Validation of Climate Model Output using Bayesian Statistical Methods

Mark A. Snyder*, Bruno Sansó† and Lisa C. Sloan‡

Abstract

The growing interest in and emphasis on high spatial resolution estimates of future climate has demonstrated the need to apply regional climate models (RCMs) to that problem. As a consequence, the need for validation of these models, an assessment of how well an RCM reproduces a known climate, has also grown. Validation is often performed by comparing RCM output to gridded climate datasets and/or station data. The primary disadvantage of using gridded climate datasets is that the spatial resolution is almost always different and generally coarser than climate model output. We have used a Bayesian statistical model derived from observational data to validate RCM output. We used surface air temperature (SAT) data from 109 observational stations in California, all with records of approximately 50 years in length, and created a statistical model based on this data. The statistical model takes into account the elevation of the station, distance from coastline, and the NOAA climate region in which the station resides. Analysis indicates that the statistical model provides reliable estimates of the mean monthly SAT at any given station. In our method, the uncertainty in the estimates produced by the statistical model are directly determined by obtaining probability density functions for predicted SATs. This statistical model is then used to estimate average SATs corresponding to each of the climate model gridcells. These estimates are compared to the output of the RCM to assess how well the RCM matches the observed climate as defined by the statistical model. Overall, the match between the RCM output and the statistical model is good with some deficiencies likely due the representation of topography in the RCM.

1 Introduction

Improved methods of evaluating the performance of regional climate models (RCMs) are needed to adequately address the uncertainty associated with projections of future climate.

*Climate Change and Impacts Laboratory, Department of Earth Sciences, University of California Santa Cruz, msnyder@es.ucsc.edu

†Department of Applied Mathematics and Statistics, University of California Santa Cruz, bruno@ams.ucsc.edu

‡Climate Change and Impacts Laboratory, Department of Earth Sciences, University of California Santa Cruz, lcsloan@es.ucsc.edu

As an initial step, the performance of a RCM must be validated against observed climate data (see, for example, Richard et al., 2002). The current method for validating the output of an RCM involves running the model for several years using data derived from observations, such as NCEP/NCAR Reanalysis, as boundary conditions for the RCM. The resulting output can then be evaluated against observational data, which ideally are independent of the driving data (Mearns et al., 1999; Liston and Pielke, 2001; Snyder et al., 2002; Snyder and Sloan, 2005). Observational data are available as either station data or as gridded datasets. Current methods of evaluation are limited and it has been suggested that more advanced statistical techniques should be applied to the validation of climate model output and to the quantification of uncertainty (Berk et al., 2001). To this end, we describe below the limitations of station data and gridded datasets and demonstrate a method of evaluating RCM output using a Bayesian statistical model.

Station data have the advantage of recording climate variables with a high degree of certainty for a limited area. There are two primary issues with station data that must be addressed when using that data for climate model validation. The first is that, in regions with complex topography, such as the Pacific Northwest and California, the area around the observation station where the value of a given climate variable is accurate is much smaller than in regions with homogenous topography. For example, since temperature changes as a function of the lapse rate, the value of temperature recorded at a station located on the slope of a mountain is only valid for a small area around the station as opposed to if the station were located on a plain. The second issue is the existence of what we will refer to as a sampling bias due to station distribution. In California, observational stations are not evenly distributed across the state. In an ideal situation stations would be evenly distributed spatially, but most stations were placed in areas near cities. This means that sparsely populated areas are underrepresented; this includes high elevation, coastal, and desert regions.

Gridded datasets are derived from station data and interpolated to a particular resolution using various methods. The advantage of these datasets is that they provide complete and uniform spatial coverage of a region. This is possible because the climate information is interpolated between stations. The primary limitation of gridded datasets is that certain assumptions about the relationship between the stations must be made in order to generate the interpolated information. In the simplest case, the interpolation is performed using a nearest-neighbor method. This method interpolates information from a set of stations to a grid based on the data available at the stations nearest to a given gridcell. Gridcells without an observational station located in them are assigned values based on nearby stations without any adjustment due to topography or other geographic factors that might affect the climate of the gridcell such as distance from the coast.

To attempt to reconcile the shortcomings of station and gridded data, we apply a new statistical approach to validation of climate model output. The goal of this study is to use Bayesian statistical methods and historical records of SAT to aid in the validation of RCM results. Two fundamental difficulties associated with this process have to be addressed. First, RCM output for a given gridcell may correspond to a spatial average over an area where several observation stations are located. Second, some model gridcells may correspond to areas where there are no stations. Our solution is to use a statistical model that allows for the interpolation of the SAT field from station data. This in turn can be used to provide an estimation of the average observed SAT corresponding to a given area in particular, to the areas covered by the RCM gridcells. The use of a Bayesian approach provides full probabilistic assessment of all uncertainties in the statistical model fit and prediction. That is, RCM data are compared to the predictive distributions of SAT provided by the statistical model, not just to single-value predictions. A related approach was presented in Sansó and Guenni (2004) for simulations of rainfall over Nebraska and more generally in Fuentes et al. (2003).

2 Methods and Models

2.1 Regional Climate Model Description and Configuration

The International Center for Theoretical Physics (ICTP) Regional Climate Model, RegCM3 (Giorgi et al., 2004a,b; Pal et al., In review), is a third generation regional scale climate model derived from the National Center for Atmospheric Research-Pennsylvania State (NCAR-PSU) MM5 mesoscale model. RegCM3 uses a compressible, finite difference scheme with hydrostatic balance and vertical sigma coordinates. Improvements to RegCM3 over previous versions include a new large-scale cloud and precipitation scheme, the Subgrid Explicit Moisture Scheme (SUBEX) (Pal et al., 2000), a new ocean flux parameterization (Zeng et al., 1998), and the availability of a new cumulus convection scheme (Betts, 1986). RegCM3 includes the Biosphere-Atmosphere Transfer Scheme (BATS1E) (Dickinson et al., 1993) for surface process representation and the CCM3 radiative transfer package (Kiehl et al., 1996).

Our RCM domain is centered over California at 37.5°N and 121.5°W with a horizontal resolution of 40 km by 40 km. This model configuration uses 18 levels in the vertical, 60 gridcells in the North-South direction, and 55 in the East-West direction. Land-surface conditions were defined using the GLCC global landuse database (Loveland et al., 2000). The sea surface temperature (SST) dataset used is the UK Met Office GISST 1 degree global dataset (Rayner et al., 1996; Parker et al., 1995). Exponential relaxation is used at the RCM boundaries with 12 gridpoints in the buffer zone. The RCM is configured with the Grell convection scheme and the Fristch and Chappell closure assumption (Grell, 1993). The RCM was run for 41 years from 1960 to 2000 using NCEP/NCAR Reanalysis 1 as the driving boundary condition data (Kistler et al., 2001). The monthly average SAT output from the RCM is used for comparison with the results of the statistical model.

2.2 Observational Data

Monthly average SAT observations for 109 stations in Northern California were obtained from the Western U.S. Climate Historical Summaries at the Western Regional Climate Center (WRCC) website (`www.wrcc.dri.edu`). Stations were chosen based on length of record (50 years or longer) and the completeness of the dataset (i.e. amount of missing data). Stations with more than two years of missing data were rejected. The station locations are shown in Figure 1.

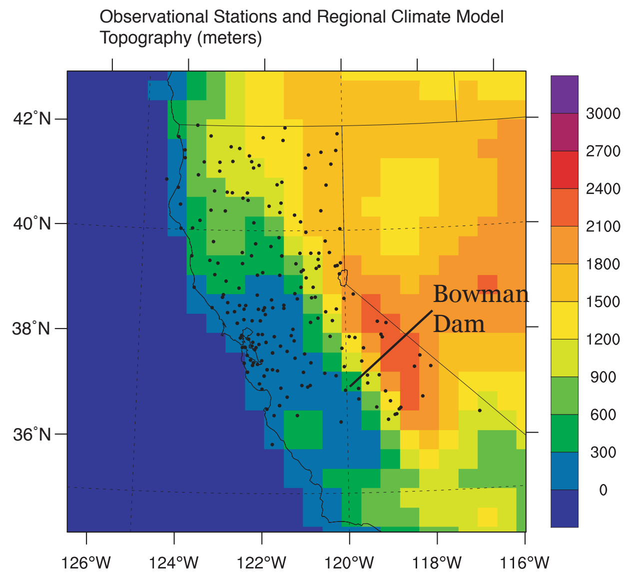

Figure 1: Locations of the observational stations used for model validation

2.3 Statistical Model

The fundamental assumption of the statistical model is that the SAT over Northern California is a Gaussian random field. Gaussian random fields provide a flexible family of models and are frequently used to model environmental variables. Traditional geostatistical applications are focused on the first and second moments of the process, so they rely on an implicit assumption of a Guassian distribution. A good account of geostatistical methods can be found in Cressie (1993). Diggle et al. (1998) discuss the advantages of a model based approach.

The statistical analysis proceeds as follows. We consider 54 years of monthly average SATs over 109 locations in Northern California. We build an empirical statistical model to represent the SAT field over the whole region for each month of the year. We validate the statistical methodology by fitting the model after leaving one station out. From this we obtain the predictive distribution of the SAT at the omitted location. We then compare the distribution to the observed values at that location. We repeat this procedure for all 109 locations. After validation, we use the statistical model to obtain the distributions of the average SATs over the areas corresponding to each of the gridcells from the RCM. We refer to these as the areal predictive distributions. Finally, we compare the RCM output to those predictive distributions and use the results as the basis of the RCM validation. Our validation of both the statistical model and the RCM output is based on comparing a distribution of SAT values to the RCM output. The idea is that the RCM output have to be likely samples from the predictive distributions. Note that assuming that the observations correspond to a Gaussian random field does not imply that the predictive distribution at a given site is a Gaussian distribution. As will be seen below, we propose a model that does not provide closed forms (i.e. specifically defining the distribution to be a Gaussian or other type of distribution) for the predictive distributions. Nevertheless we will be able to obtain samples from them. Validation methods that use a point estimate and a standard error are based on the assumption that the predictive (output from the statistical model) is normal, which may not be realistic. Moreover, since we use a Bayesian approach (see, for example Migon and Gamerman, 1999), we account for all the uncertainties due to parameter estimation.

The field of SATs is represented using convolutions of white noise with Gaussian kernels. Such models are presented in Higdon (2002). Priestley (1981) denotes these processes as general linear processes. Let the SAT at a location $s_i \in \mathbb{R}^2, i = 1, \ldots, n$, year $t, t = 1, \ldots, T$ and month $j, j = 1, \ldots, 12$ be denoted by r_{ijt}. We assume that

$$r_{ijt} = \sum_{l=1}^{m} k(s_i - u_l; \sigma^2, \rho)\omega_{lj} + \gamma h_i + \phi d_i + \varepsilon_{ijt}, \quad \varepsilon_{ijt} \sim N(0, \tau^2) \tag{1}$$

where h_i and d_i denote, respectively, the elevation and the distance from the coast of site s_i. $k(\cdot; \sigma^2, \rho)$ is a kernel depending on parameters σ^2 and ρ, given by

$$k(s_i - u_l; \sigma^2, \rho) = \frac{1}{2\pi} \frac{1}{\sigma^2} \exp\left\{ -\frac{1}{2\sigma^2}(s_i - u_l)' \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} (s_i - u_l) \right\}$$

so that $u_l \in \mathbb{R}^2, l = 1, \ldots, m$ are the locations where the kernels are centered. We then let

$$\omega_{lj} = \mu + \alpha \cos\left(\frac{2\pi}{12} j\right) + \beta \sin\left(\frac{2\pi}{12} j\right) + v_{lj}, \quad v_{lj} \sim N(0, \eta^2). \tag{2}$$

In Model (1) the kernels provide spatial dependence. σ^2 determines the range of the kernel, for a large σ^2 the kernel will average stations that are far apart. ρ determines the elongation of the kernel. For large negative values of ρ, the kernel will be strongly elongated in the NW-SE direction. A third parameter could be added to obtain different latitudinal and longitudinal ranges, but we found that the two parameter kernel produces good results. γ and ϕ are the parameters for the elevation of the station and the distance to the ocean. In Equation (2) α and β provide the strength of the seasonal deviations.

To fit the model, we split the observations into groups corresponding to the NOAA climate regions in California; thus we obtained one set of parameters for each one of the climate regions. The decision to split the observations by NOAA climate regions was reached by running a test case with the statistical model where all the stations where considered to be part of the same region. This resulted in a poor match between the statistical model output and the data at the individual stations. Splitting the data into the NOAA regions resulted in a much better match. Within a Bayesian approach, Equations (1) and (2) define the likelihood function for the parameters $\theta = (\tau^2, \sigma^2, \rho, \gamma, \phi, \mu, \alpha, \beta, \eta^2)$ based on the observed SATs, $\mathbf{R}$. We denote this by $p(\mathbf{R}|\theta)$. Inference for such parameters is done through their posterior distribution $p(\theta|\mathbf{R})$. This is obtained by multiplying $p(\mathbf{R}|\theta)$ by a chosen prior distribution, $p(\theta)$. For the data considered in this paper, the results were very similar for different choices of $p(\theta)$. We implemented a Markov Chain Monte Carlo (MCMC) method (see, for example, Gamerman, 1997) that produces samples of the posterior distribution of the parameters. Once these samples are available we can fix any point in the domain and obtain samples from the predictive distribution of the SAT at that site, for a given month. Let $r_m(s)$ denote the SAT at location s and month m, we can obtain the posterior predictive distribution of $r_m(s)$, $p(r_m(s)|\mathbf{R})$ as

$$p(r_m(s)|\mathbf{R}) = \int p(r_m(s)|\theta, R)p(\theta|\mathbf{R})d\theta. \tag{3}$$

Using the samples of θ from the MCMC output, we can draw samples from $p(r_m(s)|\mathbf{R})$ using Equation (3). In this way, we are able to produce predictions for the SAT of any given month and quantify the uncertainty of the statistical model output using a probability distribution.

In order to validate the RCM output, we need to estimate the average SAT for all model gridcells. Let g denote a given cell and $R_m(g)$ the average SAT for the cell for month m. Then

$$R_m(g) = \frac{1}{|g|}\int_g r_m(s)ds, \tag{4}$$

where $|g|$ denotes the area of g. Within the MCMC approach that we use to fit the statistical model, it is straightforward to approximate the integral in (4). We consider a set of points $s_1, \ldots, s_k$ in g, obtain samples from the predictive distribution of $r_m(s_j), j = 1, \ldots, k$ and average those samples. We repeat this process to obtain a description of the predictive distribution of $R_m(g)$. This is the distribution that is compared to the RCM output.

3 Results

3.1 Fit and validation of the statistical model

The first step of the statistical analysis consisted of fitting all the available observational data after grouping them according to the NOAA climate divisions. It is interesting to note that there has been discussion regarding the definition of the climate divisions and whether each division encompasses stations with similar climate records (Guttman and Quayle, 1996); in this case we obtained results indicative of very distinctive temperature regimes in each region. This can be observed in Figure 2 where the posterior distributions for the parameters defining the kernels in each region are shown. The plots suggest that there are significant differences between the spatial correlation of SAT in the different climate regions.

In order to validate the statistical model, we did the fitting by leaving one observational station out at a time. We then compared the actual left out observations to its predictive distribution. Figure 3 presents the results for the Bowman Dam station. Similar plots were obtained for all 109 stations. We observed that the mean of the observations was within the central range of the predictive distribution for most of the months at all stations.

To formalize the previous analysis for all stations simultaneously, we estimated the probability that the predictive distribution will be above the observed mean for each location and each month. The idea is that, if the mean of the observations is centrally located with

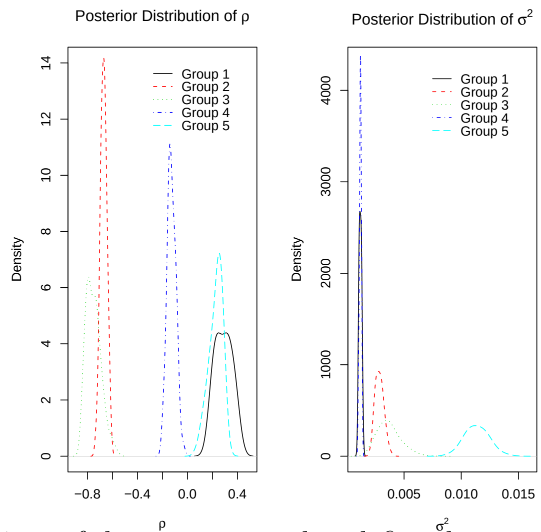

Figure 2: Posterior distributions of the parameters that define the convolution kernels. Each curve is a separate NOAA climate region. Since each curve has different distribution, this means that each climate region has a unique SAT regime.

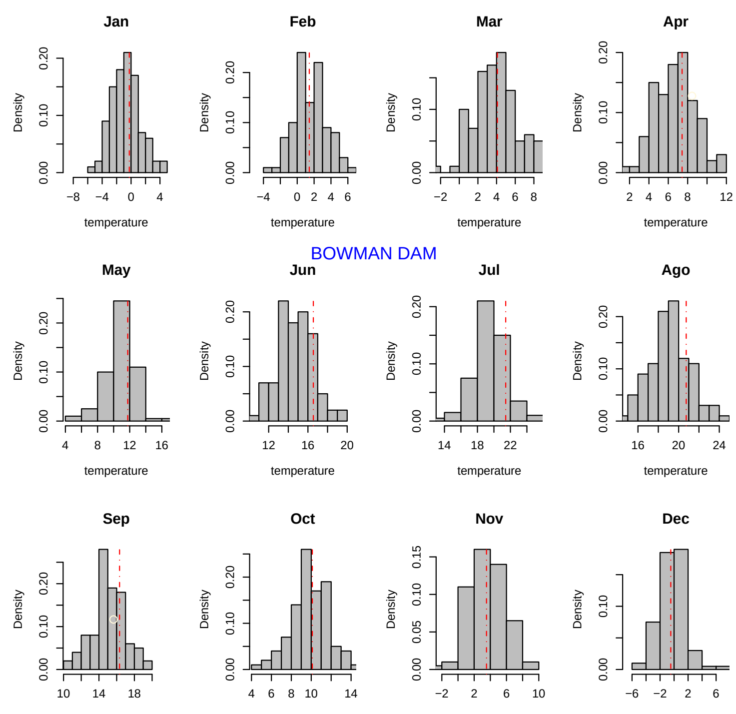

Figure 3: Histograms of the samples from posterior predictive distribution of Bowman Dam station for each month. The vertical line corresponds to the mean of the observations.

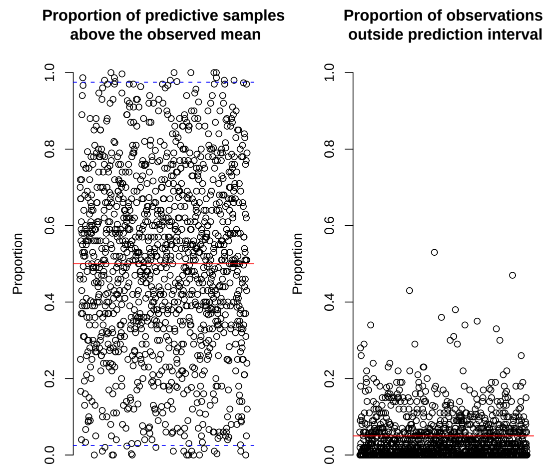

Figure 4: Probability that the predictive distribution for a given station and month will be above the observed mean (left panel). The blue dashed lines correspond to 2.5% and 97.5%. The red continuous line corresponds to 50%. The right panel shows the proportion of observed values outside the 95% probability predictive interval. The continuous red line corresponds to 5%.

respect to the predictive distribution, this probability should be close to 0.5. The left panel of Figure 4 presents such probabilities. A low probability indicates overestimation, a large probability indicates underestimation. We observe no particular over or under estimation tendency and a large proportion of the stations are between 20% and 80%. Furthermore, very few stations deviate greatly from the average, indicating that, when a station is removed, the statistical model is able to predict its SAT in a manner that is consistent with the average observed at that station. This is the case for the large majority of the stations.

A second measure of model assessment is obtained by calculating the proportion of observations that lie outside the interval given by the 2.5% and 97.5% predictive quantiles. These correspond to the boundaries of a predictive interval with 95% probability. These proportions are shown in the right panel of Figure 4. We observe that a large majority of the predictions fall within 5% of the 95% interval.

3.2 Comparison with Regional Model Output

To compare the results of the statistical model with the RCM output, we generated 100 realizations of the average SAT of each gridcell for each month. From this we obtained a description of the predictive distribution of the average SAT for each gridcell and each month. We then compared the 41 years of monthly RCM output for each gridcell with the corresponding predictive distribution. The percentage of months at each gridcell where the RCM SATs were between the 2.5% and 97.5% quantiles of the predictive distribution were calculated and are shown in Figure 5. Higher values indicate good agreement between the statistical model and the RCM. As shown in Figure 5, some regions match well with the output of the statistical model while others are poorly matched. One region with a

Matching %	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0 to 9	41	33	36	35	40	40	45	46	34	28	37	34
10 to 19	15	5	7	17	14	10	15	13	13	8	3	11
20 to 29	12	4	12	11	7	10	16	10	16	13	9	20
30 to 39	9	26	16	12	20	19	9	12	10	24	17	12
40 to 49	7	15	14	13	12	10	13	12	9	17	22	7
50 to 59	8	10	8	9	3	12	6	2	7	4	6	5
60 to 69	7	5	9	8	6	3	8	12	12	6	5	12
70 to 79	9	8	10	6	9	10	8	7	11	7	2	10
80 to 89	6	6	6	16	15	9	3	6	6	9	13	4
90 to 100	39	41	35	26	27	30	30	33	35	37	39	38

Table 1: The number of gridcells from the RCM model output that match the statistical model output by percent. There are ten categories that indicate the percentage match between the 41 RCM model output values and 100 statistical model values, on a monthly basis for each gridcell. The higher the matching percentage, the better the match between the RCM and statistcal model for those gridcells.

particularly poor match is the northern Central Valley. In this region, the RCM elevations at each gridcell are overestimated in most cases due to the pervasive topographic gradients in this region. RCM results from some gridcells along the coast are also poorly matched with the output of the statistical model. Again, this is likely due to topographic mismatch between the elevations of the individual stations used in the statistical model, and the RCM elevation for that gridcell. The mismatch along the coast is furthered complicated by the RCM representation of the coastline at 40 km resolution. At this resolution, the RCM uses a coarse representation of the coast that misses many of the finer scale details.

Table 1 shows that in almost every month over 30 gridcells (out of 153 total) fall in the 90% to 100% range, indicating an excellent match. We also see that about the same number of gridcells fall in the 0% to 9% match category. The rest of the gridcells are distributed fairly evenly through the categories in between. Since the statistical model matches quite well with the station data, the mismatch between RCM output and the statistical model in Table 1 is likely due to problems with the RCM representation of SAT. The RCM mismatch could be due to a number of factors. The primary contributor is likely the representation of topography in the RCM. Each RCM gridcell contains an average elevation value derived from a 10 minute resolution topography dataset. On a 40 km resolution grid, the preprocessing program that generates the topography for the RCM gridcells averages over approximately 4 values from the high resolution dataset. These values are also adjusted to make the topography smoother for numerical reasons. Areas with topographic gradients will be adjusted to the mean value of the slope. In areas where actual topographic features are at a finer scale than the model resolution, those features disappear at the model resolution. An example of this is the Owens Valley in the Eastern Sierra, a valley between two high mountain ranges, which is not represented in the model at 40 km resolution. The net result is lower than observed elevations in some regions (i.e. Sierra Nevada) and higher than observed elevations in others (parts of the Central Valley).

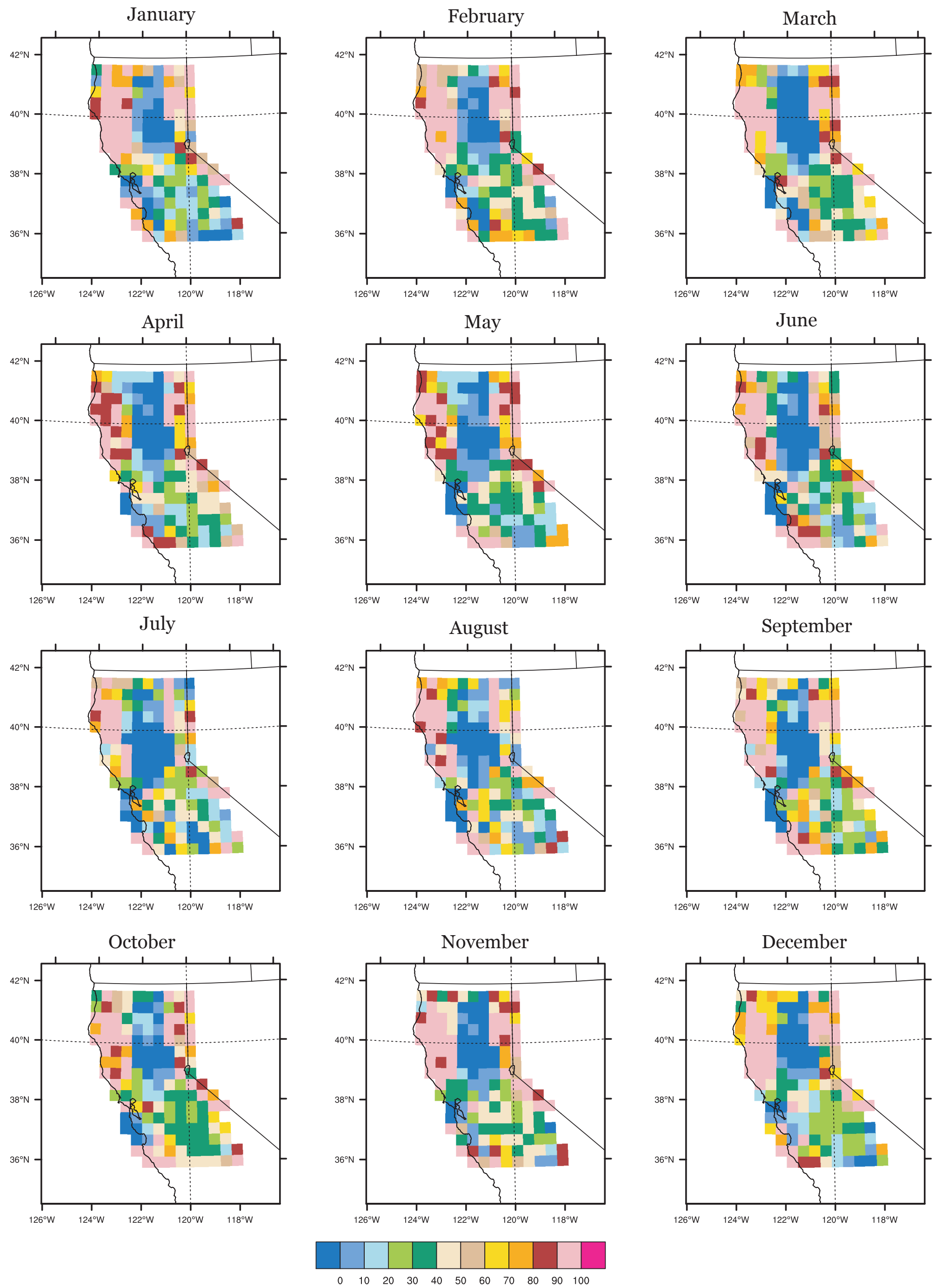

Figure 5: The percentage of monthly RCM SATs that fall within the 95% interval of the statistical model's predicted SATs. The colors indicate the percentage of monthly RCM SAT values (41 values at each gridcell per month) at each gridcell that fall between 2.5% and 97.5% of the SAT values (100 values at each gridcell per month) predicted by the statistical model. Higher percentages indicate a better match between the statistical model and the RCM.

4 Discussion

At each station, a comparison of the statistical model prediction of monthly SATs shows very good agreement with the observational record (See Figure 3 for an example from one station). For all stations, we find that the majority of the predictions from the statistical model lie within the 95% probability interval (Figure 4). This indicates that the statistical model is doing a very good job at predicting the monthly SAT values over northern California. The statistical model is empirically built from the observations; this implies that it can be used as a reliable tool for the comparison of regional model output and historical records.

The comparison of the statistical model's predictions with the output of the RCM shows good agreement for many gridcells on a monthly basis. Gridcells where the match between the RCM and the statistical model is poor are areas where the RCM is doing poorly at representing SAT. One possible reason for this is the simplification and smoothing of topography that occurs in the RCM. At 40 km resolution there is still significant underestimation of mountain elevations and overestimation of valley elevations. There may also be inherent biases introduced by the boundary conditions that the RCM may not be able to overcome. A combination of these factors is likely the cause of the mismatch and warrants a more detailed analysis of the RCM, driving conditions, and observational data.

The unique feature of the use of a Bayesian statistical method is that it can be applied to any resolution of climate model. The use of this method addresses one of the disadvantages of gridded observational climate data which is that it must be interpolated to the same resolution as the climate model output for direct comparison. In almost all cases, the interpolation method used is simplistic (i.e. nearest neighbor weighting) and doesn't take into account factors that the statistical model considers (i.e. elevation and distance from the coast). If the gridded observational data is coarser than the climate model output, the resulting interpolated data increases the spatial resolution without taking into account topography and distance from the coast. If the gridded observational data is finer than the climate model output, the interpolation may produce unrealistic values as well. Our Bayesian statistical method allows the point station data to be adjusted to any climate model grid resolution and includes a full accounting of the influence of topography and distance from the coast, as well as the uncertainty of the estimation.

This method can be used for other climate variables, although the actual statistical model may need some modifications to account for information that is relevant to that variable. For example, we found that the distance to the ocean is important to explain the geographical variability of SAT. The influence of the marine layer, especially during the summer months, has a very significant effect on SAT at coastal stations and can also influence inland stations through diurnal changes in surface winds. Distance from the coast may not be as important for other variables and there may be other factors that need to be taken into account. For example, considering precipitation introduces the difficulty of dealing with a variable with a positive probability of getting the value zero. This is an issue that has to be taken into account when fitting the statistical model for interpolation, see Sansó and Guenni (2004) and references therein for discussion.

5 Conclusions

In order to improve the accuracy and to better understand the uncertainties associated with climate model projections of future climate, there exists a need for rigorous evaluation of the performance of RCMs in simulating a known climate. By validating RCM output with observations, we will be able to better discern changes in climate due to anthropogenic climate change from biases inherent in RCMs. Our goal was to examine a method of RCM validation using statistical modeling techniques coupled with observational station data. This investigation has shown that the Bayesian statistical method holds promise for aiding in the validation of model output from RCMs. We found that the statistical model does a very good job of representing the observed SAT over the region. The comparison of the statistical model output to the RCM output revealed that deficiencies in the RCM representation of SAT lead to the mismatch between the RCM and statistical model output. The complexities of the climate of California make it an ideal place to evaluate our method. This same method could be readily applied to other RCMs, at any resolution, for different regions, and for different climate variables.

References

Berk, R. A., Fovell, R. G., Schoenberg, F. and Weiss, R. E. (2001) The use of statistical tools for evaluating computer simulations. *Climatic Change*, **51**, 119–130.

Betts, A. K. (1986) A new convective adjustment scheme. 1. observational and theoretical basis. *Quarterly Journal of the Royal Meteorological Society*, **112 (473)**, 677–691.

Cressie, N. A. C. (1993) *Statistics for Spatial Data, Revised Edition*. New York: John Wiley and Sons.

Dickinson, R., Henderson-Sellers, A. and Kennedy, P. (1993) Biosphere-atmosphere transfer scheme (BATS) version 1e as coupled to the NCAR community climate model. *Tech. rep.*, NCAR/TN-387+STR.

Diggle, P., Tawn, J. and Moyeed, R. (1998) Model based geostatistics (with discussion). *Applied Statistics*, **47**, 299–350.

Fuentes, M., Guttorp, P. and Challenor, P. (2003) Statistical assessment of numerical models. *International Statistical Review*, **71**, 201–221.

Gamerman, D. (1997) *Markov Chain Monte Carlo*. London, UK: Chapman and Hall.

Giorgi, F., Bi, X. and Pal, J. S. (2004a) Mean, interannual variability and trends in a regional climate change experiment over Europe. I. Present-day climate (1961-1990). *Climate Dynamics*, **22 (6-7)**, 733–756.

— (2004b) Mean, interannual variability and trends in a regional climate change experiment over Europe. II: Climate change scenarios (2071-2100). *Climate Dynamics*, **23 (7-8)**, 839–858.

Grell, G. (1993) Prognostic evaluation of assumptions used by cumulus parameterizations. *Monthly Weather Review*, **121**, 764–787.

Guttman, N. B. and Quayle, R. G. (1996) A historical perspective of U.S. climate divisions. *Bulletin of the American Meteorological Society*, **77**, 293–303.

Higdon, D. (2002) Space and space-time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues* (eds. C. Anderson, V. Barnett, P. C. Chatwin and A. H. El-Shaarawi), 37–56. London: Springer Verlag.

Kiehl, J. T., Hack, J. J., Bonan, G. B., Boville, B. A., Briegleb, B. P., Williamson, D. L. and Rasch, P. J. (1996) Description of the NCAR community climate model (CCM3), CGD. *Tech. rep.*, NCAR TN-420+STR.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R. and Fiorino, M. (2001) The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, **82**, 247–267.

Liston, G. E. and Pielke, R. A. (2001) A climate version of the regional atmospheric modeling system. *Theor. Appl. Climatology*, **68**, 155–173.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, J., Wang, L. and Merchant, J. W. (2000) Development of a global land cover characteristics database and IGBP discover from 1-km AVHRR data. *International Journal of Remote Sensing*, **v. 21, no. 6/7**, p. 1,303–1,330.

Mearns, L. O., Bogardi, I., Giorgi, F., Matyasovszky, I. and Pakecki, M. (1999) Comparison of climate change scenarios generated from regional climate model experiments and statistical downscaling. *Journal of Geophysical Research - Atmospheres*, **104 (D6)**, 6603–6622.

Migon, H. and Gamerman, D. (1999) *Statistical Inference: An Integrated Approach*. London, UK: Arnold.

Pal, J., Small, E. and Eltahir, E. (2000) Simulation of regional-scale water and energy budgets: Representation of subgrid cloud and precipitation processes within RegCM. *Journal of Geophysical Research-Atmospheres*, **105**, 29579–29594.

Pal, J. S., Giorgi, F., Bi, X., Elguindi, N., Solomon, F., Gao, X., Francisco, R., Zakey, A., Winter, J., Ashfaq, M., Syed, F., Bell, J. L., Diffenbaugh, N. S., Karmacharya, J., Konare, A., Martinez, D., da Rocha, R. P., Sloan, L. C. and Steiner, A. (In review) The ICTP RegCM3 and RegCNET: Regional climate modeling for the developing world. *Bulletin of the American Meterological Society*.

Parker, D. E., Folland, C. K. and Jackson, M. (1995) Marine surface temperature: Observed variations and data requirements. *Climatic Change*, **31**, 559–600.

Priestley, M. (1981) *Spectral Analysis and Time Series*. San Diego, USA: Academic Press.

Rayner, N. A., Horton, E. B., Parker, D. E., Folland, C. K. and Hackett, R. B. (1996) Version 2.2 of the global sea-ice and sea surface temperature data set, 1903-1994. *Tech. rep.*, Climate Research Technical Note 74.

Richard, J., Kirtman, B., Laprise, R., von Storch, H. and Wergen, W. (2002) Atmospheric regional climate models (RCMs): A multiple purpose tool? *Tech. rep.*, Report of the joint WGNE/WGCM ad hoc panel on Regional Climate Modelling `http://w3g.gkss.de/G/Mitarbeiter/Storch/pdf/RCM.report.040302.pdf`.

Sansó, B. and Guenni, L. (2004) A Bayesian approach to compare observed rainfall data to deterministic simulations. *Environmetrics*, **15**, 597–612.

Snyder, M. A., Bell, J. L., Sloan, L. C., Duffy, P. B. and Govindasamy, B. (2002) Climate responses to a doubling of atmospheric carbon dioxide for a climatically vulnerable region. *Geophysical Research Letters*, **29 (11), 1514**.

Snyder, M. A. and Sloan, L. C. (2005) Transient future climate over the western u.s. using a regional climate model. *Earth Interactions*. In press.

Zeng, X. B., Zhao, M. and Dickinson, R. E. (1998) Intercomparison of bulk aerodynamic algorithms for the computation of sea surface fluxes using TOGA COARE and TAO data. *Journal of Climate*, **11 (10)**, 2628–2644.