Environmental Modelling & Software 70 (2015) 113–127

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Linking watershed-scale stream health and socioeconomic indicators with spatial clustering and structural equation modeling

Georgina M. Sanchez [a], A. Pouyan Nejadhashemi [a, *], Zhen Zhang [b], Sandra Marquart-Pyatt [c], Geoffrey Habron [d], Ashton Shortridge [e]

[a] Department of Biosystems and Agricultural Engineering, Michigan State University, 225 Farrall Hall, East Lansing, MI 48824, USA
[b] Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA
[c] Department of Sociology, Michigan State University, East Lansing, MI 48824, USA
[d] Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, USA
[e] Department of Geography, Michigan State University, East Lansing, MI 48824, USA

A R T I C L E I N F O

Article history:
Received 30 December 2014
Received in revised form
20 April 2015
Accepted 22 April 2015
Available online

Keywords:
Spatial cluster
Conditional autoregressive models
Saginaw River Watershed
Stream health indicator
Socioeconomic
Natural-human system

A B S T R A C T

In this study, spatial clustering techniques were used in combination with Structural Equation Modeling (SEM) to characterize the relationships between in-stream health indicators and socioeconomic measures of communities. The study area is the Saginaw River Watershed in Michigan. Four measures of stream health were considered: the Index of Biological Integrity, Hilsenhoff Biotic Index, Family Index of Biological Integrity, and number of Ephemeroptera, Plecoptera, and Trichoptera taxa. The stream health indicators were predicted using nine socioeconomic variables that capture vulnerability in population. The results of spatial clustering showed that incorporating clustering configuration improves the model prediction. A total of 510 Confirmatory Factor Analysis (CFAs) and 85 multivariate regression models were developed for each spatial cluster within the watershed and compared with the model performance without spatial clustering (at the watershed level). In general, watershed level CFAs outperformed cluster level CFAs, while the reverse was true for the regression models.

1. Introduction

Food, water, clean air, shelter, and relative climatic constancy are provided by ecosystem services and are essential for human well-being (Corvalan et al., 2005). There is broad consensus that anthropogenic activities significantly affect ecosystems services and functions (Vitousek et al., 1997; McMichael et al., 2003; Halpern et al., 2008). Anthropogenic activities such as agriculture, fishing, livestock production, energy generation, and transportation can negatively affect the environment and do so in different ways.

Meanwhile, changes to ecosystem functions due to human development rebound on the well-being and health of human populations. Ecosystem functions and health have a direct impact on key economic and social issues: livelihood, income, local migration, political conflicts, public health, and development (Montgomery et al., 1973; Corvalan et al., 2005; IPHI, 2012). In this regard, several attempts have been made to simulate cost of ecosystems services due to anthropogenic activities. Sun and Műller (2013) developed decision making framework for carbon-based payment based on a coupled landuse and agent-based model. A more comprehensive system was developed by Ausseil et al. (2013). This indicator-based system is an approach to resource management considering different aspects of ecosystems services including: soil erosion, clean water, water availability, climate regulation, and food and fiber. Landuyt et al. (2013) introduced a semi-quantitative modeling approach (the Bayesian belief networks) capable of combing unstructured knowledge with empirical data. However, evaluation of different aspects of ecosystems services is difficult due to problems associated with integration of decision and valuation nodes.

The World Health Organization constitution states that “the enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being …”; no further justification is required to understand that every individual should have equitable protection from environmental hazards. For this reason, environmental justice should be a priority for human communities. The United States Environmental Protection Agency later expanded this definition by reinforcing that the protection from

* Corresponding author. Tel.: +1 517 432 7653; fax: +1 517 432 2892.
E-mail address: pouyan@msu.edu (A.P. Nejadhashemi).

http://dx.doi.org/10.1016/j.envsoft.2015.04.012
1364-8152/

environmental and health hazards should be fairly distributed regardless of race, color, national origin, or income (EPA, 2010). However, not all communities are identical. There are distinctive geographic distributions of racial and ethnic population (U.S. Census Bureau, 2010), income (DeNavas-Walt et al., 2013), political power, and susceptibility to natural disasters. Further, additional factors are salient for research on vulnerable populations or those communities disproportionally affected by environmental hazards and risks. Examples of these factors are economic and social status, psychosocial, cultural, and religious background, and physical, chemical or biological environmental conditions (Morris, 2010; Burger and Gochfeld, 2011). The combined effects of these factors can also lead a community to be described as vulnerable where it is unequally exposed to an environmental hazard (Burger and Gochfeld, 2011).

Improvement of water resources management will require integration of environmental, social, and economic dimensions (UNEP, 2009). Holistic watershed management accomplishes this by considering these three dimensions concurrently (Pahl-Wostl et al., 2008). However, their integration in a holistic manner is not without challenges. First, social and environmental impact assessments are generally independent (Slootweg et al., 2001; Tolun et al., 2012). Second, these systems are dynamic in nature; they vary spatially, temporally, and across organizational units (Liu et al., 2007). Third, populations are disproportionately affected by environmental hazards (Maantay, 2002; Nweke, 2011). Fourth, the interrelationships between human and natural systems are complex and largely unknown (Liu et al., 2007). In light of these barriers, spatial analyses that address the complexity of socio-ecological systems and illustrate environmental justice in watershed management are lacking (Nowak et al., 2006; Silver, 2008; Gallo and Goodchild, 2012).

To better assess environmental justice issues, various techniques have been implemented in the past few decades. Modeling is a useful approach for explaining how social and environmental factors interconnect (Chakraborty et al., 2011). Moreover, models can be a powerful technique to track environmental health disparities by characterizing complex systems that take place in multilevel transdisciplinary contexts (Gibbons et al., 2007). Environmental justice models vary with regard to categorization as spatial and conceptual models.

In a spatial model, the first step is to define the scale of analysis and an adequate spatial unit (Maantay, 2002). Commonly used spatial units are county, census tract, ZIP code, block, and neighborhood (Lobao et al., 2007). A challenge, however, is that different types of data come in different units and forms. For example, population socio-demographic data can be in aggregated areal form such as census tracts, unlike environmental data in the form of point patterns from monitoring stations or point source discharges (Chakraborty et al., 2011). The second step is to define the method of analysis. Each method presents a different understanding of where the hazard can be found, and therefore, which community is affected (Maantay, 2002; Mohai and Saha, 2006). Understanding the dynamic behavior of the study exposure hazard is a third step. Environmental hazards can have unique diffusion patterns depending on the media; for example, pollutants present in air or water travel differently through space (Kjellstrom et al., 2006). Once the hazard and the transport media have been identified, techniques such as dispersion models can be linked to proximity analysis to better estimate the exposure risk (Maroko, 2012). Finally, geographic information systems, spatial statistics, and population estimation techniques can be integrated to develop better estimations for the spatial model.

Spatial models have many advantages. They type of models integrate visualization with exploration and statistical techniques to identify trends and significant relationships for environmental justice assessments (Jerrett et al., 2010). Furthermore, spatial modeling simulates different scenarios to test the impact of given conditions. Finally, spatial modeling techniques allow simulation of multiple emission sources, allowing cumulative risk analysis (Chakraborty et al., 2011). There are some limitations to spatial models. First, it is a challenge to find data at a spatial resolution capable of correctly representing the relationship between environmental health hazards and socio-demographic characteristics. Second, the nature of the relationship between environmental hazards and social disparities can be a function of the spatial unit examined (Sanchez et al., 2014). A third limitation is that spatial models assume the hazard is distributed uniformly within the area of exposure (Maantay, 2002). Finally, socio-demographic data is generally represented in an aggregated form (e.g. country, county, census tracts, and blocks). In higher level units (e.g. country, county) the data will be less accurate than lower level units (e.g. census tracts, blocks) in serving as an indicator of environmental health disparity (Lobao et al., 2007).

In conceptual models, the first step is understanding the theory behind the system. An extensive literature review is required to link significant social and environmental factors (Helfand and Peyton, 1999; Burger and Gochfeld, 2011; Linder and Sexton, 2011). Examples of input factors that have been used in conceptual models are economic indices, social engagement, behavioral responses, psychological factors, ethnicity, and physical or biological environmental conditions (Linder and Sexton, 2011). Finally, the conceptual model is instantiated by creating pathways that connect the factors. These pathways serve as tracking systems identifying hazard sources and the impacted populations (Helfand and Peyton, 1999). Conceptual models have several advantages. First, a conceptual model can elucidate the physical processes that take place between society and nature (WHO, 2009). Second, conceptual models allow researchers to track hypothetical causes of inequalities by integrating multiple factors and processes (Diez Roux, 2012). A third advantage is that conceptual models allow tracking hazards at different locations: at the source of contamination, where environmental media and transport mechanism occurs, at exposure locations, and at the receptor population (Burger and Gochfeld, 2011). However, there is a debate among researchers regarding the implementation of conceptual models (Linder and Sexton, 2011). The model construction and data use is subjective, as well as the results, where interpretation may vary between disciplines (Diez Roux, 2012).

While a range of modeling techniques have attempted to simulate human-nature systems; many challenges remain that require innovation and multidisciplinary collaboration towards a better decision making process (Wandersee et al., 2012). The National Institute of Environmental Health Sciences combined geospatial framework and data mining to monitor the impacts of Hurricane Katrina (Pezzoli et al., 2007). This portal was successfully used to track environmental consequences of natural and man-made disasters. Hierarchical Linear Modeling is an extension of a linear regression model that can handle variability at multiple levels of nesting (Snijders and Bosker, 2012). Cheruvelil et al. (2008) used this technique to capture the variation in water quality for 479 lakes in Michigan. As a results, HLM effectively explained 3%–60% of the variation for lake water chemistry and clarity. In this study, we account for complex relations in the data and build our conceptual model by applying structural equation modeling (SEM). SEM has advantages regarding its ability to incorporate direct and indirect effects, reciprocal relations, feedback loops, and observed and latent variables (Bollen, 1989; Paxton et al., 2011; Schumacker and Lomax, 2010). We use confirmatory factor analysis (CFA) in SEM to create latent variables and examine their explanatory