Report for 2004FL59B: Ground Water Vulnerability Delineation Using Integrated GIS and Neuro-Fuzzy Methods

- Articles in Refereed Scientific Journals:
    - Application of Neural Networks, Support Vector Machines and Logistic Regression to predict contaminated wells: A comparative case study of Polk County, Florida. In preparation (to be submitted to Journal of Environmental Quality).
- Other Publications:
    - Dixon, B. 2004. Ground Water Vulnerability Mapping Tool: NN and fuzzy logic: one, the other, or both?? Presentation. AWRA's 2004 Spring Specialty Conference Geographic Information Systems (GIS) and Water Resources III Water. Nashville, May 2004.
    - Candade, N. and Dixon, B. 2004. Comparison of Neural Network and Neuro-fuzzy Techniques in Ground Water Vulnerability Mapping: A Case Study. Poster. AWRA's 2004 Spring Specialty Conference Geographic Information Systems (GIS) and Water Resources III Water. Nashville, May 2004.
    - Stetson R, Dixon, B. and Candade, N. 2005. Comparison of various krigging methods for contaminated wells in Tampa Bay region FL, Florida Society of Geographers, Annual meeting. Orlando, Feb. 2005.
    - Candade, N and Dixon, B. 2004. Integrated Vulnerability Assessment of Ground Water for Hillsborough County, Florida: A Case Study. AWRA Annual Conference, Orlando, November.
    - Dixon B. and Candade N. 2005. Groundwater Contamination Mapping Using Integrated GIS and Neural Networks: A Sensitivity Analysis. Presentation. International Conference on Environmental Science and Technology. January, New Orleans.
    - Candade, N. and Dixon B. 2005. Can Logistic Regression and/or Feature Selection Methods Be Used to Predict Contaminated Wells? a Case Study of Polk County, Florida. Poster. Soil Science Society of America. Nov. Salt lake City.

Report Follows

Ground Water Vulnerability Delineation Using Integrated GIS and Neuro-Fuzzy Methods

Barnali Dixon, University of South Florida - St. Petersburg, Dept. of Environmental Science & Policy & Geography, 140 Seventh Ave South, 210 Davis Hall, St. Petersburg, FL 33701

Detection of potentially contaminated wells is an important component of environmental protection and management. However, contamination potential mapping is not an easy task due to inherent uncertainties. This study aims at assessing suitability of various techniques in predicting contaminated wells for example, Neuro-fuzzy (NF), Neural Networks (NN), Support Vector Machines (SVM), logistic regression and feature selection.

Contamination potential depends on complex interactions of hydro-geological variables. A large number of input variables add to redundancy, cost and time. The logistic regression, feature selection methods were used to identify critical variables in transporting contaminants in and through the soil profile. NF, NN and SVM were used to identify contaminated wells. Variables identified by logistic regression and used in this study included DRASTIC parameters, soil structure (pedality), hydrologic group, landuse, organic matter and bulk density. Well data (nitrate-N) provided by FLDEP as part of the WSRP were used in this study as target class.

The objective of this study was three- fold: (a) Analyze the input variables and identify the most significant predictors of well contamination. Perform feature selection to identify the best subset of variables. (b) Use all the input variables with the NF, NN and SVM to classify wells and compare their performances. (c) Repeat the above (step b) with the variable subset from step (a) and compare results.

Classifiers were compared based on their accuracies and parameters such as sensitivity and specificity. Free Receiver Operating Curves (FROCs) were used for evaluation of classifier performance.

Preliminary results show comparable results with the NN and SVM. Feature selection did not improve accuracy. However, it helped increase the sensitivity or the true positive rate (TPR). Thus, a higher TPR was obtainable with fewer features or variables. In this study, higher TPR is desirable since the cost of detecting a contaminated well incorrectly is far higher than a non-contaminated well going undetected. In addition, obtaining comparable results using less number of variables can reduce the cost of a project. Use of NF was not adequate for TPR. Compared to NN and SVM, NF is more sensitive to the number of wells used with the model. NN and SVM performed better with increased number of wells (larger training data sets). Integration of NF, NN and SVM models to GIS facilitated sensitivity analysis over space, however, integration of GIS with SVM was not as simple as integration with NF and NN.