Fusing Point and Areal Level Space-Time Data with Application to Wet Deposition

Sujit K. Sahu[1], Alan E. Gelfand[2] and David M. Holland[3]

[1] School of Mathematics, Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, UK
[2] Department of Statistical Science, Duke University, Durham, NC, USA
[3] U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC, USA

Abstract

The combustion of fossil fuel produces a wide variety of chemicals, including such gases as sulfur dioxide and nitrogen oxides. These gases are emitted to the air, transformed to acidic compounds, and are then returned to the earth. Most of the acid deposition in the eastern U.S. can be attributed to the release of sulfur dioxide and nitrogen oxides from large fossil fueled power plants. When delivered by precipitation, such as rain, snow, or fog, the process is called wet sulfate and nitrate deposition. Wet deposition is responsible for damage to lakes, forests, and streams.

The primary objective of this study is to develop a high-resolution model for wet chemical deposition that offers better inference than is currently possible using just National Atmospheric Deposition Program (NADP, `nadp.sws.uiuc.edu`) wet deposition measurements and classical interpolation techniques. The proposed model uses deposition and precipitation data from NADP monitoring sites and output from a computer simulation model known as the Community Multi-Scale Air Quality Model (CMAQ, `epa.gov/asmdnerl/CMAQ`) on a 12 kilometer square grid. The CMAQ model uses variables, such as power station emission volumes, meteorological information, and land-use, to predict average deposition levels. However, it is well known that these predictions are biased; the monitoring data provide more accurate deposition information. The mismatch in the spatial domains for the point and grid referenced computer output is often alluded to as the 'change of support problem' and creates challenges in modeling and model fitting.

The contribution of this article is the development of a joint model by combining a conditionally auto-regressive (CAR) model for the gridded CMAQ data and a space-time process model for observed point level data. Model components are linked using latent space-time processes in a Bayesian hierarchical modeling setup. This strategy yields a better model for high

spatio-temporal resolution. All predictive inference is performed using the point level model. A key feature is avoidance of integration of the observed point level monitoring process to a grid level process.
More precisely, the average deposition level in a grid cell A_j at time t, denoted by $Z(A_j, t)$, need not be the level observed at any particular site $\mathbf{s}_i$ in A_j, denoted by $Z(\mathbf{s}_i, t)$. The change of support problem in this context addresses converting the point level $Z(\mathbf{s}_i, t)$ to the grid level $Z(A_j, t)$ through the stochastic integral,

$$Z(A_j, t) = \frac{1}{|A_j|} \int_{A_j} Z(\mathbf{s}, t)\ d\mathbf{s}, \tag{1}$$

where $|A_j|$ denote the area of the grid cell A_j. Fusion modeling, working with *block averaging* as in (1) has been considered by, e.g., Fuentes and Raftery (2005).
Our approach introduces a latent point level atmospheric process which is centered, in the form of a measurement error model (MEM), around a grid cell based latent atmospheric process. The latent processes are introduced to capture point masses at 0 with regard to deposition while the MEM circumvents the stochastic integration. In particular, the point level observed data represent ‘ground truth’ while gridded CMAQ output are anticipated to be biased. As a result, the MEM enables calibration of the CMAQ model. The opposite problem of disaggregation, i.e. converting the grid level computer output denoted by $Z(A_j, t)$ to point level ones, $Z(\mathbf{s}_i, t)$ is not required. The only assumption is that $Z(A_j, t)$ is a reasonable surrogate for $Z(\mathbf{s}_i, t)$ if site $\mathbf{s}_i$ is within grid cell A_j.
The amount of wet deposition is directly related to precipitation – there can be no deposition without precipitation. Hence, accurate predictions here require utilization of precipitation information. Note that both the precipitation and deposition data have atomic distributions, i.e., they are continuous random variables with positive mass at zero. Our proposal is to build a model for deposition based on precipitation which is able to handle these atoms. We introduce a latent space-time atmospheric process which drives both precipitation and deposition.