A knowledge-based reactive transport approach for the simulation of biogeochemical dynamics in Earth systems

D. R. Aguilera, P. Jourabchi, C. Spiteri, and P. Regnier

Department of Earth Sciences—Geochemistry, Faculty of Geosciences, University of Utrecht, P.O. Box 80.021, NL-3508 TA Utrecht, Netherlands (aguilera@geo.uu.nl)

[1] A Knowledge-Based Reactive Transport Model (KB-RTM) for simulation of coupled transport and biogeochemical transformations in surface and subsurface flow environments is presented (http://www.geo.uu.nl/~kbrtm). The scalable Web-distributed Knowledge Base (KB), which combines Information Technology (IT), an automatic code generator, and database management, facilitates the automated construction of complex reaction networks from comprehensive information stored at the level of biogeochemical processes. The reaction-centric approach of the KB-RTM system offers full flexibility in the choice of model components and biogeochemical reactions. The procedure coupling the reaction networks to a generalized transport module into RTMs is also presented. The workings of our KB-RTM simulation environment are illustrated by means of two examples of redox and acid-base chemistry in a typical shelf sediment and an aquifer contaminated by landfill plumes.

Components: 9832 words, 9 figures, 5 tables.

Keywords: biogeochemistry; information technology; Knowledge Base; numerical methods; reactive transport.

Index Terms: 1009 Geochemistry: Geochemical modeling (3610, 8410).

Received 16 December 2004; Revised 10 March 2005; Accepted 4 May 2005; Published 27 July 2005.


1. Introduction

[2] Reactive transport models (RTMs) are powerful tools for capturing the dynamic interplay between fluid flow, constituent transport, and biogeochemical transformations [Steefel and Van Cappellen, 1998]. They have been used to simulate, among others, rock weathering and soil formation [e.g., Ayora et al., 1998; Soler and Lasaga, 1998; Steefel and Lichtner, 1998a, 1998b; Thyne et al., 2001; Soler, 2003; De Windt et al., 2004], nutrient dynamics in river drainage basins and estuaries [e.g., Soetaert and Herman, 1995; Billen et al., 1994; Regnier et al., 1997; Regnier and Steefel, 1999; Vanderborght et al., 2002], reactive transport in groundwater, like contamination of aquifers [e.g., Engesgaard and Traberg, 1996; Brown et al., 1998; Hunter et al., 1998; Xu et al., 1999; Murphy and Ginn, 2000; Barry et al., 2002; Brun and Engesgaard, 2002; Thullner et al., 2004; van Breukelen et al., 2004], early diagenetic transformations in sediments [e.g., Soetaert et al., 1996; Boudreau, 1996; Van Cappellen and Wang, 1996; Dhakar and Burdige, 1996; Berg et al., 2003; Jourabchi et al., 2005], benthic-pelagic coupling in ocean systems [e.g., Soetaert et al., 2000; Archer et al., 2002; Lee et al., 2002] and hydrocarbon migration and maturation in sedimentary basins [e.g., Person and Garven, 1994]. By integrating experimental, observational and theoretical knowledge about geochemical, biological and transport processes into mathematical formulations, RTMs provide the grounds for prognosis, while diagnostic comparison between model simulations and measurements can help identify gaps in the conceptual understanding of a specific system or uncertainties
in proper parameterization of biogeochemical processes [Berg et al., 2003; Jourabchi et al., 2005].

RTMs have traditionally been developed and used to investigate the fate and transport of a selected set of chemical constituents within a given compartment of the Earth system (e.g., the early diagenetic models by Soetaert et al. [1996], Boudreau [1996], Van Cappellen and Wang [1996], Dhakar and Burdige [1996], and references cited above). As a result, they have tended to be environment and application specific with regards to the flow regime and the biogeochemical reaction network.

Although the first attempts to develop interactive software systems for automatic solution of models based on ordinary and partial differential equations date back to the creation of digital computers [e.g., Young and Juncosa, 1959; Lawrence and Groner, 1973; Mikhailov and Aladjem, 1981, and references therein], such developments have so far received little attention in the field of reactive-transport modeling. Literature review shows that RTM codes allowing for more flexible definition of state variables and processes without requiring in-depth knowledge of programming or numerical solution techniques have been developed over the last decade [e.g., Reichert, 1994; Chilakapati, 1995; Regnier et al., 1997; Chilakapati et al., 2000; MacQuarrie et al., 2001; Meysman et al., 2003a, 2003b; Van der Lee et al., 2003]. Database tools, such as that developed by Katsev et al. [2004], have also been presented recently to the reactive transport community. Model flexibility is a critical feature since a major challenge in the field of reactive transport modeling is the realistic representation of the highly complex reaction networks (RN) that characterize the biogeochemical dynamics of natural environments [e.g., Mayer et al., 2002; Berg et al., 2003; Quezada et al., 2004]. At the same time, many field- and laboratory-based experiments are also being conducted to identify novel reaction pathways, quantify reaction rates and microbial activity levels, describe ecological community structures, and elucidate interactions between biotic and abiotic processes. This rapidly growing knowledge about biogeochemical transformation processes creates a need for efficient means of transferring new experimental findings into RTMs.

Here, a unified modeling approach for implementing complex reaction networks in RTMs is presented. Our simulation environment is based on a modular approach to facilitate incorporation of new theoretical and experimental information on the rates and pathways of biogeochemical reactions. The key novel feature of the modeling environment is a Web-distributed Knowledge Base (KB) of biogeochemical processes, which acts as the evolving repository of up-to-date information gained in the field of geochemistry. The implementation of such a library within a simulation environment is a major step toward the model’s flexibility, because it is at the level of an easily accessible open resource, the KB, that process-based theoretical and experimental advances are incorporated in the modeling process. Model generation is conducted via a graphical user interface (GUI) on a Web-based “runtime” server, which allows for the development of biogeochemical reaction network modules. That is, information stored at the level of individual biogeochemical processes can be combined into mathematical expressions defining completely the (bio)chemical dynamics of the system. Since the reaction network is assembled from information stored at the process level, almost any conceivable combination of mixed kinetic and equilibrium reactions can be implemented in our model architecture. The selected RN can easily be merged with existing transport models, hence creating a flexible framework in which to assemble RTMs. The proposed approach allows the RTM community to test and compare, in close collaboration with experimentalists, alternative mathematical descriptions of coupled biogeochemical reaction networks. For example, increasingly detailed representations of biogeochemical processes can be incorporated in the Knowledge Base. Reaction network modules of increasing complexity may then be assembled and coupled to surface or subsurface flow models, in order to determine which level of biogeochemical complexity is adequate to simulate chemical system dynamics at variable spatial and temporal resolutions. By taking a “reaction-centric” approach which utilizes the unifying conceptual and mathematical principles underlying all RTMs, one-dimensional (1D) transport descriptions relevant to many compartments of the Earth system (rivers, estuaries, groundwater or sediments) can be incorporated in our simulation environment. The proposed approach should thus help overcome traditional disciplinary barriers between the different subfields of RTMs.

The paper is structured as follows: First, the mass conservation equation describing 1D coupled transport and reaction is briefly presented. A generalized continuum representation is proposed,
which allows for the simulation of reactive-transport problems characterized by different flow regimes and dispersion intensities. A brief description on how an existing Automatic Code Generator (ACG) based on symbolic programming [Regnier et al., 2002] can be used to create the model specific code necessary to the numerical solution of the governing equations is then given. We demonstrate how our Web-distributed Knowledge Base concept, which combines Information Technology with symbolic computing techniques, directs the mathematical formulation of the biogeochemical reaction network and leads to a modeling environment offering full flexibility. Finally, the workings of our KB-RTM are illustrated with two contrasting examples of complex redox and acid-base geochemistry in an aquatic sediment and an aquifer, respectively.

### 2. Mathematical Representation of Reactive-Transport Equations

A one-dimensional continuum representation of coupled mass transport and chemical reactions in Earth systems can be described mathematically by a set of partial differential equations (PDEs) in time and space of the form

\[
\xi \frac{\partial C_j}{\partial t} = \left[ \frac{\partial}{\partial x} \left( D \cdot \xi \cdot \frac{\partial C_j}{\partial x} \right) - \frac{\partial}{\partial x} (v \cdot \xi \cdot C_j) \right]_j + \sum_{k=1}^{N_r} \lambda_{k,j} \cdot \sigma_k: \quad j = 1, \ldots, m, \tag{1}
\]

where \( t \) is time and \( x \) denotes the position along the 1D spatial domain. Particular solutions of equation (1) require specification of initial and boundary conditions. Further discussion about the generalized continuum representation of the advection-dispersion-reaction (ADR) equation is given by Regnier et al. [2002] and Meile [2003]. The first two terms on the right hand side, in square brackets, compose the transport operator \( T \); the last one \( R \) represents the sum of transformation processes (e.g., reactions) affecting a species \( j \) of concentration \( C_j \). Table 1 shows that \( \xi, D \) and \( v \) are generic variables which take different meanings depending on the environment considered. \( \sigma_k \) represents, for kinetic reactions, the rate of the \( k \)-th reaction and \( \lambda_{k,j} \) is the stoichiometric coefficient of species \( j \) in that reaction. Currently, it is assumed that the reaction processes \( R \) have no effect on the physical or transport \( T \) properties of the system (i.e., \( \xi, D \) and \( v \) are unaffected by reactions). The rate \( \sigma_k \) is of arbitrary form, even nonlinear, and can be a function of several concentrations of the system. Through this coupling by the reaction terms, most multicomponent problems result in a set of coupled nonlinear partial differential equations (PDEs) of size \( m \), number of species of the reaction network. In the event that some of the reactions considered are assumed to be at equilibrium, algebraic expressions based on mass action laws are introduced into the system of equations to be solved [Regnier et al., 2002]. By replacing one or more of the \( m \) differential equations associated with reactions with algebraic relations based on a mass action expression in the local equilibrium case, the set of ODEs is transformed into a set of differential-algebraic equations (DAEs) [Chilakapati, 1995; Hindmarsh and Petzold, 1995a, 1995b; Brenan et al., 1989]. This transformation leads to a system of \( m \) equations to solve, including \( m_k \) equations associated with kinetic reactions, and \( m_e \) algebraic equations based on mass action expressions.

The numerical solution of the set of discretized PDEs and DAEs \((\partial t \rightarrow \Delta t, \partial x \rightarrow \Delta x)\) commonly requires the use of implicit methods in order to be computationally efficient [Steefel and MacQuarrie, 1996]. We currently make use of the time splitting technique, which consists of solving first the transport and then the reaction terms in sequence for a single time step. This method is referred to as the

| Table 1. Meaning of the Generalized Variables \( \xi, D, \) and \( v \) for Different Environmentsa |
|---------------------------------|---------------------------------|---------------------------------|
| Surface Flow                   | Aquatic Sediment                | Groundwater Flow Path          |
| Solutes or Suspended Solids    | Solutes                         | Solids                          |
| Solutes                        | Solutes                         | Solutes                        |
| \( \xi \)                      | \( A \)                         | \( 1 - \phi \)                  |
| \( v \)                        | \( V_{flow} \)                  | \( \omega \)                    |
| \( D \)                        | \( K_{turb} \)                  | \( 0 \)                         |
| \( \sigma_k \) \( k \)         | \( D_b \)                       | \( V_{flow} \)                  |
| \( \lambda_{k,j} \) \( j \)    | \( D_h + D_{sed} \)             | \( 0 \)                         |
| \( \lambda_{k,j} \) \( j \)    | \( D_{disp} \)                  |                                |

[a]From Meile [2003]. In porous media flow, distinction between an aqueous and a solid phase must be considered, and \( C \) is either a concentration of solute or solid. \( A \) \([L^2]\), cross-section of the area of the flow channel; \( \phi \) \([-]\), porosity; \( V_{flow} \) \([LT^{-1}]\), externally imposed flow velocity; \( \omega \) \([LT^{-1}]\), burial velocity defined with respect to the sediment-water interface (SWI); \( v_{flow} \) \([LT^{-1}]\), flow velocity acting only on solutes externally imposed or from porosity change; usually defined with respect to the SWI [e.g., Boudreau, 1997]; \( K_{turb} \) \([L^2T^{-1}]\), longitudinal turbulent dispersion coefficient; \( D_b \) \([L^2T^{-1}]\), bioturbation coefficient; \( D_{sed} \) \([L^2T^{-1}]\) = \( D_{mol} (1 - \ln(\phi^2)) \), tortuosity corrected molecular diffusion coefficient for solutes at in situ temperature and salinity [Boudreau, 1997]; \( D_{disp} \) \([L^2T^{-1}]\) = \( \sigma_k \cdot V_{flow} \), longitudinal dispersion, where \( \sigma_k \) \([L]\) is the longitudinal dispersivity [Freeze and Cherry, 1979].
sequential noniterative approach (SNIA). When solving the reaction part an iterative method is required to numerically find the roots of the function residuals, $f_j$, which correspond to mass balance equations and, if equilibrium reactions are included, mass action equations. This is because the reaction terms can be nonlinear functions of the species concentrations. By far the most common approach for finding the root of nonlinear sets of equations is the Newton-Raphson method [Press et al., 1992]. This method involves the use of a first degree Taylor series expansion to linearize the problem for every single iteration step. The function residuals, $f_j$, representing the reaction network (RN), and the Jacobian matrix, which contains the partial derivatives of the function residuals with respect to the unknown concentrations, are the most important pieces of information required to implement the Newton-Raphson algorithm [e.g., Dennis and Schnabel, 1988]. Once linearized, the resulting problem is solved using linear algebra methods, such as the LU decomposition [e.g., Strang, 1988]. A precise definition of the function residuals and the Jacobian matrix is given by Regnier et al. [2002].

3. Reactive Transport Modeling With the Knowledge Base

[5] Inspection of the transport and reaction operators $T$ and $R$ in equation (1) shows that the following information is required to define a specific reaction transport application:

[10] 1. Domain definition: spatiotemporal size and resolution ($x_{tot}$, $\Delta x$, $t_{tot}$, $\Delta t$), where $x_{tot}$ and $t_{tot}$ stand for the total domain length and simulation time, respectively, and $\Delta x$ and $\Delta t$ are the space and time step used for numerical integration, respectively.

[11] 2. Transport coefficients $\xi$, $v$ and $D$ (Table 1).


[13] 4. Boundary (BC) and initial (IC) conditions for every species of the RN.

[14] In a modeling environment offering full flexibility, this information is specific to each RTM application and needs to be automatically translated into source code. This task is relatively easy for the physical domain definition, transport parameters, BC and IC. However, if flexibility in choosing process formulations is also important, then the stiff system of differential equations using a linearization method (such as the above-cited Newton-Raphson algorithm) necessitate automatic differentiation schemes for the calculation of the terms in the Jacobian matrix. Automatic symbolic differentiation offers the advantage of producing derivatives of potentially complicated functions which are accurate up to the precision of the programming language used (e.g., FORTRAN), plus the convenience of updating the derivatives easily if the original functions are changed [Steefel and MacQuarrie, 1996]. Automated differentiation for stiff sets of differential equations is one of the key features of our Automatic Code Generation (ACG) procedure [Regnier et al., 2002].

[15] The simultaneous implementation of a library of biogeochemical processes into a Knowledge Base (KB) is an additional crucial component of the proposed simulation environment. The KB makes it possible to take full advantage of the ACG. The integration of these two components within a Web system shows how the combination of Information Technology with advanced symbolic programming allows the use of the Internet as a software provider in the area of Reactive-Transport modeling (Figure 1).

3.1. The Internet as a Software Provider

[16] An Internet system developed in PHP language (http://www.php.net) provides the adaptive JavaScript and HTML code for the Graphical User Interface (GUI), as well as the Web-based “runtime” server to our modeling environment. It is accessible at http://www.geo.uu.nl/~kbrtm. The GUI is of evolutionary nature, that is, it dynamically adapts to changes in structure and content of the KB system as well as to the selections made by the user.

[17] Initially, the user accesses an interface for the KB system in the style of a Web form for selection of desired biogeochemical processes (Figure 1, step 1; further detailed in Figure 2). Species-independent physical parameters (definition of spatiotemporal domain and most transport coefficients in Table 1) are also defined at this stage.

[18] Figure 2 gives a detailed description of the model design procedure. The KB system consists of a set of biogeochemical processes, containing default formulations for reactions, which are available to all users of the system (common KB) and which cannot be modified. However, these common processes can be edited if desired, or new...
processes created using a standard template, and stored in a “private” KB library which is inaccessible to other users. To facilitate model development, processes can be grouped in different subsets, for instance, in terms of reaction types (see below).

[19] The selected processes specify the user-defined reaction network. The latter is then analyzed to determine the list of chemical species involved in the RN. A second Web submission form is subsequently created to specify all species-dependent parameters (such as molecular diffusion coeffi-

---

**Figure 1.** Structure of the Knowledge-Based (KB) Reactive Transport Model (RTM) environment which uses the Web as a software provider: 1, form submission; 2, parsing of the ASCII files into MAPLE; 3, MAPLE/MACROFORT symbolic computing; 4, translation of MAPLE results into FORTRAN; 5, linking and compilation of the FORTRAN code; 6, transfer of the executable file to the user via e-mail.

**Figure 2.** KB-Internet model design process (detailed description of step 1 in Figure 1): (1) Process edition or creation. Templates can be used to speed up the implementation of new processes. Modified or new process files are stored in a “private” KB which is accessible only to a single user. (2) Process selection from the common and private KB process pools, and specification of the physical support. (3) Analysis of the selection and creation of a dynamically adaptive Web form to input reaction network dependent parameters. (4) Submission of complete model information (processes selected, physical parameters, and species-dependent parameters) to the ACG via our Web server.
3.2. Knowledge Base (KB)

[20] The common KB (Figure 2) contains subsets of biogeochemical process libraries and parameter values, including default formulations for biogeochemical reaction rates and equilibria. The design provides an expandable structure permitting the easy addition of new elements (processes, variables, parameters) into a “private” KB library. Several formulations can be implemented for the same biogeochemical process, simply by naming the respective process files differently. This permits a ready-made set of specialized process descriptions for a number of different scenarios, and therefore the formulation that best fits the simulation requirements can be selected.

[21] The biogeochemical processes currently implemented are distributed among the following classes, which form the architecture of our KB: primary redox reactions (PRR) involving organic matter oxidation pathways, secondary redox processes (SRR, biotic and abiotic redox reactions), aqueous complex formation (ACF), homogeneous acid-base equilibrium reactions (ABE), nonredox mineral precipitation and dissolution (MPD), adsorption and ion exchange (AIE) and gas-water reactions (GWR). These classes represent only one of the many possible categorizations. In fact, model specific categories can be added to the system. According to different criteria, one process could be assigned not solely to one, but potentially to several classes.

[22] The file structure for each process consists of the set of species (dissolved or solid) involved in the specific process, the definition of the reaction type (kinetic or equilibrium) and the stoichiometry. The corresponding rate or mass action law and default parameter values are also specified. Fields of informative nature (e.g., reaction class, keywords for searching, documentation, units for parameters and concentrations) complete the process file description. As an example, Figure 3 shows the process file for aerobic degradation of organic matter as an example. Finally, the set of variables and parameters for the entire selection of process files, along with the respective functions characterizing kinetic or equilibrium processes and their corresponding stoichiometries, fully determine the RN.

3.3. Parsing Procedure and Automatic Code Generation

[23] After the reactions of interest have been selected and their list submitted in addition to the remaining application-specific data (step 1 in Figure 1), an automated procedure is launched that parses all this information into ASCII files. This completes the model setup procedure carried out by the user on the GUI of the Web-distributed KB. It is illustrated in detail by means of examples in the next section.

[24] The parsing procedure involves reading information from files directly generated by the Internet system (e.g., definition of the physical domain) and from files originated by the KB system (reaction network). The former can easily be processed by the Automatic Code Generator (ACG), whereas the information contained in the KB files must first be concatenated in the appropriate format. The main features of the parsing involves extracting and assembling, from the set of selected processes, a number of lists containing all the species involved in the RN, along with appropriate initial and boundary conditions, the complete set of kinetic expressions and equilibrium constraints necessary to define the function residuals \( f_j \), and all the parameters appearing in \( f_j \) including their associated numerical values. Stoichiometric coefficients, which relate the changes in species concentrations to the progress in any individual reaction, is another crucial component extracted from the KB process files. A check for unit consistency for both species concentrations and parameters values is carried out at this stage.

[25] Once parsing is completed, the following operations are performed in background on our Web-based “runtime” server (Figure 1, steps 2–5). First, the information from the ASCII-files is used to assemble the stoichiometric matrix, that is, the mathematical representation of the user-defined reaction network (Figure 1, step 2). Second, a set of symbolic programming operations is performed (Figure 1, step 3) using the MAPLE software environment [Chilakapati, 1995; Regnier et al., 1997; Amberg et al., 1999; Regnier et al., 2002]. The most important one involves the construction of the Jacobian matrix from the series of function residuals. Third, the function residuals, the Jacobian matrix, plus all user-dependent information are translated into fully structured FORTRAN code.
using the MACROFORT package [Gómez, 1990] combined with our own library of MAPLE procedure (ACGLIB©). The resulting routines are then linked with our numerical engine, which contains standard routines for solving transport equations and a linear algebra solver for the sets of linearized process equations generated by the preprocessor (Figure 1, step 5). Finally, upon completion of the automated generation procedure, the “runtime” server compiles and sends the resulting model executable to the user via e-mail (Figure 1, step 6).

[26] The accuracy of the numerical engine and the ACG (referred to as Biogeochemical Reaction Network Simulator (BRNS) by Regnier et al. [2003]) has recently been evaluated by Thullner et al. [2005], from a comparison between their model results with those obtained with two well established RTMs for porous media flow applications (e.g., STEADYSED [Van Cappellen and Wang, 1996; Hunter et al., 1998] and TBC [Schafer et al., 1998a, 1998b]). Comparison with field data from Skagerrak sediments [Canfield et al., 1993] and a sand column experiment [von Gunten and Zobrist, 1993] was also performed by these authors. Here, the performance of the KB-RTM is evaluated further by means of two contrasting scenarios. However, the focus is mainly on the Web-distributed KB system and how this concept...
facilitates the incorporation and modification of complex biogeochemical reaction networks into RTM applications.

4. Applications

4.1. Introduction

Two applications of redox dynamics in porous media are presented: an early diagenetic (ED) model (saved in our server system as “EDscenario”) of shelf sediments and two simulations of a groundwater aquifer (GWA) contaminated by a landfill leachate plume (saved respectively as “GWscenario1” and “GWscenario2”). All model applications can be found on our Web-based “runtime” server as part of the KB, at http://www.geo.uu.nl/~kbrtm. In what follows, the discussion addresses only the major features, that is, similarities and differences between the applications. The user is referred to the Web site for exact parameterization.

In short, the focus is on redox processes, acid-base chemistry and carbonate/sulfide mineral phases precipitation-dissolution. Since the interplay of these processes determines the pH in both systems, which itself is sensitive to relative imbalances in reaction rates [Jourabchi et al., 2005, and references therein], computation of realistic pH profiles with mechanistic process models represents a challenging and important test application.

The structure of the reaction network (RN) is similar for the sediments and groundwater applications, with organic matter serving as the primary electron donor. The RN combines 23 species and 24 reactions for the ED scenario, and 25 species and 36 reactions for the GWA application. This includes kinetic transformations (all primary and secondary redox reactions and nonredox mineral precipitation-dissolution processes) and equilibrium acid-base reactions (see below). Almost 50 parameters are used in the definitions of rates and equilibrium constraints. Even though the set of reactions is similar in both applications, reaction parameter values may contrast. For example, the quality of the organic matter may differ, leading to faster degradation rates near the contaminant site, or differences in ionic strengths affect the apparent solubility constant in the marine setting. Note also that different units are used in both applications, with solute and solid concentration respectively expressed in mol/L_pore water and mol/g_solid in the ED application, and concentrations in moles/L_pore water for the solutes, and moles/L_total volume for the solids in the GWA examples. We will adopt the units used in the ED simulation in what follows. The reader should refer to the GWA scenario files in our server (http://www.geo.uu.nl/~kbrtm) for more details about the unit system selected. The respective RNs are complex and involve many parameters and rate formulations, yet they can be easily assembled through the KB and the Web interface. The system of DAEs is stiff in both environments, which presents a rigorous test for the proposed modeling approach.

Time and space scales of the ED and GWA simulations are fundamentally different (dm and millennia versus km and decades, respectively; Table 2). In addition, the mechanisms and relative intensities of transport processes are not the same, with dominant diffusional transport by bioturbation and molecular diffusion in the ED and dominant advective transport in the GWA. Another major difference between the two applications is the treatment of solid species in the model: in the GWA, a moving fluid percolates through a fixed solid matrix, while in the ED, solids are mixed by bioturbation.

4.1.1. Groundwater Aquifer

In the GWA application, the focus is on the biogeochemical evolution of the major constituents in a landfill leachate plume infiltrating an initially oxic, pristine aquifer of low porosity (\(\phi = 0.25\)). Simulations are carried out over a 400 m flow path

| Table 2. Web Submission Form I (Physical Parameters): Definition of the Physical Domain, Major Transport Coefficients and Forcing Functions as Well as Global Output Parameters for the Early Diagenetic and Groundwater Aquifer Applications, Respectively^a |
|---------------------------------|---------------|---------------|
| **ED**                          | **GWA**       |
| Total time                      | 1500 yr       | 22 yr         |
| Time step                       | \(1.0 \times 10^{-4}\) yr | \(5.0 \times 10^{-4}\) yr |
| Total length                    | 40 cm         | 40000 cm      |
| Number of nodes                 | 401           | 101           |
| Porosity                        | 0.85          | 0.25          |
| Cross section                   | 1.0 cm\(^2\)  | 1.0 cm\(^2\)  |
| Flow velocity                   | 0.0 cm/yr     | 250.0 cm/yr   |
| Burial velocity                 | 0.04 cm/yr    | 0.0 cm/yr     |
| Bioturbation coefficient        | 3.0 cm\(^2\)/yr | 0.0 cm\(^2\)/yr |
| Longitudinal dispersivity       | 0.0 cm        | 400 cm        |
| Temperature                     | 10\(^\circ\)C | 15\(^\circ\)C |
| Salinity                        | 35.0 PSU      | 0.1 PSU       |
| First output time               | 1.0 yr        | 1.0 yr        |
| Output interval period          | 2.0 yr        | 1.0 yr        |

^a ED, early diagenetic; GWA, groundwater aquifer. All entries are species independent and hence are not influenced by the selected RN. For each selected kinetic process, the Web interface prompts the user whether a spatial distribution of rates should be generated at output times.
which is infiltrated at the upstream boundary by the anoxic landfill-leachate recharge. In this case, the model setup (see the scenario files in our Web system) is based on Hunter et al. [1998] and Van Cappellen and Wang [1996]. On the inflow side of the domain (x = 0 cm), a specified concentration (i.e., Dirichlet boundary condition) is applied for all solute species. The chemical composition of the leachate is characterized by a high dissolved organic carbon (DOC) loading represented by two fractions of different reactivity. At the outflow (xtot = 4 · 104 cm), a concentration gradient (i.e., Neumann boundary condition) is specified for all dissolved chemical species. Solids are considered immobile and therefore are only affected by local biogeochemical transformations. Solutes are transported by fluid flow motion (250 cm · yr⁻¹) and macroscopic dispersion (dispersion coefficient D = 4 · 10³ cm² · yr⁻¹ or dispersivity αL = 400 cm). The overall redox dynamics and the resulting effect on pH evolution are investigated along a flow line after a transient simulation period of 22 years. Table 2 presents the values of the coefficients implemented to define the GWA physical framework via the Web submission form I.

4.1.2. Shelf Sediment

In the ED simulation, transient depth distributions of redox sensitive species and rates are computed for the case of a typical shelf sediment (total depth: 40 cm, T: 10°C, S: 35%) of high porosity (φ = 85%). The simulation period is 1500 years, starting from constant concentration profiles. The model setup is a slightly modified version of the shelf case presented by Van Cappellen and Wang [1996]. The exact definition can be found in our Web system (http://www.geo.uu.nl/~kbrtm). The primary driving force for biogeochemical transformation is the solid flux of organic matter (32 μmol cm⁻² · yr⁻¹) depositing at the sediment-water interface, assumed to be constant in time. The length scale of the simulation (xtot = 40 cm) is chosen such that at x = xtot, all rates are very close to zero, a condition which allows the specification of a Neumann boundary condition. At the sediment-water interface (x = 0 cm), a flux boundary condition is assigned for all solids, while typical bottom-water concentrations (Dirichlet condition) provide the upper boundary condition for solutes. Externally impressed flow is neglected here, and hence the advective velocity (0.04 cm/yr) is solely due to the burial rate. All chemical species are subject to bioturbation, and solutes are further transported by molecular diffusion. Table 2 presents the values of the coefficients implemented to define the ED physical framework via the Web submission form I.

4.2. Reaction Network

The reaction network, which is common to the two simulations, consists of the six major metabolic pathways for organic matter degradation (aerobic degradation, denitrification, Mn and Fe reduction, sulfate reduction and methanogenesis), and a set of 10 secondary reoxidation reactions (Table 3a). In addition to the microbial and chemical reduction of Mn and Fe oxides, precipitation/dissolution of MnCO₃, FeCO₃ and FeS mineral phases are also included, which have an effect on redox and acid-base chemistry. Over 45 parameters (such as kₚᵣᵣᵣᵣ, kₑₐₑ, kₛₛᵣᵣᵣᵣ, kᵰᵰᵰᵰ, kₚₚ, SSA, Kₛₛ, and Kₑₑₑₑ, see below) are extracted from the KB for these simulations.

The shelf sediment application is also forced by a constant flux of calcite (8 μmol cm⁻² · yr⁻¹). Calcium carbonate buffering is ignored in the groundwater application, as it is assumed that the landfill plume percolates through a noncalcareous groundwater aquifer. Larger pH changes are thus expected in this case. Finally, the acid-base chemistry of the ED application includes the dissociation of carbonic, sulfidic and boric acids. Since total boron is negligible in freshwater environments, the dissociation of the latter weak acid is ignored in the GWA simulation.

According to the mathematical functional expressions and the “taxonomy” proposed in our KB, four types of chemical processes are included in our RN (Table 3a). For primary redox reactions (PRR), the rates of the respective metabolic pathways (PRRᵢ) are given by [e.g., Van Cappellen and Gaillard, 1996]

\[
\text{PRRᵢ} = 0 \\
\begin{cases} \\
\left\{ \begin{array}{l}
\text{if} \ [Eₐᵢ] > Kₑₑᵢ then \\
\text{if} \ [Eₐᵢ] \leq Kₑₑᵢ \\
\end{array} \right.
\end{cases}
\]

\[
\begin{align*}
\text{if} \ [Eₐᵢ] > Kₑₑᵢ & \text{ then } PRRᵢ = kₚᵣᵣᵣᵣ \cdot [CH₂O] \cdot fᵢ \cdot \frac{[Eₐᵢ]}{Kₑₑᵢ} \\
\text{if} \ [Eₐᵢ] \leq Kₑₑᵢ & \text{ then } PRRᵢ = kₚᵣᵣᵣᵣ \cdot [CH₂O] \cdot fᵢ \cdot \frac{[Eₐᵢ]}{Kₑₑᵢ}
\end{align*}
\]
The proposed rate law assumes first-order dependency with respect to the electron donor and a pseudo Michaelis-Menten type relationship with respect to the $EA_i$. Equation (2) indicates that a specific metabolic pathway is inhibited if a more favorable $EA_i$ is present in the system at sufficiently high concentrations, that is, if a reaction yielding more free energy takes place [e.g., Van Cappellen and Wang, 1996; Berg et al., 2003].

\[ SRR = k_{SRR} \cdot [EDO] \cdot [EA], \]

where $EDO$ is an inorganic electron donor (e.g., Fe$^{2+}$, Mn$^{2+}$, H$_2$S, ...), and $k_{SRR}$ is the rate constant with units of $L_{\text{porewater}} \cdot \text{mol}^{-1} \cdot \text{yr}^{-1}$ if all reactants are solutes, or $g_{\text{solid}} \cdot \text{mol}^{-1} \cdot \text{yr}^{-1}$ if a solid species is involved in the SRR (e.g., MnO$_2$ or Fe(OH)$_3$).

\[ MPD = k_d \cdot [\text{MIN}] \cdot (1 - IAP/K_{sp})^n \quad \text{if } IAP < K_{sp} \]

\[ MPD = k_p(1 - IAP/K_{sp})^n \quad \text{if } IAP > K_{sp} \]

where $k_d [\text{yr}^{-1}]$ and $k_p [\text{mol} \cdot g_{\text{solid}}^{-1} \cdot \text{yr}^{-1}]$ are the rate constants for dissolution and precipitation, respectively, $[\text{MIN}]$ denotes the concentration of the dissolving mineral $[\text{mol} / g_{\text{solid}}]$, $IAP$ is the Ion Activity Product, $K_{sp}$ the equilibrium constant (or solubility product), and $n$ the order of the reaction. In the groundwater scenario, equilibrium constants corrected for the in-situ temperature ($15^\circ C$) are used, while in the early diagenetic simulations, apparent equilibrium constants which incorporate both temperature ($10^\circ C$) and electrolyte effects of marine solutions have been applied [Boudreau, 1997].

\[ ABE \equiv K_{eq} \cdot [H_m A^{m-}] - [H^+] \cdot [H_{m-1} A^{(n+1)-}] = 0, \]

where $A$ stands for the fully dissociated form of the protolytic species (e.g., CO$_3^{2-}$, S$^{2-}$ or BO$_3^{-}$), $m$ is the number of protons (e.g., $m = 2$ in H$_2$CO$_3$), $n$ is the negative charge of the dissociating species (e.g., $n = 0$ for H$_2$CO$_3$, $n = 1$ for HS$^-$), and $K_{eq}$ is the equilibrium constant for a specific acid-base reaction. Corrections for in-situ conditions are performed identically than for solubility products.

\[ ADS = k_{ADS} \cdot [L] \cdot (1 - IAP/K_{sp})^n \quad \text{if } IAP < K_{sp} \]

\[ ADS = k_p(1 - IAP/K_{sp})^n \quad \text{if } IAP > K_{sp} \]

where $k_{ADS} [\text{yr}^{-1}]$ and $k_p [\text{mol} \cdot g_{\text{solid}}^{-1} \cdot \text{yr}^{-1}]$ are the rate constants for adsorption and desorption, respectively, $[L]$ denotes the concentration of the solute, $K_{sp}$ the equilibrium constant (or solubility product), and $n$ the order of the reaction. The rate of adsorption depends on the concentration of the solute in the aqueous phase and the concentration of the solid phase. In the groundwater scenario, equilibrium constants corrected for the in-situ temperature ($15^\circ C$) are used, while in the early diagenetic simulations, apparent equilibrium constants which incorporate both temperature ($10^\circ C$) and electrolyte effects of marine solutions have been applied [Boudreau, 1997].
For instance, the rate of reoxidation of Fe$^{2+}$ by MnO$_2$ (reaction 14 in Table 3a) is given by $R_{14} = -d[MnO_2]/dt = k_{SRR,14}[Fe^{2+}]^{2} \cdot [MnO_2]$. Since the rate law is given in moles of MnO$_2$ reduced per unit time per unit gram of solid (i.e., $k_{SRR}$ has units of $L/(mol_{Fe^{2+}} \cdot yr)$), the appropriate stoichiometric coefficients for each species are

$$\begin{align*}
\text{HCO}_3^- + \text{Fe}^{2+} + \text{MnO}_2 &\rightarrow \text{Fe(OH)}_3 + \text{Mn}^{2+} + \text{CO}_2 \\
-2SD - 2SD &- 1 + 2 + 1SD + 2SD
\end{align*}$$

As explained in the previous section, the concatenated list of variables (Table 3b) is used to generate a new Web submission form in which all variable-specific parameters are constrained. The information to be provided is identical for every variable of the RN. As an example, Table 4 shows such parameterization for $O_2$ in the case of the ED application.

The information provided in Tables 2, 3a, 3b, and 4 completely specifies the problem at hand. This information is then parsed in the ACG for creation of the model executable via our Web-based runtime server. Examples of model results are succinctly discussed below.

### 4.3. Results and Discussion

#### 4.3.1. Groundwater Simulation

A stepwise approach is used here to investigate redox and acid-base chemistry. In the first

<table>
<thead>
<tr>
<th>Parameter Group</th>
<th>Field</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular diffusion coefficient</td>
<td>value at 0°C</td>
<td>380.45</td>
</tr>
<tr>
<td>temperature dependency</td>
<td>$\alpha$</td>
<td>0.06</td>
</tr>
<tr>
<td>Upper boundary condition</td>
<td>type</td>
<td>fixed concentration</td>
</tr>
<tr>
<td>value</td>
<td>250.0e-6</td>
<td></td>
</tr>
<tr>
<td>Lower boundary condition</td>
<td>type</td>
<td>fixed gradient</td>
</tr>
<tr>
<td>value</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Output file</td>
<td>generate name</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>o2.dat</td>
<td></td>
</tr>
</tbody>
</table>

$a$ Similar information must be provided for every chemical species of the RN (Table 3b).
simulation (scenario 1, Figures 4 and 5), only the primary redox reactions and acid-base equilibria are considered (reactions 1–10, Table 3a). In a second scenario (Figures 6 and 7), reoxidation reactions of the reduced products of organic matter degradation and precipitation/dissolution of carbonate and sulfide minerals are added to the reaction network.

**4.3.1.1. Scenario 1**

[46] In this first simulation, only the six primary biodegradation reactions as well as the carbonate and sulfide mineral precipitation/dissolution are considered, and the network of reactions is simplified to capture the main processes. This approach allows for a focused study of the primary redox reactions and their impact on the geochemical system.
and sulfide dissociation reactions are considered (reactions 1, 3 and 5–10, Table 3a). The location and magnitude of the overall rate of organic matter oxidation are subject to the value of the rate constant selected for each of the two CH$_2$O fractions. The highly reactive pool (CH$_2$O$_{lab}$, $k_{PRR} = 30$ yr$^{-1}$) is rapidly depleted within the first 12 m downstream of the recharge point (not shown) while the more refractory fraction (CH$_2$O$_{ref}$, $k_{PRR} = 0.3$ yr$^{-1}$) disappears after 300 m (Figure 4). There, it is clearly shown a general inverted sequence of degradation pathways, starting with sulfate reduction and methanogenesis at the leachate source, followed by an extensive zone with coexisting Fe(III) and Mn(IV) reduction, denitrification and, finally, a plume front dominated by

Figure 6. Pathways of CH$_2$O$_{ref}$ degradation obtained in the GWA application for scenario 2 after a transient simulation time of 22 years.

Figure 7. The pH distribution along the flow path in the GWA application for scenario 2 after a transient simulation time of 22 years.
oxic respiration. Such a reversed pattern in the order of primary redox reactions is frequently observed in groundwater systems contaminated with high loads of organic carbon [Chapelle, 2001].

[47] The individual effect of each of the PRR on the proton balance can be estimated from the modeled process rates (Figure 4), taking into account appropriate stoichiometric coefficients of protons produced or consumed per mole of metabolized organic matter. The stoichiometric coefficients have been calculated from the charge balance constraint, using the proper speciation of the aqueous carbonate and sulfur species at the in situ (computed) pH (not shown). For a detailed methodology, see Jourabchi et al. [2005]. A positive value for proton production rates leads to a decrease in pH and vice-versa. The pH distribution resulting from the effect of redox kinetic processes, combined with the buffering capacity of the carbonate-sulfur system, is shown in Figure 5. The profile is characterized by three distinct areas: In region A, a sharp increase in pH close to the point of recharge is predicted, which is caused by sulfate reduction of the CH$_2$O$_{ref}$ pool. The steady increase in the middle of the profile up to point B corresponds to a region where organic carbon degradation of CH$_2$O$_{ref}$ is dominated by Fe(III) and Mn(IV) reduction processes, both of which cause a net proton consumption. The subsequent decrease in pH from B to C results from aerobic degradation, which is the major respiratory pathway between 200 and 300 m downstream of the recharge (Figure 4).

4.3.1.2. Scenario 2

[48] In scenario 1, the reduced end-products of organic carbon degradation (NH$_4^+$, Mn$^{2+}$, Fe$^{2+}$, H$_2$S and CH$_4$) were accumulating along the flow path without possibility of further reoxidation. In scenario 2, secondary redox reactions are considered. Figure 6 shows that these reactions significantly influence the relative contributions and spatial distribution of the various organic carbon mineralization pathways. Although sulfate reduction and methanogenesis still predominate at the leachate source, the Mn reduction zone between 10–200 m is no longer observed in Figure 6. MnO$_2$ is instead reduced by the chemical reoxidation of Fe$^{2+}$ (Table 3a, reaction 14) produced by the Fe(III) reduction process. When compared to the previous simulation, the latter process dominates the other organic matter degradation pathways between 20–225 m. This trend can be explained from the fact that the rate by which Fe(OH)$_3$ is being produced through Fe$^{2+}$ reoxidation by O$_2$/MnO$_2$ (Table 3a, reactions 13 and 14) is faster than the Fe(OH)$_3$ consumption rate by total sulfide (TS) oxidation (Table 3a, reaction 17).

[49] Another significant difference between the two scenarios is the increased relative importance of denitrification at the downstream edge of the moving CH$_2$O$_{ref}$ front when secondary redox reactions are incorporated in the RN. The additional source of NO$_3^-$ is the nitrification process (reaction 11, Table 3a), which consumes O$_2$ and hence increases the relative contribution of denitrification at the expense of aerobic respiration. The contribution of O$_2$ as an oxidant for organic carbon (OC) degradation is further reduced by the other secondary redox reactions (Mn$^{2+}$, Fe$^{2+}$ and TS oxidation by O$_2$) which operate right at the edge of the CH$_2$O$_{ref}$ front (approximately 300 m).

[50] The resulting pH distribution obtained for scenario 2 is shown in Figure 7. Precipitation of FeS and FeCO$_3$ at the leachate source leads to a considerable pH drop from the boundary value of 6.5 down to 5.6 (region A). The sharp increase in pH in region B is predominantly due to the combined effect of TS reoxidation reactions by Fe(OH)$_3$ and MnO$_2$, both of which are localized at a distance of around 280 m. The slight pH drop (about 0.1 unit) in area C results from the reoxidation of Fe$^{2+}$ by O$_2$ (Table 3a, reaction 13) and MnO$_2$ (Table 3a, reaction 14), respectively. Further downstream of area C, the results are still dependent on the initial conditions since the CH$_2$O$_{ref}$ front has not yet propagated into this area.

4.3.2. Early Diagenetic Simulation

[51] Figure 8 shows a classical redox sequence of organic matter degradation pathways which develops over the first 15 cm of the sediment column. Since sulfate is present in excess, all metabolic pathways except methanogenesis are active in the system. Integration of the rates indicates the predominance of sulfate reduction, followed byoxic degradation, denitrification, iron reduction and finally manganese reduction in the overall rate of organic matter degradation. These calculations take into account the contributions of the various reoxidation reactions in the carbon balance. The latter have a predominant influence on the depth distribution of pH (Figure 9). Indeed, except for iron reduction, which is an important proton consumption reaction in the system, the major processes controlling the H$^+$ balance are nitrification, oxy-
The generation of TS and Mn$^{2+}$, reoxidation of Fe$^{2+}$ by MnO$_2$ as well as reoxidation of TS by Fe(OH)$_3$ (Table 3a). The former four reactions produce protons while the latter is a proton-consuming process. Two zones in the sediment profile can thus be distinguished: a zone of pH decrease (~0–2 cm) dominated by the aerobic oxidation of Mn$^{2+}$ and TS, and a zone of pH increase (~2–4 cm) dominated by Fe(III) reduction and reoxidation of TS by iron oxides. Note that the proton production by secondary redox reactions in the uppermost layer leads to a significant dissolution of calcite which is buffering the pH drop. Further analysis of the complex interplay between redox, pH dynamics and carbonate precipitation/dissolution is provided by Jourabchi et al. [2005].

5. Conclusions

This paper presents a simulation environment for one-dimensional reactive transport applications...
that addresses the major hurdles for the common use of complex reactive transport models. In particular, the KB-RTM approach provides complete flexibility in the choice of model components (chemical species) and the description of the interaction between them (reactions) by using a generic form of the advection-dispersion equation and by taking a “reaction-centric” approach which utilizes the unifying conceptual and mathematical principles underlying all RTMs. Thus 1-D transport descriptions relevant to many compartments of the Earth system (rivers, estuaries, groundwater or sediments) can be incorporated in our simulation environment. The proposed approach should thus help overcome traditional disciplinary barriers between the different subfields of RTM. In addition, the use of symbolic programming and automatic code generation allows testing and comparison of alternative process formulations. As a result, it facilitates the evaluation of the effect of competing model structures on predictions and uncertainties.

The shear volume of ever growing knowledge on process dynamics in the natural environment can also impede its timely incorporation into RTMs. We argue that the Knowledge Base (KB) is a particularly efficient approach for constructing reactive-transport applications of increasing complexity, as well as testing and comparing alternative process formulations. The KB system allows expertise to be stored in a dynamically evolving, Web-distributed Knowledge Base and provides a platform where both modelers and experimentalists can share expertise. The KB-RTM is available via our Web server at http://www.geo.uu.nl/~kbrtm.

Further development and maintenance of the simulation environment will involve the progressive implementation of a larger number of processes, and a continuing effort to keep the content of the KB up-to-date with new results from the scientific community. A Web-based Knowledge Book, consisting of documentation material for every process implemented in the KB, will also be implemented with complete reference to the source material, following the strategy already used for the CONTRASTE model [Regnier et al., 2002]. The Web-based system further offers the prospect for use as a forum for discussion, open to user feedback and contribution.

Here, we have illustrated the workings of our simulation environment using two model applications dealing with redox and coupled acid-base chemistry in sediments and groundwater environments. Our examples demonstrate the flexibility and potential of the proposed simulator.

Our objectives have been achieved by adhering to a modular structure, where information assemblage precedes symbolic process formulation, followed by automatic code generation and combination with numerical algorithms for solution. Finally, accessibility through the Internet is a key feature for such dissemination and combination of expert knowledge on modeling and process description. To our knowledge, this is the first attempt to use a Web-distributed flexible system to provide dynamically adaptive reactive transport models in the field of geosciences.

Acknowledgments

This research has been financially supported by several grants from the Netherlands Organization for Scientific Research (NWO), the European Commission (METROL project EVK3-CT-2002-00080), the Belgian Science Policy Program (SISCO project, EV/11/17A, EV/02/17B), and DHI-Water and Environment. We are also thankful to A. Dale and three anonymous referees for constructive comments on this paper.

References


Schäfer, D., W. Schäfer, and W. Kinzelbach (1998b), Simulation of reactive processes related to biodegradation in aquifers: 2. Model application to a column study on-


