The Annals of Applied Statistics
2019, Vol. 13, No. 2, 746–766
https://doi.org/10.1214/18-AOAS1219
© Institute of Mathematical Statistics, 2019

COMPLETE SPATIAL MODEL CALIBRATION[1]

BY YEN-NING HUANG*, BRIAN J. REICH*, MONTSERRAT FUENTES† AND A. SANKARASUBRAMANIAN*

*North Carolina State University** *and Virginia Commonwealth University*†

Computer simulation models are central to environmental science. These mathematical models are used to understand complex weather and climate patterns and to predict the climate's response to different forcings. Climate models are of course not perfect reflections of reality, and so comparison with observed data is needed to quantify and to correct for biases and other deficiencies. We propose a new method to calibrate model output using observed data. Our approach not only matches the marginal distributions of the model output and gridded observed data, but it simultaneously postprocesses the model output to have the same spatial correlation as the observed data. This comprehensive calibration method permits realistic spatial simulations for regional impact studies. We apply the proposed method to global climate model output in North America and show that it successfully calibrates the model output for temperature and precipitation.

1. Introduction. Computer models are used extensively in environmental science to understand and predict spatiotemporal processes. Weather forecasts are derived from numerical weather prediction models that are essentially computer simulations of the atmosphere based on differential equations. Predictions from computer models can cover large spatial domains with higher spatial and temporal coverage than data obtained from monitoring networks. As a result computer models can be a valuable tool to compensate for the limitations of the monitoring data. However, they are deterministic and do not carry information about the inherent uncertainty. Also, both spatial and temporal biases in these computer models need to be recognized and calibration of the model is necessary. Evaluation and bias correction of the performance of physical models are needed to obtain reliable forecasts.

Comparison of computer model outputs with observed data is difficult because they are often on different spatial scales. Since computer model outputs are often given as averages over spatial grid cells and observed data are recorded as point measurements, we must take into account the change of support problem when we compare the two data sources. Several approaches have been proposed to discuss the change of support problem for block averages. A method to deal with the

Received August 2017; revised September 2018.

[1]Supported in part by grants from the Department of the Interior (DOI-14-1-04-9) and the National Institutes of Health (R21ES022795-01A1, R01ES014843-02 and R01ES027892).

Key words and phrases. Bayesian methods, calibration, spatial statistics.

point-to-area change of support problem in geostatistics is given by block kriging (Banerjee, Carlin and Gelfand (2014), Cressie (1993)) which allows predictions of the block average of a process given observations as point measurements. Young and Gotway (2007) extended block kriging and proposed a geostatistical method for handling different types of change of support problems. Wikle and Berliner (2005) proposed a hierarchical Bayesian model that allowed combination of data observed at different spatial scales. Bradley, Wikle and Holan (2015) proposed a hierarchical Bayesian methodology to perform spatiotemporal change of support for survey data. See Gotway and Young (2002) and Cressie and Wikle (2011) for a comprehensive review of statistical methods for combining incompatible spatial data.

Methods to adjust for spatial misalignment have been applied to computer model output. Fuentes and Raftery (2005) developed a Bayesian approach that combines computer model output and station data with different spatial resolutions by introducing a latent point-level process driving both data sources. McMillan et al. (2010) and Zidek, Le and Liu (2012) presented a Bayesian hierarchical method which combines station data and numerical model output with application in $PM_{2.5}$ and ozone data respectively. Berrocal, Gelfand and Holland (2010) developed a fully model-based strategy within a Bayesian framework to downscale air quality numerical model outputs to a point level. In their static spatial model the observations are regressed on the numerical model outputs using spatially-varying coefficients that are specified through a correlated spatial Gaussian process. Gel, Raftery and Gneiting (2004) proposed the geostatistical output perturbation method for calibrating probabilistic mesoscale weather field forecasting. Berrocal, Raftery and Gneiting (2007) introduced a spatial Bayesian model averaging method to calibrate forecast ensembles of whole weather fields simultaneously. Berrocal, Raftery and Gneiting (2008) presented a two-stage spatial method for obtaining probabilistic forecasts of precipitation from a numerical forecast. Spatial postprocessing of ensemble forecasts were developed in Feldmann, Scheuerer and Thorarinsdottir (2015) and Schefzik (2017). Berrocal, Gelfand and Holland (2012) combined monitoring data at the point level with numerical model outputs at the grid cell level to obtain more accurate environmental exposure assessment. Reich, Chang and Foley (2014) proposed a computationally efficient spectral method for spatial downscaling with applications to ozone data. Uncertainty quantification of computer-model output using ensemble copula coupling is discussed in Schefzik, Thorarinsdottir and Gneiting (2013).

In the study of climate change, general circulation models (GCMs) are used to provide insights into climate changes due to anthropogenic forcings. However, there is usually a gap between large-scale outputs from GCMs and the fine-scale outputs which are required for local and regional climate impact assessments. Several dynamical and statistical downscaling methods have been developed to reduce biases between climate model outputs and monitoring data and also to address the spatial mismatch between multiple GCM outputs and the data needed for climate

application studies (Bhowmik et al. (2017), Devineni and Sankarasubramanian (2010), Seo et al. (2016)).

Most of the downscaling methods in the study of climate change impacts rely on the following approaches: a delta approach where a change (delta) is added to the observed variable of interest in order to make projections of the future (Hay, Wilby and Leavesley (2000)); a bias correction-statistical downscaling model which uses a quantile mapping approach to downscale the response variable to a regular grid (Luers et al. (2006), Wood et al. (2004)), and the linear regression approach which downscales the response variable to individual station locations (Dettinger et al. (2004)). One critical aspect in bias correcting climate change projections with monitored data is that the two data sources are asynchronous, that is, two time-varying quantities without simultaneous measurements. Hence, studies have suggested asynchronous regression approaches (Bhowmik et al. (2017), O'Brien, Sornette and McPherro (2001), Stoner et al. (2012)).

These approaches only deal with marginal distributions, but in many situations it is necessary to adjust model output so its joint distributions resemble those of the underlying process. For example, we may want to study the dependence between temperature and precipitation to forecast the probability of a specific weather event, or the joint distribution of a collection of spatial locations to study regional changes. Multivariate bias correction methods for climate model outputs were investigated in Vrac and Friederichs (2015) and Cannon (2016). The general idea of empirical copula-bias correction and the Schaake shuffling methods (Clark et al. (2004)) is to reshuffle the predictive spatiotemporal data according to some rank structure derived from training data. As a result the test dataset receives a similar dependence structure as the training data. Möller, Lenkoski and Thorarinsdottir (2012) proposed a method for post processing an ensemble of multivariate forecasts using Bayesian model averaging along with a Gaussian copula. Their method can be used to recover the dependence between multivariate weather quantities, but no spatial factors are added. Recently, Bhowmik et al. (2017) suggested asynchronous canonical correlation analysis for preserving temporal crosscorrelation between GCM outputs and monitored data; however, it did not preserve the spatial crosscorrelation that is exhibited in the monitoring data.

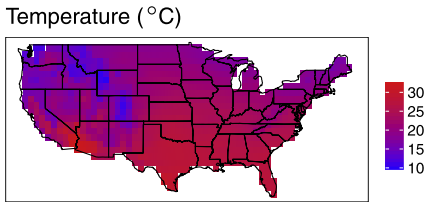
In this article we present a different approach to adjust for biases of the marginal distribution as well as the spatial correlation of GCM outputs. We model the spatially-varying marginal distributions of climate model outputs and gridded observed data using a flexible non-Gaussian assumption. Therefore, our method not only adjusts the mean but also the entire distribution including the variance and skewness and the extreme events. The Karhunen–Loéve decomposition is used to expand the two processes and isolate their differences in spatial correlation functions. Because we calibrate the entire marginal distribution and also spatial dependence structure, we call our procedure the "complete" spatial calibration method.

The article is organized as follows. The data description is provided in Section 2. The details of the proposed spatial model calibration are provided in Section 3.

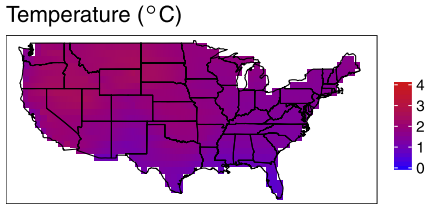
Computing details are given in Section 4. Real data applications are conducted in Section 5. A discussion is given in Section 6.

2. Data description. Monthly precipitation and average surface (2 m) temperature data from the CNRM–CM5 GCM (Centre National de Recherches Météorologiques–Coupled global climate Model version 5 General Circulation Models; nine ensemble members) are used for calibration. CNRM–CM5 was developed jointly by CNRM–GAME (Centre National de Recherches Météorologiques–Groupe d'études de l'Atmosphère Météorologique) and CERFACS (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique) to contribute toward the CMIP5 (Climate Model Inter-comparison Project 5) multimodel ensemble. The CMIP5 includes twentieth century historical runs and decadal hindcast runs (Taylor, Stouffer and Meehl (2012)), and we use the decadal hindcasts for the proposed study. Experiments forced with the representative concentration pathway (RCP) 8.5 scenario (van Vuuren et al. (2011)) are considered for future projection analysis (four ensemble members). Individual ensemble members are obtained by initializing the CNRM-CM5 model with different atmospheric and oceanic conditions to develop hindcasts and future projections of precipitation and temperature. Thus, collectively, the ensemble represents the uncertainty due to initial conditions. RCP8.5 represents the most aggressive growth scenario with the radiative forcing reaching 8.5 Watts per meter^2 by 2100. Spatial resolution of the model is $\sim$1.4° which is regridded to $\sim$1.0° using bilinear interpolation for the purpose of further downscaling. We use the gridded observed precipitation and temperature over the conterminous United States from 1950 to 1999 over the summer (June to August) at a spatial resolution of $\sim$1.0° from the Bureau of Reclamation (BOR) database. Details of the gridding procedure can be found in Maurer et al. (2002). We consider the CNRM model (nine ensemble members), 30-year hindcasts from 1981 to 2010 also over the summer at 815 locations (1.0° grid cells) in the conterminous United States.

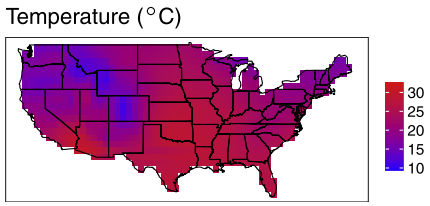
The means and standard deviations of gridded observed data and one ensemble member are given in Figures 1 and 2. For temperature the climate model captures the general mean pattern, but its pattern is smoother, and it has larger standard deviations than the gridded observed data. For precipitation the climate model has higher mean in the west and the northeast and lower mean in the midwest and the south in comparison to the gridded observed data. The climate models' standard deviations are smaller than the gridded observed data. Semivariograms given in Figures 2(c) and 2(d) indicate that for both temperature and precipitation, the two data sources have different spatial correlation functions. While these data exhibit strong spatial correlation, temporal dependence appears to be weak with less than 5% of observations having statistically significant autocorrelation. Most of the 815 locations do not have an overall increasing or decreasing trend over time. Therefore, in this analysis we ignore temporal dependence and focus on spatial modeling.



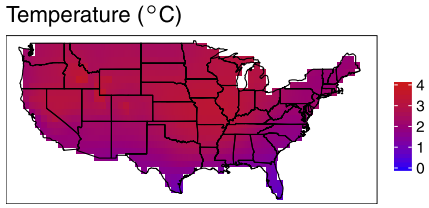
(a) Mean of gridded observed data



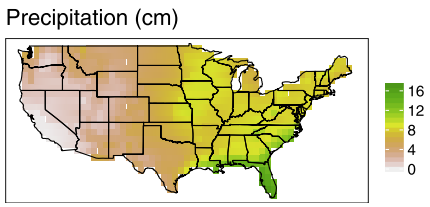
(b) Standard deviation of gridded observed data



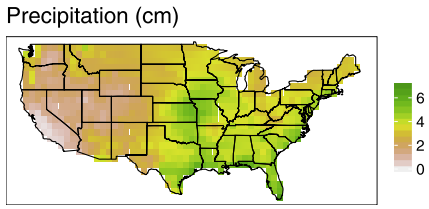
(c) Mean of model outputs



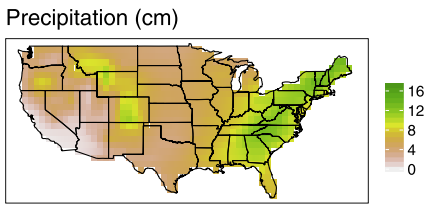
(d) Standard deviation of model outputs



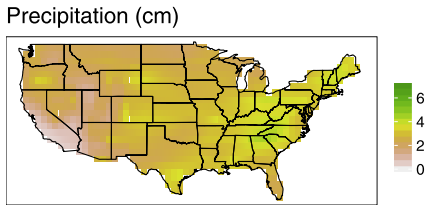
(e) Mean of gridded observed data



(f) Standard deviation of gridded observed data



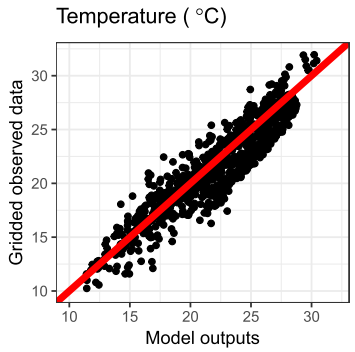
(g) Mean of model outputs

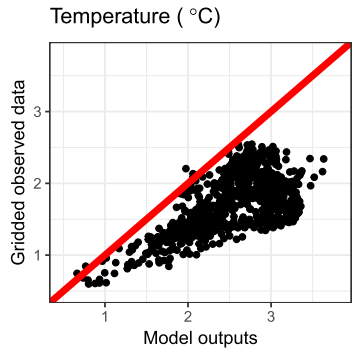


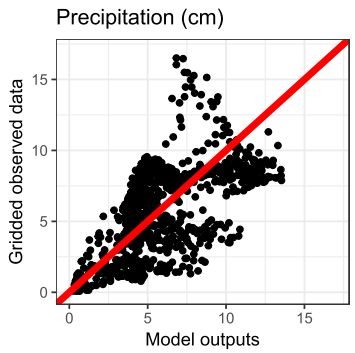
(h) Standard deviation of model outputs

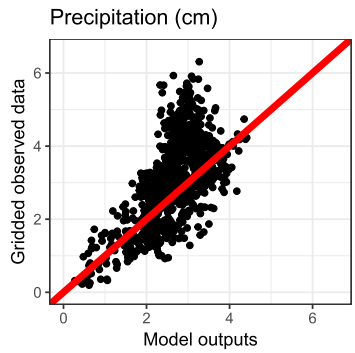
FIG. 1. *Pointwise mean and standard deviation of monthly average summer* (*June*, *July and August*) *temperature* (°C) *and precipitation* (*cm*) *from gridded observed data* (1950–1999) *and one ensemble member of climate models* (1981–2010).

We check for the distributions of gridded observed data and climate model outputs. In addition to the normal distribution, we also consider the skew-t distribution (Azzalini and Capitanio (2003), Jones and Faddy (2003), Morris et al. (2017)) which has four parameters to provide a more flexible model. Figure 3 displays



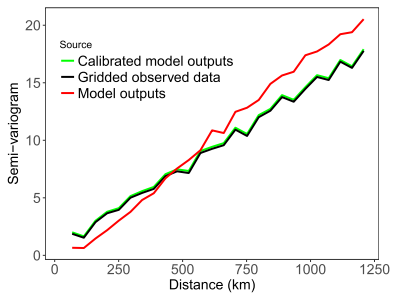


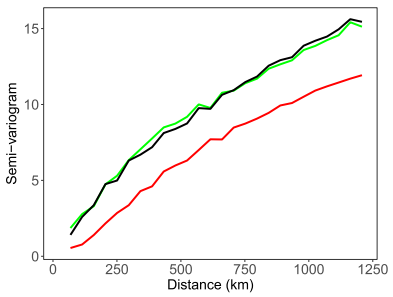




(a) Comparison of point-wise means

(b) Comparison of point-wise standard deviations

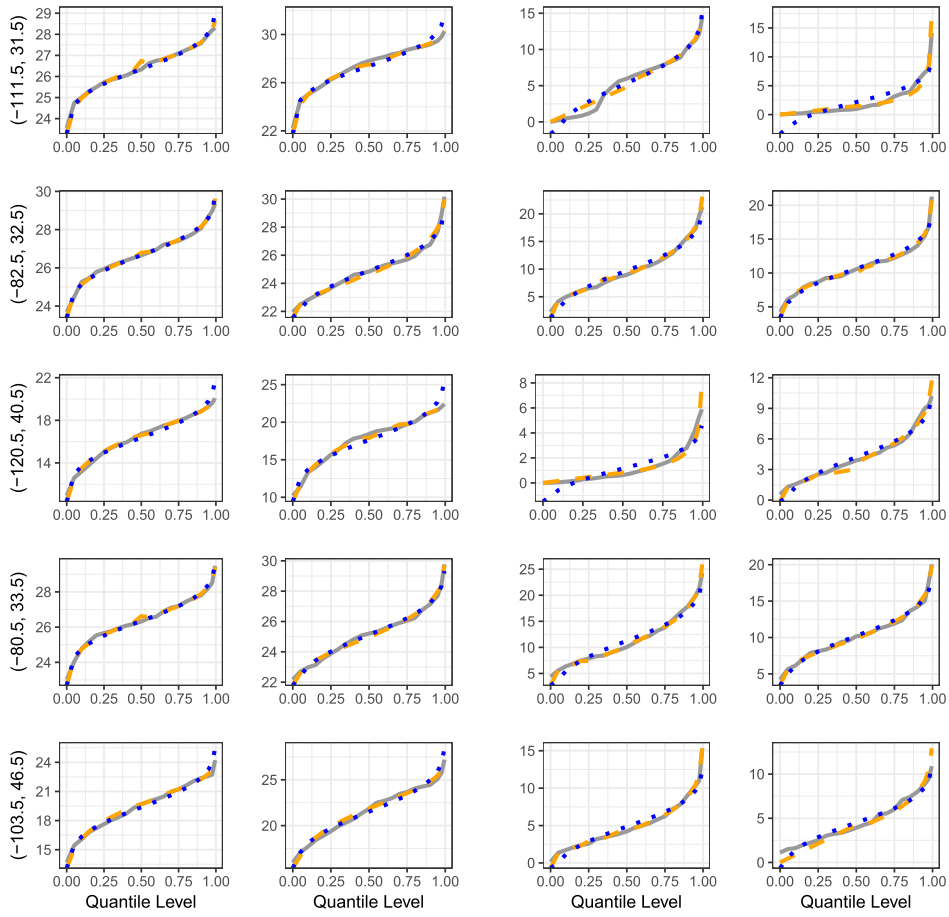




(c) Semi-variograms for temperature from gridded observed data, model outputs and calibrated model outputs (Fourier, skew-t).

(d) Semi-variograms for precipitation from gridded observed data, model outputs and calibrated model outputs (Fourier, skew-t).

FIG. 2. *Comparison of means, standard deviations and semivariograms of monthly summer* (*June, July and August*) *temperature* (°C) *and precipitation* (*cm*) *from gridded observed data* (1950–1999) *and climate models* (1981–2010, *the same ensemble member as in Figure* 1).



(a) Temperature ($^\circ C$): gridded observed data (left) and model outputs (right).

(b) Precipitation (cm): gridded observed data (left) and model outputs (right).

FIG. 3. *Quantile plots gridded observed data* (*gray*), *model outputs* (*gray*), *fitted normal* (*dotted*) *and fitted skew-t* (*dashed*) *at five arbitrarily selected locations. Longitude and latitude for each location are given beside the vertical axis in* (a).

normal and skew-t quantile plots for both data sources at five arbitrarily chosen locations. From the quantile plots for temperature and precipitation, we can see that the skew-t distribution provides a better fit than the Gaussian distribution.

3. Methods. In this section we present our method to post-process climate model output so that its marginal distributions and the spatial correlation match those of the observed data. Our approach has the following three components: first, we model the spatially-varying and potentially non-Gaussian marginal distributions of the model outputs and observed data; second, we model the spatial

covariance of both data sources; and finally, we perform calibration to rectify differences between the two data sources. Although these three components are described separately, they are in fact performed simultaneously in the Bayesian hierarchical model described in Section 4.

3.1. *Marginal distributions.* Define $X_t(\mathbf{s})$ as the climate model output for time t at site $\mathbf{s}$ and $Y_t(\mathbf{s})$ as the gridded observed data at time t at site $\mathbf{s}$. Let $F_{\mathbf{s}}(x)$ and $G_{\mathbf{s}}(y)$ be the CDFs of $X_t(\mathbf{s})$ and $Y_t(\mathbf{s})$ at location $\mathbf{s}$ respectively. The CDFs could vary with both space and time, but we focus on space here. The two processes can both be transformed to the standard normal distribution by defining $Z_{Xt}(\mathbf{s}) = \Phi^{-1}\{F_{\mathbf{s}}[X_t(\mathbf{s})]\}$ and $Z_{Yt}(\mathbf{s}) = \Phi^{-1}\{G_{\mathbf{s}}[Y_t(\mathbf{s})]\}$. Similarly, we can write the data as $X_t(\mathbf{s}) = F_{\mathbf{s}}^{-1}\{\Phi[Z_{Xt}(\mathbf{s})]\}$ and $Y_t(\mathbf{s}) = G_{\mathbf{s}}^{-1}\{\Phi[Z_{Yt}(\mathbf{s})]\}$.

For the special case when both processes are Gaussian, they can be written as

$$X_t(\mathbf{s}) = \mu_X(\mathbf{s}) + \sigma_X(\mathbf{s})Z_{Xt}(\mathbf{s}) \quad \text{and} \quad Y_t(\mathbf{s}) = \mu_Y(\mathbf{s}) + \sigma_Y(\mathbf{s})Z_{Yt}(\mathbf{s}), \tag{1}$$

where $\mu_X(\mathbf{s})$ and $\sigma_X(\mathbf{s}) > 0$ are the spatially-varying mean and standard deviation of $X_t(\mathbf{s})$, and $\mu_Y(\mathbf{s})$ and $\sigma_Y(\mathbf{s}) > 0$ are those of $Y_t(\mathbf{s})$. As mentioned in the previous section, for a richer class of marginal distributions we also consider the spatial skew-t process

$$X_t(\mathbf{s}) = \mu_X(\mathbf{s}) + \xi_X|r_{Xt}| + \sigma_{Xt}\sigma_X(\mathbf{s})Z_{Xt}(\mathbf{s}), \tag{2}$$

where $r_{Xt} \overset{\text{indep}}{\sim} N(0, \sigma_{Xt}^2)$ and $\sigma_{Xt}^2 \overset{\text{i.i.d.}}{\sim} \text{InvGamma}(\nu_X/2, \nu_X/2)$ are random effects. $Y_t(\mathbf{s})$ can be written in the same way with parameters $\mu_Y(\mathbf{s})$, ξ_Y and $r_{Yt} \overset{\text{indep}}{\sim} N(0, \sigma_{Yt}^2)$, $\sigma_{Yt}^2 \overset{\text{i.i.d.}}{\sim} \text{InvGamma}(\nu_Y/2, \nu_Y/2)$. The marginal distributions of $X_t(\mathbf{s})$ (over r_{Xt} and σ_{Xt}^2) and $Y_t(\mathbf{s})$ (over r_{Yt} and σ_{Yt}^2) both follow a skew-t distribution with four parameters $\{\mu_X(\mathbf{s}), \sigma_X(\mathbf{s}), \nu_X, \xi_X\}$ and $\{\mu_Y(\mathbf{s}), \sigma_Y(\mathbf{s}), \nu_Y, \xi_Y\}$ respectively. Here, $\mu_X(\mathbf{s})$ is the location parameter, $\sigma_X(\mathbf{s})$ is the scale parameter and $\nu_X > 0$ is the shape parameter (degrees of freedom) of the skew-t distribution in which smaller values represent heavier tails. Positive values of the skewness parameter ξ_X indicate a right-skewed distribution and vice versa.

The models in (1) or (2) do not include any measurement error. If observed data are measured with error, denote $Y_t^*(\mathbf{s})$ as the measured value and let $Y_t^*(\mathbf{s}) \overset{\text{indep}}{\sim} N[Y_t(\mathbf{s}), \tau_Y^2]$. In this case our focus is on the distribution of the underlying true value $Y_t(\mathbf{s})$, and our goal remains to match the distribution of $X_t(\mathbf{s})$ with $Y_t(\mathbf{s})$ instead of $Y_t^*(\mathbf{s})$.

3.2. *Spatial correlation.* After modeling marginal distributions we assume that $Z_{Xt}(\mathbf{s})$ and $Z_{Yt}(\mathbf{s})$ are standard Gaussian for all $\mathbf{s}$, and jointly $\mathbf{Z}_{Xt} = [Z_{Xt}(\mathbf{s}_1), \ldots Z_{Xt}(\mathbf{s}_n)]^T$, $\mathbf{Z}_{Yt} = [Z_{Yt}(\mathbf{s}_1), \ldots Z_{Yt}(\mathbf{s}_n)]^T$ are also Gaussian. In this

section we isolate the difference in their correlation functions. Assume both processes permit a Karhunen–Loéve (KL) decomposition

$$Z_{Xt}(\mathbf{s}) = \sum_{l=1}^{L} \phi_l(\mathbf{s}) x_{tl} \quad \text{and} \quad Z_{Yt}(\mathbf{s}) = \sum_{l=1}^{L} \phi_l(\mathbf{s}) y_{tl}, \tag{3}$$

where the $\phi_l(\mathbf{s})$ are orthonormal functions, and x_{tl} and y_{tl} are normal and pairwise independent loadings with mean zero and $\mathrm{Var}(x_{tl}) = f_l$ and $\mathrm{Var}(y_{tl}) = g_l$. The full KL representation has infinitely many terms, but we choose L to be a large and finite integer to produce a reasonable approximation. The covariance functions are

$$\begin{aligned} \mathrm{Cov}\big[Z_{Xt}(\mathbf{s}_i), Z_{Xt}(\mathbf{s}_j)\big] &= \sum_{l=1}^{L} \phi_l(\mathbf{s}_i)\phi_l(\mathbf{s}_j) f_l \quad \text{and} \\ \mathrm{Cov}\big[Z_{Yt}(\mathbf{s}_i), Z_{Yt}(\mathbf{s}_j)\big] &= \sum_{l=1}^{L} \phi_l(\mathbf{s}_i)\phi_l(\mathbf{s}_j) g_l. \end{aligned} \tag{4}$$

We use the same basis functions $\phi_l(\mathbf{s})$ for both $Z_{Xt}(\mathbf{s})$ and $Z_{Yt}(\mathbf{s})$. This assumption is common when using a known basis expansion such as wavelets or Fourier expansions but requires scrutiny in other cases such as empirically orthogonal functions (EOF) estimated using principal components analysis (PCA). We consider the basis functions

$$\phi_l(\mathbf{s}) = \begin{cases} \cos(\mathbf{s}'\omega_{l/2}) & l \text{ is even}, \\ \sin(\mathbf{s}'\omega_{\lceil l/2 \rceil}) & l \text{ is odd}, \end{cases}$$

where $\omega = (\omega_1, \ldots, \omega_n)$ is the set of Fourier frequencies of the form $(2\pi j_1/n_1, 2\pi j_2/n_2)$ for $(j_1, j_2) \in \mathcal{S}(n_1, n_2)$, where the data are observed within the regular grid of points $\mathcal{S}(n_1, n_2) = \{0, 1, \ldots, n_1 - 1\} \times \{0, 1, \ldots, n_2 - 1\}$. For the variances f_l and g_l we consider the spectral density of the stationary and isotropic Matérn covariance on lattice (Guinness and Fuentes (2017))

$$f(\omega_l) = f_l(\omega_{l1}, \omega_{l2}) = \frac{\sigma_f^2}{\{1 + (\frac{\alpha_f}{\delta})^2[\sin^2(\frac{\delta\omega_{l1}}{2}) + \sin^2(\frac{\delta\omega_{l2}}{2})]\}^{\nu_f + 1}}.$$

Here, $\sigma_f^2 = 1$ because σ_f^2 represents the variance of $Z_{Xt}(\mathbf{s})$ which is fixed at one to identify the scale of the marginal distribution $F_{\mathbf{s}}$, ν_f is the smoothness parameter and α_f is the range parameter. The variance g_l has the same form as f_l with parameters $\sigma_g^2 = 1$, ν_g and α_g. The lattice spacing is $\delta = 1$ for our dataset.

We also consider the nonstationary principal component decomposition under Gaussian and skew-t assumptions. In this case $\phi_l(\mathbf{s})$ are the eigenvectors of the average of sample covariance matrices for $\mathbf{Z}_{Xt}$ and $\mathbf{Z}_{Yt}$, and f_l, g_l are the corresponding eigenvalues. To approximate the covariances of $\mathbf{Z}_{Xt}$ and $\mathbf{Z}_{Yt}$, we first subtract the site's mean and divide by its standard deviation and then take the eigen decomposition of their sample covariances. We keep all the eigenvectors since our

focus is not on selection of the number L. The procedure remains the same except for the use of empirical basis functions.

Often, the climate model output represents averages over subregions while observed data are point measurements. If the climate model output is recorded as areal averages in grid squares B_j, they can be represented as

$$X_{tj} = \int_{B_j} X_t(s)\,ds = \sum_{l=1}^{L}\left[\int_{B_j} \phi_l(s)\,ds\right] x_{tl} = \sum_{l=1}^{L} \tilde{\phi}_{lj} x_{tl}$$

which has the same form as (3) except with integrated basis functions $\tilde{\phi}_{lj} = \int_{B_j} \phi_l(s)\,ds$. Similarly, the covariance between a pair of grid cell averages is $\mathrm{Cov}(X_{ti}, X_{tj}) = \sum_{l=1}^{L} \tilde{\phi}_{li}\tilde{\phi}_{lj} f_l$ which has the same form as (4). In particular the block covariance remains a linear combination of the variances f_l. With the same form for X_{tj} as in (3) and the same form for $\mathrm{Cov}(X_{ti}, X_{tj})$ as in (4), the change of support problem can be handled naturally within the KL representation. Bradley, Wikle and Holan (2015) also consider the change of support problem with KL basis expansion.

3.3. *Calibration*. Conditional on the parameters in the marginal distributions $F_{\mathbf{s}}$ and $G_{\mathbf{s}}$, the variances f_l and g_l and the latent x_{tl} and y_{tl}, the standardized climate model output is calibrated as $\tilde{Z}_{Xt}(\mathbf{s}) = \sum_{l=1}^{L} w_l \phi_l(\mathbf{s}) x_{tl}$, where $w_l = \sqrt{g_l/f_l}$. With the weight w_l the covariance of $\tilde{Z}_{Xt}(\mathbf{s})$ matches the covariance of $Z_{Yt}(\mathbf{s})$ since $\mathrm{Cov}[\tilde{Z}_{Xt}(\mathbf{s}_i), \tilde{Z}_{Xt}(\mathbf{s}_j)] = \sum_{l=1}^{L} w_l^2 \phi_l(\mathbf{s}_i)\phi_l(\mathbf{s}_j) f_l = \mathrm{Cov}[Z_{Yt}(\mathbf{s}_i), Z_{Yt}(\mathbf{s}_j)]$. The calibrated climate model output is $\tilde{X}_t(\mathbf{s}) = G_{\mathbf{s}}^{-1}\{\Phi[\tilde{Z}_{Xt}(\mathbf{s})]\}$ which has the same marginal distribution as $Y_t(\mathbf{s})$. This procedure matches the distribution of $Y_t(\mathbf{s})$ not the noisy measurements $Y_t^*(\mathbf{s})$. If matching the $Y_t^*(\mathbf{s})$ distribution were the objective, we would add $N(0, \tau_Y^2)$ error to $\tilde{X}_t(\mathbf{s})$.

If both data sources are Gaussian, then we have $Z_{Xt}(\mathbf{s})$ as in (1). Adding back the marginal (linear) transformation to $\tilde{Z}_{Xt}(\mathbf{s})$ results in the final calibrated climate model output with the same marginal distribution and spatial correlation as $Y_t(\mathbf{s})$:

$$\tilde{X}_t(\mathbf{s}) = \mu_Y(\mathbf{s}) + \sigma_Y(\mathbf{s}) \sum_{l=1}^{L} w_l \phi_l(\mathbf{s}) x_{tl}. \tag{5}$$

If both data sources follow a skew-t distribution, then the calibrated climate model output is

$$\tilde{X}_t(\mathbf{s}) = \mu_Y(\mathbf{s}) + \xi_Y |r_{Yt}| + \sigma_{Yt}\sigma_Y(\mathbf{s}) \sum_{l=1}^{L} w_l \phi_l(\mathbf{s}) x_{tl}.$$

To carry the bias adjustment on to RCP8.5, we assume the biases between model outputs and gridded observed data are stationary in time. That is, the mean difference between future observations and RCP8.5 remains the same as $\delta(\mathbf{s}) =$

$\mu_Y(\mathbf{s}) - \mu_X(\mathbf{s})$, the ratio of standard deviations remains $R(\mathbf{s}) = \sigma_Y(\mathbf{s})/\sigma_X(\mathbf{s})$ and the appropriate covariance weight remains $w_l = \sqrt{g_l/f_l}$. Then, under the Gaussian assumption the calibrated RCP8.5 becomes

$$\big[\mu_{\mathrm{RCP}}(\mathbf{s}) + \delta(\mathbf{s})\big] + \big[R(\mathbf{s})\sigma_{\mathrm{RCP}}(\mathbf{s})\big] w_l Z_{\mathrm{RCP}}(\mathbf{s}),$$

where $\delta(\mathbf{s})$, $R(\mathbf{s})$ and w_l are all adjustment factors, and $\mu_{\mathrm{RCP}}(\mathbf{s})$, $\sigma_{\mathrm{RCP}}(\mathbf{s})$ and $Z_{\mathrm{RCP}}(\mathbf{s})$ are defined similarly as for $X_t(\mathbf{s})$ and $Y_t(\mathbf{s})$. The term $\mu_{\mathrm{RCP}}(\mathbf{s}) + \delta(\mathbf{s})$ represents the mean of future observations, and $R(\mathbf{s})\sigma_{\mathrm{RCP}}(\mathbf{s})$ represents the standard deviation of future observations. The formula for the skew-t distribution can be written similarly with extra parameters.

4. Computing details. Let $\mathbf{X}_t = [X_t(\mathbf{s}_1), \ldots, X_t(\mathbf{s}_n)]^T$ denote the climate model outputs and $\mathbf{Y}_t = [Y_t(\mathbf{s}_1), \ldots, Y_t(\mathbf{s}_n)]^T$ denote the gridded observed data over the n spatial locations for time t; here, we assume the same n locations for both data sources though this assumption is not required by our approach. If the Gaussian assumption is appropriate and a nugget is included, then we have

$$\mathbf{X}_t \sim N\big(\boldsymbol{\mu}_X, D_X\big(\Phi U \Phi' + \tau_X^2 I_n\big) D_X\big) \quad \text{and}$$

$$\mathbf{Y}_t \sim N\big(\boldsymbol{\mu}_Y, D_Y\big(\Phi V \Phi' + \tau_Y^2 I_n\big) D_Y\big),$$

where $U = \mathrm{diag}(f_1, \ldots, f_L)$, $V = \mathrm{diag}(g_1, \ldots, g_L)$, Φ is an $n \times L$ matrix with $\Phi_{ij} = \phi_j(\mathbf{s}_i)$, $D_X = \mathrm{diag}[\sigma_X(\mathbf{s}_1), \ldots, \sigma_X(\mathbf{s}_n)]$ and $D_Y = \mathrm{diag}[\sigma_Y(\mathbf{s}_1), \ldots, \sigma_Y(\mathbf{s}_n)]$. The spectral densities f_l and g_l and basis functions $\phi_i(\mathbf{s}_j)$ are given in Section 3.2. The MCMC algorithm (Brooks et al. (2011)) involves a Gibbs sampling step for conjugate updates of $\boldsymbol{\mu}_X$, $\boldsymbol{\mu}_Y$, the random effects x_{tl}, y_{tl}, a Metropolis–Hasting step for updating τ_X^2, τ_Y^2 and the parameters in U and V. The random effects are obtained from the posterior distribution of $x_t = [x_{t1}, \ldots, x_{tL}]^T$ under the model $\mathbf{X}_t \sim N(\boldsymbol{\mu}_X + \Phi x_t, \tau_X^2 I_n)$. The posterior predictive distribution of $\tilde{\mathbf{X}}_t$ is then obtained within each MCMC iteration.

Under the skew-t setting we have $\mathbf{X}_t \sim N(\boldsymbol{\mu}_X + \xi_X |r_{Xt}|, D_X[\sigma_{Xt}^2(\Phi U \Phi' + \tau_X^2 I_n)]D_X)$ and $\mathbf{Y}_t \sim N(\boldsymbol{\mu}_Y + \xi_Y |r_{Yt}|, D_Y[\sigma_{Yt}^2(\Phi V \Phi' + \tau_Y^2 I_n)]D_Y)$. The MCMC algorithm under this assumption also involves a Gibbs sampling step for conjugate updates of the parameters in the mean structure as well as σ_{Xt}^2, σ_{Yt}^2, the random effects x_{tl}, y_{tl} and a Metropolis–Hasting step for updating τ_X^2, τ_Y^2 and the parameters in U and V.

Spatial Gaussian priors with exponential covariance function are assigned for the mean vectors $\boldsymbol{\mu}_X = [\mu_X(\mathbf{s}_1), \ldots, \mu_X(\mathbf{s}_n)]^T$ and $\boldsymbol{\mu}_Y = [\mu_Y(\mathbf{s}_1), \ldots, \mu_Y(\mathbf{s}_n)]^T$ and for $\boldsymbol{\theta}_X = \log(\boldsymbol{\sigma}_X)$ and $\boldsymbol{\theta}_Y = \log(\boldsymbol{\sigma}_Y)$, where $\boldsymbol{\sigma}_X = [\sigma_X(\mathbf{s}_1), \ldots, \sigma_X(\mathbf{s}_n)]^T$ and $\boldsymbol{\sigma}_Y = [\sigma_Y(\mathbf{s}_1), \ldots, \sigma_Y(\mathbf{s}_n)]^T$. We assign uniform prior within $(0, M_\nu)$ for ν_X, ν_Y and another uniform prior within $(0, M_\tau)$ for τ_X^2 and τ_Y^2. Normal priors are specified for ξ_X, ξ_Y with mean 0 and standard deviation M_ξ. Uniform hyperpriors with lower bound 0 and upper bounds 60, 20, 5 are specified for the hyperparameters M_ν, M_τ, M_ξ in the prior distributions respectively.

5. Application. In this section, we apply our method to calibrate the 30-year hindcast climate model outputs of monthly average temperature and precipitation data from 1981 to 2010 during summer (described in Section 2) and generate calibrated RCP8.5 from 2020 to 2034 to study future climate. We work with log of precipitation and transform back to the original scale for all plots and projections to avoid negative values. We ran our models for 12,000 MCMC iterations with the first 2000 as burn in.

5.1. *Model comparisons.* To compare approaches under different assumptions for the marginal distribution (Gaussian or skew-t) and covariance (Fourier or PCA), we use all model outputs (1981–2010) and gridded observed data from 1950 to 1989 as training data and the last 10 years of gridded observed data (1990–1999) as test data. Since the GCM predictions do not have temporal correspondence with gridded observed data, the overlap in the years poses no conflict in the analysis. We include another model (denoted as "normal*" in Tables 1 and 2) which is the same as (5) but without the weight w_l to demonstrate its importance. This model adjusts the marginal distributions but not the spatial correlation. Notice the PCA normal*, PCA normal, Fourier normal* and Fourier normal approaches all have the same form of marginal distribution, but different estimating procedures may cause slightly different results under each setting. We use fixed eigenvalues for f_l and g_l under the PCA setting, while under the Fourier setting we update the parameters in spectral densities f_l and g_l within each MCMC iteration. We obtain $\tilde{X}_t(\mathbf{s})$ from equation (5); $\tilde{X}_t(\mathbf{s}) = \mu_Y(\mathbf{s}) + \sigma_Y(\mathbf{s}) \sum_{l=1}^{L} w_l \phi_l(\mathbf{s}) x_{tl}$, and $\tilde{X}_t(\mathbf{s}) = \mu_Y(\mathbf{s}) + \sigma_Y(\mathbf{s}) \sum_{l=1}^{L} \phi_l(\mathbf{s}) x_{tl}$ for normal*.

To assess the marginal fit, we compute the mean squared error (MSE) of the predictive mean and standard deviation by comparing them to the sample mean and standard deviation of the test data. Let $\bar{\mu}_Y(\mathbf{s})$ be the sample mean of observations $Y_t(\mathbf{s})$ at location $\mathbf{s}$ and $\tilde{X}_t^m(\mathbf{s})$ be the calibrated model output of ensemble member m, where $m = 1, \ldots, 9$. We calculate $\mathrm{MSE}(\hat{\mu}^m) = \frac{1}{n} \sum_{i=1}^{n} [\bar{\mu}_Y(\mathbf{s}_i) - \frac{1}{T} \sum_{t=1}^{T} \tilde{X}_t^m(\mathbf{s}_i)]^2$ for each of the nine ensemble members and report the median and standard deviation of $\mathrm{MSE}(\hat{\mu}_1), \ldots, \mathrm{MSE}(\hat{\mu}_9)$ in Table 1. We define $\mathrm{MSE}(\hat{\sigma}_1), \ldots, \mathrm{MSE}(\hat{\sigma}_9)$ for the standard deviation of $Y_t(\mathbf{s})$ in a similar way and report the median and standard deviation in Table 1 as well. We also use the integrated quadratic distance (Thorarinsdottir, Gneiting and Gissibl (2013)) to compare the posterior predictive cumulative distribution F to the empirical cumulative distribution G as

$$d_{\mathrm{IQ}}(F, G) = \int_{-\infty}^{\infty} \big[F(x) - G(x)\big]^2 \, dx. \tag{6}$$

We compute integrated quadratic distance for each observation and report the average over space. We also calculated the AUC for predicted exceedances of high thresholds, where AUC denotes the area under the receiver operating characteristic (ROC) curve. For temperature we consider the performance for classifying

TABLE 1

Median (*standard deviation*) *of* MSE($\hat{\mu}$), MSE($\hat{\sigma}$), *integrated quadratic distance* (*IQ dist*) *and AUC* (*all values multiplied by* 100) *for test set prediction using climate model outputs* (9 *ensemble members*) *for* (a) *temperature and* (b) *precipitation for assessment of marginal fit* (*averaged over space and time*)

		MSE($\hat{\mu}$)	MSE($\hat{\sigma}$)	IQ dist	AUC
(a)	Model outputs	362.6 (33.0)	92.7 (9.3)	147.0 (3.3)	82.3 (3.1)
	Schaake shuffle	9.3 (0.3)	3.4 (0.1)	101.4 (0.1)	85.2 (2.3)
	PCA (normal*)	10.3 (0.1)	4.0 (0.3)	102.3 (0.3)	84.6 (2.9)
	PCA (normal)	8.4 (0.1)	2.3 (0.3)	98.6 (1.0)	86.7 (2.5)
	PCA (skew-t)	5.8 (0.3)	1.5 (0.2)	96.2 (1.3)	88.5 (3.6)
	Fourier (normal*)	8.9 (0.3)	3.1 (0.3)	99.7 (1.9)	85.7 (3.7)
	Fourier (normal)	6.9 (0.3)	1.7 (0.2)	97.9 (2.1)	88.4 (4.1)
	Fourier (skew-t)	5.9 (0.4)	1.4 (0.2)	96.8 (2.2)	89.3 (4.6)
(b)	Model outputs	322.8 (9.6)	65.2 (6.0)	97.8 (0.8)	72.7 (6.3)
	Schaake shuffle	18.2 (0.5)	16.5 (0.3)	71.1 (0.1)	75.3 (3.5)
	PCA (normal*)	19.1 (2.0)	17.0 (2.0)	71.2 (0.2)	71.2 (3.4)
	PCA (normal)	12.4 (0.3)	13.9 (1.0)	68.1 (0.7)	80.8 (2.9)
	PCA (skew-t)	9.9 (1.2)	12.9 (0.5)	67.6 (0.8)	81.6 (3.0)
	Fourier (normal*)	17.2 (2.2)	14.9 (1.7)	72.2 (0.9)	75.4 (3.1)
	Fourier (normal)	10.4 (1.7)	11.1 (0.3)	71.3 (1.0)	82.2 (4.3)
	Fourier (skew-t)	9.6 (1.7)	11.4 (0.3)	70.5 (1.0)	82.5 (3.7)

above or below 30°C; for precipitation we consider the performance for classifying above or below 15.24 cm. These two values are roughly the 99th percentile of temperature and precipitation among all locations respectively. Therefore, better results for classification indicate that we can better model the threshold exceedances. Specifically, for temperature we calculate $\frac{1}{T}\sum_{t=1}^{T} I\{X_t^m(\mathbf{s}) > 30\}$, the proportion of $X_t^m(\mathbf{s})$ above 30°C for location $\mathbf{s}$ from ensemble member m of model output and use this proportion to predict whether the gridded observed data $Y_t(\mathbf{s})$ is above or below 30°C for each t. By varying the threshold on the posterior probability used to classify the prediction as above or below 30°C, we can generate the ROC curve and calculate AUC. For each ensemble member m we calculate $\text{AUC}^m(t)$ of time t and take the average over time to obtain AUC^m. We report the median and standard deviation of $\text{AUC}^1, \ldots, \text{AUC}^9$ in Table 1. Everything is the same for precipitation, except we consider above or below 15.24 cm for prediction.

We report the median and standard deviation (both multiplied by 100) for the results of the marginal fit from the nine ensemble members in Table 1. The skew-t provides a better marginal fit than the normal distribution for all settings, but the normal distribution also works well. The model without the weight w_l (normal*) still helps to adjust biases in model outputs, but its performance is clearly inferior to the models with w_l. For the skew-t models the choice of covariance model has little effect on the marginal performance, as expected. Figure 4 displays quantile

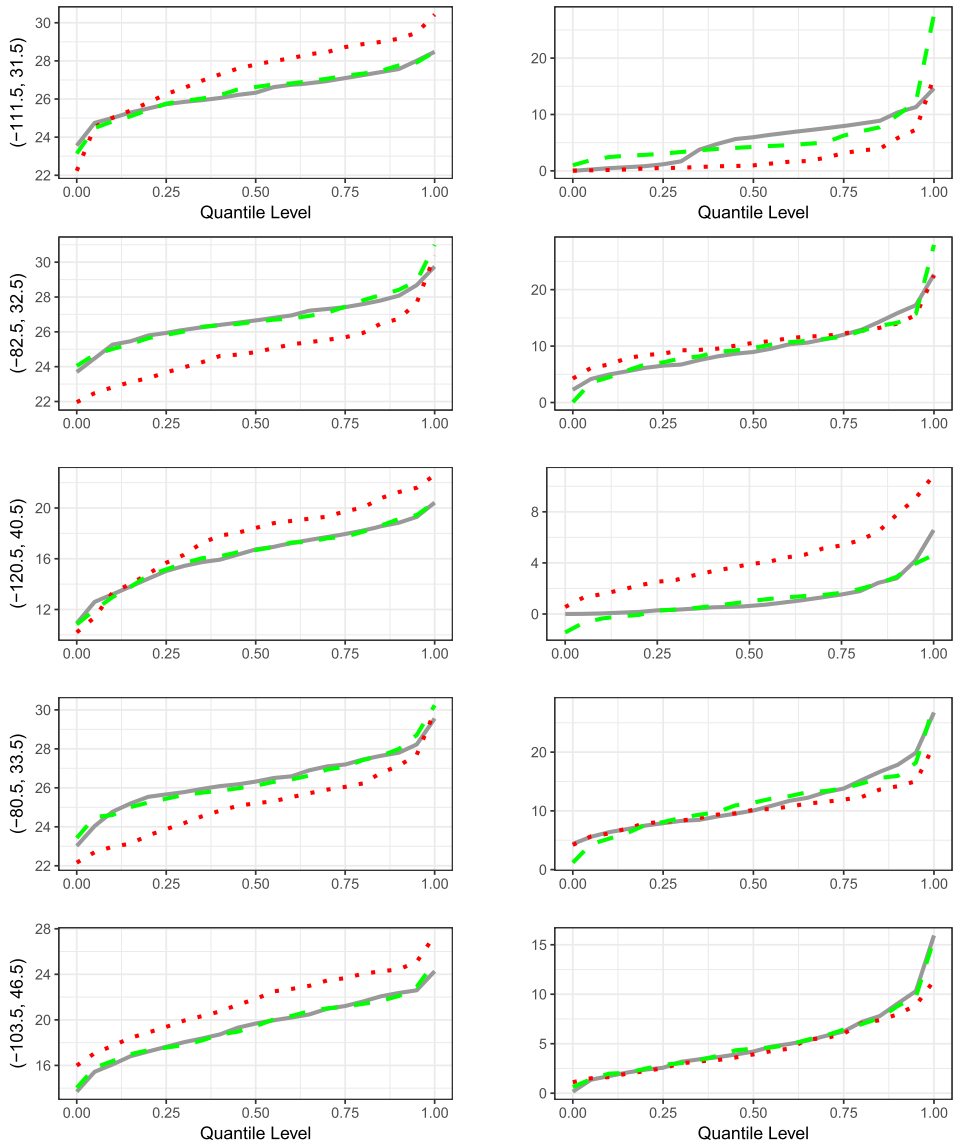
TABLE 2
Median and standard deviation (multiplied by 1000) for test set average of squared difference between semivariograms of calibrated model outputs (nine ensemble members) and gridded observed data for (a) *temperature and* (b) *precipitation over distance h for assessment of spatial fit*

		0–50 km	51–100 km	101–500 km	500–1000 km
(a)	Model outputs	1455.9 (21.1)	783.6 (9.4)	864.3 (50.3)	4111.4 (1223.9)
	Schaake shuffle	795.2 (17.5)	775.6 (10.3)	471.7 (7.8)	105.1 (10.5)
	PCA (normal*)	5.2 (0.3)	4.6 (0.4)	13.0 (1.6)	30.2 (6.5)
	PCA (normal)	4.1 (0.5)	3.6 (0.5)	10.3 (1.4)	25.9 (7.8)
	PCA (skew-t)	1.7 (0.4)	1.9 (0.5)	9.2 (1.2)	22.8 (2.3)
	Fourier (normal*)	4.3 (0.3)	4.1 (0.7)	11.3 (1.5)	26.5 (5.3)
	Fourier (normal)	3.0 (0.4)	3.4 (0.6)	8.7 (1.2)	19.8 (2.5)
	Fourier (skew-t)	1.5 (0.3)	2.2 (0.8)	7.9 (1.1)	18.8 (2.7)
(b)	Model outputs	152.3 (5.2)	680.6 (16.1)	1588.4 (134.8)	3138.7 (452.9)
	Schaake shuffle	3627.1 (36.8)	4605.8 (160.0)	1157.4 (79.4)	556.8 (85.2)
	PCA (normal*)	40.0 (2.3)	233.5 (14.8)	1196.8 (57.4)	2271.0 (136.8)
	PCA (normal)	12.3 (1.9)	56.8 (8.4)	176.8 (30.3)	852.9 (115.5)
	PCA (skew-t)	10.3 (5.2)	53.5 (5.2)	161.9 (8.4)	824.5 (92.3)
	Fourier (normal*)	35.5 (4.5)	224.5 (12.3)	1115.5 (48.4)	2381.9 (120.6)
	Fourier (normal)	11.0 (1.9)	51.6 (5.8)	98.7 (8.4)	780.0 (126.5)
	Fourier (skew-t)	9.7 (3.2)	49.0 (5.2)	96.1 (7.7)	776.8 (108.4)

plots of gridded observed data, model outputs and one of the calibrated model outputs (Fourier, skew-t) at the same locations chosen for Figure 3. We can see the quantiles of model outputs match with gridded observed data after calibration.

To assess the fit of spatial correlation, we compare the semivariogram between gridded observed data and each approach. Let $\hat{\gamma}_Y(h)$ and $\hat{\gamma}_X(h)$ be the empirical semivariogram over distance h for gridded observed data and climate model outputs respectively. We compute the average of squared difference between $\hat{\gamma}_Y(h)$ and $\hat{\gamma}_X(h)$ for distances within 0 to 50 km, 51 to 100 km, 101 to 500 km and 501 to 1000 km. The results of the spatial fit for the nine ensemble members are given in Table 2 (all values are multiplied by 1000). Our approach works very well in resolving the differences in spatial dependence between the two data sources. For the skew-t model the Fourier covariance generally outperforms the PCA covariance. Figures 2(c) and 2(d) also show the spatial correlations of climate model outputs after calibration (Fourier, skew-t) match much better to the gridded observed data than the original climate model outputs for both temperature and precipitation.

We also compare our method with the Schaake shuffle (Clark et al. (2004)). We apply the Schaake shuffle to nine standardized ensemble members and transform them back to the original scale after shuffling. The historical observations are randomly selected from the same month of all years except the year of model output. As shown in Table 1, the Schaake shuffle method adequately calibrates the



(a) Temperature ($^\circ C$) (b) Precipitation (cm)

FIG. 4. *Quantile plots for gridded observed data* (*test set*; *gray*), *model outputs* (*dotted*) *and calibrated model outputs* (*Fourier*, *skew-t*, *dashed*) *at the five arbitrarily selected locations for temperature and precipitation. Longitude and latitude for each location are given beside the vertical axis in* (a).

marginal distributions. From Table 2 we can see the shuffle method also helps to correct the biases in spatial correlation, but our method produces better results. The shuffle method does not work well within a distance of 100 km, but it manages to perform better between distances of 100–1000 km. We should mention again that we consider monthly data in summer and only select historical observations from the same month for shuffling. The shuffle method should perform better for daily data in which they can select historical observations seven days before and after.

5.2. *Future projections*. We carry the bias adjustment on to RCP8.5 for monthly mean temperature and precipitation in summer under the assumption that the biases between model outputs and gridded observed data are stationary in time, as described in Section 3.3. We choose the normal model for future projection because it works well in adjusting the biases with fewer parameters to be estimated than the skew-t model. We will benefit more from the skew-t model under the presence of extreme events, but for monthly mean data the normal model appears to be sufficient.

Our method introduces regional adjustments, since the bias in the model seems to be space dependent. While on average across space we observe an increasing trend in the mean temperature, it is not consistent at different locations and it varies more in the western United States. Furthermore, rather than adjusting the central tendency we focus on calibrating the entire distributions. Our findings provide insight and understanding regarding which corrections are needed across space and for the lower and upper quantile of temperature and precipitation.

Our adjustments for the upper quantiles are smaller than for the median and lower quantiles. On average the 95th percentile for the calibrated RCP8.5 is 0.4°C lower than for the uncalibrated RCP8.5, while the 50th percentile of the calibrated RCP8.5 is 1.9°C lower than the uncalibrated RCP8.5, and the 5th percentile of the calibrated RCP8.5 is 1.7°C lower than the uncalibrated RCP8.5. Figures 5 and 6 are the time series plots of the 5th, 50th and 95th percentiles of RCP8.5 (temperature and precipitation, four ensemble members each) for the conterminous United States in July before and after calibration (Fourier, normal). All the results indicate the need to introduce a methodology for calibration that is space and quantile dependent.

6. Discussion. In this paper we have introduced a method to adjust for biases of climate model outputs. We present our results under Gaussian and skew-t assumptions, but even more flexible models can be applied to the spatially-varying marginal distributions of the two data sources. Most of the calibration literature considers regression approaches and include model outputs as explanatory variables for estimating the gridded observed data. Mean shifting or quantile mapping are commonly used methods. In our approach we use the KL decomposition to isolate the difference in the correlations of two data sources and successfully rec-

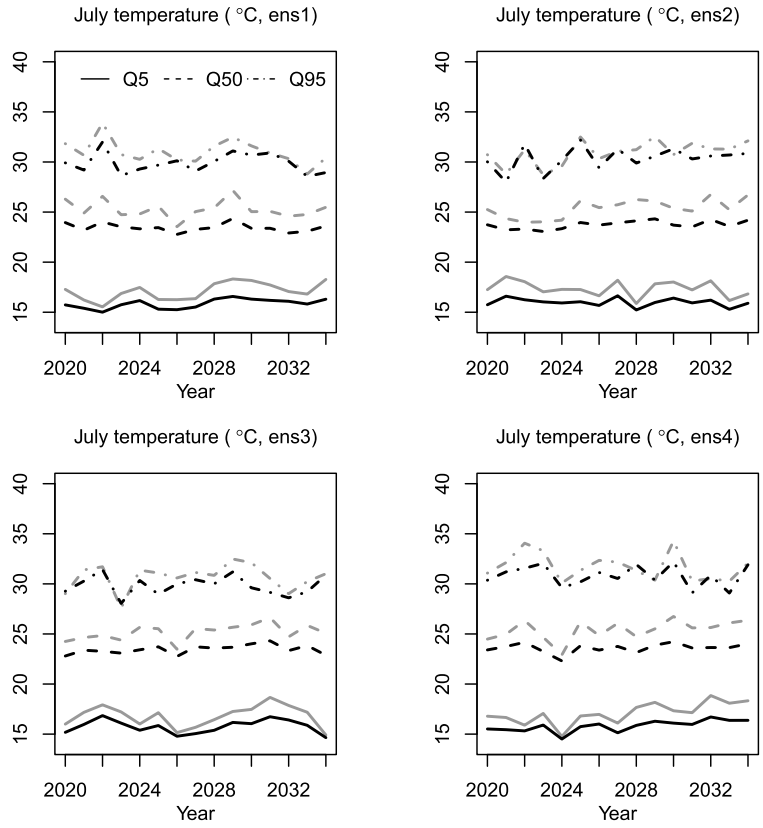


FIG. 5. *Time series plots for the* 5*th*, 50*th and* 95*th percentiles of RCP*8.5 (*four ensemble members*) *for July temperature* (°C) *from* 2020 *to* 2034 *in the United States before* (*gray*) *and after* (*black*) *calibration* (*Fourier*, *normal*).

tify it in addition to matching the marginal distributions. Our results for model comparisons have shown the importance of calibrating spatial dependence.

There are several opportunities for methodological extensions of our approach. We have not considered temporal dependence because it is relatively weak compared to spatial dependence in our data, but accounting for autocorrelation would be crucial for an analysis of daily or subdaily data. It would be possible to include a temporal trend and construct a spatiotemporal model. Another extension would be to model temperature and precipitation simultaneously and develop a multivariate spatial model calibration when we encounter correlated variables. Inclusion of covariates would be another direction to extend our model. It is possible to incorporate covariate information when we model the marginal distributions in (1) or (2). We would expect for a potential improvement to the calibration results with the information of important covariates like elevation or slope for temperature and precipitation data.

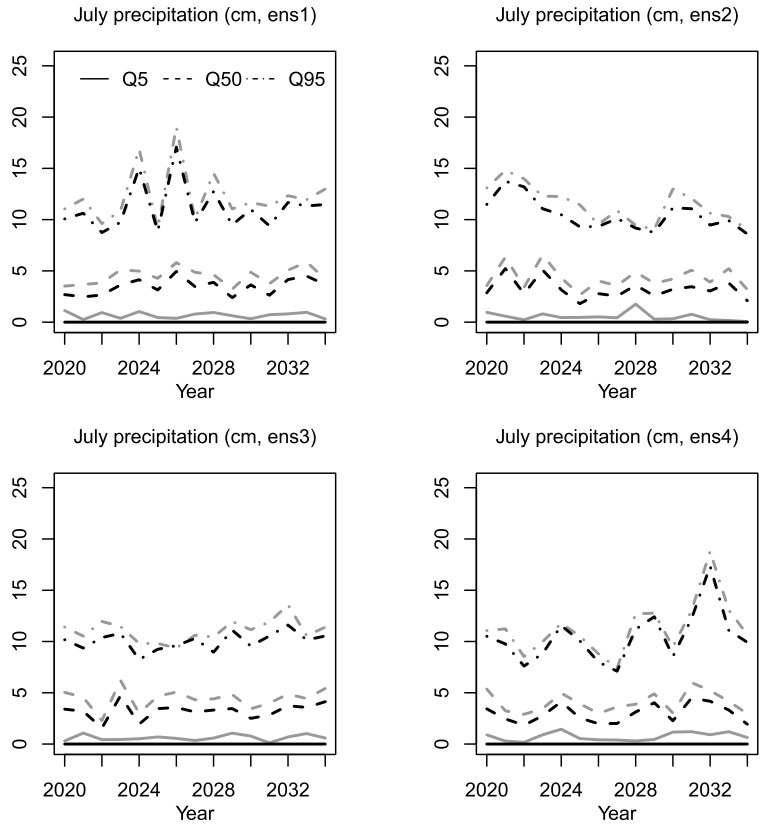


FIG. 6. *Time series plots for the* 5*th*, 50*th and* 95*th percentiles of RCP*8.5 (*four ensemble members*) *for July precipitation* (*cm*) *from* 2020 *to* 2034 *in the United States before* (*gray*) *and after* (*black*) *calibration* (*Fourier*, *normal*).

Acknowledgments. We thank the Editor, the Associate Editors and the reviewers for their constructive and insightful comments in reviewing our work which greatly helped to improve the quality of the manuscript.

REFERENCES

AZZALINI, A. and CAPITANIO, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 367–389. MR1983753

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL.

BERROCAL, V. J., GELFAND, A. E. and HOLLAND, D. M. (2010). A spatio-temporal downscaler for output from numerical models. *J. Agric. Biol. Environ. Stat.* **15** 176–197. MR2787270

BERROCAL, V. J., GELFAND, A. E. and HOLLAND, D. M. (2012). Space-time data fusion under error in computer model output: An application to modeling air quality. *Biometrics* **68** 837–848. MR3055188

BERROCAL, V. J., RAFTERY, A. E. and GNEITING, T. (2007). Combining spatial statistical and ensemble information for probabilistic weather forecasting. *Mon. Weather Rev.* **135** 1386–1402.

BERROCAL, V. J., RAFTERY, A. E. and GNEITING, T. (2008). Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *Ann. Appl. Stat.* **2** 1170–1193. MR2655654

BHOWMIK, R. D., SANKARASUBRAMANIAN, A., SINHA, T., PATSKOSKI, J., MAHINTHAKUMAR, G. and KUNKEL, K. (2017). Multivariate downscaling approach preserving cross-correlations across climate variables for projecting hydrologic fluxes. *J. Hydrometeorol.* **18** 2187–2205.

BRADLEY, J. R., WIKLE, C. K. and HOLAN, S. H. (2015). Spatio-temporal change of support with application to American Community Survey multi-year period estimates. *Stat* **4** 255–270. MR3414659

BROOKS, S., GELMAN, A., JONES, G. L. and MENG, X.-L., eds. (2011). Handbook of Markov Chain Monte Carlo. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR2742422

CANNON, A. J. (2016). Multivariate bias correction of climate model output: Matching marginal distributions and intervariable dependence structure. *J. Climate* **29** 7045–7064.

CLARK, M., GANGOPADHYAY, S., HAY, L., RAJAGOPALAN, B. and WILBY, R. (2004). The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeorol.* **5** 243–262.

CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR1239641

CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2848400

DETTINGER, M. D., CAYAN, D. R., MEYER, M. K. and JETON, A. E. (2004). Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American river basins, Sierra Nevada, California, 1900–2099. *Clim. Change* **62** 283–317.

DEVINENI, N. and SANKARASUBRAMANIAN, A. (2010). Improved categorical winter precipitation forecasts through multimodel combinations of coupled GCMs. *Geophys. Res. Lett.* **37** L24704.

FELDMANN, K., SCHEUERER, M. and THORARINSDOTTIR, T. L. (2015). Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression. *Mon. Weather Rev.* **143** 955–971.

FUENTES, M. and RAFTERY, A. E. (2005). Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics* **61** 36–45. MR2129199

GEL, Y., RAFTERY, A. E. and GNEITING, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation method. *J. Amer. Statist. Assoc.* **99** 575–583. MR2086380

GOTWAY, C. A. and YOUNG, L. J. (2002). Combining incompatible spatial data. *J. Amer. Statist. Assoc.* **97** 632–648. MR1951636

GUINNESS, J. and FUENTES, M. (2017). Circulant embedding of approximate covariances for inference from Gaussian data on large lattices. *J. Comput. Graph. Statist.* **26** 88–97. MR3610410

HAY, L. E., WILBY, R. L. and LEAVESLEY, G. H. (2000). A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States. *J. Am. Water Resour. Assoc.* **36** 387–397.

JONES, M. C. and FADDY, M. J. (2003). A skew extension of the t-distribution, with applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 159–174. MR1959820

LUERS, A. L., CAYAN, D. R., FRANCO, G., HANEMANN, M. and CROES, B. (2006). Our changing climate, assessing the risks to California. Report No. CEC-500-2006-077, 1–16. California Energy Commissionm Sacramento, CA.

MAURER, E. P., WOOD, A. W., ADAM, J. C., LETTENMAIER, D. P. and NIJSSEN, B. (2002). A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States. *J. Climate* **15** 3237–3251.

MCMILLAN, N. J., HOLLAND, D. M., MORARA, M. and FENG, J. (2010). Combining numerical model output and particulate data using Bayesian space-time modeling. *Environmetrics* **21** 48–65. MR2842223

MÖLLER, A., LENKOSKI, A. and THORARINSDOTTIR, T. L. (2012). Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Q. J. R. Meteorol. Soc.* **139** 982–991.

MORRIS, S. A., RIECH, B. J., THIBAUD, E. and COOLEY, D. (2017). A space-time skew-*t* model for threshold exceedances. *Biometrics* **73** 749–758. MR3713109

O'BRIEN, T. P., SORNETTE, D. and MCPHERRO, R. L. (2001). Statistical asynchronous regression determining: The relationship between two quantities that are not measured simultaneously. *J. Geophys. Res.* **106** 13247–13259.

REICH, B. J., CHANG, H. H. and FOLEY, K. M. (2014). A spectral method for spatial downscaling. *Biometrics* **70** 932–942. MR3295754

SCHEFZIK, R. (2017). Ensemble calibration with preserved correlations: Unifying and comparing ensemble copula coupling and member-by-member postprocessing. *Q. J. R. Meteorol. Soc.* **143** 999–1008.

SCHEFZIK, R., THORARINSDOTTIR, T. L. and GNEITING, T. (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statist. Sci.* **28** 616–640. MR3161590

SEO, S. B., SINHA, T., MAHINTHAKUMAR, G., SANKARASUBRAMANIAN, A. and KUMAR, M. (2016). Identification of dominant source of errors in developing streamflow and groundwater projection under near-term climate change. *J. Geophys. Res., Atmos.* **121** 7652–7672.

STONER, A. M. K., HAYHOE, K., YANG, X. and WUEBBLES, D. J. (2012). An asynchronous regional regression model for statistical downscaling of daily climate variables. *Int. J. Climatol.* **33** 2473–2494.

TAYLOR, K. E., STOUFFER, R. J. and MEEHL, G. A. (2012). An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93** 485–498.

THORARINSDOTTIR, T. L., GNEITING, T. and GISSIBL, N. (2013). Using proper divergence functions to evaluate climate models. *SIAM/ASA J. Uncertain. Quantificat.* **1** 522–534. MR3283896

VAN VUUREN, D. P., EDMONDS, J., KAINUMA, M., RIAHI, K., THOMSON, A., HIBBARD, K., HURTT, G. C., KRAM, T., KREY, V., LAMARQUE, J.-F., MASUI, T., MEINSHAUSEN, M., NAKICENOVIC, N., SMITH, S. J. and ROSE, S. K. (2011). The representative concentration pathways: An overview. *Clim. Change* **109** 5–31.

VRAC, M. and FRIEDERICHS, P. (2015). Multivariate—intervariable, spatial, and temporal—bias correction. *J. Climate* **28** 218–237.

WIKLE, C. K. and BERLINER, L. M. (2005). Combining information across spatial scales. *Technometrics* **47** 80–91. MR2099410

WOOD, A. W., LEUNG, L. R., SRIDHAR, V. and LETTENMAIER, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Clim. Change* **62** 189–216.

YOUNG, L. J. and GOTWAY, C. A. (2007). Linking spatial data from different sources: The effects of change of support. *Stoch. Environ. Res. Risk Assess.* **21** 589–600. MR2380677

ZIDEK, J. V., LE, N. D. and LIU, Z. (2012). Combining data and simulated data for space-time fields: Application to ozone. *Environ. Ecol. Stat.* **19** 37–56. MR2909084

Y.-N. HUANG
B. J. REICH
DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
SAS HALL
RALEIGH, NORTH CAROLINA 26795-8203
USA
E-MAIL: yenning0619@gmail.com
bjreich@ncsu.edu

M. FUENTES
COLLEGE OF HUMANITIES AND SCIENCES
VIRGINIA COMMONWEALTH UNIVERSITY
828 WEST FRANKLIN STREET
RICHMOND, VIRGINIA 23284-2019
USA
E-MAIL: mfuentes@vcu.edu

A. SANKARASUBRAMANIAN
DEPARTMENT OF CIVIL, CONSTRUCTION
AND ENVIRONMENTAL ENGINEERING
NORTH CAROLINA STATE UNIVERSITY
BOX 7908 MANN HALL
RALEIGH, NORTH CAROLINA 27695-7908
USA
E-MAIL: sarumug@ncsu.edu@ncsu.edu