Applications of High Performance Computing in Engineering VI, C.A. Brebbia, M. Ingber & H. Power (Editors)

MISER (Michigan soil-vapor extraction remediation model): development of a parallel environmental remediation application code

J. R. Lang[1], J. K. Landrum[2], C. D. Drummond[1], L. M. Abriola[1] & Q. F. Stout[2]
[1]*Dept. of Civil and Environ. Engin., Uni. of Michigan., USA*
[2]*Dept. of Elect. Engin. and Computer Sci., Univ. of Michigan, USA*

Abstract

A multidisciplinary research effort was undertaken to develop a parallel application code based on an existing serial application. MISER is a state of the art environmental remediation simulator designed to model the physical, chemical, and biological interactions occurring in field scale soil vapor extraction (SVE) and bioventing (BV) systems. The simulator is based on a series of equations describing multiphase flow, multicomponent advective diffusive transport, and bioreaction. These equations can be highly nonlinear due to the nature of the bioreaction terms, constitutive relationships, and material properties. Rate limited mass exchange between phases is modeled with linear driving force expressions. A standard Galerkin finite element method in two dimensions with linear triangular elements is employed using a set iterative approach to solve the model system of nonlinear time dependent partial differential equations. In this approach the sets of flow and transport equations are decoupled within the simulator and solved separately.

The parallel application of MISER was developed for the IBM SP2 by using Message Passing Interface directives. Parallelization was accomplished with a domain decomposition approach by partitioning the domain into blocks, each assigned to an individual processor. This parallel implementation permits the use of larger simulation domains, more refined numerical grids, and the inclusion of more complex and complete descriptions of the subsurface environment (i.e. more components, processes, and/or phases). These issues of scale and model complexity are critical to the accurate simulation of subsurface remediation. Simulations are presented which demonstrate the improved performance and scalability of the parallel application code.

Applications of High Performance Computing in Engineering VI, C.A. Brebbia, M. Ingber & H. Power (Editors)

1 Introduction

The contamination of subsurface soils and groundwater resources by hazardous chemicals is a national problem of enormous magnitude and potential impact on public health. Cleanup costs at National Priorities Listed Sites are estimated to range up to $750 billion in 1993 dollars over the next 30 years [1]. Furthermore, traditional subsurface restoration methods, commonly known as "pump-and-treat" technologies, have proven generally inefficient and ineffective, motivating the development of innovative remedial techniques. Examples of innovative remediation technologies include *in situ* microbial degradation, soil venting, surfactant or solvent flushing, and various combinations of these processes [1]. The evaluation of such emerging technologies requires the integration of physical, chemical, and biological processes in a multiphase, heterogeneous environment. While small-scale laboratory experiments are useful to explore individual processes or to demonstrate the potential performance of a particular remediation strategy, field deployment of innovative technologies will require the use of mathematical models for design and performance assessment. These models must simulate the complex and nonlinear interactions of multiple processes under the heterogeneous conditions typical of field settings.

MISER (Michigan soil-vapor extraction remediation model) is one such model, designed to simulate the related processes of soil vapor extraction (SVE) and bioventing (BV). Both processes rely on induced movement of the subsurface gas phase to remediate areas contaminated with hydrocarbons [2]. In the case of SVE, subsurface gas containing volatilized hydrocarbons is withdrawn from a well and treated above ground; in the case of BV, *in situ* biodegradation is enhanced via the injection of air. Currently the ability of MISER to simulate field scale applications is generally compromised by computational limitations, particularly at the fine discretization requirements necessary to accurately resolve heterogeneous soil properties and equation behavior. Conventional supercomputers represent one approach that has previously been used to address the computational limitations described above [3]. The use of parallel systems for the modeling of subsurface remediation processes potentially offers a number of advantages over conventional supercomputers (c.f. [4], [5], [6]) and is yet to be fully exploited [7]. In this work, the development of a parallel implementation of the transport and biodegration components of MISER using Message Passing Interface (MPI) directives and an existing parallel solver, Aztec [8], on the IBM SP2 at the University of Michigan's Center for Parallel Computing is described and preliminary performance benchmarks are discussed.

2 Conceptual model and mathematical formulation

The conceptual model, which provides the framework for development of MISER, is described briefly below. MISER uses more than a dozen coupled partial differential equations, as well as a number of ordinary differential and algebraic equations describing microbial growth and transformation, and several constitutive relationships to represent SVE and BV. Three fluid phases are

Applications of High Performance Computing in Engineering VI, C.A. Brebbia, M. Ingber & H. Power (Editors)

modeled: an entrapped non-aqueous phase liquid (NAPL), composed of an arbitrary number of organic contaminants; a mobile gas phase, composed of volatile organic contaminants, oxygen, nitrogen, and a limiting nutrient; and a mobile aqueous phase, composed of soluble organic contaminants, oxygen, water, and a limiting nutrient. The mobile phases can flow simultaneously in response to stresses induced by wells and/or density gradients. Sorption of the organic contaminants to the soil matrix is also considered. Activity by indigenous microorganisms attached to the soil matrix can result in the biotransformation of the organic contaminants, oxygen, and limiting nutrient.

In this initial parallel application, the subsurface formation is assumed to be spatially homogeneous in permeability and porosity, the mobile gas phase is assumed to move at a constant velocity, and the aqueous phase is at residual saturation and hence immobile. Due to these simplifying assumptions, no equations are needed to describe the movement of the mobile phases. Both the gas and aqueous phases are in contact with each other and with the organic phase. The aqueous phase is also in contact with the attached biophase. Because the organic liquid is assumed to be immobile in MISER, changes in organic liquid saturation result solely from interphase mass transfer; e.g., the organic liquid mass balance is [9]:

$$\frac{\partial}{\partial t}\left(\phi \rho_o^* S_o\right) = \phi \sum_i E_{o\beta i}^* \tag{1}$$

where t is time, ϕ is the porosity, ρ_o^* is NAPL mass density, S_o is the NAPL saturation, and $E_{\alpha\beta i}^*$ is the net rate of mass of component i transferred to the α-phase (NAPL: o) from all contiguous phases β (aqueous: a, and gas: g).

The composition of the fluid phases is subject to the following assumptions. The organic liquid is considered to be a mixture of an unrestricted number of organic components (in the example simulations, only one component is considered). The gas phase is assumed to be comprised of nitrogen and oxygen (i.e., the two major components of air), water vapor (neglected in this work), volatile components of the organic liquid, and a limiting nutrient (if present in the aqueous phase and volatile). The aqueous phase is limited to components of the organic liquid, water, oxygen, and a limiting nutrient (neglected in this work). The migration of individual phase components is represented by a general macroscopically averaged transport equation [9]:

$$\frac{\partial}{\partial t}\left(\phi S_\alpha \rho_\alpha x_{\alpha i}\right) + \nabla \cdot \phi S_\alpha \left(\rho_\alpha x_{\alpha i} V_\alpha - \rho_\alpha D_{\alpha i}^h \nabla x_{\alpha i}\right) = \phi E_{\alpha\beta i} \tag{2}$$

where ρ_α is phase molar density, $x_{\alpha i}$ is the mole fraction of component i in phase α, V_α is the average phase velocity, and $D_{\alpha i}^h$ is the phase hydrodynamic dispersion tensor of component i. Nonequilibrium interphase partitioning is represented with a linear driving force expression [10]:

$$E_{\alpha\beta i} = \rho_\alpha K_{\alpha\beta i} S_\alpha \left(x_{\alpha i} - x_{\alpha i}^e\right) \tag{3}$$

where α is the controlling phase, $K_{\alpha\beta i}$ is the overall α-β mass transfer coefficient, and $x_{\alpha i}^e$ is the α phase mole fraction of component i in equilibrium with the mole fraction of i in the β phase. Equation (3) is used to model sorption, volatilization and dissolution of entrapped organic liquids, and

Applications of High Performance Computing in Engineering VI, C.A. Brebbia, M. Ingber & H. Power (Editors)

gas/aqueous interchange of organic components and oxygen. The mass transfer resistance is assumed to occur in the aqueous phase for organic dissolution, gas/aqueous partitioning and sorption, and in the gas phase for organic liquid volatilization. Partitioning of the organic component(s) to the solid phase is described with Freundlich isotherms.

Biodegradation is assumed to occur only within the aqueous phase by an indigenous, spatially heterogeneous, mixed microbial population that is present as attached microcolonies. It is assumed that biomass growth does not affect soil permeability and there is no biomass transport. Monod-type kinetic equations are used to describe biomass uptake, b_{ai}, of the degradable organic contaminants, oxygen, and a limiting nutrient (not present in this work) from the aqueous phase [11], [12]:

$$b_{ai} = -F_{il}\mu_l X\left(\frac{\rho_a x_{al}}{k_{sl} + \rho_a x_{al}}\right)\left(\frac{\rho_a x_{aO_2}}{k_{sO_2} + \rho_a x_{aO_2}}\right) I_i \qquad (4)$$

where F_{il} is the use coefficient of component i with substrate l degradation ($F_{il}=1$ for $i=l$), μ_l is the maximum specific substrate utilization rate of substrate l, X is the active biomass concentration, k_{sl} is the half-saturation coefficient of component l, O_2 is the election acceptor, and I_i is an inhibition function either restricting biodegradation when substrate concentrations exceed a specified limit or when oxygen concentrations are below a specified limit. When i is O_2, Equation (4) is summed over all the substrates l. Growth of the microbial population is represented by:

$$\frac{dx}{dt} = \left[\sum_l Y_l \mu_l \left(\frac{\rho_a x_{al}}{k_{sl} + \rho_a x_{al}}\right)\left(\frac{\rho_a x_{aO_2}}{k_{sO_2} + \rho_a x_{aO_2}}\right) I_{max} - K_d I_{min}\right] X \qquad (5)$$

where Y_l is the yield coefficient for the metabolism of substrate l, I_{max} is an inhibition function limiting the biomass concentration from exceeding a maximum value, I_{min} is an inhibition function preventing the biomass concentration from decreasing below a minimum value, and K_d is the decay coefficient of the microorganisms. The model neglects any influence of growth on soil permeability. For a complete description of MISER, the reader is referred to Abriola [13] and Rathfelder [14].

3 Numerical solution approach

The coupled nonlinear flow and transport equations can be solved for a cross-sectional (x-z) and axisymetric (r-z) domains with the Galerkin finite element method, using linear triangular elements. MISER employs a modular, set-iterative solution scheme, where, the sets of flow, transport, and biodegradation equations are decoupled and solved separately. Decoupling is accomplished by lagging, either by one iteration or one time step, the coupling terms which include the phase density, interphase mass transfer, and bioreaction terms. Picard iteration is used to account for nonlinearities and the set of equations is considered converged when the max norm of the differences between successive solutions normalized by the max norm of the initial conditions is less than 10^{-6}.

Applications of High Performance Computing in Engineering VI, C.A. Brebbia, M. Ingber & H. Power (Editors)

4 Parallelization

The parallel implementation of MISER was developed for the IBM SP2 using MPI directives and a preexisting parallel solver package, Aztec [8]. The global domain was input and partitioned such that problems larger than the memory available to the 0 processor (which handles all input/output operations) are possible. Currently the partitioning routines are designed for structured grids and p^2 processors (4, 9, 16, 25 processor configurations were used in this work). The global grid is partitioned into p^2 blocks, with each block assigned to an individual processor. This partitioning was chosen to facilitate code development, to minimize interprocessor communication, and to retain relative flexiblity of processor configuration. The overlap between adjacent processors can vary, in this work it was set to one row of elements. Each subdomain is numbered locally. Once the grid is partitioned, most interprocessor communication is handled by the solver, with the exception of convergence and time step size adjustment information. These are controlled by the least convergent subdomain. This is necessary due to the strong coupling between the various partial differential equations in MISER.

Aztec is a library of iterative routines for the solution of sparse linear systems of equations. Aztec handles all updating of variables not computed on a given processor, as well as reordering of the local solutions for increased efficiency. In this work, the generalized minimum residual (GMRES) iterative method was chosen, with 3 step symmetric Gauss-Siedel preconditioning and point Jacobi scaling. The iterative method was considered converged when the residual ($\|r\|_2/\|r^0\|_2$) was reduced to less than 10^{-6}.

All simulations presented in this work were performed on the IBM SP2 at the University of Michigan Center for Parallel Computing using the thin POWER2 Super Chip (P2SC) processors with 1GB of memory per processor. The SP2 processors are superscalar (implying simultaneous execution of multiple instructions) pipelined chips and are capable of executing up to six instructions per clock cycle and four floating point operations per cycle. The application nodes run at 160 Mhz and are capable of a peak performance of 640 MFLOPS each. Each node has 8GB of local /scratch space available on disk.

5 Example Simulations

Consider the cross sectional simulation domain presented in Figure 1. Gas flow is horizontal and the aqueous phase is at an immobile residual saturation of 0.073. The entrapped NAPL source consists entirely of toluene at a residual saturation of 0.01. The top and bottom boundaries are impermeable, the left hand side boundary is third type with an influent mole concentration of 0.0 for toluene, 0.18 for oxygen, and the right hand boundary is second type or zero diffusive flux. Biokinetic parameters include: the use coefficient of oxygen with toluene degradation (2.19 moles/mole); the maximum specific substrate utilization rate of toluene (1.0/day); the half-saturation coefficient of toluene (0.5 mg/l) and of oxygen (0.1 mg/l); the biomass decay rate (0.1/day); the initial and minimum biomass (0.0162 mg/l bulk phase); the maximum biomass (20 mg/l bulk phase); and the concentration thresholds above which toluene degradation

Applications of High Performance Computing in Engineering VI, C.A. Brebbia, M. Ingber & H. Power (Editors)

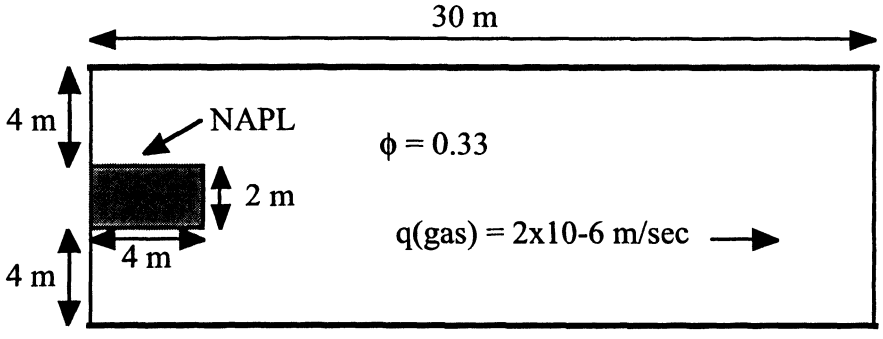

Figure 1. Simulation domain used for the performance assessment.

ceases (90% of saturation) and below which oxygen respiration ceases (1% of saturation). Freundlich sorption parameters for toluene were n = 1 and a soil capacity of 0.0727 μg/g solid with the aqueous concentration in mg/l. Mass transfer coefficients used in the presented simulations represent near equilibrium values (all coefficients were 10^{-2}/sec and then limited through an analytical expression which only allows a maximum approach to equilibrium within a single element of 95%). Equilibrium or near equilibrium conditions represent a challenging numerical problem and serve to test the code performance with realistic parameter values.

6 Discussion

Eight different herringbone grids were used to simulate this domain. They are summarized in Table 1. For grid 8 the horizontal extent of the domain was extended to 45 m. For purposes of verification, parallel and serial simulations on Grid 3 were compared at 5000 seconds of simulated time and close matches were achieved. Other scenarios were tested as well, however in the interest of

Table 1. Example grids and simulation times.

Grid	N	Grid Spacing (m)	Time for one iteration (sec) Processor number				
			1	4	9	16	25
1	1281	0.5 x 0.5	0.0361	0.020	0.035	0.034	0.046
2	3333	0.3 x 0.3125	0.1013	0.035	0.038	0.041	0.051
3	4961	0.25 x 0.25	0.1558	0.047	0.048	0.043	0.052
4	13065	0.15 x 0.15	0.4692	0.110	0.072	0.058	0.063
5	19521	0.125 x 0.125	0.7314	0.162	0.096	0.070	0.069
6	38961	0.1 x 0.08	1.6166		0.177	0.119	0.099
7	77841	0.625 x 0.625	(3.578)			0.235	0.164
8	116221	0.625 x 0.625	(5.342)				0.215

Applications of High Performance Computing in Engineering VI, C.A. Brebbia, M. Ingber & H. Power (Editors)

brevity they are not included in this work. The example simulations were run to 200 time steps for grids 1 through 5, and to 50 time steps for grids 6 through 8. The time step size was allowed to increase when the total number of Picard iterations per time step was less than 5. Results are reported as time per iteration where one iteration is taken to be one Picard iteration over a single equation, averaged among the different equations solved during each time step. Several iterations of the solver were required for each equation, the number remaining relatively constant throughout the simulations.

The example simulations are also summarized in Table 1. Runs were conducted in triplicate for the 1, 4, and 9 processor configurations and the reported times were averaged over the three runs. For the 16 and 25 processor configurations, only a single run was conducted. The times in parentheses for grids 7 and 8 on a single processor are estimated. For grid 7, an expression of the form CN^d was fit to the computed solutions, where C and d are constants and N is the number of degrees of freedom. Grids 7 and 8 have the same bandwidth, so the scaling from grid 7 to grid 8 was linear in N. This was necessary since these two simulations exceeded the memory capacity of the individual IBM SP2 processor used in this work (approximately 40,000 nodes could be allocated to an individual processor). Solutions on the finer grids advanced the time step slower than did the coarser solutions, however the increased grid resolution was one of the goals of this work and therefore time step size was not used as an element in performance assessment.

The results of the sample simulations are presented in Figures 2 and 3. Figure 2 shows that the parallel implementation of MISER is scaling well as

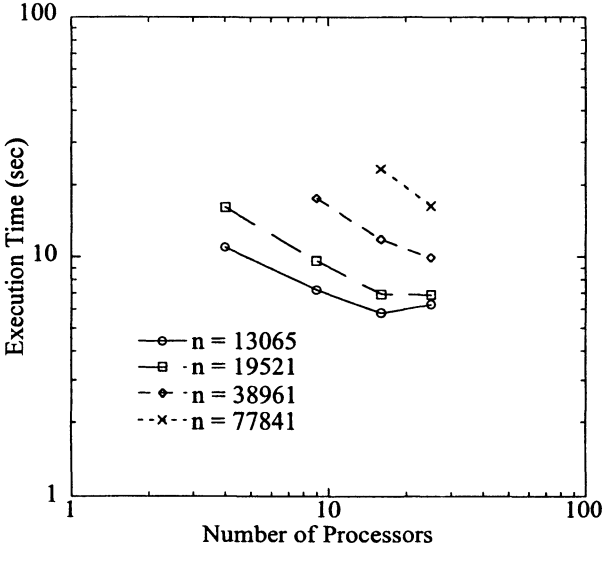

Figure 2. Plot of the log of time in seconds to perform 100 iterations vs. the log of the number of processors for four different numbers of degrees of freedom sizes.

Applications of High Performance Computing in Engineering VI, C.A. Brebbia, M. Ingber & H. Power (Editors)

long as approximately 800 nodes are allocated to each processor. Below that number, interprocessor communication overhead prevents efficient use of the multiprocessor architecture. Figure 3 shows the speedup on a given number of processors as the problem size increases. Here, speedup is defined as the time required on a single processor for one Picard iteration over a single N^2 transport equation averaged over the several different equations solved in MISER divided by the time required for the same iteration partitioned onto several processors. Ideally, the speedup should be the number of processors (i.e. for 16 processors the speedup should be 16). For this problem, a complicating factor is present. The iterative solver may require more or less iterations to converge depending on the partitioning and discretization. This is a likely cause of the speedup apparently being greater than the ideal number for the 4 processor runs on grids 4 and 5 and the 9 processor run on grid 6. Overall, Figure 3 suggests that the implementation is scaling well. The 4 and 9 processor configurations are most likely approaching their respective limits and substantial performance gains are observed when using greater numbers of processors. Additionally, finer grid resolutions are possible when using the IBM SP2, on the order of 100,000 nodes or more. This is not possible on workstations as typically configured and even if sufficient memory were available on a workstation, the time required to produce a solution would be prohibitive. In general, 20,000 to 40,000 nodes are sufficient for a wide range of two-dimensional problems, and the larger grids (100,000 nodes or more) used in this work demonstrate the feasibility of extending MISER to three dimensions.

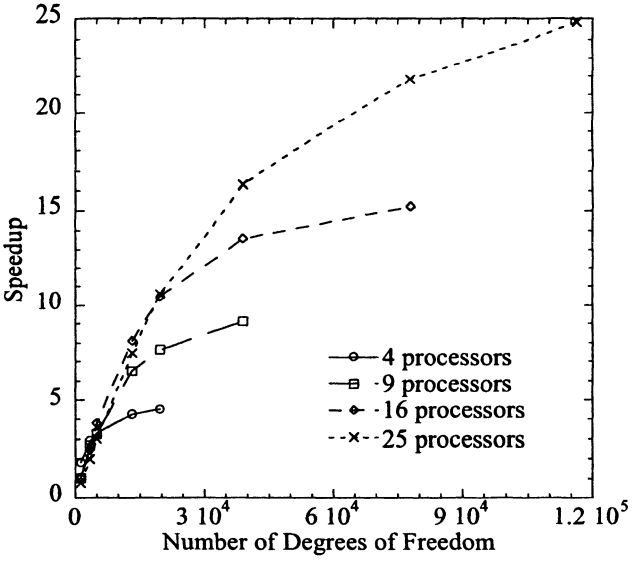

Figure 3. Plot of speedup vs. the number of degrees of freedom for four different processor numbers (4, 9, 16, and 25).

Applications of High Performance Computing in Engineering VI, C.A. Brebbia, M. Ingber & H. Power (Editors)

7 Conclusions

The parallel implementation of MISER has been shown to scale well as long as approximately 800 nodes are allocated to each processor. In addition, much larger grids are possible, allowing finer resolution at the field scale, than is currently achievable on typically configured engineering class workstations. Future work will focus upon the inclusion of the equations describing phase movement and the selection of solver options optimized for each of the different types of equations included in MISER.

Acknowledgements

This work was funded as part of the Applications thrust area in engineering supported by the National Partnership for Advanced Computational Infrastructure (NPACI). Cooperating institutions are the Center for Subsurface Modeling at the University of Texas and the Center for Parallel Computing at the University of Michigan.

References

[1] National Research Council, Alternatives for Ground Water Cleanup. National Academy Press, Washington D.C., 1994.

[2] Rathfelder, K.M., J.R. Lang, and L.M. Abriola, Soil vapor extraction and bioventing: applications, limitations, and future research directions. In R.M. Vogel (ed.). *Reviews of Geophysics, Supplement, U.S. National Report to IUGG, Contributions in Hydrology*, pp. 1067-1082, 1995.

[3] Chiang, C.Y., C.N. Dawson, and M.F. Wheeler, Modeling of in-situ biorestoration of organic compounds in groundwater. *Transport in Porous Media*, 6: 667-702, 1991.

[4] Dougherty, D.E., Hydrologic applications of the Connection Machine CM-2. *Water Resources Research*, 27(12): 3137-3148, 1991.

[5] Eppstein, M.J., and D.E. Dougherty, Comparative study of PVM workstation cluster implementations of a two-phase subsurface flow model. *Advances in Water Resources,* 17(3): 181-195, 1994

[6] Ashby, S.F., W.J. Bosl, R.D. Falgout, S.G. Smith, A.F.B. Tompson, and T.J. Williams, Numerical simulation of groundwater flow and contaminant transport on the CRAY T3D and C90 supercomputers. *International Journal of High Performance Computing Applications*, v 13 n1 Spring 1999: 80-93, 1999.

[7] Bernard, D., F. Bodini, A. Goasguen, and J.C. Fechant, Implementing a two-dimensional pore-scale flow model on different parallel machines. Computational Methods in Water Resources X, Volume 2, Peters et al. (eds.), pp. 1507-1514, 1994.

[8] Hutchinson S.A., L.V. Prevost, J.H. Shadid, C. Tong, and R.S. Tuminaro, Aztec User's Guide Version 2.0. Massively Parallel Computing Research Laboratory, Sandia National Laboratories, Albuquerque, N.M., 1998.

[9] Abriola, L.M., Modeling multiphase migration of organic chemicals in groundwater systems - A review and assessment. *Environmental Health. Perspectives 83*: 117-143, 1989.

[10] Welty, J.R., C.E. Wicks, and R.E. Wilson, Fundamentals of Momentum, Heat, and Mass Transfer. John Wiley & Sons. New York, NY, 1969.

[11] Chen, Y-M., L.M. Abriola, P.J.J. Alvarez, P.J. Anid, and T.M. Vogel, Modeling transport and biodegradation of benzene and toluene in sandy aquifer material: Comparisons with experimental measurements. *Water Resources Research 28*(7): 1833-1847, 1992.

[12] Chen, Y-M., Mathematical Modeling of *In-Situ* Bioremediation of Volatile Organics in Variably Saturated Aquifers. Ph.D. Dissertation, University of Michigan, Ann Arbor, MI, 1996.

[13] Abriola L.M., J.R. Lang, and K.M. Rathfelder, *Michigan Soil Vapor Extraction Remediation (MISER) Model: A Computer Program to Model Soil Vapor Extraction and Bioventing of Organic Chemicals in Unsaturated Geological Material.* Prepared for the U.S. Environmental Protection Agency, National Risk Management Research Laboratory, Office of Research and Development., Ada, OK, 1997.

[14] Rathfelder, K.M., J.R. Lang, and L.M. Abriola, A numerical model (MISER) for the simulation of coupled physical, chemical, and biological processes in soil vapor extraction and bioventing systems. *Journal of Contaminant Hydrology*, 1999 in press.