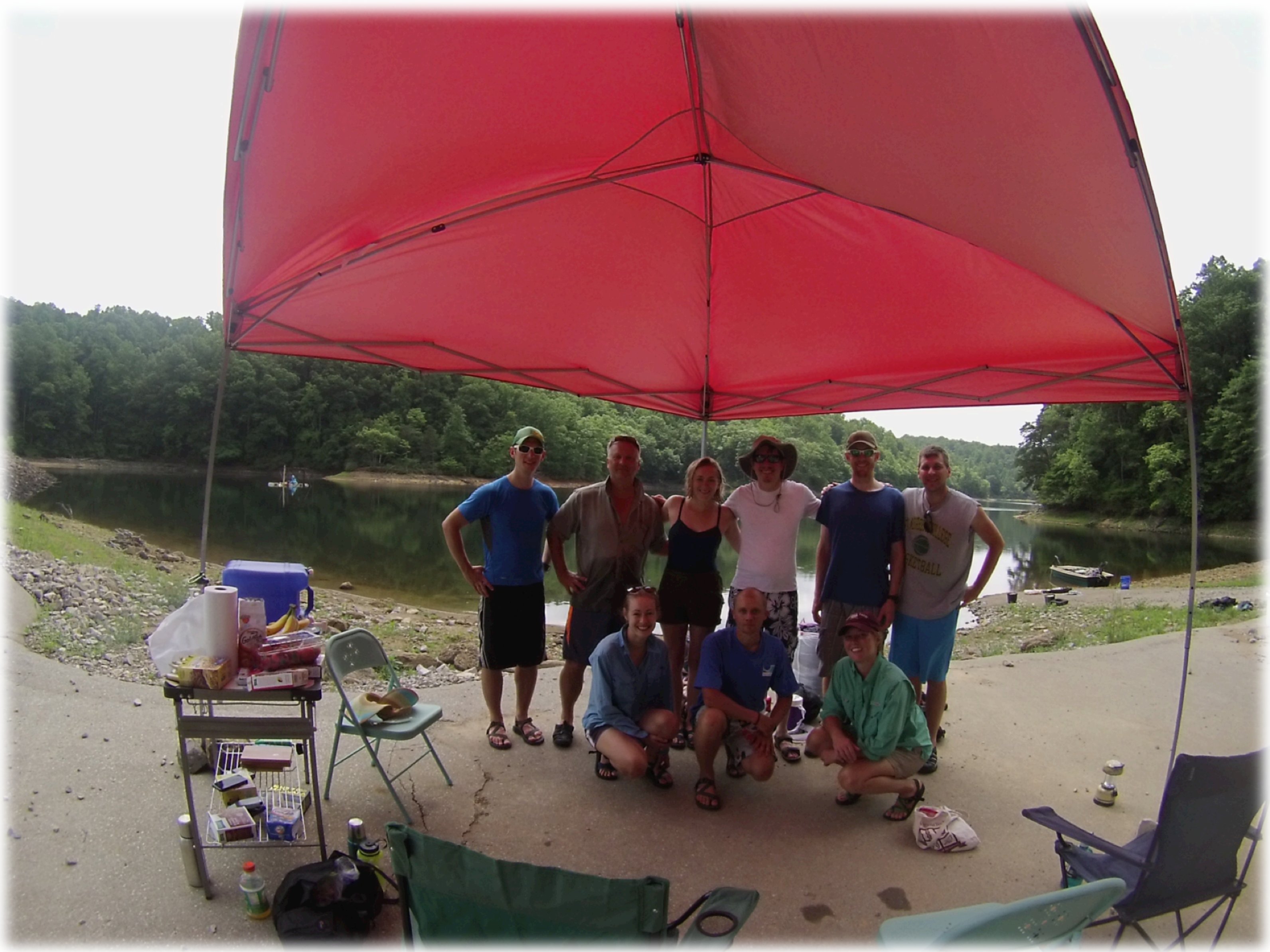
PRAGMA-GLEON Expedition (GRAPLE) - Lake Science on Overlay-networked Resources

Paul Hanson (University of Wisconsin)

Cayelan Carey (Virginia Tech)

Renato Figueiredo, Ken Subratie, Saumitra Aditya (University of Florida)

Ecologists in their natural environment

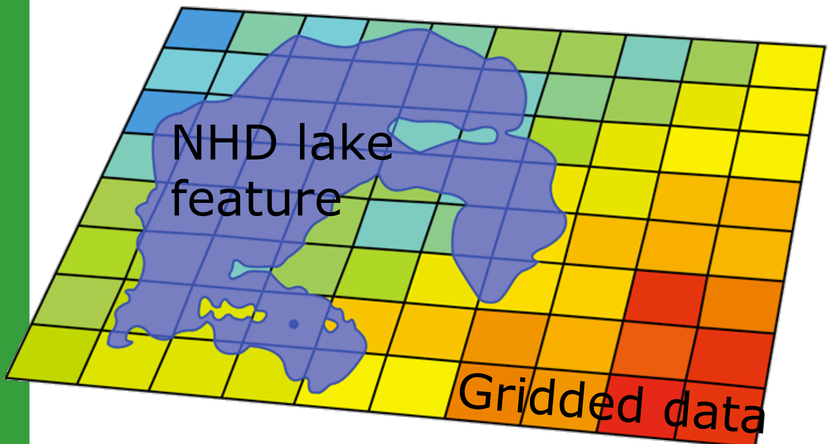


Questions like this…

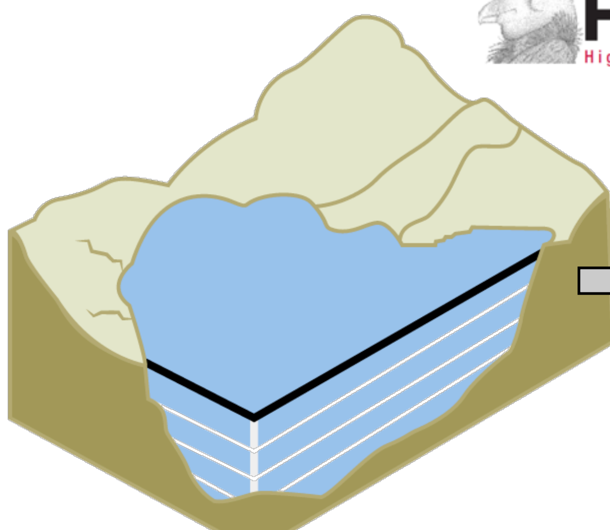
How will the availability and quality of water respond to changes in land use and climate?

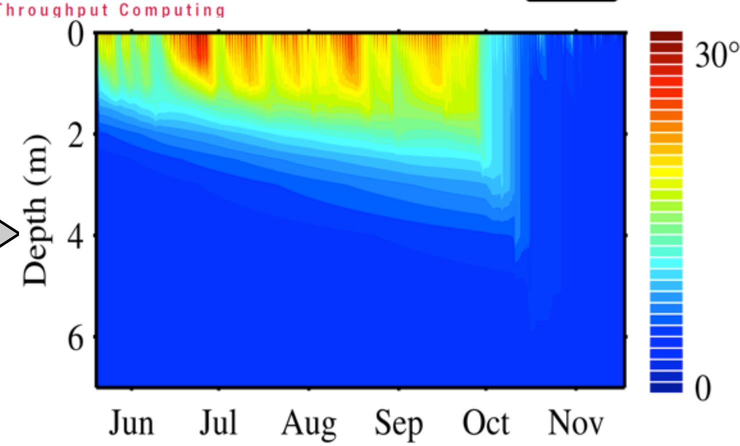
Often require…

...approaches like this.



- Parameterization of models with gridded data
- Dynamic lake temperature simulations
  - Parallel computing with HTCondor High Throughput Computing





General Lake Model: U of W AUS

(*Read et al. 2014*)

Our solution?

Landscape of numerical simulation

UWA
Matt Hipsey

Simulation software development

USGS-CIDA
Jordan Read
Luke Winslow

R-based GLM tools

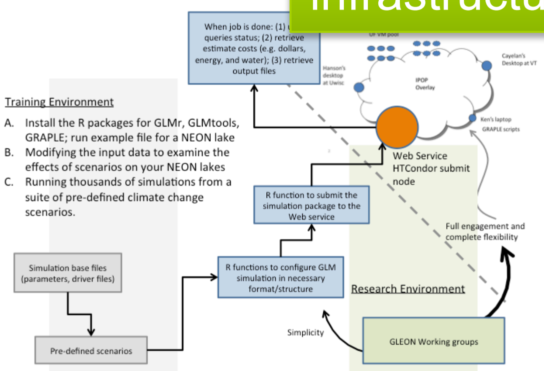
GitHub

PRAGMA

Compute infrastructure

Papers, sample parameter sets

Training & education



User community
Waikato U. (NZ)
U. Adelaide (Australia)
U. Wisconsin (USA)
Virg. Tech (USA)
NYC DEP
Cornell
USGS
UWA

Outcomes of the activities

1. Basic research in science
2. A new kind of science for ecologists
3. Water resource management
4. Education and training (project EDDIE)
5. Data standardization, workflows
6. Community development
7. Cross-discipline training

Toxic Algae Outbreak Overwhelms a Polluted Ohio River

By MICHAEL WINES SEPT. 30, 2015

Email

Share

Tweet

Save

More

The Ohio River, transformed by mining and industrial waste and sewage overflows into the nation’s most polluted major waterway, has a new and unexpected tormentor this fall: carpets of poisonous algae.

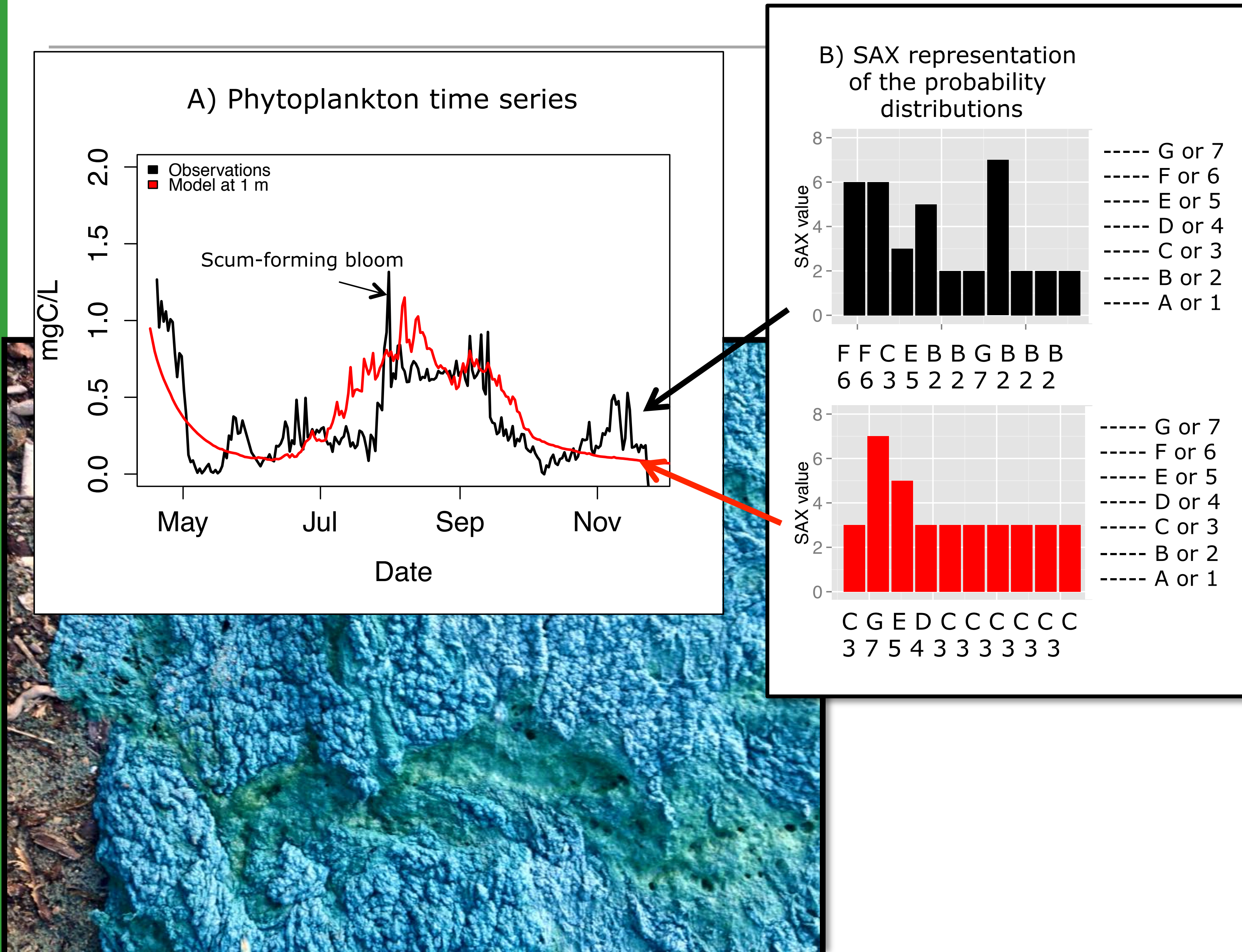
Pads of toxic blue-green algae have speckled nearly two-thirds of the 981-mile river in the last five weeks, experts say, in an outbreak that has curbed boating, put water utilities on alert and driven the river’s few hardy swimmers back to shore.

The only other recorded toxic algae bloom, in 2008, covered perhaps 40 miles of the river. In contrast, the latest bloom stretches 636 miles from Wheeling, W.Va., to Cannelton, Ind., and traces of algae have appeared as far west as Illinois.

The poisonous alga, called microcystis,

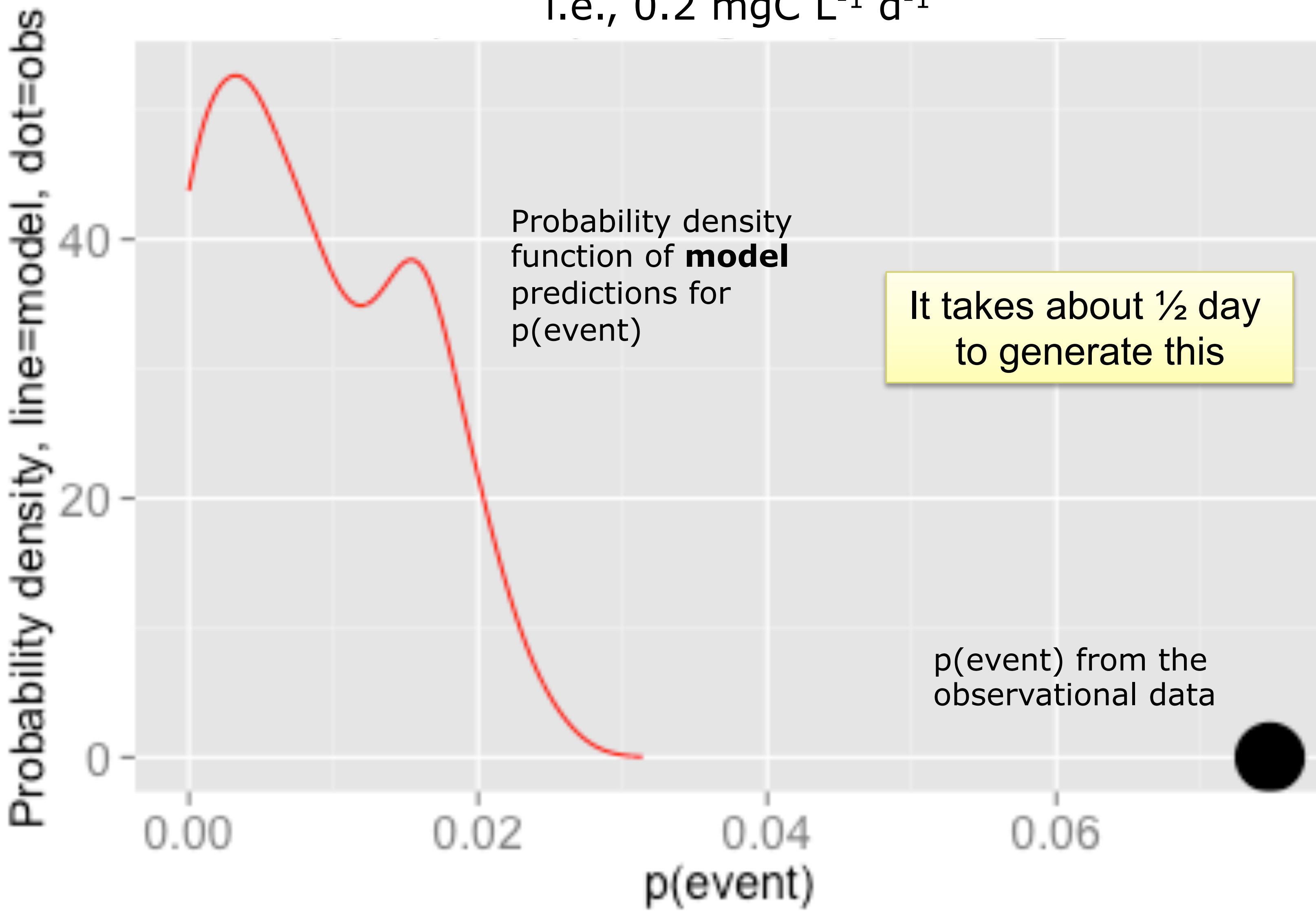
A sample of water with algae from Lake Erie near Curtice, Ohio, in July. The latest bloom of algae stretches along 636 of the Ohio River’s 981 miles.

Eric Albrecht/The Columbus Dispatch, via Associated Press


A) Phytoplankton time series
Observations
Model at 1 m
Scum-forming bloom
mgC/L
0.0
0.5
1.0
1.5
2.0
May
Jul
Sep
Nov
Date
B) SAX representation of the probability distributions
SAX value
0
2
4
6
8
----- G or 7
----- F or 6
----- E or 5
----- D or 4
----- C or 3
----- B or 2
----- A or 1
F F C E B B G B B B
6 6 3 5 2 2 7 2 2 2
SAX value
0
2
4
6
8
----- G or 7
----- F or 6
----- E or 5
----- D or 4
----- C or 3
----- B or 2
----- A or 1
C G E D C C C C C C
3 7 5 4 3 3 3 3 3 3

What is the probability of a bloom

i.e., 0.2 mgC L^{-1} d^{-1}



Goals for supplemental proposal:

1. Make available and usable new services, software and approaches to freshwater scientists
2. Expand the overlay technology beyond the existing PRAGMA-GLEON expedition
3. Grow the next generation of users by training students how to use GLM-AED to study lake water quality with hands-on modules
4. Assess the efficacy of integrated technology-science-education development in the classroom and in GLEON working groups

New PRAGMA Partners

- Yoshio Tanaka (Information Technology Research Institute, AIST), co-lead of PRAGMA Resources Working Group
- Shinji Shimojo (Osaka University)
- Weicheng Huang (National Center for High-performance Computing), who has interest in the coordination of PRAGMA testbeds and proposed simulations and applications
- Kohei Ichikawa (NAIST), with interest in interconnecting IPOP network with OpenFlow network to expand the available resources

Lake Taihu, China Photo: C. Carey

Technology

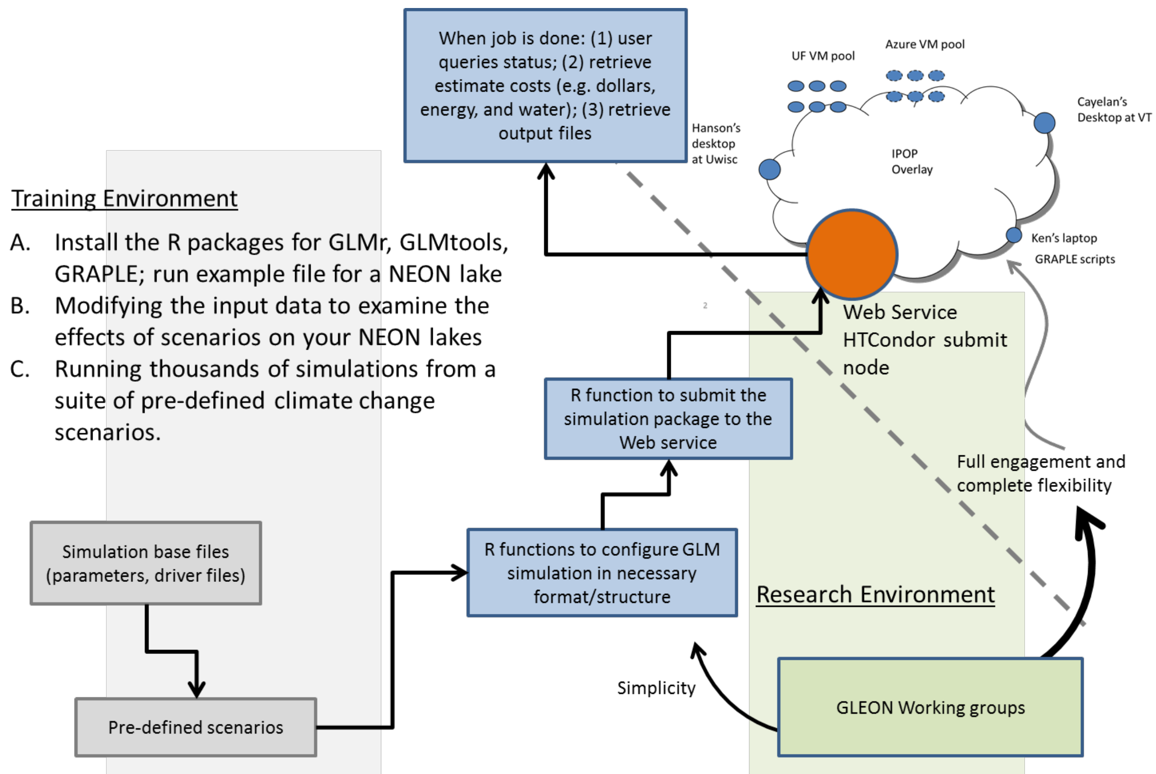
- Distributed resources are available
  - Desktops, servers, cloud; how to aggregate in user-friendly manner, with low administrative overhead?
- Local, familiar computer environments
  - Windows/MacOS, GLM, R
  - Present user’s preferred environment for ease of use
- Technical approach
  - IPOP: overlay automatically creates virtual network; only requirement is to run software on endpoints
  - HTCondor: high-throughput computing job manager
  - GRAPLEr: Web interface conveniently accessible from a user’s desktop – R

IP-over-P2P (IPOP) Overlay

- "User-Defined Networking"
  - SDN that users can deploy
    - User-level; easy to use
  - Software only needed at endpoints
    - Cloud VMs, personal computers, mobile devices
  - Map social network to virtual network links
    - E.g. GRAPLE researchers, students
  - Peer-to-peer VPN links
    - No centralized gateway
- Learn more about the open-source software:
  - http://ipop-project.org
  - Demo

Technology Updates

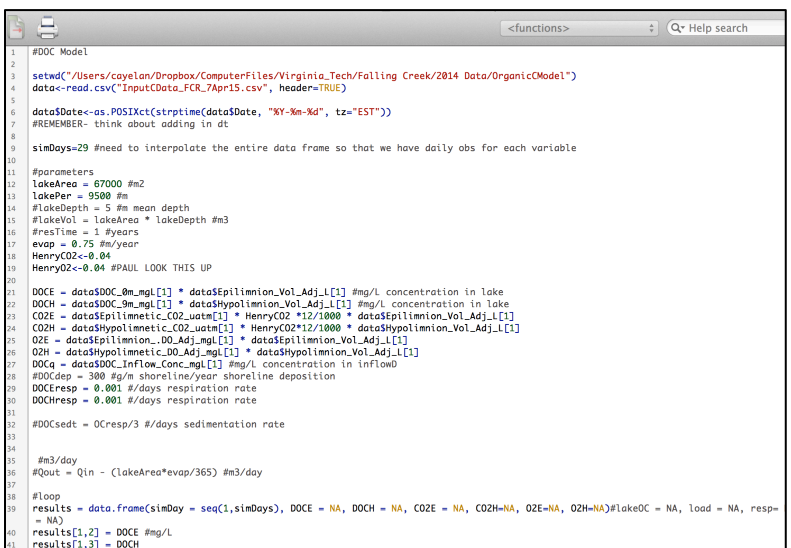
- GRAPLEr Web Service
  - Expose simple R desktop interface to users
  - Handle packaging of multiple simulations into Condor job, compression, and job submission



Demonstration

Expanding GRAPLE globally

- We have developed GRAPLEr scripts as part of a teaching module to run the lake model in undergraduate and graduate classrooms
- Students manipulate climate and land use scenarios and then run the lake model to study their effects *in silico*



Next steps

- The first cohort of students at GLEON will be formally assessed to track their responses to using GLM and Web service + overlay technology
  - Monday workshop: Cayelan + Paul + Renato
- This module will be disseminated to 8 universities across the U.S. for testing in the fall and will be released publicly in 2016
- We have partnered with pedagogical experts to study how non-CS students use this technology → will enable broad dissemination

Acknowledgments

- GRAPLE students
  - Craig Snortheim, Jon Doubek, Alex Gerling, Kate Hamre, Ryan McClure
- IPOP developers
  - Ken Subratie, Kyuho Jeong, Saumitra Aditya
- This material is based upon work supported in part by the National Science Foundation under Grants No. 1234983 and 1339737

Number of simulations in an experiment

Duration of the simulation

How frequently predictions are recorded

Output file size

Total run time

nSims	Duration	Resultion	Output	Runtime
1000	9 mo	Daily	0.01 GB	140 min
1001	9 mo	8 hour	0.12 GB	200 min
1000	9 mo	1 hour	1.00 GB	700 min

Too much for me and my laptop