APPLICATION OF A PARALLEL 3-DIMENSIONAL HYDROGEOCHEMISTRY HPF CODE TO A PROPOSED WASTE DISPOSAL SITE AT THE OAK RIDGE NATIONAL LABORATORY

Jin-Ping Gwo
Center for Computational Sciences
Oak Ridge National Laboratory[1]
P.O. Box 2008, MS6203
Oak Ridge, TN 37831-6203
email: g4p@ornl.gov

Gour-Tsyh Yeh
Department of Civil and Environmental Engineering
Pennsylvania State University
University Park, PA 16802
email: gty@arlvax.arl.psu.edu

KEY WORDS

Hydrogeochemistry, High performance FORTRAN, Bi-conjugate gradient stabilized matrix solver, Chemical equilibrium, Finite element method.

ABSTRACT

The objectives of this study are (1) to parallelize a 3-dimensional hydrogeochemistry code and (2) to apply the parallel code to a proposed waste disposal site at the Oak Ridge National Laboratory (ORNL). The 2-dimensional hydrogeochemistry code HYDROGEOCHEM, developed at the Pennsylvania State University for coupled subsurface solute transport and chemical equilibrium processes, was first modified to accommodate 3-dimensional problem domains. A bi-conjugate gradient stabilized linear matrix solver was then incorporated to solve the matrix equation. We chose to parallelize the 3-dimensional code on the Intel Paragons at ORNL by using an HPF (high performance FORTRAN) compiler developed at PGI. The data- and task-parallel algorithms available in the HPF compiler proved to be highly efficient for the geochemistry calculation. This calculation can be easily implemented in HPF formats and is perfectly parallel because the chemical speciation on one finite-element node is virtually independent of those on the others. The parallel code was applied to a subwatershed of the Melton Branch at ORNL. Chemical heterogeneity, in addition to physical heterogeneities of the geological formations, has been identified as one of the major factors that affect the fate and transport of contaminants at ORNL. This study demonstrated an application of the 3-dimensional hydrogeochemistry code on the Melton Branch site. A uranium tailing problem that involved in aqueous complexation and precipitation-dissolution was tested. Performance statistics was collected on the Intel Paragons at ORNL. Implications of these results on the further optimization of the code were discussed.

INTRODUCTION

Parallelism in computational hydrogeochemistry provides a framework for the vertical integration and modeling of biological, physical and chemical processes from microscopic to macroscopic scales. Realistic representation of large-scale multiple-objective groundwater resources management problems becomes possible and the analysis of these problems becomes achievable within economical time frames. Chemical equilibrium calculations of a hydrogeological system are intrinsically and perfectly parallel. This property, married with high performance parallel computers, eases the high memory and CPU time requirements that may otherwise overwhelm most computer platforms available today. Speed-up obtained in computation largely reduces the cost of design and planning, and scale-up in the size of problems and the underlying physical and chemical processes provides scientists a vehicle to see not only the trees but also the forest, examining the contribution of individual component processes to the behavior of a hydrogeological system. Cost of these benefits, however, may involve communication overhead among processors, high labor cost, and the lack of easy-to-

[1] managed by Lockheed Martin Energy Research, Corp. for the U.S. Department of Energy under contract DE-AC05-96OR22464

use parallelization software and parallel interpretation tools that may facilitate the design and analysis of large-scale multiple-objective problems.

This research emphatically addresses the problem of effectively solving large systems of simultaneous partial differential equations on high performance parallel supercomputers. Specifically, we focused on the parallelization of a intermediate size hydrogeochemistry code, HYDROGEOCHEM (Yeh and Tripathi, 1990), and the scale-up of the code. A state-of-the-art HPF (High Performance FORTRAN) compiler by PGI was used to parallelize the code. The insight gained from an application to a proposed waste site at ORNL and the efficiency of a parallel bi-conjugate gradient stabilized linear matrix solver was discussed.

This research was conducted on the Intel Paragons XPS5, XPS35, and XPS150, three distributed-memory massively parallel supercomputers housed by the Center for Computational Sciences at ORNL. Flow calculations were conducted on the XPS5 using PFEM (D'Azevedo et al., 1993), a parallel version of 3DFEMWATER (Yeh, 1987). The source code of PHGC3D, the parallel 3-dimensional version of HYDROGEOCHEM, and PFEM are available upon request. Details about the most recent scale-up and performance information of PHGC3D can be found at http://www.ccs.ornl.gov/staff/gwo/gwo.html. For more information about PFEM, one is referred to D'Azevedo et al., 1993.

THEORETICAL BACKGROUND OF SUBSURFACE HYDROGEOCHEMISTRY

A hydrogeochemistry system in a porous medium can be represented by three sets of equations, namely, the fluid flow mass balance equation, the solute transport equations, and the chemical equilibrium equations. To use PHGC3D, one needs to supply the velocity and pressure head fields calculated according to the following equations (Yeh, 1987):

$$\frac{\partial \theta}{\partial t} = \nabla \cdot \left(\mathbf{K} \cdot \nabla (h+z) \right) + Q, \tag{1}$$

$$\mathbf{v} = -\mathbf{K} \nabla (h+z), \tag{2}$$

where t, θ, h, z, $\mathbf{v}$, Q, and $\mathbf{K}$ are time (T), water content, pressure head (L), elevation (L), fluid velocity vector (L/T), external fluid sources ($L^3/L^3/T$), respectively, and the hydraulic conductivity tensor (L/T). The solute transport equation of an aqueous component j can be represented by the following equations (Yeh and Tripathi, 1990):

$$\begin{aligned} &\frac{\partial T_j}{\partial t} + \frac{\partial \theta}{\partial t}\left(S_j + P_j \right) = -\mathbf{v} \cdot \nabla C_j + \nabla \cdot \left(\theta \mathbf{D} \cdot C_j \right) \\ &-\theta \Lambda_j^a + M_j^a - Q C_j, \end{aligned} \tag{3}$$

where T, S, P, C are the total analytical concentration (M/L^3), the concentration (M/L^3) in the absorbed, the precipitated, and the aqueous phases, respectively; Λ and M are decay and source/sink rates ($M/L^3/T$); and $\mathbf{D}$ is the hydrodynamic dispersion tensor (L^2/T). The mass balance equation of an absorbent component j is (Yeh and Tripapthi, 1990):

$$\theta \frac{\partial W_j}{\partial t} + \frac{\partial \theta}{\partial t} W_j = -\theta \Lambda_j^s + M_j^s, \tag{4}$$

where W is the total absorbent analytical concentration (M/L^3). The balance equation of ion-exchange site is similarly represented by (Yeh and Tripathi, 1990):

$$\theta \frac{\partial N_{eq}}{\partial t} + \frac{\partial \theta}{\partial t} N_{eq} = -\theta \Lambda_{eq} + M_{eq}, \tag{5}$$

where N is the number of equivalents of the ion-exchange site per volume of solution (M/L^3). In eqs.(3)-(4), the superscript a, s, and the subscript eq represent aqueous phase, solid phase, and ion-exchange sites, respectively. The law of mass action for aqueous complex species i is (Yeh and Tripathi, 1990):

$$A_i^x = K_i^x \prod_{k=1}^{N_a} X_k^{a_{ik}^x}, \tag{6}$$

where A and X are the activities (M/L) of the complex and aqueous component species, respectively; K is the equilibrium constant, N_a is the number of aqueous components, and a is the stoichiometric coefficient. The law of mass action for absorbed species i can be represented by (Yeh and Tripathi, 1990):

$$B_i^y = K_i^y \left(\prod_{k=1}^{N_a} X_k^{a_{ik}^y} \right) \left(\prod_{k=1}^{N_s} Y_k^{b_{ik}^y} \right), \tag{7}$$

where B and Y are the activities (M/L) of the absorbed and absorbent component species, respectively, N_s is the number of absorbent species, and b is the stoichiometric coefficient. The law of mass action for ion-exchange species i is represented as (Yeh and Tripathi, 1990):

$$K_{ij} = \left(\frac{B_i}{A_i}\right)^{\nu_j} \left(\frac{A_j}{B_j}\right)^{\nu_i}, \tag{8}$$

where K_{ij} is the effective equilibrium constant of the ion-exchanged species, A and B are the activities (M/L) of an aqueous species and an ion-exchanged species, and ν is the charge of an ion-exchanged species. The law of mass action for a precipitated species is (Yeh and Tripathi, 1990):

$$1 = K_i^p \prod_{k=1}^{N_a} X_k^{a_{ik}^p}, \tag{9}$$

The superscript x, y, and p in eqs.(6), (7) and (9), suggest complexed, absorbed, and precipitated species, respectively. The subscript ik of a and b suggests the association of aqueous and absorbent component species k with the complexed, precipitated, and absorbed species i, respectively. Redox and acid-base reactions are considered as subclasses of aqueous complexation by invoking the principle of electron conservation and the proton condition, respectively. Subject to appropriate boundary conditions, eqs.(1), (3) - (9) completely define a subsurface hydrogeochemical system that was solved by a Galerkin finite element method in the parallel codes PFEM (D'Azevedo et al., 1993) and PHGC3D.

The flow equation (1) governs the movement of a fluid in a variably saturated porous medium and is widely know as the Richard's equation. The system is nonlinear because the hydraulic conductivity tensor **K** is dependent on the primary variable h; however the matrix equation is symmetric and is solved by a preconditioned conjugated gradient method. The solute transport equations (3)-(5) govern the movement of chemical species in the fluid and on the solid surface. The velocity term in eq.(3) is integrated into the matrix equation of a discretized finite element model using an upstream weighting method. This practice results in an asymmetric coefficient matrix. A bi-conjugate gradient stabilized (BI-CGSTAB) solver (Barrett, et al., 1994) is therefore used to solve the asymmetric matrix equation. The relative efficiency of various preconditioner used in the bi-conjugate gradient solver will be discussed later. The chemical equilibrium equations (6)-(9) are solved using a Newton-Raphson method and are directly coupled with the solute transport equations (3)-(5). Dimension arrays of PHGC3D are mostly linearly distributed among the processors according to the numbering of finite element nodes. The solute transport system and the chemical equilibrium equations are solved iteratively until the concentrations of the chemical species in the two systems are agreed within a user-specified error tolerance.

HYDROGEOLOGY OF THE MELTON BRANCH SUBWATERSHED AND THE URANIUM MILL TAILING PROBLEM

The Melton Branch subwatershed (MBSW) is a forested catchment with steep slopes and a shallow (<1 m) soil profile, overlying intensively folded and faulted sedimentary rocks. The vadose zone reaches a maximum thickness of ~20 m, locally decreasing to <1 m (following topographic trends). Mean annual precipitation is on the order of 1300 mm/y; episodic, high intensity recharge events (often related to precipitation events) characterize the region. Saprolite, a highly weathered shale that contains lenses of clay, comprises a 0.5-3 m zone between the soil horizon and underlying (unweathered) shale. The original structure of the parent material (e.g. lamination, bedding planes) has been well-preserved in the saprolite. The < 2 μm clay fraction of the sediment is primarily illite with lesser quantities of kaolinite, smectite, and interstratified 2:1 minerals (Kooner et al., 1995). Many of the bedding planes, fractures, and other macroporous surfaces are coated with secondary Fe- and Mn-oxides and translocated clay minerals (Wilson et al., 1993).

The Melton Branch Subsurface Facility has the characteristics of a highly fractured site that involves in a wide range of chemical reactions. These reactions include (1) ion-exchange of major cations such as Ca, Mg, K, and Na between the aqueous solution and clay minerals such as illite and kaolinite; (2) hydrolysis and polymerization of Al ions; (3) ion-exchange of major anions such as sulfate and dissolved carbons between aqueous solution and solid surface; (4) complexation of dissolved carbons with Al and Fe ions, and (5) in relatively deeper aquifers, the dissolution of $CaCO_3$. The MBSW has been the subject of experimental and theoretical research focused on preferential flow and solute transport in the vadose and shallow groundwater zone at a variety of scales (Wilson et al., 1989, 1993; Jardine et al., 1988, 1993 a,b; Gwo et al., 1995). Careful measurements of physical transport parameters (e.g. hydraulic conductivity, porosity) have been made at both the laboratory and field scales.

Topographically, the MBSW is a 0.57 hr subcatchment that looks like a baseball park (Fig. 1). To calculate the velocity and pressure fields of the watershed, we used 40% of the average annual precipitation (approximately 60% of the annual precipitation is loss to the evapotranspiration processes, see Moore and Toran, 1992) as surface infiltration into the soils. Three soil layers, namely, the B and C horizons and the limestone-shale bedrock, were used with interdispersed clay lenses in the C horizon. Hydraulic parameters of the soils can be found elsewhere (Gwo et al., 1996; James et al., 1996). The vertical sides of the watershed, except part of the outflow area near the toe of the hillslope, were modeled as no flow boundaries, and so was the bottom of the model. A vertical trench of 13 m wide and 3 m deep was previously excavated at the outflow area to collect subsurface stormflow water. The upslope face of the trench was therefore modeled as free seepage face. The entire watershed was represented by 10,500 elements, and the transient fluid velocity, pressure head and water content were calculated at 12,090 nodes. These nodes were regrouped into 16 subdomains to perform the flow calculation using PFEM.

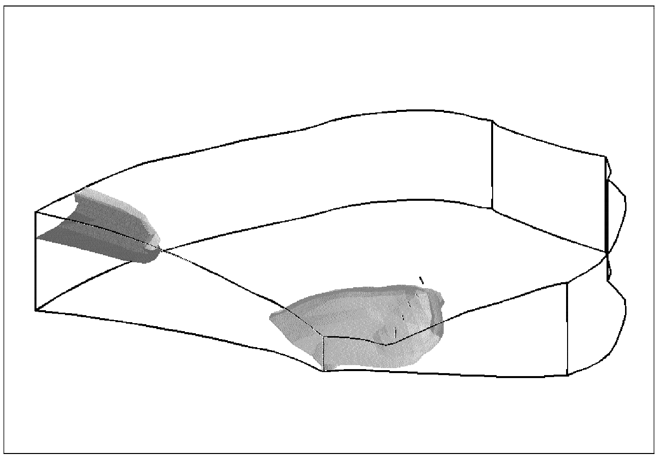

Fig. 1. Concentration isosurfaces of uranium (dark, at 1.10e-7 M) and proton (light, at 9.90e-4 M) at MBSW.

Uranium tailing problems that remain from the development of nuclear weapons by the Manhattan Engineer District, U.S. Army, and the Atomic Energy Commission are part of the cold war legacy. Various remedial actions have been taken by the U.S. Department of Energy to reduce emission of gamma radiation and radon gas and its decay products (Task Committee on Low-Level Radioactive Waste Management of the Technical Committee on Nuclear Effects, 1986). The acidic tailing waters may result in human exposure to radionuclides through various pathways and cause health hazard. Numerous incidences of human health and environmental concerns have been reported since 1966 (Task Committee on Low-Level Radioactive Waste Management of the Technical Committee on Nuclear Effects, 1986). Major concerns with groundwater contamination by the tailing water stem from direct ingestion of groundwater and groundwater discharged into surface water. Secondary ingestion from crops or stock contaminated directly and indirectly by tailing water and from the crops grown in contaminated soils were also considered possible pathways.

The uranium tailing problem considered here is a simplified hypothetical scenario of tailing water leaking through a surface treatment tank on the southeastern corner of the site (upper left corner in Fig. 1). The initial concentrations of the chemicals were similar to those used by Yeh and Tripathi (1990). In their report, they simulated the movement of 7 chemical components and 56 chemical species along a 2-dimensional hillslope. The chemical reactions in this study, including aqueous complexation, absorption/desorption, precipitation and acid-base reactions, were the same as theirs. The solute transport system therefore resulted in 84,630 degrees of freemdom in the numerical model, with 7 solute transport equations to solve simultaneously at each finite element node. Figure 1 shows the concentration isosurfaces of proton and uranium at 26.5 days after the initial leaking of the tank, which suggests clearly the vertical infiltration of solute fronts beneath the leaking tank and the depletion of proton near the outflow area due to the discharge of groundwater.

CODE PERFORMANCE AND DISCUSSION

The scale-up efficiency and performance of PHGC3D in simulating an 8-time-step hydrogeochemical uranium mill tailing problem, as of the writing of this manuscript, is shown in Fig. 2. A diagonal preconditioner with explicit message passing (during matrix multiplication) handled by native Intel NX library routines was used for these simulations. ASCII file output and input of the velocity, pressure head and water content fields were also handled by native Intel NX library calls. The chemistry module scaled linearly as expected; however, the I/O and construction of global matrix (not shown) performed poorly when the number of processors was greater than 65. The time used by the BI-CGSTAB solver was relatively small; however, increasing communication overhead at higher number of nodes also resulted in more computation time used by the solver.

The relative performance of four preconditioners, namely diagonal, LU decomposition, Gauss-Seidel, and SSOR (symmetric successive over-relaxation) in the BI-CGSTAB solver is shown in Table 1. The diagonal preconditioner was the best among the four in terms of per-iteration efficiency. LU decomposition of the asymmetric 186 x 186 local matrix proved to be very expensive and resulted in the poorest performance in both per-iteration efficiency and absolute run time among the four preconditioners. Approximately 73% of the total time required for 8 time steps was used for linear solve, of which 96% was used for

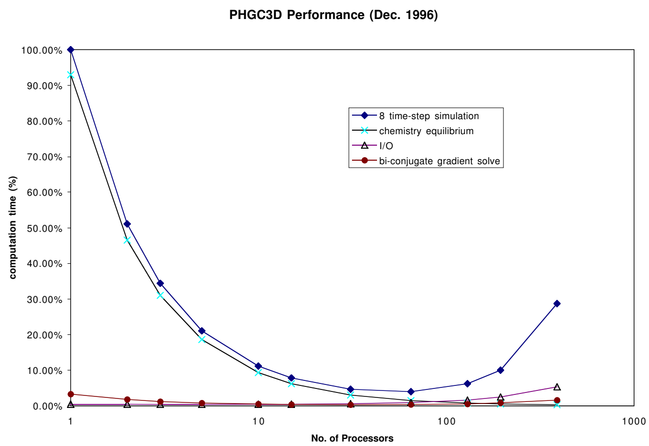

Fig. 2. Scale-up efficiency of PHGC3D using an 8-time-step simulation.

LU decomposition. In terms of linear solve time, the Gauss-Seidel preconditioner (one forward-backward sweep) outperformed diagonal preconditioner with narrow margins. The relatively smaller linear solve times as compared with the preconditioner times suggest that HPF intrinsic calls for vector inner product used in PHGC3D also consumed relatively large amount of computation time. This observation also suggests that, at relatively higher number of processors, performance of the linear solver may deteriorate because of higher HPF massage passing overhead.

Table 1. Comparison of preconditioner performance

time[*]	diagonal	LU decomp.	Gauss-Seidel	SSOR
PC/I[1]	1	338.35	4.55	4.34
LS/I[1]	1	11.60	1.13	1.48
LS[2]	1	32.75	0.96	1.37
LS time[1]	1	9.69	0.93	1.27

[*]Notation in this column: PC = preconditioner, I = iteration, LS = linear solve.

[1]Computation times were converted to the percentages of total run times and then normalized to the percentage time of diagonal preconditioner.

[2]Computation times were normalized to that of the diagonal preconditioner. Only one run using each preconditioner was used.

The relative performance of ASCII text I/O using HPF calls and native Intel NX read/write statements is shown in Table 2. No parallel I/O was used for the NX I/O calls, and the numbers shown are relevant only to ASCII text output of relatively small numbers of variables in each write statement. HPF I/O was about three times slower than NX native I/O calls in terms of performance per time step, and 14 times slower in terms of overall performance within the code. The authors observed, however, that NX FORTRAN unformatted input and output of relatively large sizes did not result in performance gain over HPF I/O (data not shown here).

Table 2. Comparison of I/O and message passing performance

time[*]	reference[!]	HPF I/O[!!]	HPF message passing[!!!]
MM/I[1]	1	-	34.91
LS/I[1]	1	-	5.48
LS[2]	1	-	8.59
I/O/T[1]	1	3.02	-
I/O[2]	1	14.16	-

[*]Notation in this column: MM = matrix multiplication, LS = linear solve, I = iteration, T = time step.

[!]The reference case used the diagonal preconditioner (see Table 1) and native Intel NX calls for file I/O and message passing.

[!!]Only output of ASCII text was compared.

[!!!]Only message passing for matrix multiplication in the linear solver was compared.

[1]Computation times were converted to the percentages of total run times and then normalized to the percentage time of the reference case.

[2]Computation times were normalized to that of the reference case. Only one run of each I/O and message passing methods was used.

In order to test message passing efficiency of HPF, we also compared HPF gather operation with NX native routines in which only data relevant to the current subdomain were gathered from its neighbors. Without the freedom of gathering data from neighboring subdomains only, the HPF gather operation in PHGC3D uses an assignment statement that copies data distributed on the processors into a local working array. As expected, this operation resulted in passing unnecessary data and relatively higher overhead. The performance of HPF gather operation may deteriorate quickly if more processors are involved. The performance data shown in Table 2, using 65 processors, suggest that NX native message passing operation was 35 times faster than HPF gather operation. This gain in message passing translated into 5.5 times performance gain during linear solve.

CONCLUSION

Using an HPF compiler to parallelize a hydrogeochemistry code, we were able to achieve reasonably high efficiency in a relatively short period of time. However, we were able to achieve only 40% of linear speed-up using 65 nodes, and there are still ample amount of room for performance optimization. Resorting to NX native library calls to gain additional performance reduces portability and is not the best way of achieving acceptable parallel efficiency and utilizing parallel computational resources. During this work and the writing of this manuscript, the PGI HPF compiler had been upgraded a few times. Compiler technology and parallel application development are evolving processes. The results presented here are meant to be used as a reference point along the evolving processes of a compiler technology. In summary, this work has shown that (1) HPF is a viable tool for achieving reasonable performance gain in a relatively short period of time, (2) chemical equilibrium calculation in hydrogeochemistry codes scales perfectly linear, (3) diagonal and Gauss-Seidel are more efficient than LU decomposition and SSOR in a bi-conjugate gradient solver, and (4) more effort is needed to address I/O and message passing performance of HPF codes using domain-decomposition.

ACKNOWLEDGMENT

This work was funded in part by the ORNL Partnership in Computational Sciences (PICS) program, supported by the Department of Energy's Mathematical, Information, and Computational Sciences (MICS) Division of the Office of Computational and Technology Research.

REFERENCE

Barrett, R., m. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozp, C. Romine, H. van der Vorst, 1994. Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods. SIAM, Philadelphia.

D'Azevedo E. F., C. H. Romine, L. E. Toran, O. R. West, and J. P. Gwo. 1993. High performance computing for environmental modeling and remediation. Supercomputing 93, Portland, Oregon, Nov. 15-19.

Gwo, J. P., L. E. Toran, M. D. Morris, and G. V. Wilson, 1996. Subsurface stormflow modeling with sensitivity analysis using a Latin-hypercube sampling technique. Ground Water, 34(5): 811-818.

Gwo, J. P., P. M. Jardine, G. V. Wilson, and G. T. Yeh, 1995. A multiple-pore-region concept to modeling mass transfer in subsurface media, J. Hydrol., 164, 217-237.

James, B. R., J. P. Gwo, and L. Toran, 1996. Risk-cost decision framework for aquifer remediation design. J. of Water Resour. Plann. and Manag., 122(6): 414-420.

Jardine, P. M., G. V. Wilson, and R. J. Luxmoore, 1988. Modeling the transport of inorganic ions through undisturbed soil columns from two contrasting watersheds, Soil Sci. Soc. Am. J. 52, 1252-1259.

Jardine, P. M., G. K. Jacobs, and G. V. Wilson, 1993. Unsaturated transport processes in undisturbed heterogeneous porous media: I. Inorganic contaminants, Soil Sci. Soc. Am. J., 57, 945-953.

Kooner, Z.S., P. M. Jardine, and S. Feldman, Competitive surface complexation reactions of sulfate and natural organic carbon on soil, J. Environ. Qual., 24, 656-662, 1995.

Moore, G. K. and L. E. Toran, 1992. Supplement to a Hydrologic Framework for the Oak Ridge Reservation, Oak Ridge, Tennessee. ORNL/TM-12191. Oak Ridge National Laboratory.

The Task Commission on Low-Level Radioactive Waste Management of the Technical Committee on Nuclear Effects, 1986. Management of inactive uranium mill tailing. J. of Environ. Engr., 112(3): 490-537.

Wilson, G. V., J. M. Alfonsi, and P. M. Jardine, 1989. Spatial variability of saturated hydraulic conductivity of the subsoil of two forested watersheds. Soil Sci. Soc. Am. J., 53, 679-685.

Wilson, G. V., P. M. Jardine, J. D. O'Dell, and M. Collineau, 1993. Field-scale transport from a buried line source in variably saturated soil, J. Hydrol., 145, 83-109.

Yeh, G.-T. and V. S. Tripathi, 1990. HYDROGEOCHEM, A Coupled Model of HYDROlogic Transport and GEOCHEMical Equilibria in Reactive Multicomponent Systems. ORNL-6371. Oak Ridge National Laboratory.

Yeh, G.-T., 1987. 3DFEMWATER: A Three-Dimensional Finite Element Model of WATER Flow Through Saturated-Unsaturated Media. ORNL-6386. Oak Ridge National Laboratory.