Biostatistics (2011), **12**, 4, *pp.* 610–623
doi:10.1093/biostatistics/kxq083
Advance Access publication on January 20, 2011

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 estimate a regression parameter of interest using the predicted exposures. This 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. One correction approach is the parametric bootstrap, which is effective but computationally intensive since it requires solving a nonlinear optimization problem for the exposure model parameters in each bootstrap sample. We propose a less computationally intensive alternative termed the "parameter bootstrap" that only requires solving one nonlinear optimization problem, and we also compare bootstrap methods to other recently proposed methods. We illustrate our methodology in simulations and with publicly available data from the Environmental Protection Agency.

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

1. INTRODUCTION

A 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 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 resulting measurement error.

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 (e.g. Dockery *and others*, 1993; Samet *and others*, 2000; Pope *and others*, 2002), and recent studies emphasize the importance of using predicted individual air pollution exposures to account for spatial variability within urban areas (Jerrett *and others*, 2005b; Kunzli *and others*, 2005; Gryparis

*To whom correspondence should be addressed.

© The Author 2011. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

and others, 2007; Szpiro *and others*, 2010). 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 stream water quality and nearby watershed land cover (Madsen *and others*, 2008; Herlihy *and others*, 1998).

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

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

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. For a universal kriging exposure model with 450 monitors (as in the examples considered here), each nonlinear optimization takes 30–60 s on an Intel Xeon processor running at 2.33 GHz, so a parametric bootstrap with only 100 samples would take approximately 1 h. This is uncomfortably long for routine usage, but it is feasible if the bootstrap is employed judiciously. If we consider, instead, a more complex spatiotemporal model of the kind being used in modern air pollution studies (Szpiro *and others*, 2010), the time required for a single optimization is an hour or more, so a full parametric bootstrap is essentially impractical unless its use is restricted to a very limited number of definitive analyses.

We describe a new method termed the "parameter bootstrap" that is a less computationally demanding approximation to the parametric bootstrap. The parameter bootstrap is consistent with a decomposition of the measurement error into 2 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. The methodology extends easily to more complex spatiotemporal exposure models that generalize universal kriging (Banerjee *and others*, 2004; Szpiro *and others*, 2010).

In Section 2, we introduce notation and formally set 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 briefly review 2 alternative strategies that have been proposed in recently published papers. In Section 5, we illustrate our methodology in a simulation study and compare it to other methods, and in Section 6, we consider an example with publicly available stream data from the Environmental Protection Agency (EPA). We conclude in Section 7 with a discussion.

2. Notation and problem setup

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 regression coefficient of interest β_X. 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.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

Inference for β_X would be straightforward if X, Y, and Z were all observed. We would estimate $\hat{\beta}_X$ by ordinary least squares (OLS) and then estimate $\hat{\theta}_\varepsilon$ from the residuals and use a sandwich-based SE (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 SE 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. N^* is the number of exposure monitors. Assume that X and X^* are jointly distributed as

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

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

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

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).

Although the exposure X is not observed directly, we can exploit the observed values X^* and the spatial model in (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

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

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 the geographic locations are fixed and known, as are the corresponding GIS covariates S and S^* and any covariates Z in the outcome model. The regression coefficients β_0, β_X, β_Z, and α and variance parameters θ_ε and θ_η are all fixed but unknown. A realization from the data-generating mechanism is obtained by drawing 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 SEs by the procedure described at the beginning of Section 2. However, these SEs 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 induced by η and η^*. Therefore, naïve SEs will typically not estimate the true sampling variability of $\hat{\beta}_X$. In addition to altering the SEs, the measurement error may introduce bias.

We decompose the measurement error into 2 components

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

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

where the Berkson-like component is

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

and the classical-like component is

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

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 change the sampling variance of $\hat{\beta}_X$, and the classical-like component can also introduce bias.

3.1 *Berkson-like component of the error*

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 *and others*, 2006, p. 9). With known exposure model parameters, it is easy to see that this holds for U_{BL}

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

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 (2.1) in the form

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

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 $E(U_{\mathrm{BL}}|W) = 0$, 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 SEs by accounting for the correlation in the new noise term $\varepsilon' = U_{\mathrm{BL}}\beta_X + \varepsilon$, using either generalized least squares or the sandwich estimator, as would be appropriate for Berkson error with a nondiagonal covariance (Gryparis *and others*, 2009; Szpiro *and others*, 2008; Carroll *and others*, 2006, p. 90). This reasoning is not completely correct, however, because it is based on treating W as fixed.

3.2 *Classical-like component of the error*

A primary feature of 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 *and others*, 2006, p. 28). 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,

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

we emphasize that U_{CL} is similar to classical measurement error but also distinct in important ways, so we cannot rely on 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 $\hat{\beta}_X$. The bias in our examples is relatively small, but it has some interesting features and for completeness we illustrate bias correction using bootstrap methods and discuss the theoretical properties of bias from this form of classical-like error in the Online Supplement 1 (see supplementary material available at *Biostatistics* online). One interesting finding is that the bias can be away from the null, rather than toward the null as in the case of standard classical measurement error.

4. CORRECTION METHODS

4.1 *Parametric bootstrap*

A natural, but computationally intensive, approach to estimating SEs and correcting bias is the parametric bootstrap (Davison and Hinkley, 1997; Madsen *and others*, 2008). The parameter estimate of interest $\hat{\beta}_X$ is calculated as in Section 2, and we wish to approximate its sampling distribution under the true data-generating mechanism. We do this by simulating bootstrap samples under our best estimate of the data-generating mechanism and calculating their empirical distribution. Given a set of observations Y and X^*, the parametric bootstrap SE based on M bootstrap samples is derived as follows:

1. Estimate the exposure model parameters $\hat{\alpha}$ and $\hat{\theta}_\eta$ by nonlinear optimization in (2.2).
2. Derive W from (2.4) and use it in place of X in (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 $j = 1, \ldots, M$
   (a) Simulate a new set of observations Y_j and X_j^* based on the models in (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 (2.2), using X_j^* in place of X^*.
   (c) Plug $\hat{\alpha}_j$,$\hat{\theta}_{\eta,j}$, and X_j^* into (2.4) to derive W_j.
   (d) Calculate $\hat{\beta}_{X,j}$ by OLS in (2.1), using W_j and Y_j in place of X and Y.
4. Calculate the parametric bootstrap SE as the empirical standard deviation of the $\hat{\beta}_{X,j}$.

Note that in step 3(a), we simulate X_j in order to obtain Y_j, but we do not use X_j in the remainder of the procedure. See the Online Supplement 3 (see supplementary material available at *Biostatistics* online) for additional implementation details.

It is straightforward to use $\hat{\beta}_{X,j}$ to estimate and correct for bias rather than to derive SEs (Davison and Hinkley, 1997). In principle, we need a nested double bootstrap to rigorously derive a bias-corrected point estimates and corresponding SEs, but such a procedure can require M^2 bootstrap samples which is very computationally intensive. Since the bias tends to be small in our examples, we approximate a nested double bootstrap by applying a bias correction and estimating SEs based on the same set of M bootstrap samples, so our SEs do not include the additional variability from bias correction.

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 3(b) above. This is feasible because we can typically obtain

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

an estimate of the sampling distribution for $\hat{\alpha}$ and $\hat{\theta}_\eta$ in step (1) without much additional computation. The procedure differs from the parametric bootstrap in the addition of step 1(a) and modification of step 3(b).

1. Estimate the exposure model parameters $\hat{\alpha}$ and $\hat{\theta}_\eta$ by nonlinear optimization in (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 (2.4) and use it in place of X in (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 $j = 1, \ldots, M$
   (a) Simulate a new set of observations Y_j and X_j^* based on the models in (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 (2.4) to derive W_j.
   (d) Calculate $\hat{\beta}_{X,j}$ by OLS in (2.1), using W_j and Y_j in place of X and Y.
4. Calculate the parameter bootstrap SE as the empirical standard deviation of the $\hat{\beta}_{X,j}$.

As described in the Online Supplement 3 (see supplementary material available at *Biostatistic* online), our implementation of step 1(a) uses a Gaussian approximation centered at the maximum likelihood value with covariance based on the estimated Hessian. In the Online Supplement 2 (see supplementary material available at *Biostatistic* online), we also describe assumptions that underlie validity of the parameter bootstrap.

4.3 *Partial parametric bootstrap*

If we neglect the classical-like error, another alternative 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 3(b). We call this the partial parametric bootstrap. Since the partial parametric bootstrap only accounts for Berkson-like error, we use it to estimate SEs but not for bias correction.

4.4 *Other correction methods*

Two alternative methods have been proposed in recently published papers. We describe these approaches briefly and compare them to our proposed bootstrap methodology in the simulation and data examples that follow. For more details, we refer an interested reader to the cited papers.

Gryparis *and others* (2009) and Madsen *and others* (2008) propose jointly modeling the exposure and outcome data in order to estimate $\hat{\beta}_X$. Since the joint model is multivariate normal given the parameters, it is possible to write down a joint likelihood and estimate all the model parameters by either maximum likelihood or Bayesian methods. We show results in a subset of our examples from a joint model fit by maximum likelihood.

Another approach proposed by Gryparis *and others* (2009) is to leave out a subset of the monitoring data for out-of-sample validation and then to use regression calibration to derive bias-corrected effect estimates. This approach is based on a classical measurement error model, which we and they have shown does not hold. It also requires fitting the exposure model with only a subset of the available data. The out-of-sample regression calibration algorithm given by Gryparis *and others* (2009) is for uncorrelated outcomes, and we implement their algorithm for a subset of our examples with uncorrelated outcomes under the optimistic assumption that 50 additional validation monitors are available.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

5. SIMULATIONS

We conduct a simulation study based on the data we analyze below in Section 6. We use the universal kriging exposure model described above and allow for correlation in the outcome residuals. In a 300 × 500 box, we randomly select locations for $N^* = 450$ exposure monitors and $N = 100$ or 2000 outcome measurements. The x and y coordinates are covariates in the universal kriging exposure model with regression coefficients $\alpha = (-25.95, -0.0035, 0.00084)^t$, and the spatial correlation has an exponential variogram structure with range $\phi_\eta = 24.13$, partial sill $\psi_\eta = 3.76$, and nugget $\tau_\eta = 1.34$ (Cressie, 1993).

The linear regression model for the outcome conditional on X is (2.1), with $\beta_0 = 5.06$, $\beta_1 = -0.322$, and no additional covariates Z. We consider the case of uncorrelated residuals ε with variance $\sigma_\varepsilon^2 = 0.76$, and we also consider the case of correlated residuals ε following an exponential variogram structure with range $\phi_\varepsilon = 80.39$, partial sill $\psi_\varepsilon = 0.26$, and nugget $\tau_\varepsilon = 0.50$.

In Table 1, we summarize the results for 2000 Monte Carlo simulations. Due to the computational intensity of the parametric bootstrap, we restrict to 100 Monte Carlo runs for these results. Scatterplots illustrating agreement between parameter bootstrap and parametric bootstrap SEs are shown in Figure 1. With uncorrelated outcomes, naïve SEs that do not correct for measurement error are too small compared

Table 1. *Simulation results for universal kriging exposure surface with range $\phi_\eta = 24.13$. The columns give the bias and standard deviation (SD) of the estimates, the mean and mode of the estimated SEs, and the coverage for* 95% *Wald confidence intervals (CIs). Results are based on* 2000 *Monte Carlo simulations, except for the parametric bootstrap which is based on 100 Monte Carlo simulations and maximum likelihood which is based on* 50 *Monte Carlo simulations*

	Bias	SD	SE (mean)	SE (mode)	95% CI Coverage
$N = 100$					
Independent outcomes					
No correction	−0.003	0.075	0.072	0.070	95
Partial parametric bootstrap	−0.003	0.075	0.072	0.071	95
Parameter bootstrap	0.000	0.075	0.079	0.074	96
Parametric bootstrap	−0.009	0.077	0.079	0.073	98
Regression calibration	−0.012	0.104	0.102	0.084	95
Maximum likelihood	−0.011	0.081	0.073	0.079	96
Correlated outcomes					
No correction	−0.004	0.093	0.097	0.085	95
Partial parametric bootstrap	−0.004	0.093	0.101	0.097	95
Parameter bootstrap	−0.001	0.093	0.105	0.099	96
Parametric bootstrap	−0.008	0.099	0.106	0.107	94
$N = 2000$					
Independent outcomes					
No correction	−0.002	0.027	0.016	0.016	78
Partial parametric bootstrap	−0.002	0.027	0.023	0.023	91
Parameter bootstrap	0.001	0.027	0.028	0.027	96
Parametric bootstrap	−0.002	0.027	0.029	0.027	97
Regression calibration	−0.010	0.075	0.066	0.045	93
Maximum likelihood	0.000	0.021	0.022	0.022	96
Correlated outcomes					
No correction	−0.001	0.064	0.068	0.053	94
Partial parametric bootstrap	−0.001	0.064	0.073	0.064	95
Parameter bootstrap	0.001	0.064	0.077	0.066	96
Parametric bootstrap	−0.008	0.067	0.081	0.065	97

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

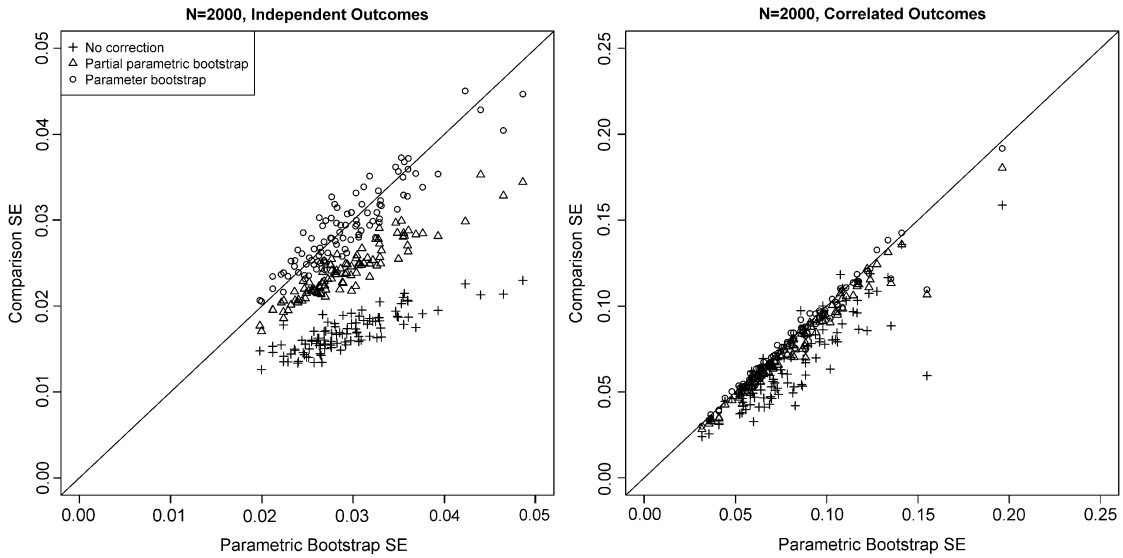

Fig. 1. Scatterplot showing agreement between parametric bootstrap SEs and parameter bootstrap approximation, based on 100 Monte Carlo simulations. Partial parametric bootstrap and uncorrected SEs are also included for comparison.

to the observed sampling distribution of $\hat{\beta}_X$, resulting in less than nominal coverage for 95% confidence intervals. The parameter bootstrap consistently gives near nominal confidence interval coverage. There is some evidence of over-coverage with the parametric bootstrap, but this may be attributable to the smaller number of simulations. In the correlated outcome model, the coverage for naïve SEs is closer to nominal. This is presumably because much of the Berkson-like error appears as additional correlated variability in the outcome and is accounted for by the sandwich form.

We also show SEs calculated with the partial parametric bootstrap. In scenarios with $N = 100$, the partial parametric bootstrap gives similar results to the parameter bootstrap. This suggests that the Berkson-like component of measurement error dominates in this situation. For $N = 2000$, however, the partial parametric bootstrap SEs are significantly smaller, indicating that the classical-like component is important in that setting.

The bias in $\hat{\beta}_X$ is a very small component of the total error in all scenarios, and it is partially corrected by the parameter bootstrap. Due to the small number of parametric bootstrap simulations, it is difficult to confirm that this approach provides an effective bias correction, but given the general agreement with the parameter bootstrap we expect it to perform similarly.

We show results for out-of-sample regression calibration and joint maximum likelihood estimation in Table 1 for the cases with uncorrelated outcomes. We do not include cases with correlated outcomes because the regression calibration method in Gryparis *and others* (2009) is not applicable, and the maximum likelihood algorithm failed to consistently converge due to difficulty in identifying the covariance structures between the exposure and outcome variables. Regression calibration results in reasonable coverage probabilities but much larger SEs than the other methods. Joint maximum likelihood estimation also gives nominal coverage probabilities, with somewhat smaller SEs than the 2-step methods with bootstrap corrections. We restricted this part of the simulation study to 50 Monte Carlo simulations due to the computational burden (it took on average 8 h to optimize the joint likelihood with $N = 2000$, compared to less than 5 min for the parameter bootstrap).

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

Table 2. *Simulation results for misspecified variance models. All simulations are fit with an exponential variogram model, but the data are generated according to either a Gaussian, spherical, or cubic model. The columns give the bias and standard deviation (SD) of the estimates, the mean and mode of the estimated SEs, and the coverage for* 95% *Wald confidence intervals (CIs). Results are based on* 2000 *Monte Carlo simulations, except for the parametric bootstrap which is based on* 100 *Monte Carlo simulations and maximum likelihood which is based on* 50 *Monte Carlo simulations*

	Bias	SD	SE (mean)	SE (mode)	95% CI Coverage
$N = 2000$					
Independent outcomes					
Gaussian variogram					
No correction	−0.013	0.021	0.014	0.014	74
Partial parametric bootstrap	−0.013	0.021	0.020	0.019	90
Parameter bootstrap	−0.011	0.022	0.024	0.022	95
Parametric bootstrap	−0.012	0.021	0.023	0.022	95
Regression calibration	−0.009	0.048	0.048	0.040	95
Maximum likelihood	−0.002	0.017	0.017	0.017	94
Spherical variogram					
No correction	−0.008	0.026	0.016	0.016	77
Partial parametric bootstrap	−0.008	0.026	0.022	0.022	91
Parameter bootstrap	−0.005	0.026	0.028	0.025	96
Parametric bootstrap	−0.007	0.025	0.028	0.026	98
Regression calibration	−0.011	0.062	0.061	0.046	95
Maximum likelihood	0.000	0.019	0.020	0.020	96
Cubic variogram					
No correction	−0.016	0.024	0.015	0.015	71
Partial parametric bootstrap	−0.016	0.024	0.021	0.021	88
Parameter bootstrap	−0.013	0.025	0.030	0.025	96
Parametric bootstrap	−0.014	0.023	0.026	0.025	94
Regression calibration	−0.009	0.057	0.054	0.044	95
Maximum likelihood	−0.003	0.017	0.018	0.018	96

Finally, in Table 2, we illustrate the impact of misspecifying the spatial correlation in the exposure model. Other types of model misspecification are also possible, but we focus on the spatial correlation in the exposure since it is particularly difficult to know in advance. We consider Gaussian, spherical, and cubic variogram models, but we assume an exponential model for the purposes of estimation. Ignoring measurement error or using the partial parametric bootstrap results in less than nominal confidence interval coverage. The parameter bootstrap and parametric bootstrap give nearly nominal coverage, as do out-of-sample regression calibration and joint maximum likelihood fitting. As in the case of a correctly specified model, regression calibration results in much larger SEs, while joint maximum likelihood gives somewhat smaller standard errors than the bootstrap approaches (at significant computational cost). There is bias that the bootstrap methods fail to correct, while maximum likelihood estimates are nearly unbiased.

6. EXAMPLE

The Environmental Monitoring and Assessment Program (EMAP) was conducted by the EPA from 1990–2006 to advance the science of ecological risk assessment and improve the EPA's ability to estimate

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

current and future risks to the health of aquatic ecological resources (US EPA monitoring and assessment program, 1999). Previous work has found that there is a strong relationship between local land use and water quality in nearby streams (Herlihy *and others*, 1998). For example, a higher percentage of local forestation has been found to be associated with improved stream water quality. Following Madsen *and others* (2008), we consider forestation level ($\text{logit}((0.98 \times \%\text{forestation}) + 1)$) as the exposure and analyze its association with chloride concentrations ($\log(\mu\text{Eq}/L)$), where elevated chloride concentrations are a marker for poor stream water quality.

We use EMAP data from the Mid-Atlantic Highlands region of the eastern United States collected during the years 1993–1996 (US EPA monitoring and assessment program, 1999). Where multiple measurements are available from different times at the same location, we use the earliest time. The outcome and the exposure are both available at a total 422 of these locations, so based on these locations, we can estimate the coefficient in a linear model without measurement error. Allowing for spatially correlated outcomes, we find a highly statistically significant negative association between local forestation and chloride concentrations ($\hat{\beta}_X = -0.346$, SE $= 0.025$).

At an additional 157 locations, only the chloride concentrations are available. To assess the impact of measurement error and our correction methods, we use a universal kriging model (with an exponential variogram and latitude and longitude as covariates) to predict the exposure at these locations and re-estimate the association using the predicted exposures and measured outcomes at 157 locations. The exposure model parameter estimates are $\hat{\alpha} = (-28.91, -0.0037, 0.0012)$, range $\hat{\phi}_\eta = 13.92$ km, partial sill $\hat{\psi}_\eta = 4.32$, and nugget $\hat{\tau}_\eta = 0.48$. A map of the respective locations is shown in Figure 2.

The results are shown in Table 3. The uncorrected effect estimate is -0.390 with a SE of 0.105, which is consistent with primarily Berkson-like measurement error since there is little change in the effect estimate compared to the case with no measurement error (only a small part of the increased SE can be attributed to the smaller sample size of 157 instead of 422). The partial parametric bootstrap does not change the estimated SE at all, which suggests that the Berkson-like error is nearly pure Berkson error. The parameter bootstrap and parametric bootstrap increase the estimated SEs slightly (0.115 and 0.111, respectively) and estimate little or no bias. Joint maximum likelihood optimization failed to converge (even after trying 20 random initial conditions) due to difficulty in distinguishing between spatial correlation in the measurement error and the outcome, indicating that joint modeling is not feasible for the present example.

In conclusion, when we use predicted exposures at stream locations where the true exposures are not available, we see statistically significant evidence of a negative association between local forestation and stream water quality as measured by chloride concentrations. The SE is inflated by the presence of measurement error, and the correct SE can be estimated at little additional computational cost by the parameter bootstrap. It turns out that a naïve analysis also gives nearly correct SEs, although we needed to do the bootstrap analysis to verify this finding.

7. Discussion

We have characterized the measurement error from using smoothing to predict exposures in environmental statistics association studies when the exposure and outcome data are misaligned in space. The resulting measurement error has a Berkson-like component 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 2 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

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

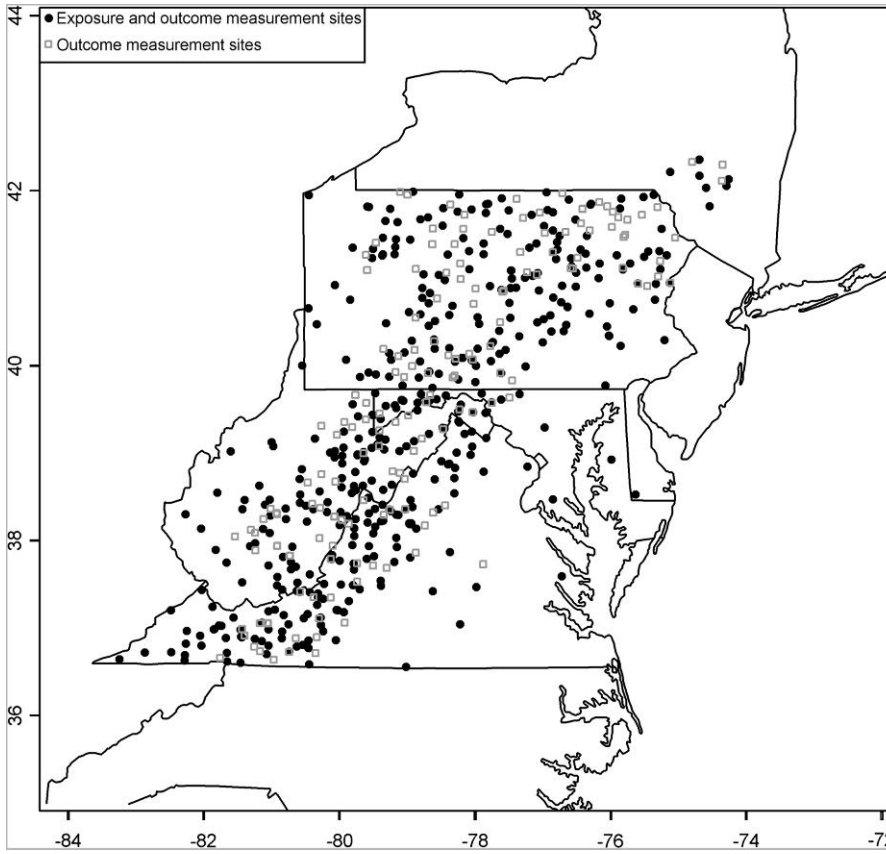

Fig. 2. EPA stream locations, for example, data analysis ($N^* = 422$ sites with exposure and outcome data, $N = 157$ locations with outcome data only).

bias and standard errors. This 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 parameters estimated from observed data. We have defined 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 physical processes that arise in environmental statistics, it is not clear what real-world phenomenon the spatial random effect represents. A promising direction for future research is to investigate the scientific validity of this and other possible data-generating mechanisms and to characterize the implications for measurement error correction.

We propose the parameter bootstrap as a less computationally intensive approximation to the parametric bootstrap. The main assumption required for the parameter bootstrap is that the estimated sampling distribution for the exposure model parameter estimates be a valid approximation to the true sampling distribution. In general, this is true for sufficiently rich exposure data, but in some applications, the available monitoring data are limited. As a computational tool, the parameter bootstrap is necessary when estimating the exposure model parameters is computationally intensive, which occurs when there is a relatively large amount of exposure data. Therefore, the choice between the parametric bootstrap and the parameter bootstrap should be informed by the amount of exposure data available, considering

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

Table 3. *Results of estimating the relationship between the log-transformed chloride level ("outcome") and the logit-transformed percent local forestation ("exposure") in streams, using EPA data from the Mid-Atlantic Highlands region of the eastern United States during the years* 1993–1996. *The true exposure result is based on* 422 *locations at which both the exposure and outcome data are available. The predicted exposure results are based on outcomes and predicted exposures at* 157 *locations for which exposure data are not available, with the other* 422 *locations used to fit the exposure model. The model allows for spatially correlated outcomes*

	N	$\hat{\beta}_X$	SE
True exposure	422	−0.346	0.025
Predicted exposure, no correction	157	−0.390	0.105
Predicted exposure, partial parametric bootstrap	157	−0.390	0.103
Predicted exposure, parameter bootstrap	157	−0.388	0.115
Predicted exposure, parametric bootstrap	157	−0.397	0.111

implications for the computational burden of the parametric bootstrap and the validity of the parameter bootstrap.

In the universal kriging model considered here with 450 exposure monitors, the parametric bootstrap is marginally feasible, requiring approximately 1 h of computing time for 100 bootstrap samples (compared to less than 5 min for 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 with the parametric bootstrap using a limited number of samples. In applications where the exposure model is more complex (e.g. the spatiotemporal air pollution model described by Szpiro *and others*, 2010), a single optimization can take on the order of an hour so the parameter bootstrap's computational advantage becomes even more important.

We compared our bootstrap methods to 2 recently proposed alternatives (Gryparis *and others*, 2009; Madsen *and others*, 2008). Regression calibration is incompatible with the theoretical properties of the measurement error, and it results in much larger SEs than any of the other alternatives. Joint estimation by maximum likelihood performed consistently well, even with a misspecified exposure model and resulted in somewhat smaller SEs than bootstrap methods. However, as Gryparis *and others* (2009) point out, the joint estimation methodology can be extremely computationally intensive and can lead to spurious feedback into the exposure when there are outliers or misspecification in the outcome model. It is also difficult to fit a joint model with spatial correlation in the outcomes, due to the challenge in distinguishing this correlation from correlation in the exposure. The parameter bootstrap is a computationally efficient alternative that works well in a wide range of settings, and further research comparing it to the joint modeling approach is needed to determine which is preferable for problems where both are feasible.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

The authors thank the associate editor and 2 anonymous referees for helpful suggestions.
conflict of Interest: None declared.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

FUNDING

US Environmental Protection Agency (R831697); Assistance Agreement (CR-83407101); National Institute of Environmental Health Sciences (R01-ES009411 and P50 ES051915). It has not been subjected to peer and policy review so no official endorsement should be inferred.

REFERENCES

BANERJEE, S., CARLIN, B. P. AND GELFAND, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman and Hall/CRC.

BRAUER, M., HOEK, G., VAN VLIET, P., MELIEFSTE, K., FISCHER, P., GEHRING, U., HEINRICH, J., CYRYS, J., BELLANDER, T., LEWNE, M. *and others* (2003). Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiology* **14**, 228–239.

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. AND CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd edition. Boca Raton, FL: Chapman and Hall/CRC.

CRESSIE, N. A. C. (1993). *Statistics for Spatial data*. New York: John Wiley & Sons.

DAVISON, A. C. AND HINKLEY, D. V. (1997). *Bootstrap Methods and Their Applications*. New York, NY: University of Cambridge Press.

DOCKERY, D. W., POPE, C. A., XU, X., SPANGLER, J. D., WARE, J. H., FAY, M. E., FERRIS, B. G. AND SPEIZER, F. E. (1993). An association between air pollution and mortality in six cities. *The New England Journal of Medicine* **329**, 1753–1759.

GRYPARIS, A., COULL, B. A., SCHWARTZ, J. AND SUH, H. H. (2007). 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**, 183–209.

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

HERLIHY, A. T., STODDARD, J. T. AND JOHNSON, C. B. (1998). The relationship between stream chemistry and watershed land cover in the Mid Atlantic region, U.S. *Water, Air, and Soil Pollution* **105**, 377–386.

JERRETT, M., ARAIN, A., KANAROGLOU, P., BECKERMAN, B., POTOGLOU, D., SAHSUVAROGLU, T., MORRISON, J. AND GIOVIS, C. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology* **15**, 185–204.

JERRETT, M., BURNETT, R. T., MA, R., POPE, C. A., KREWSKI, D., NEWBOLD, K. B., THURSTON, G., SHI, Y., FINKELSTEIN, N., CALLE, E. E. *and others* (2005). Spatial analysis of air pollution mortality in Los Angeles. *Epidemiology* **16**, 727–736.

KIM, S. Y., SHEPPARD, L. AND KIM, H. (2009). Health effects of long-term air pollution: influence of exposure prediction methods. *Epidemiology* **20**, 442–450.

KUNZLI, N., JERRETT, M., MACK, W. J., BECKERMAN, B., LABREE, L., GILLILAND, F., THOMAS, D., PETERS, J. AND HODIS, H. N. (2005). Ambient air pollution and atherosclerosis in Los Angeles. *Environmental Health Perspectives* **113**, 201–206.

LIANG, K.-Y. AND ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

MADSEN, L., RUPPERT, D. AND ALTMAN, N. S. (2008). Regression with spatially misaligned data. *Environmetrics* **19**, 453–467.

MILLER, K. A., SICOVICK, D. S., SHEPPARD, L., SHEPHERD, K., SULLIVAN, J. H., ANDERSON, G. L. AND KAUFMAN, J. D. (2007). Long-term exposure to air pollution and incidence of cardiovascular events in women. *The New England Journal of Medicine* **356**, 447–458.

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016

POPE, C. A., BURNETT, R. T., THUN, M. J., CALLE, E. E., ITO, K., KREWSKI, D. AND THURSTON, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association* **9**, 1132–1141.

SAMET, J. M., DOMINICI, F., CURRIERO, F., COURSAC, I. AND ZEGER, S. L. (2000). Particulate air pollution and mortality: findings from 20 u.s. cities. *The New England Journal of Medicine* **343**, 1742–1749.

SZPIRO, A. A., SAMPSON, P. D., SHEPPARD, L., LUMLEY, T., ADAR, S. D. AND KAUFMAN, J. D. (2010). Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics* **21**, 606–632.

SZPIRO, A. A., SHEPPARD, L. AND LUMLEY, T. (2008). Accounting for errors from predicting exposures in environmental epidemiology and environmental statistics. *UW Biostatistics Working Paper Series, Working Paper 330*. Berkeley, CA: Berkeley Electronic Press.

US EPA ENVIRONMENTAL MONITORING AND ASSESSMENT PROGRAM (1999). *Mid-Atlantic Streams 1993–96 Data Sets*. http://www.epa.gov/emap/html/data/surfwatr/data/mastreams/9396/wchem/chmval.txt, http://www.epa.gov/emap/html/data/surfwatr/data/mastreams/9396/location/watchr.txt.

[*Received January 7, 2010; revised December 15, 2010; accepted for publication December 16, 2010*]

Downloaded from http://biostatistics.oxfordjournals.org/ at University of North Texas on December 26, 2016