Biometrics , 1–24
2014

A spectral method for spatial downscaling

Brian J Reich

North Carolina State University

Howard H Chang

Emory University

and

Kristen M Foley

US Environmental Protection Agency

Summary: Complex computer models play a crucial role in air quality research. These models are used to evaluate potential regulatory impacts of emission control strategies and to estimate air quality in areas without monitoring data. For both of these purposes, it is important to calibrate model output with monitoring data to adjust for model biases and improve spatial prediction. In this paper, we propose a new spectral method to study and exploit complex relationships between model output and monitoring data. Spectral methods allow us to estimate the relationship between model output and monitoring data separately at different spatial scales, and to use model output for prediction only at the appropriate scales. The proposed method is computationally efficient and can be implemented using standard software. We apply the method to compare Community Multiscale Air Quality (CMAQ) model output with ozone measurements in the United States in July, 2005. We find that CMAQ captures large-scale spatial trends, but has low correlation with the monitoring data at small spatial scales.
Key words: Computer model output; Data fusion; Kriging; Multiscale analysis.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Research on the impacts of air quality and meteorological conditions on human health, economics, and the environment have benefited considerably from the availability of routine monitoring measurements. However monitoring networks are typically spatially sparse, preferentially located (Stuart et al., 2009), and often without complete daily measurements (Kim et al., 2013). Reliance on monitor measurements not only restricts a study's geographical region, but can also result in exposure uncertainty in the risk assessment process. Consequently, there is a growing interest in supplementing monitor measurements with additional information to increase the availability of air-quality data across space and time. Recent data fusion applications have focused particularly on simulation outputs from deterministic computer models that can provide broad spatial-temporal coverage without missing values (Fuentes and Raftery, 2005; McMillan et al., 2009; Paciorek, 2012). Generally, we refer to these supplemental data as proxy data, which may be computer model output, satellite observations, land-use variables, etc.

Air-quality simulations from computer models are known to exhibit bias due to errors in input variables, as well as inadequate mathematical representation of the underlying environmental process (Mebust et al., 2003; Lim et al., 2010). Another challenge in combining air-quality information arises from the different spatial resolutions between observations and computer models (or more generally, proxy data). Specifically, numerical model outputs are provided as average values over contiguous grid cells, while monitoring observations are only available at point locations. To address the complex missing data and spatial misalignment problems, current statistical approaches predominantly view proxy data as predictors of observations in a regression setting. After linking each monitor location to a model grid cell, associations between proxy values and observations are modeled as continuous spatial processes to resolve the bias at the point-level. This framework, known as statistical down-

scaling, allows us to calibrate model outputs at any spatial point location within a grid cell. Statistical downscaling has been successfully applied to air quality simulations (Berrocal et al., 2010b,a; Zhou et al., 2011), climate model outputs (Berrocal et al., 2013; Zhou et al., 2012), and remotely-sensed satellite images (Liu et al., 2009; Kloog et al., 2011).

The main contribution of this paper is the development of a multiscale statistical downscaler that allows distinct associations between observations and proxy data at different spatial resolutions. Nguyen et al. (2012) and Crooks and Isakov (2013) propose multiscale spatial models for the true value of the process of interest, and then relate the true process to multiple data sources while accounting for biases in each data source. Our approach is different in that we explicitly model the relationship between monitor and proxy data at different resolutions, and thus obtain a more comprehensive study of the performance of the proxy. Spatial multiscale modeling has also appeared in several applications other than downscaling. For example, Nychka et al. (2012) use nested tapered bivariate splines to model climate variables, while Morris et al. (2003) use wavelets to model functional image data. In contrast, we utilize the spectral representation of spatial processes because the environmental fields we are working with do not frequently exhibit sharp spatial gradients, where a wavelet approach is more appropriate. Also, our spectral model reduces to the usual universal Kriging model as a special case, which is commonly used in this field and provides optimal spatial prediction in certain settings.

The proposed spectral downscaler offers several key advantages compared to previous data fusion methodology. First, we utilize proxy data in the predictive model only at the appropriate spatial scales. Second, the spectral representation reduces computation considerably compared to approaches where the latent field is modeled using all available data, and avoids the problem where proxy may dominate predictions (Fuentes and Raftery, 2005; Paciorek, 2012). Finally, by considering associations at different spatial scales, we borrow information

across groups of contiguous grid cells to predict observations. Recently, Berrocal et al. (2012) proposed the use of a weighted average of grid cells around each monitor and showed that smoothed model output has better predictive power, especially when predicting observations at locations that are far from other monitoring sites. Our spectral approach can be viewed as a more general and flexible framework because it can accommodate both smoothing as in Berrocal et al. (2012), but also the opposite case where the proxy's large scale-trends do not match the observations, but its small-scale trends match the observations after removing large-scale trends. It measures and tests for dependence at different scale which provides valuable information about the utility of the proxy, and overcomes the need to select an a priori smoothing parameter for the predictor because the usefulness of the predictor at each spatial resolution can be informed by the observed data.

Our approach is also motivated by a dynamic downscaling technique known as spectral nudging (Von Storch et al., 2000). In dynamic downscaling, outputs from one numerical model are used to drive a second numerical model with higher spatial resolution. For example, in climate research, simulations from general circulation model are often used as boundary conditions for regional climate models. Spectral nudging helps maintain large-scale spatial variation provided by the driving model and has been shown to improve fine-scale prediction performance (Liu et al., 2012; Radu et al., 2008). However, the strength of nudging at different spectral frequency needs to be determined a priori and evaluated via sensitivity analysis (Alexandru et al., 2009). Here we propose a spectral statistical downscaling approach that is completely data-driven and allows the strength of association to vary flexibility across spectral frequencies. In contrast to dynamic downscaling, a statistical downscaler also allows straight-forward inference for uncertainty quantification in the downscaled products.

We apply the spectral downscaler to perform data fusion for daily ground-level ozone concentrations. Ground-level ozone is an air pollutant regulated under the National Ambient

Air Quality Standards by the US Environmental Protection Agency (USEPA). Epidemiological studies have consistently linked ozone exposure to adverse health outcomes, including premature mortality and emergency department visits for cardiovascular and respiratory diseases (Bell et al., 2004; Strickland et al., 2010). In urban settings, sources of ozone precursors include emission from industrial facilities, power generation, and vehicle exhaust. As a photochemical oxidant, ozone levels are particularly sensitive to weather conditions (Thompson et al., 2001). We combine observations from the USEPA national Air Quality System monitoring network and simulations from the Community Multiscale Air Quality (CMAQ) modeling system in July 2005. CMAQ provides daily 3-D predictions of numerous ambient air pollution concentrations based on atmospheric chemistry and physics, meteorology, and emission sources (Byun and Schere, 2006). While CMAQ's main use is to examine emission control strategies for meeting air quality standards (e.g., Foley et al., 2012), calibrated CMAQ model outputs are also being used to derive exposure estimates for health effects and health impact studies (Berrocal et al., 2011; Chang et al., 2012). Currently, the USEPA maintains a publicly available database of ozone fusion products derived from a single-resolution downscaler (http://www.epa.gov/esd/land-sci/lcb/lcb_faqsd.html).

The remainder of this paper is organized as follows. Section 2 describes the multiscale spectral downscaler. In Section 3, we present a simulation study to compare the spectral downscaler with ordinary Kriging and the standard linear downscaler. Results of the ozone data fusion application are given in Section 4. Finally, discussion and future work appear in Section 5.

2. Spectral methods for spatial downscaling

We first provide a brief review of spectral methods for univariate spatial data; for a complete review we refer to Fuentes and Reich (2010). Let $Y(\mathbf{s})$ be a continuous univariate spatial Gaussian process with mean $\mathrm{E}[Y(\mathbf{s})] = 0$, variance $\mathrm{Var}[Y(\mathbf{s})] = \sigma^2 > 0$, and stationary

correlation function $\mathrm{Cor}[Y(\mathbf{s}_1), Y(\mathbf{s}_2)] = \rho(\mathbf{s}_1 - \mathbf{s}_2)$. The spectral representation theorem states that Y can be written as the Fourier transform of Z,

$$Y(\mathbf{s}) = \int_{\mathcal{R}^2} \exp(-i\boldsymbol{\omega}^T\mathbf{s}) Z(\boldsymbol{\omega}) d\boldsymbol{\omega}, \tag{1}$$

where Z is Gaussian with $\mathrm{E}[Z(\boldsymbol{\omega})] = 0$, $\mathrm{Var}[Z(\boldsymbol{\omega})] = \sigma^2 f(\boldsymbol{\omega})$, and Z is independent over frequency $\boldsymbol{\omega} \in \mathcal{R}^2$. The spatial correlation is determined by the spectral density $f(\boldsymbol{\omega})$ (which we assume exists), which satisfies $f(\boldsymbol{\omega}) = f(-\boldsymbol{\omega}) > 0$ and $\int_{\mathcal{R}^2} f(\boldsymbol{\omega}) d\boldsymbol{\omega} = 1$. The covariance is

$$C(\mathbf{h}) = \mathrm{Cov}[Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})] = \sigma^2 \int \cos(\mathbf{h}^T\boldsymbol{\omega}) f(\boldsymbol{\omega}) d\boldsymbol{\omega}. \tag{2}$$

For example, if f is the bivariate normal density, then the covariance is squared exponential; if f is the bivariate t density, then the covariance is Matérn. Bochner's theorem states that there is a one-to-one relationship between the spectral density and the spatial covariance.

2.1 *Spectral methods for spatial downscaling*

Define $X(\mathbf{s})$ as a proxy measure at location $\mathbf{s}$ and $Y(\mathbf{s})$ as the observed measurement. We assume X is observed throughout the spatial domain and Y is observed sparsely. Our goal is to understand the relationship between X and Y at different scales, and to use X to predict Y. We begin by specifying a flexible statistical model for the joint distribution of X and Y, and then inspecting the induced conditional distribution of Y given X which is used for prediction. For notational simplicity we assume both processes have mean zero. Denote the spectral representations of these two processes as

$$\begin{aligned} X(\mathbf{s}) &= \int \exp(-i\boldsymbol{\omega}^T\mathbf{s}) Z_1(\boldsymbol{\omega}) d\boldsymbol{\omega} \\ Y(\mathbf{s}) &= \int \exp(-i\boldsymbol{\omega}^T\mathbf{s}) Z_2(\boldsymbol{\omega}) d\boldsymbol{\omega}, \end{aligned} \tag{3}$$

where Z_k is Gaussian (implying X and Y are Gaussian processes) with $\mathrm{E}[Z_k(\boldsymbol{\omega})] = 0$ and $\mathrm{V}[Z_k(\boldsymbol{\omega})] = \sigma_k^2 f_k(\boldsymbol{\omega})$ for $k = 1, 2$. To capture the potentially complex relationship

between these two processes, we model their correlation in the spectral domain. Define $\mathrm{Cor}[Z_1(\boldsymbol{\omega}), Z_2(\boldsymbol{\omega})] = \phi(\boldsymbol{\omega}) = \phi(-\boldsymbol{\omega}) \in (-1, 1)$, giving cross-covariance

$$\mathrm{Cov}\left[X(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})\right] = \sigma_1\sigma_2 \int \cos(\mathbf{h}^T\boldsymbol{\omega})\sqrt{f_1(\boldsymbol{\omega})f_2(\boldsymbol{\omega})}\phi(\boldsymbol{\omega})d\boldsymbol{\omega}. \tag{4}$$

By allowing the correlation to vary by frequency, we permit complex relationships. For example, if $\phi(\boldsymbol{\omega}) \approx 1$ for small $||\boldsymbol{\omega}||$ and $\phi(\boldsymbol{\omega}) \approx 0$ for large $||\boldsymbol{\omega}||$, then the proxy and observation processes have similar large-scale trends, but disparate small-scale variations.

To understand the resulting predictive model, we now assume that the proxy $X(\mathbf{s})$ is known at all spatial locations and inspect the conditional mean of the Y given X. Since X is observed completely, we obtain $Z_1(\boldsymbol{\omega})$ for all frequencies via the inverse Fourier transform

$$Z_1(\boldsymbol{\omega}) = \int_{\mathcal{R}^2} \exp(i\boldsymbol{\omega}^T\mathbf{s})X(\mathbf{s})d\mathbf{s}. \tag{5}$$

In the spectral domain, we have $\mathrm{E}[Z_2(\boldsymbol{\omega})|Z_1(\boldsymbol{\omega})] = \alpha(\boldsymbol{\omega})Z_1(\boldsymbol{\omega})$, where $\alpha(\boldsymbol{\omega}) = \phi(\boldsymbol{\omega})\frac{\sigma_2\sqrt{f_2(\boldsymbol{\omega})}}{\sigma_1\sqrt{f_1(\boldsymbol{\omega})}}$. Because the normal distribution is a location-scale family, conditional on Z_1 the Z_2 process has the equivalent representation $Z_2(\boldsymbol{\omega}) = \alpha(\boldsymbol{\omega})Z_1(\boldsymbol{\omega}) + Z^*(\boldsymbol{\omega})$, where $Z^*(\boldsymbol{\omega})$ is Gaussian with $\mathrm{E}[Z^*(\boldsymbol{\omega})] = 0$, $\mathrm{Var}[Z^*(\boldsymbol{\omega})] \leqslant \sigma_2^2 f_2(\boldsymbol{\omega})$, and Z^* is independent over $\boldsymbol{\omega}$. Then, conditionally,

$$Y(\mathbf{s}) = \int \exp(-i\boldsymbol{\omega}^T\mathbf{s})Z_2(\boldsymbol{\omega})d\boldsymbol{\omega} = \int \exp(-i\boldsymbol{\omega}^T\mathbf{s})\alpha(\boldsymbol{\omega})Z_1(\boldsymbol{\omega})d\boldsymbol{\omega} + \int \exp(-i\boldsymbol{\omega}^T\mathbf{s})Z^*(\boldsymbol{\omega})d\boldsymbol{\omega}. \tag{6}$$

The first term in (6) is fixed as we are conditioning on Z_1, and the final term in (6) has mean zero. Therefore the conditional mean of $Y(\mathbf{s})$ is

$$\mathrm{E}[Y(\mathbf{s})|X(\mathbf{t}) \text{ for all } \mathbf{t}] = \mu(\mathbf{s}) = \int \exp(-i\boldsymbol{\omega}'\mathbf{s})\alpha(\boldsymbol{\omega})Z_1(\boldsymbol{\omega})d\boldsymbol{\omega}. \tag{7}$$

In the simple case where X and Y have the same spatial correlation, $f_1(\boldsymbol{\omega}) = f_2(\boldsymbol{\omega})$, and dependence is constant across frequency, $\phi(\boldsymbol{\omega}) = \phi$, then $\alpha(\omega) = \alpha = \phi\sigma_2/\sigma_1$ for all frequencies and (7) reduces to the linear downscaler $\mu(\mathbf{s}) = \alpha X(\mathbf{s})$. Therefore, (7) can be viewed as a multi-resolution extension of this linear model.

Since our objective is to predict the true air quality across the entire spatial domain and not to study spatial dependence in the proxy data, rather than placing priors on the $\phi(\boldsymbol{\omega})$

and $f_1(\boldsymbol{\omega})$, we can model $\alpha(\boldsymbol{\omega})$ directly. We express $\alpha(\boldsymbol{\omega})$ as a linear combination of K known basis functions $A_k(\boldsymbol{\omega})$,

$$\alpha(\boldsymbol{\omega}) = \sum_{k=1}^{K} A_k(\boldsymbol{\omega})\alpha_k. \tag{8}$$

This gives conditional mean

$$\begin{aligned} \mu(\mathbf{s}) &= \sum_{k=1}^{K} \tilde{X}_k(\mathbf{s})\alpha_k \quad \text{where} \\ \tilde{X}_k(\mathbf{s}) &= \int A_k(\boldsymbol{\omega}) \exp(-i\boldsymbol{\omega}'\mathbf{s}_j) Z_1(\boldsymbol{\omega}) d\boldsymbol{\omega}. \end{aligned} \tag{9}$$

For interpretation purposes, we select basis functions so that $\sum_{k=1}^{K} A_k(\boldsymbol{\omega}) = 1$ for all $\boldsymbol{\omega}$. In this case, if $\alpha_k = \alpha$ for all k, then $\alpha(\boldsymbol{\omega}) = \alpha$ for all $\boldsymbol{\omega}$, and the spectral downscaler reduces to the linear downscaler. This provides a way to center our Bayesian model on this simple special case.

Computing the spectral covariates $\tilde{X}_k$ clearly requires approximation since they are stochastic integrals. Fortunately they can be approximated efficiently using the discrete Fourier transformation when the proxy data are observed at $M = m_1 m_2$ locations situated on the $m_1 \times m_2$ rectangular grid of points $\mathcal{S}_{m_1 m_2} = \{0, 1, ..., m_1 - 1\} \otimes \{0, 1, ..., m_2 - 1\}$. We can then represent the proxy data using the two-dimensional discrete Fourier transform

$$X(\mathbf{s}_j) = \sum_{l=1}^{M} \exp(-i\boldsymbol{\omega}_l^T \mathbf{s}_j) Z_l, \tag{10}$$

where $\{\boldsymbol{\omega}_1, ..., \boldsymbol{\omega}_M\}$ is the set of M frequencies of the form $(2\pi u/m_1, 2\pi v/m_2)$ for $(u, v) \in \mathcal{S}_{m_1 m_2}$, and Z_l are the complex values that result from the inverse discrete Fourier transform of X,

$$Z_l = \frac{1}{M} \sum_{j=1}^{M} \exp(i\boldsymbol{\omega}_l^T \mathbf{s}_j) X(\mathbf{s}_j). \tag{11}$$

Transformations between X and Z are very efficient using the fast Fourier transform.

The constructed covariates are then approximated as

$$\tilde{X}_k(\mathbf{s}_j) \approx \sum_{l=1}^{M} A_k(\boldsymbol{\omega}_l) \exp(-i\boldsymbol{\omega}_l^T \mathbf{s}_j) Z_l. \tag{12}$$

The Web Appendix provides computer code to efficiently compute these constructed covariates. Even for the large datasets considered here these constructed covariates can be computed in a few seconds on an ordinary PC. By the properties of the discrete Fourier transformation, this approximation retains the essential feature that if $\alpha_1 = ... = \alpha_K$ and thus $\alpha(\boldsymbol{\omega}) = \alpha$ for all $\boldsymbol{\omega}$, then $\mu(\mathbf{s}) = \alpha X(\mathbf{s})$, as with the linear downscaler.

2.2 *Aliasing*

A final technical detail common in spectral analysis is aliasing.When data are observed only on $\mathcal{S}_{m_1 m_2}$, Z_j and Z_k are complex conjugates if $\boldsymbol{\omega}_j = (\omega_{j1}, \omega_{j2})$ and $\boldsymbol{\omega}_k = (\omega_{k1}, \omega_{k2})$ satisfy $\omega_{j1} + \omega_{k1} \in \{0, 2\pi\}$ and $\omega_{j2} + \omega_{k2} \in \{0, 2\pi\}$. To see this, recall that $\exp(i\boldsymbol{\omega}^T\mathbf{s}) = \cos(\boldsymbol{\omega}^T\mathbf{s}) + i\sin(\boldsymbol{\omega}^T\mathbf{s})$ and if $\omega_{jl} = 2\pi - \omega_{kl}$, then for any integer s, $\cos(\omega_{jl}s) = \cos(\omega_{kl}s)$ and $\sin(\omega_{jl}s) = -\sin(\omega_{kl}s)$. Therefore, when all $\mathbf{s}_i$ are integers, Z_j and Z_k form a complex conjugate pair since

$$\begin{aligned} Z_j &= \frac{1}{M}\sum_{i=1}^{M}\exp(i\boldsymbol{\omega}_j^T\mathbf{s}_i)X(\mathbf{s}_i) = C + iS \\ Z_k &= \frac{1}{M}\sum_{i=1}^{M}\exp(i\boldsymbol{\omega}_k^T\mathbf{s}_i)X(\mathbf{s}_i) = C - iS \end{aligned} \tag{13}$$

where $C = \frac{1}{M}\sum_{i=1}^{M}\cos(i\boldsymbol{\omega}_j^T\mathbf{s}_i)X(\mathbf{s}_i)$ and $S = \frac{1}{M}\sum_{i=1}^{M}\sin(i\boldsymbol{\omega}_j^T\mathbf{s}_i)X(\mathbf{s}_i)$. As a result, we cannot distinguish between signals at frequencies $\boldsymbol{\omega}_j$ and $\boldsymbol{\omega}_k$, and these frequencies are aliased.

To avoid identification problems, we assume that $\alpha(\boldsymbol{\omega})$ is the same for pairs of aliased frequencies. This is done by specifying the prior for α in terms of

$$\boldsymbol{\delta} = \begin{cases} \boldsymbol{\omega} & \text{if } ||\boldsymbol{\omega}|| \leqslant ||\bar{\boldsymbol{\omega}}|| \\ \bar{\boldsymbol{\omega}} & \text{if } ||\boldsymbol{\omega}|| > ||\bar{\boldsymbol{\omega}}|| \end{cases} \in [0, 2\pi), \tag{14}$$

where $\bar{\boldsymbol{\omega}} = [I(\omega_1 > 0)(2\pi - \omega_1), I(\omega_2 > 0)(2\pi - \omega_2)]$. Therefore, the signals at both aliased frequencies $\boldsymbol{\omega}_j$ and $\boldsymbol{\omega}_k$ are attributed to the smaller frequency, denoted $\boldsymbol{\delta}$. This assumptions resolves strict identifiability problems of aliased frequencies. However, frequencies with $||\boldsymbol{\delta}|| \approx 0$ and $||\boldsymbol{\delta}|| \approx 2\pi$ remain difficult to separate, as discussed further in Section 4.

2.3 *Summary of the final model*

After computing the spectral covariates, we proceed by fitting the usual spatial model for the observational data. Y is a Gaussian process with mean $\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T\boldsymbol{\beta} + \sum_{k=1}^{K} \tilde{X}_k(\mathbf{s})\alpha_k$ and spatial covariance $\sigma^2[(1-r)I(\mathbf{h}=0) + r\rho(\mathbf{h})]$, where $\mathbf{x}(\mathbf{s})$ are spatial covariates with corresponding regression coefficients $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$, $r \in (0,1)$ is the proportion of the variance attributed to spatial variance as opposed to nugget variance, and ρ is the Matérn correlation function with range λ and smoothness ν. Therefore, despite being motivated by rather complex spectral arguments, in practice the method can be implemented by simply adding K predictors $\tilde{X}_1(\mathbf{s}), ..., \tilde{X}_K(\mathbf{s})$ to the mean of the usual spatial linear model for the observational data. As a result, this model can be fit using standard software.

In our analysis in Section 4, the proxy data $X(\mathbf{s})$ represent areal averages of ozone concentrations over a region containing location $\mathbf{s}$, whereas the monitor data $Y(\mathbf{s})$ represent the value at a specific spatial location $\mathbf{s}$. This gives a change of support problem (Gelfand et al., 2001). As in, e.g., Berrocal et al. (2010b), we assume that the resolution of the proxy data is fine enough so we may simply match $Y(\mathbf{s})$ with the nearest proxy value.

To complete the Bayesian model, we specify priors for the model's parameters and the form of the basis functions, A_k. The mean parameters have priors $\alpha_k \overset{iid}{\sim} \mathrm{N}(\bar{\alpha}, \sigma_\alpha^2)$ and $\beta_j, \bar{\alpha} \overset{iid}{\sim} \mathrm{N}(0, 100^2)$. The variances have priors $\sigma^{-2}, \sigma_\alpha^{-2} \sim$ Gamma(0.1, 0.1), the variance ratio has prior $r \sim$ Uniform(0,1), and the Matérn parameters have priors $\log(\lambda) \sim \mathrm{N}(0, 10^2)$ and $\log(\nu - 0.5) \sim \mathrm{N}(0,1)$ so that $\nu \geqslant 0.5$.

We assume isotropy in our model for $\alpha(\boldsymbol{\omega})$, so that $\alpha(\boldsymbol{\omega}) = \alpha(\omega)$, where $\omega = ||\boldsymbol{\omega}||$. Since $\delta = ||\boldsymbol{\delta}||$ is restricted to the finite interval $[0, 2\pi)$, a natural choice for the basis functions are the Bernstein basis polynomials,

$$A_k(\boldsymbol{\omega}) = A_k(\delta) = \binom{K-1}{k-1} t^{k-1}(1-t)^{K-k}, \tag{15}$$

where $t = \frac{\delta}{2\pi} \in [0,1]$. These basis functions satisfy $\sum_{k=1}^{K} A_k(\boldsymbol{\omega}) = 1$ for all $\boldsymbol{\omega}$, so that if

$\sigma_\alpha = 0$ then $\alpha(\boldsymbol{\omega}) = \bar{\alpha}$. Therefore, the key assumptions of this choice of basis expansion and prior for the basis coefficients α_k are that $\alpha(\boldsymbol{\omega})$ varies smoothly across frequency and that the prior is centered on the usual linear downscaler with $\alpha(\boldsymbol{\omega}) = \bar{\alpha}$. Recalling that the correlation $\phi(\boldsymbol{\omega})$ is a product of $\alpha(\boldsymbol{\omega})$ and the smooth spectral densities f_1 and f_2, smoothness in α also implies smoothness in ϕ.

2.4 *Spatially-varying coefficient model*

Although not our primary focus, we also consider a non-stationary model which allows the relationship between proxy data and monitor data to be different in different regions. This model is more difficult to interpret, but may lead to improved predictions in some settings. To allow for spatially-varying bias terms, we consider extending the local spectral density approach of Fuentes (2001, 2002) to the bivariate setting. We specify J spatial knots $\mathbf{t}_1, ..., \mathbf{t}_J \in \mathcal{R}^2$, and assume that near knot $\mathbf{t}_j$ the spectral density is $f_{kj}(\boldsymbol{\omega})$ for process Z_k and the correlation at frequency $\boldsymbol{\omega}$ is $\phi_j(\boldsymbol{\omega})$. These local spectral densities are weighted by smooth kernel functions $w_j(\mathbf{s})$. For example, we select the standardized Gaussian kernel $w_j(\mathbf{s}) = \tilde{w}_j(\mathbf{s}) / \sum_{i=1}^J \tilde{w}_i(\mathbf{s})$, where $\tilde{w}_j(\mathbf{s}) = \exp[-0.5(||\mathbf{s} - \mathbf{t}_j||/\psi)^2]$ and ψ is the kernel bandwidth. This gives the mean

$$\mu(\mathbf{s}) = \sum_{j=1}^{J} w_j(\mathbf{s}) \left[\beta_{0j} + \sum_{k=1}^{K} \tilde{X}_k(\mathbf{s}) \alpha_{kj} \right] = \sum_{j=1}^{J} w_j(\mathbf{s}) \beta_{0j} + \sum_{j=1}^{J} \sum_{k=1}^{K} \tilde{X}_{jk}(\mathbf{s}) \alpha_{kj} \tag{16}$$

where β_{0j} is the intercept for knot $\mathbf{t}_j$. As before, this can be fit with standard software with the constructed covariates $w_j(\mathbf{s})$ and $\tilde{X}_{jk}(\mathbf{s}) = w_j(\mathbf{s}) \tilde{X}_k(\mathbf{s})$.

3. Simulation study

We conduct a simulation study to compare the spectral downscaler with ordinary Kriging and the linear downscaler. Proxy data, $X(\mathbf{s})$, are generated as a Gaussian process with mean zero, variance one, and exponential correlation $\text{Cor}[X(\mathbf{s}), X(\mathbf{t})] = \exp(-||\mathbf{s} - \mathbf{t}||/\lambda_X)$. The

responses, $Y(\mathbf{s})$, are generated with mean

$$\mathrm{E}\left[Y(\mathbf{s})|\mathbf{X}\right] = \sum_{l=1}^{M} \alpha(\delta_l) \exp(-i\boldsymbol{\omega}_l^T \mathbf{s}_j) Z_l,$$

and covariance $\mathrm{Cov}[Y(\mathbf{s}), Y(\mathbf{t})|\mathbf{X}] = 0.5^2 \exp(-||\mathbf{s}-\mathbf{t}||/\lambda_Y)$, where $Z_1, ..., Z_M$ is the inverse discrete Fourier transform of X and λ_X and λ_Y are the spatial range parameters. We consider three α functions:

(1) $\alpha(\delta) = 1$

(2) $\alpha(\delta) = I(\delta < \pi/4) + 0.5I(\pi/4 < \delta < \pi) - 0.2I(\pi < \delta < 3\pi/2) + 0.5I(\delta > 3\pi/2)$

(3) $\alpha(\delta) = 1 - 1.25I(\pi/2 < \delta < 3\pi/2)$

The first design is the linear downscaler with mean equal to the proxy data $\mu(\mathbf{s}) = X(\mathbf{s})$. The second design emulates the ozone data analyzed in Section 4, with strong correlation at low frequencies and low correlation for high frequencies. The final design may be unrealistic, but is intended to illustrate the potential effects of model misspecification on the performance of the linear downscaler. In this design, $\alpha(\delta) = 1$ for low-resolution features with $\delta < \pi/2$ and $\delta > 3\pi/2$, and is $\alpha(\delta) = -0.25$ for remaining δ, giving a slightly negative association for high-frequency features. Despite having negative α for many frequencies, the sample correlation between $X(\mathbf{s})$ and $Y(\mathbf{s})$ (averaged over space, time, and dataset) remains positive, ranging from 0.44 and 0.77 depending on the spatial correlation parameters. We also generated data from the kernel smoother model of Berrocal et al. (2012):

(4) $\mathrm{E}[Y(\mathbf{s}_i)|X(\mathbf{s})] = \sum_{j=1}^{225} w_{ij}(\phi) X(\mathbf{s}_j)$

where the weight is $w_{ij}(\phi) = W_{ij}(\phi)/[\sum_{l=1} W_{il}(\phi)]$, $W_{ij}(\phi)$ is the Gaussian kernel $\log[W_{ij}(\phi)] = -\frac{||\mathbf{s}_i - \mathbf{s}_j||^2}{2\phi^2}$, and ϕ is the kernel bandwidth which is set to $\phi = 1$ grid cell.

Data are generated on a 15×15 regular grid of points with grid spacing one, with 10 independent replications (representing 10 days) of the spatial process for each dataset. The 225 observations are split into 50 training observations and 175 testing observations. For

each simulation design and for $\lambda_X, \lambda_Y \in \{1, 5\}$, we generate 100 data sets. For each dataset, we fit several models to the training data:

(1) **Ordinary Kriging (OK)**: $\mu(\mathbf{s}) = \beta_0$

(2) **Linear downscaler (LD)**: $\mu(\mathbf{s}) = \beta_0 + \alpha X(\mathbf{s})$

(3) **Spectral downscaler (SD)**: $\mu(\mathbf{s}) = \beta_0 + \sum_{k=1}^{K} \alpha_k \tilde{X}_k(\mathbf{s})$

(4) **Kernel smoother (S(ϕ))**: $\mu(\mathbf{s}_i) = \sum_j w_{ij}(\phi) X(\mathbf{s}_j)$

(5) **Oracle (OR)**: $\mu(\mathbf{s})$ is set at the true value used to generate the data

For the spectral downscaler, we use $K = 10$ Bernstein basis functions. For the kernel smoother we compare bandwidths $\phi = 0.5$, 1.0, and 2.0. For each model, the priors and residual spatial model are described in Section 2.3. Figure 1 gives the mean square test set prediction error for each simulated dataset. We also computed the coverage of 90% prediction intervals, which were between 0.89 and 0.91 for all methods in all cases. The models are fit using MCMC sampling in `R`. We generate 10,000 samples, discard the first 1,000 as burn-in, and thin the remaining samples by 2, leaving 4,500 samples for prediction.

[Figure 1 about here.]

The spectral downscaler performs almost as well as the linear downscaler for the first design with data generated from the linear downscaler. Therefore, little is lost by including the extra constructed covariates in the spectral downscaler in this simple case. In the second simulation design, the spectral downscaler provides a significant improvement over the linear downscaler. The largest improvements occur in the presence of strong residual spatial correlation ($\lambda_Y = 5$) for the measurement data. In these cases, it seems spatial interpolation is fairly successful at capturing large scale trends even in the absence of proxy data, and thus the ability to appropriately handle fine-scale relationships separates the models. In the pathological third example, there is substantial correlation between measurement and proxy data, but this relationship is primarily driven by shared low-resolution features. Recent literature (Reich

et al., 2006; Hodges and Reich, 2010; Hughes and Haran, 2013) shows that it is difficult to estimate regression relationships in this case. These results suggest that in this case including proxy data as a linear predictor may not improve prediction over ordinary Kriging that does not make use of the proxy data. As expected, in the fourth design with data generated from the kernel smoothed model, the true model has smaller MSE than the spectral downscaler. However, in this case the spectral downscaler remains competitive, and actually has smaller MSE than the kernel smoother with misspecified bandwidth.

4. Analysis of ozone in North America

4.1 *Data description and exploratory analysis*

To illustrate the spectral downscaler, we analyze daily model output for ozone from CMAQ version 5.0.1 (`http://www.cmaq-model.org/`) (Appel et al., 2013) and monitored ozone data for July, 2005. Model output are paired in time and space with observations obtained from EPA's Air Quality System (AQS; `http://www.epa.gov/ttn/airs/airsaqs/`) and the Clean Air Status and Trends Network (CASTNet; `http://www.epa.gov/castnet/`). AQS sites are primarily located in urban and suburban locations, while CASTNet sites are located in rural areas away from major emission sources. The ozone metric of interest in this application is the daily maximum eight-hour average ozone concentration (MDA8 O3) since it is used for determining compliance with the EPAs ozone standards. CMAQ data is available on 12km x 12km grid covering the US and AQS data are available at 1,106 monitoring stations.

Figure 2 plots the CMAQ output for one day, as well as filtered CMAQ output to illustrate the signal for different frequencies. For frequency interval $[a, b)$, the filtered values are

$$X^*_{[a,b)}(\mathbf{s}) \approx \sum_{l=1}^{M} I(a \leqslant \delta_l < b) \exp(-i\boldsymbol{\omega}_l^T \mathbf{s}_j) Z_l. \tag{17}$$

Filtered images are created using the entire dataset, and plotted only for the southwest US in Figure 2. The low frequencies with $\delta < \pi/4$ capture large-scale spatial variation, whereas

frequencies in the intervals $[\pi/4, \pi/2)$, $[\pi/2, \pi)$, and $[\pi, 3\pi/2)$ capture progressively finer resolution features. Finally, the trends corresponding to frequencies greater than $3\pi/2$ begin to resemble low-frequency trends due to aliasing.

[Figure 2 about here.]

We begin investigating the relationship between CMAQ and monitor data using a simple least squares analysis. Let $0 = a_0, ..., a_{25} = 2\pi$ be equally-spaced points that partition $[0, 2\pi]$. We create filtered images $X^*_{[a_0,a_1)}(\mathbf{s}), ..., X^*_{[a_{24},a_{25})}(\mathbf{s})$ and match these constructed covariates with the monitor value $Y(\mathbf{s})$. This is repeated each day, and the data are pooled across days. We then fit a linear regression using these constructed covariates,

$$\mathrm{E}[Y(\mathbf{s})] = \alpha_0 + \sum_{j=1}^{25} \alpha(\delta_j) X^*_{[a_{j-1},a_j)}(\mathbf{s}),$$

where $\delta_j = (a_j + a_{j-1})/2$. This regression ignores all residual spatiotemporal dependence. Figure 3 plots the estimates of the $\alpha(\delta_j)$ by frequency δ_j and period $12(2\pi)/\delta_j$ (period is multiplied by 12 because the grid cells have width 12km). The estimated slopes are near 1 for frequencies near 0 and 2π. For medium frequency trends the relationship is weaker with $\alpha(\delta)$ around 0.25 for δ between $\pi/4$ and π. The estimated slopes are near zero, in some cases less than zero, for high-frequency trends with δ between π and $3\pi/2$. These estimates suggest good agreement between CMAQ and the monitor data for large-scale features (especially those exceeding 120 km), but very little agreement for small-scale features.

[Figure 3 about here.]

Figure 3 also plots (horizontal lines) the slope α_1 for the simple regression $\mathrm{E}[Y(\mathbf{s})] = \alpha_0 + \alpha_1 X(\mathbf{s})$. The estimated slope is around 0.8, which is higher than most of the slopes from the multi-resolution regression. It appears the slope is driven largely by the low-frequency trends shared by both CMAQ and the monitor data.

We also use this plot to suggest an appropriate number of basis functions, K, to include

in our model. Figure 3 plots the fits with $K = 5, 10$, and 15 Bernstein basis functions, fit to the $\hat{\alpha}(\delta_j)$ using weighted least squares with weights inversely proportional to the squared standard error of $\hat{\alpha}(\delta_j)$. It appears $K = 15$ basis functions is sufficient to capture the variation in α across frequencies.

4.2 *Model comparisons*

To compare models, we conduct test set validation. We compare four models,

(1) **No CMAQ**: $\mu(\mathbf{s}) = \beta_0 + \beta_1 x(\mathbf{s})$

(2) **Linear downscaler**: $\mu(\mathbf{s}) = \beta_0 + \beta_1 x(\mathbf{s}) + \alpha X(\mathbf{s})$

(3) **Spectral downscaler**: $\mu(\mathbf{s}) = \beta_0 + \beta_1 x(\mathbf{s}) + \sum_{k=1}^{K} \alpha_k \tilde{X}_k(\mathbf{s})$

(4) **Kernel smoothed downscaler**: $\mu(\mathbf{s}) = \beta_0 + \beta_1 x(\mathbf{s}) + \sum_j w_{ij}(\phi) X(\mathbf{s}_j)$

where $x(\mathbf{s}) = 1$ if site $\mathbf{s}$ is a CASTNet station, and $x(\mathbf{s}) = 0$ if site $\mathbf{s}$ is an AQS station. For each model, the residuals are a mean-zero Gaussian process with Matérn covariance as described in Section 2.3. The priors are the same for all models, and also given in Section 2.3. The data are modeled as independent across the 31 days. For each model, we generate 20,000 samples, discard the first 5,000 as burn-in, and thin the remaining samples by 2.

We randomly split the data into training and testing sets. We first split the data by randomly allocating sites to the training (581) and testing (252) sets. For each model, we compute predictive mean squared error, bias, variance and coverage of 90% intervals for the observations in the test set. We refer to this as "spatial prediction". In a second analysis, we randomly allocate days to the training (19) and testing (12) sets. In this second analysis, spatial dependence is not useful for prediction because there is no training data on the same day as test set observations. Therefore, we refer to this as "non-spatial prediction".

The results are in Table 1. The spectral downscaler provides a reduction in MSE for spatial prediction compared to the linear downscaler, and modest reduction in MSE compared to the kernel smoothed downscaler. However, the spatial-only model without CMAQ is competitive

with the downscaler models due to the strong spatial dependence in ozone data. In contrast, there are very substantial differences in MSE for non-spatial prediction; the relative MSE of the spectral downscaler compared to the model without CMAQ is 145.7/339.7=0.43, and the MSE relative to the linear downscaler is 145.7/202.1=0.72. In this case, spatial dependence does not help with prediction, and thus only treatment of the CMAQ output affects prediction. By including CMAQ output only at the appropriate spatial scales, the spectral downscaler makes better use of the output. As discussed in Section 1, this is an important result because CMAQ is often used for out-of-sample prediction to determine the effects of emission control strategies.

[Table 1 about here.]

4.3 *Results*

Table 2 summarizes the three model fits to the full dataset. In all three cases we see strong spatial dependence. The posterior mean spatial range (λ) varies from 145 km for the no-CMAQ model to 243 km for the spectral downscaler, and the proportion of residual variance attributed to spatial variation (r) ranges from 0.93 for the no-CMAQ model to 0.80 for the spectral downscaler. The residual variance (σ^2) and proportion of variance attributed to spatial variation are smaller for the downscaler models because including CMAQ output explains much of the spatial pattern.

[Table 2 about here.]

The posterior of $\alpha(\delta)$ in Figure 4 resembles the least squares estimate in Figure 3, but reflects our prior belief that it varies smoothly across frequency. The function $\alpha(\delta)$ is clearly not constant for all δ, as would be the case in the linear downscaler, demonstrating the need for the multi-resolution approach. The correlation between CMAQ and monitoring data is near zero for features with 24 km periods. This is twice the size of the 12km grid cells, and

it is known that deterministic models that are based on a numerical discretization scheme, such as CMAQ or the weather model that provides the necessary meteorological inputs to CMAQ, are unable to resolve features that are less than twice the grid resolution (Pielke, 1984). CMAQ is able to explain low-resolution features well, as $\alpha(\delta)$ steadily increases from zero to one as δ increases from 24 km. Features with periods less than 24 km appear to have large $\alpha(\delta)$, however this is likely due to the aliasing effect. As discussed in Section 2.2, these terms are difficult to distinguish from low-frequency terms.

[Figure 4 about here.]

Figure 5 plots the CMAQ output, monitoring data, posterior mean of $\mu(\mathbf{s})$, and spatial predictions for the full model fit (using the entire data set) for one day in the southwest. The CMAQ output shows two prominent local features: predictions near 120 ppb for a few grid cells near Reno, Nevada and in southern California. The posterior mean for $\mu(\mathbf{s})$ (defined in Section 4.2 with $x(s) = 0$, i.e., calibrated for AQS data) is quite different between the two downscalers. The linear downscaler preserves the shape of the CMAQ output, including the local features near Reno and in southern California. The spectral downscaler filters many local features, and is thus a smoothed version of the CMAQ output.

[Figure 5 about here.]

Despite having different means, the linear and spectral downscalers give similar spatial predictions throughout most of the region due to the large number of observations in the area. For example, there are many observations near Reno, and the predicted value in the grids cells west of Reno with CMAQ predictions near 120 ppb are shrunk to around 60 ppb by both models. There are however some important differences between the two predicted surfaces. The grid cells in southern California with CMAQ near 120 ppb are not near an ozone monitor. Therefore, the predicted values for these cells are largely driven by the CMAQ value and are thus much higher for the linear downscaler than the smoother spectral downscaler.

4.4 *Sensitivity analysis*

To test for sensitivity to our model assumptions, we changed the number of Bernstein basis functions from $K = 15$ to $K = 10$ and $K = 20$. Spatial prediction mean squared error changed only slightly, from 68.8 for $K = 15$ to 55.3 for $K = 10$ and 55.3 for $K = 20$. We also fit the non-stationary model described in Section 2.4 with $J = 9$ equally-spaced knots spanning the range of CMAQ spatial locations, ψ set to the distance between adjacent knots, and the same priors as the previous fits. The test set mean squared spatial prediction error was 56.7 for the linear downscaler with $K = 1$ and 55.2 for the spectral downscaler with $k = 15$. Since these results are similar to the model without spatially-varying terms, it does not appear that this additional complication is needed for this dataset.

5. Discussion

In this paper, we propose using spectral methods for spatial downscaling. The proposed method is computationally convenient, and can be fit using standard software. For the July, 2005 ozone data we find vastly different relationships between CMAQ output and monitoring data at different frequencies, with stronger correlation at larger spatial scales. Including CMAQ output only at the appropriate scales improves spatial and non-spatial prediction.

The analysis of the ozone data in Section 4 has several limitations. First we do not account for the fact that monitor stations are preferentially located in areas with high ozone. Adding CMAQ is a step towards handling bias caused by preferential sampling. Given that there is a strong relationship between CMAQ and monitoring data and smoothing-varying CMAQ bias, CMAQ output can reveal that sampled areas generally have higher ozone than non-sampled areas, and the mean can be adjusted to account for this difference. It should be straightforward to include the proposed spectral downscaling methods in model-based approaches to preferential sampling adjustments (Diggle et al., 2010; Pati et al., 2011; Gelfand et al., 2012).

In addition, it should be possible to include the proposed spectral methods in many non-Gaussian spatial models for binary or count data (Diggle et al., 1998). Another limitation is our treatment of the change of support between point-referenced monitor data and areal-averaged CMAQ output. Given that there are very few monitor observations in the same grid cell for the ozone data we have little information on variation at this resolution. Also ozone is considered a regional pollutant mainly formed by secondary atmospheric processes, so we might expect the level to be quite homogenous within a 12km x 12km grid cell. Therefore we feel it is unlikely that this would provide an advantage in our setting. However, to account for this issue more rigorously, it may be possible to embed the proposed multi-resolution model in the melding approach of Fuentes and Raftery (2005).

An important area of future work is to extend this approach to the spatiotemporal setting. Here we focus on the spatial case because for daily ozone data the spatial dependence is far stronger than the temporal dependence. A spatio-temporal model would allow for differing relationships between monitor and proxy data at different spatial and temporal scales, which could provide further information about CMAQ performance and may improve prediction. Another area of future work is to tailor the downscaler method to capture extreme ozone events (Reich et al., 2013). In our purely Gaussian model we find that smoothing CMAQ output improves prediction, but this may not be optimal for capturing extreme events.

Supplementary Material

The Web Appendix referenced in Section 2.1 is available with this article at the Biometrics website on Wiley Online Library.

Acknowledgments

The authors thank James Hodges (University of Minnesota) and Joseph Guinness (North Carolina State University) for fruitful discussions about this work. This study was partially

supported by USEPA grants R835228 and R834799, NSF grant 1107046, and NIH grants R01ES014843, R01ES019897, and R21ES022795-01A1. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication. The United States Environmental Protection Agency through its Office of Research and Development partially funded and collaborated in the research described here. It has been subjected to Agency review and approved for publication.

References

Alexandru, A., De Elia, R., Laprise, R., Separovic, L., and Biner, S. (2009). Sensitivity study of regional climate model simulations to large-scale nudging parameters. *Monthly Weather Review* **137,** 1666–1686.

Appel, K. W., Pouliot, G. A., Simon, H., Sarwar, G., Pye, H. O. T., Napelenok, S. L., Akhtar, F., and Roselle, S. J. (2013). Evaluation of dust and trace metal estimates from the Community Multiscale Air Quailty (CMAQ) model version 5.0. *Geosci. Model Dev. Discuss.* **6,** 1859–1899.

Bell, M., Samet, J., McDermott, A., Zeger, S., and Dominici, F. (2004). Ozone and mortality in 95 U.S. urban communities from 1987 to 2000. *Journal of the American Medical Association* **292,** 2372–2378.

Berrocal, V., Craigmile, P., and Guttorp, P. (2013). Regional climate model assessment using statistical upscaling and downscaling techniques. *Environmetrics* DOI:10.1002/env.2145.

Berrocal, V., Gelfand, A., and Holland, D. (2010a). A bivariate space-time downscaler under space and time misalignment. *Annals of Applied Statistics* **4,** 1942–1975.

Berrocal, V., Gelfand, A., and Holland, D. (2010b). A spatio-temporal downscaler for output from numerical models. *Journal of Agricultural, Biological, and Environmental Statistics* **15,** 176–197.

Berrocal, V., Gelfand, A., and Holland, D. (2012). Space-time data fusion under error in computer model output: An application to modeling air quality. *Biometrics* **68,** 837–848.

Berrocal, V., Gelfand, A., Holland, D., Burke, J., and Miranda, M. (2011). On the use of a PM2.5 exposure simulator to explain birthweight. *Environmetrics* **22,** 553–571.

Byun, D. and Schere, K. L. (2006). Review of the governing equations, computational algorithms, and other components of the models-3 community multiscale air quality (cmaq) modeling system. *Applied Mechanics Reviews* **59,** 51–77.

Chang, H., Reich, B., and Miranda, M. (2012). Time-to-event analysis of fine particle air pollution and preterm birth: Results from North Carolina, 2001-2005 (with discussion). *American Journal of Epidemiology* **175,** 91–98.

Crooks, J. and Isakov, V. (2013). A wavelet-based approach to blending observations with deterministic computer models to resolve the intra-urban air pollution field. *Journal of the Air & Waste Management Association* Accepted.

Diggle, P. J., Menezes, R., and Su, T. (2010). Geostatistical inference under preferential sampling (with discussion). *Journal of the Royal Statistical Society: Series C* **59,** 191–232.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics (with discussion). *Journal of the Royal Statistical Society: Series C* **47,** 299–350.

Foley, K., Reich, B., and Napelenok, S. (2012). Bayesian analysis of a reduced-form air quality model. *Environmental Science & Technology* **46,** 7604–7611.

Fuentes, M. (2001). A high frequency kriging for nonstationary environmental processes. *Environmetrics* **12,** 1–15.

Fuentes, M. (2002). Periodogram and other spectral methods for nonstationary spatial processes. *Biometrika* **89,** 197–210.

Fuentes, M. and Raftery, A. (2005). Model evaluation and spatial interpolation by Bayesian

combination of observations with outputs from numerical models. *Biometrics* **61,** 36–45.

Fuentes, M. and Reich, B. J. (2010). *Handbook of Spatial Statistics*, chapter Spectral Analysis for Spatial Data, pages 57–78. CRC Press.

Gelfand, A., Zhu, L., and Carlin, B. (2001). On the change of support problem for spatio-temporal data. *Biostatistics* **2,** 31–45.

Gelfand, A. E., Sahu, S. K., and Holland, D. M. (2012). On the effect of preferential sampling in spatial prediction. *Environmetrics* **23,** 565–578.

Hodges, J. and Reich, B. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician* **64,** 325–334.

Hughes, J. and Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B* **75,** 139–159.

Kim, S., Sheppard, L., Hannigan, M., Dutton, S., Peel, J., Clark, M., and Vedal, S. (2013). The sensitivity of health effect estimates from time-series studies to fine particulate matter component sampling schedule. *J Expo Sci Environ Epidemiol* **23,** 581–483.

Kloog, I., Koutrakis, P., Coull, B., Lee, H., and Schwartz, J. (2011). Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment* **45,** 6267–6275.

Lim, C., Stein, M., Ching, J., and Tang, R. (2010). Statistical properties of differences between low and high resolution cmaq runs with matched initial and boundary conditions. *Environmental Modelling and Software* **25,** 158–169.

Liu, P., Tsimpidi, P., Hu, Y., Stone, B., Russell, A., and Nenes, A. (2012). Differences between downscaling with spectral and grid nudging using wrf. *Atmospheric Chemistry and Physics* **12,** 3601–3610.

Liu, Y., Paciorek, C., and Koutrakis, P. (2009). Estimating regional spatial and temporal

variability of PM2.5 concentrations using satellite data, meteorology, and land use information. *Environmental Health Perspectives* **117,** 886–892.

McMillan, N., Holland, D., Morara, M., and Feng, J. (2009). Combining numerical model output and particulate data using bayesian space-time modeling. *Environmentrics* **21,** 48–65.

Mebust, M., Eder, B., Binkowski, F., and et al (2003). Models-3 community multiscale air quality (cmaq) model aerosol component 2. model evaluation. *J Geophys Res* **180,** 4184–4202.

Morris, J., Vannucci, M., Brown, P., and Carroll, R. (2003). Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. *Journal of the American Statistical Association* **98,** 574–583.

Nguyen, H., Cressie, N., and Braverman, A. (2012). Spatial statistical data fusion for remote sensing. *Journal of the American Statistical Association* **107,** 1004–1018.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., and Sain, S. (2012). A multi-resolution Gaussian process model for the analysis of large spatial data set-shttp://www.image.ucar.edu/ nychka/manuscripts/LatticeKrigJCGS2.pdf.

Paciorek, C. (2012). Combining spatial information sources while accounting for systematic errors in proxies. *Journal of the Royal Statistical Society Series C* **61,** 429–451.

Pati, D., Reich, B. J., and Dunson, D. B. (2011). Bayesian geostatistical modeling with informative sampling locations. *Biometrika* **98,** 35–48.

Pielke, R. A. (1984). *Mesoscale meteorological modeling.* Academic Press, New York.

Radu, R., Deque, M., and Somot, S. (2008). Spectral nudging in a spectral regional climate model. *Tellus* **604,** 898–910.

Reich, B., Cooley, D., Foley, K., Napelenok, S., and Shaby, B. (2013). Extreme value analysis for evaluating ozone control strategies. *Annals of Applied Statistics* **7,** 739–762.

Reich, B., Hodges, J., and Zadnik, V. (2006). Effects of residual smoothing on estimation of the fixed effects in disease-mapping models. *Biometrics* **62,** 1197–1206.

Strickland, M., Darrow, L., Klein, M., Flanders, W., Sarnat, J., Waller, L., Sarnat, S., Mulholland, J., and Tolbert, P. (2010). Short-term associations between ambient air pollutants and pediatric asthma emergency department visits. *American Journal of Respiratory and Critical Care Medicine* **182,** 307–316.

Stuart, A. L., S., M., and Sriwatanapongse, W. (2009). The social distribution of neighborhood-scale air pollution and monitoring protection. *Journal of the Air & Waste Management Association* **59,** 591–602.

Thompson, M., Reynolds, J., Cox, L., Guttorp, P., and Sampson, P. (2001). A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmospheric Environment* **35,** 617–630.

Von Storch, H., Langenberg, H., and Feser, F. (2000). A spectral nudging technique for dynamic downscaling purposes. *Monthly Weather Review* **128,** 3664–3673.

Zhou, J., Chang, H., and Fuentes, M. (2012). Estimating the health impacts of climate change with calibrated model output. *Journal of Agricultural, Biological, and Environmental Statistics* **17,** 377–394.

Zhou, J., Fuentes, M., and Davis, J. (2011). Calibration of numerical model output using nonparametric spatial density functions. *Journal of Agricultural, Biological, and Environmental Statistics* **16,** 531–553.

Received: August 14, 2013

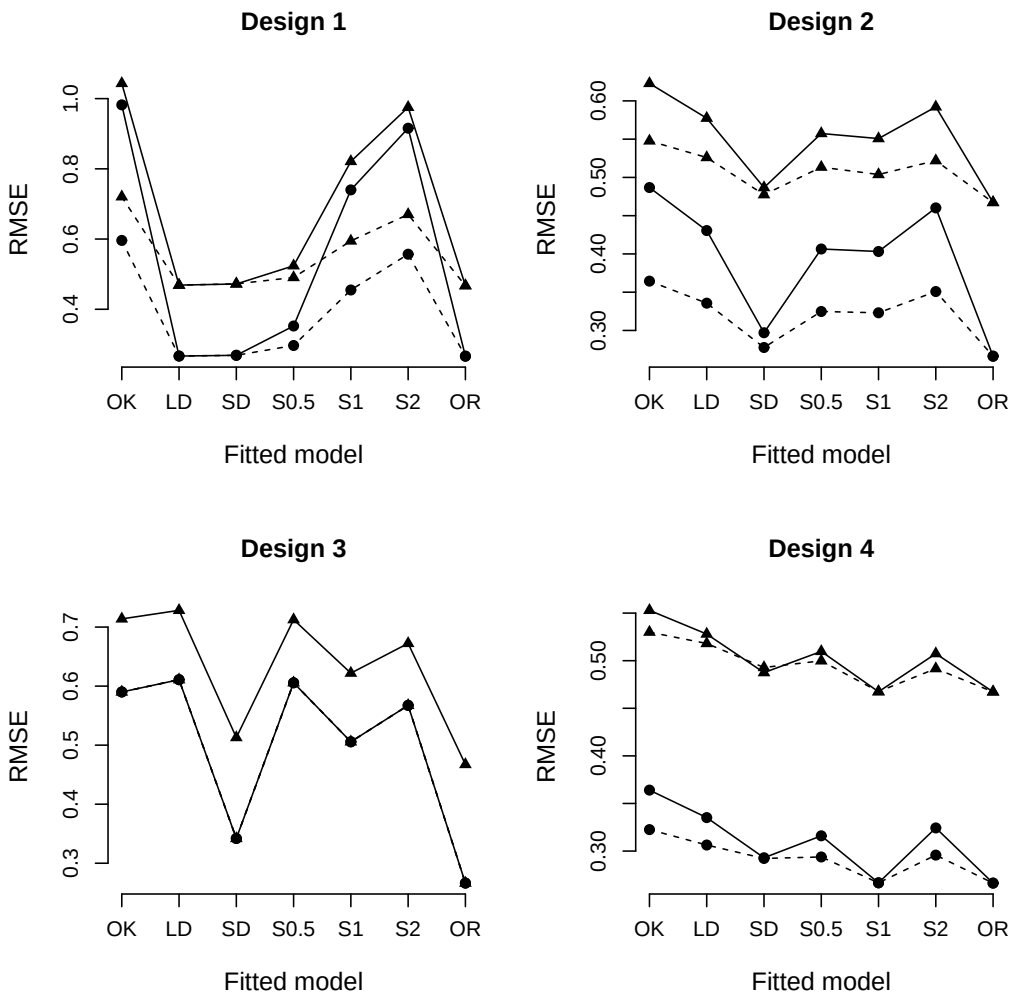

Figure 1. Square root prediction mean squared error for ordinary Kriging ("OK"), the linear downscalar ("LD"), the spectral downscaler ("SD"), kernel smoothed downscaler ("S"), and the oracle model ("OR") for simulated data. For each design, results are presented with $\lambda_X = 1$ (solid lines) and $\lambda_X = 5$ (dashed lines) and for $\lambda_Y = 1$ (triangles) and $\lambda_Y = 5$ (circles). The standard error for each value in the figure is less than 0.0246.

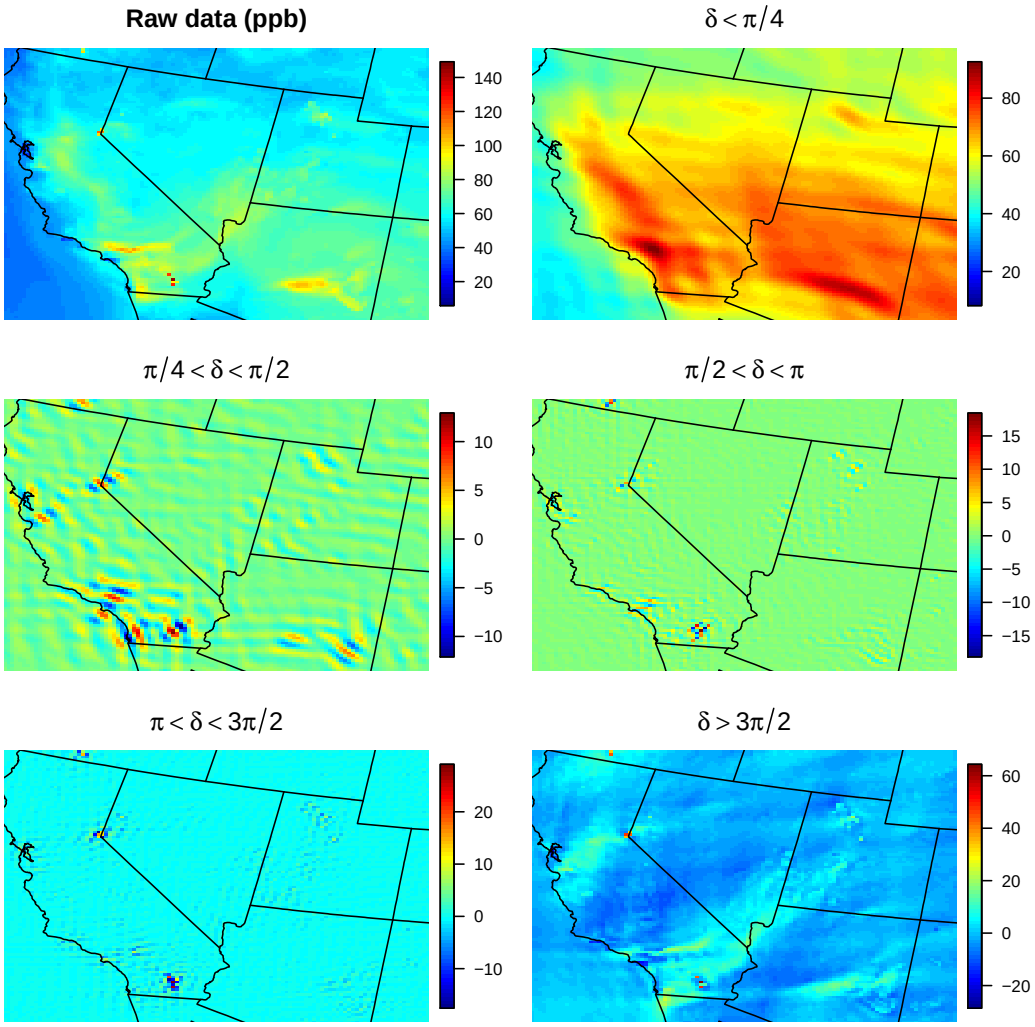

Figure 2. CMAQ output filtered at different frequencies for July 4, 2005.

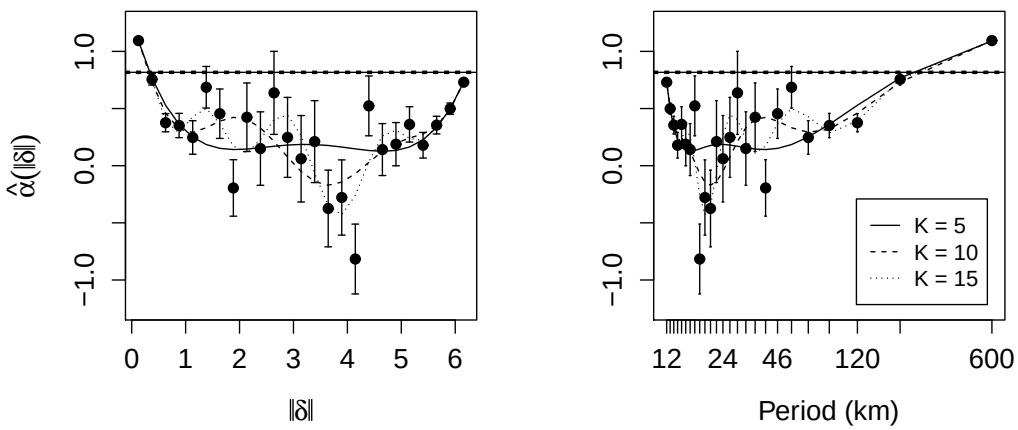

Figure 3. Least squares estimate and 95% intervals of $\alpha(\delta)$ and basis function fits with various number of Bernstein basis functions (K) to the least squares estimates, plotted by frequency δ (left) and period $12(2\pi)/\delta$ (right, log base 10 scaling). The horizontal lines near 0.8 are the estimated slope (solid) and 95% confidence interval (dashed) for the simple linear regression assuming the same effect for all frequencies, i.e., $\alpha(\delta) = \alpha$.

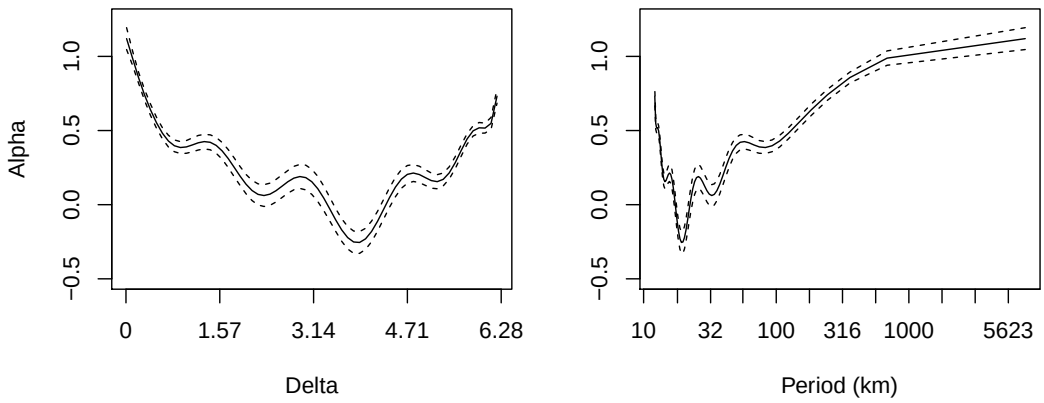

Figure 4. Posterior mean (solid) and 90% interval (dashed) of $\alpha(\delta)$, plotted by frequency δ (left) and period $12(2\pi)/\delta$ (right, log base 10 scaling).

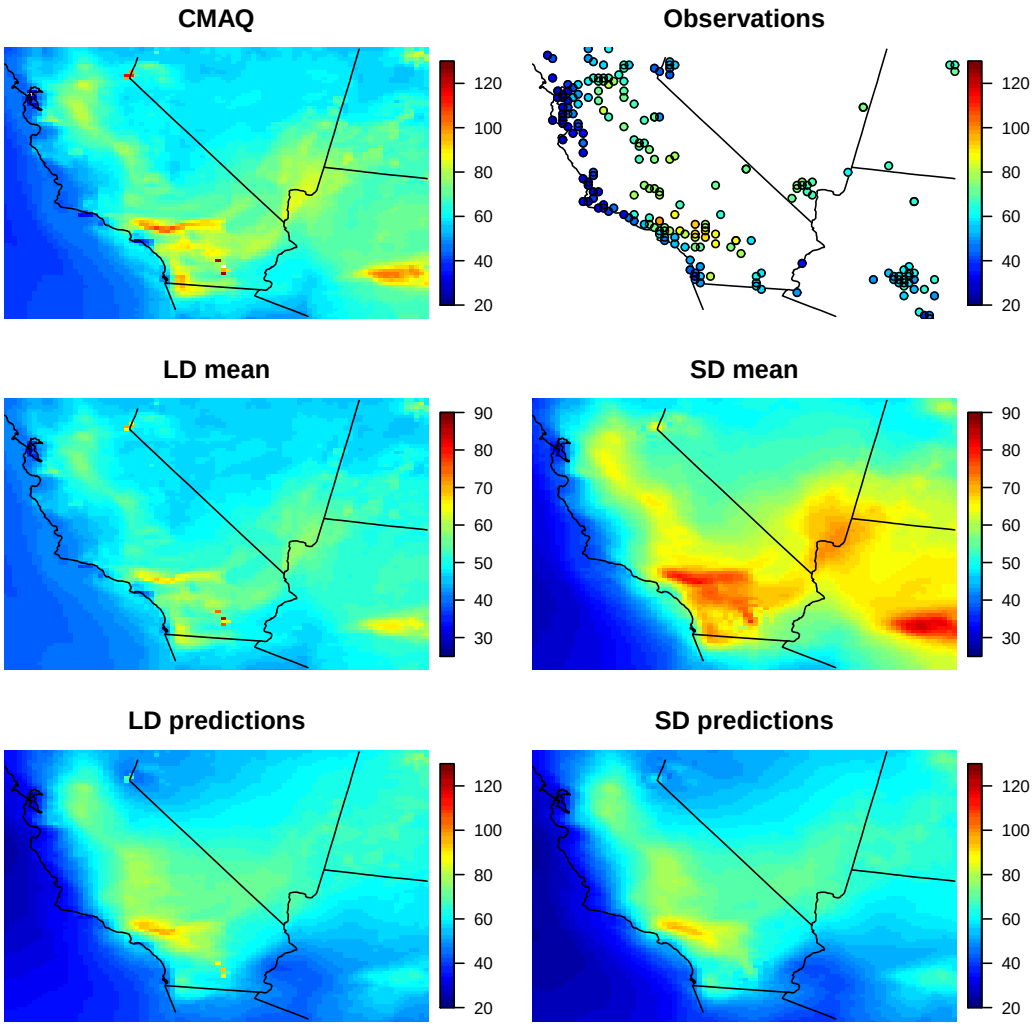

Figure 5. CMAQ output $X(\mathbf{s})$, monitor observations $Y(\mathbf{s})$, estimated mean $\mu(\mathbf{s})$, and spatial predictions under the linear (LD) and spectral (SD) downscaler models for July 4, 2005. All units are ppb.

Table 1
Test-set prediction performance, including coverage of 90% prediction intervals.

	Spatial prediction			
	MSE	Bias	Variance	Coverage
No CMAQ	62.8	-0.14	66.3	0.91
Linear downscaler	57.5	-0.26	56.2	0.91
Spectral downscaler	53.7	-0.23	53.3	0.91
Kernel smoothed downscaler with bandwidth 12 km	54.9	-0.23	54.8	0.91
Kernel smoothed downscaler with bandwidth 60 km	58.7	-0.17	59.2	0.91
Kernel smoothed downscaler with bandwidth 120 km	60.9	-0.14	62.9	0.91

	Non-spatial prediction			
	MSE	Bias	Variance	Coverage
No CMAQ	339.7	-6.17	302.0	0.89
Linear downscaler	202.1	-2.80	177.7	0.89
Spectral downscaler	145.7	0.57	129.1	0.89
kernel smoothed downscaler with bandwidth 12 km	151.2	-0.06	134.8	0.89
kernel smoothed downscaler with bandwidth 60 km	157.6	1.06	142.9	0.89
kernel smoothed downscaler with bandwidth 120 km	169.1	0.93	151.7	0.89

Table 2

Posterior mean (standard deviation) of various parameters under different models.

	No CMAQ	Linear downscaler	Spectral downscaler
Network effect (ppb), β_1	2.75 (0.16)	2.72 (0.15)	2.88 (0.15)
Standard deviation (ppb), σ	17.9 (0.35)	13.6 (0.21)	11.6 (0.13)
Variance ratio, r	0.93 (0.01)	0.85 (0.01)	0.80 (0.01)
Spatial range (km), λ	415.5 (19.9)	336.1 (14.6)	243.5 (8.85)
Smoothness parameter, κ	0.51 (0.01)	0.51 (0.01)	0.51 (0.01)