Abstract

In recent years environmental epidemiologists have begun utilizing regional-scale air quality computer models to predict ambient air pollution concentrations in health studies instead of or in addition to monitoring data from central sites. The advantages of using such models include better spatio-temporal coverage and the capability to predict concentrations of unmonitored pollutants. However, there are also drawbacks, chief among them being that these models can exhibit systematic spatial and temporal biases. In order to use these models in epidemiological investigations it is very important to bias-correct the model surfaces. We present a novel statistical method of spatio-temporal bias correction for the Community Multi-scale Air Quality (CMAQ) model that allows simultaneous bias adjustment of $PM_{2.5}$ mass and its major constituent species using publically available speciated data from ambient monitors. The method uses mass conservation and the more widespread unspeciated $PM_{2.5}$ mass observations to constrain the sum of the $PM_{2.5}$ species' concentrations in locations without speciated monitors. We develop the model in the context of an epidemiological study investigating the association between $PM_{2.5}$ species' ambient concentrations and birth outcomes throughout the state of New Jersey. Since our exposures of interest are multi-month averages we focus specifically on modeling seasonal bias trends rather than daily biases. As one would expect, our bias-corrected CMAQ results are more accurate than the original CMAQ output. More interestingly, using a cross-validation study we find that our model's predictions are improved by enforcing mass conservation, and furthermore that our model is competitive with kriging in a comparison in which the latter has the advantage.