A Methodology to Reconstruct Groundwater Contamination History with Limited Field Data[1]

Jiabao Guan[2], Morris L. Maslia[3], and Mustafa M. Aral[4]

Abstract

A health risk assessment of exposure to contaminants in drinking water frequently requires reconstructing the groundwater contamination history in aquifers. The development of groundwater flow and contaminant fate and transport models used in such an analysis is time-consuming and expensive. Therefore, new approaches need to be explored to reconstruct groundwater contamination history that may reduce the time and expense of the traditional numerical modeling approaches.

An ongoing health study at U.S. Marine Corps Base Camp Lejeune, North Carolina, will determine if an association exists between exposure to contaminated drinking water and birth defects and childhood cancers in children born to women who were pregnant while living at the base during the period 1968–1985. For this study, there are no field data on groundwater contamination at water-supply wells prior to 1985. Limited contaminant concentration data at several observation wells are available during and after 1992 when remedial investigations of the contaminated aquifers began. To assist the epidemiological study in determining historical exposure scenarios, groundwater contaminant history needs to be reconstructed at base housing areas. Interpreting the mathematical models that are used in the analysis of groundwater flow and contaminant fate and transport and using the field data available on contaminant levels, a linear control model is proposed for the purpose of reconstructing the historical exposure to contaminated groundwater. The control model includes two matrices: (i) the first matrix describes the system behavior of the contaminant movement in aquifers in a natural environment and, (ii) the second matrix reflects the effect of pumping operations on contaminant concentrations. In the control model, the system matrix is recognized by using the least-squares method and the effect of pumping operations is determined by using an optimization method and an improved genetic algorithm when solving for matrix coefficients. To test the proposed methodology and to demonstrate the effectiveness of the model and proposed algorithm, the Tarawa Terrace base housing area of Camp Lejeune, which has been modeled in detail, is chosen. Computational results show that the model proposed herein can be used as a screening method to recover

---

[1] The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the Agency for Toxic Substances and Disease Registry or the U.S. Department of Health and Human Services.

[2] Senior Research Engineer, Multimedia Environmental Simulations Laboratory, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, jg79@ce.gatech.edu.

[3] Research Environmental Engineer, Agency for Toxic Substances and Diseases Registry, Centers for Disease Control and Prevention, Atlanta, Georgia 30341, mmaslia@cdc.gov.

[4] Director, Multimedia Environmental Simulations Laboratory, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, mustafa.aral@ce.gatech.edu.

groundwater contamination history with reasonable accuracy considering the lack of data that may exist at many sites.

Key words: control model, least square method, genetic algorithm, groundwater contamination.

Introduction

An ongoing health study at U.S. Marine Corps Base Camp Lejeune, North Carolina (Figure 1), will determine if an association exists between exposure to contaminated drinking water and birth defects and childhood cancers in children born to women who were pregnant while living at the base during the period 1968–1985. In this situation there are no field data on groundwater contamination at water-supply wells prior to 1994. Limited contaminant concentration data at several observation wells are available after during and after 1992 when remedial investigations of the contaminated aquifers began. To assist the epidemiological study in determining historical exposure scenarios, groundwater contaminant history needs to be reconstructed at base housing areas. The traditional approach used to solve this problem is to develop groundwater flow and contaminant fate and transport models and then to apply them to a study area to obtain the contaminant concentration distribution in time and space. However, determining and calibrating model parameters, such as aquifer hydraulic conductivity distributions and boundary conditions, is time-consuming and expensive. Therefore, new approaches need to be explored to reconstruct groundwater contamination history. These new approaches should reduce the time and expense of the traditional numerical modeling approaches. In this paper, a linear control model is proposed to reconstruct the contamination history at the site. The control model consists of two linear parts: the first part describes the system behavior of contaminant movement in aquifers in a natural environment and the second part reflects the effect of pumping operations on contaminant concentrations. Both are expressed as matrices that are separately identified using optimization methods based on the available concentration records and pumping-schedule data. The Tarawa Terrace base housing area of Camp Lejeune which has been modeled in detail (*Maslia et al.* 2007), is used to demonstrate the effectiveness of the model and proposed algorithm described herein.

Control Model

Groundwater resources within the Hadnot Point base housing area at Camp Lejeune were contaminated by leaking landfills for more than 40 years. Contaminated water-supply wells ceased operations during 1994, but contaminant sources were not removed or remediated at the time. Pumping data for water-supply wells prior to 1994 are available, but there are no water-quality records. Some concentration data are available after 1994. For generality, it is assumed that the study period spans from the starting time t_0 to the terminal time T_f and t_a is the time at which the water-supply wells stopped operating. In the study domain there are n locations (water-supply and observation wells) where concentration distributions during the period $[t_a, T_f]$ are recorded, and m water-supply wells whose pumping rates are known within the period $[t_0, t_a]$. For consistency with control systems, the concentrations at observation wells are represented by the vector $X(t)$, denoted as $X(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$, and the pumping rates at water-supply wells by $U(t)$, denoted as $U(t) = [u_1(t), u_2(t), \ldots, u_m(t)]^T$, where $x_i(t)$ is the concentration at location i in time t, $u_j(t)$ is the pumping rate at water-supply well j in time t. The concentration states at observation wells can be determined by the system behavior of the contaminant movement in aquifers in a natural environment and the effect of pumping operations. Therefore, the concentration states in the aquifer system can be approximately described by a discrete linear system state equation as

$$\begin{cases} X(k+1) = \mathbf{A}X(k) + \mathbf{B}U(k) \\ X(t_0) = X_0 \end{cases} \tag{1}$$

where $\boldsymbol{A}$ is the $n \times n$ matrix associated with aquifer parameters including contaminant source characteristics, $\boldsymbol{B}$ is the $n \times m$ matrix associated with pumping rates, X_0 is the initial concentration vector at observation wells —it is zero in general—and k is the time step or time stress period. In the model, the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are unknown and require being identified using existing recorded data such as time series data of concentrations at observation wells and pumping rates at pumping wells. Once both matrices are identified, Equation (1) can be used to reconstruct the contaminant distribution in time and space.

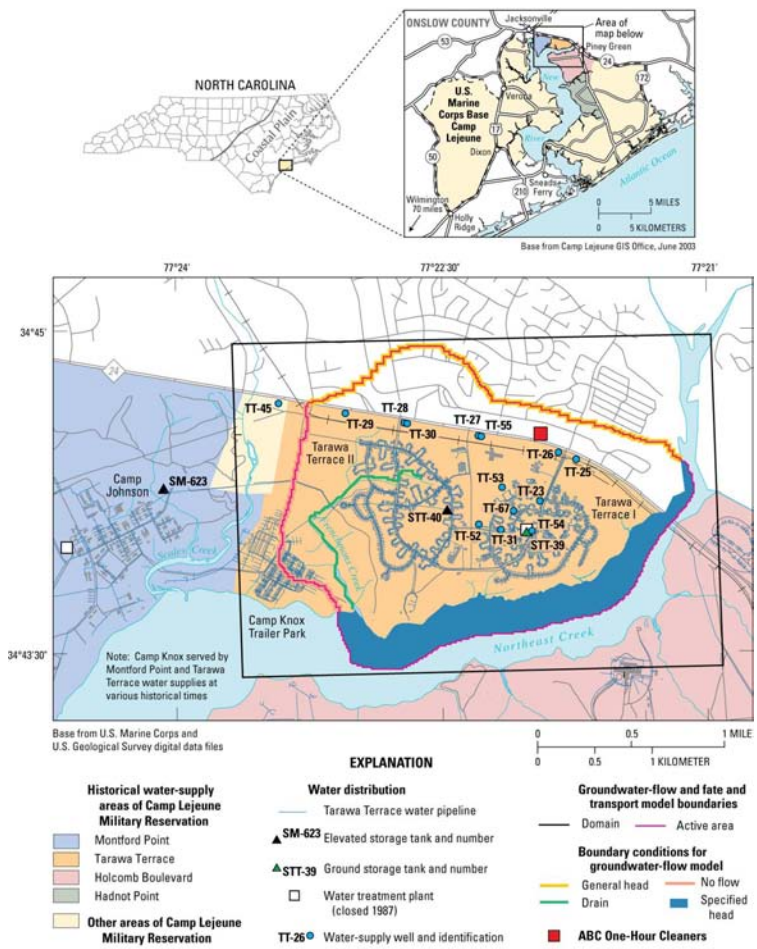

Figure 1. Water-supply areas, groundwater-flow modeling area, and water-supply facilities used for historical reconstruction analyses at Camp Lejeune.

Solution Algorithms

In order to reconstruct contamination history, the system equation described above needs to be recognized, i.e., matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ need to be identified. Based on available data, both matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can be obtained using a two-step procedure.

Step 1: Since the pumping wells cease operation after t_a, the system states at observation sites depends on just the concentrations at previous time stress period. $\boldsymbol{A}$ can be identified by using the least-squares method based on the concentration data within $[t_a, T_f]$ (*Björck*, 1996).

Step 2: The concentration data are not available during $[t_0, t_a]$ and need to be reconstructed, the least square method cannot be used to estimate matrix $\boldsymbol{B}$. However, it is obvious that concentrations at t_a are both initial concentration distributions during $[t_a, T_f]$ and also a consequence of pumping operations during $[t_0, t_a]$. In some sites there are some additional internal concentration data points during $[t_0, t_a]$. Taking advantage of these concentration data together with the pumping rate data, the estimation of matrix $\boldsymbol{B}$ can be formulated as an optimization problem that finds the matrix $\boldsymbol{B}$ by minimizing the total error between the estimated and observed concentrations at t_a as well as internal data points. The optimization model is then solved by using a genetic algorithm (*Holland*, 1975, *Guan et al*., 2007).

Numerical Applications

In this paper, the Tarawa Terrace base housing area at Camp Lejeune is used as an example to demonstrate the effectiveness of the model and proposed algorithm. The Tarawa Terrace housing area at the Camp Lejeune was constructed during 1951. Drinking water, supplied to the family housing, was contaminated with volatile organic compounds (VOCs), mostly with tetrachloroethylene (PCE) and its degradation by-products. The contamination was linked to the release of PCE from the ABC One-Hour Cleaners (Figure 1) into the groundwater system (*Faye and Green*, 2007; *Jang and Aral*, 2007; *Maslia et al*., 2007). To assist the epidemiological study (*Jang and Aral*, 2007; *Maslia et al*., 2007), detailed groundwater-flow and contaminant fate and transport models in the area were developed by the Agency for Toxic Substances and Disease Registry (ATSDR) and the Georgia Institute of Technology, Multimedia Environmental Simulations Laboratory (*Maslia et al*., 2007; *Jang and Aral*, 2007). The groundwater system at Tarawa Terrace and vicinity was modeled as 12 aquifer layers (*Jang and Aral*, 2007). Fourteen water-supply wells supplied drinking water to the base housing area from the underlying aquifer system (Figure 1). Water-supply wells, TT-26, TT-53 and TT-67 significantly impacted contaminant spreading toward the base housing area. For verification purpose, the calibrated simulation models developed for the ATSDR study are used to calculate contaminant concentration distributions within the study domain. Aquifer parameters, boundary conditions, source characteristics used in the current study (this paper) are exactly same as those used in *Jang and Aral* (2007). The simulation period is from January 1951 through December 1994 and consists of 528 monthly stress periods. To demonstrate the effectiveness of the methodology proposed herein, three Tarawa Terrace water-supply wells—TT-26, TT-53, and TT-67—are selected to operate during 408 monthly stress periods (from January 1951 to December 1984) using pumping schedules described in *Jang and Aral* (2007). Also, it is assumed that pumping operations cease after stress period 408. Five nodes around the contaminant plume are used to test the model, which are nodes 102570, 103438, 103860 (TT-26), 104066 and 104892 in the 7th layer and are closed to the contaminant source ABC Cleaners.

The resulting PCE concentration distributions at the aforementioned nodes are shown in Figure 2. The concentrations from stress period 408 to 528 (December 1984–December 1987) are used to identify system matrix ***A*** using the least-squares method. For identification of the coefficients of the matrix ***B***, two cases are presented. In case 1, only concentration data at stress period 408 (December 1984) are used and in case 2, 8 additional internal concentration data points from stress periods 1–408 are available and this concentration data along with the data at stress period 408 are used together. The PCE concentration distributions at five observation sites for cases 1 and 2, reconstructed using the identified system equation (1), are shown in Figures 3 and 4, respectively. In these figures, a solid line represents the reconstructed concentration distribution and a dashed line represents the measured concentration.

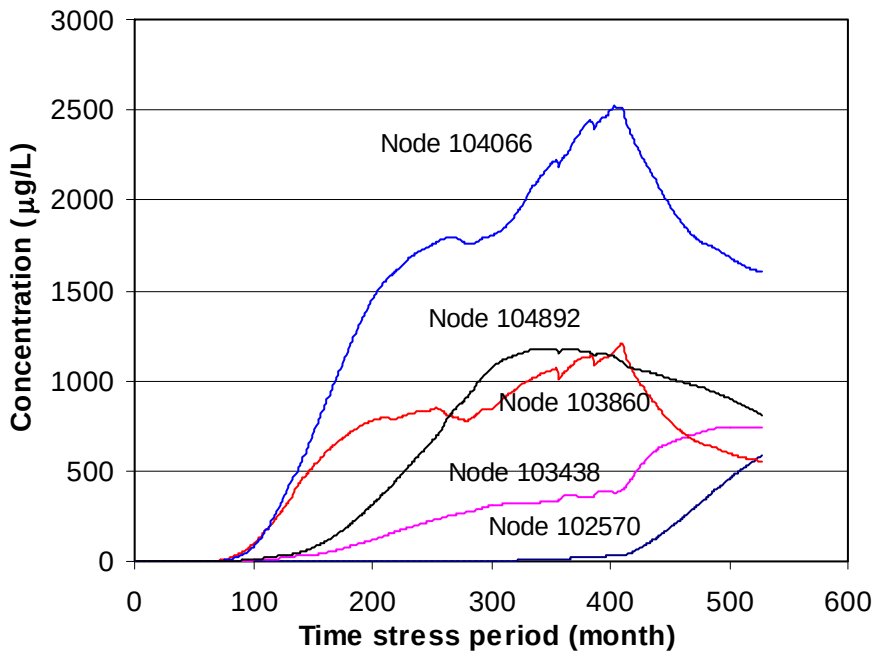

Figure 2. Resulting PCE concentration distributions in five nodes.

Results shown in Figure 3 (case 1, without additional internal data points), clearly show that the concentrations at stress period 408 are nearly identical to "measured" values. Overall, concentration distributions at observation nodes show good recovery, although some errors occur the estimation. Comparing each individual observation site, the concentrations at Node 104892 are best recovered. The maximum concentrations at Nodes 103860 and 104066 contain relatively large error compared to "measured" data. Considering the limited data availability, these results are encouraging and could be used with confidence for the epidemiological study.

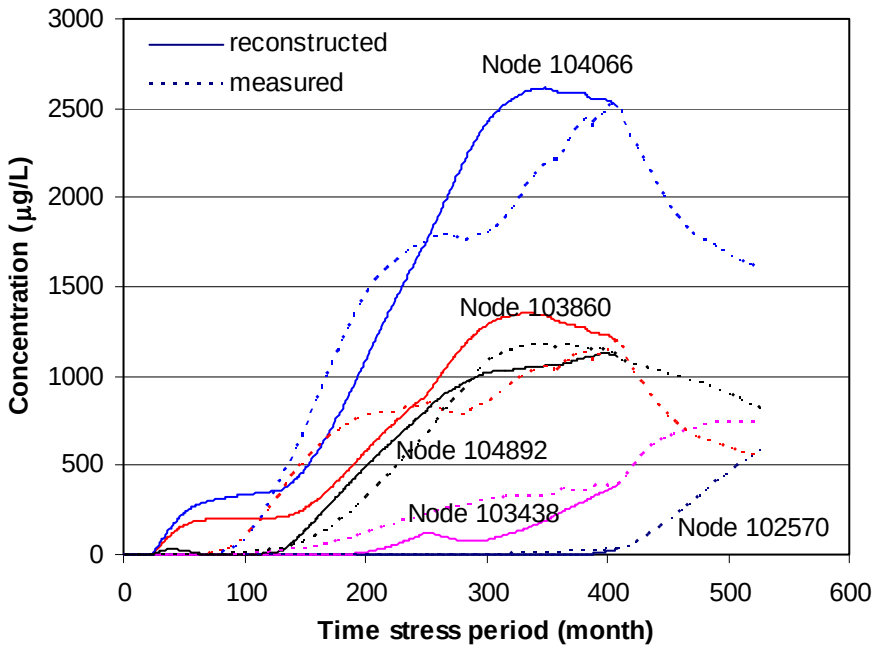

Figure 3. Concentration distributions recovered in case 1 (Without internal data points).

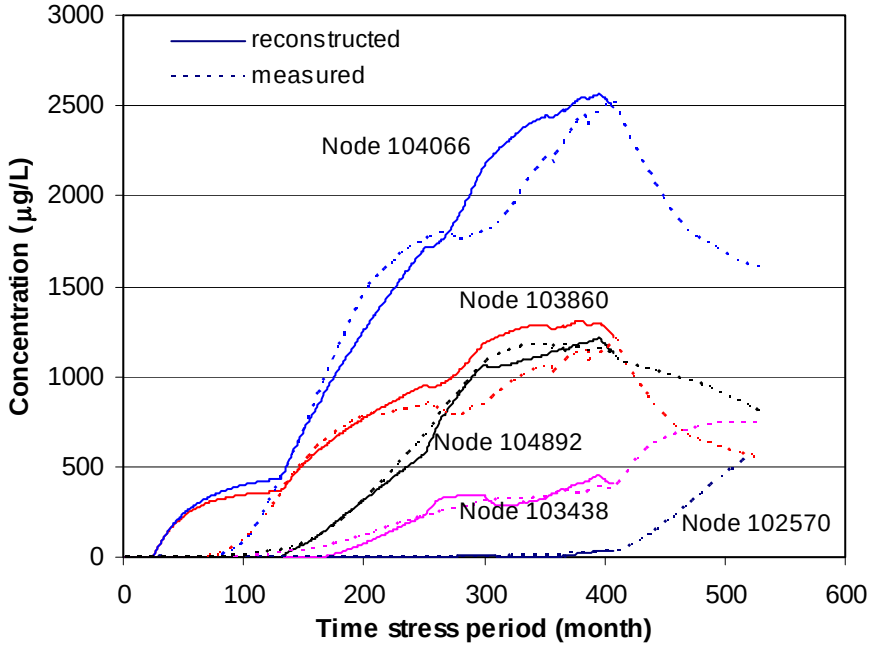

Figure 4. Concentration distributions recovered in Case 2 (With additional 8 internal data points).

If additional internal contaminant concentration data within stress periods 1–408 are available, the estimation accuracy may provide significant improvement. This is demonstrated in case 2 with the addition of 8 "measured" concentration data points within stress periods 1–408. For this scenario (case 2), calculation of the objective function value in optimization includes the additional internal concentrations in estimating coefficients of matrix ***B***. Comparing results shown in Figure 4 to those shown in Figure 3 indicates that concentrations distributions at all observation sites in Figure 4 (Case 2, 8 additional data points) are more closely reconstructed.

Conclusions

In this study a linear discrete system state equation is proposed to approximately describe the aquifer system in order to reconstruct the contamination history in the study domain. Compared to costly and time consuming groundwater-flow and contaminant fate and transport model development and calibration, the methodology proposed herein solely requires some measured concentration data at observation wells and water-supply well operation schedules to identify system and pumping effect matrices. The system and pumping effect matrices are determined by the least-squares method and a genetic algorithm, respectively. In this paper, the Tarawa Terrace base housing Camp Lejeune is used as an example to demonstrate the effectiveness of the proposed model and algorithm. The concentration distributions at the observation wells reconstructed from the control model provide a reasonable estimate of accuracy. This probably implies that the proposed model and algorithms provide an effective screening method for reconstructing groundwater contamination history when there are limited or missing historical contaminant concentration data.

References

Faye, R.E. and Green, J.W. Jr., 2007. Analyses of groundwater flow, contaminant fate and transport, and distribution of drinking water at Tarawa Terrace and vicinity, U.S. Marine Corps Base Camp Lejeune, North Carolina: Historical reconstruction and present-day conditions—Chapter E: Occurrence of contaminants in Groundwater. Agency of Toxic Substances and Disease Registry, Atlanta, GA. [cited 2009 January 23]; Available from http://www.atsdr.cdc.gov/sites/lejeune/index.html.

Guan, Jiabao, Kentel, Elcin, and Aral, Mustafa M. (2007). Genetic Algorithm for Constrained Optimization Models and Its Application in Groundwater Resources Management, Journal of Water Resources Planning and Management, Vol. 134(1), pp 64-72.

Björck, Àke, Numerical methods for least squares problems, Philadephia: SIA, c1996, 408p.

Holland, J.H. (1975). Adaptation in natural and artificial systems, University of Michigan Press, Ann Arbor, Mich., 183.

Jang, W. and Aral, M.M., 2007. Simulation of three-dimensional multi-species, multi-phase mass transport of tetrachloroethylene (PCE) and associated degradation by-products. Atlanta, GA: Multimedia Environmental Simulation Laboratory (MESL), School of Civil and Environmental Engineering, Georgia Institute of Technology, Report No.: MESL-02-07.

Maslia, M.L., Sautner, J.B., Faye, R.E., Suárez-Soto, R.J., Aral, M.M., Grayman, W.M., Jang, W., Wang, J., Bove, F.J., Ruckart, P.Z., Valenzuela, C., Green, J.W. Jr., and Krueger, A.L., 2007. Analyses of groundwater flow, contaminant fate and transport, and distribution of drinking water at Tarawa Terrace and vicinity, U.S. Marine Corps Base Camp Lejeune, North Carolina: Historical reconstruction and present-day conditions—Chapter A: Summary of Findings. Agency of Toxic Substances and Disease Registry, Atlanta, GA. [cited 2009 January 23]; Available from http://www.atsdr.cdc.gov/sites/lejeune/index.html.