GROUNDWATER VULNERABILITY ASSESSMENT USING POSITIVE AND NEGATIVE WEIGHTS-OF-EVIDENCE METHODS TO CORRECT FOR SAMPLING BIAS

Alessandro Sorichetta*[1] and Gilpin R. Robinson, Jr.[2]
1. Dipartimento di Scienze della Terra "A. Desio', Università degli Studi di Milano, Milano, Italy, alessandro.sorichetta@unimi.it
2. National Center, US Geological Survey, Reston VA, USA, grobinso@usgs.gov
*** Presenting author**

Key-words: *Groundwater vulnerability assessment; Weights of Evidence; Sampling bias; Adjusted contrast, Map reliability.*

1. Introduction

Nowadays, statistical methods are extensively used in geosciences and in many other fields for addressing spatially-related issues. The Weights-of-Evidence (WofE) method [1], following its first applications for assessing groundwater vulnerability in the early 2000's [2, 3], has been increasingly used over the years in the field of contaminant hydrogeology [4, 5, 6, 7, 8, 9].

The WofE can be defined as a data-driven Bayesian method, expressed in a log-linear form, that uses known-occurrences of an event (i.e., response variable) as training points (TPs) to define the spatial association (i.e., contrasts) between the occurrences and multiple weighted evidences (i.e., explanatory variables), in order to generate predictive probability outputs (i.e., response themes)

Its use requires to express the response variable as binary and to select a threshold distinguishing between positive and negative indicators of contamination that are usually identified as occurrences and non-occurrences, respectively.

The traditional approach when using statistical methods estimating the conditional probability of occurrence of an event, such as the WofE, uses only positive indicators as TPs in the analysis. However, this approach may be prone to unrecognized sample bias if care is not taken to control or correct for non-random variation in sampling density (for example more monitoring-well may have been placed in known contaminated areas than in other areas). Thus, in this study both positive and negative indicators were used as TPs (positive and negative TPs, respectively) in the WofE analysis and two original quantitative methodologies to recognize sample bias and correct for its effects on the resulting groundwater vulnerability maps were compared and successfully tested.

2. Study area and methods

The new approach briefly described in the Introduction was used to assess groundwater vulnerability to nitrate contamination of the shallow, unconfined, porous aquifer located within the provinces of Milan and Monza-Brianza (Fig 1).

The first methodology to correct contrasts for sampling bias consists of subtracting the ones calculated using the negative TPs (NegC; Tab. 1) from the ones calculated using the positive TPs (PosC; Tab. 1). This is similar, in some ways, to a Bayesian variation of the odds ratio formulation used in logistic regression.

The second methodology consists of subtracting the contrasts calculated using all monitoring-wells as TPs (AllC; Tab. 1) from the ones calculated using the positive TPs (PosC; Tab. 1). Indeed, since in an ideal random-sampling setting AllC values would be near zero for all evidence classes, AllC values significantly different from zero represent a measure of sample bias.

Uncorrected and corrected contrasts were then used to produce three response themes (Fig. 1) that were calibrated/validated and compared each other to evaluate the effects of sample bias on the resulting vulnerability maps.

In this study contrasts and response themes were obtained using the Spatial Data Modeler extension for ArcMAP 9.3 [10].

3. Results and conclusions

Results showed that, due to a sample bias with respect to their evidence classes, explanatory variables could appear to be good and statistically significant predictors of both types of occurrences showing an equivocal relationship with the presence of the positive and the negative indicators of contamination.

Furthermore, comparisons among the uncorrected groundwater vulnerability map (Fig 1a) and the two corrected ones (Fig. 1b, 1c) demonstrated that if sampling bias is not recognized and corrected, when assessing groundwater vulnerability by methods estimating the conditional probability of occurrence of an event, the use of such evidential themes in the analysis can produce unreliable maps. However, once the spatial associations between the TPs and the evidence classes of each explanatory variable were corrected for sampling bias effects, the WofE was found to be a reliable modeling technique for assessing groundwater vulnerability and proved to be capable of identifying areas characterized by different degrees of vulnerability.

References

[1] Bonham-Carter G.F. (1994). *Geographic Information Systems for Geoscientists – Modelling with GIS.* Pergamon Press, pp. 398.

[2] Alberti L., De Amicis M., Masetti M. & Sterlacchini S. (2001). *Bayes' rule and GIS for evaluating sensitivity of groundwater to contamination.* In: Proceedings of 2001 Annual Conference of the International Association for Mathematical Geology (IAMG), Cancun, Mexico, September 6-12, 2001.

[3] Baker A.E., Cichon J.R., Arthur J.D. & Raines G.L. (2002). *Florida vulnerability assessment.* In: Abstracts with Programs – Annual Meeting of the Geological Society of America, Denver, Colorado, October 27-30, 2002, 346-347.

[4] Antonakos A.K. & Lambrakis N.J. (2007). *Development and testing of three hybrid methods for the assessment of aquifer vulnerability to nitrates, based on the drastic model, an example from NE Korinthia; Greece.* Journal of Hydrology; 333; 288-304.

[5] Arthur J.D., Wood H.A.R., Baker A.E., Cichon J.R. & Raines G.L. (2007). *Development and implementation of a Bayesian-based aquifer vulnerability assessment in Florida;* Natural Resources Research; 16; 93-107.

[6] Masetti M., Poli S. & Sterlacchini S. (2007). *The use of the weights-of-evidence modeling technique to estimate the vulnerability of groundwater to nitrate contamination;* Natural Resources Research; 16; 109-119.

[7] Masetti M., Poli S., Sterlacchini S., Beretta G.P. & Facchi A. (2008). *Spatial and statistical assessment of factors influencing nitrate contamination in groundwater;* Journal of Environmental Management; 86; 272-281.

[8] Uhan J., Vizintin G. & Pezdic J. (2011). *Groundwater nitrate vulnerability assessment in alluvial aquifer using process-based models and weights-of-evidence method: Lower Savinja Valley case study (Slovenia);* Environmental Earth Sciences, 105; 64-97.

[9] Sorichetta A., Masetti M., Ballabio C., Sterlacchini S. & Beretta G.P. (2011). *Reliability of groundwater vulnerability maps obtained through statistical methods;* Journal of Environmental Management; 92; 1215-1224.

[10] Sawatzky D.L., Raines G.L., Bonham-Carter G.F. & Looney C.G. (2009). *Spatial Data Modeler (SDM): ArcMAP 9.3 geoprocessing tools for spatial data modelling using weights of evidence, logistic regression, fuzzy logic and neural network (http://www.ige.unicamp.br/sdm).*

Groundwater depth evidence class	Evidence class range (mm/y)	PosC	First approach		Second approach	
			NegC	Corrected PosC (PosC minus NegC)	AllC	Corrected PosC (PosC minus AllC)
1	<220	1.11	-0.22	1.33	0.58	0.53
2	221-350	1.10	0.91	0.18	1.01	0.09
3	351-1000	-1.01	-0.17	-0.84	-0.58	-0.44
4	>1000	-1.13	-0.67	-0.46	-0.88	-0.25

Tab.1 – Example of correction for sampling bias of the contrasts calculated using the positive TPs (3rd column) by applying the first (4th and 5th column) and the second methodology (6th and 7th column) described in the text. The example refers to the effective infiltration evidential classes considered in this study (1st and 2nd column).

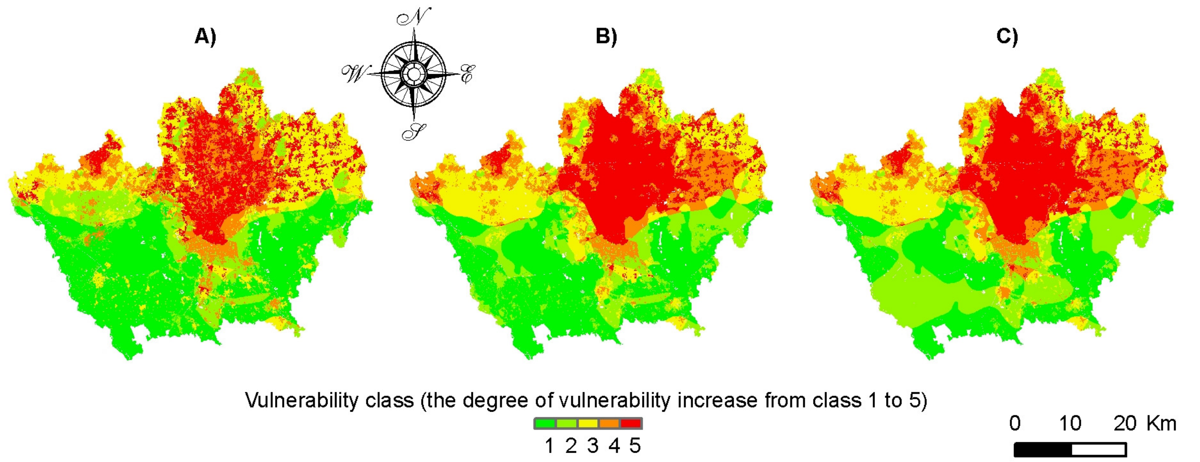

Fig. 1 – A) Groundwater vulnerability map obtained using the positive contrasts uncorrected for sampling bias. B) and C) Groundwater vulnerability maps obtained using the positive contrasts corrected for sampling bias by applying, respectively, the first and the second methodology described in the text.