USGS

WATER RESOURCES RESEARCH GRANT PROPOSAL

Title: Multiobjective Optimization of a Public Supply Wellfield using an Artificial Neural Network and Non-Linear Programming

Focus Categories: GW, WS, M&P

Keywords: Conflict Management, Decision Models, Groundwater Management, Groundwater Modeling, Groundwater Movement, Groundwater Quality, Health Effects, Hydrogeology, Institutional Relationship, Multiple Objective Planning, Mathematical Models, Optimization, Pollution Control, Risk Analysis, Risk Management, Systems Analysis, Systems Engineering, Urban Water System, Water Demand, Water Quality Management, Water Levels, Well Hydraulics

Duration: January 2000 – December 2000

Federal funds requested: $8,000.00

Non-Federal (matching) funds pledged: $27,940.00

Principal Investigator(s) name(s) and university.

Dr. Donald Davis, Professor, Hydrology and Water Resources, Univ. of Arizona

Dr. Ferenc Szidarovszky, Professor, Systems and Industrial Engineering, Univ. of Arizona

Emery A. Coppola Jr., Doctoral Candidate, Department of Hydrology and Water Resources, University of Arizona

Congressional district of university where the research is to be conducted: Fifth

Statement of critical regional or State water problems

Wellhead protection is a federal program (Safe Drinking Water Act Amendments, June 1986) mandated to protect drinking water wells from contamination. Unfortunately, many municipal wellfields (e.g. South Tucson) are in close proximity to groundwater contamination sites.Pumping of these wells produces capture zones which often draw contamination into the water supply (EPA survey of 466 randomly selected wells found 28.0 percent of the large systems contained at least one volatile organic compound, Westrick et al., 1984). Ideally, decision-makers would like to identify pumping rates that satisfy water demands without compromising water quality. Since increasing pumping rates increases the size of the capture zone areas, the objectives of maximizing supply while minimizing public health risk often conflict. Hence, a formal method of quantifying the interaction between these two objectives is critical so that decision-makers can identify pumping polices that best serve the public interest.

The Artificial Neural Network (ANN) methodology [Skapura, 1996] developed in this research will be applied as an example to the Parkway Wellfield, which supplies Toms River, New Jersey with 30 percent of its drinking water. Different wells in the wellfield have exhibited differing degrees of contamination from a neighboring Superfund site. Both historical water quality data and numerical simulation have shown that pumping rates and variable climatic conditions, such as dry summers (lower groundwater recharge), affect the risk of contamination. Accordingly, the New Jersey Department of Environmental Protection would like identification of pumping strategies that appropriately balance water supply with public health objectives in real-time. This particular test case was selected because data, models, and cooperation are available. The same methodology can be applied without modification to any groundwater system that supplies drinking water, including the Tucson area.

Statement of results or benefits

The benefits will be the development and application of a methodology for effectively balancing groundwater supply with public health risk, a multiobjective goal traditionally fraught with difficulties. The common approach of conducting trial and error simulations with a numerical model does not ensure identification of even “good” solutions, and is limited by the number of trials attempted by the modeler. More sophisticated approaches that utilize nonlinear or dynamic programming require sophisticated algorithms that often demand significant computational capabilities. The ANN methodology significantly reduces the computational demands of conventional nonlinear groundwater optimization approaches. Although not the focus of this research, the ANN could also be trained using actual field data, precluding the need for a groundwater model, which could significantly increase the accuracy of optimization results, since the accuracy of groundwater models is limited by data availability and model conceptualization.

Specifically, the ANN methodology will quantify impacts that different pumping policies at the Parkway Wellfield have on groundwater flow direction (surrogate to potential health risk) and supply objectives. This multiobjective methodology will facilitate a risk analysis and enable decision-makers to effectively trade-off short-term benefits with long-term consequences in real-time. In addition, a long-term risk management analysis can be performed to identify effective strategies for managing worst case scenarios, before they occur. For example, it may be found that the projected well replacement costs (estimated at $1,000,000 per well) are compensated by significant long-term reductions in risk. In short, this methodology will provide a rigorous and objective assessment of how different wellfield policies balance water supply with public health concerns under both short and long-term scenarios.

Nature, scope, and objectives of the research

The nature of this research is to develop an ANN methodology that can be used to both accurately estimate the dynamics of a groundwater flow system at points of interest, and effectively balance supply with public health objectives for public supply wellfields.

The scope consists of methodology development using a hypothetical case, and then its application to a complex, real world scenario (Parkway Wellfield). Methodology development required simulating groundwater flow dynamics for the hypothetical case with MODFLOW, the USGS numerical groundwater flow code. Simulation output data was used to develop and train a backpropagation ANN capable of estimating to a high degree of accuracy groundwater dynamics. The ANN architecture was programmed into Fortran, tested for accuracy, and coupled with a non-linear optimization algorithm to determine the Pareto frontier (non-dominated solutions) for the multi-objective problem. The methodology will then be applied to the Parkway Wellfield in order to test its applicability to complex, real-world problems. In this test application, a sensitivity analysis that considers the effects of variable climatic conditions (i.e. groundwater recharge) on the solution sets will also be conducted.

The objective of this research is to develop the ANN methodology at field scale and demonstrate its effectiveness in managing a real world public supply wellfield. Both different time-scales and measures of risk will be considered, and a comparison in solutions obtained from non-linear and genetic algorithms [McKinney, et al] will be conducted.

Methods, procedures, and facilities:

In order to develop the methodology, a hypothetical, heterogeneous, unconfined aquifer with three pumping wells was modeled with MODFLOW. Three hydraulic control pairs, each consisting of an upgradient and immediately downgradient node location, were selected to monitor the hydraulic gradient along a 3,000-foot boundary in the middle of the model domain. In this hypothetical case, it was assumed that a reversal of the hydraulic gradient along this boundary due to heavy pumping posed risk to the supply wells.

In order to generate data for the ANN, a continuous time-sequence of variable pumping and recharge rates was introduced into the groundwater flow model at monthly time-steps. Forty-nine years of monthly groundwater recharge data for the Toms River basin was provided by the New Jersey Geological Survey. From this data, five yearly sequences of recharge data representing both the extreme and mean recharge conditions for the basin were culled out. This data was combined with 512 different pumping patterns for the wells; each well was simulated to pump from 125 gallons per minute (gpm) up to 1,000 gpm, with 125 gpm increments. Combining the 60 months of recharge data with 512 pumping patterns produced 30,720 unique combinations, each of which constituted a single unique monthly stress period in the model. During simulation, the groundwater heads at the end of each stress period at all node locations in the model were saved.

The resulting groundwater simulation data was split in two data sets; the first was used for training the networks and the second for validating the networks. A separate ANN was trained for each month. The purpose of the ANN is to obtain, for each month, a function that accurately estimates groundwater heads at the end of the month, given heads at the beginning of the month, as well as the monthly pumping and recharge rates. In this case, the ANN architecture utilized eighteen inputs, consisting of groundwater elevations at fourteen locations in the flow model at the beginning of the stress period, pumping rates of the three wells, and recharge over the stress period. The desired output was groundwater elevations at the fourteen locations at the end of the stress period. Six of the head locations correspond to the hydraulic control pairs used to assess risk by the value of the head differences. The accuracy of the 12 monthly ANN's were validated, and the root mean square errors (difference between the network estimated head values and MODFLOW values) were all less than one percent (0.78 percent on average for the 12 monthly networks)

Following validation of the individual networks, their functional forms were linked together (programmed in Fortran) so that the evolution of the head field over any planning horizon of interest could be simulated. The MODFLOW data was then sequentially processed through the linked ANN functions. That is, for a given year, the initial heads in January and the pumping and recharge rates for that month were processed through the January function to produce final head values at the fourteen locations. These estimated final head values were inputted into the February ANN function, along with the pumping rates and recharge over this month, to estimate heads at the end of the month. This was repeated for the remaining ten months. The average head values estimated by the ANN functions for all months at the fourteen locations were compared with the MODFLOW generated values. Of the 168 head values, 71% estimated by the ANN functions matched exactly with the MODFLOW values (recorded to the nearest tenth foot), and the remaining 29% differed by only 0.1 feet (head elevations ranged from about 80 to 120 feet mean sea level). Further, it was found that the functions could accurately estimate, on average, within 0.04 feet (absolute value of the error), the head difference between the hydraulic control pairs. Given this degree of accuracy, the ANN functions correctly predicted over 96 percent of the time whether a reversal in the hydraulic gradient between the upgradient and downgradient nodes had (risk) or had not (no risk) occurred.

A nonlinear optimization algorithm was programmed and linked with the ANN functions so that optimal pumping rates could be determined for different trade-offs. In this case, a one-year planning horizon was used, with the dual objectives to minimize supply deficit and minimize risk. Deficit was measured as the difference between some desired annual water supply and the total amount of water pumped out by the three wells over the 12-month period. Risk was measured as the total sum of the head differences between the downgradient and upgradient (under non-pumping conditions) nodes at the three hydraulic control pair locations. Positive values indicate some risk since "upgradient" nodes would overall have a lower groundwater elevation than the “downgradient” nodes, resulting in a general gradient reversal. Since the two objectives measure different physical quantities (gallons per minute versus feet), they were normalized so that they could be compared. The objective function to minimize, its individual components, and normalization forms are:

Minimize [α * Risk Normed + (1 - α) * Deficit Normed]

$$1 \geq \alpha \geq 0$$

$$\text{Risk} = \sum_{i=1}^{i=12} \{(h_6 - h_5) + (h_8 - h_7) + (h_{10} - h_9)\}$$

$$\text{Deficit} = 36{,}000 - \sum_{j=1}^{j=36} P_j$$

Risk Normed = [(Risk - MinRisk)/(MaxRisk - MinRisk)]

Deficit Normed = [(Deficit-MinDefict)/(MaxDeficit-MinDeficit)]

Variables h_6 through h_{10} are the head values at the three hydraulic control risk pairs (6 nodes), with the index i corresponding to months 1 through 12. There are 36 decision variables, designated P_j, corresponding to the monthly pumping rates for each of the three wells (e.g. P_4 is the monthly pumping rate of well 3 in February).

In the objective function to be minimized, a α value of 1 considers only risk, a value of 0 only deficit, and values between some tradeoff between risk and deficit. For example, a α value of 0.5 treats weights both objectives equally. By selecting different α values, the Pareto frontier or set of non-dominated solutions was identified.

The generated Pareto frontier conforms to intuition. Since well 3 is furthest from the boundary and has the least impact on risk, this well will pump at an earlier month than the other wells when risk is weighted relatively high. Correspondingly, well 2, located closest to the boundary, will be the last to pump. As supply is weighted higher, wells 1 and 2 will pump earlier in the year.

Now that the methodology has been developed using a hypothetical case, the following tasks must be completed.

a) Apply the methodology to the Toms River test case.
b) Consider longer time periods, and evaluate how one year might effect the next year.
c) Consider other measures of risk. For example, instead of considering the cumulative risk presented above, an alternative measure would be

worst case risk (largest gradient reversal in any single month) or the number of gradient reversal that occur over the planning horizon.
d) Incorporate risk constraints directly into the optimization formulation.
e) Perform a sensitivity analysis on climate variability (extreme recharge conditions).
f) Use a genetic algorithm for obtaining optimal pumping rates and compare results with those obtained using the nonlinear algorithm.

The Toms River study area has been modeled by the New Jersey Geological Survey using MODFLOW. The model domain has been discretized into 8 layers, each consisting of 63 columns and 81 rows. This model will be run with various combinations of monthly pumping and recharge rates. Different network architectures (i.e. number and location of selected head values) will be tried until an acceptable level of accuracy is achieved. Following this, the multiobjective optimization will be conducted, and the results verified by simulation in the groundwater model.

Because of the size of this model, it will be run on the University of Arizona's supercomputer. Much of the work will be conducted utilizing the Department of Hydrology and Water Resources computer facilities, as well as Emery Coppola's personal computer.

Related Research

Linear programming was first used to optimize water supply and remediation problems [Atwood and Gorelick, 1985]. This approach is limited by linear objective functions and constraints and its application to steady-state conditions (time-varying solutions are not found). In order to consider non-linear transient conditions, non-linear programming has been used [Gorelick, et al, 1984]. Because of the computational demands associated with linking the nonlinear program with the simulation model, others have applied control theory algorithms, such as differential dynamic programming [Jones et al, 1987; Culver and Shoemaker, 1993]. Although these methods are less computationally demanding than conventional nonlinear approaches, they require sophisticated algorithms to linearize system dynamics and compute partial derivative terms. As a way of avoiding the computational demands and difficulties of non-linear programming, genetic algorithms have been used [McKinney and Lin, 1994; Cieniawski, et al, 1995]. This method utilizes a random search procedure inspired by biological evolution where only the "fittest" solutions survive and propagate to successive generations. Traditionally, these algorithms have been linked directly to a numerical flow model. An exception, and the research closest to this proposal, involved linking a genetic algorithm with a ANN that was trained to identify whether different pumping scenarios would effectively remediate contamination [Rogers and Dowla, 1994]. In this case, the ANN was not trained to estimate the dynamics of the system, but simply the success or failure of different pumping strategies.

In the proposed research, the ANN will be trained to estimate the dynamics of the groundwater flow system in response to different pumping and recharge stresses. The proposed methodology vastly reduces computational demand of non-linear optimization by replacing the system of groundwater flow equations with the ANN generated functions. In addition, these functions provide insights into system dynamics, which helps in both initializing decision variable values for optimization, and verifying results.

Lastly, the most significant difference between this proposed methodology, and the ones described above, is that this ANN approach could preclude use of a numerical model and be directly applied to field conditions. Numerical models are simplifications of the real world, and their ability to accurately simulate the real system is limited by data availability and complexity of the real system. In principle, an ANN could be trained to estimate responses in the real world at points of interest, and then used for direct optimization of the system. This would be a significant contribution to water resources management. Because of model uncertainty, many municipalities (e.g. Tucson, EPA Handbook, "Ground Water and Wellhead Protection") are reluctant to delineate wellhead protection areas with models, and rely instead on less quantitative methods, such as a fixed circle radius or groundwater vulnerability mapping.

References

Atwood, D.F., and S.M. Gorelick, Hydraulic gradient control for groundwater contaminant removal, J. Hydrol., 76, 85-106, 1985.

Cieniawski, S.E., J.W. Eheart, and S. Ranjithan, Using genetic algorithms to solve a multiobjective groundwater monitoring problem, Water Resour Res, 31(2), 399-409, 1995.

Culver, T.B., and C.A. Shoemaker, Optimal control for groundwater remediation by differential dynamic programming with quasi-newton approximations, Water Resour Res, 29(4), 823-831, 1993.

Gorelick, S.M., C.I. Voss, P.E. Gill, W. Murray, M.A. Saunders, and M. H. Wright, Aquifer Reclamation Design: The use of contaminant transport simulation combined with nonlinear programming, Water Resour. Res, 20(4), 415-427, 1984.

Jones, L.C., R. Willis, and W.W. Yeh, Optimal control of nonlinear groundwater hydraulics using differential dynamic programming, Water Resour Res, 23(11), 2097-2106, 1987.

McKinney, D.C., and M.-D. Lin, Genetic algorithm solution of groundwater management models, Water Resources Res., 30(6), 1897-1906, 1994.

Rogers, L.L., and F.U. Dowla, Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling, Water Resour Res, 30(2), 457-481, 1994.

Skapura, D. M., Building Neural Networks, ACM Press Books, 1996.