Environmental Science & Technology

Article

pubs.acs.org/est

Large Scale Air Pollution Estimation Method Combining Land Use Regression and Chemical Transport Modeling in a Geostatistical Framework

Yasuyuki Akita,[†,*] Jose M. Baldasano,[‡,§] Rob Beelen,[∥] Marta Cirach,[⊥,#,▽] Kees de Hoogh,[◊] Gerard Hoek,[∥] Mark Nieuwenhuijsen,[⊥,#,▽] Marc L. Serre,[†] and Audrey de Nazelle[⊥,⬡]

[†]Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina—Chapel Hill, Chapel Hill, North Carolina 27599-7431, Untied States

[‡]Earth Sciences Department, Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC−CNS), Barcelona 08034, Spain

[§]Environmental Modelling Laboratory, Technical University of Catalonia (UPC), Barcelona 08034, Spain

[∥]Institute for Risk Assessment Sciences, Utrecht University, Utrecht NL-3508, The Netherlands

[⊥]Centre for Research in Environmental Epidemiology (CREAL), Barcelona 08003, Spain

[#]IMIM (Hospital del Mar Research Institute), Barcelona 08003, Spain

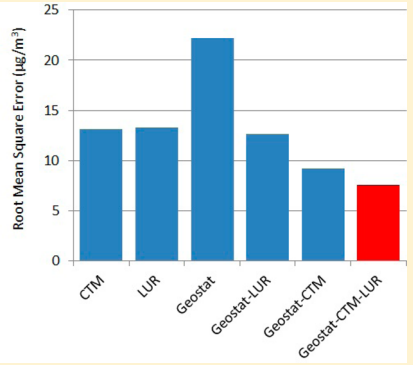
[▽]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain

[◊]MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, London WC1N 1EH, United Kingdom

[⬡]Centre for Environmental Policy, Imperial College London, London SW7 1NA, United Kingdom

Supporting Information

ABSTRACT: In recognition that intraurban exposure gradients may be as large as between-city variations, recent air pollution epidemiologic studies have become increasingly interested in capturing within-city exposure gradients. In addition, because of the rapidly accumulating health data, recent studies also need to handle large study populations distributed over large geographic domains. Even though several modeling approaches have been introduced, a consistent modeling framework capturing within-city exposure variability and applicable to large geographic domains is still missing. To address these needs, we proposed a modeling framework based on the Bayesian Maximum Entropy method that integrates monitoring data and outputs from existing air quality models based on Land Use Regression (LUR) and Chemical Transport Models (CTM). The framework was applied to estimate the yearly average NO_2 concentrations over the region of Catalunya in Spain. By jointly accounting for the global scale variability in the concentration from the output of CTM and the intraurban scale variability through LUR model output, the proposed framework outperformed more conventional approaches.



INTRODUCTION

A large body of research has shown that exposure to combustion-related ambient air pollution is associated with increased morbidity and mortality.[1−3] In recognition that intraurban exposure gradients may be as large as between-city variations, recent air pollution epidemiologic studies have become increasingly interested in capturing within-city exposure gradients.[4,5]

Among many exposure assessment approaches,[4] land use regression (LUR) modeling has become a commonly used method in air pollution epidemiologic researches.[6,7] LUR modeling is generally employed to estimate the annual average concentration of ambient air pollutant based on traffic characteristics and land use/land cover around the site. By incorporating site-specific variables into the model, the LUR model is suitable for depicting small scale spatial variability at a fine urban scale. However, this site-specificity makes it difficult to extrapolate the developed models into the wider domain.

Another approach commonly used in air pollution epidemiologic research is spatial interpolation methods.[4,8] Among many spatial interpolation approaches, geostatistical methods such as Ordinary Kriging (OK) are the most widely used, because they provide concentration estimates together

Received: December 3, 2013
Revised: February 28, 2014
Accepted: March 12, 2014
Published: March 12, 2014

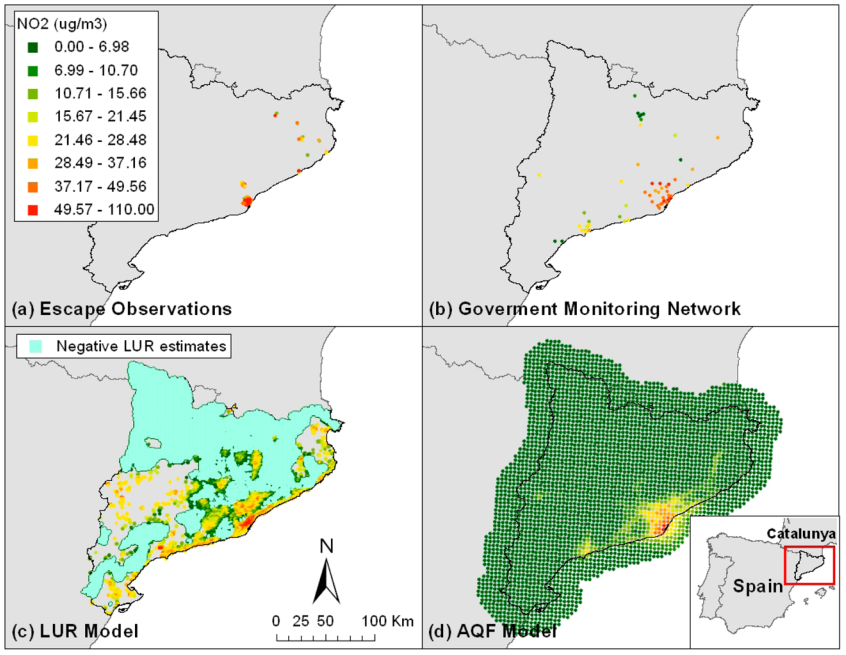


Figure 1. NO_2 concentration map of (a) 80 ESCAPE monitoring sites, (b) 53 government monitoring sites (Xarxa de Vigilància i Previsió de la Contaminació Atmosfèrica—XVPCA), (c) LUR estimates on 1 km grid, and (d) CALIOPE AQFM output at 4 km grid.

with standard error at unmonitored locations. The standard error measures the uncertainty of the interpolated results and helps identify areas with unreliable estimates. The estimation quality of OK, however, depends heavily on the availability and quality of monitoring data, since OK only uses observations. Solely relying on observations at sparse government monitoring networks, which are primarily used to assess the compliance with regulatory standards, is often not suitable to capture small-scale spatial variability, and produces overly smoothed concentration surface.

Because of the rapidly accumulating health data, recent air pollution epidemiologic studies tend to deal with large study populations distributed over large geographic domains to detect small relative risks of relatively rare diseases associated with exposure to ambient air pollution.[9] Thus, in addition to the small-scale spatial variability, there is a strong demand for a modeling framework that can cover large geographic domains.[10] Chemical Transport Models (CTM) are used to simulate the dynamics of ambient air pollutants over large spatial domains with coarse spatial resolution (between 12 and 25 km at regional scale, and between 4 and 10 km for single country applications) and fine temporal resolution (1 h) by jointly accounting for the emission patterns, meteorological conditions, and chemical reactions.[11−13] CTMs divide the atmosphere into three-dimensional grid cells. Chemical and physical transformations are simulated within each cell, and transport and diffusion processes are modeled between cells.[4] Even though CTMs generally cover large geographic domain, using CTMs in exposure assessment for epidemiologic studies is currently challenging, as their resolution is generally too coarse to capture within-city variability. In addition, high implementation costs and data requirements are limiting factors for their application in exposure assessment.[4]

Several studies have developed methods combining data from multiple sources. For example, U.S. EPA's Fused Air Quality Surfaces Using Downscaling model combined monitoring data with the output of the Models-3/Community Multiscale Air Quality (CMAQ) model using a Bayesian space-time downscaler model.[14] Beckerman et al.[15] developed a hybrid approach combining LUR model and geostatistical interpolation to estimate $PM_{2.5}$ concentration over the contiguous United States. Although several modeling approaches have been introduced to address the current needs for air pollution epidemiologic research, a modeling framework capturing within-city exposure variability to a large application domain with a low implementation cost is still lacking. The aim of this work, therefore, is to develop an estimation framework that improves the assessment of exposure gradients of traffic related air pollution at both the urban and regional scale by incorporating strengths of the different methods using existing data to reduce the implementation cost. We developed a geostatistical modeling framework based on the Bayesian Maximum Entropy (BME) method, combining the results of existing LUR model and CTM together with observational data to estimate the yearly average NO_2 concentration over the region of Catalunya in Spain. By combining a LUR model and CTM using the BME method, the framework supplements the limitation of each method. To evaluate the model performance of the proposed framework, we conducted a validation analysis using an independent set of observations at government-run monitoring sites.

METHODS AND MATERIALS

ESCAPE Monitoring Data and Land Use Regression model. Monitoring data and LUR model estimates for yearly average NO_2 concentrations originated from the European Study of Cohorts for Air Pollution Effects (ESCAPE) project. A complete description of the sampling methods and the LUR model development can be found elsewhere.[10,16,17] In brief, an LUR model was built in Catalunya, Spain, based on the yearly average NO_2 concentration observed at 80 monitoring sites with passive samplers in 2009. The monitoring sites were meant to cover a domain representing residential locations of participants in three epidemiologic cohorts, and included 40

dx.doi.org/10.1021/es405390e | Environ. Sci. Technol. 2014, 48, 4452−4459

sites in Barcelona metropolitan area, 16 sites in the neighboring town of Sabadell, and 24 in the province of Girona, as shown in Figure 1(a). Various emission-related variables, such as traffic intensity, road length, population density, and land use were constructed within a Geographic Information System (GIS). A multivariate forward supervised stepwise regression approach was employed to select the significant LUR model predictors. Model performance was evaluated by leave-one-out cross validation. Traffic intensity on major roads within a 25 m buffer (veh day^{-1} m), road length within a 1000 m buffer (m), and Natural and forested areas within a 5000 m buffer (m^2) were selected as the final LUR model predictor variables.[16] The regression coefficients of these predictors are listed in Table S1 in the Supporting Information (SI). A map of the yearly average concentration based on LUR model estimations on a 1-km grid points over Catalunya was also developed to visually inspect the pollutant distribution. When applying the LUR model, the predictor variables that fell outside the range of the observed values at the monitoring sites were truncated to the maximum (or minimum) observed value of these monitoring sites, because the linear relationship between the predictors and the concentrations might not hold outside of the range. In this work, we utilized the LUR estimates on the 1 km grid points as an input of the proposed framework, taking advantage of existing model outputs (Figure 1(c)).

Air Quality Modeling System. We obtained hourly NO_2 concentrations estimated by the CALIOPE Air Quality Forecast Modeling (AQFM) system (www.bsc.es/caliope) over Spain for the year 2009.[18] CALIOPE combines four models to forecast air quality (O_3, NO_2, SO_2, PM10, and PM2.5) over Spain with a spatial resolution of 4×4 km^2 in the Iberian Peninsula domain, within 15 vertical layers reaching up to 50 hPa, and an hourly temporal resolution. It integrates a meteorological model (WRF v3.0.1.1) and emission model (HERMES 2004), a chemical transport model (CMAQ v4.5) and a mineral dust atmospheric model (BSC-DREAM8b) together in an air quality modeling system.[19,20]

Of particular importance for high quality air pollution modeling, the High-Elective Resolution Modeling Emission System (HERMES) [18] uses state-of-the-art methodologies for emission estimations. It calculates emissions by sector-specific sources or by individual installation and stacks following a bottom-up approach for the reference year 2004. Specifically, the on-road traffic emissions module takes into account 72 diesel and gasoline vehicle categories, according to COPERT III-CORINAIR methodology.[21,22] The model includes the definition of the road network, dividing it in stretches with specific temporary disaggregating profiles, specific average speed, daily average traffic, stretch length, route type, and circulation type, and accounts for hot exhaust, cold exhaust, and evaporative emissions. Several evaluation studies[23] and near-real time evaluation (NRT) against air quality measurements on an hourly basis support the confidence on the system.

In this work, the ground level daily NO_2 concentrations within a 20 km buffer around Catalunya were extracted from CALIOPE to accurately interpolate the concentration on estimation and data points near the boundary, and averaged over the year at each grid cell. Figure 1(d) shows the map of the yearly average concentration calculated from CALIOPE AQFM.

Government Monitoring Network Data. We obtained NO_2 yearly average concentrations for the year 2009 observed at 61 government-run monitoring sites over Catalunya, equipped with reference method chemiluminiscense monitors. These data, shown in Figure 1(b), were used as an independent data set to evaluate the model performance of the proposed framework. To use stable yearly average concentrations, eight monitoring sites for which the number of daily observations was less than 75% of intended number of samples were removed from the data set. In order to construct the LUR estimates at each site predictor variables listed in SI Table S1 were computed. These estimates were also used as an input of the proposed modeling framework.

Bayesian Maximum Entropy Method. To estimate the yearly average NO_2 concentration at unmonitored location, we developed a geostatistical estimation framework that integrates all available data: ESCAPE monitoring data, LUR estimates, and CALIOPE AQFM output. In this work, the BME method was employed in order to process various sources of data with different levels of uncertainty.[24,25] The BME method provides a mathematically rigorous framework that integrates a variety of available knowledge bases (e.g., spatial dependency model, empirical relationships, scientific model, physical laws, and so forth) with data having varying levels of epistemic uncertainty. In the BME method, the data are categorized into two groups: (i) hard data, corresponding to measurements; and (ii) soft data, having an uncertainty characterized by a probability density function (PDF) of any type (e.g., Gaussian, Uniform). A full description of the BME method can be found elsewhere.[26] In brief, the BME method can be viewed as a two-stage knowledge processing procedure. At the prior stage, maximum entropy theory is used to process the general knowledge base describing global characteristics of the spatial random field representing the NO_2 yearly average concentration, such as its mean trend and the covariance function, which produces a prior PDF depicting the spatial process. Then at the posterior stage, an operational Bayesian conditionalization rule is used to update this prior PDF with respect to the site specific hard and soft data available, which produces a BME posterior PDF describing the value of the spatial process at any estimation point of interest.

BME Geostatistical Estimation Framework. In this framework, the yearly average NO_2 estimates from the AQFM on the 4 km grid cells were treated as a deterministic global offset, since it covers much larger domain than our study domain and takes into account the local emission and meteorological condition. The deterministic global offset is a function of space that describes consistent spatial patterns in the data. The word "global" emphasizes that this offset applies to the whole spatial domain encompassing all the available data. In the geostatistical analysis, the offset is generally removed from the data prior to the analysis in order to produce residuals that are as homogeneous as possible. To calculate the value of offset at each data and estimation point, the AQFM estimate in each cell was assigned to the centroid of the cell, and then, the values were linearly interpolated on the data and estimation points.

The residuals were processed by a geostatistical estimation framework based on the BME method to estimate the residual concentration at unmonitored location. The residuals concentrations at the ESCAPE monitoring sites were computed by subtracting the AQFM offset from the data and treated as hard data, since they were directly derived from the observations. The LUR estimates at the 1 km grid points and government monitoring sites were treated as soft data having an uncertainty associated with the estimation based on the predictor variables.

The PDFs characterizing this uncertainty were defined in the following forms. At the soft data points where predictor variables were within the range of those at the ESCAPE monitoring sites, we assumed that an adequate approximation for the PDF is a normal distribution truncated below zero because the concentrations cannot be negative. The mean of this truncated normal distribution was set to the concentration estimated by the LUR model, and corresponding standard deviation was derived from the prediction error. Since the LUR model was constructed based on the ordinary least-squares algorithm, the standard deviation of the estimate at a given grid point is computed by (eq 1).

$$\sigma = s\sqrt{1 + \mathbf{a}'(\mathbf{P}'\mathbf{P})^{-1}\mathbf{a}} \quad (1)$$

where vector $\mathbf{a}$ is the set of explanatory variables at the grid point, $\mathbf{P}$ is the design matrix of the LUR model, and s is the estimate of the standard deviation of the error term. At the soft data points whose LUR estimates were negative (shown in light blue in Figure 1(c)), we assumed that the PDF is approximated by a uniform distribution from zero to the minimum observed concentration at ESCAPE monitoring sites, since the true concentration is expected to be small. All of the positive LUR estimates whose predictor variables were outside the range of those at the ESCAPE monitoring sites were excluded from the framework (SI Table S6). Finally, the PDF was shifted by the AQFM offset calculated at each soft data point.

The sample covariogram was calculated from the residuals of the hard and "hardened" soft data, which are given by the mean of truncated normal distribution, and were then used to fit a positive definite covariogram model using an automated weighted least-squares procedure. In this study, exponential, Gaussian, and spherical models were tested, and the spherical model which produced the smallest mean square error was used in the analysis (SI Table S2).

The expected value of the BME posterior PDF was used as an estimate of the residual concentration at the estimation point, and the corresponding BME posterior standard deviation provides a useful characterization of the associated estimation uncertainty. The deterministic global offset interpolated on the estimation location was finally added back onto the residual estimate to obtain the yearly average NO_2 estimate.

Model Validation. To evaluate the model performance of the proposed framework, we compared the yearly average concentrations at the government monitoring stations to predictions at these stations from six estimation methods. In method (1), the concentration is estimated by the deterministic global offset based on the AQFM. In method (2), the LUR model was directly applied to estimate the concentration based on the set of predictor variables obtained for each validation point. Method (3) consisted of the BME method with an unknown constant mean based only on hard data observed at ESCAPE monitoring sites. This is equivalent to OK. In method (4), the BME method with an unknown constant mean was applied integrating the hard data (ESCAPE monitoring data) and LUR soft data at the 1 km grid and validation points. Methods (3) and (4) assumed that there is no global offset. In methods (5) and (6), first the deterministic global offset based on the AQFM was removed from the data to obtain the homogeneous residuals. In method (5), the BME method with zero constant mean based only on hard data was applied. This is equivalent to Simple Kriging (SK). In method (6), the BME method with zero constant mean integrating the hard and soft data as in method (4) was used to estimate the residual concentration. Method (6) corresponds to our proposed BME geostatistical estimation framework. Although the government monitoring data was used as a validation set for a model comparison in this work, the data can be used as hard data to increase the number of observational data. The yearly average concentration was estimated at each government-run monitoring site and compared with the observed yearly average concentration. To quantify the accuracy of each estimation method, four error statistics were calculated: root-mean-square error (RMSE), mean prediction error (MPE), Pearson's correlation coefficient, and Spearman's rank correlation.

The government-run monitoring sites were categorized into the following three types based on the main emission affecting each site: background, industrial, and traffic. In order to investigate the effect of the types of the validation point on the model performance, error statistics were also calculated separately for each type (in addition to all together).

We used ArcGIS 9.2 (ESRI Inc., Redlands, CA) for the GIS analysis and Matlab R2010a (MathWorks Inc., Natick, MA, U.S.A.) and *BMElib*, version 2.0b[26] for the geostatistical estimation.

RESULTS AND DISCUSSION

Table 1 lists the error statistics of the validation analysis obtained by the six estimation methods presented above. Our proposed framework BME with AQFM offset (method 6) outperformed all other methods in terms of all four error statistics. The RMSE of methods (1) AQFM and (2) LUR were similar. However, the MPE of method (1) is negative, whereas that of method (2) was positive. The model performance of method (3) OK without AQFM offset was the worst among all methods. The RMSE and MPE of method (3) are much higher than those of (1) and (2). The model performance of method (4) BME without AQFM offset was similar to those of (1) and (2). Method (5) SK with AQFM offset and (6) BME with AQFM offset outperformed other methods. Method (5) reduced RMSE by approximately 30% relative to the AQFM (method 1). Our proposed framework BME with AQFM offset (method 6) further reduced RMSE by 17% relative to method (5). Both methods (5) and (6) yielded Pearson's correlation coefficients and Spearman's rank correlations close to 0.9.

Table 1. Error Statistics of Validation Analysis for 53 Government-Run Monitoring Sites Based on Six Estimation Methods

method	RMSE[d]	MPE[e]	CORR[f]	RANK[g]
(1) AQFM	13.16	−9.61	0.83	0.86
(2) LUR	13.27	3.76	0.77	0.72
(3) OK[a] without offset	22.17	16.70	0.41	0.36
(4) BME without offset	12.65	7.53	0.79	0.74
(5) SK[b] with AQFM offset	9.21	−2.45	0.88	0.89
(6) BME with AQFM offset[c]	7.60	1.38	0.90	0.90

[a]Ordinary Kriging. [b]Simple Kriging. [c]Method (6) BME with AQFM offset corresponds to our proposed framework. [d]Root Mean Square Error, in $\mu g/m^3$. [e]Mean Prediction Error, in $\mu g/m^3$. [f]Pearson's correlation coefficient. [g]Spearman's rank correlation.

Table 2 lists the RMSE obtained by the six estimation methods at background, industrial, and traffic sites. Other statistics (MPE, Pearson's correlation coefficients and Spearman's rank correlations) are listed in the SI. Even though the

dx.doi.org/10.1021/es405390e | Environ. Sci. Technol. 2014, 48, 4452−4459

Table 2. Root Mean Square Error (RMSE, in $\mu g/m^3$) of the Six Estimation Methods at Each Type of Validaton Site (Background, Industrial, and Traffic) and for All Sites (All)

method	background (14)[a]	industrial (21)[a]	traffic (18)[a]	all (53)[a]
(1) AQFM	8.67	10.42	18.03	13.16
(2) LUR	12.85	11.98	14.90	13.27
(3) OK without offset	24.41	24.88	16.15	22.17
(4) BME without offset	12.99	11.16	13.94	12.65
(5) SK with AQFM offset	10.85	6.48	10.44	9.21
(6) BME with AQFM offset[a]	8.95	6.49	7.67	7.60

[a]Numbers in parentheses indicate the number of sites.

overall model performance of (1) AQFM and (2) LUR are similar, the two methods exhibit completely different patterns in model performance. Method (1) AQFM performed best at background sites, whereas the performance at the traffic sites was poor. In contrast, the performance of (2) LUR was almost the same regardless of the location setting of the validation point. Method (3) OK without AQFM offset performed much better at the traffic sites than at the industrial and background sites, whereas the performance of (4) BME without AQFM offset was almost the same regardless of the location setting, just as in method (2) LUR. Method (5) SK with AQFM offset and (6) BME with AQFM offset outperformed all other estimation methods regardless of the location setting. At the industrial and background site, model performance of both methods were similar. At the traffic sites, however, method (6) outperformed method (5) and reduced the RMSE about 27% relative to method (5).

The validation analysis provides insights into the model predictability of each method. The overall model performance of (1) AQFM and (2) LUR, for instance, were similar in terms of RMSE (Table 1). However, there was a systematic difference in model performance between them. The MPE of AQFM was negative, and its RMSE at traffic sites was much higher than that at nontraffic sites (Table 2). Pay et al.[20] reported that NO_2 daily concentrations estimated on a 12 km grid over Europe by the same CALIOPE AQFM correlated well with observations, but tended to underestimate concentrations at regional background sites. They attributed the negative bias to an underestimation of emission sources in highly polluted regions, which explains our results. In an evaluation of CALIOPE over 68 Spanish monitoring stations for 2004, Baldasano et al.[27] identified that NO_2 measurements at urban stations from small and medium-sized cities generally showed poorer agreement with modeled concentration, whereas modeled bias decreased at urban stations located in larger cities, such as Barcelona or Madrid, for which a detailed characterization of emission sources, especially from road traffic, at a higher spatial resolution (1×1 km^2) was available. Traffic sites in our study were present both in Barcelona and in rural/small town settings (SI Figure S1(b)), thus explaining our results. In contrast, the MPE of the (2) LUR method was positive and much closer to zero. Moreover, the model performance of LUR was consistent across the location settings (Table 2). Since the LUR model construction was primarily focused on traffic-related variables, depicting small-scale spatial variability in concentrations due to traffic,[10,16] it is no surprise that the LUR model outperformed AQFM at traffic site with a 17% reduction in the RMSE. The LUR model performed equally well at nontraffic sites (albeit less well than the AQFM at these sites), showing consistency across land use types, as would be hoped for (possibly a result of the natural area variable). It should be noted that the LUR model was applied to regions outside its modeling domain (more than 20 out of the 53 monitoring stations were outside of the cohort addresses domain), and to location types that are not residential (as it was developed for), and yet behaved reasonably well. This is consistent with general findings that LUR models often perform better than or as well as chemical transport or dispersion models to predict small-scale variations.[4]

The overall model performance of OK without AQFM offset (method (3)) was poorest among all the methods. The RMSE was about 68% higher than that of the (1) AQFM and (2) LUR models. Method (3) performed best at traffic sites, where the RMSE was even slightly smaller than that obtained by (1) AQFM method (Table 2). As shown in SI Figure S1, most of the government monitoring traffic sites were located near the ESCAPE monitoring sites, whereas the nontraffic sites were generally located away from these points. This indicates that the OK model performance depends heavily on the proximity between the input data points and the validation points. The BME method accounting for the LUR-based soft data, but with

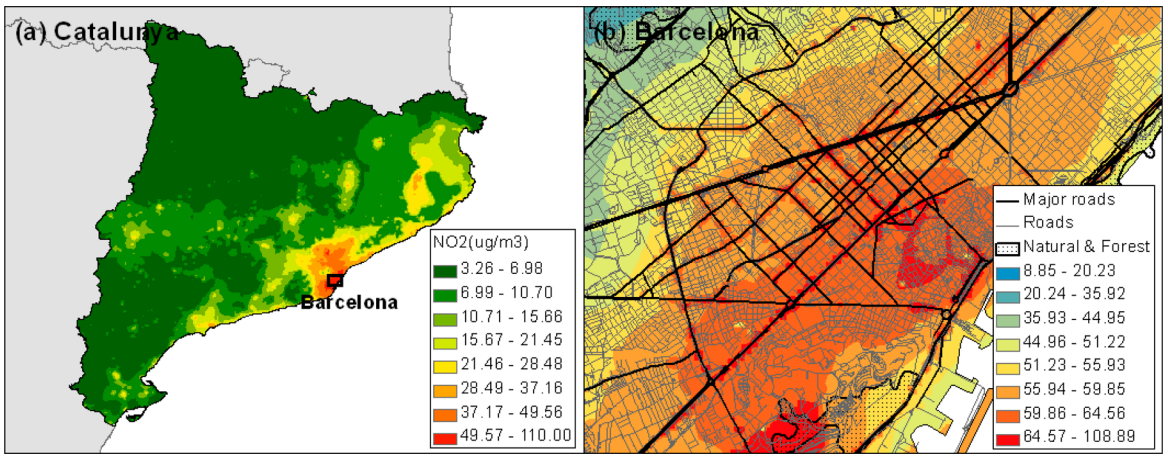


Figure 2. Maps of NO_2 yearly average concentration ($\mu g/m^{-3}$) estimated by the proposed modeling framework over (a) Catalunya and (b) Barcelona.

dx.doi.org/10.1021/es405390e | Environ. Sci. Technol. 2014, 48, 4452−4459

no global offsetting using the AQFM (method (4)), greatly improved the model performance over method (3), and the overall model performance was close to those of (1) AQFM and (2) LUR. Since the BME method complements the scarcity of the data points by accounting for the soft data based on the LUR outputs, it reduced the RMSE relative to method (3) regardless of the location settings.

Method (5), SK with AQFM offset, outperformed other methods except for method (6) and reduced the RMSE by approximately 30% relative to (1) AQFM and (2) LUR. The slightly negative MPE was probably the influence of the negative bias of the global offset based on the AQFM. Method (5) performed best at industrial sites and, greatly reduced the RMSE relative to (1) AQFM and (2) LUR at traffic and industrial sites. However, improvement at background sites was limited.

Our proposed framework, BME with AQFM offset, outperformed all other models and further reduced the RMSE by approximately 17% relative to method (5). This reduction was primarily due to the improvement in the model performance at traffic sites (27% RMSE reduction relative to method (5)). Interestingly, both methods (5) and (6) performed best at the industrial sites among all the location settings, due to LUR and AQF models performing simultaneously equally well at these sites (while as in background sites, the AQFM performed better than the LUR model, and the reverse was true at traffic sites).

In addition to the validation analysis, we created maps (Figure 2) of the estimated NO_2 concentration over (a) Catalunya and (b) Barcelona using the proposed framework to visually inspect the model performance. By jointly accounting for the global scale variability in the concentration from the output of the AQFM model and the intraurban scale variability through the soft data based on the LUR model output, both maps show not only a global trend similar to the AQFM model output, but also the influence of the small scale spatial variability influenced by the local traffic. The overall concentration distribution on the map over (a) Catalunya is similar to that of the AQFM model (Figure 1(d)). In addition, the map created by the proposed method also captures the small scale concentration gradient along major traffic road (see SI Figure S2(a) for the major road network). In Barcelona (Figure 2(b)), concentrations are generally high in the urban center (from left bottom corner to upper right corner) and gradually decrease away from the urban center. The map also shows the small-scale gradient along the road network.

Even though the proposed modeling framework performed best among all methods, there are several issues that need to be addressed. First, ESCAPE monitoring data and the validation data on the government monitoring sites were observed with different methods. The difference of the sampling techniques was not taken into account in this work. Hence, the error statistics presented above might contain additional error due to this difference, even though the difference is expected to be small. The ratio of average NO_2 concentrations measured by passive sample (Ogawa badge) and chemiluminescence monitors, colocated at the same site in Catalunya is 0.96.[17] Second, the AQFM on the 4 km grid cells were simply treated as a deterministic global offset in the proposed framework. The uncertainty associated with the AQFM was not taken into account. Third, the soft data were constructed based on the LUR estimates on 1 km grid points and on validation points (government-run monitoring sites) to fully utilize the existing study result. However, the choice of the grid size was arbitrary and finer resolutions are possible. Any shape or size of LUR estimates as soft data can be integrated in the BME framework, and it is expected that increasing the size of the soft data points would improve the model performance. Fourth, since we used the LUR estimates over all of Catalunya (our study domain), we implicitly assumed that the LUR model can be applicable to larger domain than the original data domain of the ESCAPE study, a subregion of Catalunya. As pointed out by several researchers, applicability of the LUR model outside of the original data domain is questionable.[28] In a similar vein, the LUR model was meant to predict concentrations in residential settings, but Government monitoring stations are not necessarily representative of such settings, thus stretching the applicability of the LUR model even further. The effect of the design and purpose of the existing air quality models and monitoring data were not taken into account in this work. Finally, ESCAPE monitoring data were directly used as hard data, and were also used to construct the LUR estimates, from which soft data were derived. We assumed that the spatial random field of the residual concentrations was simply characterized by spatial autocorrelation evaluated by the covariance function. However, there are correlations that are not fully taken into account in the proposed framework. It is interesting to observe that despite these limitations, the proposed BME method integrating LUR outputs still outperformed all other methods in the entire region of Catalunya.

The performance of the BME method is dependent on the quality of models and observed data used as inputs. It is also subject to modeling decisions, such as the shape and size of the grid chosen to integrate model outputs, and the influence of such decisions should be investigated in future work. Furthermore, future work can accommodate temporal patterns in addition to spatial variations.[25] The BME method has already been used as a tool for space/time interpolation in several studies. Thus, the proposed framework can be easily extended in a space/time context. For example, temporal resolution can be integrated into the LUR model by calibrating the model using the temporal trend derived from regulatory monitoring site,[29,30] or directly from CTM outputs or introducing variables with high temporal resolution, such as meteorological data.[31] These approaches can be used to construct soft data based on a LUR model with a finer temporal resolution. Thus, extension of the proposed framework into spatiotemporal estimation should be pursued in future studies.

In summary, a modeling framework based on a BME geostatistical method that integrates monitoring data and outputs from existing air quality models based on LUR and CTM was developed and applied to estimate the yearly average NO_2 concentration over Catalunya, Spain. Since recent air pollution epidemiologic studies often deal with large study population to detect small relative risks of rare diseases, an exposure assessment tool that can detect small-scale variability over large study domain is essential to reduce exposure misclassification. By jointly accounting for the global scale variability in the concentration from the output of an AQFM model and the intraurban scale variability through soft data based on a LUR model output, the proposed framework successfully captured within-city exposure variability over large geographic domain, and thus serves as an exposure assessment tool for those studies with a low implementation cost.

■ ASSOCIATED CONTENT

Supporting Information

The final LUR model predictor variables and corresponding regression coefficient (Table S1), covariogram models and model parameters (Table S2), maps of ESCAPE monitoring sites, and government-run monitoring sites (Figure S1), maps of the predictor variables used in the LUR model (Figure S2), mean prediction error (Table S3), Pearson's correlation coefficient (Table S4), and Spearman's rank correlation (Table S5) of the six estimation methods at each type of validation site and for all sites, the number of soft data points (Table S6). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (919) 966 1095; fax: (919) 966 7911; e-mail: akita@email.unc.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research received funding from CREAL; the LUR model development was funded by the European Community's Seventh Framework Program (FP7/2007e2011) under Grant Agreement No. 211250; CALIOPE air quality simulations were performed on the MareNostrum supercomputer hosted by the Barcelona Supercomputing Center (BSC-CNS), with support from the Spanish Ministry of Economy and Competitiveness (Project CGL2010/19652) and Severo Ochoa Grant SEV-2011-0006.

■ REFERENCES

(1) Brook, R. D.; Rajagopalan, S.; Pope, C. A.; Brook, J. R.; Bhatnagar, A.; Diez-Roux, A. V.; Holguin, F.; Hong, Y. L.; Luepker, R. V.; Mittleman, M. A.; Peters, A.; Siscovick, D.; Smith, S. C.; Whitsel, L.; Kaufman, J. D. Particulate matter air pollution and cardiovascular disease an update to the scientific statement from the American Heart Association. *Circulation* **2010**, *121* (21), 2331−2378.

(2) Hoek, G.; Krishnan, R. M.; Beelen, R.; Peters, A.; Ostro, B.; Brunekreef, B.; Kaufman, J. D. Long-term air pollution exposure and cardio- respiratory mortality: A review. *Environ. Health* **2013**, *12*.

(3) WHO. *Review of evidence on health aspects of air pollution—REVIHAAP Project*; 2013; p 309.

(4) Jerrett, M.; Arain, A.; Kanaroglou, P.; Beckerman, B.; Potoglou, D.; Sahsuvaroglu, T.; Morrison, J.; Giovis, C. A review and evaluation of intraurban air pollution exposure models. *J. Expo. Anal. Enviro. Epidemiol.* **2005**, *15* (2), 185−204.

(5) Jerrett, M.; Burnett, R. T.; Beckerman, B. S.; Turner, M. C.; Krewski, D.; Thurston, G.; Martin, R. V.; van Donkelaar, A.; Hughes, E.; Shi, Y.; Gapstur, S. M.; Thun, M. J.; Pope, C. A. Spatial analysis of air pollution and mortality in California. *Am. J. Resp. Crit. Care Med.* **2013**, *188* (5), 593−599.

(6) Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42* (33), 7561−7578.

(7) Ryan, P. H.; LeMasters, G. K. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhal. Toxicol.* **2007**, *19*, 127−133.

(8) Jerrett, M.; Burnett, R. T.; Ma, R. J.; Pope, C. A.; Krewski, D.; Newbold, K. B.; Thurston, G.; Shi, Y. L.; Finkelstein, N.; Calle, E. E.; Thun, M. J. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* **2005**, *16* (6), 727−736.

(9) Raaschou-Nielsen, O.; Andersen, Z. J.; Beelen, R.; Samoli, E.; Stafoggia, M.; Weinmayr, G.; Hoffmann, B.; Fischer, P.; Nieuwenhuijsen, M. J.; Brunekreef, B.; Xun, W. W.; Katsouyanni, K.; Dimakopoulou, K.; Sommar, J.; Forsberg, B.; Modig, L.; Oudin, A.; Oftedal, B.; Schwarze, P. E.; Nafstad, P.; De Faire, U.; Pedersen, N. L.; Ostenson, C. G.; Fratiglioni, L.; Penell, J.; Korek, M.; Pershagen, G.; Eriksen, K. T.; Sorensen, M.; Tjonneland, A.; Ellermann, T.; Eeftens, M.; Peeters, P. H.; Meliefste, K.; Wang, M.; Bueno-de-Mesquita, B.; Key, T. J.; de Hoogh, K.; Concin, H.; Nagel, G.; Vilier, A.; Grioni, S.; Krogh, V.; Tsai, M. Y.; Ricceri, F.; Sacerdote, C.; Galassi, C.; Migliore, E.; Ranzi, A.; Cesaroni, G.; Badaloni, C.; Forastiere, F.; Tamayo, I.; Amiano, P.; Dorronsoro, M.; Trichopoulou, A.; Bamia, C.; Vineis, P.; Hoek, G. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol.* **2013**, *14* (9), 813−822.

(10) de Nazelle, A.; Aguilera, I.; Nieuwenhuijsen, M.; Beelen, R.; Cirach, M.; Hoek, G.; de Hoogh, K.; Sunyer, J.; Targa, J.; Brunekreef, B.; Künzli, N.; Basagaña, X. Comparison of performance of land use regression models derived for Catalunya, Spain. *Atmos. Environ.* **2013**, *77* (0), 598−606.

(11) Menut, L.; Bessagnet, B. Atmospheric composition forecasting in Europe. *Ann. Geophys.* **2010**, *28* (1), 61−74.

(12) Balk, T.; Kukkonen, J.; Karatzas, K.; Bassoukos, T.; Epitropou, V. A European open access chemical weather forecasting portal. *Atmos. Environ.* **2011**, *45* (38), 6917−6922.

(13) Zhang, Y.; Bocquet, M.; Mallet, V.; Seigneur, C.; Baklanov, A. Real-time air quality forecasting, part I: History, techniques, and current status. *Atmos. Environ.* **2012**, *60*, 632−655.

(14) US EPA. *Fused Air Quality Surfaces Using Downscaling* Website; http://www.epa.gov/esd/land-sci/lcb/lcb_faqsd.html

(15) Beckerman, B. S.; Jerrett, M.; Serre, M.; Martin, R. V.; Lee, S. J.; van Donkelaar, A.; Ross, Z.; Su, J.; Burnett, R. T. A hybrid approach to estimating national scale spatiotemporal variability of PM2.5 in the contiguous United States. *Environ. Sci. Technol.* **2013**, *47* (13), 7233−41.

(16) Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli, X.; Tsai, M. Y.; Kunzli, N.; Schikowski, T.; Marcon, A.; Eriksen, K. T.; Raaschou-Nielsen, O.; Stephanou, E.; Patelarou, E.; Lanki, T.; Yli-Toumi, T.; Declercq, C.; Falq, G.; Stempfelet, M.; Birk, M.; Cyrys, J.; von Klot, S.; Nador, G.; Varro, M. J.; Dedele, A.; Grazuleviciene, R.; Molter, A.; Lindley, S.; Madsen, C.; Cesaroni, G.; Ranzi, A.; Badaloni, C.; Hoffmann, B.; Nonnemacher, M.; Kraemer, U.; Kuhlbusch, T.; Cirach, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Bellander, T.; Korek, M.; Olsson, D.; Stromgren, M.; Dons, E.; Jerrett, M.; Fischer, P.; Wang, M.; Brunekreef, B.; de Hoogh, K. Development of NO_2 and NO*x* land use regression models for estimating air pollution exposure in 36 study areas in Europe—The ESCAPE project. *Atmos. Environ.* **2013**, *72*, 10−23.

(17) Cyrys, J.; Eeftens, M.; Heinrich, J.; Ampe, C.; Armengaud, A.; Beelen, R.; Bellander, T.; Beregszaszi, T.; Birk, M.; Cesaroni, G.; Cirach, M.; de Hoogh, K.; De Nazelle, A.; de Vocht, F.; Declercq, C.; Dedele, A.; Dimakopoulou, K.; Eriksen, K.; Galassir, C.; Grauleviciene, R.; Grivas, G.; Gruzieva, O.; Gustafsson, A. H.; Hoffmann, B.; Iakovides, M.; Ineichen, A.; Kramer, U.; Lanki, T.; Lozano, P.; Madsen, C.; Meliefste, K.; Modig, L.; Moelter, A.; Mosler, G.; Nieuwenhuijsen, M.; Nonnemacher, M.; Oldenwening, M.; Peters, A.; Pontet, S.; Probst-Hensch, N.; Quass, U.; Raaschou-Nielsen, O.; Ranzi, A.; Sugiri, D.; Stephanou, E. G.; Taimisto, P.; Tsai, M. Y.; Vaskovi, E.; Villani, S.; Wang, M.; Brunekreef, B.; Hoek, G. Variation of NO_2 and NO*x* concentrations between and within 36 European study areas: Results from the ESCAPE study. *Atmos. Environ.* **2012**, *62*, 374−390.

(18) Baldasano, J. M.; Querol, X.; Pandolfi, M.; Sanz, M. J.; Diéguez, J. J.; Jiménez-Guerrero, P.; Jorba, O.; Pérez, C.; López, E.; Güereca, P.; Martín, F.; Vivanco, M. G.; Palomino, I. Caliope: An operational air quality forecasting system for the Iberian Peninsula, Balearic Islands, and Canary Islands—First annual evaluation and ongoing developments. *Adv. Sci. Res.* **2008**, *2* (Journal Article), 89−98.

(19) Baldasano, J. M.; Guereca, L. P.; Lopez, E.; Gasso, S.; Jimenez-Guerrero, P. Development of a high-resolution (1 × 1 km^2, 1 h)

emission model for Spain: The high-elective resolution modelling emission system (HERMES). *Atmos. Environ.* **2008**, *42* (31), 7215−7233.
(20) Pay, M. T.; Piot, M.; Jorba, O.; Gasso, S.; Goncalves, M.; Basart, S.; Dabdub, D.; Jimenez-Guerrero, P.; Baldasano, J. M. A full year evaluation of the CALIOPE-EU air quality modeling system over Europe for 2004. *Atmos. Environ.* **2010**, *44* (27), 3322−3342.
(21) Ntziachristos, L.; Samaras, Z. *COPERT III Computer programme to calculate emissions from road transport. Methodology and emission factors (Version 2.1)*; Technical report No 49; 2000.
(22) EEA. *EMEP/CORINAIR Atmospheric Emission Inventory Guidebook−2007*; 2007.
(23) Pay, M. T.; Jimenez-Guerrero, P.; Jorba, O.; Basart, S.; Querol, X.; Pandolfi, M.; Baldasano, J. M. Spatio-temporal variability of concentrations and speciation of particulate matter across Spain in the CALIOPE modeling system. *Atmos. Environ.* **2012**, *46*, 376−396.
(24) Christakos, G.; Bayesian Maximum-Entropy, A. View to the spatial estimation problem. *Math. Geol.* **1990**, *22* (7), 763−777.
(25) De Nazelle, A.; Arunachalam, S.; Serre, M. L. Bayesian maximum entropy integration of ozone observations and model predictions: An application for attainment demonstration in North Carolina. *Environ. Sci. Technol.* **2010**, *44* (15), 5707−5713.
(26) Serre, M. L.; Christakos, G. Modern geostatistics: Computational BME analysis in the light of uncertain physical knowledge—the Equus Beds study. *Stoch. Environ. Res. Risk Assess.* **1999**, *13* (1−2), 1−26.
(27) Baldasano, J. M.; Pay, M. T.; Jorba, O.; Gasso, S.; Jimenez-Guerrero, P. An annual assessment of air quality with the CALIOPE modeling system over Spain. *Sci. Total Environ.* **2011**, *409* (11), 2163−2178.
(28) Poplawski, K.; Gould, T.; Setton, E.; Allen, R.; Su, J.; Larson, T.; Henderson, S.; Brauer, M.; Hystad, P.; Lightowlers, C.; Keller, P.; Cohen, M.; Silva, C.; Buzzelli, M. Intercity transferability of land use regression models for estimating ambient concentrations of nitrogen dioxide. *J. Expo. Sci. Environ. Epidemiol.* **2009**, *19* (1), 107−117.
(29) Gan, W. Q.; Koehoorn, M.; Davies, H. W.; Demers, P. A.; Tamburic, L.; Brauer, M. Long-term exposure to traffic-related air pollution and the risk of coronary heart disease hospitalization and mortality. *Environ. Health Perspect.* **2011**, *119* (4), 501−507.
(30) Nethery, E.; Teschke, K.; Brauer, M. Predicting personal exposure of pregnant women to traffic-related air pollutants. *Sci. Total Environ.* **2008**, *395* (1), 11−22.
(31) Wilton, D.; Szpiro, A.; Gould, T.; Larson, T. Improving spatial concentration estimates for nitrogen oxides using a hybrid meteorological dispersion/land use regression model in Los Angeles, CA and Seattle, WA. *Sci. Total Environ.* **2010**, *408* (5), 1120−1130.