Spatial Statistics 9 (2014) 166–179

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Exploration of the use of Bayesian modeling of gradients for censored spatiotemporal data from the *Deepwater Horizon* oil spill

Harrison Quick [a], Caroline Groth [b], Sudipto Banerjee [b,*], Bradley P. Carlin [b], Mark R. Stenzel [c], Patricia A. Stewart [d], Dale P. Sandler [e], Lawrence S. Engel [f], Richard K. Kwok [e]

[a] *Department of Statistics, University of Missouri, Columbia, MO 65211, United States*
[b] *Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455, United States*
[c] *Exposure Assessments Applications, LLC, Arlington, VA 22207, United States*
[d] *Stewart Exposure Assessments, LLC, Arlington, VA 22207, United States*
[e] *National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, United States*
[f] *Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27599, United States*

A R T I C L E I N F O

Article history:
Received 10 September 2013
Accepted 7 March 2014
Available online 19 March 2014

Keywords:
Gaussian process
Gradients
Markov chain Monte Carlo
Censored data
Hierarchical modeling
Spatiotemporal data

A B S T R A C T

This paper develops a hierarchical framework for identifying spatiotemporal patterns in data with a high degree of censoring using the gradient process. To do this, we impute censored values using a sampling-based inverse CDF method within our Markov chain Monte Carlo algorithm, thereby avoiding burdensome integration and facilitating efficient estimation of other model parameters. We illustrate use of our methodology using a simulated data example, and uncover the danger of simply substituting a space- and time-constant function of the level of detection for all missing values. We then fit our model to area measurement data of volatile organic compound (VOC) air concentrations collected on vessels supporting the response and clean-up efforts of the *Deepwater Horizon* oil release that occurred starting April 20, 2010. These data contained a high percentage of observations below the detectable limits of the measuring instrument. Despite this, we were still able to make some interesting discoveries, including elevated levels of VOC near the site of the oil well on June 26th. Using the results from

* Corresponding author. Tel.: +1 612 624 0624.
E-mail addresses: quickh@missouri.edu (H. Quick), baner009@umn.edu (S. Banerjee).

http://dx.doi.org/10.1016/j.spasta.2014.03.002

this preliminary analysis, we hope to inform future research on the *Deepwater Horizon* study, including the use of gradient methods for assigning workers to exposure categories.

1. Introduction

On April 20, 2010, an explosion occurred on the *Deepwater Horizon* oil drilling rig, located about 80 km off the coast of Louisiana, resulting in millions of barrels of oil being released into the Gulf of Mexico over a period of several months. After the search, rescue and recovery efforts ended, efforts concentrated on capping the wellhead, capturing released oil and drilling a new well to relieve the pressure on the well casing and blow-out-preventer. The well was "top capped" in mid-July 2010, a measure that stopped the release of the oil. Relief wells were completed and the well was plugged, or "bottom capped", at the depths of the oil formation, to relieve pressure on the original well casing, in early August. In addition to this work, efforts were made to collect or burn the released oil that was spreading across the Gulf primarily north and east of the well site and hitting the coastlines of Louisiana, Mississippi, Alabama and Florida.

As a result of this incident, the Gulf Long-term Follow-up Study (GuLF STUDY) was initiated by the US National Institute of Environmental Health Sciences (NIEHS) to investigate possible adverse health effects to the workers associated with exposure to these spill-related chemicals. Approximately, 33,000 workers participated in the study. These workers were involved in the response and cleanup efforts with possible exposure to a variety of harmful chemicals, including volatile organic compounds (VOCs) and specific chemicals including benzene, toluene, ethyl benzene and xylene. A crucial component of this study is to estimate the exposure levels of the workers to oil-related components.

During the response and clean-up, over 150,000 measurements were collected reflecting full work-shift, primarily oil-component, exposures experienced by workers (personal measurements). In addition, over 25,000,000 measurements of VOC were collected by direct-reading instruments that provided real time air concentrations every few minutes for a variety of locations on the 38 participating vessels (area – i.e., stationary – measurements). These personal and area measurements were only collected for a limited number of study subjects and vessels, respectively; thus some method must be used to assign exposure estimates to workers or vessels that were not measured. Furthermore, many of these measurements are below the measurement instrument's limit of detection (LOD); thus we face a situation where we have a high degree of censoring.

Due to the various challenges associated with data consisting of millions of observations (e.g., computational, administrative, etc.), we chose to conduct a preliminary analysis of a much smaller set of measurement data for the purposes of informing a more thorough investigation of the full data set. Our data consists of 447 area VOC air concentrations that were collected using an AreaRae direct-reading instrument (Rae Systems, San Jose, CA) often at short, regular intervals (e.g., once every 12 min), during the clean-up and response effort at varying times within a day from June 20 to July 20, 2010; these data were collected concurrently with the remaining data, but represent a complete set of data from a particular source. In contrast to the personal exposure measurements, where observations were collected on the individual workers, these data were collected on the *vessels* carrying the workers. As illustrated in Fig. 1, we may observe data from a vessel at a variety of locations throughout the day, and these locations may vary greatly between day; as such, our data are irregular in both space and time, requiring a continuous space, continuous time framework. Furthermore, over 83% of our VOC measurements are below the detectable limit of 0.1 ppm (parts per million); thus, we require techniques for handling censoring in our analysis.

In this paper, we aim to develop a framework for identifying spatiotemporal patterns in air concentrations of VOC, characterized by a high degree of censoring, from area (or stationary) measurement data collected on the vessels supporting the response and clean-up efforts of the oil