Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: A National Application

Yadong Xu, Marc L. Serre, Jeanette Reyes, and William Vizuete*

University of North Carolina, Chapel Hill, North Carolina 27599, United States

Supporting Information

ABSTRACT: To improve ozone exposure estimates for ambient concentrations at a national scale, we introduce our novel Regionalized Air Quality Model Performance (RAMP) approach to integrate chemical transport model (CTM) predictions with the available ozone observations using the Bayesian Maximum Entropy (BME) framework. The framework models the nonlinear and nonhomoscedastic relation between air pollution observations and CTM predictions and for the first time accounts for variability in CTM model performance. A validation analysis using only noncollocated data outside of a validation radius rv was performed and the R² between observations and re-estimated values for two daily metrics, the daily maximum 8-h average (DM8A) and the daily 24-h average (D24A) ozone concentrations, were obtained with the OBS scenario using ozone observations only in contrast with the RAMP and a Constant Air Quality Model Performance (CAMP) scenarios. We show that, by accounting for the spatial and temporal variability in model performance, our novel RAMP approach is able to extract more information in terms of R² increase percentage, with over 12 times for the DM8A and over 3.5 times for the D24A ozone concentrations, from CTM predictions than the CAMP approach assuming that model performance does not change across space and time.

1. INTRODUCTION

According to EPA’s newly released Integrated Science Assessment for tropospheric ozone,¹ the evidence of public health impacts on populations residing in areas with elevated ozone levels for prolonged periods are still uncertain. A better understanding of the adverse health effects to chronic ozone requires accurate exposure estimates at multiple temporal scales and at fine spatial resolutions. Estimates of ozone concentrations typically rely on environmental data collected from two sources: monitoring networks and air quality chemical transport models (CTM). The first source gives measurement concentrations for a long temporal time, but only at a point where the monitor is located. The CTM provides predictions for all locations, but is an average concentration based on the spatial resolution of the model grid cell. Further, given the intensive resources needed to build a CTM, the numbers of days that are simulated are limited. Several categories of data integration methods, including Kalman filter methods,² variational methods,³ optimal interpolation,⁴ and Bayesian methods⁵−⁸ have been developed to integrate these two types of data and rely on their individual strengths to build a more refined air pollution estimate. In this work, we choose the BME method of modern geostatistics, a knowledge-processing framework, because of its advantage of integrating a wide variety of nonlinear, non-Gaussian knowledge bases.

We developed our data integration approach to obtain two metrics of ozone estimates, the DM8A and D24A ozone concentrations. Both of these metrics are commonly used in epidemiology studies.⁹−¹¹ Several Bayesian inference approaches⁶−⁸ provide a sophisticated statistical framework for the data integration of ozone observations and model predictions and production of multiple time averaged estimates. These approaches share the following characteristics: they parametrize the relationship between air pollution observations and predictions, use kriging to obtain air pollution estimates for any given value of the

Received: January 7, 2016
Accepted: March 21, 2016
Published: March 21, 2016

DOI: 10.1021/acs.est.6b00096
parameters, and use Bayesian inference to obtain air pollution estimates that account for parameter uncertainty. These methods, however, have the following limitations: they assume that the relationship between air pollution observations and predictions is linear and homoscedastic, they share the linear limitations of the kriging estimator, and they have a high numerical cost.

To overcome these limitations, de Nazelle et al.\(^5\) introduced an approach based on the nonlinear extension of kriging provided by the Bayesian Maximum Entropy (BME) method of modern spatiotemporal geostatistics.\(^{12,13}\) This approach uses a nonparametric methodology that fully accounts for the nonlinearity and nonhomoscedasticity of the relationship between air pollution observations and predictions. Their application of this approach showed that the BME method accounts for the nonparametric methodology that fully accounts for the nonlinearity and nonhomoscedasticity of the relationship between air pollution observations and predictions. That study applied the BME framework to integrate ozone bases that are out of the reach of kriging-based methods. That 1s. CAMx is a publicly available Eulerian grid-based model that can address tropospheric ozone, acid deposition, visibility, fine particulates, and other air pollutants issues in the context of a “1 atmosphere” perspective. The modeling simulations were created by the U.S. EPA as base-case simulations in their analysis of the final Transport Rule. These air quality-modelling simulations used the CAMx version v5.30 with gas-phase chemistry mechanism CB05, and also refined meteorological and emission fields for the year 2005 across the United States. Detailed model configurations and evaluation are discussed elsewhere.\(^{12}\) The hourly model predictions were used to compute the DM8A and D24A ozone concentrations at each grid cell. These CTM-predicted daily ozone concentrations are used to construct the soft data, as secondary information with uncertainties, consisting of the expected values of the daily concentrations and the uncertainties associated with the expected values at each grid cell. The details of soft data construction are described in Section 3.3.

3. METHODS

3.1. BME Estimation Methodology. BME is a modern geostatistical method\(^3\) for spatial–temporal interpolation that incorporates information from many different data sources. The implementation and performance of BME have been detailed in other works,\(^{12,17–19}\) and its application to the integration of O3 observations and model predictions was described by de Nazelle et al.\(^5\) In short, we model the (offset-removed) transform, which is a commonly used deterministic transformation,\(^{15}\) of air pollution as a Space/Time Random Field (S/TRF) \(X(p)\) at space/time coordinate \(p = (s,t)\), where \(s\) is the spatial coordinate and \(t\) is time. Our notation for S/TRFs consists of denoting a random variable \(X\) in capital letters, its realization, \(x\), in lower case, and vectors in bold face (e.g., \(x = [x_{1:n}']\)). The general knowledge base (G-KB) characterizing \(X(p)\) consists of the mean function \(m_{0}(p) = E[X]\), where \(E[.]\) is the stochastic expectation, describing its consistent trends, and the covariance function \(c_{0}(p,p') = E[(X(p) - m(p))(X(p') - m(p'))]\) describing its space/time dependencies. Likewise, the site-specific knowledge base (S-KB) consists of the hard data \(x_{i}\) at space/time observation points \(p_{i}\) located at the monitoring stations, and the soft data characterizing the S/TRF values \(x_{m}\) at the space/time model prediction points \(p_{m}\) in terms of a site-specific PDF \(f_{x}(x_{m})\).

Denoting the G-KB as \(G = \{m_{0}(p),c_{0}(p,p')\}\) and the S-KB as \(S = \{x_{m},f_{x}(x_{m})\}\), we can summarize the BME steps as (1) using the Maximum Entropy principle of information theory to process the G-KB in the form of a prior Probability Distribution Function (PDF) \(f_{G}(x)\) (integrating the S-KB using an epistemic Bayesian conditionalization rule to create a BME posterior PDF \(f_{K}\) characterizing the value \(x_{i}\) taken by \(X(p)\) at any estimation point \(p_{i}\) of interest, and (3) computing space/time estimates based on the BME posterior PDF. The BME posterior PDF is given by the BME equation

\[
    f_{K}(x_{i}) = A^{-1} \int d\mathbf{x}_{m} f_{G}(\mathbf{x}_{m}) f_{x}(\mathbf{x}_{m})
\]

where \(\mathbf{x}_{m} = (x_{i},\mathbf{x}_{m})\) is the value of \(X(p)\) at points \(p_{m} = (p_{i},P_{m})\) and \(A\) is a normalization constant.
Let $Z(p) = Z(x,t)$ be the Space/Time Random Field (S/TRF) representing daily ozone. In this study we define $Z(x,t)$ as the sum of a homogeneous/stationary S/TRF and a known offset as follows. We first define the transformation of the ozone observational data $z_o$ at locations $p$, as

$$x_o = z_o - o_z(p)$$

where $o_z(p)$ may be any deterministic offset that can be mathematically calculated without error as a function of the space/time coordinate $p$. We then define $X(p)$ as a homogeneous/stationary S/TRF representing the variability and uncertainty associated with the transformed data $x_o$ and we let $Z(p) = X(p) + o_z(p)$ be the S/TRF representing daily O3. We can then calculate $\hat{z}_o$, the estimated daily O3 at unmonitored location $p_i$ by obtaining the BME estimate $\hat{z}_o$ for the transformed S/TRF $X(p)$ at the estimation point $p_i$ and adding back $o_z(p)$, the offset calculated at $p_i$.

The soft data are described by the PDF $f_s(x_o)$ characterizing the offset-removed ozone values $x_o$ at the soft data points $p$ corresponding to the centroids of the $n_m$ CTM computational nodes. The offset-removed ozone model predictions $\hat{z}_o$ are calculated at these nodes. As a key conceptual aspect of our work, the generation of this soft data PDF requires not only the offset-removed ozone model predictions, but also the observation–prediction pairs where the observed and CTM predicted ozone concentrations are paired across space and time. This PDF is expressed as

$$f_s(x_o) = \prod_{i=1}^{n_m} f(x_i | \xi_i, p_i)$$

which essentially characterizes how well each CTM offset-removed ozone value $\xi_i$ predicts the true offset-removed ozone concentration $x_i$ at the computational prediction point $p_i$. Procedurally eq 3 is simply obtained by first calculating $f_i(x_i) = \prod_{j=1}^{n_m} f(x_j | \xi_j, p_j)$, where $z_j$ and $\xi_j$ are observed and CTM predicted ozone values, respectively, and then using the offset relationship $x_i = z_i - o_z(p_i)$ to obtain $f_s(x_o)$.

As described by de Nazelle et al., the PDF $f(x_i | \xi_i, p_i)$ is modeled using a parametrized statistical distribution, chosen to be the normal distribution truncated below zero with an expected value $\lambda_1(\xi_i)$ and variance $\lambda_2(\xi_i)$, such that

$$f(x_i | \xi_i) = \Phi(x_i; \lambda_1(\xi_i), \lambda_2(\xi_i))$$

In the soft data construction approach implemented by de Nazelle et al., the parameters $\lambda_1(\xi_i)$ and $\lambda_2(\xi_i)$ vary as a function of the model prediction $\xi_i$ but are constant with respect to the space/time point $p_i$, hence their implementation is based on a Constant Air Quality Model Performance (CAMP). The CAMP approach was appropriate since in their application the air quality model performance did not change across their small study geographical domain (North Carolina) and short study duration (<15 days). Our aim, however, is to extend the BME methodological framework to the national domain by modeling $\lambda_1(\xi_i, p_i)$ and $\lambda_2(\xi_i, p_i)$ as a function of both $\xi_i$ and the space and time coordinate $p_i$ as expressed below.

$$f(x_i | \xi_i, p_i) = \Phi(x_i; \lambda_1(\xi_i, p_i), \lambda_2(\xi_i, p_i))$$

Therefore, we need to investigate how the air quality model performance varies across the continental U.S.

3.2. Variability of CTM Model Performance Evaluation across the Continental U.S. Each observed daily concentration $z_i$ is paired with its corresponding CTM prediction value $\hat{z}_i$, and the error for the observation-prediction pair is defined as $e_i = z_i - \hat{z}_i$. To evaluate the air quality model performance over a given space time region $R$ of interest, we calculate error statistics such as the mean prediction error (ME), the standard deviation of the prediction error (SE), the mean normalized bias (MNB), and the mean normalized gross error (MNGE) as defined in SI eqs 1s–4s.

According to the model performance analysis of this CTM (see SI Section 1), for the DMSA O3, we find that overall the CAMx simulation with $12 \times 12$ km$^2$ grid cell resolution has a substantially lower overprediction (median ME = +1.4 ppb) than that with $36 \times 36$ km$^2$ grid cell resolution (median ME = +4.5 ppb). Furthermore, as summarized in Figures 2s and 3s, the ME, SE, MNB, and MNGE at individual monitoring sites vary over a wider range for the simulation with $36 \times 36$ km$^2$ grid cell resolution. The variability of these ME and SE values exhibit clear geographical trends (SI Figures 4s–7s for the DMSA and Figures 12s–15s for the D24A): Urban cities located in the east and west coast tend to have higher overprediction bias (i.e., higher ME) and higher imprecision (i.e., higher SE) than sites located in the central United States. We also found noticeable seasonal differences in the model performance for both CTM simulations (SI Figures 8s–9s for the DMSA and Figures 16s–17s for the D24A).
The results of this analysis provide strong evidence that the performance of CTM varies considerably across the national domain and over seasons. Therefore, there is a need to extend the implementation of the BME framework to account for this space/time variability in model performance. We used the Regionalized Air Quality Model Performance (RAMP) method to quantify how the expected value \( \lambda_1(\bar{z}_j, p_i) \) and variance \( \lambda_2(\bar{z}_j, p_i) \) for the ozone soft data derived from CTM outputs vary as a function of both the CTM prediction \( \bar{z}_j \) and the space/time computational node \( p_i \) for which that prediction was calculated. The goal of the RAMP method is to select the most relevant observation–prediction pairs to most accurately identify the CTM bias associated with the prediction value \( \bar{z}_j \) outputted for any space/time computational node \( p_i \) of interest.

### 3.3. Proposed Regionalized Air Quality Model Performance (RAMP) Evaluation Framework

In the first stage of the RAMP analysis, we pool for each monitoring site the observation–prediction pairs \((z_p, \bar{z}_j)\) that are within a time tolerance of \( \Delta T = 120 \) days of a particular time of interest \( t \).

Examples of two selected sites are shown in Figure 1. These pairs are highly relevant to the location \( s \) where the monitoring station is sited, and the 120 days time window is chosen to balance the abundance of the pairs and the intention to retain seasonal specificity in the \( \bar{z}_j - z_i \) differences. We stratify the pairs in 10 equal percentile bins of increasing predicted values \( \bar{z}_j \), and for each bin we compute the mean and variance of observed values,

\[
\bar{\lambda}_1(\bar{z}_j, s_i, t) = \frac{1}{n_0(\bar{z}_j, s_i, t)} \sum_{j=1}^{n_0(\bar{z}_j, s_i, t)} z_j \quad \text{and}
\]

\[
\bar{\lambda}_2(\bar{z}_j, s_i, t) = \frac{1}{n_0(\bar{z}_j, s_i, t)} \sum_{j=1}^{n_0(\bar{z}_j, s_i, t)} (z_j - \bar{\lambda}_1(\bar{z}_j, s_i, t))^2
\]

where \( n_0(\bar{z}_j, s_i, t) \) is the number of \((z_i, s)\) pairs in the \( b \)th bin, \( z_i \) is the \( b \)th observation value in these \( n_0(\bar{z}_j, s_i, t) \) pairs, and \( \bar{z}_j \) is the average of the predictions \( \bar{z}_j \) in these \( n_0(\bar{z}_j, s_i, t) \) pairs.

In the second stage of the RAMP analysis we obtain \( \lambda_1(\bar{z}_j, p_i) \) and \( \lambda_2(\bar{z}_j, p_i) \) for actual predicted values \( \bar{z}_j \) and space/time grid point \( p_i = (s_i, t_i) \) as follows. For each monitoring site \( s \), we perform a linear interpolation/extrapolation of the \( \bar{\lambda}_1(\bar{z}_j, s_i, t) \) and \( \bar{\lambda}_2(\bar{z}_j, s_i, t) \) values to obtain \( \hat{\lambda}_1(\bar{z}_j, s_i, t) \) and \( \hat{\lambda}_2(\bar{z}_j, s_i, t) \) (see interpolation lines in Figure 3), and then we do a spatial interpolation of these values to obtain \( \hat{\lambda}_1(\bar{z}_j, p_i) \) and \( \hat{\lambda}_2(\bar{z}_j, p_i) \) at \( p_i = (s_i, t_i) \) using the following formula

\[
\hat{\lambda}_{1/2}(\bar{z}_j, s_i, t_i) = \frac{\sum_{i=1}^{N} w_i(s_i, s_j) \hat{\lambda}_{1/2}(\bar{z}_j, s_i, t_i)}{\sum_{i=1}^{N} w_i(s_i, s_j)}
\]

where \( n = 1, ... , N \) refers to the \( N \) monitoring sites closest to the location of the computational node \( s_i \) of interest, and \( w(s_i, s_j) \) is a weight equal to the inverse of the distance between the computational node \( s_i \) of interest and the \( n \)th neighboring monitoring station.

Stated simply, \( \bar{z}_j - \hat{\lambda}_1(\bar{z}_j, p_i) \) is the bias characterizing systematic errors associated with a CTM prediction value of \( \bar{z}_j \) calculated at space/time grid point \( p_i = (s_i, t_i) \), and \( \hat{\lambda}_2(\bar{z}_j, p_i) \) is the variance characterizing the associated imprecision. The strength of the RAMP method is that it does not make any assumption on the relationship between observed and predicted values, and therefore geographical and temporal changes in nonlinear and nonhomoscedastic relationships are automatically captured in the calculation of \( \hat{\lambda}_1(\bar{z}_j, p_i) \) and \( \hat{\lambda}_2(\bar{z}_j, p_i) \), which are fully integrated in the BME soft data through eq 4.

#### 3.4. Offset Analysis

The offset is used to transform the daily O3 data into residual offset-removed data. The ozone offset \( o_{\lambda}(p_i) \) at an arbitrary location \( p_i = (s_i, t_i) \) is obtained using an exponential kernel smoothing of the surrounding observed O3 data:

\[
o_{\lambda}(s_i, t_i) = \frac{\sum_{j=1}^{N} w_j z_j}{\sum_{j=1}^{N} w_j}
\]

where \( z_j \) is the observed value at space/time observation point \( p_i = (s_i, t_i) \) within the neighborhood of the point \( p_i \) of interest, and the kernel smoothing weights are

\[
w_j = \exp \left( - \frac{\| s_i - s_j \|}{a_s} - \frac{\| t_i - t_j \|}{a_t} \right)
\]

\( a_s \) is the spatial offset kernel smoothing range, and \( a_t \) is the temporal offset kernel smoothing range.

An optimal offset (\( a_s = 50 \text{ km} \) and \( a_t = 10 \text{ day} \)) was chosen to ensure the transformed data has a low variance so that the geostatistical estimation error variance is minimized, while retaining high autocorrelation to ensure that neighboring data locations are informative at the estimation location (see SI Section 2 for details).

### 3.5. Space–Time Covariance Model

The covariance model for the homogeneous/stationary S/TRF \( \text{X}(p) \) is developed from the experimental covariance of the transformed observational data \( x_s = z_o - o_{\lambda}(p) \). The experimental covariance for a spatial lag \( r \) and a temporal lag \( \tau \) is calculated as

\[
\hat{c}_X(r, \tau) = \frac{1}{N(r, \tau)} \sum_{j=1}^{N(r, \tau)} x_{\text{head}}(r, \tau, j) m_x^2
\]

where \( N(r, \tau) \) is the number of pairs of values \((x_{\text{head}}, x_{\text{tail}})\) separated by a spatial lag \( r \) and temporal lag \( \tau \), and \( m_x \) is the mean of the \( x_s \) data. In practice, \( \hat{c}_X(r, 0) \) and \( \hat{c}_X(0, \tau) \) are calculated and plotted separately to facilitate the visualization of the space/time covariance models (SI Figures 30s and 31s). A 3-structured exponential covariance model was chosen for the subsequent BME analysis (see SI Section 2 for details).

The formula of the 3-structured exponential covariance model is given by

\[
\hat{c}_X(r, \tau) = C_0 \left[ a_x \exp \left( -\frac{3r}{a_1} \right) \exp \left( -\frac{3\tau}{a_2} \right) + b_x \exp \left( -\frac{3r}{a_1} \right) \exp \left( -\frac{3\tau}{a_2} \right) \right]
\]

\[
= \exp \left( -\frac{3r}{a_1} \right) + \left( 1 - \alpha - \beta \right) \exp \left( -\frac{3\tau}{a_2} \right)
\]

\[
\exp \left( -\frac{3\tau}{a_2} \right)
\]

\[
(11)
\]
3.6. Validation Analysis. A validation analysis is used to assess the accuracy of different BME estimation approaches. Each observed value $z_j$ at space/time point $=(s,t)$ is compared with the corresponding ozone concentration $z_j^*$ re-estimated using only noncollocated data outside of a radius $r_v$ from $s_j$. The validation error, which is the difference between each re-estimated value $z_j^*(r_v)$ and observed value $z_j$ is defined as $e_j = z_j^*(r_v) - z_j$. The estimation accuracy is quantified based on statistics of these estimation errors, which is a function of the validation radius $r_v$. They consist of the root mean square error RMSE (ppb), the $R^2$ (unitless), the mean normalized bias MNB (%), and the mean normalized gross error MNGE (%) between observations and re-estimated values, calculated as a function of $r_v$ as shown in SI eqs 7s–10s.

Using the validation error statistics RMSE($r_v$) and $R^2(r_v)$, we compare the following three BME data fusion scenarios.

1. Scenario OBS: Uses only ozone observations in the BME framework. This is the kriging limiting case of the BME data integration framework since kriging is the linear limiting case of BME when only hard data are used.

2. Scenario CAMP: Integrates both observations and CTM predictions in the BME data integration framework using CTM soft data constructed with the CAMP approach, which assumes that CTM model performance is constant across space and time.

3. Scenario RAMP: Integrates both observations and CTM predictions in the BME data integration framework, with CTM soft data obtained through the RAMP approach introduced here to account for the space/time variation in CTM model performance.

We let $R^2_{OBS}$, $R^2_{CAMP}$, and $R^2_{RAMP}$ be the coefficient of determination of estimation error for scenarios OBS, CAMP, and RAMP, respectively, and we define the percent change $PCR^2_{OBS→CAMP}$ and $PCR^2_{OBS→RAMP}$ as follows:

$$PCR^2_{OBS→CAMP} = 100 \frac{R^2_{CAMP}(r_v) - R^2_{OBS}(r_v)}{R^2_{OBS}(r_v)}$$

$$PCR^2_{OBS→RAMP} = 100 \frac{R^2_{RAMP}(r_v) - R^2_{OBS}(r_v)}{R^2_{OBS}(r_v)}$$

A positive $PCR^2$ indicates an increase in $R^2$, which corresponds to the percent improvement in estimation precision resulting in integrating air quality model predictions in the BME data integration.

4. RESULTS

4.1. BME Ozone Estimates. When incorporating CTM prediction as soft data in the BME data integration framework (scenarios CAMP and RAMP), we use the soft data with the finer grid cell resolution when it is available. That means for the areas where both 36 × 36 km$^2$ and 12 × 12 km$^2$ grid cell resolution are available, we incorporate the 12 × 12 km$^2$ CTM prediction values.

Figure 2 shows for Jul-11–2005 the BME estimates of DM8A ozone concentrations $z_j$ obtained for the three estimation scenarios. This day was chosen because it has the highest standard deviation (at 33.1 ppb) for CTM prediction errors at ozone monitoring sites, which means the CTM model performance has the highest spatial variability among sites. It is clear that on this day the BME mean estimates (in the top panel of Figure 2) in the immediate proximity of the monitoring stations (marked in circles) are at very similar levels in the three maps, being in good agreement with the observed data in their local neighborhood. As the estimation location moves away from the monitoring stations, the difference among these three maps becomes more substantial. For example, in scenario OBS we see a wider area of high ozone value, with the area above 70 ppb covering 811 296 km$^2$ across the continental United States. In scenario CAMP the ozone plume above 70 ppb only covers a much smaller area (545 184...
The analysis uses a constant offset and corresponding covariance model. \( \lambda_i \) is the validation radius around monitoring stations within which all observation points are included in the validation estimations. \( R^2_{\text{OBS}} \), \( R^2_{\text{CAMP}} \), \( R^2_{\text{RAMP}} \) are the squared Spearman’s correlation coefficients between the ozone observations and the BME estimates for the OBS, CAMP, and RAMP data integration scenarios, respectively. \( \text{RMSE}_{\text{OBS}} \), \( \text{RMSE}_{\text{CAMP}} \), and \( \text{RMSE}_{\text{RAMP}} \) are the corresponding root-mean-square errors. \( \text{PCR}_{\text{OBS-CAMP}} \) is the percent change in \( R^2 \) from OBS scenario to CAMP scenario and \( \text{PCR}_{\text{OBS-RAMP}} \) is the percent change from scenario OBS to scenario RAMP. \( p\text{-value}_{\text{OBS-CAMP}} \) is the \( p \)-value testing the hypothesis that there is no difference between the \( R^2 \) in scenarios OBS and CAMP, and \( p\text{-value}_{\text{OBS-RAMP}} \) is the \( p \)-value testing the hypothesis that there is no difference between the \( R^2 \) in scenarios OBS and RAMP.

Figure 3 shows a map of the raw CAMx modeled DM8A average ozone predictions for 11-July-2005. Also shown in the figure are the bias-corrected CTM predicted values \( \lambda (\tilde{z}_i, p_i) \) from scenarios CAMP and RAMP for the same day. Both scenarios CAMP and RAMP corrected the CTM prediction bias to some extent, especially for areas close to the monitoring sites. There are, however, substantial differences of the bias-corrected CTM predicted values \( \lambda (\tilde{z}_i, p_i) \) between scenarios CAMP and RAMP. For scenario CAMP, the map of the bias-corrected CTM predicted values \( \lambda_1 \) shows lower values than scenario RAMP, with the area above 70 ppb covering 221 616 km² across the continental United States and the highest bias-corrected CTM predicted value \( \lambda_1 \) at 105 ppb. By contrast, in scenario RAMP the size of the area with bias-corrected CTM ozone levels greater than 70 ppb is 431 856 km², with the peak bias-corrected CTM ozone level reaching 111 ppb. This substantial difference is due to the assumed homogeneity of the

Figure 3. DM8A ozone concentrations in the United States on 11-July-2005 using (left) the raw CTM Model predictions, (middle) the bias-corrected expected values \( \lambda (\tilde{z}_i, p_i) \) for the estimation scenario CAMP, and (right) the bias-corrected expected values \( \lambda (\tilde{z}_i, p_i) \) for the estimation scenario RAMP.

The uncertainty associated with the BME estimates is quantified by the corresponding BME standard estimation error (bottom panel of Figure 2). For estimation scenario OBS there is a higher estimation uncertainty, with the highest BME standard estimation error reaching 8.7 ppb for areas far away from any monitoring stations, and with an average standard estimation error of about 6.4 ppb across the continental United States. This is in contrast to estimation scenarios CAMP and RAMP, where the BME standard estimation error remains relatively low, with the highest standard estimation error reaching 6.5 and 6.3 ppb for scenarios CAMP and RAMP, respectively. This indicates that integrating both observations and model predictions improved the quality of the ozone estimates, especially for areas far away from any monitoring station. Overall, scenario RAMP has the lowest standard estimation error, with an average standard estimation error of about 4.6 ppb across the continental United States.

### 4.2. Soft Data Construction Using the RAMP Approach

The construction of the soft data using the proposed RAMP approach can be illustrated by comparing scenario RAMP that accounts for the space/time variability of CTM performance, with scenario CAMP that does not. Two important parameters that differed in these two scenarios are the bias-corrected expected values \( \lambda (\tilde{z}_i, p_i) \) and the corresponding soft data variance \( \lambda (\tilde{z}_i, p_i) \).

<table>
<thead>
<tr>
<th>RMSE_{OBS} (ppb)</th>
<th>0</th>
<th>36</th>
<th>72</th>
<th>108</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE_{CAMP}(ppb)</td>
<td>5.675</td>
<td>6.442</td>
<td>6.966</td>
<td>7.250</td>
</tr>
<tr>
<td>RMSE_{RAMP}(ppb)</td>
<td>5.445</td>
<td>6.109</td>
<td>6.531</td>
<td>6.732</td>
</tr>
</tbody>
</table>

| \( R^2 \)_{OBS} (unitless) | 0.886 | 0.853 | 0.829 | 0.817 |
| \( R^2 \)_{CAMP} (unitless) | 0.884 | 0.853 | 0.831 | 0.819 |
| \( R^2 \)_{RAMP} (unitless) | 0.893 | 0.866 | 0.849 | 0.841 |

| PCR_{OBS-CAMP} | \(-0.223\) | 0.014 | 0.197 | 0.235 |
| PCR_{OBS-RAMP} | 0.726 | 1.602 | 2.407 | 2.936 |

| \( p \)-value_{OBS-CAMP} (unitless) | 1 | 0.990 | 0.321 | 0.0009 |
| \( p \)-value_{OBS-RAMP} (unitless) | 0.980 | 0.0001 | < 0.00001 | < 0.00001 |

| \( p \)-value | < 0.00001 | < 0.00001 | < 0.00001 | < 0.00001 |
CTM model performance in scenario CAMP that forces the same correction throughout the study domain. This correction results in an over correction in some local areas such as the area covering the western and southwestern states of Nevada, Idaho, Utah, Wyoming, Colorado, Arizona, New Mexico, and California. In contrast, the scenario RAMP is better able to account for regional biases in model performance.

The maps for the corresponding square root of soft data variance $\hat{\lambda}_s(\tilde{z}_v, p_i)$ are shown in SI Figure 34s. This map characterizes the imprecision associated with the bias-corrected CTM predicted values $\hat{\lambda}_s(\tilde{z}_v, p_i)$. We find that scenario RAMP has more localized gradients for the variance $\hat{\lambda}_s(\tilde{z}_v, p_i)$, with the square root of $\hat{\lambda}_s(\tilde{z}_v, p_i)$ spanning from a low value of 2.6 ppb to a high of 20.5 ppb, and averaging about 8.7 ppb across the continental United States. By contrast, this variance has less spatial variability for the scenario CAMP; with a narrower span of $\hat{\lambda}_s(\tilde{z}_v, p_i)$ values ranging from 9.2 to 14.1 ppb, and a higher average over the continental United States of 10.3 ppb. This illustrates that the proposed RAMP method used in scenario RAMP has a greater ability to characterize regional changes in the precision of bias-corrected CTM predictions. This is an important methodological improvement explaining the improved performance in scenario RAMP in the cross validation analysis described next.

4.3. Validation Results. The validation analysis was conducted to compare estimation scenarios OBS, CAMP, and RAMP. For those monitoring site locations covered by both CTM domains, soft data with a finer grid cell resolution of 12 × 12 km$^2$ are incorporated for scenarios CAMP and RAMP.

Table 1 shows the percent change in $R^2$ as a function of validation radius $r_v$. As shown in the table, scenario RAMP has the highest $R^2$ for all radii $r_v$. Furthermore, the $R^2_{OBS→CAMP}$ representing the percent change in $R^2$ from scenario OBS to scenario CAMP, is consistently positive when $r_v$ is larger than 0 km, indicating that de Nazelle’s approach, even when applied beyond the condition for which it was developed, is still more accurate than relying on observational data alone, and its percent increase in $R^2$ consistently improves as $r_v$ increases. The $R^2_{OBS→RAMP}$ representing the percent change in $R^2$ when comparing scenarios OBS and RAMP, is also consistently positive. More importantly, it is larger than $R^2_{OBS→CAMP}$ with an overall 0.73% increase for the DM8A and 2.6% increase for the D24A O3 in $R^2$ for $r_v = 0$ km. Furthermore, there is 2.9% increase for the DM8A and 5.9% increase for the D24A in $R^2$ between scenario OBS and RAMP at locations more than 108 km away from a monitoring station. We also calculated the percent change in $R^2$ when we aggregate the BME estimates into weekly and monthly estimates (SI Table 5s). Scenario RAMP still has the highest $R^2$ increase for these aggregated metrics. More results of the cross-validation analysis are documented in SI Section 5.

These results demonstrate that integrating both observations and soft data processed through the RAMP approach improves the capability of estimating both of the DM8A and the D24A ozone concentrations compared to using only observational data and through the CAMP approach. Compared with the CAMP approach, the RAMP approach consistently results in a further improvement in estimation, as evidenced by the fact that the $R^2_{OBS→RAMP}$ values are over 12 times greater than the $R^2_{OBS→CAMP}$ values for the DM8A ozone concentrations and over 3.5 times greater than the $R^2_{OBS→CAMP}$ values for the D24A ozone concentrations.

5. DISCUSSION

We have presented an ozone estimation method that is able to integrate observations with predictions from a CTM. These predictions are weighted according to model performance that varies across space and time based on a soft data construction utilizing the newly developed RAMP method. Thus, estimates are produced that put priority on observations and take advantage of air quality model predictions based on how well they reproduce the observed values. Spatial fields generated from this approach provide an observation and CTM informed representation of ozone across space/time that is more accurate and precise than relying only on observation data. This was especially true for locations away from monitoring stations.

We developed the RAMP method by extending the CAMP framework presented by de Nazelle et al.5 We tested the RAMP method by comparing the percent change in $R^2$ and found the percent increase achieved by the RAMP method ($R^2_{OBS→RAMP}$) was 4- to 10-fold greater than that of the CAMP method ($R^2_{OBS→CAMP}$). This improvement is attributed to the RAMP ability to account for the spatial and temporal variability in model performance.

Approaches used to model the uncertainty associated with the CTM model predictions can be divided into parametric approaches that parametrize the relationship between the air pollution observations $Z$ and predictions $\hat{Z}_{\lambda}$, and nonparametric approaches such as our RAMP method that directly model air quality performance based on paired observed and predicted values. For example, Fuentes et al.6 assume that $Z(s) = \beta_0(s) + \beta_1(s) Z(s) + \epsilon(s)$, while Berrocal et al.7 assume that $Z(s) = \beta_0(s) + \beta_1(s) Z(B,s) + \epsilon(s)$, where in both cases the relation is linear and homoscedastic since the noise term is assumed to have a constant error variance, $\epsilon(s) \sim N(0, \sigma^2)$. By contrast, our novel RAMP approach is a nonparametric approach that fully accounts for the nonlinear, nonhomoscedastic relationship between observations $Z$ and predictions $\hat{Z}$, and accounts for the spatiotemporal varying nature of that relationship.

To illustrate the difference between parametric and nonparametric approaches, we also compared our nonparametric estimates to the ones generated from a cokriging estimation with a parametric relationship between the observations and the CTM model predictions (SI Section 6). In this analysis, we found that the disadvantage of cokriging is that it is limited by the parametric relationship and the final estimates tend to be heavily influenced by CTM model predictions. On the basis of the validation results (SI Table 6s), our approach outperforms the cokriging approach in terms of smaller root mean squared error, 5.45 ppb for RAMP and 6.5 ppb for cokriging, and higher spearman’s $R_s$, 0.893 for RAMP and 0.845 for cokriging for the DM8A O3.

To the best of our knowledge, our proposed framework is one of the first to fully account for the spatiotemporal variation of the nonlinear, nonhomoscedastic relationship between air pollution observations and predictions. Major strengths of our approach are that its numerical implementation is based on a straightforward analysis of paired observations and predictions, which is computationally efficient and trivially implemented on parallel computers, and it reduces the uncertainty of the mapping error by putting more weight on air quality predictions where they reproduce well the observed values.
This is particularly useful in large regulatory or health studies that need to incorporate air quality predictions with widely varying model performance across the study domain, such as studies examining the entire continental United States rather than some small portions of it, or studies combining air quality predictions from a variety of air quality model simulations with significantly varying model performance.

**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.6b00096.

Additional text, equations, figures, and tables as mentioned in the text (PDF)

**AUTHOR INFORMATION**

Corresponding Author

*E-mail: airquality@unc.edu.

Notes

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This work was supported by National Institute on Aging (NIA) of the National Institutes of Health (NIH) under award R01AG033078. It was also supported by a grant (T32ES007018) from the National Institute of Environmental Health Sciences (NIEHS) training program of NIH. The CAMx software of Environ Corporation. This work was supported by National Institute on Aging (NIA) of the National Institutes of Health (NIH) under award R01AG033078. It was also supported by a grant (T32ES007018) from the National Institute of Environmental Health Sciences (NIEHS) training program of NIH. The CAMx software of Environ Corporation. This work was supported by National Institute on Aging (NIA) of the National Institutes of Health (NIH) under award R01AG033078. It was also supported by a grant (T32ES007018) from the National Institute of Environmental Health Sciences (NIEHS) training program of NIH. The CAMx software of Environ Corporation.

**REFERENCES**


(14) U.S. EPA. Guideline on Data Handling Conventions for the 8-h Ozone NAAQS; EPA-454/R-98-017; Research Triangle Park, NC, 1998.


