The Importance of Scale for Spatial-Confounding Bias and Precision of Spatial Regression Estimators

Christopher J. Paciorek

Abstract. Residuals in regression models are often spatially correlated. Prominent examples include studies in environmental epidemiology to understand the chronic health effects of pollutants. I consider the effects of residual spatial structure on the bias and precision of regression coefficients, developing a simple framework in which to understand the key issues and derive informative analytic results. When unmeasured confounding introduces spatial structure into the residuals, regression models with spatial random effects and closely-related models such as kriging and penalized splines are biased, even when the residual variance components are known. Analytic and simulation results show how the bias depends on the spatial scales of the covariate and the residual: one can reduce bias by fitting a spatial model only when there is variation in the covariate at a scale smaller than the scale of the unmeasured confounding. I also discuss how the scales of the residual and the covariate affect efficiency and uncertainty estimation when the residuals are independent of the covariate. In an application on the association between black carbon particulate matter air pollution and birth weight, controlling for large-scale spatial variation appears to reduce bias from unmeasured confounders, while increasing uncertainty in the estimated pollution effect.

Key words and phrases: Epidemiology, identifiability, mixed model, penalized likelihood, random effects, spatial correlation, splines.

1. INTRODUCTION

Spatial confounding is likely present in many of the applied contexts in which residuals are spatially correlated, particularly in public health and social science. Consider the motivating example of the health effects of exposure to (spatially varying) air pollution, an important public health issue. Many variables that explain variability in the response, including potential confounding variables that may be correlated with exposure, also vary spatially. For example, large-scale re-

Christopher J. Paciorek is Adjunct Assistant Professor, Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, Massachusetts 02115, USA and Visiting Assistant Professor, Department of Statistics, 367 Evans Hall, University of California, Berkeley, California 94720, USA (e-mail: paciorek@alumni.cmu.edu).

gional patterns in air pollution may be correlated with regional patterns in diet, income and other risk factors for a health outcome of interest. Small-scale patterns in air pollution from local sources may be correlated with risk factors as well, for example, if lower-income people live nearer to busy roads or industrial sources. If confounding variables are not measured, it will be difficult to distinguish the effect of air pollution from residual spatial variation in the health outcome. I use the term spatial confounding to characterize this situation. Researchers have modeled the spatial structure in the outcome with the apparent goal of reducing confounding bias (e.g., Clayton, Bernardinelli and Montomoli, 1993; Pope et al., 2002; Cakmak et al., 2003; Biggeri et al., 2005). However, the statistical mechanism for reducing bias does not appear to be well understood nor investigated rigorously in the statistical or applied literature.

To consider the problem formally, start with simple linear regression with spatial structure:

$$Y_i = \beta_0 + \beta_x X_i + e_i, \quad i = 1, \ldots, n,$$

(1)

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

where each outcome, Y_i, is associated with a spatial location, $\mathbf{s}_i \in \mathfrak{R}^2$. X_i is the corresponding value of a univariate regressor of interest, which may also vary spatially, in which case we would represent X_i as $X(\mathbf{s}_i)$. $\mathbf{e} = (e_1, \ldots, e_n)^T$ is the vector of errors, whose covariance matrix, $\boldsymbol{\Sigma}$, captures any residual spatial correlation, as well as independent variation. The regression coefficients, $\beta = \{\beta_0, \beta_x\}$, are unknown, and estimation of β_x is of primary interest. Spatial statistics and regression texts note that the ordinary least squares (OLS) estimator for β_x in this setting is unbiased but inefficient, and the usual OLS variance estimator is incorrect. Assuming known $\boldsymbol{\Sigma}$, the generalized least squares (GLS) estimator is the most efficient estimator. However, little appears to be known about how the spatial scales of the residual variability and of X affect inference. Spatial structure in X is very common in applications and complicates the problem because X and the residual spatial structure compete to explain the variability in the response (Waller and Gotway, 2004). Furthermore, it would not be surprising if the spatial correlation in the residuals were caused by an unmeasured spatially varying confounder; I next introduce another representation of (1) to enable exploration of confounding. Motivated by the air pollution example, I will refer to X as the "exposure."

One can obtain the basic spatial regression model (1) using a simple mixed model,

$$(2) \quad Y_i = \beta_0 + \beta_x X(\mathbf{s}_i) + g(\mathbf{s}_i) + \varepsilon_i,$$

with random effects, $\mathbf{g} = (g(\mathbf{s}_1), \ldots, g(\mathbf{s}_n))^T$, and white noise errors, $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau^2)$. Suppose the random effects are spatially correlated, with $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \sigma_g^2 \mathbf{R}(\theta_g))$, where $\mathbf{R}(\theta_g)$ is a spatial correlation matrix parameterized by θ_g, a spatial range parameter, and σ_g^2 is the variance of the random effects. Marginalizing over $\mathbf{g}$ gives the marginal likelihood,

$$\mathbf{Y} = (Y_1, \ldots, Y_n)^T$$

(3)

$$\sim \mathcal{N}(\beta_0 \mathbf{1} + \beta_x \mathbf{X}, \sigma_g^2 \mathbf{R}(\theta_g) + \tau^2 \mathbf{I}),$$

where $\mathbf{1}$ is an n-vector of ones, $\mathbf{I}$ is the identity matrix, and $\mathbf{X} = (X_1, \ldots, X_n)^T$. Here $\boldsymbol{\Sigma}$ in (1) is explicitly decomposed into spatial and nonspatial components. An alternative formulation would specify the unknown spatial function, $g(\mathbf{s})$, as a penalized spline, where a penalty parameter plays the role of $\{\theta_g, \sigma_g^2\}$ in the marginal likelihood in penalizing complexity of the spatial structure. The exposure may itself be spatially correlated. For example, if $X(\mathbf{s})$ is a Gaussian process, then $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma_x^2 \mathbf{R}(\theta_x))$, with parameters analogous to those for $\mathbf{g}$. To demonstrate processes operating at different spatial scales, Figure 1 shows simulated spatial surfaces as one varies the spatial range parameter, θ, in a Gaussian process model.

The spatial statistics literature assumes that the error, e_i in (1), is independent of the covariate(s) (Cressie, 1993; Waller and Gotway, 2004), with little or no discussion of the possibility that the error involves variation from unmeasured confounders. Henceforth, I will refer to the errors as residuals because of the common use of the term "spatial residual" to refer to unexplained spatial variability. To explore the possibility of confounding, let's consider $\mathbf{g} \equiv \beta_z \mathbf{Z}$ to be induced as the effect, β_z, of an unmeasured variable, Z, on the outcome. $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^T$ may also be spatially
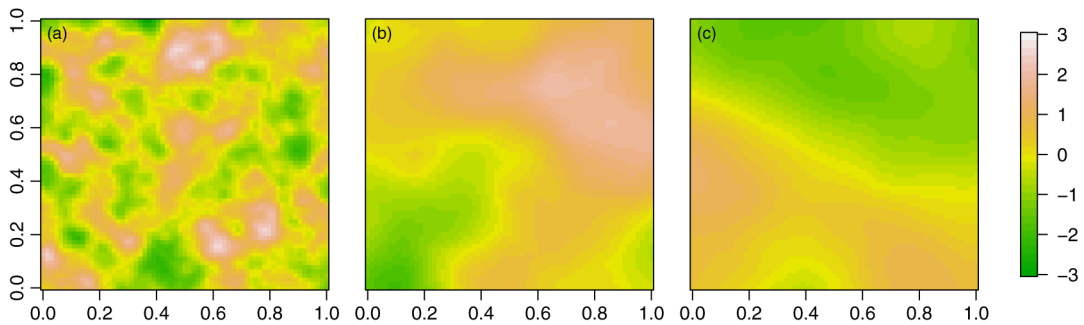

FIG. 1. *Gaussian process realizations using the Matérn covariance (see Section 2.2) for three values of θ, with* (a) *high-frequency, small (fine)-scale variability when $\theta = 0.1$,* (b) *moderate scale variability when $\theta = 0.5$, and* (c) *low-frequency, large-scale variability when $\theta = 0.9$.*

correlated, for example, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma_z^2 \mathbf{R}(\theta_z))$, such that $\sigma_z^2 = \sigma_g^2/\beta_z^2$, where θ_z is again a spatial range parameter. If Z and X are dependent, then Z is an unmeasured spatial confounder. Derivation of the marginal likelihood should be done by integrating over the (unknown) conditional distribution of $\mathbf{Z}$ given $\mathbf{X}$, whereas the integration leading to (1) ignores the dependence. Note that if $\mathbf{X}$ and $\mathbf{Z}$ are considered fixed, then association between $\mathbf{X}$ and $\mathbf{g} \equiv \beta_z \mathbf{Z}$ is known as concurvity (Buja, Hastie and Tibshirani, 1989; Ramsay, Burnett and Krewski, 2003).

In the applied literature, practitioners often recognize the need to consider residual spatial structure in the outcome, with language of "control" or "accounting" for autocorrelation, and they fit models (such as kriging or spatial random effects) that implicitly assume independence of the residual and the exposure (Burnett et al., 2001; Cakmak et al., 2003; Cho, 2003; Burden et al., 2005; Augustin et al., 2007; Molitor et al., 2007; Cerdá et al., 2009; Lee, Ferguson and Mitchell, 2009). With the recent exception of Hodges and Reich (2010), formal statements of the goals and properties of fitting such spatial models are generally absent. However, much of the interest appears to lie in using the spatial residual structure to try to account for spatial confounding, with the implicit assumption that such models reduce or eliminate confounding bias (e.g., Clayton, Bernardinelli and Montomoli, 1993; Pope et al., 2002; Cakmak et al., 2003; Richardson, 2003; Biggeri et al., 2005). One approach is to explicitly consider the spatial scales involved, hoping that accounting for variation at a relatively large spatial scale allows for identification of the parameter of interest based on exposure heterogeneity at a smaller spatial scale (e.g., Burnett et al., 2001; Cakmak et al., 2003; Zeger et al., 2007). This smaller scale variation may be less prone to confounding in a given application. However, this consideration of spatial scale is often not explicit, and effects of scale on bias reduction, while sometimes hinted at, have not been developed formally.

In the analogous context of time series modeling of air pollution, Dominici, McDermott and Hastie (2004) attempt to attribute all the temporally correlated variability in the outcome to the residual term in order to identify the effect of exposure based on the temporally uncorrelated (and presumably unconfounded) heterogeneity in the exposure. Dominici, McDermott and Hastie (2004) provide no guidance in the scenario that the exposure cannot be decomposed into autocorrelated and uncorrelated components. This issue also applies to the approach of Lombardía and Sperlich (2007), who filter out the dependence between fixed and random effects. In the spatial setting, in which measurements cannot be made at all locations, accurate estimation of the uncorrelated component, if such a component even exists, is rare: consider atmospheric phenomena such as temperature and air pollution. A common situation in which fine-scale heterogeneity is not resolved involves prediction of spatially varying exposure values using averages of nearby measurements or spatial smoothing techniques. Hence, I seek to address the problem when all of the measured components of variation in exposure are spatial.

In this paper I address estimation in simple regression models with spatial residual structure. I focus on the properties of penalized models, using a simple mixed model fit by GLS, equivalent to universal kriging, to analyze the effects of the spatially correlated residual structure on fixed effect estimators. Section 2 focuses on bias from spatial confounding. I report analytic results when the full covariance structure is known and supporting simulations when the covariance (or the amount of smoothing in penalized spline models) is estimated from the data. I assess the use of sensitivity analysis approaches based on spline models that explicitly consider the bias-variance tradeoff involved in choosing the spatial scale at which to model the residual variation. Section 3 focuses on precision of estimators when there is no association between exposure and residual (no spatial confounding). I close with a case study of the effects of air pollution on birthweight (Section 4).

2. SPATIAL CONFOUNDING AND BIAS

2.1 Identifiability

A key consideration in the basic model (2) is identifiability of β_x and $g(\mathbf{s})$. A closely-related question is how the estimation procedure attributes variability between the exposure and the spatial residual term (the random effects). In the simple linear model, attribution of variability to the covariates rather than the error term is favored because this allows the estimate of the error variance to decrease, with the normalizing constant of the likelihood favoring smaller error variance. In the spatial model, if the spatial term, $\mathbf{g}$, is unconstrained, then $\beta_x \mathbf{X}$ and $\mathbf{g}$ are not identifiable in the likelihood: one could remove the covariate from the model and redefine $g^*(\mathbf{s}) \equiv \beta_x X(\mathbf{s}) + g(\mathbf{s})$ with no change in the likelihood. Identifiability comes through constraints on $\mathbf{g}$, either by (1) penalizing lack of smoothness in $g(\mathbf{s})$, (2) considering $\mathbf{g}$ to be a random effects

term, or (3) having a prior on **g**. These approaches give higher penalized likelihood, marginal likelihood or posterior density, respectively, when variability is attributed to the unpenalized fixed effects term rather than to the spatial term. In the spatial confounding context this dynamic causes bias in estimation of β_x, for example, as seen in the simulations of Peng, Dominici and Louis (2006). An alternative constraint is to represent **g** in a reduced dimension basis, say, as a regression spline. In this case the model is identifiable if there is a component of variability in **X** that cannot be explained by the spline structure, that is, if **X** is not perfectly collinear with the columns of the chosen basis matrix.

2.2 Analytic Framework

To consider bias from unmeasured spatially varying confounders, take the following model as the data-generating mechanism,

$$(4) \qquad Y_i \sim \mathcal{N}\big(\beta_0 + \beta_x X(\mathbf{s}_i) + \beta_z Z(\mathbf{s}_i), \tau^2\big),$$

with the notation as in Section 1. For each location, **s**, suppose the correlation of $X(\mathbf{s})$ and $Z(\mathbf{s})$ over repeated sampling at the location is $\rho \neq 0$, so that Z is a confounder. Suppose further that Z is not observed and that one models the residual spatial structure in the outcome through spatially correlated random effects, $\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \sigma_g^2 \mathbf{R}(\theta_g))$ as in (2). Finally, suppose that one ignores the correlation between $\mathbf{g} \equiv \beta_z \mathbf{Z}$ and **X** and integrates over the marginal distribution for **g**, giving (3). Equivalently, $Y_i = \beta_0 + \beta_x X(\mathbf{s}_i) + \varepsilon_i^*$. The induced correlation between X and ε^* violates the usual regression assumption that the error is independent of the covariate, leading to bias. From the random effects perspective, we have (incorrectly) assumed that the random effects are independent of the covariate, a key (but often unstated) assumption of mixed effects models (Breslow and Clayton, 1993; Diggle et al., 2002, page 170).

The treatment of $X(\mathbf{s})$ and $Z(\mathbf{s})$ as random naturally induces spatial structure. However, in a given data set the most plausible repeated sampling framework may suggest that **X** and **Z** reflect spatial structure that does not arise from a stochastic data generating process. Rather, one might consider $X(\mathbf{s})$ and $Z(\mathbf{s})$ to be fixed unknown functions, particularly when **X** and **Z** vary at large scales, which mimics the partial spline/partial linear setting. This also is consistent with the treatment of large-scale variation in the mean term in traditional kriging. Consider the case when there is concurvity between the two fixed functions, reflected in a nonzero

empirical correlation, $\hat{\rho}$, between **X** and $\mathbf{g} \equiv \beta_z \mathbf{Z}$ as calculated over the collection of locations (e.g., the concurvity in the simulations of Ramsay, Burnett and Krewski, 2003; He, Mazumdar and Arena, 2006; Peng, Dominici and Louis, 2006). In the partial linear/partial spline setting it is well known that such association between the exposure and the nonparametric smooth term causes bias (Rice, 1986, Equation 28; Speckman, 1988). In any real data set, the orthogonality needed for $\hat{\rho} \approx 0$ seems particularly unlikely if both **X** and **Z** vary at large scale relative to the size of the domain (though $\hat{\rho} < 0$ may be as much a possibility as $\hat{\rho} > 0$).

The stochastic generative model is still useful under this framework of fixed functions because realizations of $X(\mathbf{s})$ and $Z(\mathbf{s})$ give plausible values for **X** and **Z** that could arise in real applications for which there is no reasonable stochastic mechanism. I choose to treat **X** and **Z** stochastically, and I use ρ to quantify explicitly the strength of association between the residual spatial variation and the exposure. This approach allows for some simple, useful analytic results and is further justified in that the variation that an unmeasured **Z** induces in **Y** is necessarily treated stochastically as part of the residual in actual applications. In some cases I report results conditional on **X**, and in others I also average over the stochastic variability in **X** and over variability in the spatial locations of the observations.

Since Z represents an unmeasured confounder, I assess the inferential properties of fitting a regression model by maximizing the marginal likelihood (3) using GLS, thereby ignoring correlation between the residual and the exposure. I assess bias as a function of the spatial scales of $X(\mathbf{s})$ and $Z(\mathbf{s})$, which I suppose to be generated as Gaussian processes with Matérn spatial correlation function

$$R(d; \theta, \nu) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{2\sqrt{\nu}d}{\theta}\right)^\nu \mathcal{K}_\nu \left(\frac{2\sqrt{\nu}d}{\theta}\right),$$

where d is the Euclidean distance between two locations, θ is the spatial range parameter, and $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind, whose order is the smoothness parameter, ν. I fix $\nu = 2$, which gives continuous and differentiable Gaussian process realizations. This reflects an assumption of some smoothness in the spatial processes under consideration, but I also consider results based on an exponential correlation function (i.e., $\nu = 0.5$). The model (3) is equivalent to both a mixed model and a universal kriging model if one knows the variance and spatial dependence parameters. Furthermore, given the extensive use of penalized splines in applications, and

the connection between penalized splines and mixed models (Ruppert, Wand and Carroll, 2003), I also consider the use of a penalized spline to represent $g(\mathbf{s})$.

In the nonspatial context, one would generally try to adjust for confounding by including relevant covariates as fixed effects; in the spatial context one could include spatial regression spline terms. The basic question that I explore in the remainder of Section 2 is the extent to which inclusion of a spatial random effect term or a penalized spline can adjust for unmeasured spatial confounding, given that these approaches do not involve a projection in the way that a regression spline does. The random effects and penalized spline approaches do estimate the residual spatial variation based on a bias-variance tradeoff (e.g., Claeskens, Krivobokova and Opsomer, 2009), and the penalized spline is a regression spline in the limit as the penalty goes to zero. So it seems plausible that these approaches may reduce bias by at least partially adjusting for the unmeasured spatial confounder. I will show that the spatial scales involved are critical.

2.3 Bias with Known Parameters

This section considers bias when I suppose that the variance parameters are known and only the regression coefficients, β_0 and β_x, are unknown. The initial results concern the situation when the exposure, X, and the unmeasured confounder, Z, vary at the same spatial scale. I then assess what happens when X varies at two scales and one is the same scale as the single-scale confounder. Finally, I consider the possibility that there is additional variability in the outcome at another scale, but uncorrelated with X.

To start, suppose that $X(\mathbf{s})$ and $Z(\mathbf{s})$ share the same spatial correlation range, θ, but may have different marginal variances, namely, $X \sim \mathcal{N}(\mu_x \mathbf{1}, \sigma_x^2 \mathbf{R}(\theta))$ and $Z \sim \mathcal{N}(\mu_z \mathbf{1}, \sigma_z^2 \mathbf{R}(\theta))$ and $\mathrm{Cov}(\mathbf{X}, \mathbf{Z}) = \rho \sigma_x \sigma_z \mathbf{R}(\theta)$. Straightforward conditional normal calculations give

$$\mathrm{E}(\hat{\beta}_x | \mathbf{X}) = \beta_x + [(\mathcal{X}^T \boldsymbol{\Sigma}^{-1} \mathcal{X})^{-1} \mathcal{X}^T \boldsymbol{\Sigma}^{-1} \mathrm{E}(\mathbf{Z}|\mathbf{X}) \beta_z]_2$$

$$\begin{aligned} = \beta_x + &\Bigg[(\mathcal{X}^T \boldsymbol{\Sigma}^{-1} \mathcal{X})^{-1} \mathcal{X}^T \boldsymbol{\Sigma}^{-1} \mathcal{X} \\ &\cdot \begin{pmatrix} \mu_z - \rho \frac{\sigma_z}{\sigma_x} \beta_z \mu_x \\ \rho \frac{\sigma_z}{\sigma_x} \beta_z \end{pmatrix} \Bigg]_2 \end{aligned}$$

(5)

$$= \beta_x + \rho \frac{\sigma_z}{\sigma_x} \beta_z,$$

where $\mathcal{X} = [\mathbf{1} \quad \mathbf{X}]$, $[\cdot]_2$ indicates the second element of the 2-vector, and $\boldsymbol{\Sigma} = \sigma_g^2 \mathbf{R}(\theta) + \tau^2 \mathbf{I}$. The resulting bias, $\rho \frac{\sigma_z}{\sigma_x} \beta_z$, is the same as if $\mathbf{X}$ and $\mathbf{Z}$ were not spatially structured and is also equal to the bias under OLS. This demonstrates that we have not adjusted for confounding at all by fitting the model that includes spatial structure. As with OLS, the model attributes as much of the variability as possible to the exposure, rather than to the spatially correlated residual term, including all of the variability in $\mathbf{Z}$ that is related to $\mathbf{X}$. If $\rho = 0$, the bias is zero in (5). This occurs because we average over stochastic variability in $\mathbf{Z}$, so any nonorthogonality between $\mathbf{X}$ and $\mathbf{Z}$ in individual realizations contributes to variance rather than bias. This contrasts with the bias terms in Rice (1986) and Dominici, McDermott and Hastie (2004), which are caused by nonorthogonality of the fine-scale variation in $\mathbf{X}$ and the nonparametric component of the model, since neither is treated stochastically.

Next, I keep the same data-generating and model-fitting framework, but explore the situation in which the exposure varies at two scales. I suppose that $X(\mathbf{s})$ is a multi-scale process and introduce correlation between $Z(\mathbf{s})$ and one of the components of $X(\mathbf{s})$. Let $\mathbf{X} = \mathbf{X}_c + \mathbf{X}_u$ be decomposed into a component, $\mathbf{X}_c$, that is at the same scale as the confounder, $\mathbf{Z}$, and a component at a different scale, $\mathbf{X}_u$, which is independent of $\mathbf{X}_c$ and $\mathbf{Z}$. Specifically, take $\mathrm{Cov}(\mathbf{X}) = \sigma_c^2 \mathbf{R}(\theta_c) + \sigma_u^2 \mathbf{R}(\theta_u)$, $\mathrm{Cov}(\mathbf{Z}) = \sigma_z^2 \mathbf{R}(\theta_c)$ and $\mathrm{Cov}(\mathbf{X}, \mathbf{Z}) = \mathrm{Cov}(\mathbf{X}_c, \mathbf{Z}) = \rho \sigma_c \sigma_z \mathbf{R}(\theta_c)$. After some straightforward algebra and matrix manipulations, we have

$$\mathrm{E}(\hat{\beta}_x | \mathbf{X}) = \beta_x + [(\mathcal{X}^T \boldsymbol{\Sigma}^{*-1} \mathcal{X})^{-1} \mathcal{X}^T$$

(6)
$$\cdot \boldsymbol{\Sigma}^{*-1} \mathrm{E}(\mathbf{Z}|\mathbf{X}) \beta_z]_2$$

$$= \beta_x + c(\mathbf{X}) \rho \frac{\sigma_z}{\sigma_c} \beta_z,$$

where

$$k(\mathbf{X}) \equiv [(\mathcal{X}^T \boldsymbol{\Sigma}^{*-1} \mathcal{X})^{-1} \mathcal{X}^T \boldsymbol{\Sigma}^{*-1} \mathbf{M}(\mathbf{X} - \mu_x \mathbf{1})]_2 p_c,$$

$$\boldsymbol{\Sigma}^* \equiv \frac{\beta_z^2 \sigma_z^2 \mathbf{R}(\theta_c) + \tau^2 \mathbf{I}}{\beta_z^2 \sigma_z^2 + \tau^2} = ((1 - p_z)\mathbf{I} + p_z \mathbf{R}(\theta_c)),$$

$$\mathbf{M} \equiv (p_c \mathbf{I} + (1 - p_c) \mathbf{R}(\theta_u) \mathbf{R}(\theta_c)^{-1})^{-1}$$

and $p_z \equiv \beta_z^2 \sigma_z^2 / (\beta_z^2 \sigma_z^2 + \tau^2)$. We see that the bias term is proportional to that in the single-scale setting, multiplied by an additional term, $k(\mathbf{X})$, that modulates the bias. $k(\mathbf{X})$ necessarily includes an extra multiplicative factor, $p_c \equiv \sigma_c^2 / (\sigma_c^2 + \sigma_u^2)$, that quantifies the magnitude of the confounded component of $\mathbf{X}$ relative to

the total variation in **X**. While the term $k(\mathbf{X})$ is complicated, we can explore its dependence on the spatial scales (θ_c and θ_u) and the magnitudes of the variance component ratios (p_z and p_c) to see how the bias compares to the same-scale setting. In the following results I average over the variability in **X**.

For a grid of $n = 100$ locations on the unit square, Figure 2 shows the average of $k(\mathbf{X})$ over 1000 simulations as a function of θ_c and θ_u, for combinations of p_c and p_z, where the empirical average approximates the expectation with respect to the distribution of **X**. There is a simple pattern to the bias modification relative to the same-scale setting. For $\theta_c = \theta_u$ (the diagonal elements on the 1:1 line), we do not need simulation: $E_X k(\mathbf{X}) = p_c$, which is equivalent to the same-scale result (5), after accounting for the proportion of variability in **X** that is confounded, p_c. Note that if one estimates the analogous bias to (6) for OLS applied to spatial data, it is nearly constant regardless of the spatial scales [$\hat{E}_X k(\mathbf{X}) \approx p_c$; not shown]. Only when

$\theta_u < \theta_c$, and particularly when $\theta_u \ll \theta_c$, do we see less bias than in the same-scale setting, with clear potential for bias reduction from modeling the residual spatial variation in the outcome (recall Figure 1 to interpret the values of θ). Above the diagonal, for $\theta_u > \theta_c$, $\hat{E}_X k(\mathbf{X}) > p_c$, indicating more bias when the scale of confounding is smaller than the scale of unconfounded variability. This situation may be of limited practical interest, because it's not clear that there are real applications in which the unconfounded variation in the exposure occurs at larger scales than the confounded variation. However, it does show that there are circumstances in which bias is larger than under OLS, a point also made in Hodges and Reich (2010). Note that the patterns in Figure 2 are qualitatively similar regardless of the values of p_c and p_z. Quantitatively, for larger values of p_c, corresponding to a larger proportion of the variation in the exposure being confounded variation, bias is larger. For larger values of p_z, corresponding to a larger proportion of the residual variation being
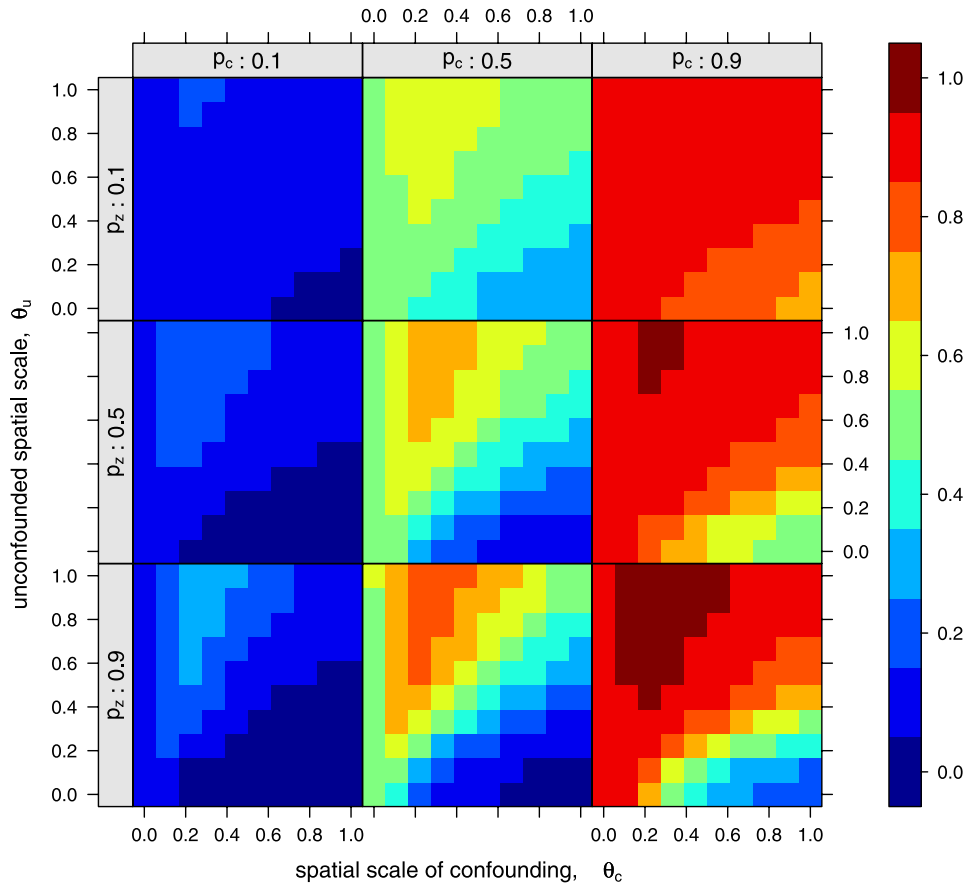

FIG. 2. *The expected value of the bias modification term, $\hat{E}_X k(\mathbf{X})$, as a function of the spatial scales of confounded (θ_c) and unconfounded (θ_u) variability for a selection of values of p_z and p_c. $k(\mathbf{X})$ quantifies the amount of bias relative to the bias in the same-scale setting or with nonspatial confounding ($\rho \sigma_z \beta_z / \sigma_x$). Along the diagonal ($\theta_c = \theta_u$) $E_X k(\mathbf{X}) = p_c$, which is equivalent to no bias reduction. Values near zero indicate substantial bias reduction.*

the contribution of the confounder, the effects of the spatial scales are more distinct. The results highlight that inclusion of the spatial residual does not give unbiased estimates and bias is substantial in many scenarios even when the covariance parameters are known. Results are very similar when I sample locations uniformly on the unit square or in a clustered fashion (using a Poisson cluster process).

The results in Figure 2 correct for the complication that the sample variance of spatial process values (calculated over the domain) decreases as θ increases. This occurs because the sample variance over the domain in a single spatial replicate underestimates population variability; see Figure 1(b)–(c) for examples. I want to have fixed ratios of average sample variances, $p_z \equiv \beta_z^2 E_Z s_z^2 / (\beta_z^2 E_Z s_z^2 + \tau^2) \in \{0.1, 0.5, 0.9\}$ and $p_c \equiv E_{X_c} s_c^2 / (E_{X_c} s_c^2 + E_{X_u} s_u^2) \in \{0.1, 0.5, 0.9\}$, for all values of θ_c and θ_u, thereby avoiding the introduction of artifacts caused solely by having ratios of sample variances change with the spatial ranges. Here s_z^2, s_c^2 and s_u^2 are the sample variances of $\mathbf{Z}$, $\mathbf{X}_c$ and $\mathbf{X}_u$, respectively. To achieve this, I generate $\mathbf{X}_c \sim \mathcal{N}(\mathbf{0}, d_c^2 \sigma_c^2 \mathbf{R}(\theta_c))$ and $\mathbf{X}_u \sim \mathcal{N}(\mathbf{0}, d_u^2 \sigma_u^2 \mathbf{R}(\theta_u))$ and modify the calculation of $k(\mathbf{X})$ in (6) accordingly. d_c and d_u are functions of θ_c and θ_u, respectively, that are chosen such that $E_{X_c} s_c^2(\theta_c) \approx \sigma_c^2$ and $E_{X_u} s_u^2(\theta_u) \approx \sigma_u^2$, where $s_c^2(\theta_c)$ is the sample variance of $\mathbf{X}_c$ for a given realization under θ_c and analogously for $s_u^2(\theta_u)$. The expectations are taken with respect to the distribution of the subscripted random vector. These manipulations allow me to present bias for scenarios that correspond to specific ratios of average sample variability of $\mathbf{X}_c, \mathbf{X}_u, \mathbf{Z}$ and $\boldsymbol{\varepsilon}$ over the spatial domain.

To have only a single scale of residual spatial variability is not very realistic. Therefore, I carried out an additional simulation study with residual spatial variability in the outcome that is independent of the exposure and at a smaller scale than the scale of $Z(\mathbf{s})$. I suppose that the data-generating model is

$$(7) \qquad \mathbf{Y} = \beta_0 \mathbf{1} + \beta_x \mathbf{X} + \beta_z \mathbf{Z} + \mathbf{h} + \boldsymbol{\varepsilon},$$

and that $\mathbf{h} \sim \mathcal{N}(\mathbf{0}, \sigma_h^2 \mathbf{R}(\theta_u))$, independent of $\mathbf{X}$, $\mathbf{Z}$ and $\boldsymbol{\varepsilon}$, with all of the other details as before. Under this data-generating model and again supposing that all variance parameters are known, simulation estimates of $E_X k(\mathbf{X})$ indicate that bias is somewhat smaller than that seen in Figure 2 for $\theta_c > \theta_u$ and somewhat larger for $\theta_c < \theta_u$ (not shown). Note that if the additional small-scale variability is correlated with the exposure, then one is back in the situation of having common scales for the exposure and the confounder, which is considered at the beginning of this section.

2.4 Bias and Precision with Estimated Parameters

To generalize the results of Section 2.3, which supposed known variance and spatial dependence parameters, I set up a simulation study to assess the impact of estimating those parameters. In addition to maximum likelihood estimation of a mixed effects/kriging model based on the marginal likelihood (3), I consider the use of penalized likelihood to fit the model (2) with a penalized thin plate spline spatial term for $g(\mathbf{s})$. I implemented the penalized spline using gam() in R, which uses generalized cross-validation (GCV) for data-driven smoothing parameter estimation (Wood, 2006). For the core simulations, I set the following parameter values, $\sigma_u^2 = \sigma_c^2 = \beta_z^2 \sigma_z^2 = 1$, $\tau^2 = 4$, $\beta_x = 0.5$, $\rho = 0.3$, and sample 100 spatial locations uniformly from the unit square. For a range of values of θ_c and θ_u, I simulate 2000 data sets for each pair $\{\theta_c, \theta_u\}$. For each simulated data set, I generate new spatial locations and new values of $\mathbf{X}$ and $\mathbf{Z}$; I then generate $\mathbf{Y}$ using (4). Again we have to account for the reduced empirical spatial variability as θ increases; these simulations have effective values of $p_c = 0.5$ and $p_z = 0.2$.

With regard to bias, the simulation results for the mixed/kriging model [Figure 3(b)] reasonably match the theoretical values with known variance parameters [Figure 3(a)]. However, when $\theta_u \ll \theta_c$, the bias is generally larger than with known variance parameters, because the fitted model sometimes estimates little or no spatial structure in the residuals, pushing bias results toward the larger bias seen under OLS [Figure 3(d)]. Results for the penalized spline model [Figure 3(c)] show smaller bias for $\theta_u < \theta_c$ than the mixed model, presumably caused by the difference between estimating the amount of smoothing by GCV compared to maximum likelihood. In either case, spatial scales are critical, and bias is smaller than with OLS only when the scale of confounding is larger than the scale of the unconfounded variability. Additional simulations indicate that as the correlation of confounder and exposure increases, or the magnitude of variation in the confounder increases, or the effect size decreases, relative bias increases (not shown). In such scenarios, substantial bias reduction occurs only for very small spatial scales in the exposure and large scales of confounding.

Figure 4 compares the mixed model with the penalized spline in the context of a bias-variance tradeoff. There is a substantial bias-variance tradeoff, with the smaller bias of the penalized spline model (for $\theta_c < \theta_u$) trading off for increased variance. The result
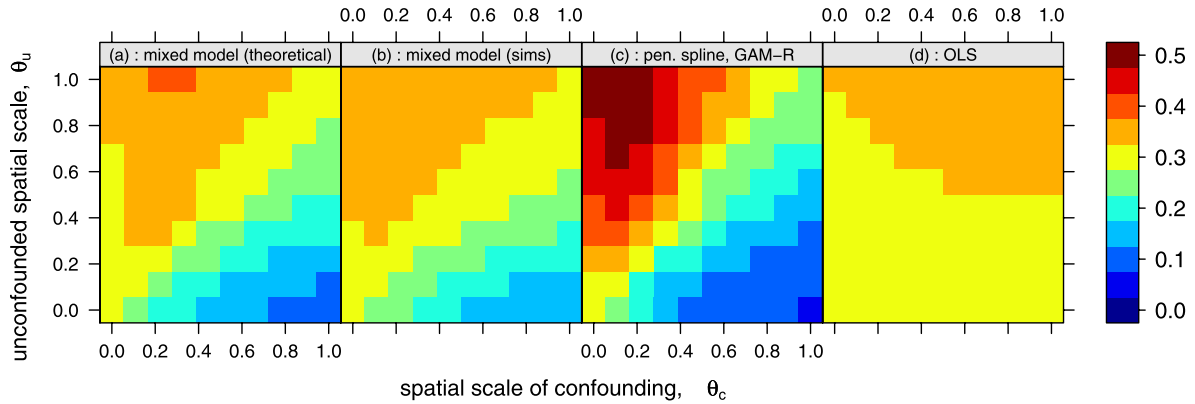
FIG. 3. *Relative bias,* $(\hat{\mathrm{E}}(\hat{\beta}_x) - \beta_x)/\beta_x$, *as a function of the spatial scales of confounded* (θ_c) *and unconfounded variability* (θ_u): (a) *the-oretical bias for the mixed/kriging model with known variance parameters, and* (b), (c), (d) *simulated bias with estimated variance/penalty parameters for* (b) *the mixed model,* (c) *a penalized spline model, and* (d) *OLS.*

is increased mean squared error (MSE) in $\hat{\beta}_x$, except when θ_u is very small. Both model variance estimates (third column) understate the variability in the coefficient estimates (second column), with particular underestimation of uncertainty and low coverage for the mixed/kriging model, and with lower coverage as one

moves away from the region of $\theta_c \gg \theta_u$. Of course the bias causes much of the poor coverage.

Fitting the mixed/kriging model by restricted maximum likelihood (REML) rather than maximum likelihood produces moderate improvement in coverage, with the average variance estimate more similar to the
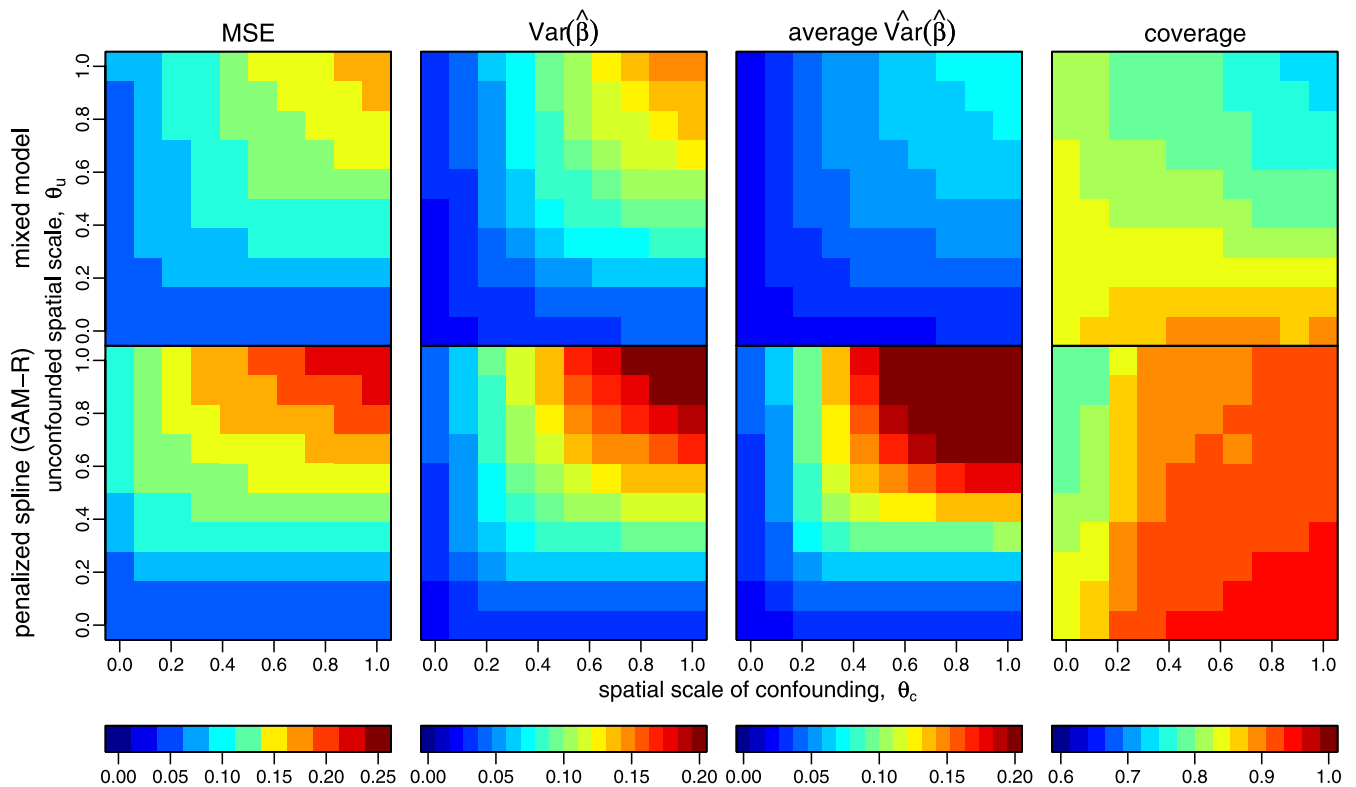

FIG. 4. *Simulation results for* (top row) *mixed model/kriging fit and* (bottom row) *penalized spline model. Each plot shows results as a function of the spatial scales of the confounded* (θ_c) *and unconfounded variability* (θ_u), *with MSE* (first column), *variance of the estimates over the simulations* (second column), *average squared standard error* (third column) *and coverage* (fourth column).

variance of the estimated coefficients. Using $\nu = 0.5$ (i.e., an exponential spatial correlation function) in the fitting rather than the true $\nu = 2$ has little effect on results. However, when I generate the unconfounded variability, $\mathbf{X}_u$, based on $\nu = 0.5$, bias is substantially smaller than the core results (particularly note that there is reduced bias relative to OLS when $\theta_c = \theta_u$), apparently because the nondifferentiable sample paths of processes with exponential covariance play the role of very fine-scale, unconfounded variability. There is little change in results when using spatial locations simulated using a Poisson cluster process with an average of seven children per cluster and cluster kernel standard deviation of 0.03. Finally, simulations with $\rho = 0$, that is, no confounding, indicate no bias for either model, as expected.

Our bias results when $\rho \neq 0$ are analogous to the bias seen with penalized spline models in He, Mazumdar and Arena (2006) and Peng, Dominici and Louis (2006). There, concurvity (i.e., $\hat{\rho} \neq 0$) between the smooth temporal term (analogous to our spatial residual) and the exposure emerged from the fixed basis coefficients chosen based on empirical data examples (R. Peng, personal communication; He, Mazumdar and Arena, 2006). Similar results are seen in the spatial settings of Ramsay, Burnett and Krewski (2003).

The presence of small-scale independent variation in the residual (7) reduces bias for $\theta_u < \theta_c$ (not shown), relative to the results presented above. This occurs through an increase in the number of degrees of freedom estimated from the data to capture residual variability, that is, undersmoothing with respect to the variation at the θ_c scale, analogous to undersmoothing in the partial spline setting (Rice, 1986; Speckman, 1988). This scenario seems quite likely in applications: if there is large-scale residual spatial structure, there is likely to be finer-scale structure as well. Thus, analyses that attempt to best fit the data may in the process reduce bias from confounding at the larger scales.

2.5 The Bias-Variance Tradeoff

We have seen that even when all covariance parameters are known and the scale of confounding is much larger than the scale of unconfounded variability in $\mathbf{X}$, bias remains, albeit at a much reduced level. In principle, if the structure at the confounded scale could be exactly fit using a set of basis functions, such as a regression spline (e.g., Dominici, McDermott and Hastie, 2004), then the exposure effect estimate would be unbiased, as in any multiple regression. The partial residual kernel smoothing approach of Speckman

(1988) reduces bias in a similar fashion, albeit without using a projection, through the technique of twicing. However, in a real application, one has to choose the basis functions, and if the basis functions do not fully explain the confounded, large-scale variability, even with a basis of seemingly sufficient dimension, this will induce a bias. One could instead consider a penalized spline approach with penalty parameter chosen in advance to give the desired effective degrees of freedom (e.d.f.). For fixed e.d.f., since the penalized spline smoother is not a true projection (Speckman, 1988; Peng, Dominici and Louis, 2006), one would expect the penalized spline approach to have more bias than the regression spline approach. Heuristically, bias in this approach occurs because the estimated spatial term does not fully explain the confounded component of the variability in the outcome, causing a bias analogous to that seen in the partial spline setting (Rice, 1986; Speckman, 1988). However, we would expect the penalized spline to be less sensitive to the exact form of the basis functions and number and placement of knots, as is seen in the example (Section 4). Furthermore, one can always undersmooth to reduce the bias, following the recommendation in the partial spline literature (Rice, 1986; Speckman, 1988). Thus, using a penalized spline seems reasonable, albeit without the clean interpretation of a projection. I show below that simulations comparing regression spline and penalized spline models support these theoretical results from the literature, in the spatial context considered here, with the regression spline having reduced bias and increased variance relative to penalized modeling.

The primary issue in an application is choosing the amount of smoothing to reduce bias, since inference about β_x is the goal rather than best fitting the data. Data-driven smoothing might reduce bias (if there is small-scale residual correlation) or might have little effect on bias (if the data suggest only large-scale residual correlation). Thus, the reduction in bias will depend on the scales involved and the actual amount of smoothing done, and the analysis will reveal little about the sensitivity of estimation to scale. Instead, one could explicitly assess the bias-variance tradeoff by varying the amount of smoothing and assessing the sensitivity of the exposure effect inference. One approach is a spatial analogue to the sensitivity analysis approaches of Peng, Dominici and Louis (2006): fit a model with spatial basis functions and vary the e.d.f. (e.g., Zeger et al., 2007). Plotting $\hat{\beta}_x$ and uncertainty intervals as a function of e.d.f. (or some other metric) provides an assessment of the robustness of results to

potential spatial confounding at various scales. If one is concerned about confounding at a particular scale, then one can report the results for an e.d.f. that would undersmooth with respect to that scale to reduce bias, accepting the tradeoff of increased uncertainty.

Motivated by this analysis strategy, I set up a simulation under the settings of Section 2.4, using a regression spline (i.e., unpenalized fixed effects) and varying the e.d.f. by changing the dimension of the basis in gam() in R. Figure 5(a) shows relative bias as a function of the spatial scales involved. As before, I focus on the results below the 1:1 diagonal ($\theta_c = \theta_u$), as this is the scenario of practical interest. By choosing a large number of e.d.f., one can decrease bias more effec-

tively than when estimating the amount of smoothing from the data [i.e., Figure 3(c)]. However, with moderate and large scale variability, the variance of the estimates in this fixed effects model increases dramatically [Figure 5(b)]. This causes a concordant increase in the MSE (not shown), highlighting the bias-variance tradeoff. In contrast, using a penalized spline with fixed e.d.f. [fixing the smoothing parameter in gam() in R] shows much more stable results. As expected, for a given e.d.f. bias is not reduced as much as with a regression spline [Figure 5(a)], but there is much less variability [Figure 5(b)].

A diagnostic approach to understanding whether the residual may include variation from an unmeasured
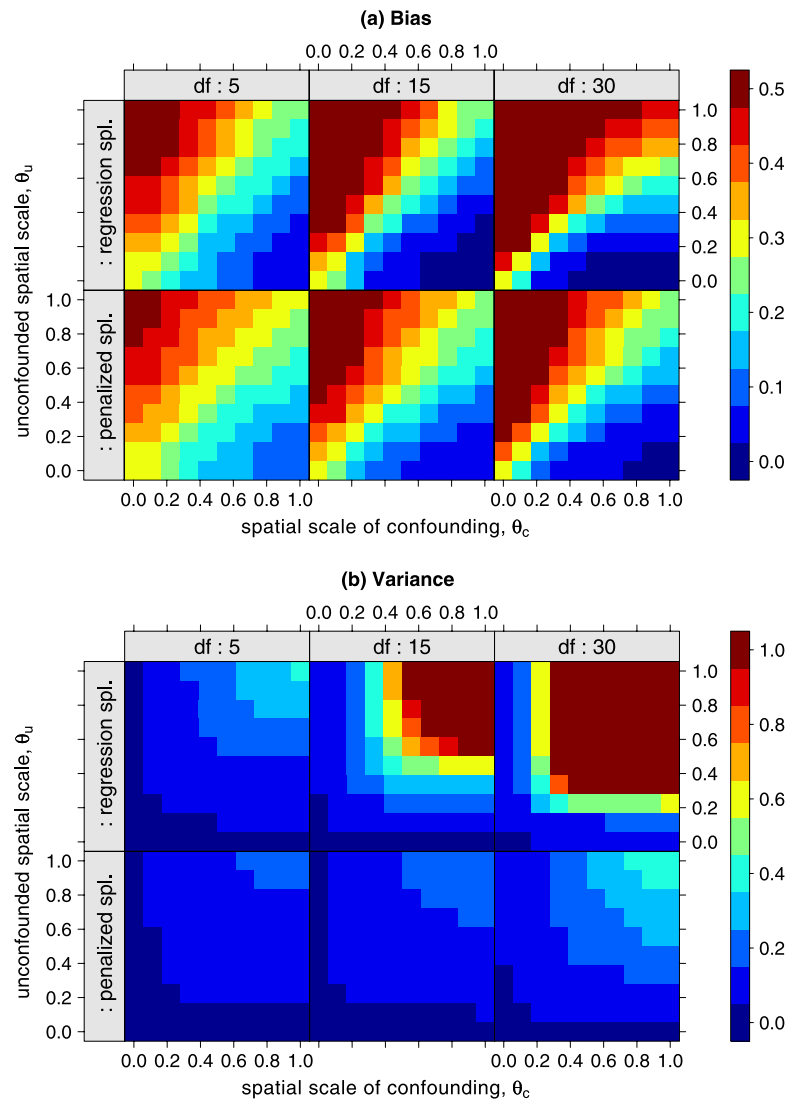

FIG. 5. *Simulation results for relative bias* (a) *and variance* (b) *of* $\hat{\beta}_x$ *as a function of the spatial scales of confounded* (θ_c) *and uncon-founded* (θ_u) *variability for regression splines* (top rows) *and penalized splines* (bottom rows) *with* 5, 15 *and* 30 *e.d.f., where the e.d.f. are pre-specified, rather than estimated based on the data.*

confounder is to assess the correlation between the residual and the exposure. Not knowing β_x, one might use a variety of plausible values of β_x to estimate **g** and then calculate the correlation with **X** (and potentially with filtered versions of **X** that exclude small-scale variation).

2.6 Accounting for Residual Spatial Correlation

If one accounts for large-scale variation as a means of reducing bias, there may still be small-scale residual variation, such as fine-scale correlation in health outcomes related to residential sorting. As I have shown, one can reduce potential confounding bias from this fine-scale variation through explicit spatial modeling only if there is variability in the exposure at an even smaller scale. If there is not, then one is effectively assuming that the fine-scale variation is uncorrelated with the exposure. Given this assumption, one may need to account for the fine-scale residual spatial variation so that uncertainty estimation for β_x is not compromised (but note the results of Section 3.3). One possibility would be to use an analysis robust to misspecification of the residual variance, for example, using an estimating equation with uncertainty based on the sandwich estimator, with regression spline terms in the mean to account for large-scale spatial confounding bias. Alternatively, one could fit a penalized model with the amount of smoothing determined from the data. This has two effects. First, it naturally accounts for the effect of the spatial structure on uncertainty estimation. Second, in the presence of small-scale residual variability, the model will naturally undersmooth with respect to large-scale variability that may cause confounding, thereby reducing bias from confounding at the larger scale, as discussed previously.

3. SPATIALLY CORRELATED RESIDUALS AND PRECISION

In this section I suppose that the residual and the exposure are independent ($\rho = 0$ in the framework of Section 2), which results in unbiased estimation of β_x. I consider effects of spatial scale on the following questions about efficiency of estimators for β_x (henceforth simply β) and quantification of uncertainty:

(1) Given a fixed amount of residual variation, how is efficiency affected by the proportion of that variation that is spatial?
(2) What is the magnitude of the improvement in efficiency when accounting for residual spatial variation, relative to OLS?
(3) If one uses the naive estimator for the variance of the OLS estimator, $\hat{\beta}_{\mathrm{OLS}}$, what is the magnitude of

the error in uncertainty estimation compared to the correct variance estimator for $\hat{\beta}_{\mathrm{OLS}}$?

The first question does not appear to have been raised in the literature. With regard to the second, while we know that GLS is the most efficient estimator when the residuals are correlated, here I investigate the magnitude of this efficiency advantage as a function of the spatial scales involved. Regarding the third, the conventional wisdom in the statistical and applied literature appears to be that not accounting for spatial structure leads to underestimation of uncertainty (e.g., Legendre, 1993; Burnett et al., 2001; Schabenberger and Gotway, 2005, page 324). However, I have not seen a formal quantification of this underestimation for a regression coefficient, in contrast to our understanding of the potentially severe underestimation of uncertainty for the mean of a spatial process (Cressie, 1993, Section 1.3; Schabenberger and Gotway, 2005, Section 1.5).

Note that there are three variance estimators (i.e., estimators for the sampling variability of the estimated regression coefficient) under consideration here: the true GLS variance estimator, and the true and naive OLS variance estimators. When $\rho = 0$, OLS is unbiased, so it makes sense to consider OLS for estimation, provided we adjust the usual OLS variance estimator to account for the residual spatial correlation. While actual applications will likely involve more complicated modeling, consideration of these questions in this simple setting, and with known variance components, helps to understand the basic issues.

3.1 Relationship Between Spatial Scale and GLS Efficiency

Given a fixed amount of residual variation, how is efficiency affected by the proportion of variation that is spatial? I quantify efficiency in terms of precision rather than variance, as this allows for closed form derivations.

LEMMA 3.1. *Consider the model* (3) *and suppose that all parameters are known except β_0 and $\beta \equiv \beta_x$. The expectation of the precision of $\hat{\beta}_{\mathrm{GLS}}$, with respect to the sampling distribution of* **X**, *is*

$$\mathrm{E}_X(\mathrm{Var}(\hat{\beta}_{\mathrm{GLS}})^{-1})$$

(8)
$$= \frac{\sigma_x^2}{\tau^2 + \sigma_g^2}\left(\mathrm{tr}\{\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{R}(\theta_x)\} - \frac{\mathbf{1}^T\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{R}(\theta_x)\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}^T\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{1}}\right),$$

where $\tilde{\Sigma} \equiv (1 - p_g)\mathbf{I} + p_g\mathbf{R}(\theta_g)$, $p_g \equiv \sigma_g^2/(\sigma_g^2 + \tau^2)$, *and the remaining notation follows that in previous sections. See the Appendix for the proof.*

Note that the term in parentheses is an effective sample size, analogous to $n - 1$ in the nonspatial problem. Here the adjustment is for spatial structure in residual and exposure, with the second component in the parentheses analogous to the degree of freedom lost for estimating a mean.

Figure 6(a) shows Monte Carlo estimates of the expected precision as a function of θ_x and θ_g, averaging (8) over 500 sets of $n = 100$ locations simulated uniformly on the unit square. I report the expected precision divided by a baseline of $\sigma_x^2(n - 1)/(\tau^2 + \sigma_g^2)$, which is the expected precision in the nonspatial setting, supposing that the total residual variation, $\tau^2 + \sigma_g^2$, remains constant. Compared to the nonspatial setting, lower precision occurs unless the exposure varies at small spatial scale. When the exposure varies
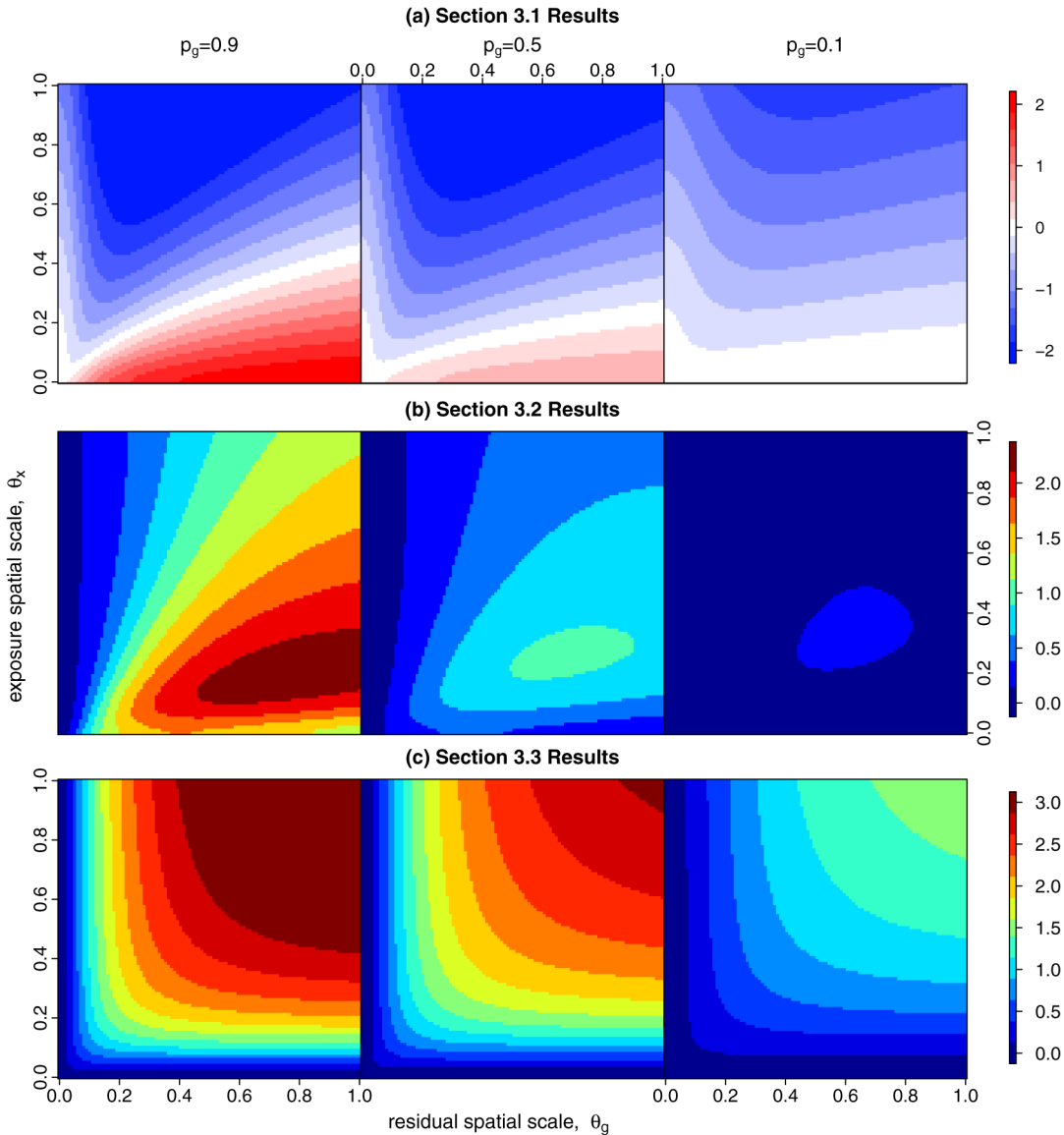

FIG. 6. *Efficiency and precision results for three values of $p_g = \sigma_g^2/(\sigma_g^2 + \tau^2)$ (columns) as a function of the spatial scales of the residual (θ_g) and the exposure (θ_x). (a) The log of the expected precision of the GLS estimator (8), relative to the expected precision in the nonspatial setting with equivalent total residual variation. (b) Relative efficiency of GLS and OLS estimation, quantified as the log of the expected ratio of GLS to OLS precision. (c) The log of the expected ratio of the correct and naive OLS variance estimators (9). The results are based on 500 simulations for each set of parameter values, with a Matérn correlation with $\nu = 2$ and 100 locations sampled uniformly over the unit square.*

at small spatial scale and the residual at larger spatial scale, precision can be substantially greater than in the nonspatial setting. The model is able to account for part of the residual variance through the spatial structure, as if the spatial structure were an additional covariate to which variation in the response is attributed. GLS implicitly estimates the process, $\mathbf{g}$ in (2), that gives rise to the marginalized model (3). This reduces the remaining "unexplained" residual variability and thereby improves efficiency relative to having independent errors but equivalent overall residual variability. In contrast, when $\mathbf{X}$ varies only at large spatial scales, then efficiency decreases because of difficulty in distinguishing $\beta X(\mathbf{s})$ from $g(\mathbf{s})$. Results are similar using points on a regular grid or clustered based on a Poisson cluster process.

3.2 Efficiency of GLS and OLS Estimators

Here I consider how spatial scale affects the relative efficiency of spatial and nonspatial estimators, comparing the precisions of the OLS and GLS estimators. Since the true OLS variance, $[(\mathcal{X}^T\mathcal{X})^{-1}(\mathcal{X}^T(\tau^2\mathbf{I} + \sigma_g^2\mathbf{R}_g)\mathcal{X})(\mathcal{X}^T\mathcal{X})^{-1}]_{2,2}$, is a complicated function, it is difficult to derive closed form expressions for efficiency relative to the GLS estimator. Instead I conduct a small simulation study. For a regular grid of values of θ_x and θ_g, I carry out 500 simulations for each pair of values, with $n = 100$ observations whose spatial locations are drawn uniformly over the unit square domain. Note that I consider the ratio of the GLS precision to the OLS precision, so the values of σ_x^2 and $\sigma_g^2 + \tau^2$ cancel out of the ratio and do not affect the results.

Figure 6(b) shows the Monte Carlo estimates of the expected relative precision, as a function of the spatial scales, θ_g and θ_x, and the proportion of the residual variability that is spatial. When little of the residual variability is spatial ($p_g = 0.1$), there is little gain in precision, as expected. When more is spatial, the gains in precision are small when $\mathbf{g}$ varies at a small scale, but substantial when $\mathbf{g}$ varies at a large scale. Unfortunately, this is also precisely the case in which one would be concerned about spatial confounding. If we suppose that the large-scale structure in the residual has been controlled for in an effort to reduce the potential for bias, then with the remaining residual variability being fine scale, there is limited gain in precision regardless of the spatial scale of the exposure. With locations on a regular grid, the gains in precision are slightly less for small values of θ_g, while with Poisson cluster process sampling, the gains are somewhat larger for small values of θ_g. See also Dow, Burton and White

(1982) for similar simulation results when a Markov random field structure induces the correlation.

3.3 Underestimation of Uncertainty by the Naive OLS Variance Estimator

Applied analyses often ignore residual spatial correlation, raising the question of how strongly uncertainty estimates are affected. One can express the ratio of the true OLS variance to the incorrect naive OLS variance as follows. First define $\mathbf{W} \equiv (\mathbf{X} - \bar{X}\mathbf{1})/s$ where $s^2 \equiv \frac{1}{n}\sum(X_i - \bar{X})^2$. After expressing $\hat{\beta}_x = [(\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\mathbf{Y}]_2 = (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\mathbf{Y}$ where $\tilde{\mathbf{X}} = \mathbf{X} - \bar{X}\mathbf{1}$, we have

$$\frac{\mathrm{Var}_{\text{true}}(\hat{\beta}_x)}{\mathrm{Var}_{\text{naive}}(\hat{\beta}_x)} = \frac{(\sigma_g^2 + \tau^2)^{-1}(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})}{(\sigma_g^2 + \tau^2)^{-1}(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})(\tilde{\mathbf{X}}^T\tilde{\Sigma}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}}$$

$$= \frac{\tilde{\mathbf{X}}^T\tilde{\Sigma}\tilde{\mathbf{X}}}{\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}} = \frac{1}{n}\mathbf{W}^T\tilde{\Sigma}\mathbf{W}.$$

Averaging over the sampling distribution of $\mathbf{X}$, we have

$$(9) \qquad \mathrm{E}_X\left(\frac{1}{n}\mathbf{W}^T\tilde{\Sigma}\mathbf{W}\right) = \frac{1}{n}\mathrm{tr}(\tilde{\Sigma}\,\mathrm{Cov}(\mathbf{W})).$$

So for $\tilde{\Sigma} \approx \mathbf{I}$ or $\mathrm{Cov}(\mathbf{W}) \approx \mathbf{I}$, that is, when either θ_g or θ_x is close to zero, we expect the ratio to be near one. Note also that with spatial correlation functions that are nonnegative, the only negative contribution to the ratio can be from negative covariances induced by standardizing $\mathbf{X}$. Such negative covariances should diminish as the sample size increases, so we expect the ratio to generally be no smaller than one, indicating that the naive variance does underestimate uncertainty. Finally, the largest values of the ratio would occur with large positive correlations in corresponding elements of $\tilde{\Sigma}$ and $\mathrm{Cov}(\mathbf{W})$, which is to be expected when both $\mathbf{g}$ and $\mathbf{X}$ show large-scale variation.

Figure 6(c) supports these heuristic results, showing the average ratio of variances in simulations, where the simulations are conducted as in Section 3.2. The ratio is close to one when either of the spatial terms has fine-scale variability and far from one when both have large-scale behavior. This result is similar to that of Bivand (1980) for inference about a correlation coefficient and to (Johnston and DiNardo, 1997, page 178) under serial autocorrelation in a regression setting. As expected, when the proportion of residual variability is smaller (moving from the bottom left to bottom right panels), the expected ratio gets closer to one. This indicates that when nonspatial variation dominates the

residual and the spatial structure in the residual or exposure is not too large in scale, the naive variance estimator may be reasonable. A lack of large-scale residual structure might result from having accounted for large-scale variation in attempting to reduce spatial confounding bias. Results with gridded locations show ratios slightly closer to one, and with clustered locations, ratios further from one. Note that the uncertainty estimate in any given naive analysis may be larger than when fitting a spatial model because the more sophisticated model both corrects the variance estimate, which increases the estimated uncertainty, and uses a more efficient estimator, which decreases the fundamental uncertainty.

Simple simulations with spatial ranges and sampling designs specific to an analysis could be easily carried out for further guidance in a given setting, allowing one to assess whether ignoring the spatial structure has substantial impact on uncertainty estimation. Accounting for small-scale spatial correlation requires estimation of the spatial structure and is often computationally burdensome, so an assumption of independence can have an important practical benefit. Of course in some analyses, any underestimation of variability may be cause for concern, in which case use of the naive variance estimator would not be tenable.

4. CASE STUDY: BIRTHWEIGHT AND AIR POLLUTION

Chronic health effects of ambient air pollution in developed countries involve small relative risks, but are of considerable public health importance because of widespread exposure. Epidemiologic studies attempt to estimate a small effect from data with high levels of variability and stronger effects from other covariates, including potential confounders such as socioeconomic status, so spatial confounding bias is of critical concern.

I reanalyze data on the association between ambient air pollution (estimates of black carbon, a component of particulate matter) and birthweight in eastern Massachusetts (Zeka, Melly and Schwartz, 2008; Gryparis et al., 2009). These analyses found significant negative effects of traffic proxy variables and black carbon, respectively, on birthweight. Gryparis et al. (2009) used several methods to try to account for effects of measurement error in the predicted black carbon concentrations, which are based on a regression model that accounts for spatial and temporal structure and key covariates.

I follow these analyses in using an extensive set of covariates to try to account for potential confounding. I use smooth terms for mother's age, gestational age and mother's cigarette use, to account for nonlinearities, a linear term for census tract income, and categorical variables for the following: presence of a health condition of the mother, previous preterm birth, previous large birth, sex of baby, year of birth, index of prenatal care and maternal education. The exposure of interest is the estimated nine-month average black carbon concentration at the geocoded address of the mother, based on a black carbon prediction model (Gryparis et al., 2007). Following Gryparis et al. (2009), for simplicity, I exclude the 13,347 observations with any missing covariate values, giving 205,713 births.

In Gryparis et al. (2009) we found no evidence of residual spatial correlation based on a spatial semivariogram. Further analysis here indicates that there is significant residual spatial variation but that nonspatial variation overwhelms the magnitude of this variation. Figure 7(a) is a semivariogram showing no evidence of spatial structure, while a spatial smooth of model residuals [Figure 7(b)] indicates clear spatial structure. While individual nonspatial variability among babies swamps the spatial variation (hence the flat semivariogram), it is large relative to the estimated pollution effect (note the surface values in the range of -40 to 40, for comparison with effect estimates in Figure 8). Thus, if the residual spatial variation is caused by spatially varying confounders, it could bias estimation of the pollution effect.

To include a spatial term in models of birthweight, I consider a regression spline, an unpenalized approach, and a penalized spline, both with e.d.f. chosen in advance (see Section 2.5), as well as a penalized spline with data-driven smoothing parameter estimation based on GCV, all implemented in gam() in R, using the thin plate spline basis. Note that the thin plate regression spline approach implemented in gam() should minimize sensitivity to knot placement (Wood, 2006).

I first add a spatial term to the model with the full set of covariates to assess whether some of the estimated effect may be biased by spatial confounding. Figure 8(a) shows how the estimated effect of black carbon varies with the e.d.f. and the spatial smoothing approach. The estimate attenuates somewhat as more e.d.f. are used to account for the spatial structure. For the penalized spline, as more than about 10 e.d.f. are used, the upper confidence limit exceeds zero, and for larger e.d.f., the upper limit increases further. GCV
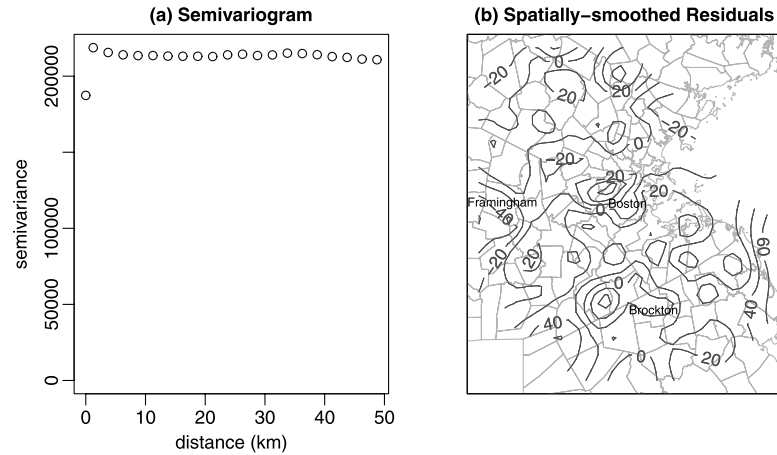
FIG. 7. (a) *Semivariogram of full model residuals, with the first point representing births to mothers living at the same location.* (b) *Spatial smooth of residuals with town boundaries in grey. The spatial smooth, with* 129 *e.d.f. chosen by GCV, is highly significant.*

chooses 157 e.d.f., indicating fairly small-scale spatial structure in the data. For context note that with 129 e.d.f. in Figure 7(b) we see spatial features at the scale of individual towns. While the regression spline approach implemented here avoids having to choose the knots, the empirical results are still very sensitive to e.d.f., in contrast to the stability of the penalized spline solution as the e.d.f. varies. For both penalized and regression splines, there is a clear bias-variance trade-off, with increasing variance as the number of e.d.f. increases. However, for this problem with a very large sample size, the confidence intervals do not increase drastically, nor is there much difference in the uncertainty between the regression and penalized spline ap-

proaches. The spatial confounding assessment suggests that while we have somewhat reduced confidence in the black carbon effect, the effect estimate is reasonably stable even when using a spatial term with a large number of degrees of freedom.

Next I consider what might have happened if most of the covariates (particularly the ones related to socioeconomic status) were not measured, potentially inducing serious confounding. Figure 8(b) indicates that without any spatial term in the model, the effect estimate is -23.0 with a 95% confidence interval of $(-26.8, -19.2)$, indicating a much more substantial effect of black carbon than the fully adjusted model. As soon as one accounts for spatial structure, even with
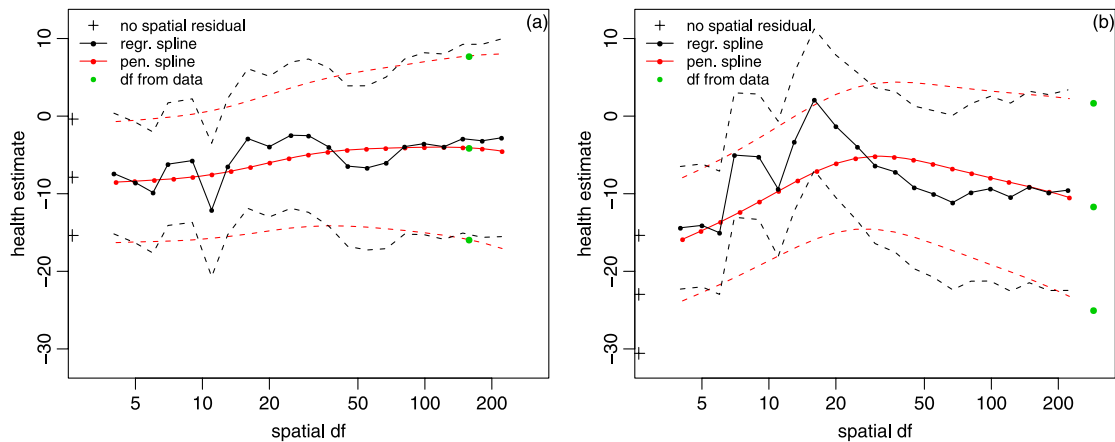

FIG. 8. *For the model with the full set of covariates* (a) *and the reduced set of covariates* (b), *black carbon effect estimates and* 95% *confidence intervals based on different specifications for the spatial term in an additive model: black pluses indicate the model with no spatial term and green dots with the e.d.f. chosen by GCV, while black (regression spline) and red (penalized spline) dots indicate results when fixing the degrees of freedom at a set of discrete values. The lines through the points and corresponding dashed lines are taken by connecting the effect estimate and confidence interval bounds for the discrete set.*

a small number of e.d.f., the estimate attenuates, approaching the fully adjusted estimate, with the upper confidence limit rising above zero. The reduced model appears to suffer from serious confounding, with the estimated pollution effect apparently driven by large-scale association of pollution and birthweight. The spatial analysis is able to account for much of this apparent confounding, substituting for a rich set of covariates.

Ideally one would fit a model that accounts for fine-scale spatial structure to improve one's confidence in the uncertainty estimation. However, with 205,713 observations, this is a computational challenge that I do not take up here. Given the results in Section 3.3 that indicate that large-scale structure causes most of the variance underestimation, one can hope that the uncertainty at the larger values for the spatial e.d.f. in Figure 8 may reasonably approximate the true uncertainty.

5. DISCUSSION

Considerations of scale are critical in spatial regression problems. Standard spatial regression models, which use spatial random effects, kriging specifications or a penalized spline to represent the spatial structure, are penalized models with inherent bias-variance tradeoffs in estimating the smooth function. Under unmeasured spatial confounding, the bias carries over into estimating the coefficient for the exposure of interest, but the degree of bias depends on the spatial scales involved. Inclusion of a spatial residual term accounts for spatial correlation in the sense of reducing bias from unmeasured spatial confounders only when there is unconfounded variability in the exposure at a scale smaller than the scale of the confounding. If the variation in exposure is solely at large scales, there is little opportunity to reduce spatial confounding bias, but with a component of small-scale exposure variability, large-scale spatial confounding bias can be reduced substantially. Accounting for large-scale residual correlation is also important for improving precision of regression estimators and for correctly estimating uncertainty. In contrast, when residual correlation occurs at small scales, there is little opportunity for reducing spatial confounding bias at those scales or improving regression estimator precision. However, under the assumption of no small-scale confounding, fitting such residual structure can reduce bias from larger scale confounding by causing undersmoothing with respect to the large-scale structure. While the results here are limited to the simple setting of linear regression/additive models with a single covariate and

single unmeasured confounder, I expect that the qualitative results and principles hold in more complicated settings, with no reason to believe that the bias results would improve in more complicated models.

Sensitivity analyses that show the bias-variance tradeoff as a function of the scale at which the spatial residual structure is modeled (Peng, Dominici and Louis, 2006; Zeger et al., 2007) offer one approach that helps to frame the issue of bias in the context of the spatial scales involved. In choosing a spline formulation to carry out such an analysis, while a regression spline has an appealing interpretation and in theory result in less bias in estimating the effect of interest, a penalized spline with a fixed effective degrees of freedom may give more stable results. Of course the sensitivity analysis approach does not answer the question of how to get a single estimate of the effect of interest. One might also consider an approach similar to that of Beelen et al. (2007) and explicitly decompose the exposure into multiple scales, including exposure at each scale as a separate covariate and focus causal interpretation on the effect estimates for the smaller scales (e.g., Janes, Dominici and Zeger, 2007). Lu and Zeger (2007) use matching estimators for each pair of observations and assess how effect estimates vary with spatial lag between the pairs to assess sensitivity. Note that estimating equation approaches are not capable of reducing bias from unmeasured spatial confounding because the marginal variance is assumed to be unrelated to the exposure and variation is not attributed to a spatial term.

From the econometric perspective, spatial confounding bias might be seen as a type of endogeneity bias, with exposure the endogenous variable and the unconfounded component of exposure, or some proxy for it, an exogenous variable. Since the unconfounded component is not measured directly, some sort of scale decomposition appears necessary. Standard endogenous variable techniques such as two-stage least squares and instrumental variable methods (Johnston and DiNardo, 1997) do not appear directly useful but do share commonalities with approaches mentioned above.

Others have noted the identifiability problems in spatial models, with sensitivity of effect estimates to inclusion of a spatial residual term when the covariates vary spatially (Breslow and Clayton, 1993; Clayton, Bernardinelli and Montomoli, 1993; Burden et al., 2005; Lawson, 2006, page 187; Augustin et al., 2007; Wakefield, 2007). A different methodologic perspective than that presented here has been taken by Reich, Hodges and Zadnik (2006) and Houseman,

Coull and Shine (2006), who estimate the effect of exposure, $\mathbf{X}$, by forcing the spatial residual to be orthogonal to $\mathbf{X}$, attributing as much variability as possible to $\mathbf{X}$. This approach makes a very strong assumption of no confounding to avoid overadjustment bias from accidentally accounting for some of the effect of the covariate in the residual. Note that the residuals and covariates are not orthogonal under GLS estimation (Schabenberger and Gotway, 2005, page 349). Gustafson and Greenland (2006) confront a similar problem of modeling systematic residual confounding in a context with identifiability problems, finding that imposing structure through a prior distribution in a nonidentified model can help account for a portion of the confounding, improving bias and precision of estimators.

Note that measurement error in the exposure is of critical concern, because reducing bias relies on estimating variability in exposure at scales smaller than the confounding. In many contexts, measurement error becomes an increasing concern at small scales because of limitations in measurement resources. In contrast, large-scale exposure variation may be well estimated using spatial smoothing and regression models, thereby inducing Berkson-type error through what is effectively regression calibration (Gryparis et al., 2009). To the extent to which accounting for bias forces one to rely on exposure estimates more likely contaminated by classical measurement error, one may find oneself reducing bias from confounding only to increase it from measurement error. To the extent small-scale variation is affected by Berkson error, one would increase variance but not incur bias by relying on the small-scale variation.

Finally note that in many settings one has aggregated exposure and outcome data, so one has limited ability to identify effects of exposure based on fine-scale variation because the aggregation eliminates the fine-scale variation (e.g., Janes, Dominici and Zeger, 2007). This suggests that accounting for spatial confounding with areal data, for which researchers often use standard conditional auto-regressive models, is likely to be ineffective when aggregating over large areal units, which is consistent with the bias seen in Richardson (2003). In work concurrent with that presented here, Hodges and Reich (2010) have investigated bias in the areal setting under a variety of perspectives on the spatial random effects, also making the case for the approach taken in Reich, Hodges and Zadnik (2006).

APPENDIX: PROOF OF LEMMA 3.1

From the definition of the GLS estimator, we have

$$\mathrm{Var}(\hat{\beta}_{\mathrm{GLS}}) = [\boldsymbol{\mathcal{X}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{X}}]_{2,2}^{-1}$$

$$= \frac{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} - \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}}.$$

Using the definitions of $\tilde{\boldsymbol{\Sigma}}$ and p_g, and taking the reciprocal, we have

$$\mathrm{Prec}(\hat{\beta}_{\mathrm{GLS}}) = \frac{1}{\sigma_g^2 + \tau^2} \Big(\mathbf{X}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{X}$$

$$- \frac{\mathbf{X}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{1} \mathbf{1}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{X}}{\mathbf{1}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{1}} \Big).$$

Conclude by taking the expectation with respect to the sampling distribution of $\mathbf{X}$, using the expectation of a quadratic form, and rearranging the matrices inside the second trace to give a scalar:

$$\mathrm{E}_X(\mathrm{Prec}(\hat{\beta}_{\mathrm{GLS}}))$$

$$= \frac{\sigma_x^2}{\sigma_g^2 + \tau^2} \Big(\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{R}(\theta_x))$$

$$- \frac{\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{1} \mathbf{1}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{R}(\theta_x))}{\mathbf{1}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{1}}$$

$$+ \mu_x^2 \mathbf{1}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{1} - \frac{\mu_x^2 \mathbf{1}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{1} \mathbf{1}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{1}}{\mathbf{1}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{1}} \Big)$$

$$= \frac{\sigma_x^2}{\sigma_g^2 + \tau^2} \Big(\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{R}(\theta_x)) - \frac{\mathbf{1}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{R}(\theta_x) \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{1}}{\mathbf{1}^T \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{1}} \Big).$$

REFERENCES

AUGUSTIN, N., LANG, S., MUSIO, M. and VON WILPERT, K. (2007). A spatial model for the needle losses of pine-trees in the forests of Baden–Wurttemberg: An application of Bayesian structured additive regression. *J. Roy. Statist. Soc. Ser. C* **56** 29–50. MR2339161

BEELEN, R., HOEK, G., FISCHER, P., BRANDT, P. and BRUNEKREEF, B. (2007). Estimated long-term outdoor air pollution concentrations in a cohort study. *Atmospheric Environment* **41** 1343–1358.

BIGGERI, A., BONANNINI, M., CATELAN, D., DIVINO, F., DREASSI, E. and LAGAZIO, C. (2005). Bayesian ecological regression with latent factors: Atmospheric pollutants, emissions, and mortality for lung cancer. *Environmental and Ecological Statistics* **12** 397–409. MR2196329

BIVAND, R. (1980). A Monte Carlo study of correlation coefficient estimation with spatially autocorrelated observations. *Quaestiones Geographicae* **6** 5–10.

BRESLOW, N. E. and CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88** 9–25.

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17** 453–510. MR0994249

BURDEN, S., GUHA, S., MORGAN, G., RYAN, L., SPARKS, R. and YOUNG, L. (2005). Spatio-temporal analysis of acute admissions for ischemic heart disease in NSW, Australia. *Environmental and Ecological Statistics* **12** 427–448. MR2196331

BURNETT, R., MA, R., JERRETT, M., GOLDBERG, M., CAKMAK, S., POPE III, C. and KREWSKI, D. (2001). The spatial association between community air pollution and mortality: A new method of analyzing correlated geographic cohort data. *Environmental Health Perspectives* **109** 375–380.

CAKMAK, S., BURNETT, R., JERRETT, M., GOLDBERG, M., POPE III, C., MA, R., GULTEKIN, T., THUN, M. and KREWSKI, D. (2003). Spatial regression models for large-cohort studies linking community air pollution and health. *Journal of Toxicology and Environmental Health, Part A* **66** 1811–1823.

CERDÁ, M., TRACY, M., MESSNER, S., VLAHOV, D., TARDIFF, K. and GALEA, S. (2009). Misdemeanor policing, physical disorder, and gun-related homicide: A spatial analytic test of "broken-windows" theory. *Epidemiology* **20** 533–541.

CHO, W. (2003). Contagion effects and ethnic contribution networks. *American Journal of Political Science* **47** 368–387.

CLAESKENS, G., KRIVOBOKOVA, T. and OPSOMER, J. (2009). Asymptotic properties of penalized spline estimators. *Biometrika* **96** 529–544. MR2538755

CLAYTON, D., BERNARDINELLI, L. and MONTOMOLI, C. (1993). Spatial correlation in ecological analysis. *International Journal of Epidemiology* **22** 1193–1202.

CRESSIE, N. (1993). *Statistics for Spatial Data*, rev. ed. Wiley-Interscience, New York. MR1239641

DIGGLE, P., HEAGERTY, P. J., LIANG, K.-Y. and ZEGER, S. (2002). *Analysis of Longitudinal Data*, 2nd ed. Oxford Univ. Press, Oxford. MR2049007

DOMINICI, F., MCDERMOTT, A. and HASTIE, T. (2004). Improved semiparametric time series models of air pollution and mortality. *J. Amer. Statist. Assoc.* **99** 938–949. MR2113312

DOW, M., BURTON, M. and WHITE, D. (1982). Network autocorrelation: A simulation study of a foundational problem in regression and survey research. *Social Networks* **4** 169–200.

GRYPARIS, A., COULL, B., SCHWARTZ, J. and SUH, H. (2007). Latent variable semiparametric regression models for spatio-temporal modeling of mobile source pollution in the greater Boston area. *J. Roy. Statist. Soc. Ser. C* **56** 183–209. MR2359241

GRYPARIS, A., PACIOREK, C., ZEKA, A., SCHWARTZ, J. and COULL, B. (2009). Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics* **10** 258–274.

GUSTAFSON, P. and GREENLAND, S. (2006). The performance of random coefficient regression in accounting for residual confounding. *Biometrics* **62** 760–768. MR2247204

HE, S., MAZUMDAR, S. and ARENA, V. (2006). A comparative study of the use of GAM and GLM in air pollution research. *Environmetrics* **17** 81–93. MR2222035

HODGES, J. and REICH, B. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. Technical Report 2010-002, Division of Biostatistics, Univ. Minnesota. Available at http://www.biostat.umn.edu/ftp/pub/2010/rr2010-002.pdf.

HOUSEMAN, E., COULL, B. and SHINE, J. (2006). A nonstationary negative binomial time series with time-dependent covariates: Enterococcus counts in Boston harbor. *J. Amer. Statist. Assoc.* **101** 1365–1376. MR2307571

JANES, H., DOMINICI, F. and ZEGER, S. (2007). Trends in air pollution and mortality: An approach to the assessment of unmeasured confounding. *Epidemiology* **18** 416–423.

JOHNSTON, J. and DINARDO, J. (1997). *Econometric Methods*, 4th ed. McGraw-Hill, New York.

LAWSON, A. (2006). *Statistical Methods in Spatial Epidemiology*, 2nd ed. Wiley, New York. MR2243369

LEE, D., FERGUSON, C. and MITCHELL, R. (2009). Air pollution and health in Scotland: A multicity study. *Biostatistics* **10** 409–423.

LEGENDRE, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology* **74** 1659–1673.

LOMBARDÍA, M. J. and SPERLICH, S. (2007). Multi-level regression between fixed effects and mixed effects models. Technical report, Georg-August-Univ. Göttingen. Available at http://www.zfs.uni-goettingen.de/index.php?id=54.

LU, Y. and ZEGER, S. (2007). Decomposition of regression estimators to explore the influence of "unmeasured" time-varying confounders. Technical Report 159, Dept. Biostatistics, Johns Hopkins Univ., Baltimore, MD. Available at http://www.bepress.com/jhubiostat/paper159.

MOLITOR, J., JERRETT, M., CHANG, C. ET AL. (2007). Assessing uncertainty in spatial exposure models for air pollution health effects assessment. *Environmental Health Perspectives* **115** 1147–1153.

PENG, R., DOMINICI, F. and LOUIS, T. (2006). Model choice in time series studies of air pollution and mortality. *J. Roy. Statist. Soc. Ser. A* **169** 179–203. MR2225539

POPE III, C., BURNETT, R., THUN, M., CALLE, E., KREWSKI, D., ITO, K. and THURSTON, G. (2002). Lung cancer, cardiopulmonary mortality and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association* **287** 1132–1141.

RAMSAY, T., BURNETT, R. and KREWSKI, D. (2003). Exploring bias in a generalized additive model for spatial air pollution data. *Environmental Health Perspectives* **111** 1283–1288.

REICH, B., HODGES, J. and ZADNIK, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* **62** 1197–1206. MR2307445

RICE, J. (1986). Convergence rate for partially linear splined models. *Statist. Probab. Lett.* **4** 203–208. MR0848718

RICHARDSON, S. (2003). Spatial models in epidemiological applications. In *Highly Structured Stochastic Systems* (P. Green, N. Hjort and S. Richardson, eds.) 237–259. Oxford Univ. Press, Oxford. MR2082412

RUPPERT, D., WAND, M. and CARROLL, R. (2003). *Semiparametric Regression*. Cambridge Univ. Press, Cambridge. MR1998720

SCHABENBERGER, O. and GOTWAY, C. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman & Hall, Boca Raton, FL. MR2134116

SPECKMAN, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B* **50** 413–436. MR0970977

WAKEFIELD, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics* **8** 158–183.

WALLER, L. and GOTWAY, C. (2004). *Applied Spatial Statistics for Public Health Data*. Wiley, Hoboken, NJ. MR2075123

WOOD, S. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall, Boca Raton, FL. MR2206355

ZEGER, S., DOMINICI, F., MCDERMOTT, A. and SAMET, J. (2007). Mortality in the Medicare population and chronic exposure to fine particulate air pollution. Technical Report 133, Dept. Biostatistics, Johns Hopkins Univ., Baltimore, MD. Available at http://www.bepress.com/jhubiostat/paper133.

ZEKA, A., MELLY, S. and SCHWARTZ, J. (2008). The effects of socioeconomic status and indices of physical environment on reduced birth weight and preterm births in eastern Massachusetts. *Environmental Health* **7** 60.