ELSEVIER

Computers & Geosciences 32 (2006) 64–72

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

A general simulator for reaction-based biogeochemical processes☆

Yilin Fang[a,*], Steven B. Yabusaki[a], Gour-Tsyh Yeh[b]

[a]*Pacific Northwest National Laboratory, P.O. Box 999, MS K9-36, Richland, WA 99352, USA*
[b]*Department of Civil and Environmental Engineering, University of Central Florida, Orlando, FL 32816, USA*

Received 15 October 2004; received in revised form 28 April 2005; accepted 3 May 2005

Abstract

As more complex biogeochemical situations are being investigated (e.g., evolving reactivity, passivation of reactive surfaces, dissolution of sorbates), there is a growing need for biogeochemical simulators to flexibly and facilely address new reaction forms and rate laws. This paper presents an approach that accommodates this need to efficiently simulate general biogeochemical processes, while insulating the user from additional code development.

The approach allows for the automatic extraction of fundamental reaction stoichiometry and thermodynamics from a standard chemistry database, and the symbolic entry of arbitrarily complex user-specified reaction forms, rate laws, and equilibria. The user-specified equilibria and kinetic rates (i.e., they are not defined in the format of the standardized database) are interpreted by the Maple V (Waterloo Maple) symbolic mathematical software package. FORTRAN 90 code is then generated by Maple for (1) the analytical Jacobian matrix (if preferred over the numerical Jacobian matrix) used in the Newton–Raphson solution procedure, and (2) the residual functions for governing equations, user-specified equilibrium expressions and rate laws. Matrix diagonalization eliminates the need to conceptualize the system of reactions as a tableau, which comprises a list of components, species, the stoichiometric matrix, and the formation equilibrium constant vector that forms the species from components (Morel and Hering, 1993), while identifying a minimum rank set of basis species with enhanced numerical convergence properties. The newly generated code, which is designed to operate in the BIOGEOCHEM biogeochemical simulator, is then compiled and linked into the BIOGEOCHEM executable. With these features, users can avoid recoding the simulator to accept new equilibrium expressions or kinetic rate laws, while still taking full advantage of the stoichiometry and thermodynamics provided by an existing chemical database. Thus, the approach introduces efficiencies in the specification of biogeochemical reaction networks and eliminates opportunities for mistakes in preparing input files and coding errors. Test problems are used to demonstrate the features of the procedure.

Keywords: Reaction network; BIOGEOCHEM; Thermodynamic database; Symbolic programming; Diagonalization

1. Introduction

Biogeochemical modeling can be used by investigators to test hypotheses that may explain experimental observations or, coupled with hydrologic transport, to evaluate the impact of complex biogeochemical

---

☆Code on server at http://www.iamg.org/CGEditor/index.htm.

*Corresponding author. Tel.: +1 509 372 4051; fax: +1 509 372 6089.
E-mail address: yilin.fang@pnl.gov (Y. Fang).

0098-3004/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2005.05.003

processes on natural systems. One can use well-established biogeochemical software (e.g., MINTEQ (Allison et al., 1991), OS3D (Steefel and Yabusaki, 1996), KEMOD (Yeh et al., 1995) and Parkhurst and Appelo (1999)) for a large spectrum of problems. Alternatively, one can develop his/her own code by either modifying an existing code or writing a problem specific code.

The common aspects of the biogeochemical models are related to information collection and the manner in which the problem is solved. Specifically, the information related to a biogeochemical reaction network, such as the reaction stoichiometries and thermodynamics, geochemical constraints (e.g., total concentration, fixed activity), kinetic rate laws (e.g., simple kinetics, transition state theory, Monod kinetics, and others) and parameters has to be collected, and then the user must solve the problem numerically by either directly integrating the system of ordinary differential equations or using the mixed differential/algebraic equation (DAE) approach which applies also to purely algebraic equation, distinctions between which can be found in Fang et al. (2003).

There are several biogeochemical simulators that can model mixed kinetic and equilibrium reactions (cf. Yeh et al., 1995; Steefel and Yabusaki, 1996; Lichtner, 1996; Parkhurst and Appelo, 1999; Tebes-Stevens et al., 1998). However, many investigators still favor their own specific models. This is because most biogeochemical simulators allow only a few limited forms to express reactions and rate equations for kinetic reactions and reaction types. Furthermore, the reactions are often required to be specified in tableau form (i.e., identifying separate primary components and secondary species) (e.g., Yeh et al., 1995; Steefel and Yabusaki, 1996; Salvage and Yeh, 1998). If components need to be reselected due to numerical considerations (for example, the concentration of one of the components falls so low during the simulation that the solution does not converge) or some reactions should be considered as kinetic instead of equilibrium, the tableau has to be rewritten with a new set of components and thermodynamic data.

Some simulators have embedded a thermodynamic database sweep capability that, given a set of components, can automatically identify all possible minerals and secondary species in the database and generate reaction stoichiometries. For simulators without this capability, more effort will need to be expended, and there will be more opportunities for mistakenly omitting important species. Moreover, the lack of this capability can become a constraint for the application of the simulators when one doesn't know a priori which species may be important in a system, (e.g., whether or not to consider the presence of certain minerals and secondary species in the system). Moreover, by sweeping through the database, one can obtain a more complete reaction network that may include various types of redox reactions. When a complete suite of redox reactions is included, the evolution of redox state is internally and automatically achieved with the simulation of the network, and a readjustment as to whether a reduced or an oxidized species is important from one region to the others is not necessary.

In terms of accommodating a variety of reaction rates and equilibrium reaction types, a limitation of most reactive transport simulators (e.g., Fang et al., 2003) is that the program has to be recoded to incorporate user-specified reaction types and rate laws in the reaction system. This often happens in biogeochemical studies where different rate laws are proposed to adequately describe the system. Therefore, this limits the generality of the simulator if the user has to write a new source code every time a new reaction form or rate law is proposed. It is also something that typical investigators would avoid because (1) they may not know how to program, or (2) they may make a mistake while programming, and/or (3) it takes too long to develop a new code and test whether it works. However, this constraint can be overcome using a symbolic mathematical software package such as Maple (Waterloo Maple) to generate necessary subroutines such as residual functions, reaction rates, and analytical Jacobian matrices (a Jacobian matrix is the derivative of a certain set of functions with respect to the unknown variables, i.e., species concentrations in this case needed for the Newton–Raphson nonlinear solution scheme) when accuracy is a major concern. Some numerical models, such as RAFT (Chilakapati et al., 1998) and CONTRASTE (Regnier et al., 2002), already employ this automatic code generation for functionals and Jacobian matrices.

The tableau restriction can also be relieved using a general paradigm to model reaction-based biogeochemical processes using a diagonalization procedure (Fang et al., 2003; Yeh et al., 2005a, b). The salient aspects of this procedure are that (1) linearly independent kinetic reactions can be segregated, and can then be used to formulate and parameterize all reactions individually if reaction networks are appropriately designed, (2) systematic operations can be performed to remove linearly dependent equilibrium reactions and kinetic reactions that depend on equilibrium reactions, (3) chemical basis species can be systematically identified to explicitly enforce mass conservation using algebraic equations, and (4) all equilibrium reactions can be eliminated by substitution from the set of ordinary differential equations governing the evolution of kinetic variables to reduce problem stiffness. This approach does not require chemical reactions to be written in tableau form, i.e., chemical components must be selected a priori. Instead, this approach reads stoichiometric