Comparing and Blending Regional Climate Model Predictions for the American Southwest

Esther Salazar[1], Bruno Sansó[2], Andrew Finley[3], Dorit Hammerling (Presenter, doritmh@umich.edu)[4], Ingelin Steinsland[5], Xia Wang[6], Paul Delamater[3]

[1]Duke University, Durham, NC 27708 [2]University of California, Santa Cruz, CA 95064 [3]Michigan State University, East Lansing, MI 48824 [4]University of Michigan, Ann Arbor, MI 48109-2125
[5]Norwegian University of Science and Technology, Trondheim, Norway [6]National Institute of Statistical Sciences, Research Triangle Park, NC 27709

Abstract

We consider the challenge of forecasting future regional climate by blending different members of an ensemble of regional climate model (RCM) simulations while accounting for the discrepancies these simulations have shown with observational records for the recent past. To this end, we develop Bayesian space-time models that assess the discrepancies between different 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.

Methodology and Approach

The developed model is a Bayesian space-time model using predictive Gaussian processes to model spatial non-stationarity and to make computation feasible. Let $y_t(s)$ be the observed temperatures at time t and site s and $y_{tj}^{CM}(s)$ 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, 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:

$$y_t(s) = x_t^T(s)\eta + \xi(t - t_0) + \omega_t(s) + \varepsilon_t(s), \quad t = 1,\ldots,t_0,$$
$$y_{jt}^{CM}(s) = x_t^T(s)\eta + \xi(t - t_0) + \omega_t(s) + d_j(s) + \varepsilon_{jt}(s), \quad t = 1,\ldots,t_0,\ldots,T,$$

where $\varepsilon_t(s) \sim N(0,\sigma^2)$, $\varepsilon_{jt}(s) \sim N(0,\sigma_j^2)$, for $j = 1, 2, 3$, correspond to observational errors. $x_t(s)$ are the k-dimensional vectors of covariates, and η the corresponding coefficients. ξ is the slope of the long-term temperature trend. $\omega_t(s)$ is a Gaussian process that captures spatially correlated variability. $d_j(s)$ are also Gaussian processes, and are associated with the discrepancy terms. To specify $\omega_t(s)$ and $d_j(s)$ we use a modified predictive process and a predictive process as proposed in Finley et al. (2009a,b).

Omitting a detailed description of the specification of the modified predictive process (Salazar et al., 2011), the proposed model can be written as

$$Y_t = X_t^*\eta^* + \Psi\alpha_t + \widetilde{\varepsilon}_t + \varepsilon_t \qquad \varepsilon_t \sim N(0,\sigma^2 I_N), \quad t = 1,\ldots,t_0,$$
$$Y_{jt}^{CM} = X_t^*\eta^* + \Psi(\alpha_t + \alpha_j) + \widetilde{\varepsilon}_t + \varepsilon_t \qquad \varepsilon_{jt} \sim N(0,\sigma_j^2 I_N), \quad t = 1,\ldots,T,$$

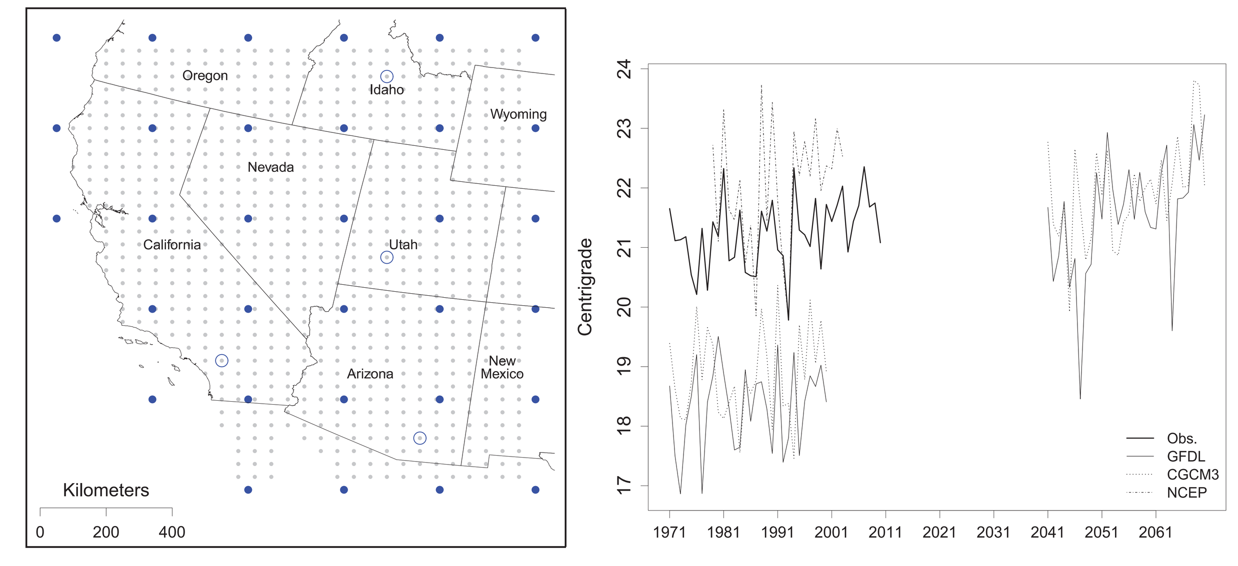
where $Y_t = (y_t(s_1), \ldots, y_t(s_N))^T$ is a $N \times 1$ vector, X_t^* is the $N \times (k+1)$ matrix of covariates with last column given by $(t - t_0)1_N$, $\eta^* = (\eta^T, \xi)^T$ is a $(k+1) \times 1$ vector and $\Psi = (\psi(s_1),\ldots,\psi(s_N))^T$ is an $N \times M$ matrix. The time-varying coefficient α_t evolves as $\alpha_t \sim N(\varphi\alpha_{t-1}, \mathrm{L})$ with $\alpha_0 \sim N(0, C^\alpha)$. The discrepancies α_j are assumed constant in time. The parameters are estimated in a Bayesian framework using Markov chain Monte Carlo methods.

Regional climate model and observational data

The application presented in this poster 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 regional climate model output simulated using RegCM3 (Pal et al., 2007) under NCEP, GFDL and CGCM3 forcings. In addition, we use observational data derived from weather station data. The hourly weather station data from 1971-2010 was drawn from the National Climatic Data Center database and used to predict mean daily temperature at the RCM pixel centroids. These predictions were generated using a spatial regression model that included elevation as a covariate, spatial random effects, and a nugget. Model parameters were estimated simultaneously using maximum likelihood. As a result, our analysis uses four sources of information. These are summarized in Table 1.

Table 1: Available data

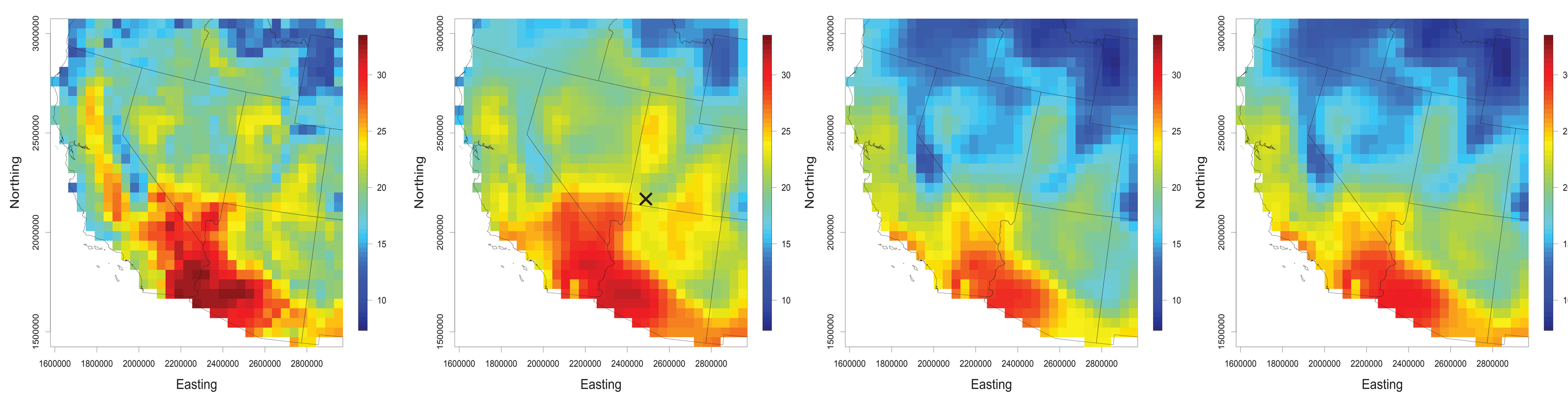
	Current scenario		Future scenario	
Source	Period	Years	Period	Years
Observations	1971–2010	40	–	–
GFDL	1971–2000	30	2041–2070	30
CGCM3	1971–2000	30	2041–2070	30
NCEP	1979–2003	25	–	–



(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 is used to build the predictive process. The circles correspond to the locations used as examples in the results. (b) Time series of spatially averaged data.

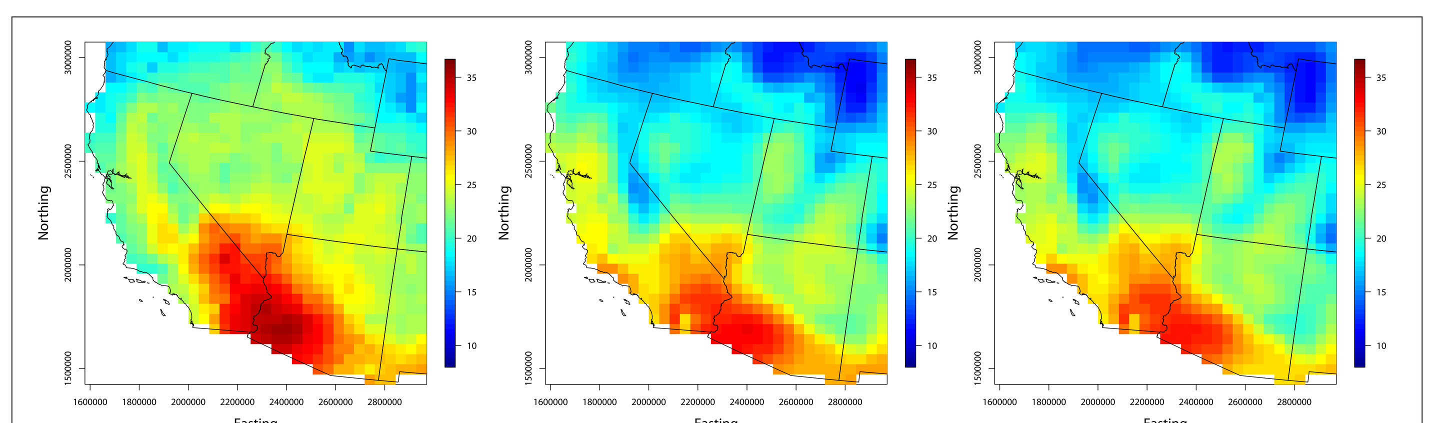
Figure 2: (a) Temporal average of the observations; (b) Temporal average of the RCM simulations under NCEP forcings; (c) Temporal average of the RCM simulations under GFDL forcings; (d) Temporal average of the RCM simulations under CGM3 forcings. Period considered: 1979-2000.



(a) Observations (b) RegCM3 « NCEP (b) RegCM3 « GFDL (b) RegCM3 « CGM3

Predictions across space

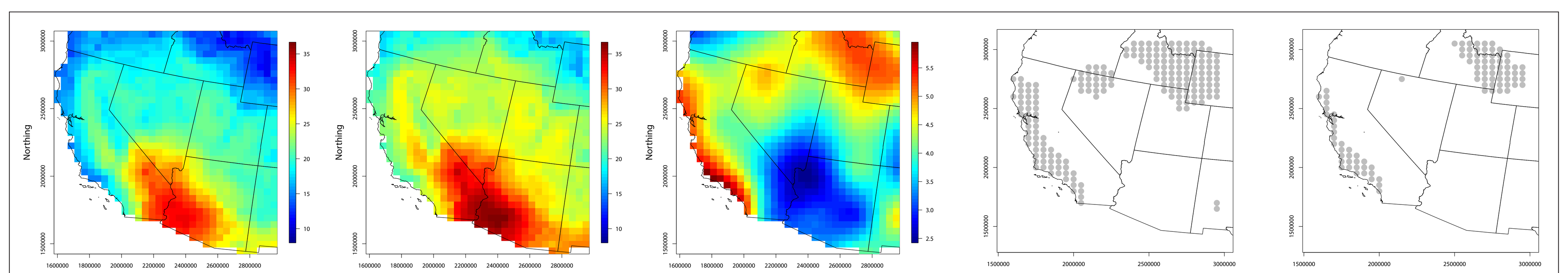
The summer mean temperatures averaged over 2041 to 2070 are similar between the RCM simulations under GFDL and CGM3 forcings, but show marked difference between the RCM simulations and the model predictions. The model predictions are higher in most areas with the exception of the California coastline. The differences between the RCM simulations and the model predictions stem from the discrepancies between the RCM simulations and the observations. These discrepancies are location-dependent and are propagated into the future.



(a) Predictions (b) RegCM3 « GFDL (c) RegCM3 « CGM3

Figure 4: (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.

Figure 5 indicates locations where large temperature increases are highly probable, namely the California coast line and a region covering eastern Idaho into Wyoming. These are areas where the RCMs have comparatively similar or even higher temperatures than observations for past conditions.

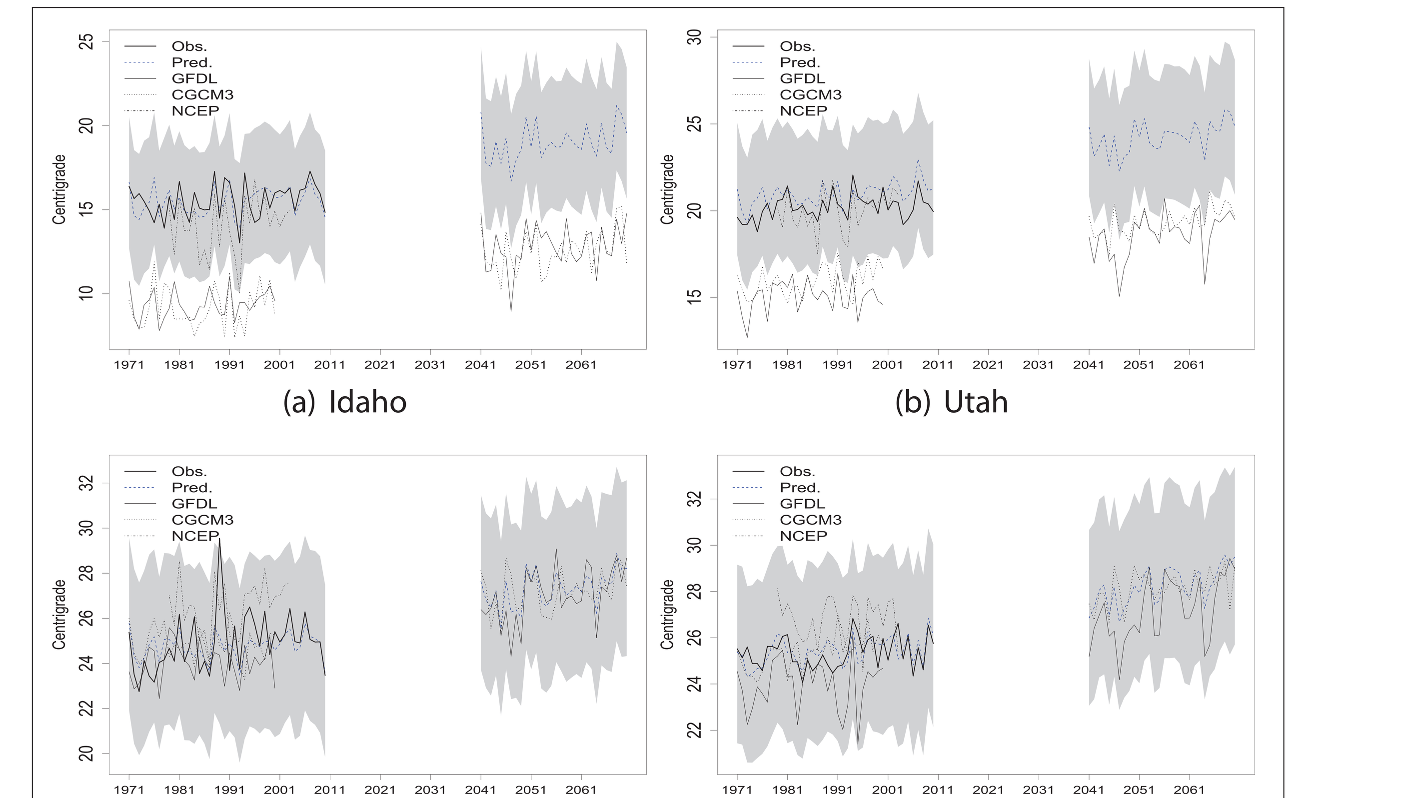


(a) Pred. 2010 (b) Pred. 2070 (c) Diff. 2010/2070 (d) Prob(increase>3°C) >85% (e) Prob(increase>3°C) > 90%

Figure 5: (a) Predictions for 2010 (b) Predictions for 2070 (c) Difference of predictions between 2010 and 2070. (d) Probability of exceeding a 3°C increase more than 85%. (e) Probability of exceeding 3°C increase more than 90%.

Predictions across time

A detailed analysis of the resulting predictions across time can be obtained by focusing on specific locations. The chosen locations are indicated by circles in Figure 1. Figure 3 presents the predictions from our model along with all sources of information used in the model. The shadowed areas correspond to 95% probability intervals. The uncertainty around the predictions is rather large spanning about 7°C. It can be noted that the differences between the RCM predictions driven by the GFDL and CGCM3 models and the predictions of our model are strongly dependent on location. For the locations in Idaho and Utah, the RCM predictions driven by the GFDL and CGCM3 models fall outside the 95% probability intervals of our model predictions for both the current and future time periods. This is due to the large discrepancies between the RCM predictions and the observed values at these locations. Our model propagates these discrepancies into the future, so they are also accounted for in the future predictions. The location-dependency of the discrepancies visible in Figure 3 indicates the necessity to allow for spatially varying discrepancy terms in the model formulation.



(a) Idaho (b) Utah (c) California (d) Arizona

Figure 3: Predictive values across time at four selected locations. The gray shadows correspond to 95% probability intervals.

References

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

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.

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.

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. 12

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.

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.

Mearns, L.O., et al., 2007, updated 2011. The North American Regional Climate Change Assessment Program dataset, National Center for Atmospheric Research Earth System Grid data portal, Boulder, CO. Data downloaded <insert date here>.

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 ICTPRegCM3 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.

Salazar, E., Sansó, B., Finley, A., Hammerling, D., Steinsland, I., Wang, X., and Delamater, P. (2011). "Comparing and Blending Regional Climate Model Predictions for the American Southwest." (in prep.).

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.