A Stochastic Simulation of Natural Organic Matter and Microbes in the Environment

Xiaorong Xiang          Gregory Madey
Yingping Huang         Steve Cabaniss
(University of New Mexico)
Department of Computer Science and Engineering
University of Notre Dame
Sponsored by NSF/ITR-DEB
Objectives

- New approach for NOM modeling
  - Agent-based modeling
- E-Science on the Web
- Intelligent interface
- The NOM Collaboratory
Outline

- Introduction
- Modeling
- Core simulation engine
- Intelligent Web-based interface
- The NOM collaboratory
- Conclusion
- Future work
Introduction

- What is Natural Organic Matter (NOM)?
- Role of NOM in various science disciplines
  - Mobility and transport of pollutants
  - Availability of nutrients for microorganisms and plant communities
  - Affects quality of drinking water
- Need to understand the evolution and heterogeneous structure of NOM
Forest Service Bog [DOC] 7 MW 2200

Twomile Creek [DOC] 17 MW 1500

Nelson Creek [DOC] 79 MW 900

UNDERC
University of Notre Dame Environmental Research Center
Previous models

Two examples:

- **Daisy** (S. Hansen, H. E. Jensen, and N. E. Nielsen 1990-present): a soil plant atmosphere system model

- **StochSim** (C. J. Morton-Firth 1998-present): Stochastic simulation of cell signaling pathways
Our model

- Agent-based modeling (Individual-based modeling)
  - Agent-based model
  - Each molecule as an individual object with spatial properties
  - Bottom-up approach
  - Stochastic model
  - Trace changes of the system → Database and data mining technologies
Our model (cont.)

- Web-based simulation model
  - Serve as an example for E-Science
    - R. M. Jakobovits, J. F. Brinkley, C. Rosse, and E. Weinberger (1998): Enabling clinicians, researchers, and educators to build custom Web-based biomedical information system
  - Support the collaborations, data and information sharing between scientists
  - No installation, expensive computation resources needed by scientists
Outline

- Introduction
- **Modeling**
- Core simulation engine
- Intelligent Web-based interface
- The NOM collaboratory
- Conclusion
- Future work
Modeling

- A complex system
  - Consists of a large number of objects
    - Molecules, Microbes
      - Heterogeneous properties
      - Individual behaviors
      - Interaction between each other
      - Objects behavior and interaction are stochastically determined by:
        - Attributes (intrinsic parameters)
        - Reactions rates
        - Environment condition (extrinsic parameters)
  - No central control
  - Emergent properties
Modeling (cont.)

- **Agent Attributes**
  - More specific than “percent carbon” but less detailed than a molecular connectivity map
  - **Elemental composition**
    - (C, H, O, N, S, P)
  - **Functional groups**
    - (double bonds, ring structure, alcohols …)
  - **The origin of objects**
    - spatial position in the system
    - Precursor type of molecule
  - **Probability table of physical and chemical reactions**
  - **Molecule weight**
Modeling (cont.)

- Agent Behaviors (reactions and processes):
  - Transport through soil pores by water (spatial mobility)
  - Adsorb onto or desorbed from mineral surfaces
  - Chemical reactions
    - Total 10 types in current model
    - First order
    - Second order
  - Stochastically determined

- Space:
  - 2-D grid
Modeling (cont.)

- **Environmental parameters**
  - Temperature
  - pH
  - Light intensity
  - Microbe density
  - Water flow rate
  - ...

Outline

- Introduction
- Modeling
- **Core simulation engine**
- Intelligent Web-based interface
- The NOM collaboratory
- Conclusion
- Future work
Core simulation engine

- Implementation
  - Swarm toolkit (Santa Fe Institute)
  - Java programming language (Sun JDK 1.4.1_01)
- GUI version
  - View the animation of molecules
  - Easy for debugging and modeling
- Web-based version
Core simulation engine (cont.)

- Read simulation parameter from the database (JDBC)
  - Environmental parameters (pH, temperature, light intensity, and so on)
  - Molecule types and distributions
- User defined time has been separated to a large number of equal size time steps
- Write relevant data into the database every time step (JDBC)
  - Trace the dynamic properties of individuals and the system over time
Visualization

Black - No Adsorption
Grays - Levels of Adsorption
Red - Lignins
Green - Cellulose
Blue - Proteins
Yellow - Reacted
Orange - Adsorbed
Visualizations — NOM molecules in solution and adsorption
Outline

- Introduction
- Modeling
- Core simulation engine
- Intelligent Web-based interface
- The NOM collaboratory
- Conclusion
- Future work
Web-based model

- Distributed, Web-based scientific application model
- Based on Sun Java 2 Enterprise Edition (J2EE)
  - Standard HTML Forms / JSP
  - Java Servlets, Java Beans
  - JDBC - Oracle
  - Oracle Database, Oracle Forms, and Reports
- Three parts:
  - Intelligent Web-based interface
  - Core simulation engine
  - Data analysis packages, Data mining technologies
Access NOM simulation from Web
Web-based interface

- Sign up
  - New user
  - Login

- Submitted simulations
- Simulation reports
- Dynamic running time prediction
- Email notification

- Start a new one
- Parameter input
- Invoke simulation
- Static running time prediction
- Find similar simulations
- Restarter

- Database

- Simulation engines
Example of Interface

Welcome to NOM Research Group!  
Leilani Arthurs

(* Required fields)

<table>
<thead>
<tr>
<th>Molecule Name</th>
<th>Make it available to public</th>
</tr>
</thead>
<tbody>
<tr>
<td>SnicFA</td>
<td></td>
</tr>
</tbody>
</table>

Atoms of the molecule
Each field must be a non-negative integer. Default value is 0.

<table>
<thead>
<tr>
<th></th>
<th>(Atom) C</th>
<th>(Atom) H</th>
<th>(Atom) N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>66</td>
<td>64</td>
<td>0</td>
</tr>
<tr>
<td>(Atom) O</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Functional groups of the molecule
Each field must be a non-negative integer. Default value is 0.

<table>
<thead>
<tr>
<th></th>
<th>DoubleBond</th>
<th>Rings</th>
<th>Phenyl</th>
<th>Alcohols</th>
<th>Phenols</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ethers</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Esters</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ketones</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Aldehydes</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Acids</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Arylacids</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Amines</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RingN</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Amides</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Thioethers</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Thiols</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Phosphoesters</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IIPhosphoesters</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Phosphates</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Create New Molecule Type
Report example

Please provide the following information.
Current Time Step is: 1969

Time Step: 1969
Status: adsorbed
Start Y: 0
End Y: 299

Generate Reports
Generate XML file

Weight Distribution

<table>
<thead>
<tr>
<th>Count</th>
<th>361</th>
<th>695</th>
<th>1390</th>
<th>2780</th>
<th>5560</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Intelligent interface components

- Email notification
- Running time prediction
  - Static
    - number of molecules
    - number of time steps
  - Dynamic
    - current time step
    - current wall clock time
Intelligent interface components (cont.)

- Find similar simulations
  - Environment parameters
  - Molecule types and distributions
  - Retrieve the data sets from database
  - Points on a high dimension space
  - Euclidean distance
  - Ordered list
  - Review the simulation results or restart
Intelligent interface components (cont.)

- Automatic reater
  - Save the state of each objects in the system to database every check point
  - Load the state to the core simulation engine
Outline

- Introduction
- Modeling
- Core simulation engine
- Intelligent Web-based interface
- **The NOM collaboratory**
- Conclusion
- Future work
Previous work


- BioCoRE: [http://ks.uiuc.edu/Research/biocore](http://ks.uiuc.edu/Research/biocore)

The NOM Collaboratory

- Interdisciplinary project
  - Chemists
  - Biologists
  - Ecologists
  - Computer Scientists

- Build and integrate software using J2EE
  - NOM modeling & simulation software
  - XML-based standard data format definitions
  - Data querying and manipulation tools
  - Electronic communication tools
**NOM Collaboratory**

- **NOM Simulator**: Provide an intelligent interface to facilitate using the NOM simulator.
- **Search Engine**: Provide an interface to search simulation information.
- **Discussion Board**: Provide a threaded discussion board.
- **Chat Room**: Provide a real-time chat box.
- **Molecule Editor**: Provide an interface to define new molecule type.
- **Molecule Validator**: Provide an administration role to validate the newly added molecule for public usage.
- **XML-based Markup Language (XMLML)**: Provide a XML-based Markup Language definition.
- **File Upload**: Provide an interface to upload publications.
XML-based NOM Markup Language (NOML)

- NOML:
  - Standard data format
    - Environment.dtd, Molecules.dtd, Setup.dtd
  - Facilitate communication
    - User === User
    - Application === Application
  - Extension
Other tools

- **Molecule editor**
  - Define new molecule type

- **Molecule validator**
  - Authorized persons (Chemists) to validate data
  - Share the molecule type

- **Search engine**
  - Ad-hoc query
  - View results of the completed simulations
  - Restart some simulations
Outline

- Introduction
- Modeling
- Core simulation engine
- Intelligent Web-based interface
- The NOM collaboratory
- Conclusion
- Future work
Conclusion

- Agent-based stochastic model for simulating the NOM evolution with discrete temporal and spatial properties
- A Web-based simulation architecture (multiple simulation servers, database servers, and data mining technologies)
- Database technologies
- A Web-based intelligent configuration interface
- The NOM collaboratory
Outline

- Introduction
- Modeling
- Core simulation engine
- Intelligent Web-based interface
- The NOM collaboratory
- Conclusion
- Future work
Future work

- Model testing (this summer)
  - Testing of the sorption process
  - More features need to be added into the core simulation engine
- Model validation (this summer)
- Collaboratory:
  - More communication tools
  - More simulation models for NOM study
  - NOML extension
Thank You!

Questions?
GUI version
Work flow

Init

Add molecule

Randomly pick one

Sorption

move

yes

Leave

All checked

Randomly pick one

yes

Get reaction probabilities

Random number

First order

no

Find nearest one

Find nearest empty cell

no

Update the probability table & Molecule structure

no

All checked

Completed

no

yes

Terminate

With split

Without split

disappear

yes