

Health Effects of Long-Term Air Pollution: Influence of Exposure Prediction Methods

Author(s): Sun-Young Kim, Lianne Sheppard and Ho Kim

Source: *Epidemiology*, May, 2009, Vol. 20, No. 3 (May, 2009), pp. 442-450

Published by: Lippincott Williams & Wilkins

Stable URL: <https://www.jstor.org/stable/20485735>

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/20485735?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



Lippincott Williams & Wilkins is collaborating with JSTOR to digitize, preserve and extend access to *Epidemiology*

JSTOR

Health Effects of Long-term Air Pollution

Influence of Exposure Prediction Methods

Sun-Young Kim,^a Lianne Sheppard,^{a,b} and Ho Kim^c

Background: Air pollution studies increasingly estimate individual-level exposures from area-based measurements by using exposure prediction methods such as nearest-monitor and kriging predictions. However, little is known about the properties of these methods for health effects estimation. This simulation study explores how 2 common prediction approaches for fine particulate matter (PM_{2.5}) affect relative risk estimates for cardiovascular events in a single geographic area.

Methods: We estimated 2 sets of parameters to define correlation structures from 2002 data on PM_{2.5} in the Los Angeles area, and selected additional parameters to evaluate various correlation features. For each structure, annual average PM_{2.5} was generated at 22 monitoring sites and 2000 preselected individual locations in Los Angeles. Associated survival time until cardiovascular event was simulated for 10,000 hypothetical subjects. Using PM_{2.5} generated at monitoring sites, we predicted PM_{2.5} at subject locations by nearest-monitor and kriging interpolation. Finally, we estimated relative risks of the effect of PM_{2.5} on time to cardiovascular event.

Results: Health effect estimates for cardiovascular events had higher or similar coverage probability for kriging compared with nearest-monitor exposures. The lower mean square error of nearest monitor prediction resulted from more precise but biased health effect estimates. The difference between these approaches dramatically moderated when spatial correlation increased and geographic characteristics were included in the mean model.


Conclusions: When the underlying exposure distribution has a large amount of spatial dependence, both kriging and nearest-monitor predictions gave good health effect estimates. For exposure with little spatial dependence, kriging exposure was preferable but gave very uncertain estimates.

(*Epidemiology* 2009;20: 442–450)

Submitted 28 December 2007; accepted 15 September 2008; posted 13 March 2009.

From the Departments of ^aEnvironmental and Occupational Health Sciences, and ^bBiostatistics, University of Washington, Seattle, WA; and ^cDepartment of Biostatistics and Epidemiology, Seoul National University, Seoul, Korea.

Supported by the Korean Government (MOEHRD) Korean Research Foundation Grant (KRF-2005-214-E00041).

 Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Sun-Young Kim, Department of Environmental and Occupational Health Sciences, University of Washington, 4225 Roosevelt Way NE, Suite 100, Seattle, WA 98105. E-mail: puha0@u.washington.edu.

Copyright © 2009 by Lippincott Williams & Wilkins
ISSN: 1044-3983/09/2003-0442
DOI: 10.1097/EDE.0b013e31819e4331

Studies of health effects of long-term air pollution have typically assigned exposures to individuals using measurements made at a few fixed-site monitoring stations.^{1,2} This approach was based on the assumption that the spatially averaged ambient air pollution concentrations measured at monitoring sites in an administrative area (such as a community or a metropolitan area) is representative of the exposure of individuals residing at various locations in the area. However, such exposure data do not capture all the individual spatial heterogeneity in exposure and may result in less accurate or reliable health effect estimates.

To improve exposure assessment, recent studies have predicted individual air pollution exposure instead of using an average area-wide monitored concentration. One prediction approach assigns exposure based on the monitor nearest to the participant's residential location.^{3–6} Another approach is to interpolate predictions by applying a geostatistical method such as kriging.^{7–10} Kriging predicts individual concentrations corresponding to residential locations after estimating the parameters of a model of the spatial surface of air pollution concentration. Although both methods have been used, little is known about how these prediction methods might influence health effect estimation. We conducted a simulation study to investigate the impact on the health effect estimate of predicted exposure from these 2 exposure prediction approaches, using assumptions derived from an analysis of annual average concentration of particulate matter less than or equal to 2.5 μm in aerodynamic diameter (PM_{2.5}) and a previous analysis of the effect of particulate matter on cardiovascular events.⁴

METHODS

Data and Assumptions

Population

We selected 2000 residential locations from the Los Angeles (LA)-Long Beach-Santa-Ana urbanized area of California (LA urbanized area) as defined by the US census classification (eFigure, <http://links.lww.com/A980>). Locations were randomly selected using a uniform distribution stratified by census tract to obtain up to 3 residential locations per tract, with the number of individual sites proportional to the population size. To obtain 10,000 subjects, 5 subjects

were assumed to live around each residential site and to have identical PM_{2.5} exposure.

PM_{2.5} Data and Analysis

The 22 PM_{2.5} monitoring stations used in the simulation were located in counties (Los Angeles, Orange, Riverside, San Bernardino, and Ventura) surrounding the LA urbanized area. These stations were nearly identical to those used by Jerrett et al.⁸ Using annual average concentrations from 2002, we estimated parameters for a spatial model of PM_{2.5} by fitting mean and covariance models. We fit a covariance model to the empirical semi-variogram, half the squared differences of PM_{2.5} against distance. The covariance model has a specific functional form with 3 parameters: the range, partial sill, and nugget. The range is the distance at which the semi-variogram curve reaches the maximum variability. The total variance, named sill, is partitioned into the partial sill and nugget. The partial sill is the component of variance due to spatial variability, whereas the nugget is considered measurement error. The mean model is a regression model with covariates assumed to affect PM_{2.5}. In this analysis, we considered only longitude (X) and latitude (Y) as possible covariates. After exploration of several covariance and mean models, the best-fitting model, chosen by cross-validation, was a spherical covariance model with a second-order polynomial mean model, using the XY coordinates as predictors. This model is referred to as TEM 1 in the tables and figures. It is similar to the one used in previous studies in the LA area.^{8,9} For sensitivity analysis, we also estimated spatial parameters from a constant mean model (TEM 4).

Disease

We used an exponential relative risk (RR) model for time to first cardiovascular event and used the relative risk, baseline rate, and drop-out patterns consistent with a previous study of postmenopausal women in the Women’s Health Initiative cohort.⁴ Miller et al⁴ reported 2113 cardiovascular and cerebrovascular events over a median follow-up of 6 years among 65,893 women and estimated overall hazard

ratio of 1.24 for a 10 µg/m³ increase in long-term average PM_{2.5} concentration. In the simulation, we assumed a baseline incidence rate of 0.032 per year and a β coefficient of 0.0215 (=log(1.24)/10) for the relative risk parameter.

Simulation

Given the residence locations and model parameters, we generated true long-term PM_{2.5} exposures for 2000 residences and 22 monitor locations, and 10,000 outcomes (times to cardiovascular event) in 1000 simulations. We made the simplifying assumption that PM_{2.5} concentration equals PM_{2.5} exposure. Using only monitoring data, we then predicted PM_{2.5} exposure by 2 methods: nearest monitor and kriging. The effect of PM_{2.5} on cardiovascular event incidence was estimated conditional on predicted or true exposure in a Cox proportional hazards regression analysis. Across all simulations, locations were fixed for the 22 monitors (real locations) and 2000 residences; the true exposures and health outcomes were generated in every simulation.

True Models

Underlying Exposure Model

We assumed that exposure to PM_{2.5} followed a multivariate normal distribution with variability consisting of spatial and nonspatial components parameterized by the partial sill (σ²), range (ϕ), and nugget (τ²) in the spherical covariance model. True exposure (W) at location s_i was generated as

W(s_i) = μ(s_i) + v(s_i) + ε(s_i) (1)

The true exposure was composed of the mean (μ(s_i)) and 2 error terms, the spatial (v(s_i)) and nonspatial errors (ε(s_i)). Variances of these errors are the partial sill (σ²) and nugget (τ²), respectively. Correlation of spatial errors is a function of the range (ϕ).¹¹ To represent different spatial correlation scenarios, we focused primarily on 6 true exposure models with different mean and covariance parameters (Table 1). Parameters for 2 models were estimated from analyses of PM_{2.5} data in LA (section PM_{2.5} Data and Analysis). In the

TABLE 1. Source of Spatial Variability and Parameter Value in the 6 True Exposure Models

| True Exposure Model | Parameter Value | | | |
|-------------------------------|--------------------------|--------------------------------|----------------|---|
| | Nugget (τ ²) | Partial Sill (σ ²) | Range (km) (ϕ) | Mean (µg/m ³) (µ[s _i]) |
| 1 Geographic characteristics | 0.43 | 5.81 | 44 | −46771 + 6.69X + 24.07Y −0.09X ² − 0.31Y ² − 0.0016XY ^a |
| 2 No spatial correlation | 46.9 | 0 | 0 | 18.35 |
| 3 Low spatial correlation | 0 | 46.9 | 10 | 18.35 |
| 4 Medium spatial correlation | 0 | 46.9 | 120 | 18.35 |
| 5 High spatial correlation | 0 | 46.9 | 277 | 18.35 |
| 6 Highest spatial correlation | 0 | 46.9 | 500 | 18.35 |

^a“X” and “Y” indicate X and Y coordinates, respectively. Original latitude and longitude coordinates were converted into kilometers in the Universal Transverse Mercator coordinate system.

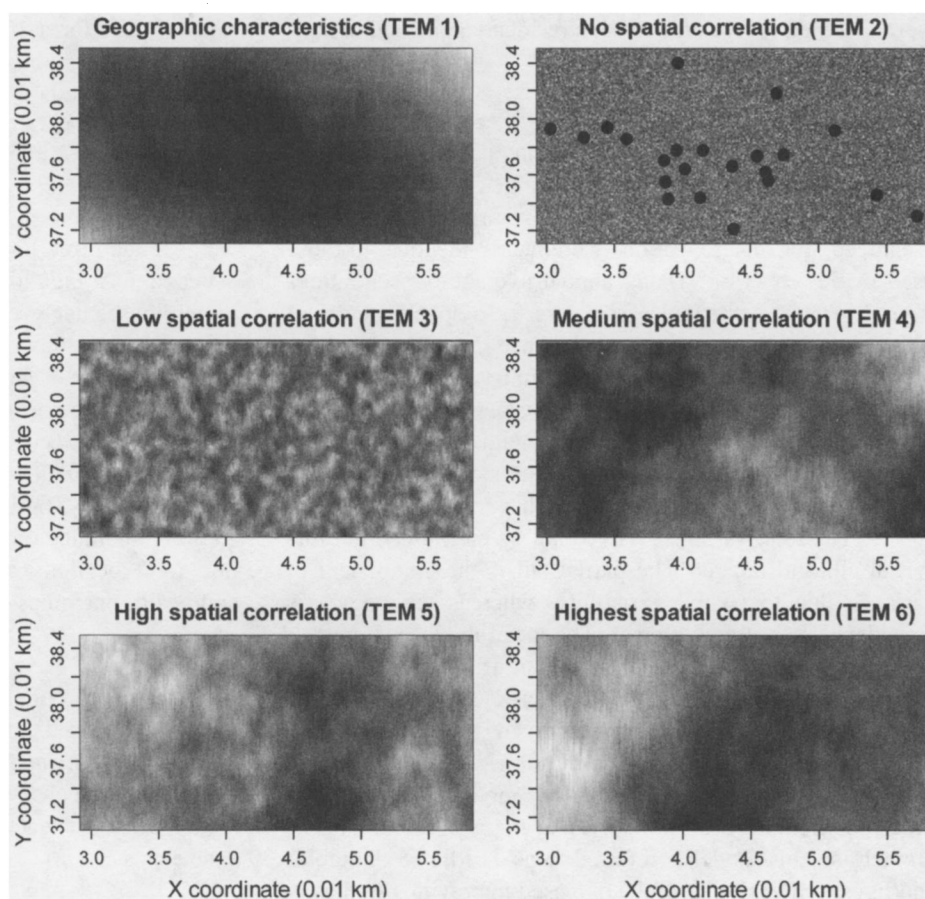


FIGURE 1. Spatial surface of the 6 true exposure models. Darker shading indicates higher concentration. Black dots on the spatial surface for the TEM 2 represent the locations of the 22 PM_{2.5} monitoring sites in LA.

first true exposure model (TEM 1), most of the variability of PM_{2.5} was dominated by the geographic covariates. The mean was determined by a second-order polynomial function of XY coordinates, $(\mu(s_i) = (-46771) + (6.69 * X) + (24.07 * Y) + (-0.09 * X^2) + (-0.31 * Y^2) + (-0.0016 * X * Y))$, while the partial sill, range, and nugget were relatively small ($\sigma^2 = 5.81$, $\phi = 44$ km, $\tau^2 = 0.43$). The other model (TEM 4) had a constant mean ($\mu(s_i) = 18.35 \mu\text{g}/\text{m}^3$) so all the spatial dependence was parameterized in the covariance model. Thus, the partial sill and range were larger ($\sigma^2 = 46.9$, $\phi = 120$ km) and there was no nugget ($\tau^2 = 0$). To explore sensitivity to the exposure model structure, arbitrary parameter values were chosen for 4 other models (TEM 2, TEM 3, TEM 5, and TEM 6). These models had the same constant mean and partial sill as the TEM 4 but different range parameters to represent a range of spatial correlations from none to high. Figure 1 shows examples of each exposure surface; the spatial surface is smoother in the first, fourth, fifth, and sixth models with large spatial dependence from the variogram or the second-order mean models. To better understand how modeled exposure affected health effect estimates, we conducted a sensitivity analysis by simulating an additional set of 40 true exposure models. We assigned a constant mean to half and a polynomial mean to the other half

(mean identical to the first true exposure model). Each of the 20 pairs had range parameters (ϕ) between 10 and 500 km, partial sill parameters (σ^2) between 0.1 and 90, and no nugget ($\tau^2 = 0$).

Underlying Disease Model

Survival time to cardiovascular event occurrence (T) followed an exponential distribution with mean,

$$\gamma = \frac{1}{\lambda_0 \exp(\beta(W - \bar{W}))}, \text{ ie,}$$

$$T \sim \text{Exp}(\gamma) \quad (2)$$

Censoring was based on 2 mechanisms: cardiovascular disease-free survival for 10 years (study duration) and loss to follow-up (drop-out). Ten percent of all subjects dropped out in the first 10 years with time determined by the uniform distribution so the probability of drop-out was constant over all study years.

Fitted Models

Predicted Exposure Model

For each of the 6 exposure models, 2 versions of predicted PM_{2.5} were estimated from the simulated ("mea-

sured") PM_{2.5} concentration at the 22 monitor locations. Nearest-monitor exposures (W^*) assigned the measured PM_{2.5} at the closest monitor to each individual residential location. Kriged exposures (W^{**}) interpolated measured PM_{2.5} by universal kriging methods. Kriging prediction consists of 2 parts: estimating the spatial structure and interpolating the spatial surface. We estimated parameters from the monitor measurements under a semi-variogram with a spherical covariance and a second-order polynomial mean model. Given the estimated parameters, we predicted PM_{2.5} at individual residence locations.

Fitted Disease Model

The effect of PM_{2.5} on time to cardiovascular event was estimated by the Cox proportional hazards model, conditional on true ($W' = W$) or predicted exposure ($W' = W^*$ or W^{**}),

$$\lambda(t) = \lambda_0 \exp(\hat{\beta}(W' - \bar{W}')) \quad (3)$$

Assessment of Validity and Reliability

To evaluate validity and reliability of health effect estimates, bias, variance, mean square error, and 95% coverage probability were estimated. Coverage probability is the proportion of 95% confidence intervals of health effect estimates that include the true β .

To gain insight into properties of health effect estimates based on properties of exposures and their prediction, we also investigated validity and reliability of predicted exposure by summarizing average prediction error (difference of predicted PM_{2.5} minus true PM_{2.5}), variance of true and predicted exposures at subject locations, mean square prediction error, and 1 minus mean square prediction error relative to true exposure variance (constrained to always be greater than or equal to 0), which is an estimate of R^2 .

RESULTS

Summary statistics of true and predicted long-term PM_{2.5} exposure in the first simulation are shown in the eTable (<http://links.lww.com/A980>). Across the exposure models, the true PM_{2.5} exposure means for the 2000 residential sites were between 17.1 and 26.1 $\mu\text{g}/\text{m}^3$. Predicted PM_{2.5} had smaller standard deviations and ranges than true PM_{2.5}, particularly for kriging prediction. Relative to the variation in the true exposure data, kriged PM_{2.5} varied much less in the second and third true exposure models (TEM 2 and TEM 3), which assumed little spatial correlation. A more detailed picture emerges from the true versus predicted exposure scatter plots displayed in Figure 2. Predicted PM_{2.5} was more correlated with true PM_{2.5} when the true exposure models had

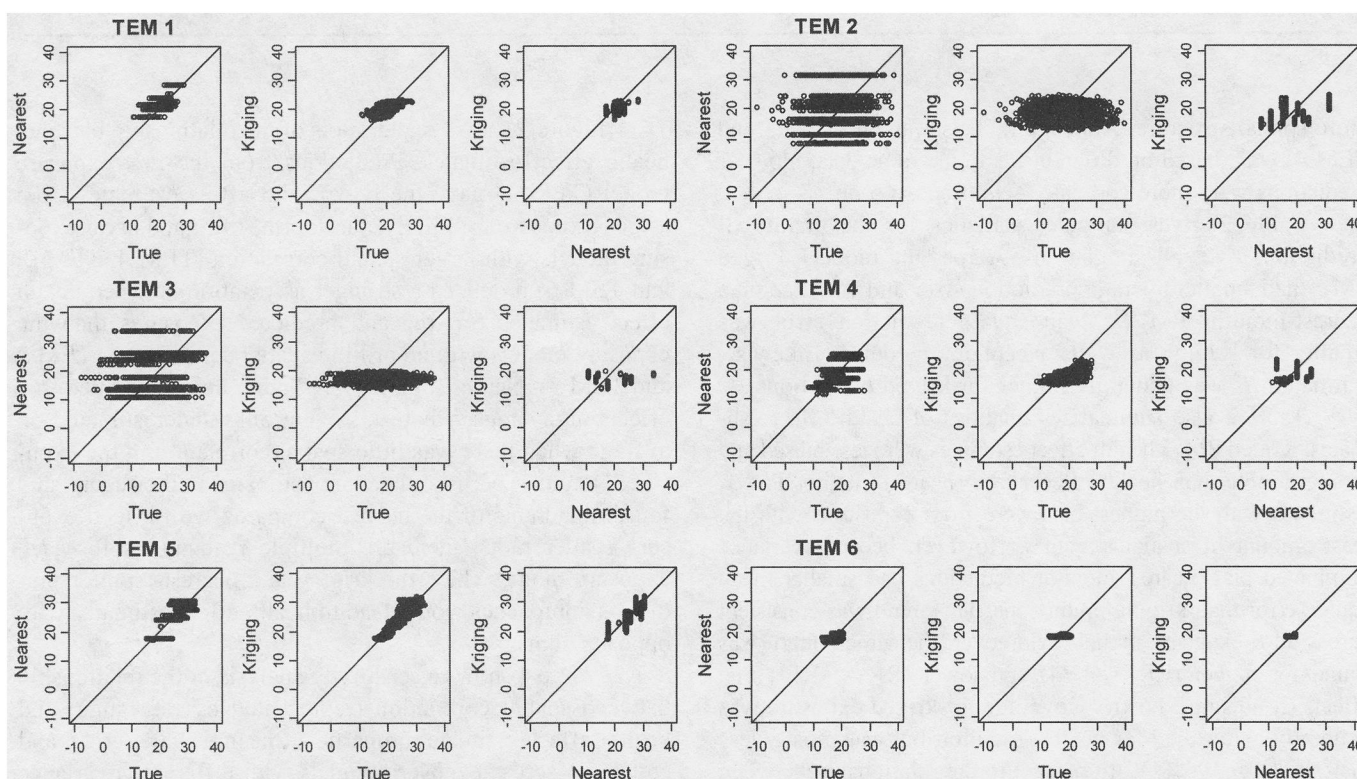


FIGURE 2. Relationship between true and predicted exposures for the 6 true exposure models in which sources of spatial variability come from geographic characteristics (TEM 1), little spatial correlation (TEM 2 and 3), medium spatial correlation (TEM 4), and high spatial correlation (TEM 5 and 6). Solid line is the X–Y identity.

TABLE 2. Summary Estimates of Validity and Reliability of Predicted Exposure and Health Effect Estimate by the 6 True Exposure Models Across 1000 Simulations

| Exposure Model | | Exposure ^a | | | | Health Effect ^b | | | |
|----------------|---------|-----------------------|------------------|---------------------|---|----------------------------|----------|------|-----------------|
| True | Fitted | MAPE ^c | RMV ^d | RMMSPE ^e | Mean of Estimated R ² ^f | Bias ² | Variance | MSE | CP ^g |
| TEM 1 | True | | 2.93 | | | 0 | 47 | 48 | 0.97 |
| | Nearest | 0.43 | 2.59 | 2.44 | 0.29 | 36 | 93 | 129 | 0.83 |
| | Kriging | 0.00 | 2.29 | 2.07 | 0.47 | 0 | 136 | 136 | 0.93 |
| TEM 2 | True | | 6.84 | | | 0 | 8 | 8 | 0.95 |
| | Nearest | 0.06 | 6.31 | 9.67 | 0 | 463 | 13 | 476 | 0.00 |
| | Kriging | 0.04 | 2.07 | 7.45 | 0 | 472 | 291 | 763 | 0.51 |
| TEM 3 | True | | 6.79 | | | 0 | 8 | 8 | 0.96 |
| | Nearest | 0.06 | 6.44 | 8.92 | 0 | 326 | 21 | 347 | 0.02 |
| | Kriging | 0.03 | 2.01 | 7.26 | 0 | 339 | 686 | 1025 | 0.54 |
| TEM 4 | True | | 4.25 | | | 0 | 27 | 27 | 0.95 |
| | Nearest | −0.01 | 4.03 | 3.22 | 0.35 | 34 | 54 | 88 | 0.75 |
| | Kriging | −0.02 | 2.42 | 3.91 | 0.20 | 0 | 727 | 728 | 0.71 |
| TEM 5 | True | | 2.88 | | | 0 | 59 | 59 | 0.96 |
| | Nearest | 0.01 | 2.74 | 2.13 | 0.36 | 30 | 108 | 139 | 0.85 |
| | Kriging | 0.01 | 2.09 | 2.15 | 0.40 | 1 | 400 | 401 | 0.87 |
| TEM 6 | True | | 2.16 | | | 0 | 108 | 108 | 0.94 |
| | Nearest | −0.01 | 2.06 | 1.58 | 0.37 | 32 | 143 | 175 | 0.92 |
| | Kriging | −0.04 | 1.71 | 1.44 | 0.47 | 0 | 332 | 332 | 0.95 |

^aSummary of 1000 realizations (average prediction error, variance, mean square prediction error, and estimated R²) for 2000 exposures in each simulation.
^bBias, variance, and mean square error (MSE) for health effect estimates in 1000 simulations were multiplied by 10⁶.
^cMean of average prediction error for the difference of predicted PM_{2.5} − true PM_{2.5}.
^dRoot mean of variance.
^eRoot mean of mean square prediction error.
^fMean of 1 − mean square prediction error divided by variance of true exposure. Estimates were constrained to have a minimum of 0.
^gCoverage probability of health effect estimate.

more spatial structure. Although models 1 and 4 (TEM 1 and TEM 4) were based on different fits to the same data, model 1 predictions were more correlated with true exposures.

Table 2 gives summary statistics for the quality of predictions over all simulations. Across the models, kriged PM_{2.5} had smaller average prediction error and variance than nearest-monitor PM_{2.5}. Mean square prediction error was smaller for kriging as well, except for model 4. Likewise, estimated R² was generally higher for kriged predictions.

Table 2 also summarizes qualities of health effect estimates. Kriged PM_{2.5} health effect estimates were less biased and more variable than health effects from nearest-monitor PM_{2.5}. Estimates with the highest bias were from exposures with the least amount of spatial structure. However, because variance dominated bias, nearest-monitor predictions had smaller mean square error than kriging predictions. This trend was consistent across all 6 exposure models. Moreover, the same pattern was found for higher (RR = 1.34) and lower (RR = 1.14) true effects (results not shown). Coverage for kriged exposure was generally better than for nearest monitor, but was mostly less than the target 95%. With respect to the relationship between predicted exposure and health effect estimates, spatial exposure structures with higher estimated R² produced higher coverage of health effect estimates.

Figure 3 shows scatter plots of the relationship between health effect estimates. Predictions from the first exposure model (TEM 1) gave the best health effect estimates from kriged exposure. Similarly, the fourth, fifth, and sixth exposure models with larger spatial correlation (TEM 4, TEM 5, and TEM 6) produced a stronger association between health effect estimates for true and predicted exposures than the exposure models with no or little spatial dependence (TEM 2 and TEM 3). Nearest-monitor prediction underestimated true effects on average, whereas kriging only underestimated on average when there was little spatial correlation in the exposure. Note that because there are estimates in the second and fourth quadrants of the nearest compared with kriged exposure scatter plots, there are multiple realizations for each exposure model where the kriging and nearest-monitor prediction approaches would lead to health effect estimates with opposite signs.

Figure 4 shows a sensitivity analysis of the relationship between spatial correlation (represented by the range) and health effect estimate properties (mean square error and coverage) across exposure models with different covariance parameters and mean structure. The models that produced the best health effect estimates (as defined by good coverage and small mean square error) had large range parameters relative

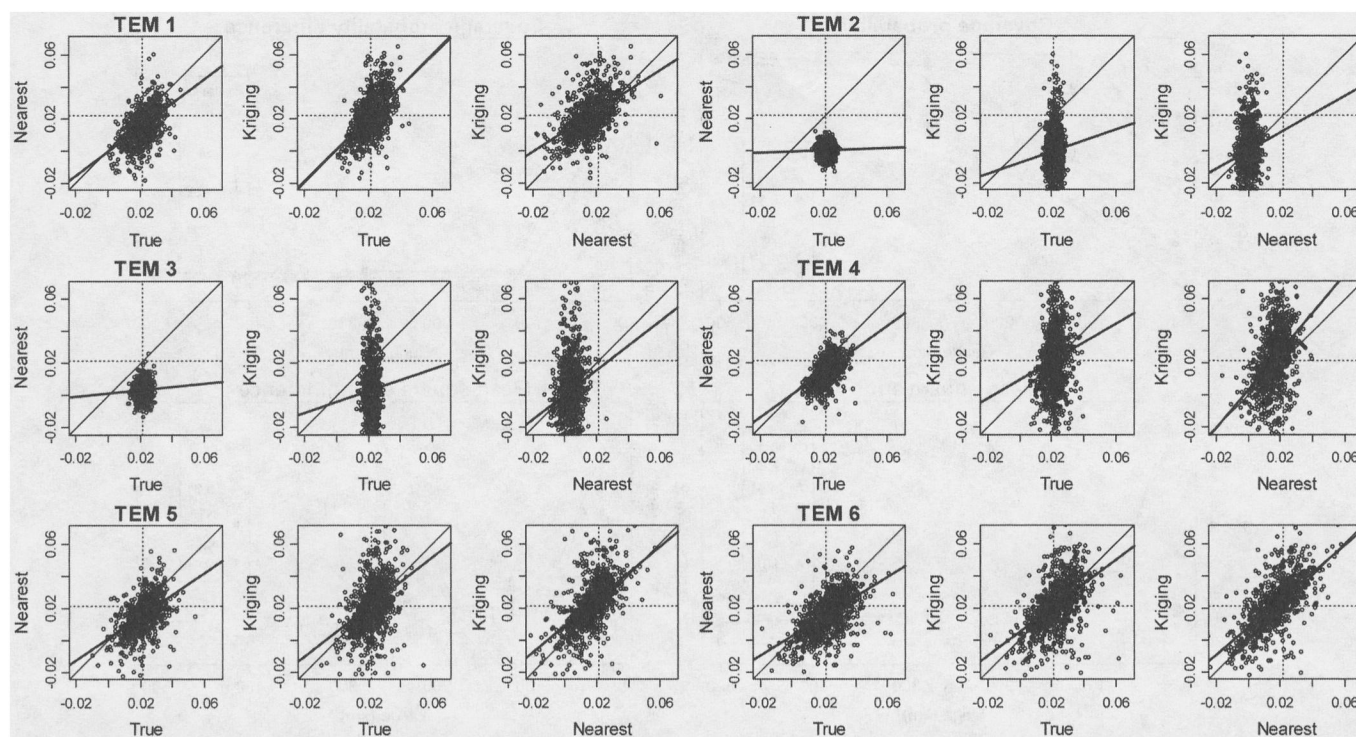


FIGURE 3. Relationship between estimated health effects from true and predicted exposures for the 6 true exposure models in which sources of spatial variability come from geographic characteristics (TEM 1), little spatial correlation (TEM 2 and 3), medium spatial correlation (TEM 4), and high spatial correlation (TEM 5 and 6) after excluding outlying estimates. Thinner and thicker solid lines display the X-Y identity line and the best-fit line between health effect estimates of true and predicted exposures, respectively. Dotted horizontal and vertical lines present the true effect size (0.0215).

to the partial sill. As the range increased, coverage increased toward 95% and mean square error decreased. Moreover, the difference between prediction approaches for both coverage and mean square error became smaller. Addition of mean structure had a dramatic effect on mean square error but relatively a small impact on coverage. In particular, with mean structure in the model there was only a small difference in mean square error between kriging and nearest-monitor exposures for all range parameters.

DISCUSSION

We explored 2 exposure prediction approaches in a simulation study to estimate the effect of long-term PM_{2.5} exposure on cardiovascular events. Kriging prediction estimated more accurate exposure because it had smaller average mean square prediction error. However, kriged exposure predictions were also less variable, which has implications for health effect estimates. Overall, health effect estimates were better with kriged exposure when comparing results based on coverage probability. In contrast, for mean square error kriging generally performed worse than nearest-monitor prediction due to greater variance in the health effect estimates resulting from smaller exposure variance. Differences between the 2 approaches diminished in more spatially de-

pendent exposure structures where the dependence is represented by a longer range or the polynomial mean model.

Health effect estimates for both exposure prediction approaches had better properties as spatial correlation in the covariance model increased and a spatially varying mean structure was included. Coverage converged to 95% and mean square error decreased as the range became larger. This improvement was consistent across different partial sill parameters (Fig. 4). In addition, when the spatially varying mean model was included, coverage improved slightly, whereas mean square error reduced more dramatically. In particular, mean square error of kriging prediction dramatically declined and even became smaller than that of nearest-monitor prediction.

The mean square error and coverage criteria led to generally different conclusions about the preference of nearest-monitor versus kriging predictions. Coverage is a very important property of health effect estimates because interval estimates are the basis for inference. Coverage of 95% means that the 95% confidence interval estimated in an observational study has the correct interpretation, that is, that the interval has a 95% probability of covering the true effect. Although nearest-monitor prediction had generally smaller estimated mean square error, this was due to less variable but

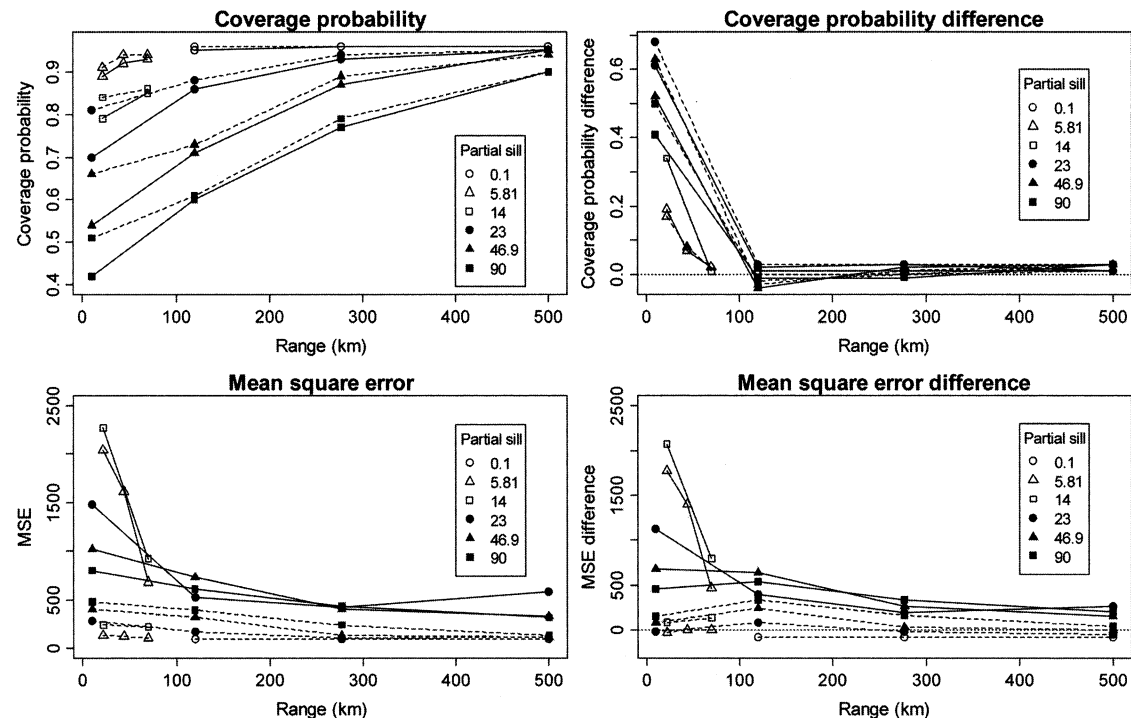


FIGURE 4. Relationship of range versus coverage probability and mean square error (MSE) of health effect estimates for kriging exposure and differences of kriging from nearest-monitor exposures for all 40 true exposure models. Solid and dotted lines represent constant and polynomial mean models, respectively. Horizontal line for the difference figures is at 0 to show no difference. Several extreme values for mean square error and the mean square error difference occurred when partial sill is equal to 0.1 and are not shown. Differences are for estimates (of mean square error or coverage) for kriging exposure minus nearest-monitor exposure.

mostly biased effect estimates that produced generally smaller coverage below 95%. When we examined the expanded set of exposure models with different range, partial sill, and mean parameters, all models except 5 showed better coverage for kriged compared with nearest-monitor exposures. Four models among them had medium range (120 km) and large partial sill (46.9 and 90), with and without mean structure. The coverages of the health effect estimates of nearest monitor were slightly better (by 1%–4%). In the remaining models, coverage in kriging was much better (> 50%) when spatial correlation was small with the range of 10 km and became slightly better (by 1%–3%) as spatial correlation increased up to the range of 500 km (Fig. 4).

Health effect estimates from the 2 prediction methods showed features typically attributed to classic and Berkson measurement errors depending on the spatial structure. Nearest-monitor health effect estimates were always attenuated, as seen from the lower fitted regression line for nearest monitor compared with true exposure (Fig. 3). Attenuation was also present for kriged exposure with no spatial correlation or little range (TEM 2 and TEM 3). These underestimated effects behave like attenuation due to classic measurement error. In contrast, health estimates from kriged predictions of exposure models with strong spatial structure showed little or no

attenuation but the estimates were more variable than for true exposures. This is behavior typically attributed to a Berkson measurement error model, although in this application with a spatially correlated surface the error structure is not independent and identically distributed so the Berkson analogy is only approximate. Because we also showed kriged exposures in models with poor spatial structure gave health effect estimates that were both attenuated and more variable, our results suggest that there is additional deviation from the Berkson error structure due to uncertainty in estimating the prediction model parameters. Szpiro et al¹² propose a method to correct the standard error estimates that accounts for this additional source of uncertainty. In future work, we plan to evaluate the performance of their methods for our examples, and we expect the corrected coverage probabilities will be closer to the target 95%.

There are other ways to incorporate prediction uncertainty in health analyses. In a 2-stage approach, exposure can be simulated to capture the full exposure distribution. Then multiple realizations of the simulated exposures can be used in health analyses and the estimates combined as with multiple imputation. However, the approach is not true multiple imputation because the outcome is not included in the simulation (imputation) step. Gryparis et al¹⁴ have shown that this

“exposure simulation” approach leads to notable attenuation bias in the health effect estimate. A priori, many researchers would argue that the best approach would use a joint model for both the exposure and the disease outcome. It is notable that a few recent simulation studies have not achieved consistently good coverage for the health effect estimate obtained from a joint model, fit either with maximum likelihood¹⁴ or using a fully Bayesian model,¹³ contrary to expectations. In addition, joint models allow feedback between health and exposure, and this feedback could influence inference.¹⁵ It is a defensible scientific choice to keep the exposure modeling separate from the health effect estimation, thus ensuring there is no feedback between the health and exposure models. Tools are being developed to correct second-stage health effect estimates for exposure prediction uncertainty.¹²

One of the features of our approach is that we considered a monitoring network and exposure models based on existing data to ensure the relevance of our results. Health effect estimates from exposure models derived from analysis of LA data (TEM 1 and TEM 4) showed good properties. Although the Los Angeles area has a great deal of spatial variability of $PM_{2.5}$ and a relatively large number of monitors compared with other areas, the monitoring network is still relatively sparse from a spatial statistics perspective. Thus, it may be challenging in practice to model the spatial structure. More work is needed to determine the best approaches to spatial prediction for air pollution exposure data. As a sensitivity analysis for one aspect of this question, we looked at the impact of monitor density on our simulation results. First we increased the number of monitors by 5 up to 42 in the first true exposure model (TEM 1). Locations of additional monitors were uniformly distributed within the area covered by the existing 22 monitors. The denser network did produce marginally better exposure predictions and health effect estimates. With 20 additional monitors, mean square error for health effect estimates decreased by 16% and 15% in nearest-monitor and kriging predictions, respectively. However, the coverage probability was almost identical to that of the original number of monitors. We also explored the influence of a sparser monitoring network, a more frequent occurrence in air pollution epidemiology, by reducing the number of monitors by 5 down to 12 in total. With only 12 monitors (10 less than original monitoring network), our health effect estimates yielded similar coverage probability estimates and a small increase of mean square error estimates for kriged exposures. In contrast, nearest monitor prediction performed much worse. Mean square error increased by over 50%. Coverage probability dropped by about 20%. This sensitivity analysis suggests that a large number of monitors may not be necessary for kriging prediction in a health effect analysis, at least when there is adequate spatial variability and the form of the spatial model is not in question. Health effect estimates

were more sensitive to monitor reduction when individual exposure was predicted using nearest monitor.

We observed that for the kriged exposure, a small percentage of health effect estimates resulted in large outliers (omitted from Fig. 3). In further evaluation of those simulations, we found the exposure model covariance parameter estimates were unreasonably large. Under the spatial structure with very small range, poorly estimated covariance parameters induced the most extreme outlier occurrences. As part of our analysis, we examined both universal and ordinary kriging fitting procedures for all exposure models. We found that estimation using ordinary kriging (ie, with a constant mean) produced consistently poorer covariance parameter estimates across all exposure models. The universal kriging approach gave more stable parameter estimates even when there was no underlying mean structure in the true exposure model. Although it is not practical to investigate a large number of mean and covariance models in a simulation study, further observational research should be done to determine whether other spatial statistical models or fitting algorithms would produce more reasonable parameter estimates in such cases.

To focus this research on features of spatially dependent exposure and prediction that affect health effect parameter estimates, we made some simplifying assumptions. Most important, we assumed ambient air pollution concentration was identical to ambient source exposure. This assumption does not reflect the difference of exposure from concentration due to time activity and local sources or the attenuation of ambient exposure by infiltration. Future research will need to incorporate these exposure features. We also assumed the multivariate normal distribution for the underlying true exposure. In sensitivity analyses, we relaxed this assumption and introduced skewness in the exposure data by changing the underlying distribution to a log-normal. In the few cases we examined, the pattern of health effect estimates remained and kriging still performed better than nearest-monitor prediction. Finally, we assumed there was no additional spatial structure in the disease model. There may be spatial dependence in the health outcome in addition to the spatial structure induced from the exposure distribution. Work is needed to incorporate spatial dependence in the disease model and determine how separable the 2 sources of spatial structure may be in the health analysis.

Although the focus of this paper is specific to the field of air pollution epidemiology, this research is broadly relevant to inference in epidemiologic studies in the presence of limited exposure data. We present a problem in which there are no exposure measurements for study subjects but one in which it is possible to use an external dataset to develop an exposure prediction model to plug into the health effect analysis. We evaluate the merits of 2 exposure prediction approaches directly in terms of the parameter of interest in the epidemiologic study. We learned that it is not sufficient to

assess the accuracy and precision of the exposure predictions when the target of inference is the health effect parameter in the disease model. We showed that the bias and precision of the resulting health effect estimates depend on the approach to exposure prediction and the features of the underlying exposure distribution. Although the details of our results are specific to the air pollution epidemiology context, our general approach and overarching conclusions are relevant to other epidemiologic study settings.

In summary, based on the generally better coverage and lower bias properties of the health effect parameter estimate of interest, we conclude that kriging exposure prediction is preferable to nearest-monitor prediction for estimation of long-term air pollution health effects, particularly when the underlying true exposure is less spatially correlated. Both prediction approaches perform well when there is good spatial structure in the underlying exposure distribution.

ACKNOWLEDGMENT

We thank Kristin A. Miller for her constructive suggestions and comments.

REFERENCES

1. Dockery DW, Pope CA III, Xu X, et al. An association between air pollution and mortality in six US cities. *N Engl J Med*. 1993;329:1753–1759.
2. Pope CA III, Burnett RT, Thun MJ, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*. 2002;287:1132–1141.
3. Basu R, Woodruff TJ, Parker JD, Saulnier L, Schoendorf KC. Comparing exposure metrics in the relationship between PM_{2.5} and birth weight in California. *J Expo Anal Environ Epidemiol*. 2004;14:391–396.
4. Miller KA, Siscovick DS, Sheppard L, et al. Long-term exposure to air pollution and incidence of cardiovascular events in women. *N Engl J Med*. 2007;356:447–458.
5. Ritz B, Wilhelm M, Zhao Y. Air pollution and infant death in Southern California, 1989–2000. *Pediatrics*. 2006;118:493–502.
6. Wilhelm M, Ritz B. Local variations in CO and particulate air pollution and adverse birth outcomes in Los Angeles County, California, USA. *Environ Health Perspect*. 2005;113:1212–1221.
7. Finkelstein MM, Jerrett M, Sears MR. Environmental inequality and circulatory disease mortality gradients. *J Epidemiol Community Health*. 2005;59:481–487.
8. Jerrett M, Burnett RT, Ma R, et al. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology*. 2005;16:727–736.
9. Kunzli N, Jerrett M, Mack WJ, et al. Ambient air pollution and atherosclerosis in Los Angeles. *Environ Health Perspect*. 2005;113:201–206.
10. Leem JH, Kaplan BM, Shim YK, et al. Exposures to air pollutants during pregnancy and preterm delivery. *Environ Health Perspect*. 2006;114:905–910.
11. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall/CRC Press; 2004; 21–68.
12. Szpiro AA, Sheppard L, Lumley T. Accounting for errors from predicting exposures in environmental epidemiology and environmental statistics. *UW Biostatistics Working Paper Series*. 2008. Working paper 330.
13. Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA. Measurement error caused by spatial misalignment in environmental epidemiology. *Harvard University Biostatistics Working Paper Series*. 2007. Working paper 59.
14. Madsen L, Ruppert D, Altman NS. Regression with spatially misaligned data. *Environmetrics*. 2008;19:453–467.
15. Wakefield J, Shaddick G. Health-exposure modeling and the ecological fallacy. *Biostatistics*. 2006;7:438–455.