Continental-scale remote monitoring of invasive species dynamics through petascale high performance computing system

Chunyuan Diao

Assistant Professor
Department of Geography and GIScience
University of Illinois at Urbana-Champaign
I use Blue Waters to prototype a parallel computational framework to handle massive amount of satellite data for large-scale invasive species monitoring.
Saltcedar is an exotic shrub species invading riparian zones of the United States

- Alter stream hydrology
- Increase soil salinity
- Degrade habitats for native species

Annual economic losses from saltcedar in the US are estimated to be $133-285 million
Remote Sensing

- Machine/deep learning to map saltcedar distribution
Saltcedar Phenology

Detect saltcedar at leaf coloration stage

Saltcedar Phenology

Satellite time series

Challenges

1) Leaf coloration timing cannot be predicted using current phenological models

![Satellite time series with images for Jan. to Dec.]

2) Massive volume of satellite data cannot be adequately handled by traditional remote sensing systems

![Graph showing temporal and spatial dimensions]
Objective

- Develop a parallel computational framework to model the spatio-temporal dynamics of saltcedar over the past 40 years

1) Develop computational algorithms that can model the leaf coloration stage of invasive saltcedar using satellite time series

2) Devise a high-performance parallel system to prototype the data- and compute-intensive satellite invasive species monitoring system
Parallel computational framework

1. Leaf coloration computational algorithms

2. High-performance parallel system
Computational algorithms

- To model and predict the timing of saltcedar coloration


To model the timing of saltcedar coloration with sparse time series

Multiyear Spectral Angle Clustering Model

- Landsat time series (Fall phenology, 2003)
- Landsat time series (Fall phenology, 2004)
- Landsat time series (Fall phenology, 2005)

Component 1:
- Time series spectral outlier removal (Angle-based outlier detection)
- Pre-processed Landsat time series of spectral signatures

Component 2:
- Time series spectral clustering (Cosine distance-based k-means clustering)
- Synthesized fall phenology across three years

Component 3:
- Time series spectral matching (Spectral angle mapper-based moving average)
- Phenological transition date

Composite image (2004)
To model the timing of saltcedar coloration with sparse time series
Multiyear Spectral Angle Clustering Model

1) **Time series spectral outlier removal**
   (Angle-based outlier detection method)

   Time series of spectral signature
   - Fall Phenology 2003
   - Fall Phenology 2004
   - Fall Phenology 2005

2) **Time series spectral clustering**
   (Cosine distance-based k-means clustering)

   Synthesized Fall Phenology
   Time series of spectral signature

Leaf Coloration?
3) **Time series spectral matching**
(Spectral angle mapper-based moving average method)

![Diagram showing time series spectral matching and leaf coloration](image)

**Leaf Coloration**

**Synthesized Fall Phenology**

- **Time series of spectral signature**
- **Matching Score**

Mathematical expressions for MSAC_{SM1}\(t\) and MSAC_{SM2}\(t\):

\[
MSAC_{SM1}(t) = \frac{\sum_{i=1}^{n_1} SAM(t, t - i)}{n_1}
\]

\[
MSAC_{SM2}(t) = \frac{\sum_{i=1}^{n_2} SAM(t, t + i)}{n_2}
\]

**Day of Year**

- **2003**
- **2004**
- **2005**

**Spectral Matching Score**

**12/08/2004**

- **12/08/2004**

- **Leaf Coloration**
To model the timing of saltcedar coloration with dense time series
Network representation of saltcedar phenological progress
- Node: spectral reflectance obtained on each date of the time series
- Edge: spectral similarity between the spectral nodes

Pheno-network with three groups, namely the pre-transition, transition, and post-transition groups.
Network measures of saltcedar leaf coloration

- Betweenness Centrality: the transition node serves as the hub connecting the nodes across phenological stages
- Clustering Coefficient: the neighbors of the transition node are sparsely connected to each other

\[
\text{Bridging Coefficient} = \frac{\text{Betweenness Centrality}}{\text{Clustering Coefficient}}
\]
Composite Landsat Image

Overall Accuracy: 81.25%
Kappa: 0.65
Producer’s Accuracy: 76%
User’s Accuracy: 83%

Overall Accuracy: 74.25%
Kappa: 0.49
Producer’s Accuracy: 66%
User’s Accuracy: 79%

Image acquisition date at leaf senescence (2004)
Parallel computational framework

1. Leaf coloration computational algorithms

2. High-performance parallel system
Conventional remote sensing system

- Conventional remote sensing systems analyze entire remote sensing imagery as a whole
  - High memory requirements and low scalability

- Large-scale remote sensing monitoring is challenging
  - Massive amount of satellite imagery
  - High demands for computational resources
The leaf coloration algorithms are designed at the pixel level.

The parallel system decomposes the remote sensing imagery into a multitude of sub-tiles:
- Reduce memory requirement
- Optimize I/O and computation time
High performance parallel system

The parallel system adopts hybrid computation models

- Node-level data distribution model – MPI
- Core-level computation model - OpenMP
The two-level data distribution model: massive data and I/O operations are evenly distributed among all computing nodes.
Parameter calibration in pheno-network models

Core-level computation model increases computation efficiency while decreasing memory requirement.
Blue Waters facilitates the processing of massive amount of satellite data with high spatial, temporal and spectral dimensions

- Large storage space
- Access to a large number of nodes
- High-speed simultaneous access to a large number of images
- Large network bandwidth to increase data distribution speed
Scalability of parallel system

![Graph showing the scalability of a parallel system with Blue Waters as the dataset, plotting machine time (sec) against the number of nodes. The graph indicates a decrease in machine time as the number of nodes increases.]
Saltcedar distribution map in 2005
<table>
<thead>
<tr>
<th>Class</th>
<th>1985</th>
<th>1995</th>
<th>2005</th>
<th>Overall Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saltcedar (ha)</td>
<td>4497</td>
<td>5027</td>
<td>5607</td>
<td>+1110</td>
</tr>
<tr>
<td>Native woody riparian</td>
<td>2352</td>
<td>2201</td>
<td>2021</td>
<td>-331</td>
</tr>
<tr>
<td>vegetation (ha)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (ha)</td>
<td>13743</td>
<td>13364</td>
<td>12964</td>
<td>-779</td>
</tr>
</tbody>
</table>
Conclusions

1) The multiyear spectral angle clustering and phenonetwork models can model the leaf coloration stage of invasive saltcedar

2) The high performance parallel system can efficiently process massive satellite time series with high scalability

3) Invasive saltcedar is displacing native riparian vegetation over time and space


Acknowledgement

- This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.

- This research is partially supported by NSF Office of Advanced Cyberinfrastructure award and Campus Research Board at the University of Illinois at Urbana-Champaign.
Thank You!