CDI-Type II: Resolving Computational Intensity to Understand Spatially Explicit Dynamics of Food Supply in Coupled Crop-Land Systems

Week of April 13, 2009
Major Homework – Getting Scoped Better and More Specific

- Domain-specific case study scenarios enabled by
  - Preferably representative of
    - Local
    - Regional
    - Global

- Parallel computing focus
  - More specific
  - Preliminary results
The Upper Mississippi River Basin

- Drains approximately 189,000 square miles, including large parts of the states of Illinois, Iowa, Minnesota, Missouri, and Wisconsin. Small portions of Indiana, Michigan, and South Dakota are also within the basin.
- More than 30 million people live in the basin.
- The basin has 30,700 miles of streams.
- Over 60 percent of the basin is cropland or pasture.
- There are over 3,000 reservoirs in the basin.

Concerning Food Supply

- The Upper Mississippi River System is the only water body in the nation that has been recognized by Congress as a "nationally significant ecosystem and a nationally significant commercial navigation system." (Section 1103 of the Water Resources Development Act of 1986, P.L. 99-662)

- In 2000, barges transported 122 million tons of commodities on the Upper Mississippi River, over half of which was grain for world export.

- Approximately 52 percent of the nation’s corn and 41 percent of the nation’s soybean exports are carried on the Upper Mississippi River System, which includes the Illinois River.

Spatially-Explicit Dynamics

- Crop growth
  - Climate/environmental changes
- Hydrological
  - Water quantity and quality
- Global food-water dynamics
  - IMPACT-WATER
    - Economics
- Human/environment
  - Agent-based models
Transition to Computational Thinking

- **Optimize**
  - Food production
  - Land use
  - Etc.

- **Simulate**
  - Spatial synthesis

- **Discover spatial patterns**
  - Massive geographic data
  - GIS

- Crop growth
  - Climate/environmental changes

- Hydrological
  - Water quantity and quality

- Global food-water dynamics
  - IMPACT-WATER

- Human/environment
  - Agent-based models
Formulation of Food Supply Optimization Problem

\[
Y = \max \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij}x_{ij} \quad \text{(Profit objective: yield)}
\]

\[
R = \min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}x_{ij} \quad \text{(Cost objective: risk)}
\]

\[
\min(-w_1 Y + w_2 R) \quad \text{(Global objective: weighted sum)}
\]

such that:

\[
\sum_{j=1}^{n} d_{ij}x_{ij} \leq a_i, \ i = 1..m \quad \text{(crop capacity constraint)}
\]

\[
\sum_{j=1}^{n} e_{ij}x_{ij} \leq b_i, \ i = 1..m \quad \text{(virtual water constraint)}
\]

... ... (soil constraint)
... ... (CO2 constraint)
... ... (OZone constraint)
... ... (Human constraint)

\[
\sum_{i=1}^{m} x_{ij} \leq 1, \ j = 1..n \quad \text{(crop selection constraint)}
\]

\[x_{ij} \in \{0, 1\}, \ i = 1..m, \ j = 1..n\]

Where:

\( p_{m \times n} \): crop yield matrix. Objective: maximize the sum of each land parcel’s yield

\( e_{m \times n} \): risk matrix. Objective: minimize the sum of each land parcel’s risk

\( d_{m \times n} \) and \( a_i, i = 1..m \): crop capacity constraint

\( e_{m \times n} \) and \( b_i, i = 1..m \): virtual water constraint
Computational Challenges

- **Scale**
  - Upper Mississippi River Basin covers 189,000 square miles
    - 189,000 1mile*1mile land parcels
    - 189,000 problem variables
  - 30,700 streams provide 40,000,000 acre-feet volume of water
  - While the number of crops is small, each crop has a set of types
  - ...

- **Complexity**
  - Some constraints are capacity constraints (e.g., crop capacity, water)
    - Each capacity constraint forms a bin, each variable is an item
    - Generalized Assignment Problem (GAP)
  - It belongs to the class of knapsack problems (NP-hard)
Evolutionary Algorithm Approach

- **Sequential algorithm**
  - 2-D encoding requires spatially-explicit GA operator
    - Selection, crossover, mutation, fitness evaluation
  - Infeasible solution handling
    - Repair operator
    - Penalty-based fitness evaluation

- **Parallel evolutionary algorithms**
  - Global parallelization
    - Parallelize fitness evaluation function
  - Dividing population
    - Island model: loosely-connected islands (subpopulations) via migration
    - Cellular evolutionary algorithms: singleton population on each processor
GAPPGA: Sequential Algorithm Performance

- So far the best sequential genetic algorithm for GAP (OR-LIB benchmark datasets)

<table>
<thead>
<tr>
<th>Prob. Type</th>
<th>Size</th>
<th>IP/LP</th>
<th>CPLEX</th>
<th>CPLEX %-gap</th>
<th>CRH-GA (best sol'n)</th>
<th>Our version (best sol'n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t[s]</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>5</td>
<td>25670</td>
<td>0.00</td>
<td>0.44</td>
<td>3495381 3609.7</td>
<td>25722 0.20 4646590</td>
</tr>
<tr>
<td>D</td>
<td>10</td>
<td>25670</td>
<td>0.18</td>
<td>1.57</td>
<td>4061620 7042.7</td>
<td>25445 0.67 10136804</td>
</tr>
<tr>
<td>D</td>
<td>20</td>
<td>25670</td>
<td>0.51</td>
<td>1.97</td>
<td>4041905 13049.6</td>
<td>24870 1.32 75665408</td>
</tr>
<tr>
<td>D</td>
<td>40</td>
<td>6092</td>
<td>3.96</td>
<td>3.72</td>
<td>1361463 2072.9</td>
<td>6257 2.71 1477306</td>
</tr>
<tr>
<td>D</td>
<td>100</td>
<td>6092</td>
<td>2.22</td>
<td>3.06</td>
<td>2357294 7116.2</td>
<td>62493 2.03 4487719</td>
</tr>
<tr>
<td>D</td>
<td>200</td>
<td>6092</td>
<td>1.10</td>
<td>2.81</td>
<td>3574368 23374.1</td>
<td>24877 2.07 3550984</td>
</tr>
<tr>
<td>D</td>
<td>80</td>
<td>6110</td>
<td>6.60</td>
<td>7.01</td>
<td>846320 2878.6</td>
<td>6493 6.26 3468876</td>
</tr>
<tr>
<td>D</td>
<td>100</td>
<td>6110</td>
<td>2.87</td>
<td>3.76</td>
<td>230429 13075.6</td>
<td>12427 2.43 3697777</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>7757</td>
<td>0.00</td>
<td>0.24</td>
<td>552212 147.1</td>
<td>7760 0.04 6611226</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>15611</td>
<td>0.00</td>
<td>0.23</td>
<td>1281834 669.8</td>
<td>15929 0.12 1179025</td>
</tr>
<tr>
<td>E</td>
<td>20</td>
<td>30794</td>
<td>0.00</td>
<td>0.28</td>
<td>2744401 3026.7</td>
<td>30844 0.16 1686314</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>7387</td>
<td>0.61</td>
<td>0.91</td>
<td>1005960 447.3</td>
<td>7437 0.67 1272077</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>15039</td>
<td>0.00</td>
<td>0.96</td>
<td>1807355 1557.2</td>
<td>15128 0.59 2659767</td>
</tr>
<tr>
<td>E</td>
<td>100</td>
<td>29977</td>
<td>0.00</td>
<td>0.94</td>
<td>3360020 6136.3</td>
<td>30158 0.60 5581238</td>
</tr>
<tr>
<td>E</td>
<td>200</td>
<td>7348</td>
<td>1.32</td>
<td>1.74</td>
<td>1458153 1711.4</td>
<td>7425 1.05 3252759</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>14765</td>
<td>0.89</td>
<td>1.70</td>
<td>1955937 3119.4</td>
<td>14903 0.93 3150038</td>
</tr>
<tr>
<td>E</td>
<td>100</td>
<td>29500</td>
<td>0.34</td>
<td>1.60</td>
<td>4525547 14462.3</td>
<td>29789 0.98 6719421</td>
</tr>
<tr>
<td>E</td>
<td>200</td>
<td>7387</td>
<td>3.32</td>
<td>3.11</td>
<td>976572 1573.5</td>
<td>7486 2.32 1478596</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>14630</td>
<td>1.85</td>
<td>2.20</td>
<td>1889566 6061.5</td>
<td>14856 1.54 4633217</td>
</tr>
<tr>
<td>E</td>
<td>100</td>
<td>29186</td>
<td>0.69</td>
<td>2.14</td>
<td>2977713 20536.8</td>
<td>29575 1.33 3706508</td>
</tr>
<tr>
<td>E</td>
<td>200</td>
<td>7316</td>
<td>2.00</td>
<td>3.11</td>
<td>737647 2614.1</td>
<td>7666 0.21 1335258</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>14566</td>
<td>2.17</td>
<td>2.93</td>
<td>1584730 11078.3</td>
<td>14873 2.10 3665464</td>
</tr>
<tr>
<td>E</td>
<td>100</td>
<td>29161</td>
<td>1.57</td>
<td>2.49</td>
<td>3063944 56250.8</td>
<td>29727 1.94 2054368</td>
</tr>
<tr>
<td>F</td>
<td>5</td>
<td>2755</td>
<td>0.00</td>
<td>0.41</td>
<td>507714 123</td>
<td>2755 0.00 880156</td>
</tr>
<tr>
<td>F</td>
<td>5</td>
<td>5294</td>
<td>0.00</td>
<td>0.35</td>
<td>647177 293.2</td>
<td>5304 0.19 279463</td>
</tr>
<tr>
<td>F</td>
<td>10</td>
<td>10745</td>
<td>0.00</td>
<td>0.25</td>
<td>1235605 1130</td>
<td>10755 0.09 3252880</td>
</tr>
<tr>
<td>F</td>
<td>10</td>
<td>2276</td>
<td>1.99</td>
<td>3.95</td>
<td>1019383 386.4</td>
<td>2357 3.52 955639</td>
</tr>
<tr>
<td>F</td>
<td>100</td>
<td>4644</td>
<td>1.13</td>
<td>3.12</td>
<td>1565513 1127.4</td>
<td>4748 2.23 2706156</td>
</tr>
<tr>
<td>F</td>
<td>100</td>
<td>9372</td>
<td>0.46</td>
<td>2.78</td>
<td>2897607 4207.9</td>
<td>9503 1.39 7147851</td>
</tr>
<tr>
<td>F</td>
<td>20</td>
<td>2145</td>
<td>8.15</td>
<td>8.38</td>
<td>1514783 1010.3</td>
<td>2264 5.54 2811951</td>
</tr>
<tr>
<td>F</td>
<td>200</td>
<td>4310</td>
<td>4.73</td>
<td>6.85</td>
<td>2227943 2878.4</td>
<td>4557 5.73 4424268</td>
</tr>
<tr>
<td>F</td>
<td>200</td>
<td>8479</td>
<td>2.38</td>
<td>7.15</td>
<td>3864825 10390</td>
<td>8993 6.06 10054778</td>
</tr>
<tr>
<td>F</td>
<td>400</td>
<td>2110</td>
<td>21.28</td>
<td>18.27</td>
<td>1191719 1572.9</td>
<td>2462 16.68 4757323</td>
</tr>
<tr>
<td>F</td>
<td>400</td>
<td>4086</td>
<td>10.14</td>
<td>12.86</td>
<td>2461344 6228.2</td>
<td>4507 10.29 2806139</td>
</tr>
<tr>
<td>F</td>
<td>400</td>
<td>8274</td>
<td>4.05</td>
<td>10.36</td>
<td>3676777 19192.9</td>
<td>9027 9.10 5388651</td>
</tr>
<tr>
<td>F</td>
<td>800</td>
<td>2064</td>
<td>31.37</td>
<td>26.77</td>
<td>1620919 3640.1</td>
<td>2583 25.12 710516</td>
</tr>
<tr>
<td>F</td>
<td>800</td>
<td>4123</td>
<td>19.05</td>
<td>17.33</td>
<td>1738801 9828.6</td>
<td>4754 15.29 3778932</td>
</tr>
<tr>
<td>F</td>
<td>800</td>
<td>8167</td>
<td>9.18</td>
<td>12.55</td>
<td>2666127 32432.1</td>
<td>9230 12.68 2780104</td>
</tr>
</tbody>
</table>
GAPPGA: Parallel Implementation

- **MPI implementation**
  - Island model: each processor handles a subpopulation
  - Asynchronous communication for migration
    - MPI_Ibsend(), MPI_Iprobe(), MPI_recv()
  - Network topology: regular grid

- **Experiments**
  - TeraGrid clusters: Mercury and Abe
  - Up to 2048 cores
  - Communication overhead is marginal: < 0.05%
GAPPGA: Preliminary Results
Formulation of Food Supply Simulation
Real World Factors Implying Significant Computational Intensity

- Spatial extent: 800 km x 1,500 km
- # parcel: 4,000 x 7,500 = 30,000,000 (upper limit)
  - Average parcel size: 200 m x 200 m (Google map)
  - Including urban area and water
- # counties: 500
  - Use this to determine # organizations/policymakers at the county level
- # population: 30 million
Computational Intensity of SAM – Inherently Spatial

A pattern snapshot of a parallel mobile agent-based model; left: a geospatial domain overlaid with the structure of a computational domain; middle: the corresponding computational intensity map; right: processor pattern (Compu-Cells: computational intensity map cells; CI: normalized computing time-based computational intensity, a function of agent and environment characteristics.)
Performance Evaluation

Landscape Size: \(51,200 \times 51,200\); #Agents: 3,276,800
Sequential time: 1126.25 seconds
Computing resources: Mercury @ NCSA
- 1.3/1.5 GHZ and 4/12 GB RAM per processor
- 1774 processors
Rectilinear domain decomposition

Decompose environment into rectangular sections where each processor owns a unique rectangular section of the environment and all agents residing on that section. Similar to cartesian (uniform) grid, except uniform spacing is *not* required.

Properties
- Linearly aligned
- Rectangular area
- Flexible granularity
Spatial Interpolation and Pattern Detection

Analysis execution management
Domain decomposition
Query pattern detection

ComputationalIntensity map

Computing control
InputSplit control
Task scheduling
MapReduce

Spatial data control
Bigtable composition
Tablet allocation
Bigtable
Hadoop

Storage control
Partitioning & allocation
Replication
GFS

Hadoop
Deliverables

- Coupled modeling for global food supply prediction
- Spatially explicit multi-scale modeling and optimization
- Incorporation of social and environmental dimensions

------------------------
SAM-----------------------------------

- Computational intensity map
- A unified parallel computing framework
  - Application-level parallel primitives
  - Synthesize data- and compute-driven parallel paradigms

-------------------------
GIS-----------------------------------

- GISolve-Food
  - Cyberenvironment
  - GISolve
  - EOT
Discussions for the Week of April 20

- Writing tasks of individual team members
- Info needed from each individual team member
  - 2-page bio
  - Current & pending support