Presented at the 11th Annual CMAS Conference, Chapel Hill, NC, October 15-17, 2012

CHARACTERIZATION OF AIR QUALITY OZONE MODEL PERFORMANCE USING LAND USE REGRESSION MODEL: AN APPLICATION IN EXPOSURE ASSESSMENT FOR EPIDEMIOLOGY STUDIES

Yadong Xu
William Vizuete
Marc Serre
The University of North Carolina, Chapel Hill, NC, USA

1. Abstract

Bayesian Maximum Entropy (BME), as a spatiotemporal epistemic knowledge-processing framework, has been recently applied to map criteria pollutants[1, 2]. Using BME to integrate air monitoring observations and numerical model predictions has been used to improve spatial predictions of ozone concentrations[3]. In our study, this approach is used to provide ozone exposure assessment for epidemiology studies. To obtain optimal estimates of exposure, the observed ozone concentrations, treated as an error free proxy, are integrated with the model predictions weighed according to how well they reproduce the observed values. The construction/generation of soft data from air quality model predictions is an important process that eventually will be integrated into the BME framework together with the hard observed data to provide a representation of space-time ozone distribution. Therefore, how we characterize the performance of the air quality model is critical to the accuracy and the precision of ozone estimations.

In our study, CMAQ outputs of ozone annual simulations, at different grid cell resolutions (range from 4k to 36k) for the domain of the continental U.S, is evaluated against observed concentrations extracted from the Air Quality System (AQS) network and The Clean Air Status and Trends Network (CASTNET). Hourly Observations for each monitoring site and model results, paired in space and time, are used to construct soft data.

In the previous study using BME integrating ozone observations and model predictions [3], model performance has been assumed to be homogeneous for the small study domain. This approach might not be applicable as the study domain expands because we observed the spatial heterogeneity and temporal variability of ozone model performances across the U.S. According to our exploratory model performance analysis, the spatial distribution of the ozone model performance at the monitoring sites, measured in the statistical metrics ME (the mean error) and SDE (the standard deviation of the prediction errors), are not homogenous. The ozone model performances for urban sites are worse than for rural sites. The statistical metrics also show the model performances during photochemical hours are much better than predictions made during the night. Rather than assuming the model performance is homogeneous across the domain as in previous studies, we developed a stochastic land use regression model to characterize the prediction errors and also the distribution of the prediction errors that change across space/time.

The development of this model performance land use regression model uses the following predictor variables:

(1) A representation of the population density for the surrounding areas of each site calculated from the U.S. census block group population data obtained through geographic information systems (GIS), which is used to quantify the urbanization of

the location.

(2) The gridded fractional land use/land cover data processed from National Land Cover Data (NLCD) and USGS land use category data, which has great importance not only for chemical surface fluxes calculation, but also for characterization of the land surface exchange processes in meteorology modeling and for emission inventory spatial allocation in emission modeling. Therefore, it has great influences on ozone model performance.

(3) The gridded NOx (=NO+NO2) level or H2O2/HNO3 ratio extracted from the same air quality model as O3 values, considering the photochemical inter-conversions between NOx and O3.

We investigate the correlations between the ozone model performance metrics and the individual sets of above predictor variables first. Then we develop a multiple linear regression model that can predict the model performance at any location where there is no air-monitoring site.

We apply this new methodology to generate soft data for ozone concentrations across the U.S. By conducting cross-validation analysis, we present results comparing our approach to an analysis that assume ozone model performance is homogenous and the soft data is only a function of model predictions. We show that our new approach, by incorporating a land use regression model to characterize the ozone model performance, improves the accuracy and the precision of ozone estimations across the U.S.

2. References

1. Yu H-L, Chen J-C, Christakos G, Jerrett M: BME Estimation of Residential Exposure to Ambient PM and Ozone at Multiple Time Scales. Environmental Health Perspectives 2009, 117(4):537-544.

2. Bogaert P, Christakos G, Jerrett M, Yu HL: Spatiotemporal modelling of ozone distribution in the State of California. Atmospheric Environment 2009, 43(15):2471-2480.

3. Nazelle Ad, Arunachalam S, Serre ML: Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: An Application for Attainment Demonstration in North Carolina. Environmental Science & Technology 2010, 44(15):5707-5713.