Journal of Environmental Management 180 (2016) 264–271

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Research article

Geospatial estimation of the impact of Deepwater Horizon oil spill on plant oiling along the Louisiana shorelines

Pierre Goovaerts [a,*], Cameron Wobus [b], Russell Jones [b], Matthew Rissing [b]

[a] PGeostat LLC, Jerome, MI 49249, USA
[b] Abt Associates, Boulder, CO 80302, USA

ARTICLE INFO

Article history:
Received 9 February 2016
Received in revised form
13 April 2016
Accepted 17 May 2016
Available online 1 June 2016

Keywords:
Marsh
Geographically-weighted regression
Cokriging
Indicator kriging
False positives

ABSTRACT

Stranded oil covering soil and plant stems in fragile Louisiana marshes was one of the most visible impacts of the 2010 Deepwater Horizon (DWH) oil spill. As part of the assessment of marsh injury after the DWH spill, plant stem oiling was broken into five categories (0%, 0–10%, 10–50%, 50–90%, 90–100%) and used as the independent variable for estimating death of vegetation, accelerated erosion, and other metrics of injury. The length of shoreline falling into each of these stem oiling categories was therefore a key measure of the total extent of marsh injury, and its accurate estimation is the focus of this paper. First, we used geographically-weighted logistic regression (GWR) to explore and model spatially varying relationships between stem oiling field data and secondary information (oiling exposure category) collected during shoreline surveys. We then combined GWR probability estimates with field data using indicator cokriging to predict the probability of exceeding four stem oiling thresholds (0, 10, 50, and 90%) at 50 m intervals along the Louisiana shoreline. Cross-validation using Receiver Operating Characteristic (ROC) Curves demonstrate the greater prediction accuracy of the multivariate geostatistical approach relative to either aspatial regression or indicator kriging that ignores secondary information.

1. Introduction

Studies of vegetation death and accelerated marsh erosion following *Deepwater Horizon* (DWH) have shown that both of these impacts (or "injuries," when assessing natural resource damages) can be related to the percent of oiling on the stems of marsh vegetation (e.g., Hester et al., 2015; Silliman et al., 2015). Spatial quantification of these injuries thus relies on estimates of how many kilometers of shoreline fell into each of the four stem oiling categories on which these injury determinations were based (0–10%, 10–50%, 50–90%, 90–100%). Vegetation oiling from the DWH spill was unevenly distributed across Louisiana marsh environments, however, and quantitative measurements of stem oiling were collected only at discrete points (Deepwater Horizon NRDA, 2010a). Spatially continuous observations of shoreline oiling were collected as part of response activities and the natural resource damage assessment (NRDA), and these data were combined into a "shoreline exposure" database for the NRDA (Deepwater Horizon NRDA, 2010b; Nixon, 2015; NOAA, 2013). However, the oiling categories within the shoreline exposure database are qualitative, and do not contain direct information on stem oiling. Furthermore, due to the scope of the DWH spill and the difficulty of finding oil in marshes, these qualitative shoreline surveys sometimes documented segments as "no oil observed" (NOO) in places where more detailed surveys documented oiling at other points in time. Recognizing the relative strengths and limitations of both of these oiling datasets, the goal of this study was to test and apply geospatial methods for combining quantitative point observations of stem oiling with continuous, qualitative observations in the shoreline exposure database to estimate the length of shoreline falling into each of the five stem oiling categories.

One way of quantifying the length of shoreline falling into each stem oiling category would be to assume that measured stem oiling values were evenly distributed within each exposure category, and to calculate the length of each stem oiling category based on proportional assignments within the shoreline exposure framework. However, stem oiling data were often clustered in space, violating the assumption of equal distribution within exposure categories. This is a particular concern for apportionment of stem oiling data that were recorded within the NOO category (i.e. false negatives), since there were thousands of kilometers of shoreline within this

* Corresponding author. PGeostat LLC, 11487 Highland Hills Drive, MI 49249, USA.
E-mail address: goovaerts.pierre@gmail.com (P. Goovaerts).

http://dx.doi.org/10.1016/j.jenvman.2016.05.041
0301-4797/

category but nonzero stem oiling observations within this category were generally clustered in space.

An alternative is to use a geostatistical approach that can account for the geographical location of the environmental data and their spatial correlation, as modeled by variograms (e.g., Goovaerts et al., 2008; Kitsiou and Karydis, 2011). The application of geostatistics to this particular dataset, however, presented several challenges. First, field data were collected with different spatial resolutions and degrees of reliability. Some data represent precise measurements of percentage of stem oiling at specific locations ("hard data"). At other specific locations, we only know whether the vegetation was oiled or not; we do not know the percentage of stem oiling ("soft data"). Finally, although the shoreline exposure database provides a more comprehensive spatial coverage, the oiling descriptors in this dataset are more qualitative with respect to stem oiling ("secondary information").

The second major challenge arises from the complex geometry of the site. Louisiana has a deeply dissected and crenulated marsh coastline, and oil was transported into this region via the bays and inlets that dissect it. As a result, spatial correlation of stem oiling may be more directly related to over-water distance than straight line (Euclidian) distance (Barabás et al., 2001; Money et al., 2009). Third, the heterogeneity of the shoreline and the fact that surveys were conducted by different teams at different times likely impacts locally the relationship between secondary data (oiling exposure category) and percentage of stem oiling measured in the field (i.e. non-stationary relationships). Fourth, the sheer size of the datasets analyzed (e.g. 1100 field data and 118,151 shoreline grid nodes to predict) precluded the use of Bayesian methodologies based on traditional MCMC (Markov chain Monte Carlo) schemes (Gelfand et al., 2003), while more powerful approaches, e.g. the INLA (integrated nested Laplace approximations) methodology, rely heavily on numerical methods and computer programming that are beyond the scope of this study (Martins et al., 2013).

This paper describes the procedure developed to estimate the expected lengths of mainland herbaceous shoreline in Louisiana falling into four stem oiling categories: 0–10%, 10–50%, 50–90%, and 90–100%. The varying reliability of the different pieces of information was integrated using a soft and hard indicator coding of the data (Goovaerts, 1997; Hu et al., 2005), whereas geographically-weighted logistic regression (Fotheringham et al., 2002; Goovaerts et al., 2015; van Donkelaar et al., 2015) was used to explore and model spatially varying relationships between stem oiling field data and secondary information collected during shoreline surveys. Probabilities of exceeding stem oiling thresholds estimated from field measurement and survey data were combined using indicator cokriging (Goovaerts and Journel, 1995). Sensitivity analysis and cross-validation helped guide the choice of optimal sets of parameters and investigate the impact of search strategy and distance metrics on prediction accuracy.

2. Materials and methods

2.1. Datasets

Detailed measurements of stem oiling were collected at 911 discrete points within mainland herbaceous marshes of coastal Louisiana (Fig. 1). These marshes are located along the edges of saline to brackish estuaries and bays throughout Louisiana, and are dominated by the marsh vegetation *Spartina alterniflora*. Stem oiling measurements were collected in the late summer and early fall of 2010, as part of a study referred to as the Marsh Pre-Assessment Study (Deepwater Horizon NRDA, 2010a). At each of these 911 locations, field data were recorded as the percent of stem height oiled, and these raw measurements were then condensed into one of the five stem oiling categories described above (0%, 0–10%, 10–50%, 50–90%, or 90–100%). The marsh pre-assessment dataset also includes 185 additional sites where stem oiling was simply categorized as "oiled" or "not oiled." These 185 "soft" data provide additional information on the spatial distribution of oiled plant stems (see Fig. 1).

Between 2010 and 2013, spatially continuous descriptions of oiling were also collected along the Louisiana shoreline as part of the shoreline cleanup and assessment technique (SCAT) program (Michel et al., 2013; NOAA, 2013). These observations were collected primarily to inform response activities, and summarized oiling along the shoreline using qualitative, categorical descriptors. As a supplement to the SCAT data, spatially continuous, qualitative descriptions of shoreline oiling were also collected for the NRDA as part of the Rapid Assessment program (Deepwater Horizon NRDA, 2010b). Although both of these data sources provide spatially continuous coverage, they did not include detailed measurements of stem oiling.

During the injury assessment phase of the NRDA, the SCAT and Rapid Assessment datasets were combined into a single database referred to as the shoreline exposure database (Nixon, 2015) displayed in Fig. 2. Oiling exposure in this dataset is classified into one of the following four categories: NOO, lighter oiling, heavier oiling, and heavier persistent oiling. For the purposes of our study, the categorical descriptors of shoreline oiling within this dataset are referred to as "secondary information" on stem oiling. 729 of the 911 locations with hard data on stem oiling were co-located with secondary information from the shoreline exposure database (Fig. 3).

The geospatial analysis was thus based on four main types of data:

(1) Measurement of the percentage of plant stem oiling from the pre-assessment dataset (911 "hard" data)
(2) Indicators of presence/absence of plant stem oiling from the pre-assessment dataset (185 "soft" data)
(3) Oiling exposure category ("secondary information") surveyed along approximately 1600 km of mainland herbaceous marsh coastline and at 729 of the 911 hard data locations.
(4) Length of shoreline located within 118,151 50 × 50 m squares discretizing the Louisiana coastline.

2.2. Methodology

The analysis was conducted using the following software: 1) SpaceStat 4.0 (Jacquez et al., 2014) for geographically-weighted regression and variogram modeling, 2) SAS 9.3 (SAS Institute Inc., 2011) for aspatial logistic regression and the creation of ROC curves, 3) SGeMS (Remy et al., 2008) and Gslib (Deutsch and Journel, 1998) for cross-variogram modeling and indicator cokriging, and 4) code written by Dr. Goovaerts for data manipulation and computation of expected lengths of shoreline in different categories of plant stem oiling. The flowchart in Fig. 4 illustrates the main steps in the analysis, as described below.

2.2.1. Indicator coding of plant stem oiling data

The analysis started with the coding of each percentage of stem oiling data into a vector of indicators of exceedance of four thresholds $z_c = 0, 10, 50,$ and 90%. Let $\mathbf{u}_\alpha = (x_\alpha, y_\alpha)$ be a vector of UTM coordinates representing the geographical location of a stem oiling data point, denoted $z(\mathbf{u}_\alpha)$ for hard data and $s(\mathbf{u}_\alpha)$ for soft data. The set of four indicators at any hard data location $\mathbf{u}_\alpha$ was then constructed as: