On Lakes and Clouds: A Retrospective on the PRAGMA/ GLEON Lake Expedition

Renato Figueiredo
University of Florida

Cayelan C. Carey
Virginia Tech

Paul C. Hanson
University of Wisconsin

Why a Lake Expedition?

Lakes and Reservoirs - Critical to Sustaining Humans

- Food: Available freshwater will limit food production for a growing world population
- Water: Most of the available and manageable drinking water is in lakes and reservoirs
- Climate change: Organic carbon locked up in the terrestrial biosphere is in peat and lake sediments

What is the future of freshwater quality and quantity in lakes and reservoirs around the globe?

Lake science + Cyber-infrastructure

- Access and use of data
  - Requires connecting networks
  - Automation, workflows - connection to models
- Computational thinking and practice
  - Leveraging cyber-infrastructure resources and expertise
  - Prioritization of technology development
- A science work force
  - Inherently inter-disciplinary teams: CS and lake science

Lake expedition origins

- GLEON (Global Lake Ecological Observatory Network)
  - “conducts innovative science by sharing and interpreting high-resolution sensor data to understand, predict and communicate the role and response of lakes in a changing global environment”
- PRAGMA
  - “community of practice comprising individuals and institutions from around the Pacific Rim that actively collaborate with, and enable, small- to medium-sized groups to solve their problems with information technology"

PRAGMA-26 (Taiwan, 2014)

GLEON-16 (Canada, 2014)

Engaging our team

- In-person meetings: rapport, learning terminology, engaging students, postdocs
  - Workshops (PRAGMA, GLEON, VT, UF, Wisc)
  - Lake sampling-a-thon
  - Video calls between meetings
- Brainstorming, converging on problems of mutual interest

Gainesville, USA, 2017

Nara, Japan, 2015

Chuncheon, South Korea, 2015

Blacksburg, USA, 2015

The First Product - GRAPLEr

GRAPLEr: Motivations

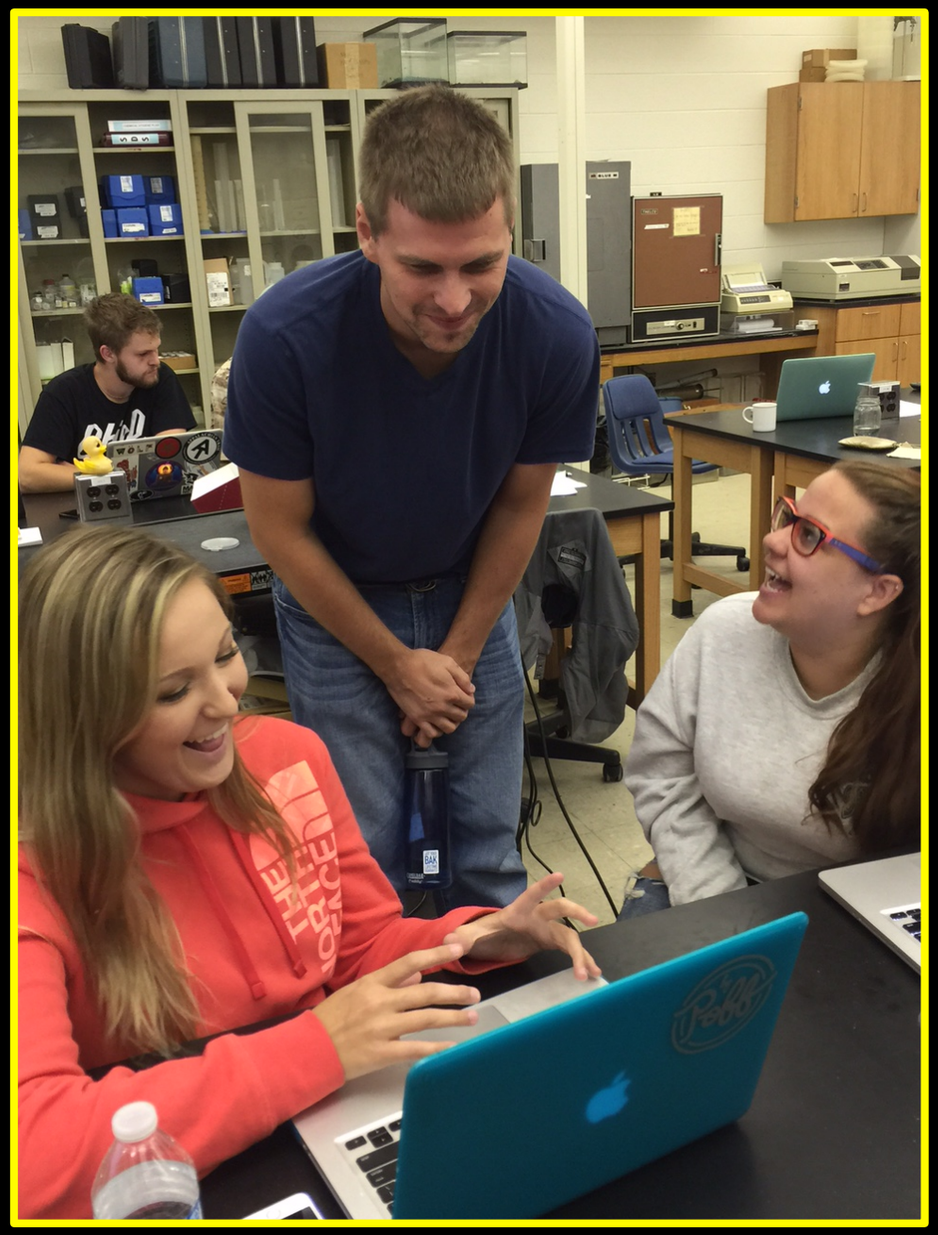
- Data + numeric lake models:
  - Key to understanding future of freshwater quality, quantity in lakes and reservoirs
- Many parameters/scenarios - many model runs
  - Climate change: 100s of thousands to millions
- High-throughput computing not new technology
  - But new to the lake science community

First approach

- Community adopted model: GLM - open-source
- HTCondor solves “run lots of models” problem
- Which resources? Desire to connect resources across PRAGMA - IPOP overlay network
- Integrate IPOP and HTCondor in deployable VMs to create a distributed virtual cluster
  - It works! Done!! Right?

Not so fast…

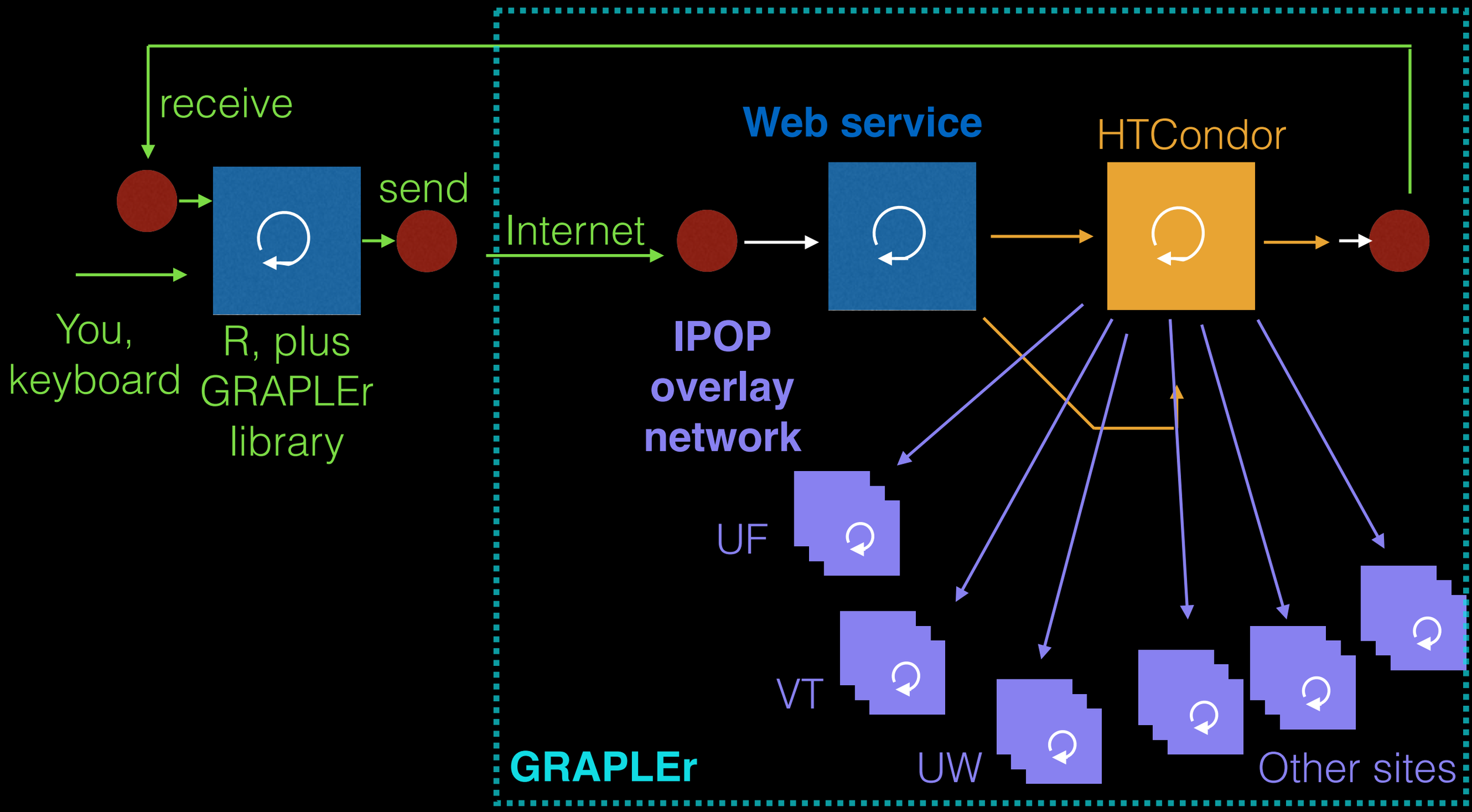
"Does PRAGMA run on my Mac?"



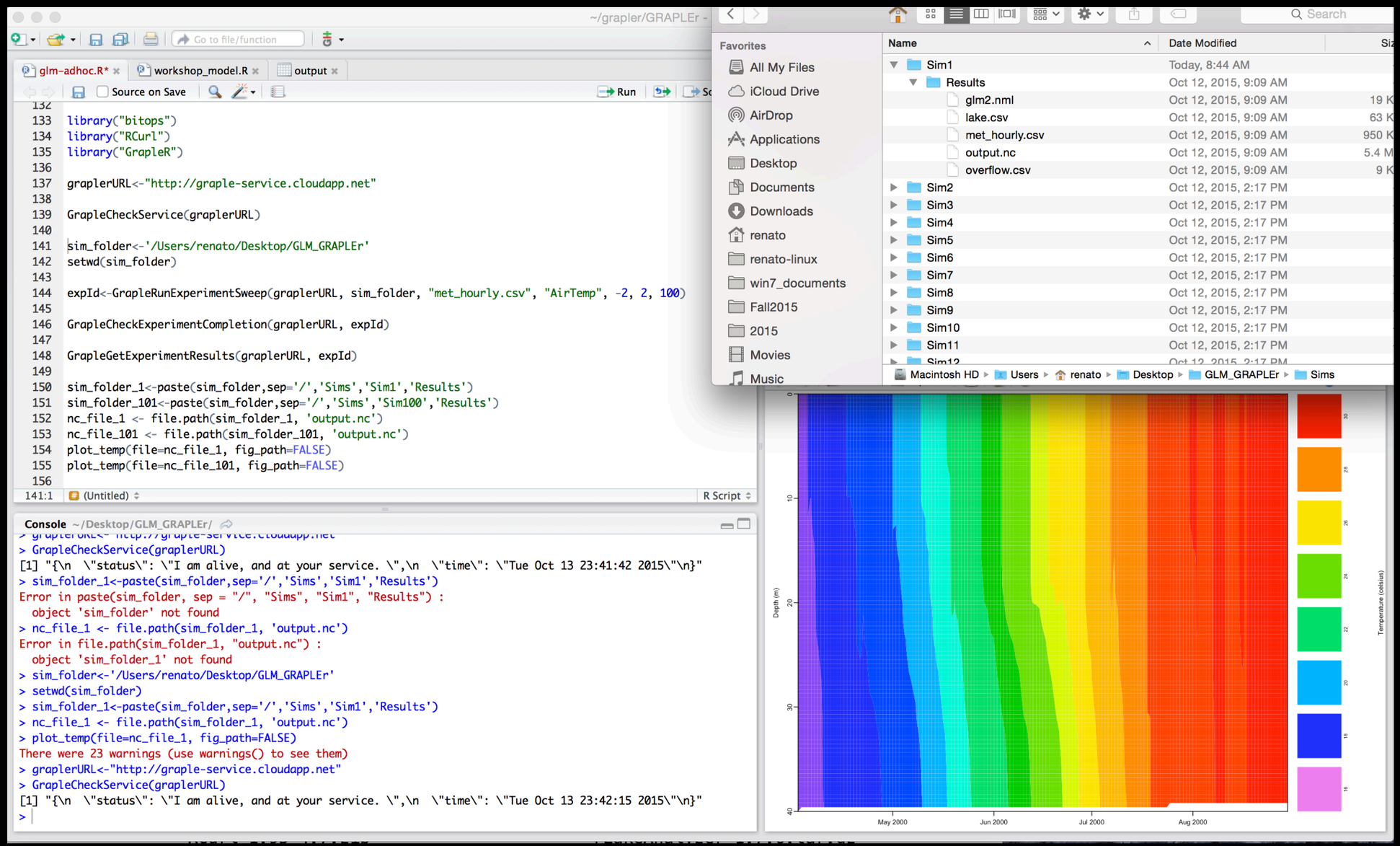
Revisiting GRAPLEr

- Overall back-end architecture - we had that right
- *User-facing front-end had to match familiar environment*
- Required working closely together to communicate goals, requirements, limitations - iterative, collaborative process
  - Web service accessible through R/Rstudio
    - Very simple software installation on client
  - Collaboration in defining interfaces (API)
    - Intuitive to users, yet powerful and expressive

GRAPLEr in a nutshell



GRAPLEr in Action



Local input/ output folders

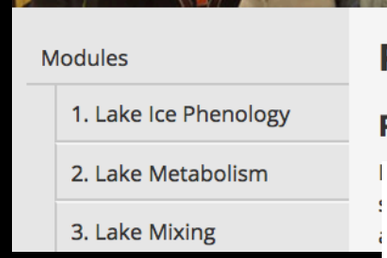
Output visualization

GRAPLEr API calls:
submit, check, download

GRAPLEr - Activities and Impact

Training, Education

- Hands-on workshops at GLEON meetings: G16,17,18
- Project EDDIE: graduate and undergraduate students
- From handful of models on laptop to 100K+ models in the cloud through GRAPLEr
    - Transforming a branch of lake ecology to utilize high-throughput computing




Project EDDIE: Lake Modeling Module

Project Summary

J Sci Educ Technol (2017) 26:1–11
DOI 10.1007/s10956-016-9644-2

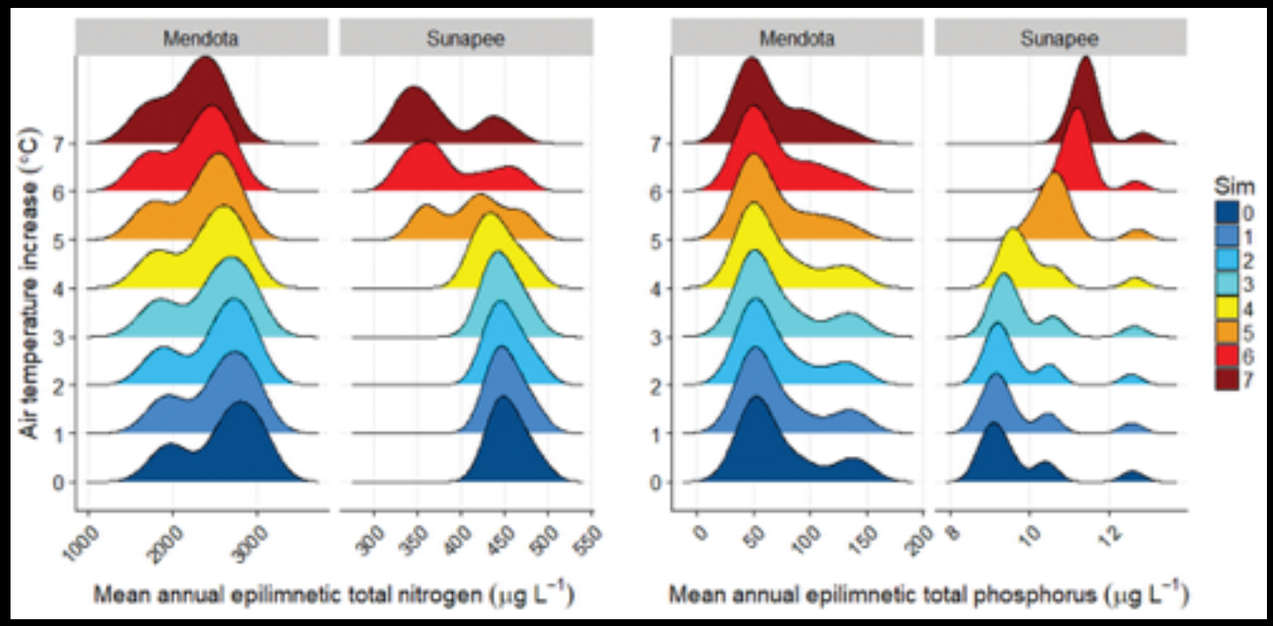
Simulation Modeling of Lakes in Undergraduate and Graduate Classrooms Increases Comprehension of Climate Change Concepts and Experience with Computational Tools

Science

- GRAPLEr scenarios show that low-oxygen conditions in the bottom waters of lakes are sensitive to small changes in weather (Snortheim et al. 2017)

- Thousands of climate scenarios applied to different GLEON lakes reveal that lakes with initially high water quality are more sensitive to climate change than lakes with already-degraded water quality (Farrell et al. 2018)

- Algal blooms are more likely to occur in lakes with greater temperature variability (Krinos et al., in prep)



As air temperatures increase (from 0 to 7°C, shown by the blue to red scale), simulation modeling indicates that while epilimnetic (surface water) total nitrogen concentrations will generally not change, phosphorus concentrations will increase, but only in lakes with initially low levels of nutrients

Publications, presentations

- Kensworth C. Subratie, Saumitra Aditya, Srinivas Mahesula, Renato Figueiredo, Cayelan C. Carey, Paul C. Hanson, “GRAPLEr: A distributed collaborative environment for lake ecosystem modeling that integrates overlay networks, high-throughput computing, and WEB services”, Concurrency and Computation, Apr. 2017

- Farrell, K.J., and C.C. Carey. Power, pitfalls, and potential for integrating computational literacy into undergraduate ecology courses. In press at Ecology and Evolution. DOI: 10.1002/ece3.4363

- Carey, C.C., N.K. Ward, K.J. Farrell, M.E. Lofton, A.I. Krinos, R.P. McClure, K. Subratie, R.J. Figueiredo, J.P. Doubek, P.C. Hanson, P. Papadopoulos, and P. Arzberger. Enhancing collaboration between ecologists and computer scientists: lessons learned and paths forward. In review at Ecosphere.

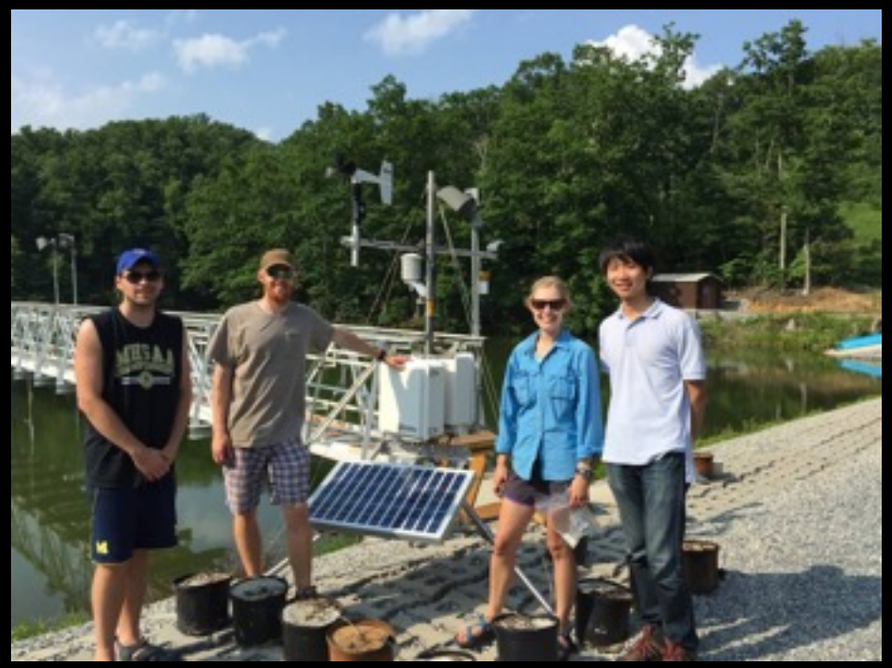
- Farrell, K.J., C.C. Carey, A.I. Krinos, N.K. Ward, P.C. Hanson, R.J. Figueiredo, V. Daneshmand, and K. Subratie. Increasing air temperatures differentially alter intra- and inter-annual nitrogen and phosphorus cycling in a eutrophic and an oligotrophic lake. Oral presentation. Ecological Society of America, New Orleans, Louisiana, August 2018.

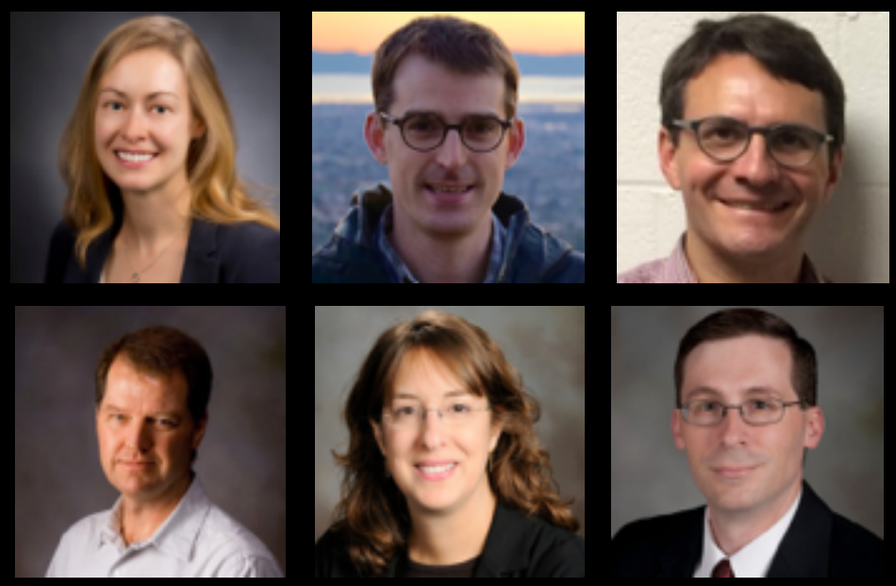
- Carey, C.C., A.I. KrinosU, K.J. FarrellP, J. SukumarG, K. SubratieG, A. Hetherington, P.C. Hanson, and R. Figueiredo. Simulation modeling reveals non-linear lake water quality responses to climate and land use change. Oral presentation. Ecological Society of America Conference, Portland, Oregon, August 2017.

Expanding to the Edge: Sensor Gateways

Motivation

- Smart and Connected Communities (SCC) inter-disciplinary project
  - http://smartreservoir.org/
- Integrating data from sensors with model run workflows for novel model-data fusion predictive modeling approach
- Leverages overlay network technology used in PRAGMA, brings it to the “edge”

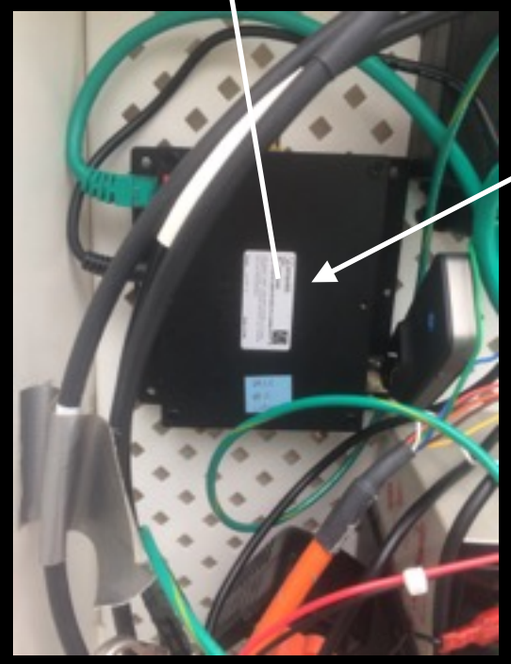




Smart reservoir faculty team: Cayelan Carey (PI), Quinn Thomas, Francois Birgand, John Little, Maddie Schreiber, Michael Sorice, Renato Figueiredo

Sensor Gateway Design

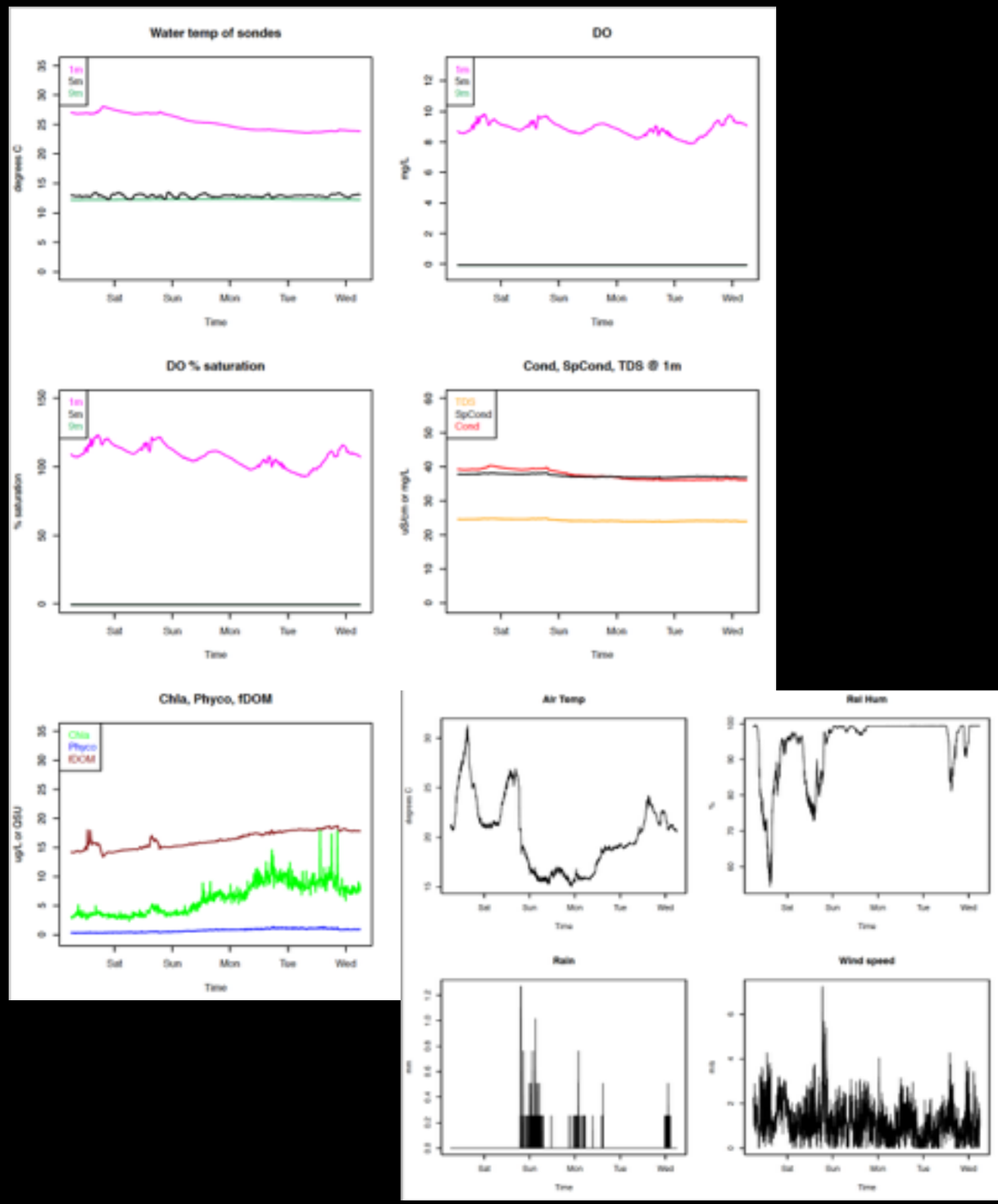
- Fanless mini-PC, open-source software
- Reads/stores data from logger (Campbell)
- Security, NAT traversal:
  - Connects to IPOP overlay VPN via 4G cellular link
  - Data publishing, remote maintenance
- Pushes data updates (diffs) to private and/or public repository using git



fitlet2

Prediction workflow

- Data collected from multiple sensors (air temperature, rain, wind, water temperature, DO, chlorophyll) committed to git data repos over IPOP daily
- Triggers pull execution of ensemble GLM forecast model runs
    - Dedicated servers, or GRAPLEr

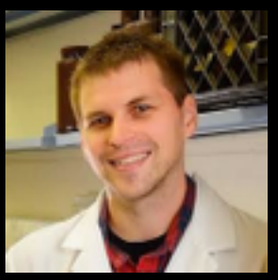


Lessons Learned

Making Interdisciplinary Research Work

- Trust, engagement, and genuine interest in each other's disciplines are crucial - and need to be fostered
    - Embedding in each other's communities:
        - Face-to-face meetings, demonstrations, engaging students
    - Leveraging existing networks and building connections
- Need to understand opportunities, incentives and pain points
    - Listening, working together, and designing iteratively - prescribing solutions is not a solution

Acknowledgments



PRAGMA/GLEON collaborators: Cayelan Carey, Paul Hanson

Students and post-docs: Arianna Krinos, Mary Lofton, Nicole Ward, Ryan McClure, Craig Snortheim, Jon Doubek, Alex Gerling, Kate Hamre, Ken Subratie, Jaikrishna Tanjore, Saumitra Aditya, Vahid Daneshmand, Srinivas Mahesula, Kait Farrell, Youna Jung, Amy Hetherington, Hilary Dugan

Acknowledgments

- PRAGMA is supported by its member institutions and the U.S. National Science Foundation (NSF OCI 1234983), involving researchers at UCSD, U Florida, Indiana U, and U Wisconsin

- Smart Water Systems is supported by the US National Science Foundation as part of the Smart & Connected Communities program, Award CNS-1737424

- IPOP is based upon work supported in part by the National Science Foundation under grants 1339737 and 1527415.

Thank you!

Questions?