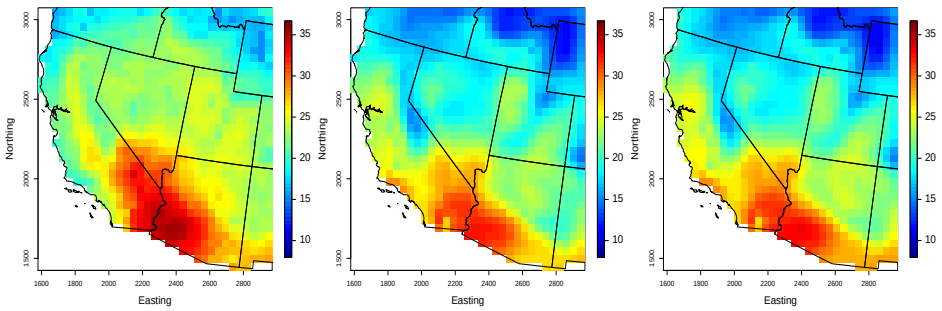
(a) Idaho (b) Utah

(c) California (d) Arizona

Figure 6: Predictive mean summer temperature values across time at some selected locations. The gray shadows correspond to 95% probability intervals.

Another way of comparing our model prediction and the RCM simulations for the future is by contrasting the predictions across space. Figure 7 shows the temporal averages over the future period for the model predictions and the RCM simulations. The mean summer temperatures are similar between the RCM simulations under GFDL and CGM3 forcings, but show marked difference between the RCM simulations and the model predictions. The model predictions are higher in most areas with the exception of the California coastline. As mentioned above, the differences between the RCM simulations and the model predictions stem from the discrepancies between the RCM simulations and the observations in the past time period. Figure 4(a) and (b) also illustrate how the California coastline features large positive discrepancies, whereas most of the remainder of the domain has large negative discrepancies.

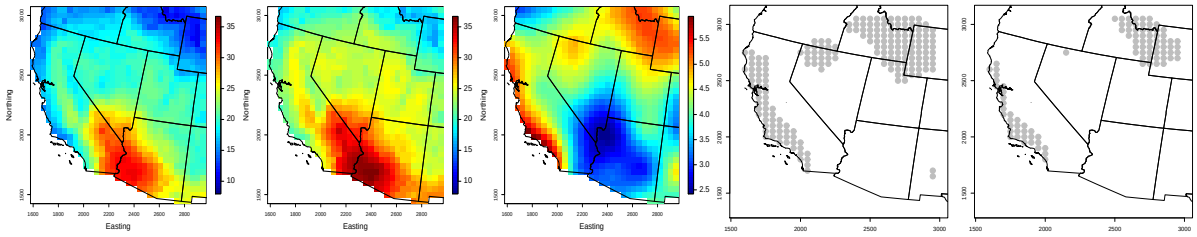
Figure 8(a) and (b) show the predictions of our model for 2010 and 2070, respectively; with Figure 8(c) showing the difference of the predictions between 2010 and 2070. It should be noted that the scale in Figure 8(c) ranges from 2.5 °C to 6 °C, so even areas that are dark blue are indicative of a temperature increase. Figure 8(d) and (e) show locations where the probability of a temperature increase of more than 3 °C is more than 85% and 90%, respectively. These locations include the California coastline and a region covering eastern Idaho and its border area with Wyoming and Montana. The sample-based approach adopted in this paper lends itself to probabilistic evaluations of blended predictions such as the ones



(a) Predictions (b) GFDL (c) CGM3

Figure 7: (a) Temporal average of the model predictions; (b) Temporal average of the RCM simulations under GFDL forcings; (c) Temporal average of the RCM simulations under CGM3 forcings. Period considered: 2041-2070.

shown in Figure 8(d) and (e). This is an attractive feature of the developed model and allows for probabilistic statements accounting for model uncertainties.



(a) Pred. 2010 (b) Pred. 2070 (c) Diff. 2010/2070 (d) Prob(Diff.>3°C) >85%) (e) Prob(Diff.>3°C) >90%

Figure 8: (a) Predictions for 2010. (b) Predictions for 2070. (c) Difference of predictions between 2010 and 2070. (d) Probability of exceeding 3°C difference more than 85%. (e) Probability of exceeding 3°C difference more than 90%. (c)

4 Discussion

We have proposed and developed a space-time model for the assessment of regional climate model predictions. Our model allows for the comparison of different members of a climate model ensemble. Comparisons can be done both at the local level in space and time, using space-time fields of relevant parameters. They can also be done at a global level, in functional ANOVA style, using a decomposition of the covariance structure induced by predictive Gaussian processes. The model also allows for blending of the different sources of information, to obtain unified predictions of future regional climate. The model relies on a factor analysis to reduce the dimension of the parameter space, a key element for computability.

The proposed model can be extended in a number of ways. While we have focused in the Southwest region of the USA, we anticipate that the model can be extended to the whole NARCCAP domain. Regarding the NARCCAP study, the most interesting extension would be to consider, not only different forcings for the same RCM, but also different RCMs. That is, to move from a one way ANOVA, to a two way ANOVA. Another interesting extension would be to consider, jointly, more than one atmospheric variable. This will shed light on how the associations between variables affect the discrepancies between RCM and observational records. Generally speaking, we believe that the proposed model can serve as a template for the analysis of physical quantities that vary in space and time, and for which several sources of information are available.

Acknowledgments

The present paper was developed as part of the Program on Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change that took place at the Statistical and Applied Mathematical Sciences Institute (SAMSI) during 2009–2010. We acknowledge the support given by SAMSI to the research presented here. We also acknowledge the involvement of Steve Sain, as the person who suggested the NARCCAP topic to our working group, and provided information and access to the data. Salazar was supported in part by National Science Foundation Grant DMS-06-35449. Sansó was partially funded by the National Science Foundation grant DMS-0906765.

A Full conditional distributions

The notation $\theta|\ldots$ denotes the full conditional distribution of θ given all other parameters.

- Sampling η^*: $\eta^*|\ldots \sim N(\mu^*_\eta, \Sigma^*_\eta)$ where $\Sigma^*_\eta = \left[\sum_{t=1}^T \boldsymbol{X}_t^{*T}\left(\Sigma^{-1}1_{t\leq t_0} + \sum_{j=1}^3 \Sigma_j^{-1}\right)\boldsymbol{X}_t^* + \Sigma_\eta^{-1}\right]^{-1}$, $\mu^*_\eta = \Sigma^*_\eta\left[\sum_{t=1}^T \boldsymbol{X}_t^T\left(\Sigma^{-1}\boldsymbol{Y}_t^*1_{t\leq t_0} + \sum_{j=1}^3 \Sigma_j^{-1}\boldsymbol{Y}_{jt}^*\right) + \Sigma_\eta^{-1}\mu_\eta\right]$ such that $\boldsymbol{Y}_t^* = \boldsymbol{Y}_t - \boldsymbol{\Psi}\boldsymbol{\alpha}_t$ and $\boldsymbol{Y}_{jt}^* = \boldsymbol{Y}_{jt}^{CM} - \boldsymbol{\Psi}(\boldsymbol{\alpha}_t + \boldsymbol{\alpha}_{jt})$.

- Sampling $\boldsymbol{\alpha}_t$: Let $\boldsymbol{Y}_t^{(l)} = \boldsymbol{Y}_t - \boldsymbol{X}_t^*\boldsymbol{\eta}^* - \sum_{m\neq l}\boldsymbol{\Psi}_{(m)}\alpha_{mt}$ for $t = 1,\ldots,t_0$ and $\boldsymbol{Y}_{jt}^{(l)} = \boldsymbol{Y}_{jt}^{CM} - \boldsymbol{X}_t^*\boldsymbol{\eta}^* - \boldsymbol{\Psi}\boldsymbol{\alpha}_{jt} - \sum_{m\neq l}\boldsymbol{\Psi}_{(m)}\alpha_{mt}$ for $t = 1,\ldots,T$. Then, each element of $\boldsymbol{\alpha}_t$ is sampled from $\alpha_{lt}|\ldots \sim N(m_{lt}, C_{lt})$, $l = 1,\ldots,M$, where

$$m_{lt} = \begin{cases} C_{lt}[\boldsymbol{\Psi}_{(l)}^T \boldsymbol{S}_t + \alpha_{l,t+1}\varphi\lambda_l^{-1} + m_{0,l}C_{0,l}^{-1}] & ,t=1 \\ C_{lt}[\boldsymbol{\Psi}_{(l)}^T \boldsymbol{S}_t + (\alpha_{l,t-1}+\alpha_{l,t+1})\varphi\lambda_l^{-1}] & ,t=2,\ldots,t_0 \\ C_{lt}[\boldsymbol{\Psi}_{(l)}^T \tilde{\boldsymbol{S}}_t + (\alpha_{l,t-1}+\alpha_{l,t+1})\varphi\lambda_l^{-1}] & ,t=t_0+1,\ldots,T-1 \\ C_{lt}[\boldsymbol{\Psi}_{(l)}^T \tilde{\boldsymbol{S}}_t + \alpha_{l,t-1}\varphi\lambda_l^{-1}] & ,t=T \end{cases}$$

$$C_{lt} = \begin{cases} [\boldsymbol{\Psi}_{(l)}^T(\Sigma^{-1}+\sum_{j=1}^3\Sigma_j^{-1})\boldsymbol{\Psi}_{(l)} + \varphi^2\lambda_l^{-1} + C_{0,l}^{-1}]^{-1} & ,t=1 \\ [\boldsymbol{\Psi}_{(l)}^T(\Sigma^{-1}+\sum_{j=1}^3\Sigma_j^{-1})\boldsymbol{\Psi}_{(l)} + (1+\varphi^2)\lambda_l^{-1}]^{-1} & ,t=2,\ldots,t_0 \\ [\boldsymbol{\Psi}_{(l)}^T(\sum_{j=1}^3\Sigma_j^{-1})\boldsymbol{\Psi}_{(l)} + (1+\varphi^2)\lambda_l^{-1}]^{-1} & ,t=t_0+1,\ldots,T-1 \\ [\boldsymbol{\Psi}_{(l)}^T(\sum_{j=1}^3\Sigma_j^{-1})\boldsymbol{\Psi}_{(l)} + \lambda_l^{-1}]^{-1} & ,t=T \end{cases}$$

where $\boldsymbol{S}_t = \Sigma^{-1}\boldsymbol{Y}_t^{(l)} + \sum_{j=1}^3 \Sigma_j^{-1}\boldsymbol{Y}_{jt}^{(l)}$ and $\tilde{\boldsymbol{S}}_t = \sum_{j=1}^3 \Sigma_j^{-1}\boldsymbol{Y}_{jt}^{(l)}$.

- Sampling $\boldsymbol{\alpha}_{jt}$: Let $\tilde{\boldsymbol{Y}}_{jt}^{(l)} = \boldsymbol{Y}_{jt}^{CM} - \boldsymbol{X}_t^*\boldsymbol{\eta}^* - \boldsymbol{\Psi}\boldsymbol{\alpha}_t - \sum_{m\neq l}\boldsymbol{\Psi}_{(m)}\alpha_{j,mt}$ for $t=1,\ldots,T$. Then, each element of $\boldsymbol{\alpha}_{jt}$ is sampled from $\alpha_{j,lt}|\ldots \sim N(m_{j,lt}, C_{j,lt})$, $l=1,\ldots,M$, where

$$m_{j,lt} = \begin{cases} C_{j,lt}[\boldsymbol{\Psi}_{(l)}^T\Sigma_j^{-1}\tilde{\boldsymbol{Y}}_t^{(l)} + \alpha_{j,l,t+1}\varphi_j\lambda_{j,l}^{-1} + m_{0j,l}C_{0j,l}^{-1}] & ,t=1 \\ C_{j,lt}[\boldsymbol{\Psi}_{(l)}^T\Sigma_j^{-1}\tilde{\boldsymbol{Y}}_t^{(l)} + (\alpha_{j,l,t-1}+\alpha_{j,l,t+1})\varphi_j\lambda_{j,l}^{-1}] & ,t=2,\ldots,T-1 \\ C_{lt}[\boldsymbol{\Psi}_{(l)}^T\Sigma_j^{-1}\tilde{\boldsymbol{Y}}_t^{(l)} + \alpha_{j,l,t-1}\varphi_j\lambda_{j,l}^{-1}] & ,t=T \end{cases}$$

$$C_{j,lt} = \begin{cases} [\boldsymbol{\Psi}_{(l)}^T\Sigma_j^{-1}\boldsymbol{\Psi}_{(l)} + \varphi_j^2\lambda_{j,l}^{-1} + C_{0j,l}^{-1}]^{-1} & ,t=1 \\ [\boldsymbol{\Psi}_{(l)}^T\Sigma_j^{-1}\boldsymbol{\Psi}_{(l)} + (1+\varphi_j^2)\lambda_{j,l}^{-1}]^{-1} & ,t=2,\ldots,T-1 \\ [\boldsymbol{\Psi}_{(l)}^T\Sigma_j^{-1}\boldsymbol{\Psi}_{(l)} + \lambda_{j,l}^{-1}]^{-1} & ,t=T \end{cases}$$

- Sampling $\boldsymbol{\alpha}_t$ and $\boldsymbol{\alpha}_{jt}$ via FFBS: For $t=1,\ldots,t_0$, let $\boldsymbol{Y}_t^{**} = \boldsymbol{Y}_t - \boldsymbol{X}_t^*\eta^*$, $\boldsymbol{Y}_{jt}^{**} = \boldsymbol{Y}_{jt}^{CM} - \boldsymbol{X}_t^*\eta^* - \boldsymbol{\Psi}\boldsymbol{\alpha}_{jt}$ and $\tilde{\boldsymbol{Y}}_t = (\boldsymbol{Y}_t^{**T}, \boldsymbol{Y}_{1t}^{**T}, \boldsymbol{Y}_{2t}^{**T})^T$. Then, we can rewrite the observation equation as $\tilde{\boldsymbol{Y}}_t = \boldsymbol{F}\boldsymbol{\alpha}_t + \tilde{\boldsymbol{\epsilon}}_t$ where $\tilde{\boldsymbol{\epsilon}}_t \sim N(0,\boldsymbol{\Omega})$, $\boldsymbol{F} = \mathbf{1}_3 \otimes \boldsymbol{\Psi}$ and $\boldsymbol{\Omega} = \text{diag}(\sigma^2, \sigma_1^2, \sigma_2^2) \otimes \boldsymbol{I}_N$. In order to sampled $\{\boldsymbol{\alpha}_t\}_{t=1}^{t_0}$ jointly, we use the well known FFBS scheme. In this scheme, we need to invert the huge $3N\times 3N$ matrix $\boldsymbol{F}(C_{t-1}+\boldsymbol{\Lambda})\boldsymbol{F}^T + \boldsymbol{\Omega}$. This operation can be done efficiently by using the well known Sherman-Morrison-Woodbury matrix identity that only requires the inversion of $M\times M$ matrices such that the inverse is given by $\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\boldsymbol{F}\left\{(C_{t-1}+\boldsymbol{\Lambda})^{-1} + \boldsymbol{F}^T\boldsymbol{\Omega}^{-1}\boldsymbol{F}\right\}^{-1}\boldsymbol{F}^T\boldsymbol{\Omega}^{-1}$. For $t=t_0+1,\ldots,T$, we use the FFBS scheme considering $\boldsymbol{Y}_t^{**} = \boldsymbol{\Psi}\boldsymbol{\alpha}_t + \boldsymbol{\epsilon}_t$ where $\boldsymbol{\epsilon}_t \sim N(0,\sigma^2 I_N)$ and initial prior $\boldsymbol{\alpha}_{t_0} \sim N(\boldsymbol{\alpha}_{t_0-1}, \boldsymbol{\Lambda})$.

  For $\boldsymbol{\alpha}_{jt}$, $t=1,\ldots,T$, let $\tilde{\boldsymbol{Y}}_{jt} = \boldsymbol{Y}_{jt}^{CM} - \boldsymbol{X}_t^*\eta^* - \boldsymbol{\Psi}\boldsymbol{\alpha}_t$. Now, we have that $\tilde{\boldsymbol{Y}}_{jt} = \boldsymbol{\Psi}\boldsymbol{\alpha}_{jt} + \boldsymbol{\epsilon}_t$ where $\boldsymbol{\epsilon}_{jt} \sim N(0,\sigma_j^2 I_N)$. Analogously to $\boldsymbol{\alpha}_t$, we use the FFBS scheme to sample $\{\boldsymbol{\alpha_{jt}}\}_{t=1}^T$ jointly.

- Sampling $\tilde{\boldsymbol{\epsilon}}_t$: Let $\hat{\boldsymbol{Y}}_t = \boldsymbol{Y}_t - \boldsymbol{X}_t^*\boldsymbol{\eta}^* - \boldsymbol{\Psi}\boldsymbol{\alpha}_t$ and $\hat{\boldsymbol{Y}}_{jt}^{CM} = \boldsymbol{Y}_t - \boldsymbol{X}_t^*\boldsymbol{\eta}^* - \boldsymbol{\Psi}(\boldsymbol{\alpha}_t + \boldsymbol{\alpha}_{jt})$ such that $\hat{\boldsymbol{Y}}_t \sim N(\tilde{\boldsymbol{\epsilon}}_t, \sigma^2\boldsymbol{I}_N)$ and $\hat{\boldsymbol{Y}}_{jt}^{CM} \sim N(\tilde{\boldsymbol{\epsilon}}_t, \sigma_j^2\boldsymbol{I}_N)$. Then, each component of $\tilde{\boldsymbol{\epsilon}}_t$ is

sampled from $\tilde{\epsilon}_{it}|\ldots \sim N(m^{\epsilon}_{it}, C^{\epsilon}_{it})$ where $m^{\epsilon}_{it} = C^{\epsilon}_{it}[\hat{y}_t(s_i)\sigma^{-2}1_{t\leq t_0} + \sum_{j=1}^{3}\hat{y}_{jt}(s_i)\sigma_j^{-2}]$ and $C^{\epsilon}_{it} = [\sigma^{-2}1_{t\leq t_0} + \sum_{j=1}^{3}\sigma_j^{-2} + (\tau^2 + v(s_i)^T H^{-1}v(s_i))^{-1}]^{-1}$ for $t = 1, \ldots, T$.

- Sampling σ^2 and σ_j^2: For σ^2 we have that $\sigma^2|\ldots \sim IG(a^*, b^*)$ where $a^* = t_0N/2 + a$ and $b^* = b + \sum_{t=1}^{t_0}(\boldsymbol{Y}_t^* - \boldsymbol{X}_t^*\eta^* - \tilde{\boldsymbol{\epsilon}}_t)^T(\boldsymbol{Y}_t^* - \boldsymbol{X}_t^*\eta^* - \tilde{\boldsymbol{\epsilon}}_t)/2$. Analogously for σ_j^2, considering T, a_j, b_j and $\boldsymbol{Y}_{jt}^*$ instead of t_0, a, b and $\boldsymbol{Y}_t^*$, respectively.

- Sampling τ_j^2: $\tau_j^2|\ldots \sim IG(c_j^*, d_j^*)$ where $c_j^* = (T-1)M/2 + c_j$ and $d_j^* = d_j + \sum_{t=2}^{T}(\boldsymbol{\gamma}_{jt} - \boldsymbol{\gamma}_{j,t-1})^T\boldsymbol{\rho}^{-1}(\boldsymbol{\phi})(\boldsymbol{\gamma}_{jt} - \boldsymbol{\gamma}_{j,t-1})/2$.

- Sampling φ and φ_j: $\varphi|\ldots \sim Ntr_{(-1,1)}(\mu_{\varphi}^*, \sigma_{\varphi}^*)$ where $\mu_{\varphi}^* = \sigma_{\varphi}^*(\sum_{t=2}^{T}\boldsymbol{\alpha}_{t-1}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{\alpha}_t + \mu_{\varphi}/\sigma_{\varphi})$ and $\sigma_{\varphi}^* = 1/(\sum_{t=2}^{T}\boldsymbol{\alpha}_{t-1}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{\alpha}_{t-1} + 1/\sigma_{\varphi})$. Analogously for φ_j, considering $\boldsymbol{\alpha}_{j,t}$ and $\boldsymbol{\Lambda}_j$ instead of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\Lambda}$, respectively.

- Draw τ^2 and ϕ_1 using Metropolis-Hastings steps. Specifically, for the i-current MCMC iteration, candidate draws are generated from a lognormal distribution such that $\pi(\tau^{2(i)} \rightarrow \tau^2) \sim \log N(\log(\tau^{2(i)}), \delta_{\tau})$ and $\pi({\phi_1}^{(i)} \rightarrow \phi_1) \sim \log N(\log({\phi_1}^{(i)}), \delta_{\phi_1})$.

For the case of $\boldsymbol{\alpha}_{jt} = \boldsymbol{\alpha}_j$, each element of $\boldsymbol{\alpha}_j$ is sampled from $\alpha_{j,l}|\ldots \sim N(m_{j,l}, C_{j,l})$, $l = 1, \ldots, M$, where $m_{j,l} = C_{j,l}(\boldsymbol{\Psi}_{(l)}^T\Sigma_j^{-1}\sum_{t=1}^{T}\tilde{\boldsymbol{Y}}_{jt}^{(l)} + \mu_j\lambda_{j,l}^{-1})$, $C_{j,l} = (T\boldsymbol{\Psi}_{(l)}^T\Sigma_j^{-1}\boldsymbol{\Psi}_{(l)} + \lambda_{j,l}^{-1})^{-1}$ and $\tilde{\boldsymbol{Y}}_{jt}^{(l)} = \boldsymbol{Y}_{jt}^{CM} - \boldsymbol{X}_t^*\boldsymbol{\eta}^* - \boldsymbol{\Psi}\boldsymbol{\alpha}_t - \sum_{m\neq l}\boldsymbol{\Psi}_{(m)}\alpha_{j,m}$. Regarding to $\boldsymbol{\mu}_j$, we consider a prior $N(\boldsymbol{e}, \boldsymbol{F})$ so $\boldsymbol{\mu}_j|\ldots \sim N(\boldsymbol{e}^*, \boldsymbol{F}^*)$ where $\boldsymbol{e}^* = \boldsymbol{F}^*(\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\alpha}_j + \boldsymbol{F}^{-1}\boldsymbol{e})$ and $\boldsymbol{F}^* = (\boldsymbol{\Lambda}_j^{-1} + \boldsymbol{F}^{-1})^{-1}$. Finally, τ_j^2 can be sampled from $\tau_j^2|\ldots \sim IG(c_j^*, d_j^*)$ where $c_j^* = M/2 + c_j$ and $d_j^* = d_j + (\boldsymbol{\gamma}_j - \boldsymbol{P}\boldsymbol{\mu}_j)^T\boldsymbol{\rho}^{-1}(\boldsymbol{\phi})(\boldsymbol{\gamma}_j - \boldsymbol{P}\boldsymbol{\mu}_j)/2$.

References

Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis of Spatial Data*. New York: Chapman and Hall.

Carlin, B. P., Polson, N. G., and Stoffer, D. (1992). "A Monte Carlo approach to nonnormal and nonlinear state-space modelling." *Journal of American Statistical Association*, 87, 493–500.

Carter, C. K. and Kohn, R. (1994). "On Gibbs sampling for state space models." *Biometrika*, 81, 541–553.

Finley, A., Banerjee, S., Waldman, P., and Ericsson, T. (2009a). "Hierarchical Spatial Modeling of Additive and Dominance Genetic Variance for Large Spatial Trial Datasets." *Biometrics*, 65, 441–451. DOI: 10.1111/j.1541-0420.2008.01115.x.

Finley, A. O., Sang, H., Banerjee, S., and Gelfand, A. E. (2009b). "Improving the Performance of Predictive Process Modeling for Large Datasets." *Computational Statistics and Data Analysis*, 53, 2873–2884.

Frühwirth-Schnatter, S. (1994). "Data augmentation and dynamic linear models." *Journal of Time Series Analysis*, 15, 2, 183–202.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*. 2nd ed. London, UK: Chapman and Hall.

Giorgi, F. and Mearns, L. O. (1999). "Introduction to special section: Regional climate modeling revisited." *Journal of Geophysical Research - Atmospheres*, 104 (D6), 6335–6352.

Gneiting, T., Balabdaoui, F., and Raftery, A. (2007). "Probabilistic forecasts, calibration and sharpness." *Journal of the Royal Statistical Society, B*, 69, 2, 243–268.

Gneiting, T. and Raftery, A. (2007). "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Society*, 102, 477, 359–378.

Gneiting, T., Stanberry, L., Grimit, E., Held, L., and Johnson, N. (2008). "Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds." *Test*, 17, 2, 211–235.

Gschlößl, S. and Czado, C. (2005). "Spatial modelling of claim frequency and claim size in insurance."

IPCC (2007). *Climate Change 2007 – The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC*. Solomon, S. et al. (eds.), Cambridge University Press. 996 pp.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D. (1996). "The NCEP/NCAR 40-year reanalysis project." *Bulletin of the American Meteorological Society*, 77, 437–471.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. (2010). "Challenges in Combining Projections from Multiple Climate Models." *Journal of Climate*, 23, 2739–2758. Doi: 10.1175/2009JCLI3361.1.

Lemos, R. and Sansó, B. (2011). "Conditionally linear models for non-homogeneous spatial random fields." *Statistical Methodology*. In press.

Mearns, L., Gutowski, W., Jones, R., Leung, L., McGinnis, S., Nunes, A., and Qian, Y. (2009). "A regional climate change assessment program for North America." *EOS*, 90, 311–312.

Meehl, G., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J., Stouffer, R., and Taylor, K. (2007). "The WCRP CMIP3 multimodel dataset: A new era in climate change research." *Bull. Amer. Meteor. Soc.*, 88, 1383–1394.

Pal, J. S., Giorgi, F., Bi, X. Q., Elguindi, N., Solmon, F., Gao, X. J., Rauscher, S. A., Francisco, R., Zakey, A., Winter, J., Ashfaq, M., Syed, F. S., Bell, J. L., Diffenbaugh, N. S., Karmacharya, J., Konare, A., Martinez, D., da Rocha, R. P., Sloan, L. C., and Steiner, A. L. (2007). "Regional climate modeling for the developing world - The ICTP RegCM3 and RegCNET." *Bulletin of the American Meteorological Society*, 88, 1395–.

Rougier, J. (2007). "Probabilistic inference for future climate using an ensemble of climate model evaluations." *Climatic Change*, 81, 247–264.

Rougier, J., Goldstein, M., and House, L. (2010). "Assessing climate uncertainty using evaluations of several different climate simulators." Tech. rep., Dept. of Mathematics, University of Bristol.

Smith, R., Tebaldi, C., Nychka, D., and Mearns, L. (2009). "Bayesian modeling of uncertainty in ensembles of climate models." *Journal of the American Statistical Association*, 97–116.

Tebaldi, C. and Sansó, B. (2008). "Joint Projections of Temperature and Precipitation Change from Multiple Climate Models: A Hierarchical Bayes Approach." *Journal of the Royal Statistical Society, A*, 172, 83–106.

Whittle, P. (1954). "On stationary processes in the plane." *Biometrika*, 41, 434–449.