Rejoinder

Peter F. Craigmile[*], Catherine A. Calder[†], Hongfei Li[‡], Rajib Paul[§], and Noel Cressie[¶]

We would like to thank Christopher David Barr, Francesca Dominici, David B. Dunson, and Alexandra M. Schmidt for taking the time to discuss our article. We have grouped our responses under the various topics that were raised.

1 Data, data management, and software

We agree with Schmidt that it is essential that researchers from many diverse areas have access to affordable, but still trustworthy, software. In this research project, substantial effort went into preparing datasets. Much of the data came from different government agencies, with databases arranged in multiple formats, often including variables that were not immediately relevant to our scientific pursuits. In our work on this project, the use of SAS was essential to producing clean datasets.

A further issue that complicates data formation is that government agencies tend to arrange their data in a format fit for a single purpose, usually monitoring compliance in our case. In earlier work (Cressie et al. 2007), we investigated including global water information into our study of Environmental Protection Agency (EPA) Region 5 (consisting of six midwestern states). An example of the problems we encountered was that we could not find recorded arsenic concentrations for public water systems throughout EPA Region 5; while we were able to obtain data for Ohio, for the remainder of the states we only had the dates for which public water systems were out of compliance.

1.1 Information on time in the NHEXAS dataset

As Schmidt mentions, the temporal dimension of exposure-related measurements is often key to understanding dose-response relationships. With the exception of a few individuals who were monitored over multiple time periods, all NHEXAS measurements for an individual were confined to a seven-day period, and the collection of samples occurred at fixed time points within this sampling window. This aspect of the study design precluded us from exploring the strength of the media/exposure associations at different temporal lags and over different periods of integration. We note that despite these inherent limitations of our primary data source, substantial day-to-day variation

[*]Department of Statistics, The Ohio State University, Columbus, OH, USA, http://www.stat.osu.edu/~pfc/

[†]Department of Statistics, The Ohio State University, Columbus, OH, USA, http://www.stat.osu.edu/~calder/

[‡]IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, http://www.stat.osu.edu/~hongfei/

[§]Department of Statistics, Western Michigan University, Kalamazoo, MI, USA

[¶]Department of Statistics, The Ohio State University, Columbus, OH, USA, http://www.stat.osu.edu/~ncressie/