Arrays Revisited

SBML Forum #11

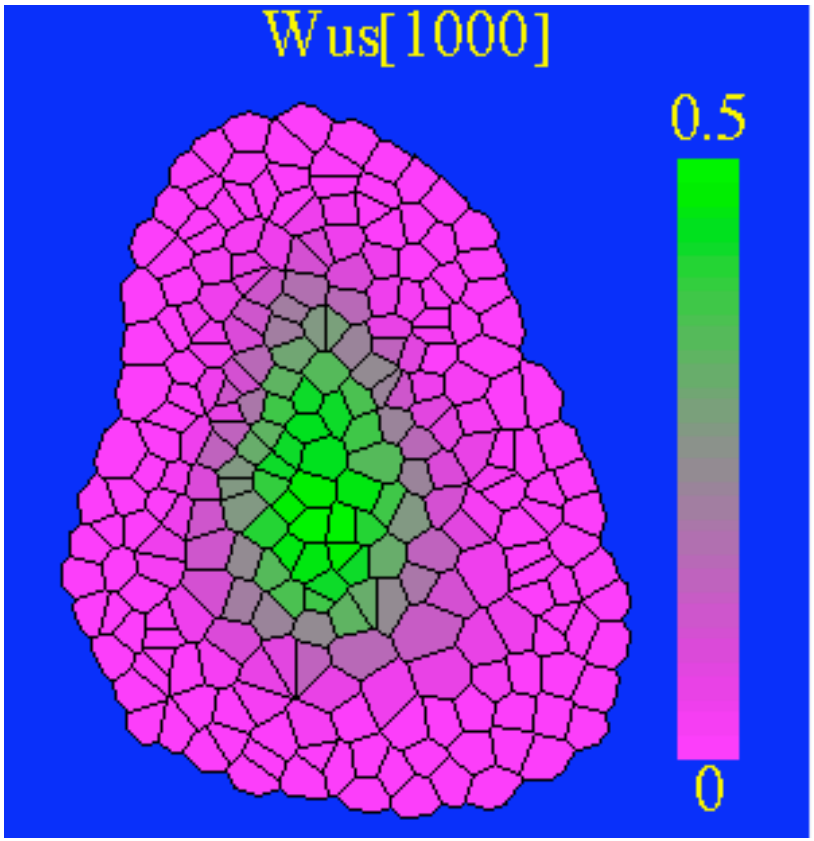
13 October 2006

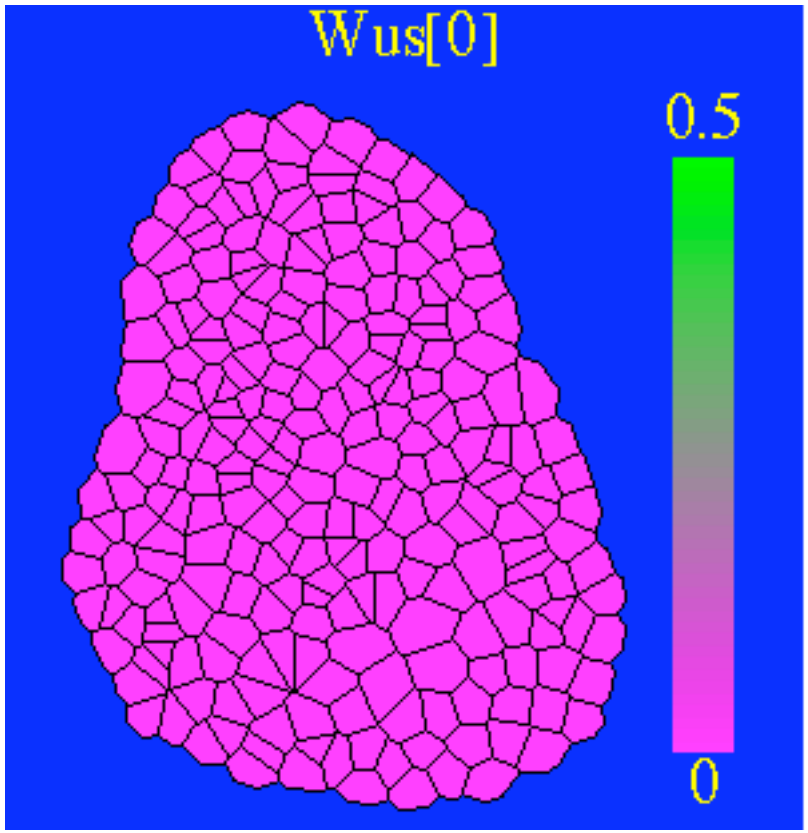
Bruce Shapiro

Sets, Lists, Bags, Collections, Vectors, Matrices, Bunches, Piles, Groups, Agglomerations, Clumps, Clusters, Medleys, Caboodles, Shebangs, Multitudes, Passels, Bevys, Oodles,…

Arrays allow us to describe a bunch of stuff without listing every item explicitly every time

http://computableplant.org





Jönsson et al (2005) Bioinformatics, 21: i232-i240.

A Simple Model

$$\left.\begin{array}{c} \varnothing \xrightarrow{Y,\,A} W \\ W \rightarrow \varnothing \\ \varnothing \xrightarrow{L1} Y \\ Y \rightarrow \varnothing \\ A \xrightarrow{Y} \varnothing \\ 2A+B \rightarrow 3A \\ A \rightarrow B \end{array}\right\} \times 253 \text{ cells} + \text{ diffusion} \begin{cases} Y_i \rightleftarrows Y_j \\ A_i \rightleftarrows A_j \\ B_i \rightleftarrows B_j \end{cases} \times 705 \text{ cell walls}$$

$$7 \times 253 + 6 \times 705 = 6001 \text{ reactions}$$

plus: 253 compartment volumes

plus: 253 cell centers (x, y coords)

```
<reaction id="reaction2608"
      name="Diffusion of Y from Cell 201 to Cell 218"
      reversible="false" fast="false">
   <listOfReactants>
    <speciesReference species="Y_201"/>
   </listOfReactants>
   <listOfProducts>
    <speciesReference species="Y_218"/>
   </listOfProducts>
   <kineticLaw timeUnits="time" substanceUnits="substance">
    <math xmlns="http://www.w3.org/1998/Math/MathML">
     <apply>
      <times/>
      <ci>Dy</ci>
      <apply>
       <minus/>
       <ci>Y_201</ci>
       <ci>Y_218</ci>
      </apply>
     </apply>
    </math>
   </kineticLaw>
  </reaction>
```

134,000 lines
3.4 MB SBML File

Typical Current Model (Unpublished)

- 3D Geometry
  - 805 polyhedral cell volumes
  - 805 x 3 = 2415 cell nuclear coordinates
  - 31115 x 3 = 93345 facet vertex coordinates
    - Plus associated rate rule for each vertex
  - [Side Remark: Most cells in nature are NOT polyhedral
    - Finite element descriptions are MUCH larger]
- 14 species per cell, 11270 Total
- 27 reactions per cell, 21735 Total
- 5 species diffuse, 40370 reactions

Useful MathML “Array” Objects

- Vectors
  - X[1], X[2], …
- Matrices
  - Y[1,1], Y[1,2], …
- Sets (nested)
  - {{X,Y,Z}, {X,Y,Z}, …}
- Lists = sets where the order is significant

But arrays are not biological entities …

Neither are variables

Variables describe biological entities

Arrays (sets, lists, bags, etc.) describe collections of biological entities

Biologists won't accept arrays (sets, lists, collections, bags, etc.)

Biologists are not stupid.

Proposals

- http://sbml.org/wiki/arrays
- Finney et al (2003) Arrays
- Finney et al (2003) Sets
  - Strongly Tied to Model Composition
- Computable Plant (2004)

Experience

- “Arrays” are not supported by MathML
  - Vectors and matrices are not sufficient
  - Sets and/or Lists (better if both are supported) needed to easily describe things with multiple “indices” e.g., tensors, multi-component species
- Need some way to described connectivity
  - E.g., “sum over” all cells connected to cell j
- Static arrays are not sufficient
  - Size of object MUST be dynamic
    - Upper/lower bounds as formulas
  - Birth/death rules for object elements

Example

- A mersitem is collection of ≈500 cells
  - cell[j], j=1, 2, …
  - cell[j] has k polyhedral faces
  - face[j,k] has m vertices and n edges
  - any object described by finite element covers (triangulation)

```
<listOfParameters>
<parameter id="n" value="10"/>
<parameter id="x">
    <listOfDimensions>
        <dimension id="i" lowerLimit="1" upperLimit="n" />
        <dimension id="j" lowerLimit="0">
        <upperLimit>
            <math xmlns="http://www.w3.org/1998/Math/MathML">
            <apply>
            <plus/>
                <apply>
                <times/>
                    <cn type="integer">2</cn>
                    <ci>m</ci>
                </apply>
                <cn type="integer">1</cn>
            </apply>
            </math>
        </upperLimit>
        </dimension>
    </listOfDimensions>
</parameter> …
```

Implied Array of Rules

```
<rateRule variable="x">
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      <apply>
      <times/>
         <ci>z</ci>
         <ci>y</ci>
      </apply>
    </math>
</rateRule>
```

$$\frac{dx_i}{dt} = y_i z_i$$

Implied Reaction Array

$$A_i + B_i \rightarrow C_i$$

```
<listOfCompartments>
        <compartment id="cell">
        <listOfDimensions>
        <dimension id="i" lowerLimit="1" upperLimit="100"/>
        </listOfDimensions>
</compartment>
</listOfCompartments>
<listOfSpecies>
        <species id="A" compartment="cell" />
        <species id="B" co
        <species id="C" co
</listOfSpecies>
```

```
<reaction id="reaction1">
  <listOfReactants>
    <speciesReference species="A"/>
    <speciesReference species="B"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="C"/>
  </listOfProducts>
</reaction>
```

<selector>

$$A_2 v_k$$

```
<apply>
<times/>
    <apply>
       <selector/>
          <ci>A</ci>
          <cn type="integer">2</cn>
    </apply>
    <apply>
       <selector/>
          <ci>v</ci>
          <ci>k</ck>
    </apply>
</apply>
```

MathML Functions Required

- Constructors: `set, list, vector, matrix, matrixrow`
- Containers: `lowlimit, uplimit, selector, domainofapplication`
- Operators: `union, intersect, in, notin, subset, prsubset, notsubset, notprsubset, setdiff, card, cartestianproduct, sum, product, vectorproduct, scalarproduct, outerproduct, transpose, determinant`
- Calculus: `curl, divergence, grad, laplacian`
- Other: `emptyset`

Discussion