Journal of Environmental Management 92 (2011) 1215–1224

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Reliability of groundwater vulnerability maps obtained through statistical methods

Alessandro Sorichetta [a], Marco Masetti [a,*], Cristiano Ballabio [b], Simone Sterlacchini [c], Giovanni Pietro Beretta [a]

[a] Dipartimento di Scienze della Terra - Ardito Desio, Università degli Studi di Milano, Via Mangiagalli 34, 20133 Milan, Italy
[b] Dipartimento di Scienze Ambientali e del Territorio, Università degli Studi di Milano-Bicocca, Piazza della Scienza 1, 20126 Milan, Italy
[c] Instituto per la Dinamica dei Processi Ambientali, Centro Nazionale delle Ricerche (CNR-IDPA), Piazza della Scienza 1, 20126 Milan, Italy

ARTICLE INFO

Article history:
Received 13 May 2010
Received in revised form
9 November 2010
Accepted 11 December 2010
Available online 5 January 2011

Keywords:
Groundwater vulnerability
Statistical methods
Land use management
Spatial agreement
Validation procedure

ABSTRACT

Statistical methods are widely used in environmental studies to evaluate natural hazards. Within groundwater vulnerability in particular, statistical methods are used to support decisions about environmental planning and management. The production of vulnerability maps obtained by statistical methods can greatly help decision making. One of the key points in all of these studies is the validation of the model outputs, which is performed through the application of various techniques to analyze the quality and reliability of the final results and to evaluate the model having the best performance. In this study, a groundwater vulnerability assessment to nitrate contamination was performed for the shallow aquifer located in the Province of Milan (Italy). The Weights of Evidence modeling technique was used to generate six model outputs, each one with a different number of input predictive factors. Considering that a vulnerability map is meaningful and useful only if it represents the study area through a limited number of classes with different degrees of vulnerability, the spatial agreement of different reclassified maps has been evaluated through the kappa statistics and a series of validation procedures has been proposed and applied to evaluate the reliability of the reclassified maps. Results show that performance is not directly related to the number of input predictor factors and that is possible to identify, among apparently similar maps, those best representing groundwater vulnerability in the study area. Thus, vulnerability maps generated using statistical modeling techniques have to be carefully handled before they are disseminated. Indeed, the results may appear to be excellent and final maps may perform quite well when, in fact, the depicted spatial distribution of vulnerability is greatly different from the actual one. For this reason, it is necessary to carefully evaluate the obtained results using multiple statistical techniques that are capable of providing quantitative insight into the analysis of the results. This evaluation should be done at least to reduce the questionability of the results and so to limit the number of potential choices.

1. Introduction

Statistical methods are widely used in environmental studies to evaluate natural hazards to support decisions necessary for environmental planning and management. These studies are generally based on an analysis of the spatial correlation between a specific event and a series of predisposing factors that can affect the occurrence of the event. Among these studies, landslide hazard and groundwater vulnerability are probably the most common topics (Carrara and Pike, 2008; Twarakavi and Kaluarachchi, 2006). The final results of these studies are represented by maps of the study area identifying zones with different degree of susceptibility to hazardous events. Various methods, such as logistic regression (Tesoriero and Voss, 1997; Nolan et al., 2002; Gardner and Vogel, 2005), discriminant analysis (Carrara, 1983), likelihood ratio functions (Chung, 2006) and Weights of Evidence (van Westen et al., 2003; Neuhäuser and Terhorst, 2007; Poli and Sterlacchini, 2007; Masetti et al., 2008) have been used either as a single approach to a specific study area or for comparison (Harris et al., 2003).

One of the key features in all of these studies is the validation of the model outputs (Brenning, 2005). Validation is usually performed through the application of several techniques to analyze the quality and reliability of the final results and to evaluate the model that has the best performance. These techniques are crucial to

* Corresponding author. Tel.: +39 0250315507; fax: +39 0250315494.
E-mail address: marco.masetti@unimi.it (M. Masetti).

0301-4797/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2010.12.009

evaluating the quality of the results. As reported by Fabbri and Chung (2008), the current emphasis in the field should be on the development and application of reliable and robust validation procedures rather than in new methods.

Although many validation techniques can be used (Fabbri and Chung, 2008), little attention has been paid to the comparison of results obtained using the same technique, especially in terms of the spatial agreement of predicted patterns. Final maps from different combinations of predictors can lead to maps with different spatial distributions of hazard zones but that have a quality that is apparently similar (Fabbri et al., 2003). This fact is extremely important because a situation may occur when two or more maps with similar predictive power show poor agreement in terms of the predicted spatial patterns. This discrepancy generates numerous problems when disseminating the final results to the users including how to use maps to develop efficient land use policies.

Groundwater vulnerability assessment requires a particular approach to evaluating the quality of the final results. In landslide hazard assessment, validation procedures can be rightfully concentrated on the ability of the method to correctly identify classes with the highest hazard values, with less interest on the classes with low values. However the situation is different in groundwater vulnerability assessment, which is not a property that can be directly measured or surveyed in the field (Gogu and Dassargues, 2000). Every aquifer can be considered vulnerable, even if some areas are more vulnerable to groundwater contamination than others (Vrba and Zaporozec, 1994). Therefore, the reliability of a model that allows vulnerability assessment should be evaluated for the entire study area and not only on its ability to correctly identify the most vulnerable zones.

In this paper a groundwater vulnerability study has been performed that analyzes two main factors: a) the application of a series of validation techniques to evaluate the best map; and b) the spatial agreement among maps showing similar performances, but different predicted patterns, in order to demonstrate that a rigorous validation procedure can help minimize uncertainty in the choice of the best representative map.

The procedure has been applied to an unconfined aquifer in northern Italy, which is located within an area that has been classified as vulnerable by the European Community since 1991 (EU, 1991)

The aim of the study is to evaluate the spatial agreement among multiple maps and to analyze the application of a series of validation techniques to estimate the reliability of each one of them. This was done in order to identify the map best representing the actual distribution of the nitrate contamination in the study area. Such map has been indicated as the most suitable to be used for environmental planning and management.

2. Methods

2.1. Statistical method

The Weights of Evidence (WofE) modeling technique combines different spatial datasets in a Geographical Information System (GIS) environment to analyze and describe their interactions and generate predictive patterns (Bonham-Carter, 1994, p. 398; Raines et al., 2000). WofE can be defined as a data-driven Bayesian method in a log-linear form that uses known occurrences representing the response variable (i.e., impacted wells) as training sites (training points) to produce predictive probability maps (response themes) from multiple weighted evidences (evidential themes representing explanatory variables or predictor factors) influencing the spatial distribution of the occurrences in the study area (Raines, 1999). WofE is generally applied on a raster basis, using equal area pixels. Training points (TPs) are used in WofE to calculate the prior probability, the weights for each class representing a different range of values of each generalized evidential theme, and the posterior probability values in the response theme. Prior probability is based on prior knowledge of the TPs location in the study area. Prior probability is simply defined by the ratio between the area containing occurrences (i.e., the number of pixels containing a training point) and the total area (i.e., the total number of pixels). Thus, the prior probability represents the probability that a pixel within the study area contains an occurrence (i.e., impacted well) without considering any evidential themes, and can be expressed as:

$$P\{D\} = \frac{ND}{NT} \tag{1}$$

where ND and NT are respectively the number of pixels containing a training point and the total number of pixels in the study area.

For each class of each evidential theme, a positive and a negative weight, W^+ and W^-, are computed based on the location of the TPs with respect to the study area. Thus, for a given class B, W^+ and W^- would be, respectively, positive and negative or negative and positive depending on whether B has more or fewer TPs than would be expected by chance.

The weights can be expressed as:

$$W^+ = \log_e \frac{P\{B|D\}}{P\{B|\overline{D}\}} \tag{2}$$

$$W^- = \log_e \frac{P\{\overline{B}|D\}}{P\{\overline{B}|\overline{D}\}} \tag{3}$$

where $P\{B|D\}$ and $P\{B|\overline{D}\}$ are respectively the probability of a pixel of being in the class B when the same pixel contains or does not contain a training point, and $P\{\overline{B}|D\}$ and $P\{\overline{B}|\overline{D}\}$ are respectively the probability of a pixel of not being in the class B when it contains or does not contain a training point.

The contrast (positive weight minus negative weight) represents the overall degree of spatial association between each class of a given evidential theme and TPs and, thus, it is a measure of the usefulness of the considered class in predicting the location of TPs (Raines, 1999).

A confidence value for the ratio between the contrast and its standard deviation must be selected to provide a useful measure of the significance of the contrast (Raines, 1999). For this study, a confidence value of 1.654, corresponding approximately to a 95% level of significance, was chosen as the minimum acceptable value to consider an evidential theme class as statistically significant.

The posterior probability represents the relative probability that a pixel contains an occurrence (i.e., impacted well) based on the evidences provided by the evidential themes (i.e., the calculated weights). The posterior probability can be expressed as:

$$\log_e O\left\{D \middle| B_1^k \cap B_2^k \cap B_3^k \ldots B_n^k\right\} = \sum_{j=1}^{n} W_j^k + \log_e O\{D\} \tag{4}$$

where k is either + or - if the pixel is, respectively, inside or outside the class B_n (k refers to), and $O\{D\}$ equals $P\{D\}/(1-P\{D\})$ is the odds form of $P\{D\}$.

Relative probability means that a pixel of higher posterior probability is more likely to contain an occurrence than a pixel of lower one (Raines, 1999). Thus, the posterior probability does not represent the actual probability that a pixel contains an occurrence.

Using this modeling technique, a groundwater vulnerability assessment of nitrate (NO_3^-) contamination has been performed in the shallow unconfined aquifer of the Province of Milan (northern Italy), where groundwater nitrate concentrations, constantly monitored by a net of about 300 wells (Fig. 1), reaches values higher than 50 mg/l.