# SIMULATION OF REACTIVE GEOCHEMICAL TRANSPORT PROCESSES IN CONTAMINATED AQUIFERS USING SURROGATE MODELS

Hamed Koohpayehzadeh Esfahani<sup>1</sup> and Bithin Datta<sup>1,2</sup>

<sup>1</sup>School of Engineering and Physical Sciences, James Cook University, Australia;
<sup>2</sup>CRC for Contamination Assessment and Remediation of the Environment (CRC-CARE), University of South Australia, Mawson Lakes SA 5095, Australia

**ABSTRACT**: Transport of contaminant species undergoing chemical reactions in groundwater aquifers is a complex physical and biochemical process. Simulating this transport process involves solving complex nonlinear equations and requires huge computational time for a given aquifer study area. Development of optimal remediation strategies in aquifers may require repeated solution of such complex numerical simulation models. To overcome this computational limitation and improve the computational feasibility of large number of repeated simulations, Genetic Programming based trained surrogate models are developed to approximately simulate such complex transport processes. Transport process of acid mine drainage, a hazardous pollutant is first simulated using a numerical simulated model: HYDROGEOCHEM 5.0 for a study area resembling a mine site. Simulation model solution results for an illustrative contaminated aquifer site is then approximated by training and testing a Genetic Programming (GP) based surrogate model. To decrease the total number of GP formulations, the coordinates of observation locations are implemented as input data in the surrogate models. Comparison of the surrogate models and numerical simulation results show that the surrogate models can provide acceptable approximations of this complex transport process in contaminated groundwater aquifers.

*Keywords: Biochemical transport simulation, Genetic programming, Acid mine drainage, contaminated aquifers, surrogate models.* 

# 1. INTRODUCTION

Simulation of flow and transport processes for chemically reactive species in contaminated groundwater aquifers generally require extensive computational time. When repeated simulations are necessary to implement and solve optimization based decision model, e.g., to develop optimal aquifer management strategies, the computational burden may determine the feasibility of any methodology. To address this, trained surrogate models approximating the simulation model can be developed. These surrogate models are generally computer programs describing the relationship between output values (e. g., pollutant concentration at different locations and times) and input values (e.g., pollutant flux at potential pollutant source locations). Simulation of reactive species transport process (i.e. Acid mine drainage (AMD)) in contaminated groundwater aquifers is complex and computationally intensive. Therefore it is useful to develop trained and tested surrogate models to approximately simulate the transport processes. In this study, the flow and biochemical transport simulation model is replaced by trained genetic programming (GP) as surrogate models which can reduce consumption time.

An accurate description of the contaminant transport in aquifers is obtained if both chemical and

physical behaviors of contaminant species are incorporated. Solution of the transport process with chemical reactive species in groundwater was addressed by [1] and developed by [2-4]. One approach couples the non-reactive transport model MT3DMS [5] with various chemical reactive transport simulators [6-10]. HYDROGEOCHEM [11] was the first comprehensive three dimensional hydrogeologic simulator of transport and geochemical reaction in saturated-unsaturated media [11]. This code was developed to solve comprehensive heat, reactive geochemical and biochemical transport [12]. The proposed methodology uses HYDROGEOCHEM 5.0 (HGCH) as the simulation model for AMD transport process with chemically reactive pollutants for an illustrative study area with synthetic hydro-geochemical data. These simulation results are then utilized to train and test a GP based surrogate model.

AMD, which is the result of wastewater from metal mines or coal mines containing sulphur compounds [13], is hazardous contaminant sources for groundwater. Various sulphide minerals constitute AMD based on their chemical weathering reactions such as pyrite (FeS2), pyrrhotite (Fe1-xS), chalcopyrite (CuFeS2), arsenopyrite (FeAsS), etc. [14]. Beside mining activities, rocks' surface weathering in presence of water, oxygen and microorganisms produces AMD. These contaminants are considerable threats for water resources. In this study, the transport process of copper and sulphate, hazardous AMD's compounds, along with their chemical reactions through the aquifer is considered.

A response matrix approach as an initial linear surrogate model was earlier used to simulate the aquifer responses [15], [16]. More recently proposed non-linear surrogate models include Artificial Neural Network (ANN)[17] and Genetic Programming (GP) based surrogate models [18]. References [19-21] developed groundwater simulation methodology by replacing numerical simulation model with ANN-base surrogate models. Using trained ANN-based surrogate modelling approach obtains the optimal model formulation by trial and error [19]. Reference [20] predict the complex flow and transport processes using ANN as an approximate simulation for replaces the three dimensional numerical simulation model in a coastal aquifer. Reference [21] replaced the three dimensional numerical simulation model with trained ANN-based surrogate models for approximating density depended saltwater intrusion process in coastal aquifer.

Reference [22] implemented ANN as a surrogate model to replace the flow and transport simulation in the non-dominated front search process as well as to save a huge amount of computational time. References [23] and [24] used GP as meta-model for simulation of pumping saltwater patterns in an optimization framework. Reference [24] demonstrated GP model have several advantages compared with conventional surrogate models and ANN surrogate models. These advantages are: simpler surrogate models, optimizing the model structure more efficiently, and parsimony of parameters. Replacing simulation groundwater model by GP-based ensemble surrogate models in linked simulation-optimization developed methodology was addressed by [24, 25] which improve the computational efficiency and obtains reasonably accurate results under aquifer hydrogeologic uncertainties. Use of GP surrogate model for groundwater contamination management, and development of a monitoring network design feedback methodology to identify unknown source characteristics was addressed by [26].

In this study our main objective is to develop genetic programming based surrogate models to approximately simulate the complex transport process in a complex hydrogeologic system with reactive chemical species, and to illustrate its efficiency and reliability in a contaminated aquifer resembling an abandoned mine site.

# 2. METHODOLOGY

The current study proposes and evaluates a

methodology, which includes two steps. In the 1st step, a numerical simulation model with specified boundary conditions, specified existing initial hydraulic and geochemical conditions, and with hydrogeologic estimated and geochemical parameter values is used to simulate flow and biochemical transport processes for hazardous compounds of AMD such as sulfate and copper. in a given contaminated aquifer. The specified illustrative aquifer resembles an abandoned (no longer in operation) mine site in Queensland, Australia, where similar contaminants are present and similar topologic and hydrogeologic conditions exist. In the second step, trained GP models are implemented to obtain pollution concentration at specified locations of the contaminated aquifer. These concentrations obtained using the surrogate model and those obtained by solving the implemented numerical three dimensional transient reactive contaminant transport simulation model (HGCH) are compared to evaluate the potential applicability of the GP based surrogate models.

### 2.1 Simulation Model for Groundwater Flow and Biogeochemical Transport

The HGCH flow and transport simulation model consisting of the flow simulation model and physiochemical transport model to obtain numerical solutions. It is a computer program that numerically solves the three-dimensional groundwater flow and contaminant transport equations for a porous medium. The finite-element method is used in this simulation model. The general equations for flow through saturated-unsaturated media are obtained based on following equations [11]:

$$\frac{\rho}{\rho_0} \mathbf{F} \frac{\partial \mathbf{h}}{\partial t} = \nabla \cdot \left[ \mathbf{K} \cdot \left( \nabla \mathbf{h} + \frac{\rho}{\rho_0} \nabla \mathbf{z} \right) \right] + \frac{\rho^*}{\rho_0} \mathbf{q}$$
(1)

K is the Hydraulic conductivity tensor (L/T) and F is the generalized storage coefficient (1/L) defined as:

$$F = \alpha' \frac{\theta}{n_e} + \beta' \theta + n_e \frac{ds}{dh}$$
(2)  

$$K = \frac{\rho g}{\mu} k = \frac{(\rho/\rho_0)}{(\mu/\mu_0)} \frac{\rho_0 g}{\mu_0} k_s k_r = \frac{(\rho/\rho_0)}{(\mu/\mu_0)} K_{so} k_r$$
(3)

Where:  $\theta$ : effective moisture content (L<sup>3</sup>/L<sup>3</sup>); h: pressure head (L); t: time (T); z: potential head (L); q: source or sink of fluid [(L<sup>3</sup>/L<sup>3</sup>)/T];  $\rho_0$  : fluid density without biochemical concentration (M/L<sup>3</sup>);  $\rho$  : fluid density with dissolved biochemical concentration (M/  $L^3);$ 

 $\rho^*$ : fluid density of either injection (=  $\rho^*$ ) or withdraw (=  $\rho$ ) (M/L3);

 $\mu_0$ : fluid dynamic viscosity at zero biogeochemical concentration [(M/L)/T];

 $\mu$ : the fluid dynamic viscosity with dissolved biogeochemical concentrations [(M/L)/T];

 $\alpha$ : modified compressibility of the soil matrix (1/L);

 $\beta$ : modified compressibility of the liquid (1/L);

 $n_e$ : effective porosity (L<sup>3</sup>/L<sup>3</sup>);

S: degree of effective saturation of water;

G: is the gravity  $(L/T^2)$ ;

k: permeability tensor (L<sup>2</sup>);

 $k_s$ : saturated permeability tensor (L<sup>2</sup>);

 $K_{so}$ : referenced saturated hydraulic conductivity tensor (L/T);

k<sub>r</sub>: relative permeability or relative hydraulic conductivity (dimensionless);

The general transport equation using advection, dispersion/diffusion, source/sink, and biogeochemical reaction as the major transport processes can be written as follows:

# $$\begin{split} & \frac{D}{Dt} \int_{V} \ \theta C_{i} dv = - \int_{\Gamma} \ n. \left(\theta C_{i}\right) V_{i} d\Gamma - \int_{\Gamma} \ n. J_{i} d\Gamma + \\ & \int_{V} \ \theta r_{i} dv + \int_{V} \ M_{i} dv, i \in M \\ & (4) \end{split}$$

 $C_i$ : the concentration of the i-th species in mole per unit fluid volume (M/L<sup>3</sup>);

v: the material volume containing constant amount of media  $(L^3)$ ;

 $\Gamma$ : the surface enclosing the material volume v (L<sup>2</sup>);

n:the outward unit vector normal to the surface;

 $J_i$ : the surface flux of the i-th species due to dispersion and diffusion with respect to relative fluid velocity [(M/T)/L<sup>2</sup>];

 $\theta r_i$ : the production rate of the i-th species per unit medium volume due to all biogeochemical reactions [(M/L<sup>3</sup>)/T];

 $M_i$ : the external source/sink rate of the i-th species per unit medium volume [(M/L<sup>3</sup>)/T];

M: the number of biogeochemical species;

 $V_i$ : the transporting velocity relative to the solid of the i-th biogeochemical species (L/T).

#### 2.2 Surrogate Model

Genetic Programming (GP) models are used in this study to develop surrogate models to approximately simulate flow and transport processes. Trained GP models are developed using the simulated response of the aquifer to randomly generated source fluxes. The selected GP models can replace the numerical simulation model to obtain concentration of contaminants in observation wells. GP, a branch of genetic algorithms, is an evolutionary algorithm-based methodology inspired by biological evolution to find computer programs that perform a user-defined task. Essentially, GP is a set of instructions and a fitness function to measure how well a computer model has performed a task.

GP utilizes a set of input-output data which are generated randomly by simulation model. The numerical Simulation model creates M number of out-put sets from M number of input sets, which is generated by using random Latin hypercube sampling in a defined range. The performance of each GP program is evaluated in terms of training, testing, and validation using the set of input-output patterns. The testing data evaluates the model performance for new data without developing a new fitness function. To compare the GP and HGCH results at the same location, the normalized error is used as defined by the following equation:

$$f = \sum_{k=1}^{nt} \sum_{iob=1}^{nob} \frac{ABS(Chgch_{iob}^k - Cgp_{iob}^k)}{Chgch_{iob}^k}$$
(5)

ABS: the absolute value of

**Chgch** $_{iob}^{k}$ : the concentration simulated by the HGCCH model at observation monitoring location iob and at the end of time period k (ML<sup>-3</sup>).

 $Cgp_{iob}^{k}$ : the concentration estimated by the GP models at observation monitoring location iob and at the end of time period k (ML<sup>-3</sup>).

nt: the total number of monitoring time steps. nob: the total number of observation wells.

#### 2.3 Performance Evaluation of Developed Methodology

To evaluate the performance of the proposed surrogate models and compare it with actual simulation model solution results, a hypothetical homogeneous and isotropic aquifer is utilized as an illustrative example as shown in Fig. 1. The grey area represent the contaminant sources S(i) which include distributed and point source. The monitoring networks are shown in Fig 2. Cells marked with green circle, wells set 1, are the grid locations containing a monitoring well which their data are used to train, test and validate GP models' formulations. Moreover, Cells marked with yellow and blue circle are the grid locations containing a monitoring well with their coordinates within, and beyond range of location coordinates implemented in GP formulations, respectively. Groundwater flow and solute transport process is simulated with hydro-geological parameters as given in Table 1. The synthetic concentration measurement data used for the specified polluted aquifer facilitates evaluation of the developed methodology.



Fig.1 Example definition and flow boundary conditions (Total Head: A=37 m, B=40 m, C=33m, D=30 m; Level (F)=37m)



Fig. 2 Location of concentration measurements In the incorporated scenario, copper  $(Cu^{++})$  and sulfate  $(SO_4^{--})$  are introduced as initial pollutants in sources, which are involved chemical reactions, which are showed in table 2.

# Table 1 Aquifer's properties

Unit	Value
m	100
m	100
m	50
	387
	1432
m/d	10
	0.3
m/d	6
m/d	3
	1
g/lit	0-100
	0
	Unit m m m/d m/d m/d g/lit

Table 2Chemicalreactionsduringthecontaminants' transport

Chemical Reaction Equations	Constant Rate
	(Log k)
$Cu^{++} + H_2O \leftrightarrow Cu(OH)^+ + H^+$	-9.19
$Cu^{++} + 2H_2O \leftrightarrow Cu(OH)_2 + 2H^+$	-16.19
$Cu^{++} + 3H_2O \leftrightarrow Cu(OH)_3^- + 3H^+$	-26.9
$Cu^{++} + SO4^{2-} \leftrightarrow CuSO_4$	2.36

The source activities are specified for eight similar time intervals of 100 days each. The actual pollutant concentration from each of the sources is presumed to be constant over a stress period. The pollutant concentration of copper as well as sulfate in the pit is represented as Cpit(i) and Spit(i) respectively, where i represents the stress period number, and also C(i) and S(i) represent copper and sulfate concentrations in the point sources, respectively at different time steps. An overall of sixteen concentration values for each contaminant are considered as explicit variables in the simulation model. The concentration measurements are simulated for 800 days since the start of the The pollutant concentration simulation. measurement at the observation wells starts at time t=100 days and are measured after every 100 days at all the observation locations till t=800 days. Figure 3 shows the pollutant concentration profile in the study area.

#### Genetic Programming formulation

Copper and sulfate concentration in sources are the two main input data sets which consist of sets of concentration values for each of the 32 values, Cpit(i) and C(i) as well as Spit (i) and S(i) (i=1 to 8), representing two sources, two kind of contaminants, and eight active stress periods. Although 128 GP formulations need to be generate in this scenario, the well coordinates are used as input data for GP models to decrease the number of GP models required. Therefore the numbers of models are reduced to 8 models for copper, as well as eight models for sulfate which are generated based on pollution concentration in sources and locations of data measurement. The corresponding output data consists of the resulting pollutant concentration measurements due to these source fluxes at all the 10 monitoring well at time t1 = 100, t2 = 200, t3 = 300, t4 = 400, t5 = 500, t6 = 600, t7 = 700 and t = 800 days. 1,000 data patterns comprising of inputs and the corresponding outputs are used in the GP models. Out of total data patterns, 40 % is used for training, 30 % for validation, and the remaining 20 % for testing. A Latin Hypercube distribution (MATLAB R2020b) was used for generating the random pollution values ranging between 0 g/lit and 100 g/lit, as the input. The corresponding output data was simulated using HGCH code. Discipulus<sup>TM</sup> (RMLTechnologies, Inc.) is used for training, validation and testing of the GP models.

#### 3. RESULTS AND DISCUSSION

The output data from HGCH are compared with GP model results at three arbitrary monitoring networks. The coordinates of first set of wells in the first network are implemented for GP models creation, and the coordinates of second and third monitoring networks are within and beyond the range of first well's location, respectively. These comparison results are shown in Figs 3, 4 and 5. Each time step is marked on the *x*-axis. Each of the bars corresponds to contaminant concentration in each well, obtained by HGCH and GP models. Fig 3 shows the HGCH and GP results in different time steps for wells set 1. Fig 4 and Fig 5 demonstrate these results for wells set 2 and set 3, respectively.

Figures 3, 4 and 5 show that the results obtained from GP models are very close to the simulated results obtained using a numerical simulation model. Although few wells' coordinates are used as input data for GP models, these models can estimate the concentration for all locations in the aquifer such as wells set 3, which shows acceptable results (Fig 5).

Figure 6 shows the summation of normalized error for each monitoring network in each period of time. As expected, generally GP models provide relatively accurate results for well set 1. However, errors for data set 2 and 3 are also small even though complex contaminant transport process with chemical reaction of species is involved in the evaluated scenario.



Fig. 3 Comparison data in wells set 1



Fig. 4 Comparison data in wells set 2

#### 4. CONCLUSION

The developed methodology based on GP models for approximate simulation of the chemically reactive multiple species transport process appears to result is acceptable approximate representation of the transport process in a contaminated aquifer resembling an abandoned mine site. The developed GP models result in reducing the computational time and complexity, and appear to provide acceptable results. However from the limited results in this study, it cannot be concluded if the surrogate models can replace the simulators in all situations. This method can be applied to real scenarios of contaminated aquifers where especially repeated running of numerical simulation models is required, e.g., in linked simulation-optimization model where computational time is a major constraint. GP based surrogate models can increase efficiency and feasibility of developing optimal management strategies for complex contaminated aquifers such as mine sites.



Fig. 5 Compare data in wells set 3



Fig. 6 Normalized errors for all wells sets

### 5. REFERENCES

- Parkhurst, D.L., Thorstenson, D.C. and Plummet, L.N., PHREEQE - A computer program for geochemical calculations. Water Resour. Invest., 1982. 210: p. 16.
- Herzer, j., Kinzelbach, W.,, Coupling of Transport and Chemical Processes in Numerical Transport Models. Geoderrna, 1989. 44: p. 13.
- 3. Tebes-Stevensa, С., Valocchia, A. J., VanBriesenb Rittmannb, J. М., Β. Е., Multicomponent transport with coupled geochemical and microbiological reactions: model description and example simulations. Journal of Hydrology, 1998. 209: p. 16.
- 4. Prommer, H., Barry, D.A., Davis, G.B., Modelling of physical and reactive processes during biodegradation of a hydrocarbon plume

under transient groundwater flow conditions. Journal of Contaminant Hydrology, 2002a. 59: p. 19.

- Zheng, C., Wang, P.P., MT3DMS: A modular three-dimensional multispecies model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems; Documentation and User's Guide. . Contract Report SERDP-99-1, US Army Engineer Research and Development Center, Vicksburg,MS., 1999.
- Prommer, H., PHT3D—A reactive multicomponent transport model for saturated porous media. . Version 1.0 User's Manual, Technical report, Contaminated Land Assessment and Remediation Research Centre, The University of Edinburgh., 2002b.
- Parkhurst, D.L., Appelo, C.A.J., User's guide to PHREEQC—A computer program for speciation, reac-tion-path, 1D-transport, and inverse geochemical calculations. . Technical Report 99-4259, US Geol. Survey Water-Resources Investigations Report., 1999.
- Waddill, D.W., Widdowson, M.A., SEAM3D: A numerical model for three-dimensional solute transport and sequential electron acceptor-based bioremediation in groundwater, . Technical report, Virginia Tech., Blacksburg, Virginia., 1998.
- Parkhurst, D.L., Kipp, K.L., Engesgaard, P., Charlton, S.R., PHAST e A Program for Simulating Ground-water Flow, Solute transport, and Multicomponent Geochemical Reactions. . U.S. Geological Survey, Denver, Colorado., 2004.
- Mao, X., Prommer, H., Barry, D.A., Langevin, C.D., Panteleit, B., Li, L., Threedimensional model for multi-component reactive transport with variable density groundwater flow. Environmental Modelling & Software, 2006. 21 ( (5)): p. 14.
- 11. Yeh, G.T., Tripathi, V. S, HYDROGEOCHEM: A coupled model of hydrologic transport and geochemical multicomponent equilibria in reactive systems. Environmental Science Division Publication, 1991. No. 3170, Oak Ridge National Laboratory, Oak Ridge, TN, .
- 12. Sun, J., A Three-Dimensional Model of Fluid Flow, Thermal Transport, and Hydrogeochemical Transport through Variably Saturated Conditions. . M. S. Thesis. Department of Civil and Environmental Engineering, University of Central Florida, Orlando, FL 32816, 2004.
- 13. Kalin, M., Fyson, A., Wheeler, W. N., Review The chemistry of conventional and alternative treatment systems for the neutralization of acid mine drainageScience of the Total Environment 2006. 366: p. 14.

- 14. Nordstrom, D.K.a.A., C.N., Geochemistry of acid mine waters, in Plumlee, G.S., and Logsdon, M.J., eds., The environmental geochemistry of mineral deposits, Part A: Processes, techniques, and health issues, Reviews Economic Geology,, 1999. 6A: p. 28.
- Zhou, X., Chen, M., Liang, C., Optimal schemes of groundwater exploitation for prevention of seawater intrusion in the Leizhou Peninsula in southern China. Environmental Geology, 2003. 43: p. 8.
- Abarca, E., Vazquez-Sune, E., Carrera, J., Capino, B., Gamez, D., Battle, F., Optimal design of measures to correct seawater intrusion. Water Resource Research., 2006. 42.
- Ranjithan, S., Eheart, J. W., and Garrett, J. H., Neural network based screening for groundwater reclamation under uncertainty. Water Resour. Res., 1993. 29(3): p. 12.
- Koza, J.R., Genetic programming as a means for programming computers by natural-selection. Statistics and computing, 1994. 4: p. 26.
- 19. Bhattacharjya, R., and Datta, B.,, Optimal management of coastal aquifers using linked simulation optimization approach. Water Resour.Manage., 2005. 19(3): p. 25.
- 20. Bhattacharjya, R., K., Datta, B., and Satish, M.,, Artificial neural networks approximation of density dependent saltwater intrusion process in coastal aquifers. Journal of Hydrologic Engineering, 2007. 12(3): p. 10.
- 21. Bhattacharjya, R.K., and Datta, B.,, ANN-GAbased model for multiple objective management of coastal aquifers. Water Resour. Planning Manage., 2009. 135(5): p. 8.
- Dhar, A., Datta, B., Saltwater Intrusion Management of Coastal Aquifers. I: Linked Simulation-Optimization. Journal Hydrology Engineering, 2009. 14: p. 9.

- 23. Sreekanth, J., Datta, B.,, Multi-objective management models for optimal and sustainable use of coastal aquifers. Journal of Hydrology, 2010. 393: p. 11.
- 24. Sreekanth, J., Datta, B.,, Comparative Evaluation of Genetic Programming and Neural Network as Potential Surrogate Models for Coastal AquiferManagement Water Resour Manage 2011a(25): p. 18.
- 25. Datta, B., Prakash, O., Sreekanth, J, Application of genetic programming models incorporated in optimization models for contaminated groundwater systems management. EVOLVE -A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V Advances in Intelligent Systems and Computing, 2014. 288: p. 16.
- 26. Datta, B., Prakash, O., Campbell, S., Escalada, G.,, Efficient Identification of Unknown Groundwater Pollution Sources Using Linked Simulation-Optimization Incorporating Monitoring Location Impact Factor and Frequency Factor. Water Resour Manage, 2013. 27: p. 18.

Int. J. of GEOMATE, March, 2015, Vol. 8, No. 1 (Sl. No. 15), pp. 1190-1196.

MS No. 4263 received on June 15, 2014 and reviewed under GEOMATE publication policies. Copyright © 2015, International Journal of GEOMATE. All rights reserved, including the making of copies unless permission is obtained from the copyright proprietors. Pertinent discussion including authors' closure, if any, will be published in March 2016 if the discussion is received by Sept. 2015.

Corresponding Author: Hamed Esfahani