Journal of Exposure Science and Environmental Epidemiology (2015) 25, 138–144

www.nature.com/jes

ORIGINAL ARTICLE

Consequences of kriging and land use regression for PM2.5 predictions in epidemiologic analyses: insights into spatial variability using high-resolution satellite data

Stacey E. Alexeeff[1,2], Joel Schwartz[3], Itai Kloog[3,4], Alexandra Chudnovsky[3], Petros Koutrakis[3] and Brent A. Coull[1]

Many epidemiological studies use predicted air pollution exposures as surrogates for true air pollution levels. These predicted exposures contain exposure measurement error, yet simulation studies have typically found negligible bias in resulting health effect estimates. However, previous studies typically assumed a statistical spatial model for air pollution exposure, which may be oversimplified. We address this shortcoming by assuming a realistic, complex exposure surface derived from fine-scale (1 km × 1 km) remote-sensing satellite data. Using simulation, we evaluate the accuracy of epidemiological health effect estimates in linear and logistic regression when using spatial air pollution predictions from kriging and land use regression models. We examined chronic (long-term) and acute (short-term) exposure to air pollution. Results varied substantially across different scenarios. Exposure models with low out-of-sample R^2 yielded severe biases in the health effect estimates of some models, ranging from 60% upward bias to 70% downward bias. One land use regression exposure model with >0.9 out-of-sample R^2 yielded upward biases up to 13% for acute health effect estimates. Almost all models drastically underestimated the SEs. Land use regression models performed better in chronic effect simulations. These results can help researchers when interpreting health effect estimates in these types of studies.

Journal of Exposure Science and Environmental Epidemiology (2015) **25,** 138–144; doi:10.1038/jes.2014.40; published online 4 June 2014

Keywords: air pollution; kriging; land use regression; measurement error; PM2.5; spatial models

INTRODUCTION

There is strong epidemiological evidence that both short-term and long-term exposures to air pollution are related to cardiovascular morbidity and mortality.[1] In particular, much of the air pollution research shows that exposure to ambient particulate matter (PM) with aerodynamic diameter $\leq 2.5\,\mu g/m^3$ ($PM_{2.5}$) is associated with many adverse cardiovascular outcomes. In addition, ambient levels of $PM_{2.5}$ often vary within a given city or region, and traffic sources may contribute to this variation.[2,3] However, levels of $PM_{2.5}$ are typically measured only at a small number of stationary monitoring sites, which makes this regional heterogeneity hard to fully characterize.

Spatial modeling of air pollution levels is becoming widespread in air pollution epidemiology research. Kriging (also called ordinary kriging or simple kriging, with a constant mean) and land use regression (also called universal kriging, with a mean function that depends on spatial covariates) have been used to predict $PM_{2.5}$ exposures and study relationships with health, such as the assessment of the short-term relationship between $PM_{2.5}$ and cardiac responses[4] and associations between $PM_{2.5}$ and cancer mortality.[5]

The use of spatially predicted air pollution exposures in an epidemiologic analysis can be viewed in a measurement error framework, where the predicted exposures represent imperfect surrogates of the true exposures. In general, the naive plug-in of the individual-specific exposure estimates can lead to biased health effect estimates and overstated confidence in the resulting risk assessments.[6] However, in the statistical literature, several simulation studies have shown that direct use of the predicted exposures often induces little to no bias.[7–10] One explanation for those findings is that the exposure surfaces are simulated from spatial fields in a well-characterized statistical model. In real data scenarios, the actual performance of the naive plug-in estimator and the degree to which bias and variance adjustments need to be made are unknown.

A gold standard for the fine-scale spatial distribution of air pollution throughout an entire region is not available. Thus, the extent to which this exposure measurement error may be affecting health effect analyses is largely unknown because of inherent lack of validation data to study such an issue. In particular, there is no complete spatial representation of ambient air pollution exposure surfaces. A recent development is the availability of satellite measurements of aerosol optical depth (AOD) at the 10 km × 10 km resolution,[11] which can be calibrated to reflect $PM_{2.5}$ concentrations.[12,13] In addition, new satellite AOD measurements are now available at 1 km × 1 km resolution.[14] We propose that calibrated high-resolution satellite data at 1 km × 1 km could be viewed as a 'silver standard' of comparison to evaluate the performance of health effect estimators based on spatial air pollution predictions from kriging and land use regression.

[1]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA; [2]National Center for Atmospheric Research, Institute for Mathematics Applied to Geosciences, Boulder, Colorado, USA; [3]Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, USA and [4]Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Be'er Sheva, Israel. Correspondence: Dr. Stacey E. Alexeeff, National Center for Atmospheric Research, Institute for Mathematics Applied to Geosciences, PO Box 3000, Boulder, CO 80307-3000, USA.
Tel.: +1 303 497 2493. Fax: +1 303 497 2483.
E-mail: salexeef@ucar.edu
Received 28 October 2013; revised 29 April 2014; accepted 30 April 2014; published online 4 June 2014

In this study, we investigate the consequences of measurement error on health effect estimates via a simulation study, in which the true exposure surface is based on high-resolution calibrated satellite data. Under common scenarios of linear and logistic health models, we examine the magnitude and direction of the bias in health effect parameter estimates as well as the coverage of naive 95% confidence intervals (CIs). This analysis yields new insight on the practical implications of epidemiological analyses that use spatial model predictions in place of real air pollution surfaces.

MATERIALS AND METHODS

Satellite AOD Data

Daily spectral AOD data were obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) on the Aqua satellite for the year 2003. A new algorithm called Multi-angle Implementation of Atmospheric Correction (MAIAC) has been developed to process MODIS data.[14,15] MAIAC retrieves aerosol parameters over land at 1 km resolution simultaneously with parameters of a surface bidirectional reflectance distribution function. This is accomplished by using the time series of MODIS measurements and simultaneous processing of groups of pixels. The MAIAC algorithm ensures that the number of measurements exceeds the number of unknowns, a necessary condition for solving an inverse problem without empirical assumptions typically used by current operational algorithms. The MODIS time series accumulation also provides multi-angle coverage for every surface grid cell, which is required for the bidirectional reflectance function retrievals from MODIS data. The improved accuracy of the MAIAC algorithm results from using the explicit surface characterization method in contrast to the empirical surface parameterization approach. Further, MAIAC incorporates a cloud mask algorithm based on spatiotemporal analysis that augments traditional pixel-level cloud detection techniques.[16] Daily values of AOD were assigned to the grid cell where the AOD retrieval centroid was located. One feature of the AOD data is that some of the grid-specific AOD values are missing on some days due to cloud cover or snow cover.[12,17] Thus, the spatial coverage of the AOD data varies considerably by day.

Air Pollution Monitors

Data for daily $PM_{2.5}$ mass concentrations across New England for the year 2003 were obtained from the US Environmental Protection Agency (EPA) Air Quality System (AQS) database as well as the IMPROVE (Interagency Monitoring of Protected Visual Environments) network. IMPROVE monitor sites are located in national parks and wilderness areas, whereas AQS monitoring sites are located across New England including urban areas such as downtown Boston. There were 71 monitors with unique locations operating in New England during the study period.

Spatial and Temporal Covariates

Spatial covariates included major roads, point emissions and area emissions.

Data on the density of major roads were based on A1 roads (hard surface highways including Interstate and US numbered highways, primary State routes, and all controlled access highways) data obtained through the US census 2000 topologically integrated geographic encoding and referencing system. As the distributions of covariates representing density of major roads were highly right-skewed, they were log-transformed.

Temporal covariates included wind speed, humidity, visibility, and height of the planetary boundary layer. All meteorological variables (temperature, wind speed, humidity, and visibility) were obtained through the national climatic data center. Height of the planetary boundary layer data was obtained from the North American Regional Reanalysis. Further details on spatial and temporal covariates are given in Kloog *et al.*[12,13,18]

Calibration of AOD

A description of the method used to calibrate the AOD values to represent $PM_{2.5}$ concentrations is given in Kloog *et al.*[12,13,18] Briefly, the relationship between $PM_{2.5}$ and AOD at the monitoring sites was modeled using a mixed-effects regression model in which $PM_{2.5}$ was the dependent variable and AOD was the main explanatory predictor. The model included spatial covariates for major roads, point emissions and area emissions, and temporal covariates for wind speed, visibility, and height of the planetary boundary layer, with interactions between AOD and random intercepts for each day.

Kloog *et al.*[12] also includes a third stage of modeling, which imputes $PM_{2.5}$ at the missing AOD locations. In this study, we restricted to only days with ample AOD present to leverage the observed spatial variability in the data to minimize the use of exposures imputed from a land use regression model.

Simulation Setup

A simulation study was conducted to assess the performance of kriging and land use regression methods under the assumption that the true pollution surface follows that represented by the highly resolved 1-km satellite-derived predictions. Separate simulation studies were conducted to consider studies of chronic health effects due to long-term air pollution exposures and acute health effects due to short-term air pollution exposures. We restricted our simulation studies to the 32 days with at least 50,000 grid cells of AOD data available.

We considered two types of health outcomes: a binary health outcome and a continuous health outcome. A linear regression health model was assumed for the continuous health outcome, where the outcome depends linearly on the exposure. For the binary health outcome, a logistic regression health model was assumed, where the outcome depends linearly on the exposure through a logit link function applied to the probability of the outcome. No other confounding variables were included in the health model. We explored exposure models with a Matern covariance function and two levels of smoothness ($\kappa = 0.5$ is rough and $\kappa = 2.0$ is smooth). We also contrasted two settings for the number of monitors where $m = 100$ represents a realistic setting (although still higher than the actual number of monitors in this region during the study period), and $m = 500$ represents an even-better-than-realistic scenario. This latter sample size was chosen to illustrate the degree to which the problems in health parameter estimates could be attributed to a relatively sparse number of monitors *versus* underlying model misspecification. This extremely dense monitoring network will have monitors much closer to the locations where exposure is predicted; however, any systematic problems in the exposure model will still induce some bias in the health effect parameters.

Acute Effects Simulation

We designed our acute effect simulation to mimic the setting of a health study of the short-term effects of PM. Using the 32 days of calibrated $PM_{2.5}$ predictions, we considered the relevant exposure period of interest to be 1 day of $PM_{2.5}$ exposure. For each simulation, we generated 1000 subjects' residential locations by randomly sampling the day of the exposure and then sampling the health locations by population density. Once the date and grid cell were randomly chosen, we assigned the corresponding calibrated $PM_{2.5}$ exposure at the grid cell. The health outcomes were generated to depend on the assigned exposure using the chosen health model type with no confounders. The 1000 subjects per simulation corresponded to approximately 30 subjects sampled from each of the 32 days. The monitor locations were chosen by a random uniform distribution across the exposure surface, and the corresponding daily calibrated $PM_{2.5}$ value at the monitor location was used as the observed exposure for each day.

Using the measured exposure at the monitor locations, the kriging or land use model was fit to the data by day and exposure predictions were generated for each day at the residential locations of the subjects. We considered four different modeling strategies. The C1 acute kriging models had a constant daily mean and a Matern covariance. The D1 acute land use regression models had a mean that depended on land use and temporal covariates: distance to nearest A1 road, density of major roads within 1 km, and temporal term humidity, wind speed, height of the planetary boundary layer, and vegetation. Note that these covariates are the same as those used in the satellite calibration procedure, so that this scenario represents the desirable setting in which the correct predictors are used in the land use regression. The D2 land use regression models had a mean that depended on only spatial covariates: distance to nearest C1 road, density of major roads within 1 km. The D3 land use regression models used a two-stage approach where we first subtract the daily mean across the monitors, then fit the spatial model to the centered daily data, and add the daily mean onto the spatial predictions. The predicted exposures were then fit to the health outcomes to estimate the association.

Chronic Effect Simulation

To emulate the setting of a health study of the chronic effects of PM, we generated a chronic exposure surface by averaging the calibrated $PM_{2.5}$ data at each grid cell over the 32 days of exposure. In this scenario, all subjects' exposures were sampled from this one common exposure surface. Thus, the spatial variability of the surface provided the only variability in the exposures of different subjects.

For each simulation, we generated 500 subjects' exposure and outcome measurements. To assign the exposure, we first generated each subjects' residential location by population density. Population density sampling was approximated using the geocoded locations of births during 2003 from a previous study.[18] We then assigned the corresponding average (over 32 days) calibrated $PM_{2.5}$ value at the subjects' residential location as the exposure. The health outcome was generated to depend on the assigned exposure using the chosen health model type with no confounders. The monitor locations were chosen by a random uniform distribution across the exposure surface, and the corresponding calibrated $PM_{2.5}$ value at the monitor location was used as the observed exposure.

Using the measured exposure at the monitor locations, the kriging or land use model was fit to the data and chronic exposure predictions were generated at the residential locations of the subjects. We used three different modeling strategies to predict the long-term exposures. The A1 chronic kriging models had a constant mean and a Matern covariance, and we applied one kriging fit to the monitor averages. The A2 chronic kriging models were fit to the daily monitor values and then averaged. The B1 chronic land use regression models had a mean that depended on land use covariates and a Matern covariance. Land use regression models for the chronic setting included terms for distance to nearest A1 road and density of major roads within 1 km. The predicted exposures were then fit to the health outcomes to estimate the association.

Supplementary Simulations

To address some related questions of interest, we ran a number of additional simulations. First, we examined the performance of all models under the null to see whether the size of the alpha = 0.05 test was inflated to a rate greater than 5%. We also considered simulated surfaces that had greater proportions of non-spatial Berkson error, representing the case of more instrument error in the actual monitoring measurements. Finally, we considered a simulated chronic surface fit with a misspecified kriging model to try to emulate some of the results seen in the chronic satellite scenarios. The results of these simulations are given in the Supplementary Material.

RESULTS

The average daily $PM_{2.5}$ levels from the calibrated AOD data ranged from 1.98 to 16.82 $\mu g/m^3$, with a mean of 7.47 $\mu g/m^3$. The $PM_{2.5}$ levels on all days at all locations ranged from 0.002 to 20.0 $\mu g/m^3$. Between-day variability accounted for 92% of the total variation in $PM_{2.5}$, whereas the within-day variability accounted for 8% of the total variation in $PM_{2.5}$ levels. A table summarizing the daily mean, SD, and number of grid cells for the $PM_{2.5}$ concentrations for each of the 32 days used in the study is given in the Supplementary Material section. Figure 1 shows the spatial $PM_{2.5}$ levels for one date, 10 September 2003, and the spatial $PM_{2.5}$ levels for the chronic average surface.

The results from the simulations of chronic pollution effects are shown in Tables 1 and 2, where Table 1 shows the results for a linear model relating chronic air pollution exposure to a continuous health outcome, and Table 2 shows the results for a logistic model relating chronic air pollution exposure to a binary health outcome. The A1 chronic kriging models result in notable upward bias and highly inflated empirical SEs in both the linear and logistic health regression models. The A2 chronic kriging models, which implement daily kriging, result in slight upward bias in the linear health model and notable attenuation bias in the logistic health model. These opposite direction effects are the result of the fact that the logistic model mean and variance are both estimated by a single parameter. For the chronic kriging model, we found that the alternative model using daily kriging reduced the overall bias to a level of 4% upward bias to 15% downward bias.

The B1 land use regression model, which included two land use terms, showed an improved exposure R^2 compared with the chronic kriging models and exhibited 1–5% upward bias in the health effect estimates in both the linear and logistic health models. There was still significant undercoverage in the linear health effect model. Overall, the results of these analyses showed that the estimation of the health effect parameter shows considerable sensitivity to different model setups.

The results for the simulations of acute effects of air pollution are shown in Tables 3 and 4, where Table 3 shows the results for a linear model relating acute air pollution exposure to a continuous health outcome, and Table 4 shows the results for a logistic model relating acute air pollution exposure to a binary health outcome. The C1 daily kriging models show negligible bias (1% to 2%) and only slightly inflated empirical SEs compared with using the true acute exposure, for both the linear and logistic health models. In contrast, there was considerable downward bias and inflation of empirical SEs in both the linear and logistic health effect setting for the D1 land use regression, which included both temporal weather covariates and spatial land use terms. This led to naive CIs that typically missed the true effect completely due to both the bias and the discrepancy between the naive model-based SE and the empirical SE. The main problem with this exposure model including both temporal and spatial terms is that the underlying atmospheric processes are too complex to be approximated by a simple statistical model. Given the large amount of day-to-day variation compared with spatial variation in the true levels, use of a daily spatial interpolation with a smoothing factor is more effective than attempting to model the underlying temporal process.

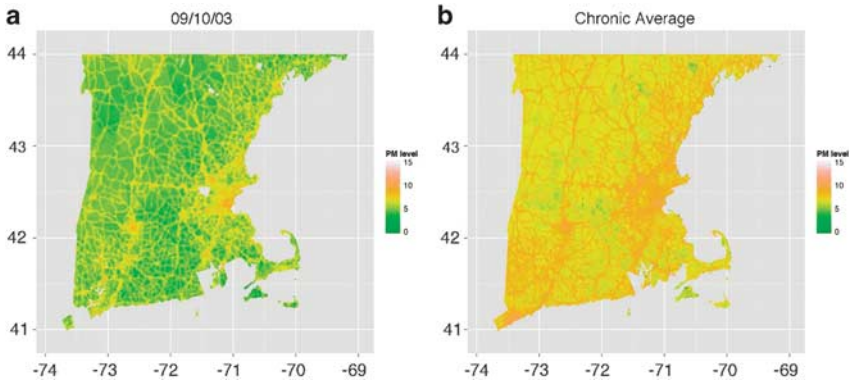

Figure 1. $PM_{2.5}$ concentrations with satellite grid cells at 1 km × 1 km resolution for (**a**) one day 10 September 2003, (**b**) average surface over 32 days of available AOD data used for the chronic exposure in simulations.

Table 1. Linear regression health model with chronic exposure to air pollution, fit using the true exposure, and fit using the predicted exposures from several different kriging and land use regression models.

Exposure scenario	κ	m	*Predicted exposure* R^2	*Effect estimate,* β	*Empirical SE*	*Model SE*	*Mean square error*	*95% CI coverage*
Chronic, true exposure				1.001	0.030	0.030	0.001	95.0
A1. chronic, kriging	0.5	100	0.27	1.603	0.871	0.180	1.122	31.1
A1. chronic, kriging	2	100	0.26	1.533	0.765	0.473	0.868	32.5
A1. chronic, kriging	0.5	500	0.44	1.240	0.202	0.084	0.098	35.0
A1. chronic, kriging	2	500	0.41	1.221	0.208	0.088	0.092	40.9
A2. chronic, kriging	0.5	100	0.25	1.043	0.371	0.104	0.139	43.2
A2. chronic, kriging	2	100	0.24	1.033	0.362	0.107	0.132	45.0
A2. chronic, kriging	0.5	500	0.36	0.818	0.143	0.063	0.054	28.2
A2. chronic, kriging	2	500	0.34	0.810	0.141	0.065	0.056	24.3
B1. chronic, LUR	0.5	100	0.72	1.050	0.140	0.047	0.022	47.7
B1. chronic, LUR	2	100	0.71	1.041	0.144	0.047	0.022	48.0
B1. chronic, LUR	0.5	500	0.84	1.014	0.077	0.038	0.006	67.9
B1. chronic, LUR	2	500	0.84	1.013	0.079	0.038	0.006	68.0

Abbreviation: CI, confidence interval.

Table 2. Logistic regression health model with chronic exposure to air pollution, fit using the true exposure, and fit using the predicted exposures from several different kriging and land use regression models.

Exposure scenario	κ	m	*Predicted exposure* R^2	*Odds ratio*	*Empirical SE*	*Model SE*	*Mean square error*	*95% CI coverage*
Chronic, true exposure				2.028	0.167	0.165	0.028	95.2
A1. chronic, kriging	0.5	100	0.27	2.570	0.769	0.513	0.654	89.3
A1. chronic, kriging	2	100	0.26	2.513	0.699	1.199	0.540	90.4
A1. chronic, kriging	0.5	500	0.44	2.170	0.293	0.284	0.092	94.8
A1. Chronic, Kriging	2	500	0.41	2.119	0.298	0.289	0.092	94.9
A2. chronic, kriging	0.5	100	0.25	1.815	0.397	0.314	0.167	83.1
A2. chronic, kriging	2	100	0.24	1.806	0.417	0.321	0.184	83.7
A2. Chronic, Kriging	0.5	500	0.36	1.657	0.232	0.204	0.089	75.4
A2. chronic, kriging	2	500	0.34	1.611	0.252	0.207	0.110	72.9
B1. chronic, LUR	0.5	100	0.72	2.111	0.270	0.228	0.076	91.0
B1. chronic, LUR	2	100	0.71	2.087	0.264	0.226	0.071	91.3
B1. chronic, LUR	0.5	500	0.84	2.076	0.215	0.196	0.047	93.8
B1. chronic, LUR	2	500	0.84	2.072	0.216	0.196	0.048	93.7

Abbreviation: CI, confidence interval.

Table 3. Linear regression health model with acute exposure to air pollution, fit using the true exposure, and fit using the predicted exposures from several different kriging and land use regression models.

Exposure scenario	κ	m	*Predicted exposure* R^2	*Effect estimate,* β	*Empirical SE*	*Model SE*	*Mean square error*	*95% CI coverage*
Acute, true exposure				1.000	0.006	0.006	0.000	95.2
C1. acute, kriging	0.5	100	0.91	1.020	0.016	0.013	0.001	61.4
C1. acute, kriging	2	100	0.91	1.020	0.017	0.013	0.001	61.8
C1. acute, kriging	0.5	500	0.94	1.024	0.012	0.011	0.001	41.8
C1. acute, kriging	2	500	0.93	1.023	0.013	0.011	0.001	44.3
D1. acute, LUR	0.5	100	0.31	0.451	0.068	0.023	0.307	0.0
D1. acute, LUR	2	100	0.23	0.365	0.078	0.023	0.410	0.0
D1. acute, LUR	0.5	500	0.53	0.641	0.038	0.021	0.130	0.0
D1. acute, LUR	2	500	0.48	0.604	0.035	0.022	0.158	0.0
D2. acute, LUR	0.5	100	0.96	1.081	0.034	0.010	0.008	0.8
D2. acute, LUR	2	100	0.94	1.134	0.056	0.011	0.021	0.4
D2. acute, LUR	0.5	500	0.98	1.021	0.010	0.008	0.001	25.0
D2. acute, LUR	2	500	0.98	1.038	0.011	0.008	0.002	1.4
D3. acute, LUR	0.5	100	0.97	1.025	0.016	0.009	0.001	27.8
D3. acute, LUR	2	100	0.97	1.026	0.017	0.009	0.001	25.6
D3. acute, LUR	0.5	500	0.98	1.016	0.009	0.008	0.000	46.6
D3. acute, LUR	2	500	0.98	1.016	0.010	0.008	0.000	42.1

Abbreviation: CI, confidence interval.

Table 4. Logistic regression health model with acute exposure to air pollution, fit using the true exposure, and fit using the predicted exposures from several different kriging and land use regression models.

Exposure scenario	κ	m	*Predicted exposure* R^2	*Odds ratio*	*Empirical SE*	*Model SE*	*Mean square error*	*95% CI coverage*
Acute, true exposure				2.000	0.048	0.050	0.002	96.2
C1. acute, kriging	0.5	100	0.91	1.886	0.045	0.045	0.005	71.4
C1. acute, kriging	2	100	0.91	1.890	0.048	0.046	0.005	69.8
C1. acute, kriging	0.5	500	0.94	1.926	0.047	0.047	0.004	83.8
C1. acute, kriging	2	500	0.93	1.924	0.047	0.046	0.004	83.8
D1. acute, LUR	0.5	100	0.31	1.261	0.041	0.020	0.214	0.0
D1. acute, LUR	2	100	0.23	1.204	0.042	0.018	0.260	0.0
D1. acute, LUR	0.5	500	0.53	1.428	0.037	0.027	0.115	0.0
D1. acute, LUR	2	500	0.48	1.390	0.032	0.025	0.133	0.0
D2. acute, LUR	0.5	100	0.96	2.078	0.057	0.053	0.005	90.2
D2. acute, LUR	2	100	0.94	2.165	0.073	0.056	0.012	72.4
D2. acute, LUR	0.5	500	0.98	2.008	0.050	0.050	0.002	95.0
D2. acute, LUR	2	500	0.98	2.044	0.051	0.052	0.003	94.4
D3. acute, LUR	0.5	100	0.97	1.994	0.050	0.050	0.003	94.2
D3. acute, LUR	2	100	0.97	1.994	0.050	0.050	0.003	94.2
D3. acute, LUR	0.5	500	0.98	2.001	0.050	0.050	0.002	95.2
D3. acute, LUR	2	500	0.98	2.012	0.050	0.050	0.002	95.2

Abbreviation: CI, confidence interval.

In the D2 land use regression model in the acute scenario, we found that the model that excluded the temporal covariates and included only the roadway covariates reversed the direction of bias, showing a level of 1–13% upward bias. Interestingly, although the exposure R^2 is high (0.94–0.98), the spatial variability is not explained well; this yields upward bias in the acute health effect estimates similar to the upward bias seen in the B1 chronic land use regression models. In the D3 acute scenario with the two-stage land use regression model, we found that the bias was negligible, up to 3% at most, although the undercoverage of the 95% CIs in the linear model was still severe.

The results of the supplementary simulation analyses are given in the Supplementary Material. The performance of all models under the null showed very little inflation of type-I error rates, at most 6% across all simulations. The results for the simulated surfaces with greater proportions of non-spatial Berkson error showed that for the linear health model, simulated chronic exposures can have a wide range of exposure R^2 from 0.43 to 0.87 and still be unbiased. This demonstrates the separate issues of total variability explained and exposure model misspecification. Finally, the results of a simulated chronic surface fit with a misspecified kriging model shows 4–39% upward bias.

DISCUSSION

In this study, we found that there may be substantial bias of health effect estimates in models using exposures predicted by kriging or land use regression. We found that the direction of bias may be either toward or away from the null, and the degree of bias varies by the type of exposure model and the study design, with some exposure predictions working well in certain situations. We also found substantial undercoverage where the true effect was often not included in the naive 95% CI. We gained these insights into the spatial variability of $PM_{2.5}$ predictions by using high-resolution satellite data on AOD, which were calibrated to reflect $PM_{2.5}$ concentrations.

In the chronic simulations where exposure variation was purely spatial, kriging alone on the average surface was insufficient to model and predict exposures and resulted in unacceptable bias. The chronic models with daily kriging worked better in the linear health model than the logistic health model. This highlights the difference between the effects of measurement error in a linear model *versus* a logistic model, which is a result of the parameterization of the mean and variance.[6] The improved performance of the chronic exposure model with land use terms may be related to the exposure R^2 of the prediction model, which includes covariates used in the calibration of AOD. We also observed that the predictions from the chronic kriging model had the smallest variability, whereas predictions from the two-stage model varied more, and consequently better reflected the variability of the true exposures (see Supplementary Figure 1). Hence, the shrinking of the exposure distribution in the chronic kriging predictions may partly explain its poor performance compared with the other predictions.

In the acute setting, the model incorporating spatial and temporal covariates performed very poorly; the addition of the temporal covariates, which could not correctly model the complex underlying temporal process, resulted in an exposure model that explained very little exposure variability and yielded substantial bias in the health effect parameter. The other acute exposure models performed better in terms of both the exposure R^2 and the health effect estimates. However, there were still notable differences in bias and coverage, despite the high exposure R^2 for those models.

Overall, our study shows that the exposure R^2 is certainly a helpful tool in assessing model performance, and models with poor exposure R^2 tend to yield the worst biases, yet even a high R^2 does not guarantee that the health parameters will be unbiased. This is because the degree of exposure model misspecification depends on how much of the *true spatial variability* is explained by the model, but the proportions of spatial variability and non-spatial Berkson error (seen in the 'nugget' of spatial models) is always unknown for real exposures. Our supplementary simulations with different proportions of Berkson error also demonstrate this phenomenon. This observation that parameter bias does not directly depend on exposure R^2 is consistent with a recent brief report suggesting that predicted exposures with higher R^2 in the exposure model may not always improve the quality of health effect estimates.[19] Moreover, a high R^2 does not guarantee good coverage for the resulting CIs.

Other factors of the exposure model such as the covariance model chosen did not have a strong effect on overall performance, as evidenced by similar results in each setting across varying κ. Hence, the choice of spatial covariance model may not have as strong a role in the effectiveness of using exposure predictions in health effect analyses as the choices concerning how the spatial and temporal variation is accommodated in the mean model.

Other statistical studies assessing performance of kriging and land use regression models have not examined the performance under real-world pollution fields. In the current literature, studies using simulated exposure surfaces have found that use of exposure predictions in health effect models often induces little to no bias.[7–10] However, Madsen *et al.*[9] and Szpiro *et al.*[10] assume smooth exposure surfaces that can be fit well using kriging methods, finding no need for bias correction. Our study found that in some cases model misspecification in spatial exposure models can lead to severe biases. The issue of model misspecification has not been a focus of previous statistical research in the area of measurement error in air pollution epidemiology. A recent study on the effects of measurement error in land use regression finds that realistic land use regression scenarios can result in severe attenuation of the health effect parameter in a linear regression model.[20] Our findings are consistent with these results, demonstrating the importance of the choice of statistical exposure model. Additional innovations of our study are the use of calibrated high-resolution satellite AOD measurements and the inclusion of acute and chronic exposure scenarios with both linear and logistic health regression models. Another recent study characterizes the complex form of measurement error induced by two-stage modeling approaches and proposes a correction approach that can be used when the exposure has a misspecified mean model.[21] This type of method that can correct for model misspecification could be particularly beneficial to correct the cases of severe model misspecification seen in this paper.

Limitations

Any simulation study will need to focus on a finite set of well-defined simulation scenarios. Thus, it is not possible to represent every scenario one might envision. However, we have attempted to provide a range of simulations with varying degrees of temporal and spatial variability.

There are many other potential sources of measurement error in air pollution epidemiology studies not considered here. Zeger *et al.*[22] provides a framework for considering a number of sources of exposure measurement error in air pollution research. We also assumed no confounding to isolate problems stemming from the measurement error in exposure modeling. The combination of misspecified exposure models and incomplete control for confounding variables may introduce different problems and is not yet known.

The days in which we have the most complete coverage of AOD retrievals represent days with clear-sky conditions and limited snow coverage. Hence, these days are not a representative sample of all days throughout the year. Other days that are under partly cloudy conditions may have a different spatial distribution of AOD and of $PM_{2.5}$.

This study does not suggest that satellite-calibrated AOD measurements are a perfect measure of true $PM_{2.5}$ exposure. It is difficult to evaluate how well such measurements reflect true spatial variation in $PM_{2.5}$ exposures without considerably more spatial coverage of air pollution monitoring data. There remains no gold standard for the entire fine-scale spatial distribution of PM throughout a region. This study can lend insight into potential performance of kriging, land use regression, and spatiotemporal modeling by using a more realistic representation of a regional $PM_{2.5}$ surface; however, it is not generalizable to all possible true air pollution surfaces. Rather, these simulations serve as examples of potential scenarios in which kriging and land use regression may perform better or worse.

CONCLUSIONS

This simulation study uses high-resolution satellite data to provide several settings with realistic exposure surfaces and suggests that (i) kriging and land use regression models sometimes work well in health effect models but sometimes introduce substantial biases, (ii) the success in using modeled exposures varies by the spatial and temporal properties of the underlying data and the exposure model chosen, and (iii) future statistical research is needed to understand the implications of misspecifying exposure models, to provide appropriate diagnostic procedures, and to implement effective measurement error correction strategies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We greatly appreciate A. Lyapustin (NASA Goddard Space Flight Center, Baltimore, Maryland, USA) and Y. Wang (University of Maryland, Baltimore) for their work in providing the MAIAC data set for 2003. This work was supported by US EPA grant 834798 and NIH grants ES007142, ES016454, ES020871, and ES000002. This publication's contents are solely the responsibility of the grantee and do not necessarily represent the official views of the US EPA. Further, US EPA does not endorse the purchase of any commercial products or services mentioned in the publication.

REFERENCES

1 Brook RD, Rajagopalan S, Pope 3rd CA, Brook JR, Bhatnagar A, Diez-Roux AV *et al.* Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation* 2010; **121**: 2331–2378.
2 Brauer M, Hoek G, van Vliet P, Meliefste K, Fischer P, Gehring U *et al.* Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiology* 2003; **14**: 228–239.
3 Clougherty JE, Wright RJ, Baxter LK, Levy JI. Land use regression modeling of intra-urban residential variability in multiple traffic-related air pollutants. *Environ Health* 2008; **7**: 17.
4 Liao D, Peuquet DJ, Duan Y, Whitsel EA, Dou J, Smith RL *et al.* GIS approaches for the estimation of residential-level ambient PM concentrations. *Environ Health Persp* 2006; **114**: 1374–1380.
5 Jerrett M, Burnett RT, Ma R, Pope 3rd CA, Krewski D, Newbold KB *et al.* Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 2005; **16**: 727–736.
6 Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu C. *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall/CRC, 2006.
7 Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics* 2009; **10**: 258–274.
8 Lopiano KK, Young LJ, Gotway CA. A comparison of errors in variables methods for use in regression models with spatially misaligned data. *Stat Methods Med Res* 2011; **20**: 29–47.
9 Madsen L, Ruppert D, Altman NS. Regression with spatially misaligned data. *Environmetrics* 2008; **19**: 453–467.
10 Szpiro AA, Sheppard L, Lumley T. Efficient measurement error correction with spatially misaligned data. *Biostatistics* 2011; **12**: 610–623.
11 Remer LA, Kaufman YJ, Tanre D, Mattoo S, Chu DA, Martins JV *et al.* The MODIS aerosol algorithm, products, and validation. *J Atmos Sci* 2005; **62**: 947–973.
12 Kloog I, Koutrakis P, Coull BA, Lee HJ, Schwartz J. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos Environ* 2011; **45**: 6267–6275.
13 Kloog I, Nordio F, Coull BA, Schwartz J. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM(2.5) exposures in the Mid-Atlantic states. *Enviro Sci Technol* 2012; **46**: 11913–11921.
14 Lyapustin A, Martonchik J, Wang YJ, Laszlo I, Korkin S. Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. *J Geophys Res Atmos* 2011; **116**: D03210.
15 Lyapustin A, Wang Y, Laszlo I, Kahn R, Korkin S, Remer L *et al.* Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *J Geophys Res Atmos* 2011; **116**: D03211.
16 Lyapustin A, Wang Y, Frey R. An automatic cloud mask algorithm based on time series of MODIS measurements. *J Geophys Res Atmos* 2008; **113**: D16207.
17 Chudnovsky A, Tang C, Lyapustin A, Wang Y, Schwartz J, Koutrakis P. A critical assessment of high resolution aerosol optical depth (AOD) retrievals for fine

particulate matter (PM) predictions. *Atmos Chem Phys Discuss* 2013; **13**: 14581–14611.

18 Kloog I, Melly SJ, Ridgway WL, Coull BA, Schwartz J. Using new satellite based exposure methods to study the association between pregnancy PM(2).(5) exposure, premature birth and birth weight in Massachusetts. *Environ Health* 2012; **11**: 40.

19 Szpiro AA, Paciorek CJ, Sheppard L. Does more accurate exposure prediction necessarily improve health effect estimates? *Epidemiology* 2011; **22**: 680–685.

20 Basagana X, Aguilera I, Rivera M, Agis D, Foraster M, Marrugat J *et al.* Measurement error in epidemiologic studies of air pollution based on land-use regression models. *Am J Epidemiol* 2013; **178**: 1342–1346.

21 Szpiro AA, Paciorek CJ. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics* 2013; **24**: 501–517.

22 Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D *et al.* Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ Health Persp* 2000; **108**: 419–426.

Supplementary Information accompanies the paper on the Journal of Exposure Science and Environmental Epidemiology website (http://www.nature.com/jes)