

# Blending Multiple Nitrogen Dioxide Data Sources for Neighborhood Estimates of Long-Term Exposure for Health Research

Ivan C. Hanigan\*

Centre for Air Quality and Health Research and Evaluation, Woolcock Institute of Medical Research, University of Sydney, Sydney, Australia  
University of Canberra, Canberra, Australia

Grant J. Williamson

Centre for Air Quality and Health Research and Evaluation, Woolcock Institute of Medical Research Sydney, Australia & School of Biological Sciences, University of Tasmania, Hobart, Australia

Luke D. Knibbs

Centre for Air Quality and Health Research and Evaluation, Woolcock Institute of Medical Research Sydney, Australia & School of Public Health, The University of Queensland, Herston, Australia

Joshua Horsley

School of Public Health, University of Sydney, Sydney, Australia

Margaret I. Rolfé

School of Public Health, University of Sydney, Sydney, Australia

Martin Cope

Centre for Air Quality and Health Research and Evaluation, Woolcock Institute of Medical Research Sydney, Australia & CSIRO, Melbourne, Australia

Adrian G. Barnett

Institute of Health and Biomedical Innovation & School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia

Christine T. Cowie

Centre for Air Quality and Health Research and Evaluation, Woolcock Institute of Medical Research, University of Sydney; South West Sydney Clinical School, University of NSW & Ingham Institute for Applied Medical Research, Sydney, Australia

Jane S. Heyworth

Centre for Air Quality and Health Research and Evaluation, NESP Clean Air and Urban Landscapes, School of Population and Global Health, The University of Western Australia, Perth, Australia

Marc L. Serre

University of North Carolina, Chapel Hill, United States

Received: June 13, 2017

Revised: September 22, 2017

Accepted: September 26, 2017

Published: September 26, 2017



Bin Jalaludin

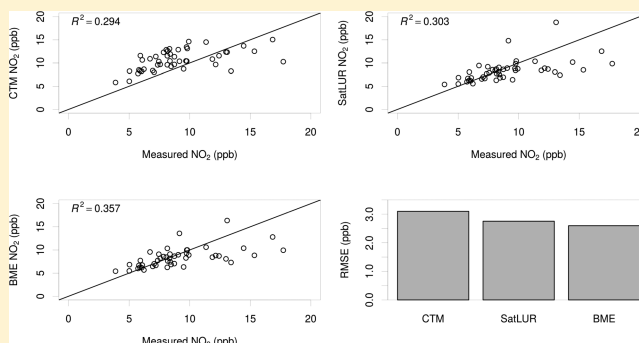
Centre for Air Quality and Health Research and Evaluation, Woolcock Institute of Medical Research, University of Sydney; South West Sydney Clinical School, University of NSW & Ingham Institute for Applied Medical Research, Sydney, Australia

Geoffrey G. Morgan

Centre for Air Quality and Health Research and Evaluation, Woolcock Institute of Medical Research & University Centre for Rural Health, North Coast, School of Public Health, University of Sydney, Sydney, Australia

**S** Supporting Information

**ABSTRACT:** Exposure to traffic related nitrogen dioxide ( $\text{NO}_2$ ) air pollution is associated with adverse health outcomes. Average pollutant concentrations for fixed monitoring sites are often used to estimate exposures for health studies, however these can be imprecise due to difficulty and cost of spatial modeling at the resolution of neighborhoods (e.g., a scale of tens of meters) rather than at a coarse scale (around several kilometers). The objective of this study was to derive improved estimates of neighborhood  $\text{NO}_2$  concentrations by blending measurements with modeled predictions in Sydney, Australia (a low pollution environment). We implemented the Bayesian maximum entropy approach to blend data with uncertainty defined using informative priors. We compiled  $\text{NO}_2$  data from fixed-site monitors, chemical transport models, and satellite-based land use regression models to estimate neighborhood annual average  $\text{NO}_2$ . The spatial model produced a posterior probability density function of estimated annual average concentrations that spanned an order of magnitude from 3 to 35 ppb. Validation using independent data showed improvement, with root mean squared error improvement of 6% compared with the land use regression model and 16% over the chemical transport model. These estimates will be used in studies of health effects and should minimize misclassification bias.



## INTRODUCTION

There is evidence that short-term acute exposures to ambient air pollution cause adverse health effects such as hospitalizations or deaths from cardiovascular disease and respiratory disease.<sup>1</sup> Several studies have shown an association between long-term exposure to air pollution and increased risks of various health outcomes.<sup>2</sup> However, evidence is lacking at the lower end of the exposure-response function, because most existing studies have been conducted in cities where pollutant levels are relatively high.<sup>3</sup> At lower concentrations it is important that errors associated with exposure are minimized. Sydney, Australia's most populous city (~5 million), and its surrounding regions have relatively low concentrations of air pollution compared with similar economically developed cities. For example, the annual average nitrogen dioxide ( $\text{NO}_2$ ) concentration was 7.6 parts per billion (ppb) in 2011 (the focal year of our study) (source: New South Wales Government air pollution data). This is lower than concentrations found in many health studies in other cities around the world. In a review of 15 cohort studies, Hoek et al.<sup>4</sup> found much higher long-term annual average  $\text{NO}_2$  concentrations for their cohorts with a mean of 18.7 ppb (8.9 to 35.0 ppb). Rome, for example, is a similar population size to Sydney but long-term average  $\text{NO}_2$  in that cohort was 23.0 ppb, three times higher than the average in Sydney.

Considerable research effort has been spent evaluating statistical methods which are sensitive to subtle variation in exposures.<sup>5</sup> However, there is no consensus concerning the most appropriate methods to estimate long-term, spatially varying exposures, especially when the risk of exposure misclassification

bias is high. Valid estimates of the spatial distribution of air pollution are required so that exposure-response associations can be investigated in epidemiological studies. Additional research is necessary that can provide more accurate and precise estimates of exposure, and to ascertain the key sources of the uncertainties that remain.

$\text{NO}_2$  is a marker of traffic-related air pollution and more spatially heterogeneous compared with particulate matter in urban areas where traffic is the main source,<sup>6</sup> hence our interest in better exposure estimation for this pollutant. We aimed to provide improved concentration estimates by improving the representativeness of the modeled spatial patterns in order to better estimate exposure at the residential address. This improved characterization of the variation in exposures should produce more accurate estimates of risk-response functions.

Progress in the development of spatially resolved air pollution exposure assessment models has become increasingly sophisticated, both nationally and internationally, using new data sources such as satellite data and new statistical techniques. For instance Akita et al.,<sup>7</sup> Xu et al.,<sup>8</sup> and Buteau et al.<sup>9</sup> have used Bayesian blending methods to improve air pollution models in North America. Meanwhile other approaches have developed such as that used by Knibbs et al.<sup>10</sup> who produced improved  $\text{NO}_2$  air pollution maps for Australia using a Land Use Regression (LUR) model that combined satellite images with ground based predictor variables. A different approach used by Cope et al.<sup>11</sup> in Australia applies physics and chemistry principles to model the dispersion of emissions to predict pollution. Relatedly, Shaddick et al.<sup>12</sup> recently developed a global Bayesian-based  $\text{PM}_{2.5}$  model.

While several novel air pollution exposure assessment methods are therefore now available, few studies have sought to combine methods to leverage the best attributes of each, particularly in low exposure settings where uncertainty of the measurements will be a key challenging point (i.e., close to detection limits). Therefore, the purpose of this study was to blend together information from multiple data sources to estimate annual average  $\text{NO}_2$  concentrations at the level of Sydney neighborhoods (small areas incorporating housing within hundreds of meters of roads). We implemented the Bayesian Maximum Entropy (BME) approach<sup>13</sup> to integrate modeled predictions with measurements of  $\text{NO}_2$ , to provide a new spatial model. Our goal was to blend data from multiple sources based on BME (a method which has been recently used internationally<sup>7</sup>), but our aim was not to compare BME against any other blending methods. To date no Bayesian blending methods have previously been implemented in the Sydney study area specifically, and only occasionally in air pollution modeling more generally, therefore this work is filling a key knowledge gap.

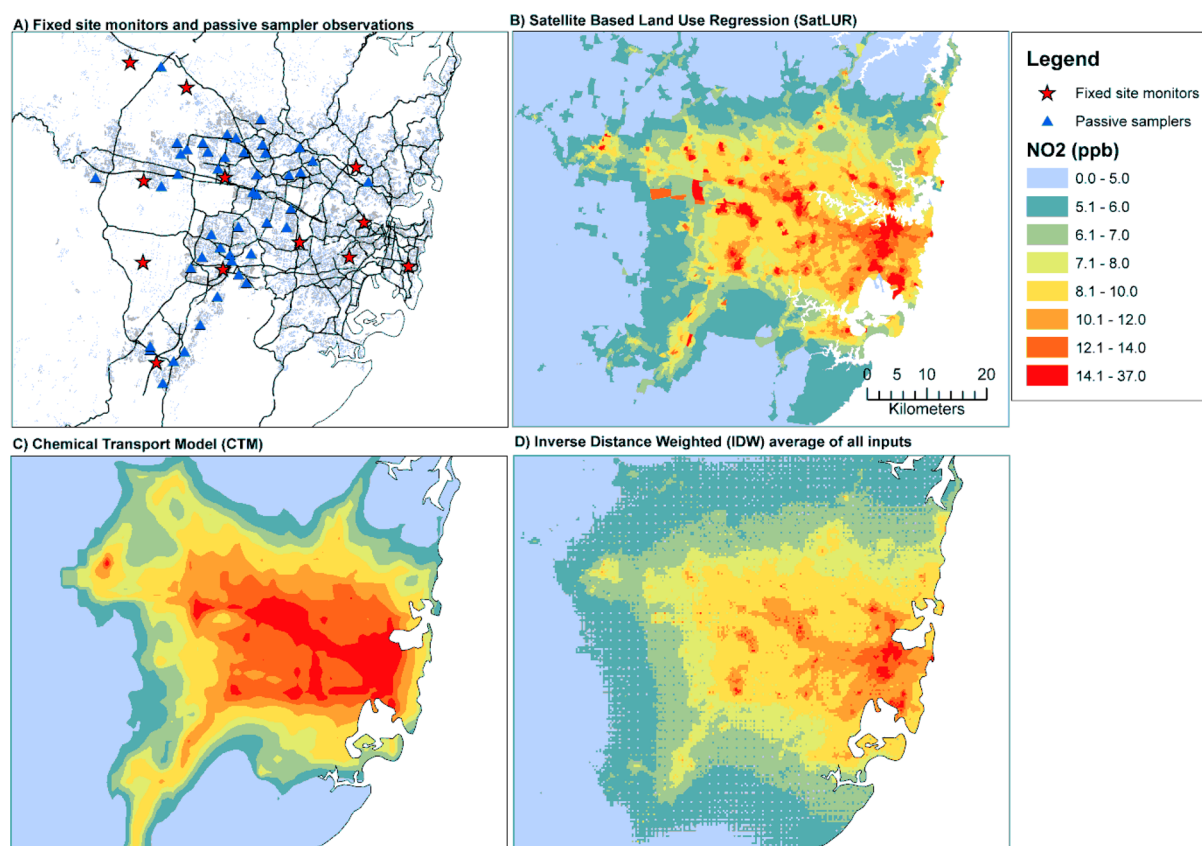
## METHODS AND MATERIALS

Our approach involved blending fixed-site monitor data with a satellite-based Land Use Regression model (SatLUR) and Chemical Transport Model (CTM). The resulting estimates were validated against passive samplers because they were independent of all other data inputs. Our input data are shown in panels A, B, and C of Figure 1. Panels B and C show that  $\text{NO}_2$  varies substantially across Sydney, and SatLUR estimates display greater variation than CTM. Panel D shows the derived trend surface of all inputs, used to adjust for regional spatial autocorrelation prior to our spatial modeling.

**Study Area and Period.** Our study area was the Sydney greater metropolitan region of New South Wales (NSW), Australia. This is the most populous region in Australia. The study area covered over 17 000  $\text{km}^2$ , and had an estimated population in 2015 of 5.8 million (Australian Bureau of Statistics, LGA population estimates 2015). Our study period comprised the years 2011 (for the predictive modeling) and 2013–2014 (for the validation data set) which we selected based on the availability of pollution data and because it coincides with the five-year follow up of the longitudinal “4SandUp” cohort study with a sample size of 99 317 persons in the greater Sydney region which adds to the applicability of the study to support health research (see [www.saxinstitute.org.au](http://www.saxinstitute.org.au) for more information on the cohort). The comparability of the two periods was supported by our exploratory analyses that showed that the spatial pattern of annual averages between these periods was stable (see the [Supporting Information \(SI\)](#) document).

The Sydney region has relatively low concentrations of air pollution when compared with similar economically developed cities around the world; the mean of annual average daily  $\text{NO}_2$  concentrations at 14 fixed-site monitors in 2011 was 7.6 ppb, ranging from 1.8 to 13.0 ppb at individual monitors (NSW Office of Environment and Heritage monitor data, <http://www.environment.nsw.gov.au/AQMS/search.htm>).

**Fixed-Site Monitoring  $\text{NO}_2$  Measurements.** During the study period The NSW Office of Environment and Heritage (OEH) monitored the concentrations of  $\text{NO}_2$  at 14 sites across the Sydney Metropolitan area for regulatory compliance. We obtained daily average ground-level  $\text{NO}_2$  measurements for January to December 2011 from the OEH data portal (<http://www.environment.nsw.gov.au/AQMS/search.htm>).



**Figure 1.** Maps of (A)  $\text{NO}_2$  monitoring sites with roads, (B) Satellite based Land Use Regression model (SatLUR), (C) Chemical Transport Model (CTM), and (D) the inverse distance weighted smoothed surface used as an offset for our Bayesian Maximum Entropy (BME) modeling.

NO<sub>2</sub> concentrations were measured using the standard chemiluminescence methods. The measurements had undergone quality assurance procedures via internal processing prior to public release of data online (see <http://www.environment.nsw.gov.au/resources/air/nepm-air-monitoring-plan.pdf> for details).

We computed annual averages for each of the 14 monitors using data for all days where concentrations had been recorded between January and December 2011. All sites had more than 90% of days where concentrations had been recorded. The annual average was calculated using the mean of 24 h daily averages as originally provided.

**Satellite-Based Land Use Regression Model.** We used a national satellite-based NO<sub>2</sub> Land Use Regression model (SatLUR) developed by Knibbs et al.<sup>10</sup> The model development and validation are described extensively elsewhere (Knibbs et al.;<sup>10</sup> Knibbs et al.<sup>14</sup>). The SatLUR model used data from fixed-site government-run monitors from networks around Australia as the outcome variable, and the model incorporated high-resolution land-use predictors across Sydney (including roads, impervious surfaces, industrial point sources, and industrial land use) and satellite observations of NO<sub>2</sub> from the ozone monitoring instrument aboard the Aura satellite. Two estimates of SatLUR derived NO<sub>2</sub> concentrations were available: one based on surface data (estimated using surface-to-column ratios from the WRF-Chem model), and the second based on the estimates of total tropospheric NO<sub>2</sub> column density. The two LUR models developed using the separate satellite estimates produced similar results but the column-density model estimates were used in this study due to ease of implementation compared with the surface model.<sup>10</sup> The column model captured up to 81% of spatial variability in annual average NO<sub>2</sub> concentrations when cross-validated, and up to 66% when validated against an independent set of NO<sub>2</sub> measurements.<sup>14</sup> Predicted values of NO<sub>2</sub> were estimated for Mesh Block centroids from the 2011 Census. Mesh Blocks were the smallest geographical statistical unit used by the Australian Bureau of Statistics in the 2011 census. In NSW in 2011 Mesh Blocks had a median population of 83 (range 3 to 1,932 persons) and mean area of 8.8 km<sup>2</sup> (range 0.0005 to 12,541 km<sup>2</sup>) (Australian Bureau of Statistics data).

**Chemical Transport Model.** Chemical Transport Models (CTMs) were used to estimate concentrations of NO<sub>2</sub> as one of our data inputs. The CTMs used emissions data and physical dispersion modeling to predict NO<sub>2</sub> air pollution concentrations. We used NO<sub>2</sub> predictions from the CTMs developed by Cope et al.<sup>11</sup> for July 2010 through June 2011. The model comprised a prognostic meteorological model, the 2008 NSW OEH air emissions inventory (developed to describe the emissions from the shipping industry as well as traffic and industry emissions), and a chemical transport and particle dynamics model.<sup>11</sup> The model used a 1 × 1 km grid cell for the inner region around the central urban area of Sydney and a 3 × 3 km square grid cell for the larger area (Greater Sydney Metropolitan Region). Daily predicted NO<sub>2</sub> average concentrations from the CTM were used to calculate monthly concentrations, which were then combined to calculate annual average NO<sub>2</sub> concentrations.

**Passive Samplers.** The resulting estimates from the BME model were validated against passive samplers data because they were independent of all other data inputs. Forty-seven Ogawa passive samplers were deployed at sites located within the metropolitan area of the city that were selected to capture within-city and near-road variability in NO<sub>2</sub>. Samples were taken for two week periods during July–August (winter) 2013,

November–December (summer) 2013, and March–April (autumn) 2014. We used data for this period because no other data were collected during 2011. The fixed-site monitors showed minimal change in concentration between 2011 and 2014 (see SI document). Passive sampling locations were chosen to capture the full range of expected NO<sub>2</sub> concentrations across the city, based on knowledge of land use, traffic conditions and industrial sources. Passive sampler siting followed the method as outlined in the European Study of Cohorts for Air Pollution Effects (ESCAPE) protocol.<sup>15</sup> For more details see the SI document.

Conversion to annual mean NO<sub>2</sub> was conducted using standard protocols.<sup>14</sup> Passive sampler measurements were adjusted to an estimated annual mean using the ratio of mean NO<sub>2</sub> measured by fixed-site monitors during each measurement period, compared with its annual mean. The ratio was calculated based on three separate fixed-site monitors in the study area. The selection criteria for the fixed-site monitors and the adjustment process are described in the original paper.<sup>14</sup>

**Bayesian Maximum Entropy Spatial Model.** We implemented the BME geostatistical approach.<sup>13</sup> BME is a statistical framework for spatial estimation that can process a wide range of general and site specific knowledge based on entropy maximization and epistemic Bayesian processing rules. We used the BMELib<sup>13</sup> software which is a numerical implementation of the BME framework in the case where the general knowledge includes mean and covariance, and the site specific knowledge consist of hard and soft data. Where we talk specifically about the implementation of BME, we refer to the numerical package in the BMEprobaMoments function found in the BMELib<sup>13</sup> software. We created an exposure surface by blending the measured concentrations of NO<sub>2</sub> at the fixed-site monitors with the two models (SatLUR and CTM). These data all included various sources of uncertainty such as measurement error, emission inventory assumptions, and spatial uncertainty about the geographical predictors (such as industrial land use classes). Each data point therefore contained information about the uncertainty that the BMELib model used when estimating the concentration.

The model approach is similar to spatial interpolation, which is a method used to generate estimates across an entire region (e.g., the city of Sydney) from a scattered set of data points in that region (e.g., monitors and model estimation points). BMELib is similar to the kriging algorithm common in GIS software which estimates the interpolated values using a statistical model governed by the covariance relationships between data points across space.

The BME framework allows prior information to be weighted, based on a Probability Density Function (PDF). A PDF represents the known or assumed likelihood of all the possible values. The PDF is used to specify the probability of the value falling within a particular range of values. PDFs can be combined to give estimates that are based on all the available knowledge. To do this, the BME framework uses maximum entropy, which is a concept used in information theory. Maximum entropy involves a modest claim being made regarding certainty about the probabilities of expected values, and PDFs can be combined to achieve a maximally informative model, given the uncertain general knowledge base. Thus, the “informativeness” requirement is mathematically expressed in terms of a maximum entropy condition of the type based on the Shannon information concept.<sup>13</sup> The estimated PDF is the one that best represents the current state of knowledge asserted by the prior PDFs. This is referred to as the posterior PDF.



When the simplest constraints are used in BME, no PDF is assigned as all the data points are assumed to be the “true” value at that spatial location. The BMEprobaMoments algorithm then reduces to the kriging algorithm. However, if additional information is included to apply probabilistic constraints then each datum is assigned a PDF. The PDF explicitly models the measurement error in each datum using a distribution to reflect uncertainty that is centered on the observed datum. Incorporating these distributions is straightforward in a Bayesian paradigm and including measurement error in the model more closely captures reality. This should give a better estimate of the mean pollutant concentration as well as be better at capturing uncertainty than a standard geostatistical model.<sup>16</sup>

In this study, we applied probabilistic constraints on our prior expectations of the uncertainty around that central estimate (i.e., the range and shape of the PDF). Then using the BMelib program,<sup>13</sup> these prior PDFs are blended together and a posterior PDF produced. The mean of the BME posterior PDF minimizes the mean square estimation error and was used as our estimate of NO<sub>2</sub> in the output result.

We categorized our data into two groups: (i) hard data: the fixed-site monitor measurements with high precision or minimal uncertainty; and (ii) soft data: the SatLUR and CTM model estimates with an uncertainty characterized by a PDF.

**Parametrization for Prior Probability Density Functions of Soft Data.** We created the PDFs with a triangular distribution where the mean was set to the concentration estimated by either of our two models (the SatLUR or CTM model), and the corresponding range (the minimum and maximum for the triangular distribution) was estimated from the uncertainties associated with the expected values at each location, based on how well they reproduce the colocated observed values from the NO<sub>2</sub> passive samplers within each location. This meant that areas with less observed measurement error had a narrower PDF. We used the root mean squared error (RMSE) of these paired predictions and observations to determine the range of the prior PDFs. This was 2.7 ppb for the SatLUR and 3.1 ppb for the CTM.

The SatLUR estimates for Mesh Block centroids were used as one type of soft data input. We extended the spatial coverage of these by also incorporating a grid of points equally spaced at 2.5 km across the region (because all points within a Mesh Block give the same value), so that each estimation node in the modeled output would have guaranteed data inputs within the search radius used by the BMEprobaMoments algorithm.

**Detrending Data Inputs Using a Global offset.** Similar to kriging, BMelib relies on the assumption that the random errors have zero mean and constant variance, and so trend removal can help satisfy assumptions of normality and stationarity. The first stage in our analysis was to develop a smoothed trend surface. We used an inverse distance weighted average smoother with 300 × 300 m grid cell resolution as an offset to detrend the input data sets. We chose this resolution so that the smooth trend should produce low variance in the residuals while maintaining adequate residual autocorrelation. As Xu et al.<sup>8</sup> point out this is because when a trend surface is smoother and less variable across space, the transformed residuals retain more variability. Conversely as the surface becomes less smooth the residuals are less autocorrelated. Residuals for each data input and this offset surface were calculated, along with the upper and lower limits of the PDFs.

**Determination of the Spatial Covariance Model.** The experimental covariance model was estimated based on

the residuals of all hard and soft data within a 10 × 10 km square that included high expected concentrations of NO<sub>2</sub>. We used the ArcMap 10.2.1 kriging cross-validation tool to optimize the parameters for an exponential spatial covariance model. The covariance sill reported by ArcGIS was 1.04. The covariance range was 1238.5 m. Please see the SI document for further details on the spatial covariance model.

**Estimation of the BME Posterior PDFs.** We computed the BME posterior PDFs at the nodes of a two-dimensional network of estimation points on the Australian Map Grid (GDA94, zone 56), with regular spacing every 100 m within 10 × 10 km squares. We chose this resolution as a trade-off between computational time and our aim to model NO<sub>2</sub> at the level of street addresses. We did this within these subsets to improve the efficiency of computational work and allow more rapid testing and revision of the parametrizations. We then combined all resultant subset squares into a mosaic for the full study region. We used the mean estimate of NO<sub>2</sub> from the posterior PDFs as our predicted concentration. During computation we allowed the inclusion of data points external to the squares out to a buffer distance of 1200 m to minimize any discontinuity at the boundaries when the squares were combined. This buffer distance means that all estimation nodes within each subset square has a complete set of input data points within the search radius, even when the point is on the boundary of the subset square. We set the maximum search radius for each estimation node at 1,240 m so that the experimental covariance range was met, and to ensure that each node would include some data points (e.g., soft data for SatLUR were at the equally spaced 2500 m grid in addition to each of the Mesh Block centroids). We assumed that the mean was constant over the estimation neighborhood, after assessing the frequency distribution of our detrended residuals from the global offset. The BMelib model returned estimates that represent the mean value which minimizes the estimation error variance of the posterior PDF. This value is on the scale of the residuals and so when added to the global offset the result provides the estimated long-term NO<sub>2</sub> concentrations.

**Validation Methods.** We used an external validation method to assess our model. Our validation data set comprised NO<sub>2</sub> passive sampler observations (adjusted for seasonality using the method described in Knibbs et al.<sup>14</sup>). In this approach we assessed the agreement between the modeled estimates and a set of unrelated, independent measurements. We calculated the RMSE for our BMelib model as well as the two models used as data inputs (SatLUR and CTM).

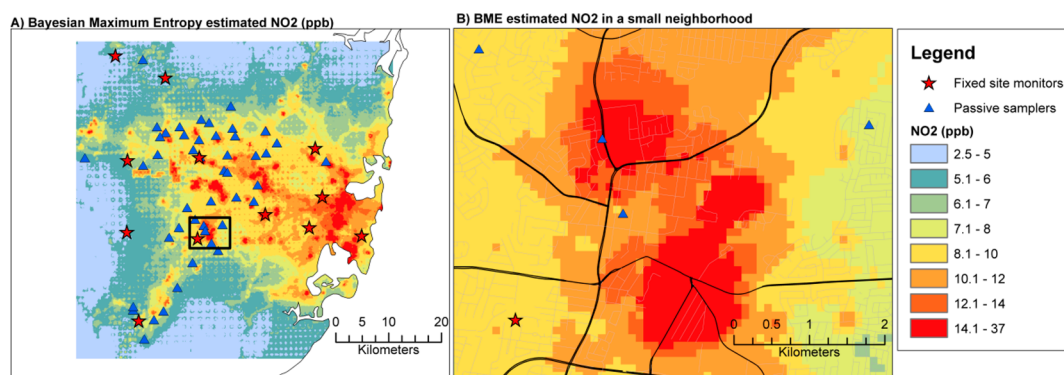
We used Matlab R2016a, BMelib2.0b, ArcGIS 10.2.1, and R version 3.2.5 for data preparation and analyses.

## RESULTS AND DISCUSSION

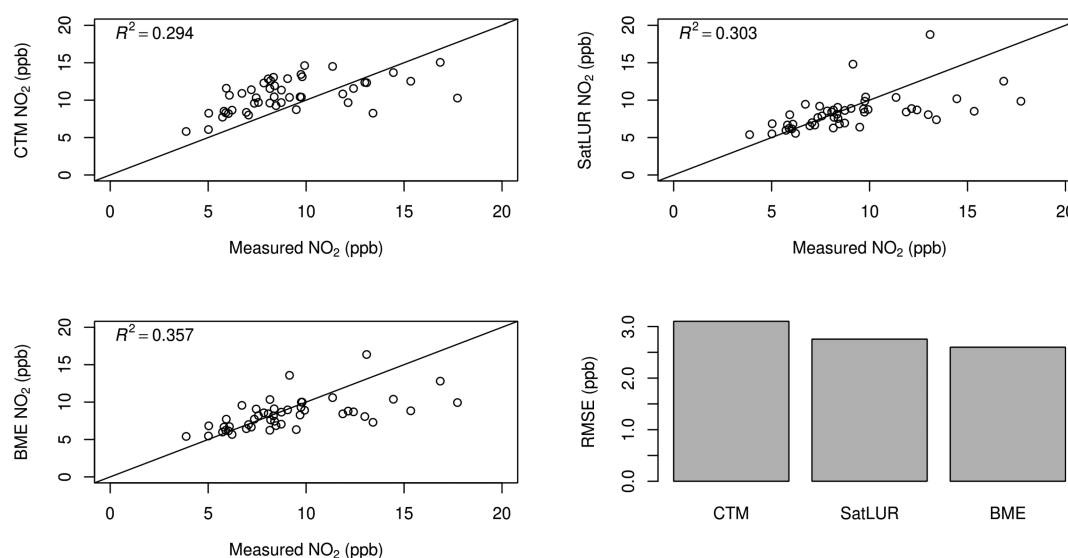
The BMelib model estimated annual average NO<sub>2</sub> concentrations are shown in Figure 2.

**Validation Statistics.** We compared the RMSE for the CTM, SatLUR, and BMelib output (Figure 3D). In all models there was an indication of under-prediction at levels above 10 ppb (Figure 3A–C).

**Key Results and Comparison with Other Studies.** We found that the BMelib resulted in a reduced RMSE against the validation data set when compared to the RMSE for the CTM and SatLUR. See Table 1. The magnitude of the difference between the three RMSE was 0.5 ppb (BMelib compared to CTM) and 0.2 (BMelib compared to SatLUR). As a percentage difference this was a 16.2% improvement over the CTM and 5.7% improvement over the SatLUR.



**Figure 2.** Maps of NO<sub>2</sub> annual average (ppb) estimated by Bayesian Maximum Entropy blending of various data inputs for (A) Sydney and (B) Liverpool (a built-up southwestern suburb of Sydney).



**Figure 3.** Scatter plots of measured versus estimated NO<sub>2</sub> annual averages (ppb) estimated by (A) chemical transport model (CTM), (B) satellite-based land use regression (SatLUR) and (C) Bayesian maximum entropy (BME) integration of fixed-site monitors, CTM and SatLUR data. Panel D shows the RMSE in ppb for each modeling approach.

**Table 1. Root Mean Squared Error for Each Model against the Validation Dataset**

modeling method	root mean squared error (ppb)
chemical transport model (CTM)	3.1
satellite-based land use regression (SatLUR)	2.8
Bayesian maximum entropy (BME)	2.6

A similar study by Akita et al.<sup>7</sup> in Spain found larger improvements in predictive accuracy of NO<sub>2</sub> using BME compared with their CTM and LUR (not Satellite-based). Akita et al. found that their BME model reduced the RMSE for validation data by approximately 40% from both their CTM and LUR models, and over 60% from an ordinary kriging model of sensor data. A possible reason for this is that NO<sub>2</sub> concentrations in Sydney were lower than concentrations measured in Spain. Our maximum predicted concentration was 35 ppb in Sydney, whereas Spain had a maximum concentration of 58 ppb (converted from the original predicted NO<sub>2</sub> value reported in Akita et al. using the conversion factor of 1.9125).

Alternatively, the difference may reflect higher underlying precision of our input model data (our CTM and SatLUR). The CTM used by Akita et al. was the CALIOPE Air Quality Forecast Modeling (AQFM) system and was reported to have a

RMSE against the fixed-site monitors of 6.9 ppb. In contrast, the CTM from Cope et al.<sup>11</sup> which we used has a smaller RMSE of 3.1 ppb against our passive sampler validation data set. Therefore, the BME model RMSE of 4.0 ppb (7.6  $\mu\text{g}/\text{m}^3$ ) reported by Akita et al. represented a substantial improvement over their CTM. An additional likely explanation is that there were only 14 observations used as hard data here, while there were a lot more observations ( $N = 80$ ) in the Akita et al. model, therefore we would expect a much greater drop in RMSE in that study than here. While our relative improvements were more modest, they are still important for epidemiological studies because the small relative risks arising from air pollution exposures (e.g., of the order of  $\sim 1.04$ ), are assigned to large population numbers, which results in a large population attributable risk from air pollution.

**Strengths and Weaknesses.** Our project demonstrates an alternative to similar applications of the BME in two important ways. First, we detrended the input data sets using an inverse distance weighted smoother to obtain our offset, which is a robust alternative to other options and requires fewer assumptions regarding the spatial autocorrelation structure. To the best of our knowledge this is a new contribution to the approach. Second, we defined the range of the prior probability intervals using a triangular density function based on the agreement of the input

data with the external validation data set. This is an important parameter and using our approach we estimate the uncertainty against an independent set of external observed data, rather than from internal model cross-validation.

In general, the strengths of the BME approach include the ability to combine all available data inputs despite their different temporal and spatial attributes into one estimate, making use of all available prior information regarding uncertainties. Weaknesses of our study include that the workload for preparing the data and conducting the analysis was very labor intensive. In addition, the actual calculations themselves were computationally intensive and required many hours of computer time to complete.

A limitation of this study was the incomplete seasonal coverage of the passive sampler data used as the external validation. The seasonal adjustments made by Knibbs et al.<sup>14</sup> of the three  $\times$  fortnightly samples is a widely used method for passive samplers which is likely to make them representative of annual (long-term) averages. Even so, the potential remains that the sampled seasonal coverage of the measurements might not fully capture the annual average conditions.

One major conclusion of our study is that the blending done in the BMelib model showed an improvement in predictive performance in comparison to either the CTM or SatLUR alone. However, the three modeled outputs are not exactly comparable because the SatLUR was developed with a training data set that spanned multiple cities across the Australian continent. It is possible that a SatLUR model built exclusively from the Sydney data might perform better than this national model,<sup>17–19</sup> however no city-specific SatLUR currently exists that could be used for our study, so we have made use of all information at the spatial extent and resolution that was available. However, this remains a limitation and it is not possible to conclude that the BMelib is the optimal method based on our study. Rather, we can only conclude that BMelib with a SatLUR model at a continental extent had the greatest performance to-date as measured by our validation set for Sydney.

**Future Directions.** Future research should attempt to ameliorate a key weakness of any blending model, which is that they cannot overcome limitations that are common across all the data inputs (i.e., observations, SatLUR and CTM). For example, it is likely that  $\text{NO}_2$  in reality may be much more spatially variable than what is displayed in any of the input data sets<sup>20</sup> and so the spatial variability of the original data inputs places limits on the extent that Bayesian blending can produce spatial variability in the predictions. However, the aim of such blending is to get an estimate that is at least as good as the best of the available inputs, or even better. BMelib can theoretically achieve this outcome by incorporating all available information and accounting for uncertainty to combine all the available data into one single exposure metric, without losing information from each individual data set.

There are a number of parameters in the model specification that might be optimized in future research. One of the most important assumptions relates to the range of the interval used to construct the PDF for the soft data. Also, the shape of the PDF used may be important. We used a triangular PDF because of the ease of specification (only the mean, minimum and maximum are needed) however other functional forms such as a normal distribution may result in further improvements if they capture the prior information better, but this was beyond the scope of this study. In future research we aim to experiment with ways that the prior PDF parameters can be used to optimize the predictions

based on validation (i.e., change the parameters and compare the impact these changes have on the validation statistics).

**Implications.** This paper presents a combination of novel approaches to the operationalization of the BME spatial modeling framework for air pollution estimation and shows the validation statistics for this approach provided better results than some more classical alternatives.

## ■ ASSOCIATED CONTENT

### ⑤ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.est.7b03035](https://doi.org/10.1021/acs.est.7b03035).

BME\_NO2\_Sydney\_SL.docx describes the full details of our data preparation, parameter selection procedure and some technical detail about our data sets (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Phone: +61 2 6201 5298; fax: +61 2 6201 5999; e-mail: [ivan.hanigan@canberra.edu.au](mailto:ivan.hanigan@canberra.edu.au).

### ORCID

Ivan C. Hanigan: 0000-0002-6360-6793

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We gratefully acknowledge contributions from individuals within Government departments and private institutes involved in the measurement and modeling of air pollutant concentrations. Funding for this work was provided by the Centre for Air quality and health Research and evaluation (CAR), an Australian National Health and Medical Research Council Centre for Research Excellence, administered by the Woolcock Institute of Medical Research, University of Sydney.

## ■ REFERENCES

- (1) Pope, C. A.; Dockery, D. W. Health effects of fine particulate air pollution: lines that connect. *J. Air Waste Manage. Assoc.* **2006**, *56* (6), 709–742.
- (2) Brunekreef, B. Health effects of air pollution observed in cohort studies in Europe. *J. Exposure Sci. Environ. Epidemiol.* **2007**, *17*, S61–S65.
- (3) Beelen, R.; Raaschou-Nielsen, O.; Stafoggia, M.; Andersen, Z. J.; Weinmayr, G.; Hoffmann, B.; Wolf, K.; Samoli, E.; Fischer, P.; Nieuwenhuijsen, M.; et al. Effects of long-term exposure to air pollution on natural-cause mortality: An analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* **2014**, *383* (9919), 785–795.
- (4) 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* (1), 43.
- (5) Peng, R. D.; Dominici, F. *Statistical Methods for Environmental Epidemiology with R. A Case Study in Air Pollution and Health*; Springer Science & Business Media: New York, 2008.
- (6) Health Effects Institute Panel on the Health Effects of Traffic-Related Air Pollution. *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, And Health Effects*. Health Effects Institute Special Report 17; Boston, MA, 2010.
- (7) Akita, Y.; Baldasano, J. M.; Beelen, R.; Cirach, M.; de Hoogh, K.; Hoek, G.; Nieuwenhuijsen, M.; Serre, M. L.; de Nazelle, A. Large Scale Air Pollution Estimation Method Combining Land Use Regression and



Chemical Transport Modeling in a Geostatistical Framework. *Environ. Sci. Technol.* **2014**, *48* (8), 4452–4459 (8).

(8) Xu, Y.; Serre, M. L.; Reyes, J.; Vizuete, W. Bayesian Maximum Entropy integration of ozone observations and model predictions: a national application. *Environ. Sci. Technol.* **2016**, *50* (8), 4393–4400.

(9) Buteau, S.; Hatzopoulou, M.; Crouse, D. L.; Smargiassi, A.; Burnett, T.; Logan, T.; Deville, L.; Goldberg, M. S. Comparison of spatiotemporal prediction models of daily exposure of individuals to ambient nitrogen dioxide and ozone in Montreal, Canada. *Environ. Res.* **2017**, *156* (2), 201–230.

(10) Knibbs, L. D.; Hewson, M. G.; Bechle, M. J.; Marshall, J. D.; Barnett, A. G. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. *Environ. Res.* **2014**, *135*, 204–211.

(11) Cope, M.; Keywood, M.; Emmerson, K.; Galbally, I.; Boast, K.; Chambers, S.; Cheng, M.; Crumeyrolle, S.; Dunne, E.; Fedele, R.; et al. *The Centre for Australian Weather and Climate Research Sydney Particle Study-Stage-II*; CSIRO and Office of Environment and Heritage: Sydney, Australia, 2014; ISBN: 978-1-4863-0359-5.

(12) Shaddick, G.; Thomas, M. L.; Jobling, A.; Brauer, M.; van Donkelaar, A.; Burnett, R.; Chang, H. H.; Cohen, A.; Van Dingenen, R.; Dora, C.; et al. Data Integration Model for Air Quality: A Hierarchical Approach to the Global Estimation of Exposures to Ambient Air Pollution. *J. R. Stat. Soc. C* **2017**, DOI: [10.1111/rssc.12227](https://doi.org/10.1111/rssc.12227).

(13) Christakos, G.; Bogaert, P.; Serre, M. *Temporal GIS: Advanced Functions for Field-Based Applications*; Springer-Verlag: New York, 2002.

(14) Knibbs, L. D.; Coorey, C. P.; Bechle, M. J.; Cowie, C. T.; Dirgawati, M.; Heyworth, J. S.; Marks, G. B.; Marshall, J. D.; Morawska, L.; Pereira, G.; et al. Independent Validation of National Satellite-Based Land-Use Regression Models for Nitrogen Dioxide Using Passive Samplers. *Environ. Sci. Technol.* **2016**, *50* (22), 12331–12338.

(15) Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli, X.; Tsai, M.-Y.; Künzli, N.; Schikowski, T.; Marcon, A.; et al. Development of NO<sub>2</sub> and NO<sub>x</sub> land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. *Atmos. Environ.* **2013**, *72* (2), 10–23.

(16) Diggle, P. J.; Tawn, J. A.; Moyeed, R. A. Model-Based Geostatistics. *J. R. Stat. Soc. C* **1998**, *47* (3), 299–350.

(17) Larkin, A. J.; Geddes, A.; Martin, R. V.; Xiao, Q.; Liu, Y.; Marshall, J. D.; Brauer, M.; Hystad, P. A Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. *Environ. Sci. Technol.* **2017**, *51* (12), 6957–6964.

(18) Hoek, G.; Eeftens, M.; Beelen, R.; Fischer, P.; Brunekreef, B.; Boersma, K. F.; Veeffkind, P. Satellite NO<sub>2</sub> data improve national land use regression models for ambient NO<sub>2</sub> in a small densely populated country. *Atmos. Environ.* **2015**, *105*, 173–180.

(19) Bechle, M. J.; Millet, D. B.; Marshall, J. D. National Spatiotemporal Exposure Surface for NO<sub>2</sub>: Monthly Scaling of a Satellite-Derived Land-Use Regression, 2000–2010. *Environ. Sci. Technol.* **2015**, *49* (20), 12297–305.

(20) Apte, J. S.; Messier, K. P.; Gani, S.; Brauer, M.; Kirchstetter, T. W.; Lunden, M. M.; Marshall, J. D.; Portier, C. J.; Vermeulen, R. C. H.; Hamburg, S. P. High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data. *Environ. Sci. Technol.* **2017**, *51* (12), 6999–7008.