Ecological Indicators 45 (2014) 340–349

Contents lists available at ScienceDirect

Ecological Indicators

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

Classification of watersheds into integrated social and biophysical indicators with clustering analysis

Alex Mayer[a,*], Richelle Winkler[b], Lauren Fry[c]

[a] Department of Civil & Environmental Engineering, Michigan Technological University, 1400 Townsend Drive, Houghton, MI, USA
[b] Department of Social Sciences, Michigan Technological University, Houghton, MI, USA
[c] Cooperative Institute for Limnology and Ecosystems Research, University of Michigan, At NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, MI, USA

ARTICLE INFO

Article history:
Received 4 November 2013
Received in revised form 17 April 2014
Accepted 21 April 2014

Keywords:
Watershed hydrology
Watershed indicators
Water quality
Socio-ecological systems
Multivariate analysis

ABSTRACT

In this work, we classify watersheds in the US portion of the Great Lakes basin according to a wide range of social and environmental characteristics. Classified watershed indicators serve to provide organizing principles for prescribing effective management strategies and for developing regional scale monitoring and modeling efforts. Classifications also provide a means for synthesizing seemingly disparate ecological attributes into powerful indicators. We use a robust watershed classification scheme based on cluster analysis that integrates a set of 12 social and environmental factors chosen to reflect the state of water resources in the Great Lakes basin. We found five statistically distinct classified watershed indicators: Urban Centers, Intensive Agriculture, Cultivated Rural, Northwoods, and Lakes Destinations. Within these classifications, we distinguished relationships between impacts on water resources and biophysical, demographic, land-use, and social characteristics of the landscape. We found that agricultural areas can be divided into those with high and low water impact, and that watersheds with considerable influence of seasonal homes are further distinguished into watersheds with inland lakes and relatively high socioeconomic status, contrasted with watersheds with wetlands and relatively low socioeconomic status.

1. Introduction

Watersheds increasingly serve as organizing units for assessing and managing human impacts on the environment. Watersheds are, for instance, the locus of natural resource management efforts (i.e. O'Neill, 2005; Duram and Brown, 1999), biophysical analyses of water systems (e.g. Serveiss, 2002), and analyses of social-environment interactions (e.g. Floress et al., 2011). However, the processes through which biophysical and social variables affect water systems are complex and vary across space and time, complicating efforts to understand general relationships and practices to conserve and restore water resources (McDonnell et al., 2007; O'Neill, 2005).

Indicators of the ecological status of watersheds can be useful for synthesizing the interacting biophysical and social structures of the landscape. Like all ecological indicators, watershed indicators can be used to assess the condition of the environment to monitor temporal trends in conditions or to detect the basis of an environmental problem (Dale and Beyeler, 2001). Ultimately, the value of a watershed indicator is to enable data-driven management decisions (Turnhout et al., 2007). In this work, we develop an indicator that is a synthetic classification of watersheds, based on social and environmental attributes, to inform management decisions and design of regional assessments of watershed health (Wolock et al., 2004). Each watershed is unique with specific social and environmental characteristics that affect hydrologic processes, which in turn influence social systems (Chess and Gibson, 2001, O'Neill, 2005). Biophysical characteristics (e.g. topography, soil, geology, and vegetation), anthropogenic influence (e.g. urban development, agricultural cultivation, and water withdrawals), and social structure (population size, socioeconomic, and institutional differences) each contribute to their uniqueness (Beven, 2000). Watersheds are complex social–ecological systems with multifaceted, co-evolving biophysical, social, and technological processes that constitute the interactions between ecosystem function, ecosystem services, socioeconomic systems, and the stressors they generate (Berkes et al., 2003; Ostrom, 2009). This heterogeneity and complexity makes it difficult to develop generalizable integrated science

* Corresponding author. Tel.: +1 906 487 3372; fax: +906 487 2643.
E-mail address: asmayer@mtu.edu (A. Mayer).

http://dx.doi.org/10.1016/j.ecolind.2014.04.030
1470-160X/

approaches for assessing and managing watersheds. One approach to seeking order in heterogeneous systems such as watersheds is through systematic classification (McDonnell and Woods, 2004; Wagener et al., 2008). Scientists routinely categorize features of the world into taxonomic groups as a means of reducing the complexity of biophysical and social systems into meaningful and comprehensible types to facilitate communication and promote understanding.

While every watershed is distinct, classifying watersheds into meaningful groups of relatively similar cases can provide useful indicators of biophysical and social conditions and their impacts on water systems. This strategy can be useful for developing hydrologic models, designing observational networks, and conducting socio-economic analyses of regional watersheds. Classification of watersheds allows for the selection of "model" watersheds to be studied intensively, followed by regional scaling up of study results. Classification also allows for the characterization of watershed behavior, e.g. flood flows, at data-poor sites by using information aggregated from nearby hydrologically watersheds with extensive data records (Ilorme and Griffis, 2013). Classification of watersheds can also be useful in management contexts. For instance, these classifications could facilitate the borrowing of effective management policies across watersheds that experience similar social and environmental conditions (Wardrop et al., 2007).

Natural scientists have suggested several watershed classification schemes. Hydrologic classification schemes for watersheds have been proposed on the basis of fundamental hydrologic process descriptions (Sivakumar, 2008; Wagener et al., 2007) and spatial distributions of selected hydrologic parameters (Wolock et al., 2004; Wardrop et al., 2007). River system classification schemes that describe ecological characteristics and function in rivers are based on landscape-scale attributes, such as climate, topography, geology, and land cover (Hawkins et al., 2000) and fluvial geomorphology attributes (Snelder et al., 2004). In the Gages II dataset, which provides geospatial data and classifications for U.S. Geological Survey (USGS)-maintained stream gages, a "hydrologic disturbance index" is attributed to each basin (Falcone, 2011). In these cases, classifications provide an important organizing principle to complement hydrologic modeling and experimental approaches by providing guidance on the similarities and differences between watersheds, as well as to offer assessments of the potential impacts of human activities and climate change at the watershed scale (Wagener et al., 2007).

These efforts integrate sophisticated biophysical watershed dimensions, but there has been little effort to integrate multidisciplinary watershed attributes into a holistic framework. Since ecological indicators are applied to balance social, economic and ecological dimensions in decision making (Turnhout et al., 2007), it is appropriate to include not only biophysical attributes, but also socioeconomic characteristics in classified watershed indicators. Social attributes and related human impacts on water resources vary considerably across watersheds and are crucial to understanding policy contexts, but these components have generally been ignored in watershed classifications. Hutchinson et al. (2010) is one important exception. They integrate basic demographic, land cover, and water use variables with climate and physical attributes to classify hydrologically similar units. We are not aware of any watershed classification approach that incorporates social structural variables (beyond population counts/densities and land use) that could, for example, affect the capacity and interest of the population to respond to environmental impact or to effectively collaborate on watershed management teams. Yet we know from the cumulative findings of a rich social science literature in environmental sociology (more generally) and watershed governance (more specifically) that human communities are diverse in how they use water resources, their interest in conserving and restoring environmental resources, and their capacity to facilitate change (O'Neill, 2005; Rathwell and Peterson, 2012).

This study is unique in that we specifically integrate biophysical, demographic, land-use, and social structural variables in a classified watershed indicator approach aimed at facilitating policy development and implementation. We apply multivariate cluster analysis to quantitatively classify watersheds across the US portion of the Laurentian Great Lakes basin. The Great Lakes basin offers a rich diversity of biophysical and socioeconomic conditions and is a regional socio-ecological system of national and global significance. While we focus on a specific region, a strength of this research is that we introduce a methodologically and conceptually rigorous approach to cluster analysis that can overcome potential issues of subjectivity and challenges for interpretation that have been raised for this statistical method (e.g. Caratti et al., 2004) and can be applied to study any regional basin or employed with alternative geographic units. The classification scheme reveals important insights into the connections between socioeconomic attributes and ecological conditions of watersheds. The classifications also result in the synthesis of biophysical and socioeconomic attributes into a single environmental indicator, which is a critical need (King et al., 2005).

2. Methods

A wide variety of multivariate statistical methods have been employed for classification, such as multivariate discriminant analysis, logistic regression, neural networks, and regression or classification trees. Clustering algorithms are a variety of data mining methods that attempt to group objects (in this case, watersheds) of similar kind into respective classifications by assessing the degree to which objects within the same class are more similar to each other than objects in different classes. Clustering offers advantages over other multivariate statistical analysis methods in that the underlying principles are relatively easy to grasp, it is capable of finding structures directly from the given data, without relying on prescribed hierarchies, and it can robustly integrate attributes with heterogeneous properties, such as continuous vs. categorical variables. Several clustering algorithms have been used to classify watersheds (primarily based on biophysical attributes), including the iterative self-organizing data analysis technique (Hutchinson et al., 2010), hierarchical agglomerative clustering (Caratti et al., 2004), mean similarity dendrograms (Snelder et al., 2004), nearest neighbor chain algorithm (Wolock et al., 2004), and Ward's clustering technique (Caratti et al., 2004).

While clustering algorithms are attractive, they should be used with caution. The number of clusters is usually set by the investigator because the internal structure of the data may not lend itself to an obvious number of classifications (Gauch and Whittaker, 1981). Guidelines for choosing the best clustering technique are unclear. Since no classification algorithm is perfect, the success of a classification technique can only be judged by comparing its results with those derived from other similar evaluations. Ideally, this requires that several classification techniques be applied to the same dataset and the results compared (Gauch, 1982; Nathan and McMahon, 1990). In practice, classification remains partially an art that relies on the investigator's experience and insight (Caratti et al., 2004).

Following best practices (Romesburg, 2004; Morton and Padgitt, 2005; Everitt et al., 2011), we rely on theory and prior research in making decisions about which variables to include in the cluster analysis and in judging the interpretability of clustering results. Furthermore, we conduct sensitivity analyses of multiple classification schemes and number of clusters, testing each against robust performance measures.