Statistical Methods for Estimating Rate of Denitrification in a Shallow Ground Water Aquifer at Jacksonville

Raoul Fernandes (rif08@fsu.edu), Department of Earth, Ocean and Atmospheric Sciences, Florida State University

Ming Ye and Liying Wang, Department of Computational Science, Florida State University

Abstract

Nitrates (NO_3^-) are one of the principal contaminants in ground water. Excess nitrate in ground water is known to cause serious illnesses such as methemoglobinemia, and cancer. In addition to the adverse impact on the health of humans, excess nitrate is known to have unfavorable effects on the ecosystem. One of the major contributors to nitrates in the system are septic tanks. Approximately one-third of Florida's population uses Onsite Wastewater Treatment System (OWTS). In order to quantify the nitrate load to a water body several models have been developed, these models always ignore nitrate form normally working septic tanks and denitrification that occurs between the septic tank drain field and the water body. Additionally these models are often complex and developed specifically for a given site.

The aim of this project is to develop a simplified model that can estimate nitrate fate and transport form an OWTS to a targeted water body. The Simplified model is developed in two parts, the first to estimate the fate and transport of nitrate and the second the development of a denitrification rate (R_{dn}). This work focuses on the development of a model to estimate the rate of denitrification using easily available parameters.

To estimate the denitrification rate, data was first collated form existing literature values and other sources. The data collected included the main factors that controlled denitrification i.e. Texture, Temperature, Water Filled Porosity, Organic Carbon, pH, Bulk Density, Soil Depth, Nitrate Concentration and the denitrification rate. A total of 1129 distinct set of parameters and denitrification rates were collected and then statically analyzed to determine the relationships between the factors and the denitrification rate.

Three statistical methods were used to estimate Rdn, , linear regressions with Monte Carlo simulation, Multi Regression analysis and the development of a neural network to estimate. Eventually it was found that due to the complexity of the process it was the Neural Network that was able to best estimate the denitrification rate. The resulting network was then used with existing parameters to predict denitrification rates in the Jacksonville area. In addition isotope data was used to predict the percent of nitrate removed due to denitrification. This method serves as an alternative to estimate the loss of due to denitrification but is unable to estimate a rate of removal..

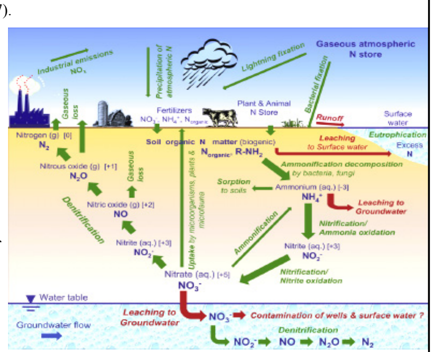
Introduction

Widespread pollution of ground and surface waters from Nitrate (NO_3^-) is of global concern to human health and the environment. The presence of NO_3^- in drinking water is hazardous to health. Serious illnesses such as methemoglobinemia are related to increased nitrate levels. In addition exposure to high levels of nitrate can cause diuresis and hemorrhaging of the spleen. (Department of Ecology). To ensure the safety of the public the EPA (2009) has established the following standards for Nitrate: Maximum Contaminant Level Goal (MCLG): 10 mg/l; Maximum Contaminant Levels (MCL): 10 mg/l. In addition to the adverse impact on the health of humans using the contaminated water, excess nitrate is known to have unfavorable effects on the ecosystem as well. Excess NO3- is known to cause eutrophication in many aquatic systems (Turner & Rabalais, 1996; Fenn et al., 2003).

It is widely accepted that once nitrate is leached into the soils there are four accepted pathways for its removal (Rivet et al., 2008; DeBernardi et al., 2007).

- Microbial biomass/ plant uptake
- Dilution
- Denitrification
- Dissimilatory nitrate reduction to ammonia, i.e. Ammonification (DNRA)

Denitrification is the reduction of NO_3^- to nitrogen gas (N_2) and it considered by most to be the only acceptable nitrate attenuation method that can completely remove nitrates from the system. Denitrification refers to the dissimilatory reduction, by essentially anaerobic bacteria, of one or both of the ionic nitrogen oxides (nitrate, (NO_3^-), and nitrite, (NO_2^-)) to the gaseous oxides (nitric oxide, (NO), and nitrous oxide, (N_2O)), which may then be further reduced to di-nitrogen (N_2).



$$2\,NO_3^- + 12\,H^+ + 10\,e^- \rightarrow N_2 + 6\,H_2O --------- (1)$$

Figure 1 : Overview of the Nitrogen Cycle

The following variables are the major factors controlling denitrification. (Rivett et al., 2008; Hiscock et al., 1991; Knowles, 1982).

- Nitrate Concentration (electron acceptor)
- Electron donor concentration (OC)
- Oxygen Concentration
- Nutrient and micro-nutrient activity
- pH
- Temperature
- Salinity
- Inhibitory substances
- Sediment pore size
- Microbial acclimation
- Hydraulic retention time

A number of different approaches have been used to develop denitrification as sub-models in N cycling models (Parton et al., 1996),

(1)Microbial growth models,

(2) Soil structural models, and

(3) Simplified process models.

Within the variety of simplified models reviewed, there seems to be general consensus about the mathematical formulation of the model.

$$D_a = \alpha \;\; f_N \;\; f_S \;\; f_T \;\; f_{pH} \quad --- (2)$$

The framework for the model as described by Heinen (2000) is ideal with a slight modification. Instead of having the denitrification rate as a function of potential denitrification rate and other additional reduction functions, it is perhaps easier to have the denitrification rate expressed directly in terms of these controlling factors.

$$R_{dn} = f(Tx,\ Temp,\ WFP,\ pH,\ OC,\ Bulk\ Density,\ Nitrate\ Conc,\ Thickness) \quad ------- (3)$$

Data Set

The data set is compromised of a total of 1129 sets of data that were obtained form various sources. The data is divided into two sets, set "A" which is composed of data gathered by Tucholke (2007) and set "B" obtained form Oehler (2010). The data is combined into one data set and the results from the combined data is represented here. The dataset composed by Tucholke (2007) was extracted for literature reviews. Missing pH data form data set A was filled in based on the mean value of the dataset, 20 pH values were filled in this represented 3.3% of dataset A. Since a direct correlation was available between the bulk density and the organic carbon this approach was used for estimating the organic carbon (OC) content of a soil based on the reported bulk density. In total, 21 missing OC values were filled in (3.5% of all values). Dataset B was essentially left unaltered, except for changes to the units that were needed to facilitate a combined dataset.

Methods

Three statistical methods were used to estimate Rdn,

- Hierarchical linear regression
- Multi Regression analysis and
- A set of neural networks to estimate Rdn

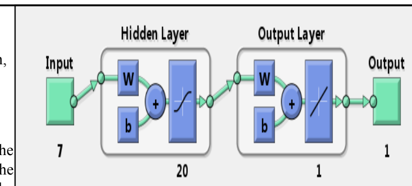


Figure 2 : Neural network developed.

To aid in the development of the ANN, the dataset was divided based on texture before the information was fed into the ANN. This led to the development of 13 distinct subsets which could be used to develop networks. The ANN was developed in MATLAB using the available toolbox and GUI. Each database was randomly divided into a training (70%), validation (15%) and test (15%) subsets. A two layer feed-forward network with sigmoid hidden neurons and linear output neurons was developed using 7 Input Nodes, 20 hidden nodes and one output node. This corresponded to the 7 input parameters that were compiled for each denitrification value.

The network was then trained with Levenberg-Marquardt back-propagation algorithm. The network was trained till the Mean Squared Error was as close to zero as possible. In addition the target Rdn Vs. output Rdn for three divided datasets and all data was plotted measure the correlation between outputs and targets. This allowed a direct comparison of the accuracy of the networks. The final network for each texture was selected based on a combination of three factors, the lowest MSE, the best fit between the output and target and a regression value as close to 1 as possible. Where possible the network with the least amount of input nodes and the least amount of hidden nodes was selected. In general all of the textures has seven input nodes.

Due to the paucity of data not all textures could be represented by a network, for example Textural class 4 (Loamy sand) and Texture 6 (Sandy clay loam) were unable to be developed. Fortunately the one Textural class of concern , Texture 5 (Sand) had a 103 sets of data, unfortunately due to some missing values the dataset was reduced to 96. this was divided into training (68), validation (14) and testing (14).

Several networks were created and evaluated; to keep track of the networks the convention used was to label the network as ANN 5-7-20, where the first number represented the texture, the second the number of input nodes and thus the number of input data required and the third number represented the number of hidden nodes. This allowed other networks to be used in situation where there was a deficiency of data. The use of a network with a lower number of inputs would expectedly result in a prediction with a lower accuracy.

Results and discussion

The network developed here is for Texture 5 (Sand). As can be seen form Figure 3, the network has been trained well. This is evident form the fact that the all four graphs on the figure show a good correlation between the target and the output. In addition the MSE was 0.6594 for the test data set, 2.49 for the validation dataset and 0.6999 for the training dataset. This indicates that the network could be used to predict the denitrification of a sandy soil under any conditions fairly accuracy.

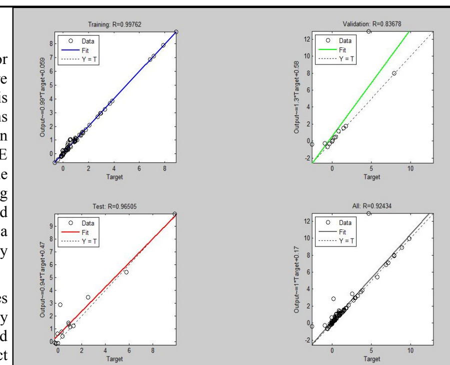


Figure 3 : Results of Training, Validating and Testing ANN 5-7-20

Data from the Natural Resources Conservation Service (NRCS), soil Survey Geographic (SSURGO) database and additional sources was then used to collect input information for the Study area in Jacksonville. The information was then input into the ANN 5-7-20.

The information used to calculate the denitrification rates was, Texture 5 (Sand), Temperature 20ºC, WFP (100%) as we are concerned with the saturated zone, pH 5.2 (average pH of soils in the region), bulk density 1.5 gm cm-3, saturated thickness of 1000 cm, Organic carbon percentage of 2.1% and an average concentration of 10.3 µg N gm-1soil. All of the values were derived based on averages.

Based on the above data it is estimated that the rate of denitrification in the Lake Shore region of the Jacksonville City in Duval county Florida is 6.6949 Kg N ha^{-1} d^{-1}.

Conclusion and Future Work

The application of ANN 5-7-20 to the Lake shore region seems to be valid. When compared with removal rates form similar aquifers it can be seen that this method provides a simple and accurate enough method and model to predict the rate of denitrification using existing available data. Brettar and Hofle (2002) have measures an denitrification rate of 8.5 Kg N ha^{-1} d^{-1}.

A comparison of the ANN denitrification rate could not be made with the estimated nitrate removed due to denitrification as a rate factor was unavailable. At a cursory glance however it does seem that the rates obtained by the two methods may be comparable.

While it seems that the ANN methodology is superior to other currently available simplified models caution must be exercised when using the predicted rates for application. The networks were mainly trained on laboratory data due to the lack of available field data. While the applicability of these results to field data seems to be valid, however it should me kept in mind that the results probably reflect a optimistic or a maximum possible denitrification rate.

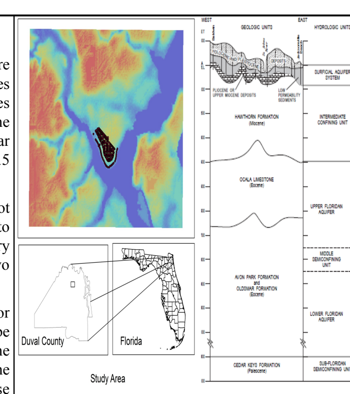


Figure 4 : Study area (Left) and Geology of the Aquifers (right)

In order to improve the accuracy of the ANN's , it is recommended that they be trained with a much wider database and if possible solely on a dataset that represents field measurements of denitrification rates

Selected References

Korom, S. F. (1992). *Natural denitrification in the saturated zone: A review. Water Resources Research, 28*(6), 1657.

Oehler, F., Rutherford, J., & Coco, G. (2010). *The use of machine learning algorithms to design a generalized simplified denitrification model. Biogeosciences, 7*(10), 3311-3332.

Rivett, M. O., Buss, S. R., Morgan, P., Smith, J. W. N., & Bemment, C. D. (2008). *Nitrate attenuation in groundwater: A review of biogeochemical controlling processes. Water Research, 42*(16), 4215.

Tucholke, M. B. (2007). *Statistical assessment of relationships between denitrification and easily measurable soil properties: A simple predictive tool for watershed-scale modeling.* Colorado School of Mines.