TWRI REPORT

Modeling Nitrate Transport in Aquifer System with Data Assimilation

Sean Tolle

Mills Scholarship 2008-2009

Introduction

Elevated nitrate concentrations in drinking water from an aquifer source can cause methemoglobinemia and could interact with nitrosatible drugs to cause birth defects (Johnson et al. 1987)(Brender et al., 2004). Because of the health concerns, the EPA set the maximum concentration level for nitrate at 10 mg/L NO_3-N for public drinking water (EPA, 1996). Private drinking water wells are unregulated and a study by Hudak (2000) showed nitrate concentrations were elevated in the west Texas aquifers. Nitrate in an aquifer comes mostly from surface sources (fertilizers, manure, etc.) and nitrate flux into an aquifer can be influenced by nitrogen application rates, land use, soil hydraulic properties, climatic conditions, and depth of aquifer (Gburek and Folmar, 1999)(Hudak, 2000). Nitrate in the aquifer is mobile with denitrification for the removal of nitrogen from the aquifer (Ruiz et al., 2002).

Nitrate concentrations in an aquifer are measured on a sparse temporally and spatial scale, because of the expense involve in drilling a sample well. Nitrate concentrations between measured values are determined with numerical models of groundwater flow, nitrate transport and transformation such as MODFLOW-2005 and MT3DMS (Harbaugh, 2005)(Zheng, 2006). A common method is to calibrated aquifer flow and transport parameters for the models with one set of data and validated the parameters with another set. If the calibrated model is used outside the calibrated time, errors in the nitrate concentrations are not affected by the numerical and parametric errors, but also errors in the estimated fluxes of water and nitrate. One method to decrease the modeling error is to use a data assimilation technique like Ensemble Kalman Filter to update the models state variables.

Data assimilation is a technique where modeled results and measured values are combined with a filter to produce updated results used as initial values for the next time step of the model (Drecourt et al., 2005). Kalman filter is one data assimilation technique developed by Kalman (1960) and was extended to nonlinear system by Evensen (1994) with the Ensemble Kalman filter. The Ensemble Kalman filter uses an ensemble of models with different parameters representing the uncertainty of the parameters. The models are simultaneously solved and the Ensemble Kalman filter combines the results with measured values to create updated results used in the next time step. Ensemble Kalman filter can cause the results to converge to single result and must be perturb the data with each update (Houtekamer and Mitchell, 1998). Whitaker and Hamill (2002) modified the Ensemble Kalman filter to avoid the need for perturbing the data and called the filter the Square Root Ensemble Kalman filter (SQEnKF).

The SQEnKF data assimilation process starts by subtracting the ensemble models simulated results (forecast) from the forecast mean in equation 1 (Whitaker and Hamill, 2002).

$$\boldsymbol{X}''^{f} = \boldsymbol{X}^{f} - \overline{\boldsymbol{X}}^{f} \qquad \textbf{(1)}$$

Where $\boldsymbol{X}'^{f}$ is the ensemble forecast difference, $\boldsymbol{X}^{f}$ is the ensemble forecast results, and $\overline{\boldsymbol{X}}^{f}$ is the forecast mean for the modeled results. After calculating the difference, the Kalman gain $\boldsymbol{K}$ and variance constant α are calculated with equation 2 and 3.

$$\boldsymbol{K} = \frac{\boldsymbol{PH}^T}{\boldsymbol{HPH}^T + \boldsymbol{R}^2} \tag{2}$$

$$\alpha = \left(1 + \sqrt{\frac{\boldsymbol{R}}{\boldsymbol{HPH}^T + \boldsymbol{R}}}\right)^{-1} \tag{3}$$

Where ***P*** is the forecast ensemble covariance matrix, ***R*** is the observation error matrix, and ***H*** is a matrix for selecting ensemble result with matching observations. The updated ensemble results used as initial values in the next simulations are calculated with equations 4, 5, and 6.

$$\overline{\boldsymbol{X}}^a = \overline{\boldsymbol{X}}^f + \boldsymbol{K}\left(y - \boldsymbol{H}\overline{\boldsymbol{X}}^f\right) \tag{4}$$

$$\boldsymbol{X}'^a = \boldsymbol{X}'^f + \alpha\boldsymbol{K}\left(-\boldsymbol{HX}'^f\right) \tag{5}$$

$$\boldsymbol{X}^a = \boldsymbol{X}'^a + \overline{\boldsymbol{X}}^a \tag{6}$$

Where *y* is the observation values, $\overline{\boldsymbol{X}}^a$ is the updated mean, $\boldsymbol{X}'^a$ is the ensemble updated difference, and $\boldsymbol{X}^a$ is the update ensemble values. My research used to SQEnKF data assimilation process as a method to calibrate and model nitrate transport. This method is compared to the more traditional method with simulated dataset and southern portion of the Ogallala Aquifer in Texas.

Methods and Material

My research into using data assimilation with nitrate concentration in groundwater was begun by using statistical analysis to develop a conceptual model of several aquifers in Texas. Data assimilation was then tested with a synthetic dataset based on the properties of an aquifer with a conceptual model with recharge fluxes calculated yearly. After testing with the synthetic dataset, the data assimilation technique was tested on the selected aquifer actual properties and sample data.

Development of Aquifer Conceptual Model from Statistical Analysis

Before the kalman filter was used with model nitrate transport, a statistical analysis of the Ogallala, Trinity, and Seymour Aquifers was performed to help develop a conceptual model of each aquifer. The statistical analysis examined the correlation of nitrate with climate data, soil hydraulic properties and well properties. The properties that have stronger correlation with nitrate probably influence the nitrate levels and need to be account for in the transport model. Three basic categories of conceptual models were considered for the aquifer system:

1. Deeper aquifer where travel time through vadose zone causes the surface properties to have little influence on nitrate levels. Conceptual model is groundwater model with water and nitrate recharge fluxes calculated over long term.
2. Median deep aquifer where soil properties influence nitrate level, but depth causes a delay and average of nitrate and water fluxes into the groundwater. Conceptual model is a couple vadose zone/aquifer models. Vadose zone can be model with a coarse resolution after the root zone. MODFLOW
3. Shallow aquifer where seasonal trends in the nitrate concentrations exist. Conceptual model is a closely linked groundwater and vadose model. The timing of nitrogen uptake and geochemistry uses is important. Smaller domains in the area of interest should be

modeled. If a river is near the area of interest, the model could become a linked river-groundwater riparian system.

The statistical analysis was used to determine Ogallala, Seymour, and Trinity Aquifers conceptual model. Trinity Aquifer was divided between the outcrop and downdip regions.

The statistical analysis included a cluster analysis, nonparametric correlation, and principle component analysis (PCA). Cluster analysis is a statistical test that separates the samples into clusters that have similar behavior for the properties. Nitrate concentrations (TWDB, 2008), climate data (PRISM Group, 2008), and land use (Vogelmann et al., 2000) was used for cluster analysis to divide the aquifers into smaller subdivisions. Soil (NRCS, 2008) and well properties were added to the correlation and PCA analysis. The correlation was used to determine the properties correlated with nitrate. Higher correlated properties were assumed to be needed in the conceptual model. PCA analysis reduced the properties beside nitrate to their principal components. The properties with higher association with the same principal component can be related to each other and needed in the conceptual model. Based on the statistical analysis, each cluster was classified into one of the general three conceptual models. The cluster with the first conceptual model was used to test the Square Root Ensemble Kalman filter data assimilation technique.

Data assimilation with synthetic dataset.

The selected aquifer cluster was used to develop a synthetic dataset. The synthetic dataset was used to test the effectiveness of data assimilations to calibrate and run an aquifer model. The synthetic dataset was initially modeled at a fine resolution to determine nitrate concentrations throughout the run. Random Samples of the aquifer properties and results were then used at a coarser resolution to the data assimilation in calibration of the aquifer properties. Finally, the aquifer properties were fixed to compare the results of data assimilation model to a normal model.

A spatial covariance model was fitted to the hydraulic conductivity, porosity, storativity, base elevation, and water recharge from TWDB water availability model. Nitrate dispersivity, first-order decay rate, and recharge fluxes range of values were estimated from literature. Three synthetic datasets of aquifer and nitrate transport properties were randomly generated with the appropriate spatial covariance model or parameter ranges. These three synthetic sets were modeled with MODFLOW 2005 and MT3DMS at a 400x400 resolution for eight years. The results were to act as the true values for the synthetic datasets and compared to data assembled results modeled at a higher resolution. Random samples of nitrate concentration, hydraulic head elevations were taken from the results and used as observation data from the data assimilation. Random samples of aquifer properties were taken to be used as initial properties while calibrating of the model.

The Square Root Ensemble Kalman filter (described in Introduction) was used as the data assimilation technique. The Kalman filter was combined with MODFLOW and MT3DMS to create data assimilation model. Calibration of the aquifer properties was accomplished by treating the aquifer properties as state variables that can be used in the data assimilation. Calibration was done with the first five years of sample nitrate concentrations and hydraulic heads. First, the water properties were calibrated by running the data assimilation with

MODFLOW. The model was considered calibrated when the mean of the ensemble for each property stabilized. If the means did not stabilized when the model completed the five-year simulations, then the results were inputted as initial values and model repeated. After the water properties stabilized, the nitrate transport properties were calibrate in a similar faction, but MODFLOW and MT3DMS were both used and water properties were fixed. After calibration, the final three years of sample data was modeled with all properties fixed. MODFLOW and MT3DMS were used with and without data assimilation techniques. The final two years results were compared with the true values from earlier.

Data assimilation with actual dataset

The actual aquifer properties and sample measurements from 1985-1990 were used in process to calibrate the aquifer properties. After calibrating the aquifer properties, the sample measurements from 1991-1995 randomly selected with a two-thirds used as observation values and the other third used for validation of results. As with the synthetic dataset, the aquifer was modeled with and without data assimilation results. The observation data was the two-thirds sample measurements. Results of both models were compared to the remaining third measurements.

Results

Cluster Analysis divided the Ogallala Aquifer into three clusters while the Seymour and Trinity Aquifers were divided into two clusters (figure 1). Ogallala Aquifer's clusters are divided between the southern, northwestern, and northeastern sections. Seymour aquifer is similar divided between the eastern and western sections with most of the isolated areas. Downdip first cluster occurs where another major or minor aquifer overlays the Trinity Aquifer. The second cluster has a confining layer above it, but not a major or minor aquifer. Trinity Outcrop region has both clusters divided into two sections with most of the clustering based on the nitrate concentrations with the first cluster having higher mean and median nitrate concentration.

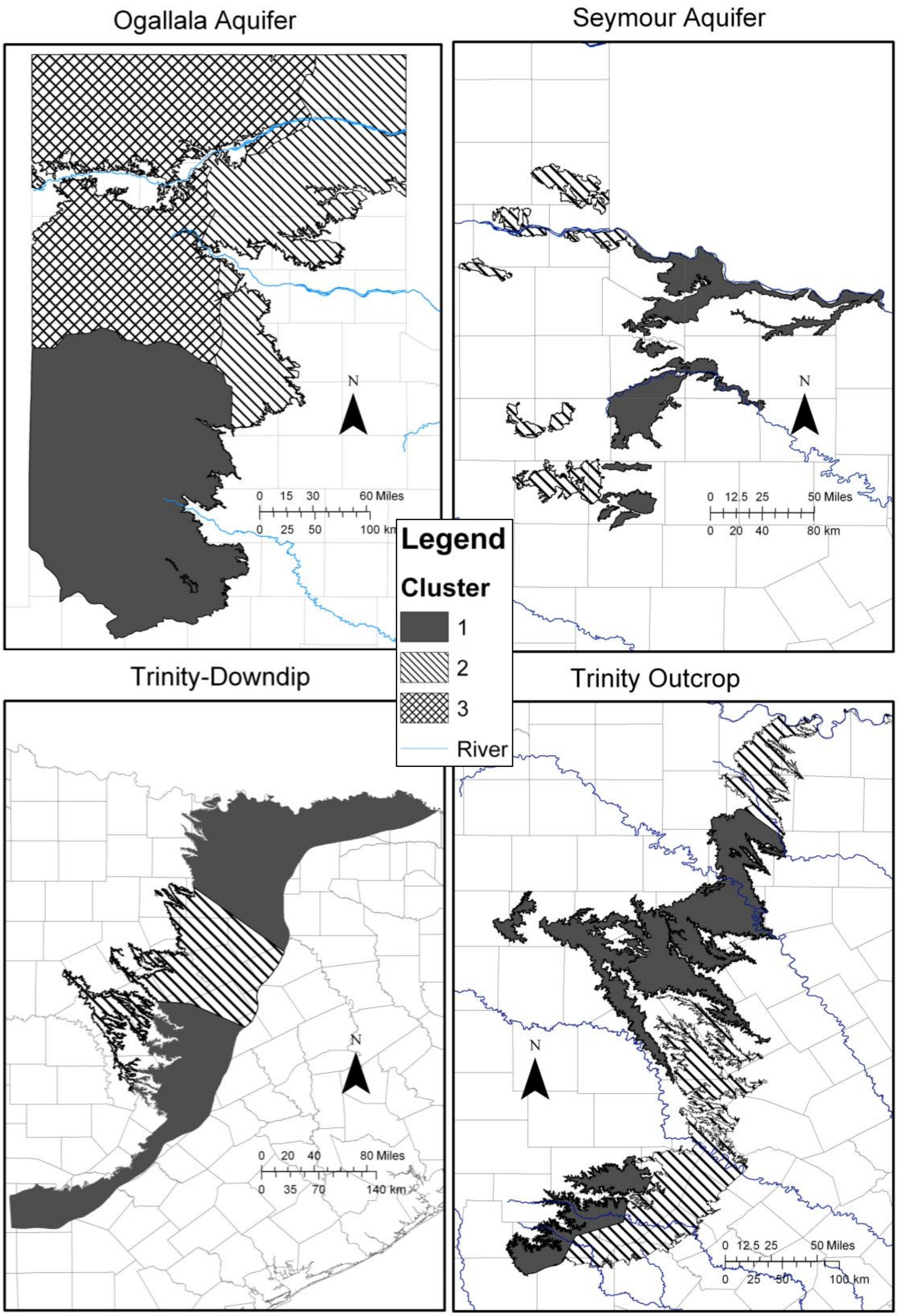

Figure 1. Maps of Ogallala, Seymour, and Trinity Aquifers with clusters of each aquifer.

For the Ogallala Aquifer, the southern cluster has the highest mean and median nitrate concentrations. Correlation and PCA analysis shows that the southern cluster nitrate concentrations are strongly related to the depth of the well and lesser extend to the climatic conditions. Statistical analysis showed that northeast cluster nitrate concentrations are related to soil properties, while the northwest cluster nitrate concentrations are related to soil and climatic properties. From the statistical analysis, the southern cluster can be model as the first conceptual model deep aquifer. The northern two clusters are the second conceptual model with vadose zone recharge fluxes included.

For the Seymour Aquifer, statistical analysis indicates that nitrate concentration is related on the soil properties and annual mean temperature for the first cluster, and soil properties, date and season sample was taken for the second cluster. Seymour first cluster can be modeled with the second conceptual model, but an alluvial/groundwater system should be used for the proportion of the aquifer near the one of the two major rivers. The alluvial system is probably most important for the long narrow strips of the Seymour Aquifer next to the river. The addition of the importance of the season sample was taken indicates that the third conceptual model should be used with timing of the nitrogen cycle important at the surface with an alluvial system when a river is near. A difficult of the Seymour Aquifer conceptual model analysis is the discontinuity of the aquifer. Most of the sections have low number of nitrate samples that prevents significant analysis for each individual section.

For the Trinity Aquifer Downdip, statistical analysis showed that nitrate concentrations for the first cluster are related to average temperature and watershed with the relationship depended on location (north or south spilt). Because of the aquifers above the first cluster, the first cluster is not directly hydraulically connected to the surface and can be modeled with a simple spatial regression model. Statistical analysis of the second cluster indicates depth of the well, year sample was taken, and climate are important factors for nitrate concentration. The second cluster nitrate concentration is increases with time, but the year of sample has a higher correlation with nitrate concentration than the data of sample with nitrate concentration. This indicates that nitrate fluxes can be calculated over the long term. The median nitrate concentration for the second cluster is 0.52 mg/L of NO_3-N. The low levels and confining layer indicates that the spatial regression model can be used for the second cluster.

The Trinity Aquifer Outcrop statistical analysis indicates that well depth, soil properties, and climate are important for nitrate concentrations for both clusters and the overall outcrop area. This indicates that entire outcrop area should be modeled with the second conceptual model with division of the outcrop based on know boundary conditions.

Full details of the statistically analysis will be published in a paper and included in my dissertation. Data assimilation results and analysis not completed at the time, because the simulations are still running and some minor alterations to the calibration techniques. Results and analysis will be published in peer review paper and dissertation. Copies of all the peer review papers and dissertation will be sent as completed.

References

Brender, J.D., J.M. Olive, M. Felkner, L. Suarez, W. Marckwardt, and K.A. Hendricks. 2004. Dietary Nitrites and Nitrates, Nitrosatable Drugs, and Neural Tube Defects. *Epidemiology*. 15(3):330-336.

Drecourt, J., H. Madsen, and D. Rosbjerg. 2006. Calibration framework for a Kalman filter applied to a groundwater model. *Advances in Water Resources*. 29:719-734.

EPA. 1996. Drinking Water Regulations and Health Advisors. EPA, Washington, DC.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Resources*. 99:10143-10162.

Gburek, W.J. and G.J. Folmar. 1999. Patterns of contaminant transport in a layered fractured aquifer. *Journal of Contaminant Hydrology*. 37:87-109.

Harbaugh, A.W. 2005. MODFLOW-2005, the U.S. Geological Survey modular ground-water model -- the Ground-Water Flow Process. U*.S. Geological Survey Techniques and Methods 6-A16.*

Hudak, P.F. 2000. Regional trends in nitrate content of Texas groundwater. *Journal of Hydrology.* 228:37-47.

Houtekamer, P.L. and H.L. Mitchell. 1998. Data Assimilation using an ensemble Kalman Filter Technique. *Monthly Weather Review*. 126: 796-811.

Johnson, C.J., P.A. Borod, T.I. Dosch, A.W. Kilness, K.A. Senger, D.C. Busch, M.R. Meyer. 1987. Fatal outcome of methemoglobinemia in infant. *Journal of the American Mdeical Association.* 257:2796-2797.

Kalman, R.E. 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*. 82: 35-45

PRISM Group. 2008. Oregon State University. http://www.prismclimate.org. Created 23 June 2008.

Ruiz, L.S., Abiven, C. Martin, P. Durand, V. Beaujouan, and J. Molénat. 2002. Effect on nitrate concentration in stream water of agricultural practices in small catchments in Brittany : II. Temporal variations and mixing processes. *Hydrology and Earth System Sciences*. 6:507-513.

Vogelmann, J.E., S.M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and J. N. Van Driel. 2001. Completion of the 1990's National Land Cover Data Set for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*: 67:650-662.

Whitaker, J.S. and T.M. Hamill. 2002. Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review*. 130: 1913-1924

TWDB. 2008. *Texas Groundwater Database*. <www.twdb.state.tx.us> Accessed on 5 Jan 2009. Updated on 28 Dec 2008.

Zheng, C. 2006. *MT3DMS v5.2 Supplemental User's Guide.* Technical Report to the U.S. Army Engineer Research and Development Center, Department of Geological Sciences, University of Alabama, 24 p