

Efficient Measurement Error Correction with Spatially Misaligned Data

Adam A. Szpiro*, Lianne Sheppard, Thomas Lumley

Department of Biostatistics, University of Washington, Seattle WA 98195, USA
aszpiro@u.washington.edu

SUMMARY

Association studies in environmental statistics often involve exposure and outcome data that are misaligned in space. A common strategy is to employ a spatial model such as universal kriging to predict exposures at locations with outcome data and then to estimate the regression parameter of interest based on the predicted exposures. This procedure results in measurement error because the predicted exposures do not correspond exactly to the true values. We characterize the measurement error by decomposing it into Berkson-like and classical-like components. An important effect of both components is to change the variability of parameter estimates, so naïve standard errors are incorrect. The classical-like component can also introduce bias in the parameter estimates. We focus on deriving corrected standard errors, and one approach to doing this is the parametric bootstrap. While effective, the parametric bootstrap is computationally intensive since it requires solving a nonlinear optimization problem to estimate the exposure model parameters in each bootstrap sample. We propose a less computationally intensive alternative termed the “parameter bootstrap” that exploits our decomposition of the measurement error into

*To whom correspondence should be addressed.

Berkson-like and classical-like components. The primary advantage of the parameter bootstrap is that it only requires solving one nonlinear optimization problem. We illustrate our methodology in simulations and with forestry data from the Environmental Protection Agency.

Keywords: Measurement error; Kriging; Exposure modeling; Environmental epidemiology; Environmental statistics

1. INTRODUCTION

A major challenge for association studies in environmental statistics is that we cannot directly measure the exposure at every location where there is outcome data. Modern Geographic Information System (GIS) technology makes it feasible to sample environmental exposures at a relatively small number of monitoring locations and then to predict exposures at unmonitored locations using a statistical model such as universal kriging that exploits dependence on GIS covariates and incorporates spatial smoothing (Cressie 1993). The overall strategy is to use predicted exposures in place of the true exposures at locations with outcome data in order to estimate the parameter of interest in a regression model. The problem that we address in this paper is how to ensure valid inference in light of the measurement error that results from using predicted exposures in place of the exact values.

An example application in environmental epidemiology is evaluating the relationship between exposure to ambient air pollution and adverse health outcomes. Many studies have documented adverse effects of air pollution (Dockery et al. 1993; Samet et al. 2000; Pope et al. 2002), but until recently most studies have assigned exposures based on area-wide monitored concentrations. More recent studies have emphasized the importance of using predicted individual air pollution exposures to account for spatial variability of pollution levels within urban areas (Jerrett et al. 2005b; Kunzli et al. 2005; Gryparis et al. 2007; Szpiro et al. 2009).

Other environmental applications that do not involve human health effects are analogous from a statistical perspective. An example that we will return to later in this paper involves assessing the relationship between chloride levels in streams and nearby watershed land cover (Madsen et al. 2008; Herlihy et al. 1998). Another application involves assessing the relationship between changes in annual rainfall and vegetation cover. This has been studied in the African Sahel, where remotely sensed vegetation levels are available everywhere but rainfall levels are only monitored at a limited number of locations (Los et al. 2006; Lindström and Lindgren 2008).

Various methods have been employed for predicting exposures, including nearest neighbor interpolation (Miller et al. 2007), regression based on GIS covariates (Brauer et al. 2003; Jerrett et al. 2005a), interpolation by a geostatistical method such as kriging (Jerrett et al. 2005b; Kunzli et al. 2005), and semi-parametric smoothing (Gryparis et al. 2007; Kunzli et al. 2005). All of these methods result in measurement error that does not fit into the standard categories of classical or Berkson error (Carroll et al. 2006). In this paper, we focus on characterizing and correcting measurement error for exposures predicted by universal kriging.

In a recent paper, Kim et al. (2009) have shown that using predicted exposures from kriging performs better than nearest neighbor interpolation, but depending on the range of spatial correlation in the exposure surface, significant errors may remain resulting in confidence intervals that do not provide correct coverage. Gryparis et al. (2009) review the relevant measurement error literature and compare several correction strategies in a simulation study, and Madsen et al. (2008) apply a version of the parametric bootstrap to obtain corrected standard errors.

The parametric bootstrap is effective, but it is computationally intensive since it requires solving a nonlinear optimization problem to estimate the exposure model parameters in each bootstrap sample. We

describe a new method termed the “parameter bootstrap” that can be viewed as a much less computationally demanding approximation to the parametric bootstrap. The parameter bootstrap is motivated by a decomposition of the measurement error into two approximately independent components, one of which is similar to Berkson error (“Berkson-like”) and the other of which is similar to classical measurement error (“classical-like”). We develop our methodology in a setting where we use universal kriging to predict the exposure and where we model the association of interest with linear regression, including the possibility of spatially correlated residuals. While our primary focus is on calculating standard errors, the classical-like component of measurement error can also result in biased effect estimates. The bias is small in the examples we consider, but it is important to note that it can be away from the null, rather than toward the null as in classical measurement error. We discuss the origin and properties of the bias further in Appendix A.

The remainder of this paper is organized as follows. We begin in Section 2 by introducing notation and formally setting out the problem. In Section 3 we characterize the measurement error by decomposing it into Berkson-like and classical-like components, and in Section 4 we define the parametric and parameter bootstraps and give some details on implementation. In Section 5 we illustrate our methodology in a simulation study, and in Section 6 we consider an example with publicly available forestry data from the Environmental Protection Agency. We conclude in Section 7 with a discussion.

2. NOTATION AND PROBLEM SETUP

We consider an association study with the $N \times 1$ vector of observed outcomes Y , $N \times 1$ vector of exposures X , and $N \times m$ matrix of covariates Z . Assume a linear regression model

$$Y = \beta_0 + X\beta_X + Z\beta_Z + \varepsilon, \tag{2.1}$$

with intercept β_0 , regression coefficient β_X for the exposure, and $m \times 1$ vector of coefficients β_Z for the covariates. Assume that ε is an $N \times 1$ random vector distributed as $N(0, \Sigma_\varepsilon(\theta_\varepsilon))$, for a positive definite matrix function $\Sigma_\varepsilon(\cdot)$ and unknown parameter θ_ε .

The goal is inference for β_X , which would be straightforward if X , Y , and Z were all observed. In that case, we could estimate $\hat{\beta}_X$ by ordinary least squares (OLS) and then estimate $\hat{\theta}_\varepsilon$ from the residuals and use a sandwich-based standard error (Liang and Zeger 1986). If the ε are independent and we estimate $\hat{\theta}_\varepsilon$ by the method-of-moments, the sandwich form reduces to the classical standard error estimate.

We are interested in the situation where Y and Z are observed, but instead of X we observe the $N^* \times 1$ vector X^* of exposures at different locations. We refer to N^* as the number of exposure monitors. Assume that X and X^* are jointly distributed as

$$\begin{pmatrix} X \\ X^* \end{pmatrix} = \begin{pmatrix} S \\ S^* \end{pmatrix} \alpha + \begin{pmatrix} \eta \\ \eta^* \end{pmatrix}. \quad (2.2)$$

In this expression, S and S^* are known $N \times k$ and $N^* \times k$ dimensional matrices of GIS covariates, α is an unknown $k \times 1$ vector of coefficients, and

$$\begin{pmatrix} \eta \\ \eta^* \end{pmatrix} \sim N\left(0, \Sigma_{(\eta\eta^*)}(\theta_\eta)\right), \quad (2.3)$$

independent of ε , for a positive definite matrix function $\Sigma_{(\eta\eta^*)}(\cdot)$ and unknown parameter θ_η . It is useful to introduce the decomposition

$$\Sigma_{(\eta\eta^*)}(\cdot) = \begin{pmatrix} \Sigma_\eta(\cdot) & \Sigma_{\eta\eta^*}(\cdot) \\ \Sigma_{\eta^*\eta}(\cdot) & \Sigma_{\eta^*}(\cdot) \end{pmatrix}.$$

Universal kriging is a special case if θ_η comprises the range, partial sill, and nugget parameters from a geostatistical model (Cressie 1993). More general spatio-temporal correlation structures also fit naturally in this framework (Banerjee et al. 2004; Szpiro et al. 2009).

Although the exposure X is not observed directly, we can exploit the observed values X^* and the spatial model in equation (2.2) to estimate $\hat{\beta}_X$ as follows. First, we estimate the exposure model parameters $\hat{\alpha}$ and $\hat{\theta}_\eta$ based on X^* by maximum likelihood or another nonlinear optimization approach, and then we define the estimated exposure by

$$\begin{aligned} W &= \text{E}(X|X^*; \hat{\alpha}, \hat{\theta}_\eta) \\ &= \Phi(X^*, \hat{\alpha}, \hat{\theta}_\eta) \end{aligned} \tag{2.4}$$

where

$$\Phi(X^*, \hat{\alpha}, \hat{\theta}_\eta) = S\hat{\alpha} + \Sigma_{\eta\eta^*}(\hat{\theta}_\eta)\Sigma_{\eta^*}^{-1}(\hat{\theta}_\eta)(X^* - S^*\hat{\alpha}).$$

Since we are interested in frequentist sampling properties of an estimator for β_X , we take care to specify the assumed data-generating mechanism. All of the geographic locations are assumed to be fixed and known, as are the corresponding GIS covariates S and S^* and any covariates Z in the outcome model. The regression coefficients $\beta_0, \beta_X, \beta_Z$, and α and variance parameters θ_ε and θ_η are all fixed but unknown. A random realization from the data-generating mechanism is obtained by drawing a sample from the joint distribution of ε , η , and η^* .

3. MEASUREMENT ERROR

If we ignore the measurement error from using W in place of X , we can derive naïve standard errors by the procedure described at the beginning of Section 2. However, these standard errors are based on the assumption that all sampling variability in $\hat{\beta}_X$ is induced by ε , and they ignore the additional sampling variability from $U = X - W$ that is in turn induced by η and η^* . Therefore, naïve standard errors will typically not estimate the true sampling variability of $\hat{\beta}_X$, and a correction procedure is needed in order

make accurate inferential statements.

We decompose the measurement error into two components

$$\begin{aligned}
 U &= X - W \\
 &= X - E(X|X^*; \hat{\alpha}, \hat{\theta}_\eta) \\
 &= \left(X - E(X|X^*; \alpha, \theta_\eta) \right) + \left(E(X|X^*; \alpha, \theta_\eta) - E(X|X^*; \hat{\alpha}, \hat{\theta}_\eta) \right) \\
 &= U_{BL} + U_{CL},
 \end{aligned}$$

where the Berkson-like component is

$$U_{BL} = X - E(X|X^*; \alpha, \theta_\eta), \quad (3.1)$$

and the classical-like component is

$$U_{CL} = E(X|X^*; \alpha, \theta_\eta) - E(X|X^*; \hat{\alpha}, \hat{\theta}_\eta).$$

The Berkson-like component U_{BL} accounts for variability from η and η^* , conditional on known exposure model parameters, and the classical-like component U_{CL} incorporates additional variability from η^* in estimating the exposure model parameters. Both of these components inflate the sampling variance of $\hat{\beta}_X$. The classical-like component can also introduce bias, but we focus on the sampling variability since our simulations in Section 5 indicate this is the dominant error in the scenarios we have considered.

3.1 Berkson-Like Component of the Error

In this subsection, we assume that the exposure model parameters α and θ_η are known so that U_{BL} is the only source of measurement error. A primary feature of Berkson error is that it has mean zero conditional on the estimated exposure W (Carroll et al. 2006, page 9). With known exposure model parameters, it is

easy to see that this holds for U_{BL}

$$\begin{aligned}
 E(U_{BL}|W) &= E\left(E(U_{BL}|W)|X^*\right) \\
 &= E\left(U_{BL}|X^*\right) \\
 &= E\left(X - E(X|X^*; \alpha, \theta_\eta)|X^*\right) \\
 &= 0.
 \end{aligned}$$

The second line holds since W is deterministic conditional on X^* , the third line is the definition of U_{BL} , and the final line holds since α and θ_η are the parameters in the data-generating mechanism for X .

Since we can rewrite equation (2.1) in the form

$$Y = \beta_0 + W\beta_X + Z\beta_Z + U_{BL}\beta_X + \varepsilon,$$

it is easy to see that $\hat{\beta}_X$ derived by OLS with W in place of X is unbiased for estimating β_X . We verify this by conditioning on W , exploiting the fact that

$$E(U_{BL}|W) = 0,$$

and then taking the expectation of $\hat{\beta}_X$ over the sampling distribution of W . As in the case of Berkson error, the effect of U_{BL} is to make W less variable than the true exposure X , effectively adding to the variance of the noise in the outcome model and resulting in increased variability of $\hat{\beta}_X$.

It is tempting to carry the Berkson analogy further and argue that we can derive valid standard errors by accounting for the correlation in the new noise term

$$\varepsilon' = U_{BL}\beta_X + \varepsilon, \tag{3.2}$$

using either generalized least squares or the sandwich estimator, as would be appropriate for Berkson

error with a non-diagonal covariance (Gryparis et al. 2009; Szpiro et al. 2008; Carroll et al. 2006, page 90). This reasoning is not completely correct for our scenario, however, because it is based on treating W as fixed, which is equivalent to treating the monitoring data X^* as fixed. While one could plausibly posit such a data-generating mechanism, it is not suitable when we also want to consider the variability that results from estimating the exposure model parameters. As an alternative, in Section 4 we describe a bootstrap-based method that is consistent with our assumed data-generating mechanism and that includes the pure Berkson-like error situation as a special case.

3.2 Classical-Like Component of the Error

A primary feature associated with classical measurement error is that it increases the variability of W relative to X , introducing variation that is not correlated with the outcome Y (Carroll et al. 2006, page 28). It is easy to see that U_{CL} is analogous since it comprises the error from estimating the exposure model parameters, which introduces variability that is not informative for Y . Strictly speaking, U_{CL} is not independent of Y since $\hat{\alpha}$ and $\hat{\theta}_\eta$ are derived from X^* which is correlated with X . It is also not independent across locations. Therefore, we emphasize that U_{CL} is similar to classical measurement error but also distinct in important ways, so we cannot simply apply standard measurement error correction techniques like regression calibration.

Our simulation results in Section 5 suggest that the dominant effect of U_{CL} is to increase the sampling variability of $\hat{\beta}_X$, so we focus on methods to derive corrected standard errors which are sufficient to obtain valid confidence intervals for inference. Bias is typically also a concern with classical measurement error, and in principle U_{CL} has the potential to introduce bias in estimating $\hat{\beta}_X$. The bias in our examples is relatively small, but it has some interesting features and for completeness we discuss it further in Ap-

pendix A. One interesting finding is that the bias from this form of classical-like measurement error can be away from the null, rather than toward the null as in the case of standard classical measurement error.

4. PARAMETER BOOTSTRAP

4.1 *Parametric Bootstrap*

A natural, but computationally intensive, approach to estimating standard errors that account for measurement error in our scenario is the parametric bootstrap (Davison and Hinkley 1997; Madsen et al. 2008). The parameter estimate of interest $\hat{\beta}_X$ is calculated as in Section 2, and we wish to derive a standard error that approximates the sampling distribution of $\hat{\beta}_X$ under the assumed true data-generating mechanism. We do this by simulating bootstrap samples under our best estimate of the data-generating mechanism and calculating their empirical standard deviation.

Given a set of observations Y and X^* , the parametric bootstrap procedure is as follows. Let M be the number of bootstrap samples.

1. Estimate the exposure model parameters $\hat{\alpha}$ and $\hat{\theta}_\eta$ by nonlinear optimization based on the model in equation (2.2).
2. Derive W from equation (2.4) and use it in place of X in equation (2.1) to estimate the outcome model parameters $\hat{\beta}_0, \hat{\beta}_X, \hat{\beta}_Z$, and $\hat{\theta}_\varepsilon$.
3. Repeat the steps below for each $j = 1, \dots, M$
 - (a) Simulate a new set of observations Y_j and X_j^* based on the models in equations (2.1) and (2.2), using $\hat{\alpha}, \hat{\theta}_\eta, \hat{\beta}_0, \hat{\beta}_X, \hat{\beta}_Z$, and $\hat{\theta}_\varepsilon$ in place of the unknown true parameters.
 - (b) Estimate new exposure model parameters $\hat{\alpha}_j$ and $\hat{\theta}_{\eta,j}$ by nonlinear optimization based on the

model in equation (2.2), using X_j^* in place of X .

(c) Plug $\hat{\alpha}_j, \hat{\theta}_{\eta,j}$, and X_j^* into equation (2.4) to derive W_j .

(d) Calculate $\hat{\beta}_{X,j}$ by OLS in equation (2.1), using W_j and Y_j in place of X and Y .

4. Calculate the parametric bootstrap standard error

$$\hat{\sigma}_{\beta_X} = \sqrt{\frac{1}{M-1} \sum_{j=1}^M \left(\hat{\beta}_{X,j} - \frac{1}{M} \sum_{j=1}^M \hat{\beta}_{X,j} \right)^2}.$$

Note that in step (3a) we simulate X_j in order to obtain Y_j , but we do not use X_j in the remainder of the procedure. In step (3d) we treat β_0 and β_Z as unknown rather than using the estimated values $\hat{\beta}_0$ and $\hat{\beta}_Z$. See Section 4.4 for some additional details about the implementation in steps (1), (2) and (3b).

Kriging is not a completely automated procedure since there are several standard parametric covariance models from which to choose, and if a particular model does not fit the data well the typical approach is to fit a different parametric covariance model (or combination of models) in order to obtain the best predictions (Laslett 1994; Zimmerman et al. 1999; Goovaerts 1997). It is not practical to follow this procedure in the parametric bootstrap, but it is also not desirable to include realizations in step (3a) for which the kriging model fits very poorly. As a pragmatic way to deal with this in our examples, we exclude any realization that leads to an estimate of the nugget that is very near zero (< 0.05) and any realization where the likelihood is extremely flat at the ML parameter estimates (any diagonal element of the inverse Hessian > 9 , with the Hessian based on ML estimation of $\log \theta_{\eta}$). If a realization is excluded from the parametric bootstrap because of poor model fit, we simulate an additional realization in order to ensure that there are still M samples to calculate the bootstrap standard error.

Our simulation results in Section 5 suggest that the parametric bootstrap gives valid standard errors, but it is extremely computationally intensive since it requires solving a nonlinear optimization problem to

estimate the exposure model parameters for each bootstrap sample.

4.2 Parameter Bootstrap

The idea of the parameter bootstrap is to decrease the computational burden by eliminating the nonlinear optimization that is repeated M times in step (3b) above. This is feasible because we can typically obtain an estimate of the sampling distribution for $\hat{\alpha}$ and $\hat{\theta}_\eta$ in step (1) without much additional computation (assuming the parameter estimation is done using a standard procedure such as maximum likelihood or Markov chain Monte Carlo). For completeness, we list in full below the steps of the parameter bootstrap. The procedure differs from the parametric bootstrap only in the addition of step (1a) and the modification of step (3b). The name “parameter bootstrap” refers to the separate bootstrap of exposure model parameter estimates in step (3b).

1. Estimate the exposure model parameters $\hat{\alpha}$ and $\hat{\theta}_\eta$ by nonlinear optimization based on the model in equation (2.2).
 - (a) Estimate a density function $\hat{p}(\cdot, \cdot)$ corresponding to the sampling distribution of $\hat{\alpha}$ and $\hat{\theta}_\eta$.
2. Derive W from equation (2.4) and use it in place of X in equation (2.1) to estimate the outcome model parameters $\hat{\beta}_0, \hat{\beta}_X, \hat{\beta}_Z,$ and $\hat{\theta}_\varepsilon$.
3. Repeat the steps below for each $j = 1, \dots, M$
 - (a) Simulate a new set of observations Y_j and X_j^* based on the models in equations (2.1) and (2.2), using $\hat{\alpha}, \hat{\theta}_\eta, \hat{\beta}_0, \hat{\beta}_X, \hat{\beta}_Z,$ and $\hat{\theta}_\varepsilon$ in place of the unknown true parameters.
 - (b) Sample $\hat{\alpha}_j$ and $\hat{\theta}_{\eta,j}$ from the probability distribution defined by $\hat{p}(\cdot, \cdot)$.
 - (c) Plug $\hat{\alpha}_j, \hat{\theta}_{\eta,j},$ and X_j^* into equation (2.4) to derive W_j .

(d) Calculate $\hat{\beta}_{X,j}$ by OLS in equation (2.1), using W_j and Y_j in place of X and Y .

4. Calculate the parameter bootstrap standard error using the same formula as the parametric bootstrap.

In the supplementary material (see below) we describe the assumptions that underlie validity of the parameter bootstrap as an approximation to the parametric bootstrap, and we give evidence from our simulation examples to support these assumptions.

4.3 *Partial Parametric Bootstrap*

If we are willing to neglect the classical-like component of the error U_{CL} , another alternative procedure is to modify the parameter bootstrap by using $\hat{\alpha}$ and $\hat{\theta}_\eta$ in each bootstrap sample instead of drawing new values from the estimated sampling distribution as in step (3b). We call this procedure the partial parametric bootstrap.

4.4 *Implementation Details*

We describe here our approach to implementing the individual estimations required for the parametric and parameter bootstraps. Steps (1) and (3b) in the parametric bootstrap require estimating the exposure model parameters $\hat{\alpha}$ and $\hat{\theta}_\eta$ based on the model in equation (2.2). We estimate the maximum likelihood parameter values by using the constrained L-BFGS-B algorithm implemented in the `optim()` function in R (Byrd et al. 1995; R Development Core Team 2008), first log-transforming the variance parameters to make the optimization easier. In the parameter bootstrap we also need an estimate of the sampling distribution for $\hat{\alpha}$ and $\hat{\theta}_\eta$. We use a Gaussian approximation centered at the maximum likelihood value with covariance based on the estimated Hessian produced by the `optim()` function.

Sometimes the L-BFGS-B algorithm finds a solution on the boundary that is not the actual maximum,

so to guard against this we repeat the numerical optimization five times and pick the parameter values corresponding to the largest value of the likelihood. In addition, for each of the five candidate maxima, if the estimated range or nugget parameters are very close to zero or if they are extremely large we perturb the initial values and repeat the optimization. This somewhat ad-hoc procedure provides a practical approach to ensuring that we find the actual maximum without more calls to `optim()` than is necessary, as this step is very computationally intensive.

The other nonlinear optimization is contained in step (2), where we estimate the parameters θ_ε in the outcome model (2.1). If we knew X without error, this could be accomplished in a relatively straightforward manner by fitting model (2.1) using OLS and then estimating the variance parameters θ_ε from the residuals. The complicating factor is that the Berkson-like component of the measurement error introduces an additional covariance contribution into the residual distribution, as in the definition of ε' in equation (3.2). If we condition on X^* , the covariance for the residuals is

$$\Sigma_{\varepsilon'} = \Sigma_{U_{BL}}\beta_X + \Sigma_\varepsilon(\theta_\varepsilon) \quad (4.1)$$

where $\Sigma_{U_{BL}}$ can be computed based on equation (3.1) if we treat the estimated exposure model parameters as known.

In order to estimate θ_ε we fit the model in equation (4.1) to the residuals from the OLS fit to equation (2.1), optimizing over the nuisance parameter β_X . Maximum likelihood optimization turns out to be unstable, so instead we find the least-squares fit to the empirical binned variogram. We again use the L-BFGS-B algorithm implemented in the `optim()` function in R (Byrd et al. 1995; R Development Core Team 2008), repeating the numerical optimization 20 times with random initial conditions and picking the parameters estimates based on the best fit from these 20 iterations.

5. SIMULATIONS

An example application is the problem considered by (Madsen et al. 2008; Herlihy et al. 1998) of estimating the association between the log of chloride levels in streams

$$Y = \log_cl$$

and the logit of the fraction of forestation in the local watershed

$$X = \text{logit_for.}$$

The following simulations are based on data collected by the Environmental Protection Agency in the Mid-Atlantic Highlands region of the eastern United States during the years 1993-1996 (U.S EPA Environmental Monitoring and Assessment Program 1999). In a 400×500 box we randomly select locations for $N^* = 200$ exposure monitors and $N = 300$ or 2000 outcome measurements the universal kriging exposure model is

$$\begin{pmatrix} X \\ X^* \end{pmatrix} = \begin{pmatrix} S \\ S^* \end{pmatrix} \alpha + \begin{pmatrix} \eta \\ \eta^* \end{pmatrix} \quad (5.1)$$

where each row of S and S^* is of the form $(1, x, y)$, x and y are the coordinates, and the regression coefficients are $\alpha = (-25.95, -0.0035, 0.00084)^t$. The residual term $(\eta^t, \eta^{*t})^t$ has an exponential variogram structure with parameters

$$\theta_\eta = (\phi_\eta, \psi_\eta, \tau_\eta),$$

where $\phi_\eta = 24.13$ is the range, $\psi_\eta = 3.76$ is the partial sill, and $\tau_\eta = 1.34$ is the nugget (Cressie 1993).

We also consider a modified version of this model with a longer range $\phi_\eta = 90.0$. Example realizations of the exposure surface for the two different ranges are shown in Figure 1. As in the parametric bootstrap, we exclude realizations that exhibit very poor fit with the assumed universal kriging model since in practice

an analyst would attempt to identify a different parametric correlation structure that is more consistent with the observed data.

The linear regression model for the outcome conditional on X is

$$Y = \beta_0 + X\beta_X + \varepsilon$$

where $\beta_0 = 5.06$ and $\beta_1 = -0.322$. We consider the case of uncorrelated residuals ε with variance

$$\theta_\varepsilon = (\sigma_\varepsilon^2),$$

where $\sigma_\varepsilon^2 = 0.76$, and we also consider the case of correlated residuals ε following an exponential variogram structure with parameters

$$\theta_\varepsilon = (\phi_\varepsilon, \psi_\varepsilon, \tau_\varepsilon),$$

where $\phi_\varepsilon = 80.39$ is the range, $\psi_\varepsilon = 0.26$ is the partial sill, and $\tau_\varepsilon = 0.50$ is the nugget.

In Table 1, we summarize the results for 2000 Monte Carlo runs based on a universal kriging exposure model with range $\phi_\eta = 24.13$. Due to the computational intensity of the full parametric bootstrap, we restricted to 100 Monte Carlo runs for these results. The bias in $\hat{\beta}_X$ is a very small component of the total estimation error in all scenarios. In the model with uncorrelated residuals, naïve standard errors that do not correct for measurement error are too small compared to the observed sampling distribution of $\hat{\beta}_X$, resulting in significantly less than nominal coverage for 95% confidence intervals. In the correlated outcome model, where the uncorrected standard errors account for correlated residuals using a sandwich form, the coverage is closer to nominal. This is presumably because much of the Berkson-like error appears as additional correlated variability in the residuals and is accounted for by the sandwich form.

Both the parameter bootstrap and the parametric bootstrap give corrected standard errors that result in some over-coverage, with coverage probabilities of 97% – 98%. The expected value of the standard errors

from parameter simulation tends to be larger than the sampling standard deviation of $\hat{\beta}_X$, but the mode of the standard errors tends to align well with the observed standard deviation. This suggests that there are some large outliers, which can be confirmed by examining scatterplots of the estimated standard errors (results not shown). If we replace the estimated sampling distribution of $\hat{\alpha}$ and $\hat{\theta}_\eta$ used in the parameter bootstrap with the true sampling distribution, then the outliers go away and the coverage probabilities are much closer to 95% (results not shown).

In addition to results for the parametric bootstrap and the parameter bootstrap, we show standard errors calculated with the partial parametric bootstrap. The partial parametric bootstrap accounts only for the Berkson-like component of the measurement error, neglecting the classical-like component. In scenarios with only 300 measured outcomes ($N = 300$), the partial parametric bootstrap gives similar results to the parameter bootstrap and results in nearly nominal coverage probabilities. This suggests that the Berkson-like component of measurement error dominates in this situation. For $N = 2000$, however, the partial parametric bootstrap standard errors are significantly smaller than those from the parameter bootstrap and result in less than nominal coverage probabilities, indicating that the classical-like component of the error is more important relative to the Berkson-like component. This is consistent with our expectations, since the magnitude of the classical-like component does not change for large N , but the first-order contribution of the Berkson-like component diminishes since its effect is analogous to increasing the random noise in the outcome model.

We report an analogous set of results in Table 2 for scenarios with a longer correlation range ($\phi_\eta = 90.0$) in the exposure model. The same general patterns are observed as for $\phi_\eta = 24.13$, with the one difference being that the partial parametric bootstrap provides adequate measurement error correction even for $N = 2000$. This may be because introducing additional spatial correlation in the exposure

surface makes it easier to estimate some of the exposure model parameters, decreasing the contribution of the classical-like component of measurement error.

Based on our simulation results, we can directly evaluate the assumptions that are required for validity of the parameter bootstrap as an approximation to the parametric bootstrap. These validation results are given in the supplementary material (see below).

6. EXAMPLE

We conclude by applying our methodology to the EPA data that formed the basis for our simulation study in the previous section (Madsen et al. 2008; Herlihy et al. 1998; U.S EPA Environmental Monitoring and Assessment Program 1999). We use the universal kriging exposure model described above and allow for correlation in the outcome residuals.

The outcome (log of chloride level) and the exposure (logit of percent forestation) are both available at a total 422 of these locations (we have restricted to streams with percent forestation strictly between 0% and 100% in order to allow for the logit transformation). Where multiple measurements are available from different times, we use the earliest time. In the remainder of the example, we assume that the exposure is only available at $N^* = 200$ randomly selected locations and that the outcome is available at the remaining $N = 222$ locations. A map of the respective locations is shown in Figure 2. Distances are calculated using a flat-earth approximation, with one degree of latitude equal to 111.3 km and one degree of longitude equal to 85.9 km.

In Figure 3 we show the empirical binned variogram and the fitted variogram model for the exposure monitoring data, after detrending by a linear model in latitude and longitude. This figure is consistent with a good fit between the data and the assumed exposure model. The estimated values for the exposure model

parameters based on X^* are $\hat{\alpha} = (-42.56, -0.0049, 0.0026)$, range $\hat{\phi}_\eta = 11.00$, partial sill $\hat{\psi}_\eta = 2.35$, and nugget $\hat{\tau}_\eta = 2.08$. We show scatterplots and the linear regressions of the outcome against the true and predicted exposures in Figure 4.

The results of estimating $\hat{\beta}_X$ and its standard error based on W and X are shown in Table 3. The estimated effect size based on the predicted exposure W is -0.460 , while the value obtained by using the true exposure X is -0.322 . The standard error estimate based on using W without correcting for measurement error is 0.252 . The partial parametric bootstrap gives a similar value, but the standard error estimates from the parameter bootstrap and the parametric bootstrap are both larger, taking values of 0.305 and 0.337 , respectively. This suggests that the parameter bootstrap is a reasonable approximation to the more computationally intensive parametric bootstrap for accounting for measurement error in estimating the standard error for $\hat{\beta}_X$ in this dataset.

7. DISCUSSION

We have characterized the measurement error that results from using smoothing to predict exposures in an environmental statistics association study when the exposure and outcome data are misaligned in space. The resulting measurement error has a Berkson-like component that results from information lost in smoothing and a classical-like component that is related to uncertainty associated with estimating the smoothing parameters.

The measurement error structure we have identified is complex because it is a mixture of two types of error, neither one of which fits exactly into the traditional categories of Berkson or classical. Therefore, standard measurement error correction methods are not appropriate. If we are willing to assume that the exposure and outcome models are correctly specified, we can use a parametric bootstrap to estimate

corrected standard errors. This approach requires that we be precise about the assumed data-generating mechanism since the idea is to draw multiple samples from an approximation to the data-generating mechanism, with the model parameters estimated based on observed data. We have chosen to define a data-generating mechanism that is consistent with the geostatistical kriging model we use for smoothing. Although it is well known that geostatistical methods are useful for interpolating the kinds of physical processes that arise as exposure variables in environmental statistics, it is not clear what real-world phenomenon the correlated variance in this model represents. A promising direction for future research is to investigate the scientific validity of this and other assumed data-generating mechanisms, and to characterize the implications for measurement error correction.

Our decomposition of the measurement error into Berkson-like and classical-like components enables us to propose the parameter bootstrap as a less computationally intensive approximation to the parametric bootstrap. The parameter bootstrap is valid under assumptions that appear to hold in our simulation scenarios, but care should be taken in applying this method in other settings where the assumptions have not been verified. In particular, it is important that the estimated sampling distribution for the exposure model parameter estimates be a valid approximation to the true sampling distribution. In general, this should be true for sufficiently rich exposure monitoring data, but in many applications the available monitoring data are limited. As a computational tool, the parameter bootstrap is only necessary when estimating the exposure model parameters is computationally intensive, and this is only the case if there is a relatively large amount of exposure data. Therefore, the decision of whether to use the parametric bootstrap or the parameter bootstrap should be informed by the amount of exposure data available, considering the implications for the computational burden of the parametric bootstrap and the validity of the parameter bootstrap. One practical compromise is to use the parameter bootstrap with a larger number of bootstrap samples as the

primary correction and to validate it by comparing with results of the parametric bootstrap based on a limited number of samples.

8. ACKNOWLEDGMENTS

Research described in this manuscript has been funded wholly or in part by the United States Environmental Protection Agency through R831697 to the University of Washington. It has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

REFERENCES

- S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall, CRC, 2004.
- M. Brauer, G. Hoek, P. van Vliet, K. Meliefste, P. Fischer, U. Gehring, J. Heinrich, J. Cyrys, T. Bellander, M. Lewne, and B. Brunekreef. Estimating long-term average particulate air pollution concentrations: Application of traffic indicators and geographic information systems. *Epidemiology*, 14(2):228–239, 2003.
- R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16:1190–1208, 1995.
- R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*. Chapman and Hall/CRC, 2006.
- N. A. C. Cressie. *Statistics for Spatial data*. John Wiley and Sons, New York, 1993.
- A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Applications*. University of Cambridge Press, 1997.
- D. W. Dockery, C. A. Pope, X Xu, J. D. Spangler, J. H. Ware, M. E. Fay, B. G. Ferris, and F. E. Speizer. An association between air pollution and mortality in six cities. *New England Journal of Medicine*, 329(24):1753–1759, 1993.
- P. Goovaerts. *Geostatistics for Natural Resource Evaluation*. Oxford University Press, New York, 1997.

- A. Gryparis, B. A. Coull, J. Schwartz, and H. H. Suh. Semiparametric latent variable regression models for spatio-temporal modeling of mobile source particles in the greater Boston area. *Journal of the Royal Statistical Society, Series C*, 56(2):183–209, 2007.
- A. Gryparis, C. J. Paciorek, A. Zeka, J. Schwartz, and B. A. Coull. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics*, 10(2):258–274, 2009.
- A. T. Herlihy, J. T. Stoddard, and C. B. Johnson. The relationship between stream chemistry and watershed land cover in the Mid Atlantic region, U.S. *Water, Air, and Soil Pollution*, 105:377–386, 1998.
- M. Jerrett, A. Arain, P. Kanaroglou, B. Beckerman, D. Potoglou, T. Sahsuvaroglu, J. Morrison, and C. Giovis. A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology*, 15:185–204, 2005a.
- M. Jerrett, R. T. Burnett, R. Ma, C. A. Pope, D. Krewski, K. B. Newbold, G. Thurston, Y. Shi, N. Finkelstein, E. E. Calle, and M. J. Thun. Spatial analysis of air pollution mortality in Los Angeles. *Epidemiology*, 16(6):727–736, 2005b.
- S. Y. Kim, L. Sheppard, and H. Kim. Health effects of long-term air pollution: influence of exposure prediction methods. *Epidemiology*, 20(3):442–450, 2009.
- N. Kunzli, M. Jerrett, W. J. Mack, B. Beckerman, L. LaBree, F. Gilliland, D. Thomas, J. Peters, and H. N. Hodis. Ambient air pollution and atherosclerosis in Los Angeles. *Environmental Health Perspectives*, 113(2):201–206, 2005.
- G. M. Laslett. Kriging and splines: An empirical comparison of their predictive performance in some applications. *Journal of the American Statistical Association*, 89(426):391–400, 1994.
- K-Y Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.
- J. Lindström and F. Lindgren. A Gaussian Markov random field model for total yearly precipitation over the African Sahel. *Preprints in Mathematical Statistics*, (LUTFMS–5074–2008), 2008.
- S. O. Los, G. P. Weedon, P. R. J. North, J. D. Kaduk, C. M. Taylor, and P. M. Cox4. An observation-based estimate

- of the strength of rainfall-vegetation interactions in the Sahel. *Geophysical Research Letters*, 33:L16402, 2006.
- L. Madsen, D. Ruppert, and N. S. Altman. Regression with spatially misaligned data. *Environmetrics*, 19:453–467, 2008.
- K. A. Miller, D. S. Sicovick, L. Sheppard, K. Shepherd, J. H. Sullivan, G. L. Anderson, and J. D. Kaufman. Long-term exposure to air pollution and incidence of cardiovascular events in women. *New England Journal of Medicine*, 356(5):447–458, 2007.
- C. A. Pope, R. T. Burnett, M. J. Thun, E. E. Calle, K. Ito D. Krewski, and G. D. Thurston. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association*, 9(287):1132–1141, 2002.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- J. M. Samet, F. Dominici, F. Curriero, I. Coursac, and S. L. Zeger. Particulate air pollution and mortality: Findings from 20 u.s. cities. *New England Journal of Medicine*, 343(24):1742–1749, 2000.
- A. Szpiro, L. Sheppard, and T. Lumley. Accounting for errors from predicting exposures in environmental epidemiology and environmental statistics. *UW Biostatistics Working Paper Series*, (Working Paper 330), 2008.
- A. Szpiro, P. D. Sampson, L. Sheppard, T. Lumley, S. D. Adar, and J. D. Kaufman. Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics*, In Press (DOI: 10.1002/env.1014), 2009.
- U.S EPA Environmental Monitoring and Assessment Program. Mid-Atlantic Streams 1993-96 Data Sets, 1999.
<http://www.epa.gov/emap/html/data/surfwatr/data/mastrems/9396/wchem/chmval.txt>,
<http://www.epa.gov/emap/html/data/surfwatr/data/mastrems/9396/location/watchr.txt>.
- D. Zimmerman, C. Pavlik, A. Ruggles, and M. P. Armstrong. An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Mathematical Geology*, 31(4):375–390, 1999.

APPENDIX

A. BIAS FROM CLASSICAL-LIKE MEASUREMENT ERROR

Consider a simplified version of the model from Section 2. In particular, assume

$$Y = X\beta_X + \varepsilon$$

and

$$X = S\alpha + \eta,$$

where Y , X , and S are N -vectors and ε and ν are independent Normally distributed N -vectors with zero means and variances σ_ε^2 and σ_ν^2 . We regard β_X and α as fixed but unknown, S as fixed and known, and Y and X as drawn at random. Only Y is observed so we need to estimate X in order to estimate β_X .

Notice that we have assumed a scalar exposure without an intercept in the outcome model and uncorrelated residuals for both the exposure and the outcome. These assumptions simplify the calculations that follow and allow us to clearly illustrate the origin of any bias that may result from estimating parameters in the exposure model.

Since we do not directly observe X , we want to make inference about β_X based on the observations Y and an approximation to X denoted by W . Specifically, we will use

$$W = S\hat{\alpha}$$

where $\hat{\alpha}$ is an estimate of α . Assume that in the sampling distribution

$$\hat{\alpha} = \alpha + \delta,$$

where δ is Normally distributed with mean zero and variance σ_δ^2 . In practice, $\hat{\alpha}$ could be an estimate of

α based on monitoring data X^* at different locations which satisfy the same linear model as X and for which the corresponding geographic covariates S^* are also known.

If we knew X without error, the OLS fit for β_X would be

$$\hat{\beta}_{X,true} = (X^t X)^{-1} X^t Y.$$

Since we do not know X directly, we approximate X by plugging in W as defined above and find the corresponding OLS fit for β_X

$$\begin{aligned} \hat{\beta}_X &= (W^t W)^{-1} W^t Y \\ &= (\alpha + \delta)^{-1} (S^t S)^{-1} S^t Y. \end{aligned}$$

We are interested in the direction of the bias

$$\pi = E \left[\hat{\beta}_X - \beta_X \right].$$

To analyze this quantity, it is useful first to recall that if we knew α without error, then we could define

$$W_{berkson} = S\alpha$$

and the corresponding estimator

$$\begin{aligned} \hat{\beta}_{X,berkson} &= (W_{berkson}^t W_{berkson})^{-1} W_{berkson}^t Y \\ &= \alpha^{-1} (S^t S)^{-1} S^t Y. \end{aligned}$$

Since the measurement error in $W_{berkson}$ is purely Berkson, this estimator would be unbiased for β_X .

Therefore, we can re-write the bias for $\hat{\beta}_X$

$$\begin{aligned}\pi &= E \left[\hat{\beta}_X - \hat{\beta}_{X,berkson} \right] \\ &= E \left[((\alpha + \delta)^{-1} - \alpha^{-1}) (S^t S)^{-1} S^t Y \right].\end{aligned}$$

Now, since the mean of δ is zero, Taylor series expansion of $(\alpha + \delta)^{-1}$ around α^{-1} gives

$$\begin{aligned}\pi &\approx E \left[(-\delta\alpha^{-2} + \delta^2\alpha^{-3}) (S^t S)^{-1} S^t Y \right] \\ &= \sigma_\delta^2 \alpha^{-3} E \left[(S^t S)^{-1} S^t Y \right] \\ &= \sigma_\delta^2 \alpha^{-2} E \left[\hat{\beta}_{X,berkson} \right] \\ &= \sigma_\delta^2 \alpha^{-2} \beta_X.\end{aligned}$$

Thus, we have shown that the direction of bias is away from the null, at least in the limit for small δ . The magnitude of the bias is proportional to the squared coefficient of variation for α . In our simulations the bias is a relatively small component of the overall error, but it tends to be away from zero as predicted by this analysis. This calculation also clearly exposes the source of bias in $\hat{\beta}_X$, since even if W is an unbiased estimate of X , nonlinearity in the OLS solution operator results in bias for the parameter estimate.

	$N = 300$					$N = 2000$				
	Bias	SD	E(SE)	Mode(SE)	Cov	Bias	SD	E(SE)	Mode(SE)	Cov
Independent Outcomes										
No correction	-0.009	0.068	0.053	0.049	89%	-0.006	0.049	0.021	0.019	65%
Partial parametric bootstrap	-0.009	0.068	0.065	0.058	95%	-0.006	0.049	0.037	0.031	88%
Parameter bootstrap	-0.009	0.068	0.083	0.073	98%	-0.009	0.049	0.058	0.046	96%
Parametric bootstrap	-0.010	0.069	0.081	0.073	98%	0.000	0.045	0.053	0.042	93%
Correlated Outcomes										
No correction	-0.009	0.099	0.099	0.082	94%	-0.004	0.090	0.087	0.067	92%
Partial parametric bootstrap	-0.009	0.099	0.102	0.079	94%	-0.004	0.090	0.090	0.075	92%
Parameter bootstrap	-0.009	0.099	0.124	0.106	97%	-0.004	0.090	0.112	0.092	96%
Parametric bootstrap	-0.003	0.101	0.122	0.102	99%	-0.005	0.081	0.115	0.120	98%

Table 1. Simulation results for universal kriging exposure surface with range $\phi_\eta = 24.13$. The columns give the bias and standard deviation of the estimates, the mean and mode of the estimated standard errors, and the coverage for 95% Wald confidence intervals. Results are based on 2000 Monte Carlo simulations, except for the parametric bootstrap results which are based on 100 Monte Carlo simulations.

	$N = 300$					$N = 2000$				
	Bias	SD	E(SE)	Mode(SE)	Cov	Bias	SD	E(SE)	Mode(SE)	Cov
Independent Outcomes										
No correction	-0.003	0.048	0.040	0.036	90%	-0.003	0.032	0.016	0.014	69%
Partial parametric bootstrap	-0.003	0.048	0.050	0.046	96%	-0.002	0.032	0.029	0.026	93%
Parameter bootstrap	-0.003	0.048	0.055	0.049	98%	-0.003	0.032	0.036	0.030	97%
Parametric bootstrap	-0.005	0.053	0.062	0.054	98%	0.004	0.031	0.035	0.035	94%
Correlated Outcomes										
No correction	-0.002	0.082	0.083	0.068	93%	-0.001	0.077	0.076	0.057	93%
Partial parametric bootstrap	-0.002	0.082	0.085	0.072	92%	-0.001	0.077	0.079	0.066	91%
Parameter bootstrap	-0.002	0.082	0.091	0.076	94%	-0.001	0.077	0.084	0.070	93%
Parametric bootstrap	0.002	0.084	0.100	0.089	96%	0.001	0.073	0.094	0.082	98%

Table 2. Simulation results for universal kriging exposure surface with range $\phi_\eta = 90.0$. The columns give the bias and standard deviation of the estimates, the mean and mode of the estimated standard errors, and the coverage for 95% Wald confidence intervals. Results are based on 2000 Monte Carlo simulations, except for the parametric bootstrap results which are based on 100 Monte Carlo simulations.

	$\hat{\beta}_x$	SE
True exposure	-0.322	0.034
Predicted exposure, no correction	-0.460	0.252
Predicted exposure, partial parametric bootstrap	-0.460	0.263
Predicted exposure, parameter bootstrap	-0.460	0.305
Predicted exposure, parametric bootstrap	-0.460	0.337

Table 3. Results of estimating the relationship between chloride levels in streams (log-transformed) and local forestation levels (logit-transformed percent forestation) based on EPA data from the Mid-Atlantic Highlands region of the eastern United States during the years 1993-1996. The exposure is assumed to be available at $N^* = 200$ randomly selected locations, and the outcome is available at the remaining $N = 222$ locations.

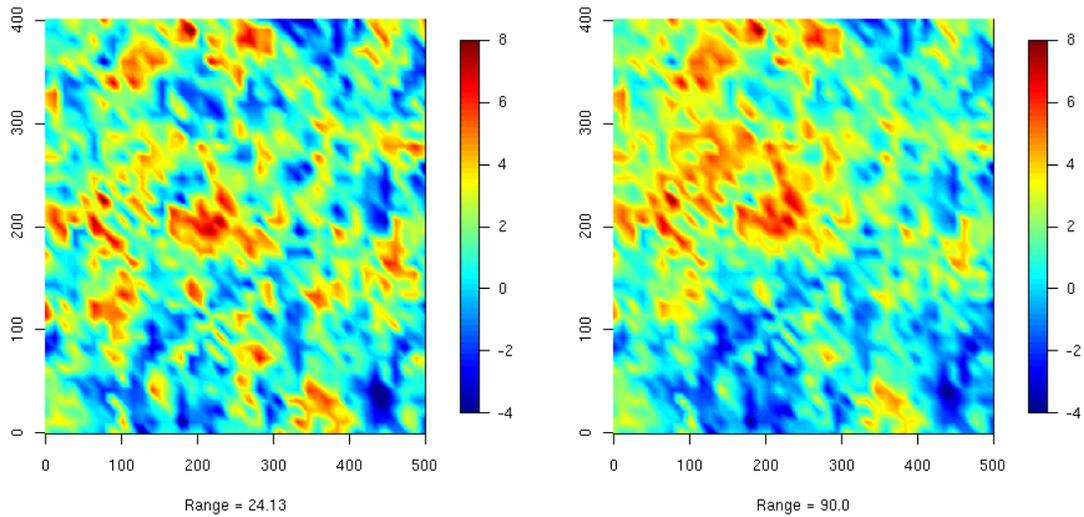


Fig. 1. Examples of simulated spatially correlated exposure surfaces.

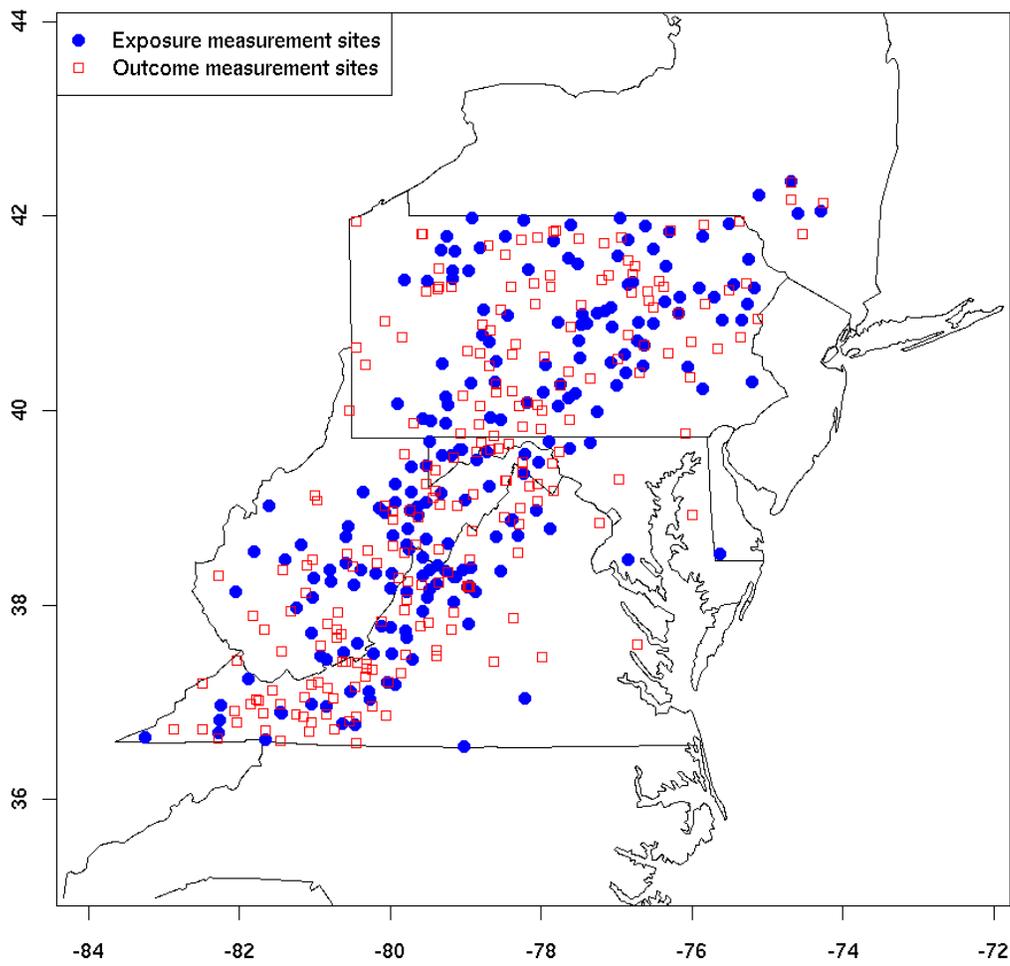


Fig. 2. Partitioning of EPA stream locations for example data analysis ($N^* = 200$ monitors and $N = 222$ outcomes).

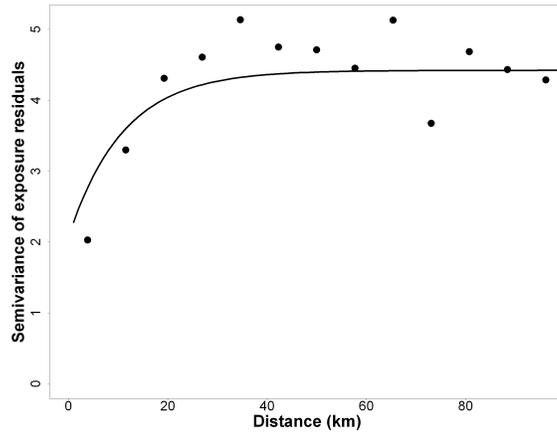


Fig. 3. Empirical binned variogram and fitted variogram based on the logit-transformed percent forestation measurements in the EPA data from the Mid-Atlantic Highlands region of the eastern United States during the years 1993-1996 (200 locations).

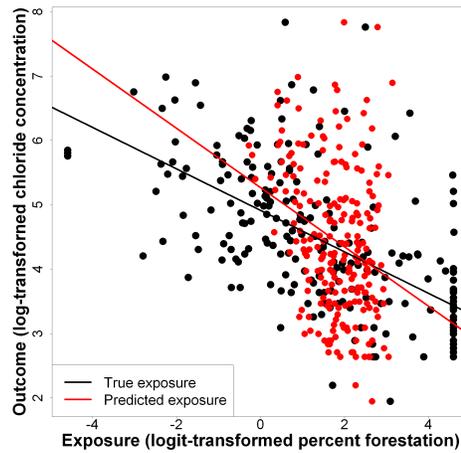


Fig. 4. Scatterplots and the linear regressions of the outcome (log-transformed chloride concentration) against the true and predicted exposures (logit-transformed percent forestation) based on EPA data from the Mid-Atlantic Highlands region of the eastern United States during the years 1993-1996. The black points and line show the true exposure while the red points and line are for the predicted exposures.

ONLINE SUPPLEMENT: THEORETICAL BASIS FOR PARAMETER BOOTSTRAP

The parameter bootstrap is valid insofar as it approximates the true sampling distribution of $\hat{\beta}_X$, which requires three main assumptions. The first assumption, as in the more standard parametric bootstrap, is that the model in equations (2.1) and (2.2) is a valid representation of the true data-generating mechanism. The second assumption is also relatively straightforward to state, namely that the density function $\hat{p}(\cdot, \cdot)$ provides a good approximation to the sampling distribution of $\hat{\alpha}$ and $\hat{\theta}_\eta$. This can generally be expected to hold for a large number of exposure monitors N^* . In our simulation scenarios in Section 5 where we use a Gaussian approximation for $\hat{p}(\cdot, \cdot)$, we find that the assumption holds reasonably well under practical assumptions about how the exposure model will be fit to monitoring data. In Figures S1 and S2, we compare the estimated sampling distributions to the true sampling distributions for each parameter. The density plots for the estimated distributions are a composite of the individual distributions estimated in each Monte Carlo simulation, with each of these distributions centered at the average of the estimated means before constructing the composite. The agreement is reasonably good, although we note that there is significant variation between the estimated sampling distributions in the individual Monte Carlo simulations (results not shown).

The third and final assumption is more subtle to state. In the parameter bootstrap, we have replaced the derivation of $\hat{\alpha}_j$ and $\hat{\theta}_{\eta,j}$ by nonlinear optimization with a draw from the estimated marginal distribution of $\hat{\alpha}$ and $\hat{\theta}_\eta$. Therefore, it seems that the validity of the parameter bootstrap requires assuming in the true data-generating mechanism that $\hat{\alpha}$ and $\hat{\theta}_\eta$ are independent of the data Y and X^* , which is obviously false since $\hat{\alpha}$ and $\hat{\theta}_\eta$ are defined as nonlinear functions of X^* . However, we only need to be able to treat $\hat{\alpha}$ and $\hat{\theta}_\eta$ as if they were independent of the data Y and X^* for the purpose of evaluating the marginal sampling distribution of $\hat{\beta}_X$.

In order to make this statement precise, we introduce some additional notation. We write

$$\begin{aligned}\hat{\beta}_X &= \Psi(Y, W) \\ &= \Psi(Y, \Phi(X^*, \hat{\alpha}, \hat{\theta}_\eta))\end{aligned}$$

where $\Psi(\cdot, \cdot)$ is defined to be the ordinary least squares solution for β_X from equation (2.1) (suppressing additional covariates Z in this notation), and $\Phi(\cdot, \cdot, \cdot)$ is given by equation (2.4). Let $q_1(\cdot, \cdot)$ be the density function corresponding to the distribution of Y and X^* . Then we can define a density function $q(\cdot, \cdot, \cdot, \cdot)$ for the joint sampling distribution of Y , X^* , $\hat{\alpha}$, and $\hat{\theta}_\eta$ by

$$q(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta) = \begin{cases} q_1(Y, X^*) & \varphi(X^*) = \hat{\alpha}, \hat{\theta}_\eta \\ 0 & \text{otherwise} \end{cases}$$

where $\varphi(\cdot)$ is the nonlinear solution operator for estimating $\hat{\alpha}$ and $\hat{\theta}_\eta$ based on equation (2.2). Finally, let $q_2(\cdot, \cdot)$ correspond to the marginal sampling distribution of $\hat{\alpha}$ and $\hat{\theta}_\eta$. Now we can formally state the assumption that underlies the parameter bootstrap.

ASSUMPTION A.1 Sampling from distributions with the following two densities induces a common distribution for $\hat{\beta}_X = \Psi(Y, \Phi(X^*, \hat{\alpha}, \hat{\theta}_\eta))$

$$q(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta)$$

and

$$q'(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta) = q_1(Y, X^*)q_2(\hat{\alpha}, \hat{\theta}_\eta).$$

This assumption is plausible because X^* contributes variability to W in two fundamentally different ways. Its contribution through the parameter estimates $\hat{\alpha}$ and $\hat{\theta}_\eta$ can be viewed as an overall characterization of the structure of the exposure surface, while its direct contribution as the object of smoothing

incorporates local information based on the assumption that exposures at nearby locations are similar. We are able to test Assumption A.1 in our simulation examples by drawing samples directly from the distributions corresponding to $q(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta)$ and $q'(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta)$ and comparing the empirical distributions induced for $\hat{\beta}_X$.

We test the validity of Assumption A.1 in our simulation examples by sampling $\hat{\beta}_X$ under each of the densities $q(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta)$ and $q'(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta)$. We use the empirical distribution of $\hat{\alpha}$ and $\hat{\theta}_\eta$ in order to draw samples from $q'(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta)$. The density functions for $\hat{\beta}_X$ are compared in Figures S3 and S4. While there appears to be slightly more spread in the distributions under $q'(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta)$, the densities for $\hat{\beta}_X$ are very similar, which is consistent with the fact that the parameter bootstrap and the parametric bootstrap agree very well.

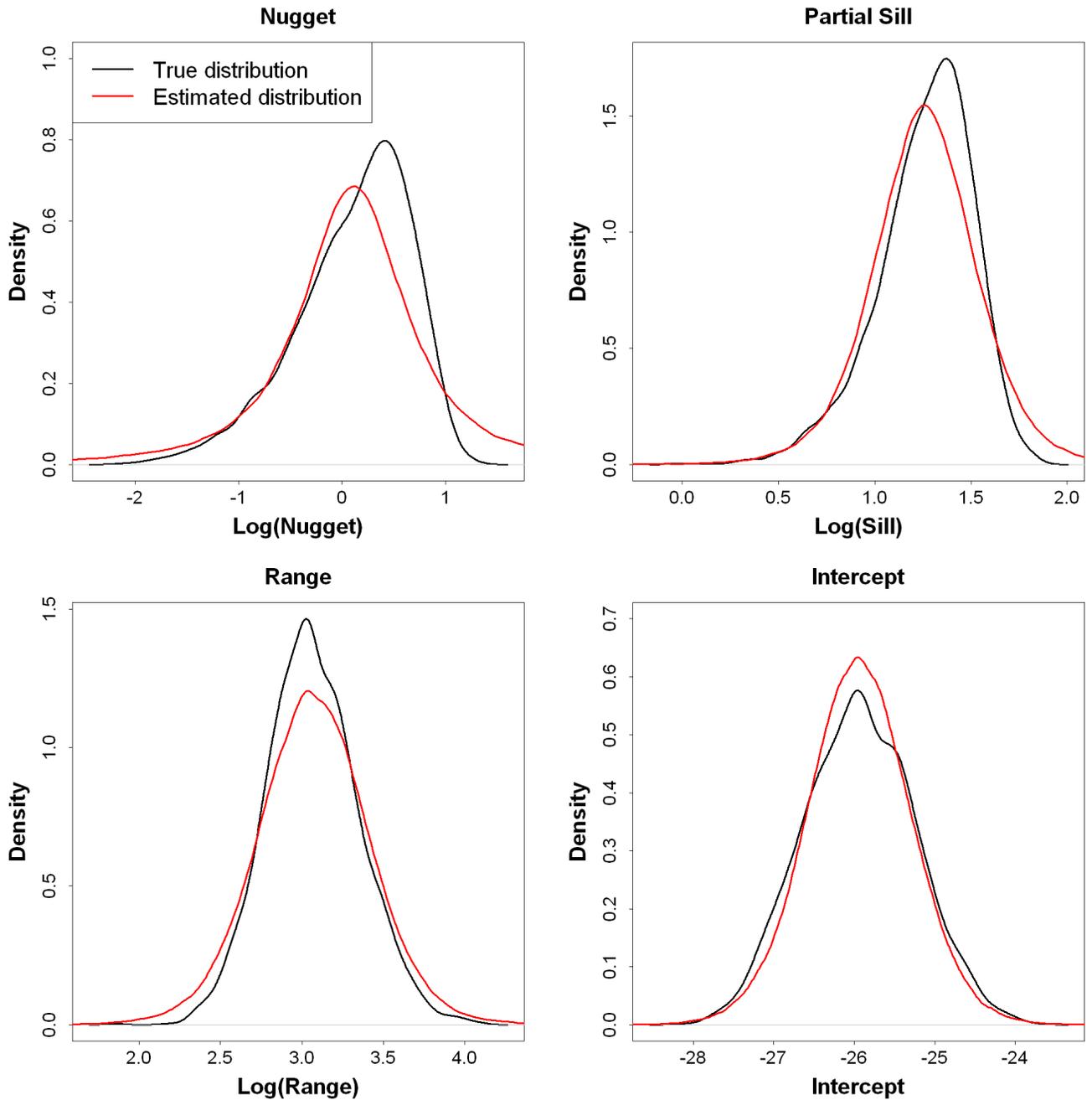


Fig. S1. Estimated and true sampling distributions for exposure model parameters with universal kriging. The density shown for the estimated sampling distribution is a composite of the estimated distributions from 2000 Monte Carlo simulations, with each distribution centered at the average of the estimated means. (Exposure model range $\phi_\eta = 24.13$.)

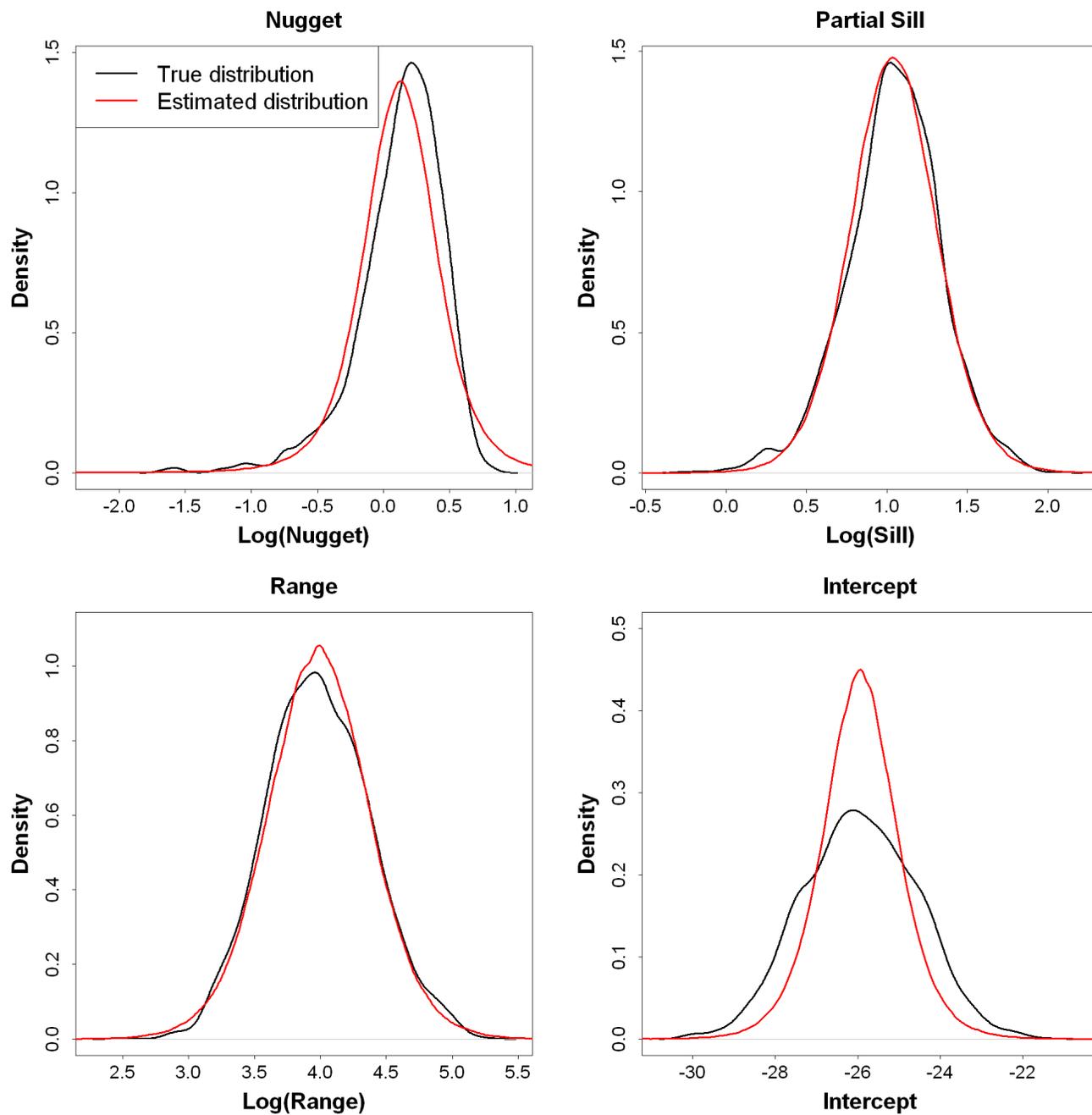


Fig. S2. Estimated and true sampling distributions for exposure model parameters with universal kriging. The density shown for the estimated sampling distribution is a composite of the estimated distributions from 2000 Monte Carlo simulations, with each distribution centered at the average of the estimated means. (Exposure model range $\phi_\eta = 90.0$.)

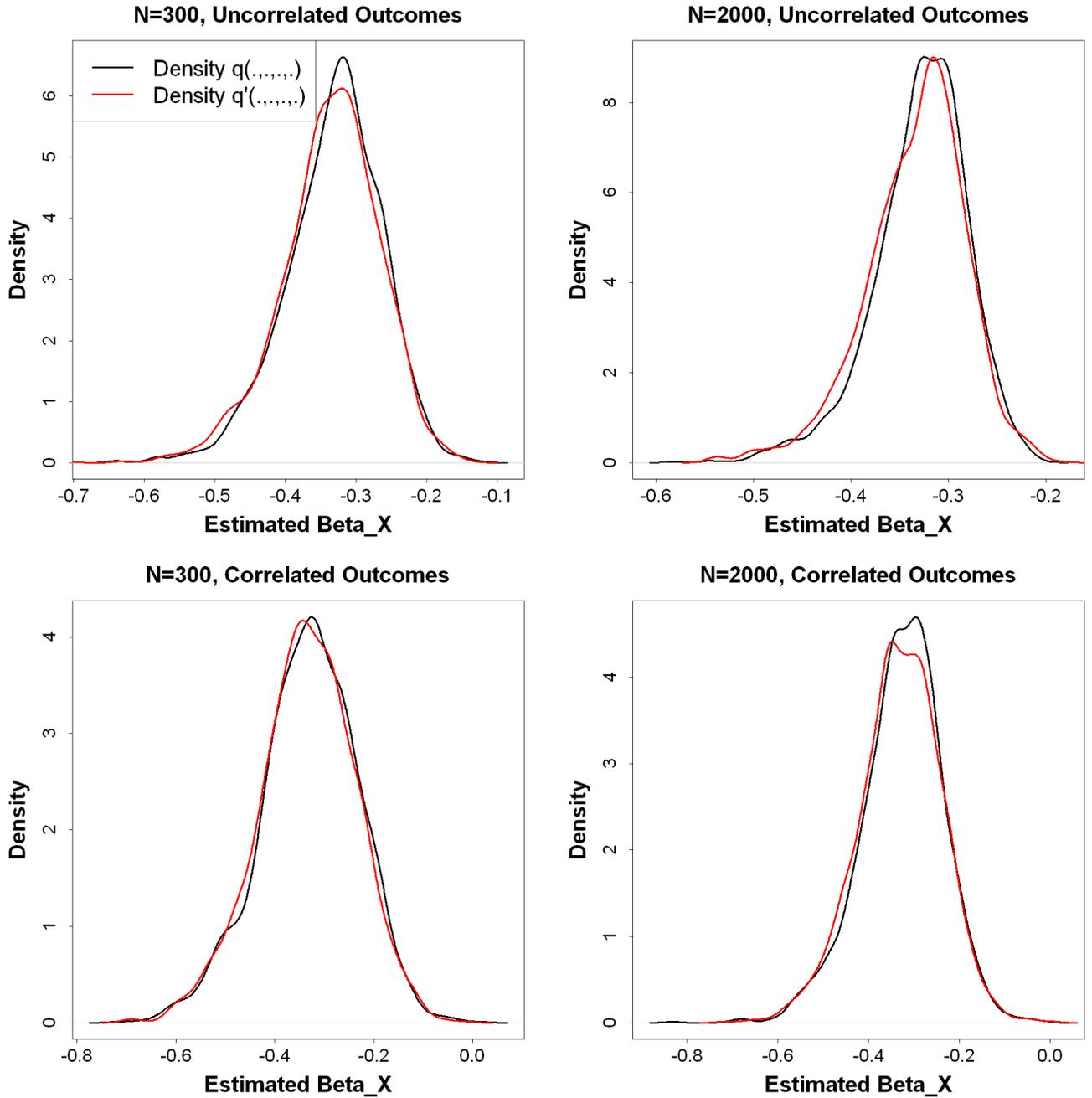


Fig. S3. Comparison of the sampling distributions of $\hat{\beta}_X$ under the true density $q(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta)$ and the approximate density $q'(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta)$ defined in Assumption A.1. (Universal kriging with exposure model range $\phi_\eta = 24.13$.)

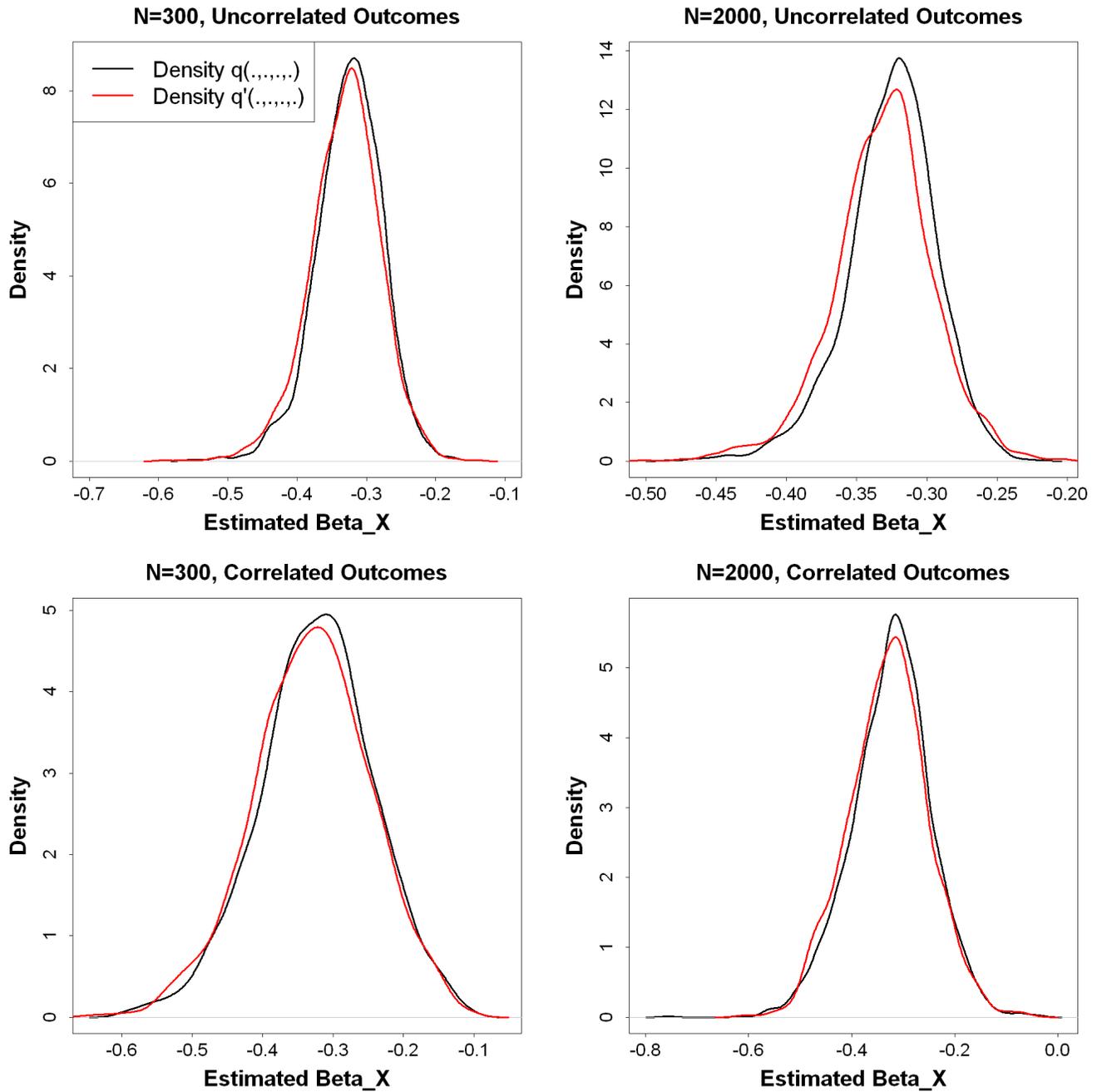


Fig. S4. Comparison of the sampling distributions of $\hat{\beta}_X$ under the true density $q(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta)$ and the approximate density $q'(Y, X^*, \hat{\alpha}, \hat{\theta}_\eta)$ defined in Assumption A.1. (Universal kriging with exposure model range $\phi_\eta = 90.0$.)