Stochastic Synthesis of Natural Organic Matter

Steve Cabaniss, UNM
Greg Madey, Patricia Maurice, Yingping Huang, ND
Laura Leff, Ola Olapade KSU
Bob Wetzel, UNC
Jerry Leenheer, Bob Wershaw USGS

Fall 2002
What is NOM? Sources:

- Plant and animal decay products
  - Terrestrial- woody and herbaceous plants
  - Aquatic- algae and macrophytes
- Structures
  - cellulose, lignins, tannins, cutin
  - proteins, lipids, sugars
What is NOM?  Composition

45-55 Wt% Carbon
35-45 Wt% Oxygen
3-5 Wt% Hydrogen
1-4 Wt% Nitrogen
Traces P, S
MW 200-20,000 amu
Equiv. Wt. 200-400 amu
10-35% aromatic C
What is NOM?

A mixture of degradation and repolymerization products from aquatic and terrestrial organisms which is heterogeneous with respect to structure and reactivity.
NOM Interactions with sunlight

Direct photoredox
Fe(III)-NOM $\rightarrow$ Fe(II) + NOM + CO$_2$

Photosensitizer
NOM + O$_2$ $\rightarrow$ H$_2$O$_2$
+ OH·
+ O$_2$·
+ etc.

Light attenuation

Absorption
Wavelength
NOM Interactions with mineral surfaces

Adsorption  Hemi-micelle formation  Acid or complexing dissolution  Reductive dissolution

Adsorbed NOM coatings impart negative charge and create a hydrophobic microenvironment
NOM Interactions with microbes

Ingestion:
Energy and Nutrients

Electron shuttle

Metal ion complexation and de-toxification
NOM Interactions with pollutants

Binding to dissolved NOM increases pollutant mobility
NOM in water treatment

\[\text{NOM} + \text{HOCl} \rightarrow \text{CHCl}_3 + \text{CHCl}_2\text{Br} + \text{CCl}_3\text{COOH} \text{ and other chlorinated by-products}\]
Why study NOM?

Natural ecosystem functions
  Nutrition, buffering, light attenuation
Effects on pollutants
  Radionuclides, metals, organics
Water treatment
  DBP’s, membrane fouling, Fe solubility
Carbon cycling & climate change
NOM Questions:

• How is NOM produced & transformed in the environment?
• What is its structure and reactivity?
• Can we quantify NOM effects on ecosystems & pollutants?
Environmental Synthesis of Natural Organic Matter

Cellulose
Lignins
Proteins
Cutins
Lipids
Tannins

O₂, light, bacteria, H⁺, OH⁻, metals, fungi

NOM
Humic substances & small organics

CO₂
Simulating NOM Synthesis
Deterministic Reaction Kinetics

For a pseudo-first order reaction

\[ R = \frac{dC}{dt} = k' \ C \]

- \( R \) = rate (change in molarity per unit time)
- \( C \) = concentration (moles per liter)
- \( k' \) = pseudo-first order rate constant (units of time\(^{-1}\))

Based on macroscopic concentrations
Deterministic Reaction Kinetics: Solve a system of ODE’s

• Begin with initial $C_i$ for each of $N$ compounds, $k_j$ for each of $M$ reactions
• Apply Runge-Kutta or predictor-corrector methods to calculate $C_i$ for each time step (use Stiff solvers as needed)
• Repeat for desired length of simulation, obtaining results as $C_i$ versus time
Problem w/ ODE approach: Size and Computation Time

- Assuming $N > 200$ (different molecules)
- Assume $M = 20 \times N$ (20 reactions per molecule)
- Total set of >4000 very stiff ODE’s is impractical (transport eqns not included)
Problem w/ODE Approach: Knowledge Base

- Structures of participating molecules unknown
- Pertinent reactions unknown
- Rate constants $k_j$ unknown
Simulating NOM Synthesis
Probabilistic Reaction Kinetics

For a pseudo-first order reaction

\[ P = k' \_t \]

\( P = \) probability that a molecule will react
with a short time interval \( t \)
\( k' = \) pseudo-first order rate constant
units of time\(^{-1}\)

*Based on individual molecules*
Stochastic algorithm: Initialization

- Create initial pseudo-molecules (objects)
  - Composition (protein, lignin, cellulose, tannin)
  - Location (top of soil column, stream input)
  - Input function (batch mode, continuous addition, pulsed addition)
- Create environment
  - specify pH, light, enzyme activity, bacterial density, humidity, T°, flow regime
Stochastic Algorithm: Reaction Progress

• Chemical reaction: For each time-slice, each pseudo-molecule
  – determine which reaction (if any) occurs
  – modify structure, reaction probabilities
• Transport: For each time-slice, each pseudo-molecule
  – Determine mobility
  – Modify location, reaction probabilities
• Repeat, warehousing ‘snapshots’ of pseudo-molecules and aggregate statistics
Stochastic Algorithm: Advantages

- Computation time increases as # molecules, not # possible molecules
- Flexible integration with transport
- Product structures, properties not pre-determined
Stochastic synthesis: Data model

Pseudo-Molecule

- Elemental Composition
- Functional Structural
- Calculated Chemical Properties and Reactivity
- Location Origin State
- State
**Average Lignin Molecule:**
Oligomer of 40 coniferyl alcohol subunits

<table>
<thead>
<tr>
<th>Numbers of atoms</th>
<th>Numbers of functional groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 Carbon</td>
<td>40 Total ring structures</td>
</tr>
<tr>
<td>322 Hydrogen</td>
<td>40 Phenyl rings</td>
</tr>
<tr>
<td>81 Oxygen</td>
<td>1 Alcohol</td>
</tr>
<tr>
<td></td>
<td>1 Phenol</td>
</tr>
<tr>
<td></td>
<td>118 Ether linkages</td>
</tr>
</tbody>
</table>
Model reactions transform structure

Ester Hydrolysis

Ester Condensation

Amide Hydrolysis

Dehydration

Microbial uptake
Reaction Probabilities: P calculated from

- Molecular structure
- Environment (pH, light intensity, etc.)
- Proximity of near molecules
- State (adsorbed, micellar, etc.)
- Length of time step, _t
Example: Ester Hydrolysis

\[ P = (\# \text{ Esters}) \, A \, e^{-\frac{E_a}{RT}} \left(1 + b[H^+] + c[OH^-]\right) \]

Where
- \( A = \) Arrhenius constant
- \( E_a = \) activation energy
- \( R = \) gas constant
- \( T = \) temperature, Kelvins
- \( b = \) acid catalyzed pathway
- \( c = \) base catalyzed pathway
Property prediction

**Environmental**
- Light absorbance
- Molecular weight
- Acid content & pKₐ
- Bioavailability
- Kₗw
- Metal binding K

**Analytical**
- Elemental %
- Titration curves
- IR Spectra
- NMR spectra
Property Calculation Methods

- Trivial - MW, elemental composition, Equivalent weight
- Simple QSAR - $pK_a$, $K_{ow}$
- Interesting
  - Bioavailability
  - Light absorption
  - Metal binding
Presentation and Analysis

- Spatial mapping of molecules
- Results stored in Oracle database
- Remote query via WWW interface
- Standard graphs of reaction frequency, molecular properties versus time
Trial: Can we convert lignin oligomer (MW ~6000) in “NOM”?

Atmospheric O₂  No light
Neutral pH            No surfaces
Moderate enzyme activity  No transport
27 months reaction time
Lignin in dark water

Number of Molecules

Time (hours)
Lignin in dark water

% Carbon

% Oxygen

Time (hours)
Lignin in dark water

Time (hours)

Mw

Mn
Lignin in dark water

Percent aromatic carbon

Time (hours)
Lignin in dark, acidic soil

![Graph showing the number of molecules over time in dark, acidic soil.](image-url)
Lignin in dark, acidic soil
Lignin in dark, acidic soil

Fraction aromatic carbon

Time (hours)
Lignin -> NOM conversion

- Elemental composition similar to whole water NOM
- Average MW within range for aquatic NOM, soil NOM respectively
- Aromaticity lower than normal
Stochastic synthesis
Preliminary tests

- Chromatography-like NOM movement in soils and sub-surface
- Log-normal distribution of NOM molecular weights
- Rapid consumption of proteins
Current development

- Expanding reaction set
- Determination of reaction probabilities
- Best method of spatial mapping
  - Discrete grid vs Continuous space
- Remote query capability
Next Steps-

- Property prediction algorithms
- Data mining capabilities
- Comparison with lab and field results
Stochastic Synthesis of NOM

Goal: A widely available, testable, mechanistic model of NOM evolution in the environment.
Financial Support

NSF Division of Environmental Biology and Information Technology Research Program

Collaborating Scientists

Steve Cabaniss (UNM) Greg Madey (ND)
Jerry Leenheer (USGS) Bob Wetzel (UNC)
Bob Wershaw (USGS) Patricia Maurice (ND)
Laura Leff (KSU)