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

Paul Hanson (University of Wisconsin)
Cayelan Carey (Virginia Tech)
Renato Figueiredo, Ken Subratie (University of Florida)

California Imposes First Mandatory Water Restrictions to Deal With Drought

By **ADAM NAGOURNEY** APRIL 1, 2015

Sao Paulo’s main drinking water reservoir. Photo: Mauricio Lima

We are facing an unprecedented global water crisis

- Across the world, climate change and other human activities are threatening freshwater quality and quantity
- These stressors will likely only increase in the future, especially for lakes and reservoirs, our primary drinking water sources
- Ultimately, we need to better manage lake water quality to best leverage the water resources we do have access to

The implications are profound…

...And will affect all of us.

- Undrinkable water due to contaminants, pharmaceuticals, heavy metals, and more
- Toxic algal blooms
- Fish kills
- Loss of recreation
- Long-term health effects
- Altered food prices and economies
- Water refugees and conflict

Ultimately, we need to understand how lake water quality will respond to human demands and changing climate.

Addressing these problems in a globally significant way requires use of distributed resources:

data, models, people, compute

+

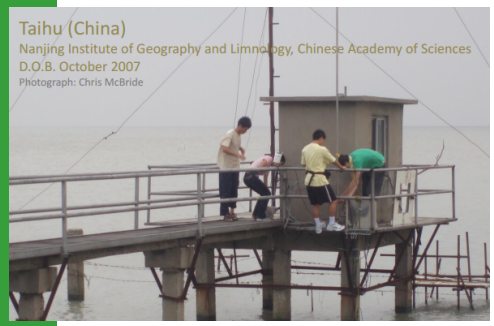
GLEON

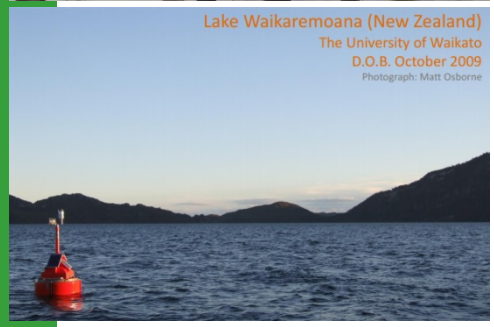
GLEON GLOBAL LAKE ECOLOGICAL OBSERVATORY NETWORK

Search

Understand, Predict and Communicate the Role and Response of Lakes in a Changing Global Environment

Data





GLEON dataset archive

DataONE GLEON is a DataONE Member Node. GLEON data may be search on the DataONE portal or on the GLEON specific portal.

CUAHSI HIS
Sharing hydrologic data

Login

Home | All Data Services | Forum

Dorset Environmental Science Centre

Dorset Environmental Science Centre
GLEON_Dorset
http://hydroportal.cuahsi.org/GLEON_Dorset/cuahsi_1_1.asmx?WSDL

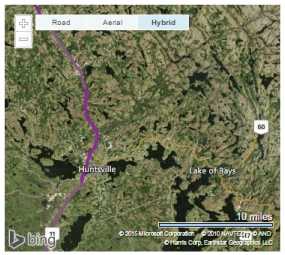
Contact: Jim Rusak
Jim.Rusak@ontario.ca
705-766-0659

Service Statistics:

Sites: 1
Variables: 16
Values: 443,718

Geographic Extent:
45.37999
-79.13396 -79.13396
45.37999

Last Harvested on 3/14/2015 6:01:15 PM



Contributing Organizations

Abstract

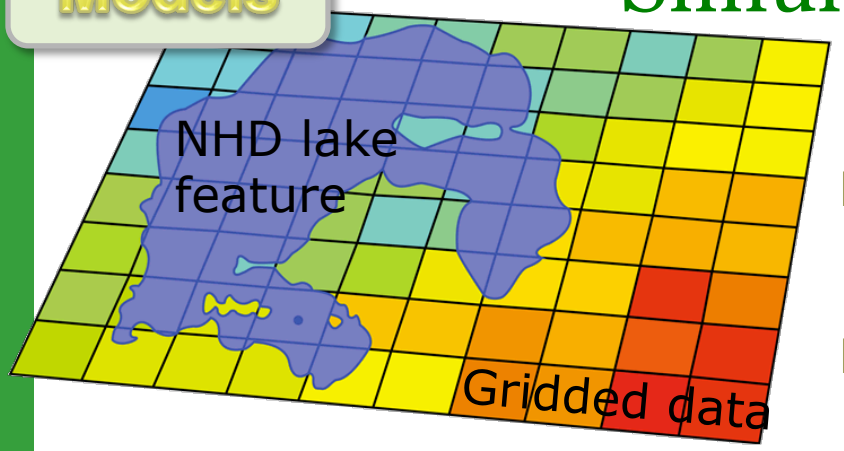
Focusing on headwater lakes and streams near Dorset, Ontario, this program gathers chemical, physical and biological data to investigate long-term patterns and processes in aquatic ecosystems. A watershed scale approach is used to investigate the effects of multiple stressors, including climate change, on water quality and quantity in Ontario's inland aquatic ecosystems and sustain the functions and services that these systems provide. Currently CUAHSI is hosting 10-minute data, for a wide variety of parameters, from the Dorset Environmental Science Centre's automated sensing platform on Harp Lake in the Muskoka Region of Ontario.

Citation

Data provided by the Dorset Environmental Science Centre of the Ontario Ministry of the Environment

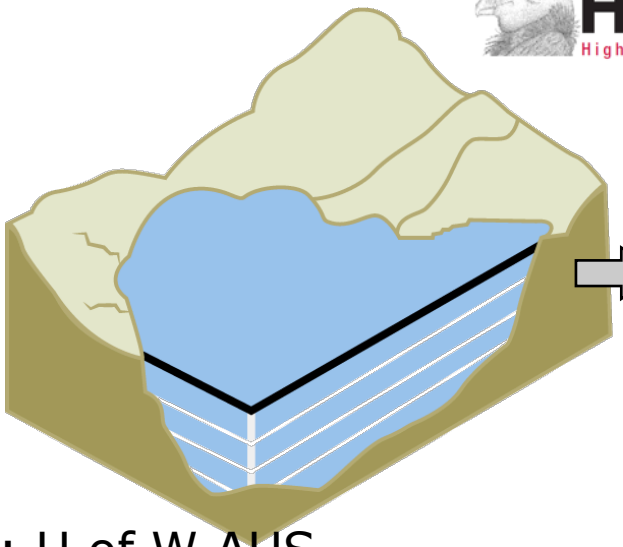
Models

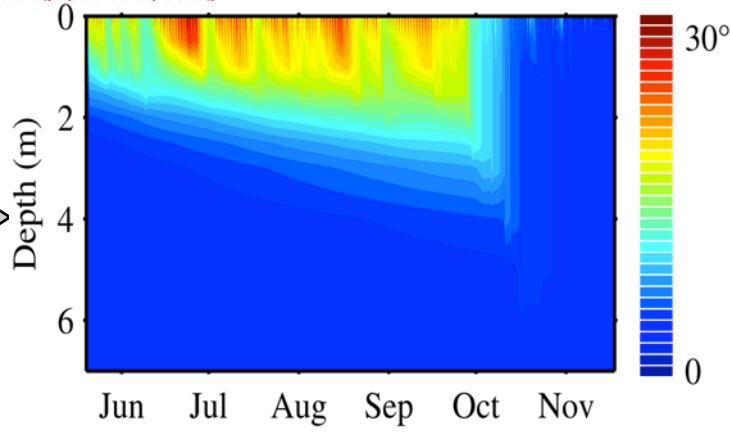
Simulating lakes in the landscape



- 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*)

Inter-disciplinary activities

- PRAGMA'26, '27, '28
  - does PRAGMA run on a Mac?
- GLEON'16 workshop
  - cyanobacteria = blue-green algae?
  - IP over… P?
- GLM Week
  - We want more computers! And save the world!

Meta-Goals

- Exploit lake sensor network observations (GLEON) with overlay network technology (PRAGMA) to expand our modeling capacity =(GRAPLE)
- **GRAPLE** = **G**LEON **R**esearch **A**nd **P**RAGMA **L**ake **E**xpedition: integrating disparate technology, knowledge, and researchers to advance lake science
- Our ultimate long-term goal: expand the lake expedition to include cohorts of graduate and undergraduate students through targeted training in both computer science and limnology

Technology

- Middleware: we need an architecture that is poised for growth: flexible, general-purpose, scalable
- Application: we need to support applications poised for growth in the community
- Technical approach
  - IPOP: overlay network aggregates distributed resources – servers, desktops, cloud
  - HTCondor: manages jobs across resources
  - GRAPLE scripts: application-specific code, exposing familiar computer environments, emphasis on ease of use for GLM-FABM jobs

IP-over-P2P (IPOP) Overlay

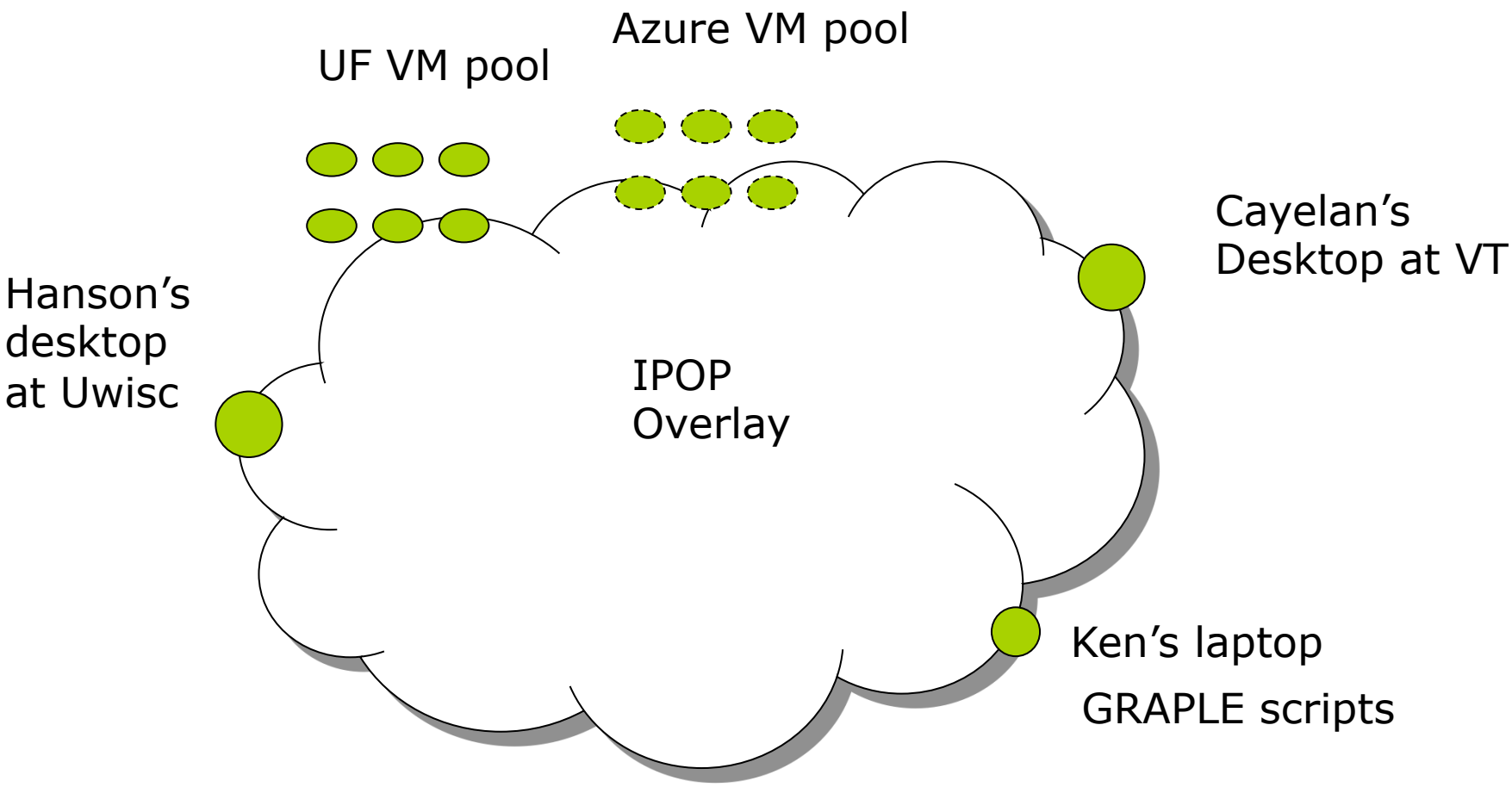
- "User-Defined Networking"
  - SDN that users can deploy across their own resources
    - User-level; easy to use
  - Software only needed at endpoints
    - Cloud VMs, personal computers, mobile devices
  - Map social network to virtual network links
    - E.g. GRAPLE researchers, students
  - Peer-to-peer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

Technology Updates

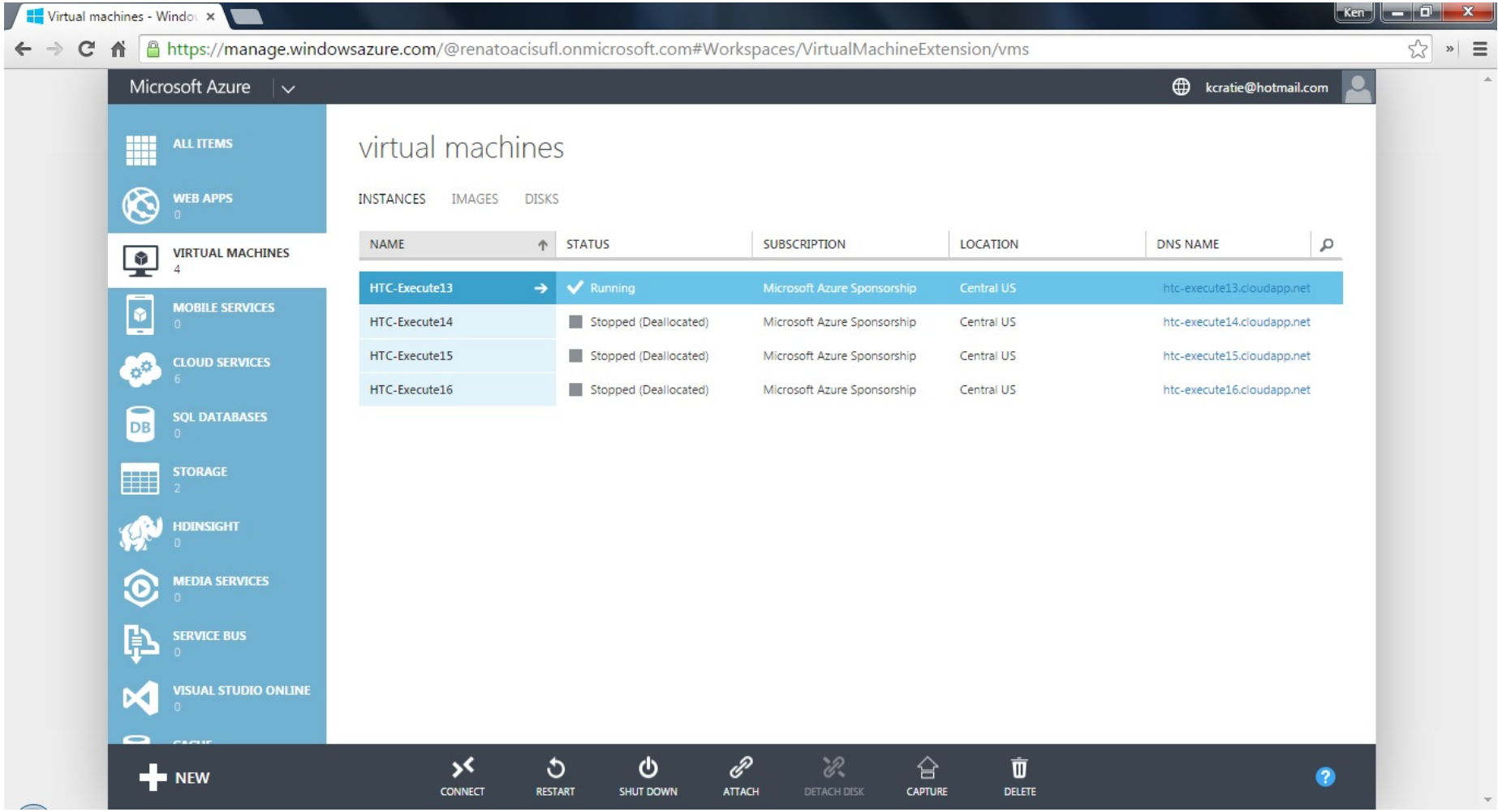
- Application: GRAPLE scripts
  - Expose simple interface to users
  - Handle packaging of multiple simulations into Condor job, compression, and job submission
- Middleware: IPOP enhancements
  - Ganglia monitoring
  - Switch mode, opportunities for integration with SDNs
  - Engagement in Google summer of code 2015
- Resources: private and public clouds
  - UF, UW, VT: prototype cluster & desktops
  - Azure: successful $20k proposal; on-demand cycles

Demonstration - overview

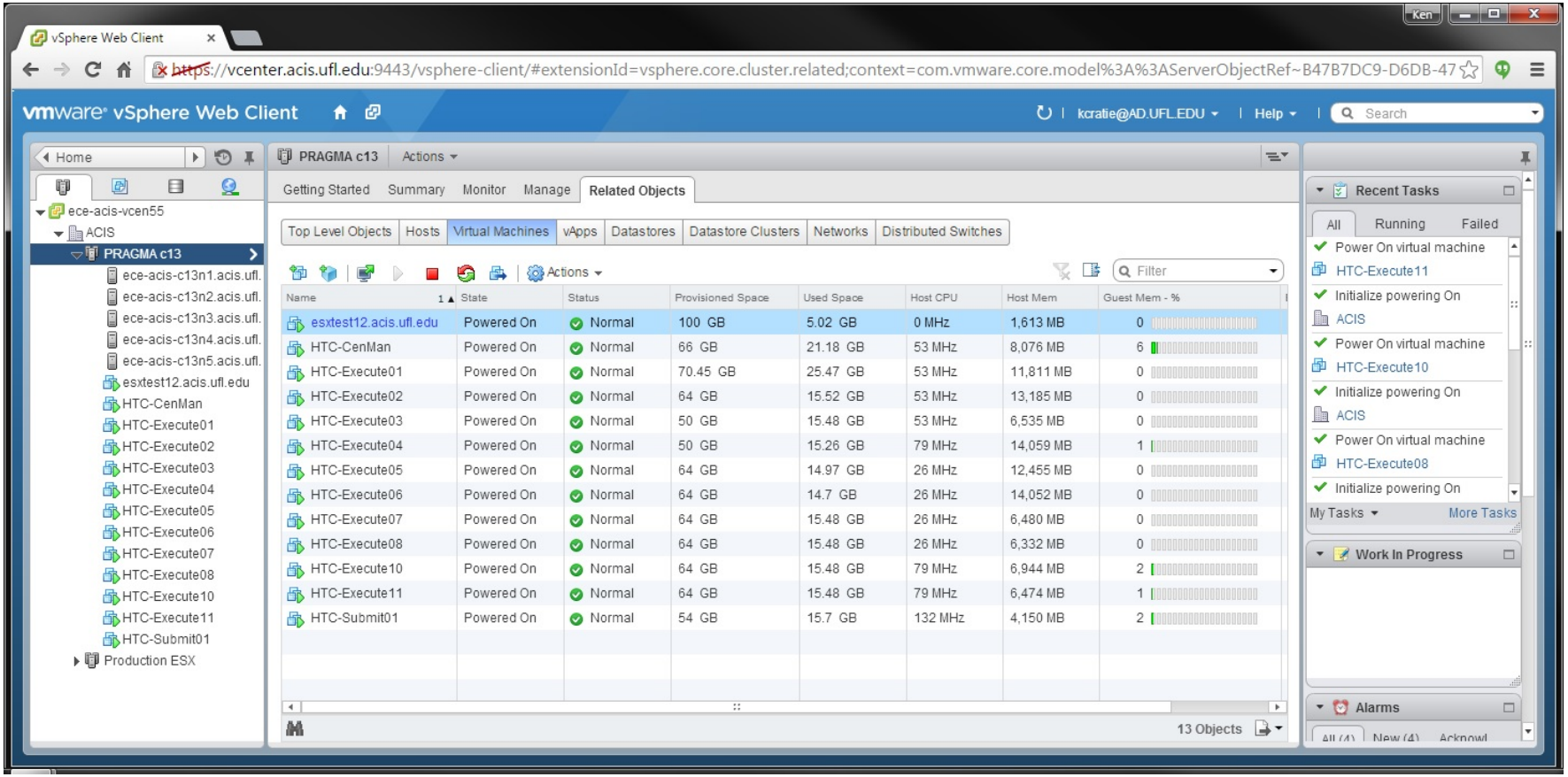
A distributed HTCondor pool spanning multiple domains



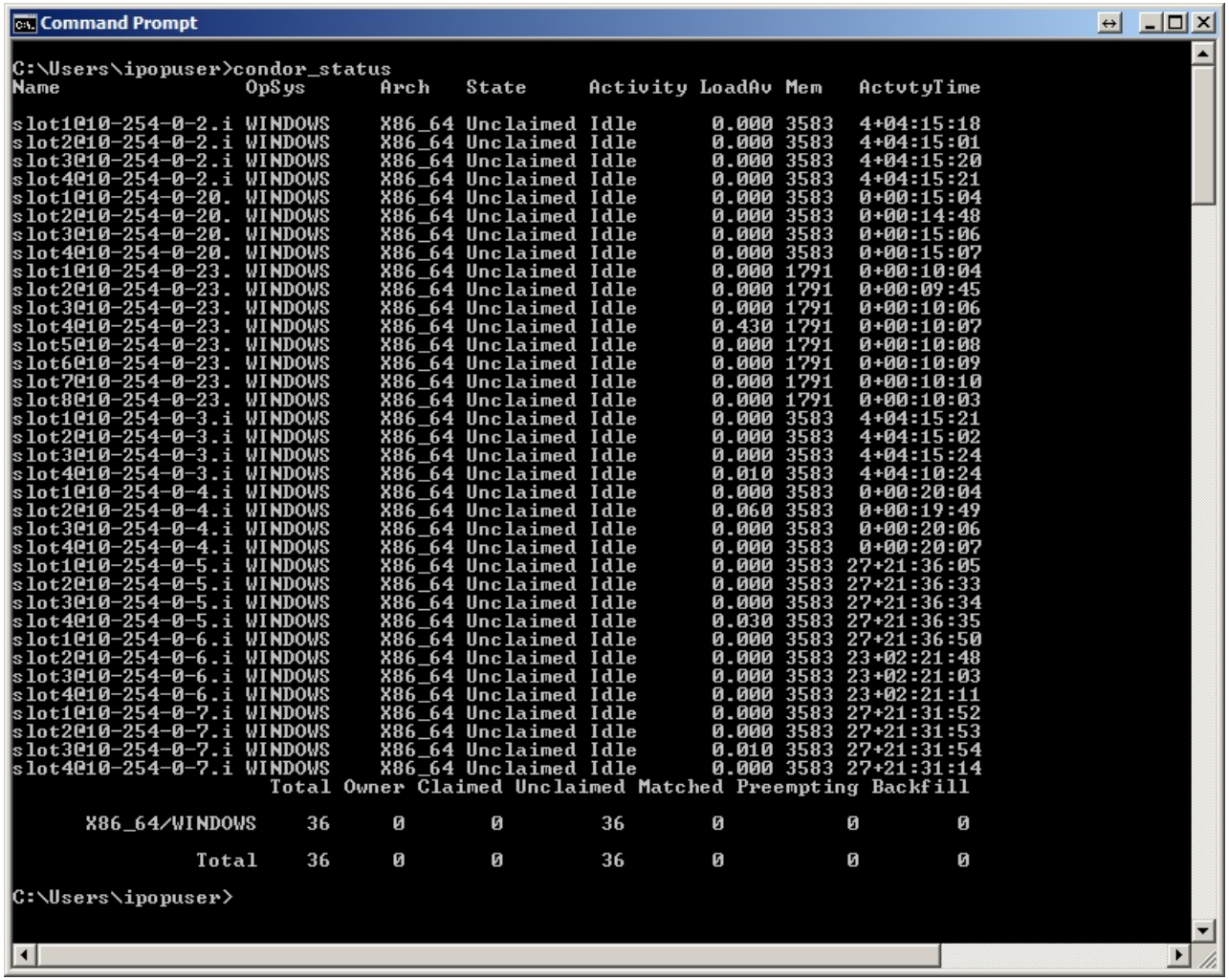
Azure VM


Virtual machines - Windo
https://manage.windowsazure.com/@renatoacisufl.onmicrosoft.com#Workspaces/VirtualMachineExtension/vms
Microsoft Azure
kcratie@hotmail.com
ALL ITEMS
WEB APPS
0
VIRTUAL MACHINES
4
MOBILE SERVICES
0
CLOUD SERVICES
6
SQL DATABASES
0
STORAGE
2
HDINSIGHT
0
MEDIA SERVICES
0
SERVICE BUS
0
VISUAL STUDIO ONLINE
0
virtual machines
INSTANCES
IMAGES
DISKS
NAME
STATUS
SUBSCRIPTION
LOCATION
DNS NAME
HTC-Execute13
Running
Microsoft Azure Sponsorship
Central US
htc-execute13.cloudapp.net
HTC-Execute14
Stopped (Deallocated)
Microsoft Azure Sponsorship
Central US
htc-execute14.cloudapp.net
HTC-Execute15
Stopped (Deallocated)
Microsoft Azure Sponsorship
Central US
htc-execute15.cloudapp.net
HTC-Execute16
Stopped (Deallocated)
Microsoft Azure Sponsorship
Central US
htc-execute16.cloudapp.net
NEW
CONNECT
RESTART
SHUT DOWN
ATTACH
DETACH DISK
CAPTURE
DELETE

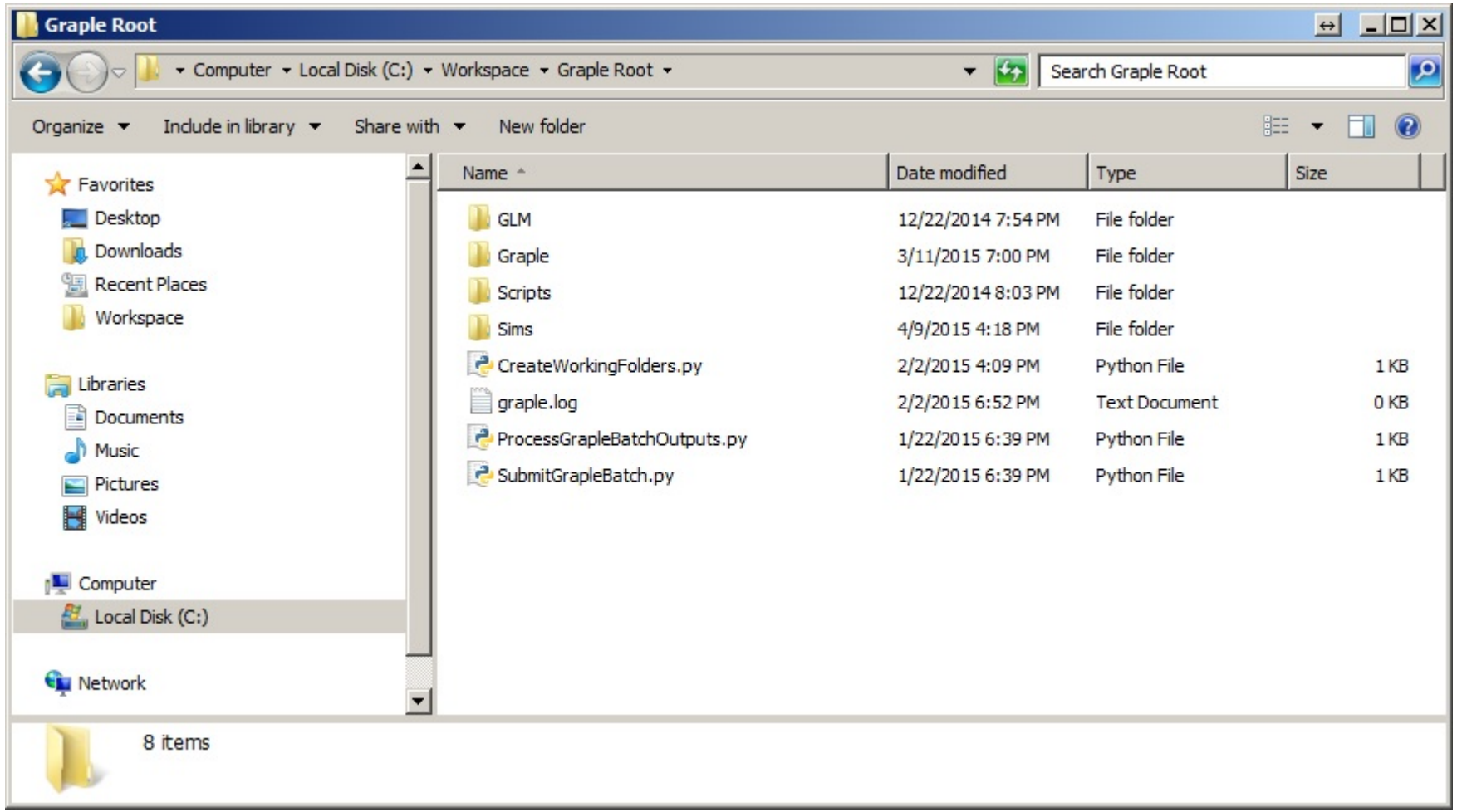
VMware ESX VMs



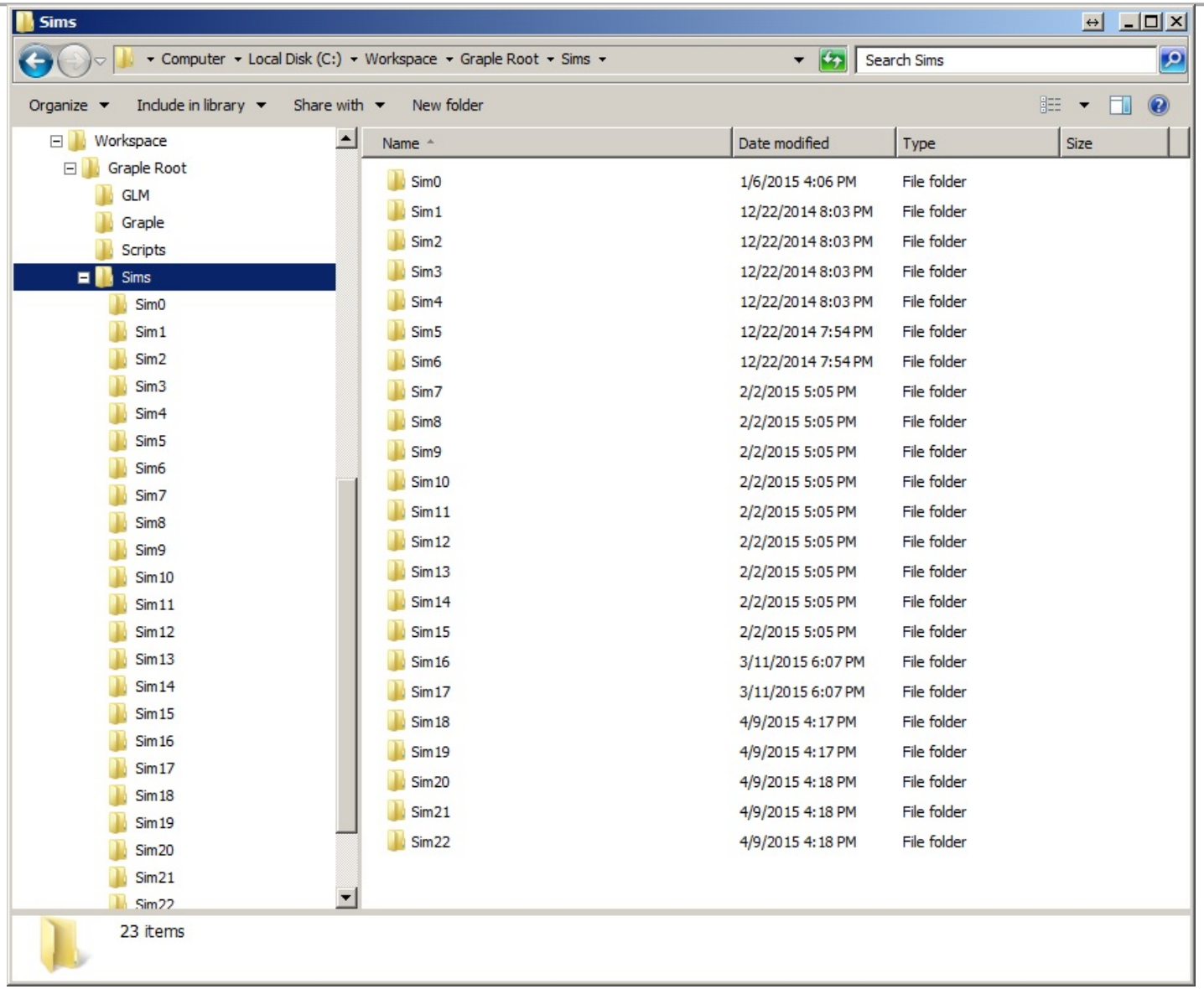
HTCondor+overlay



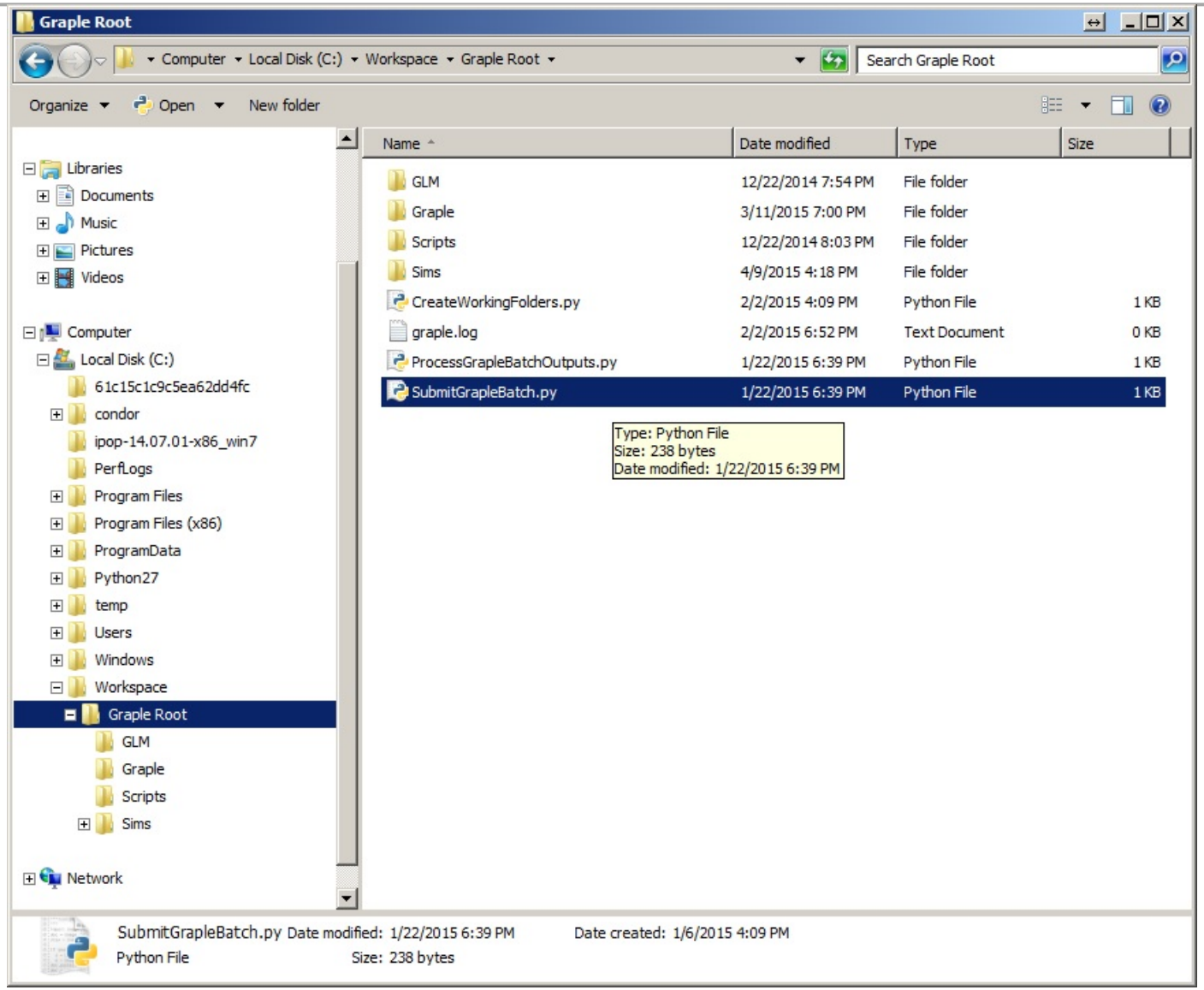
GRAPLE directories



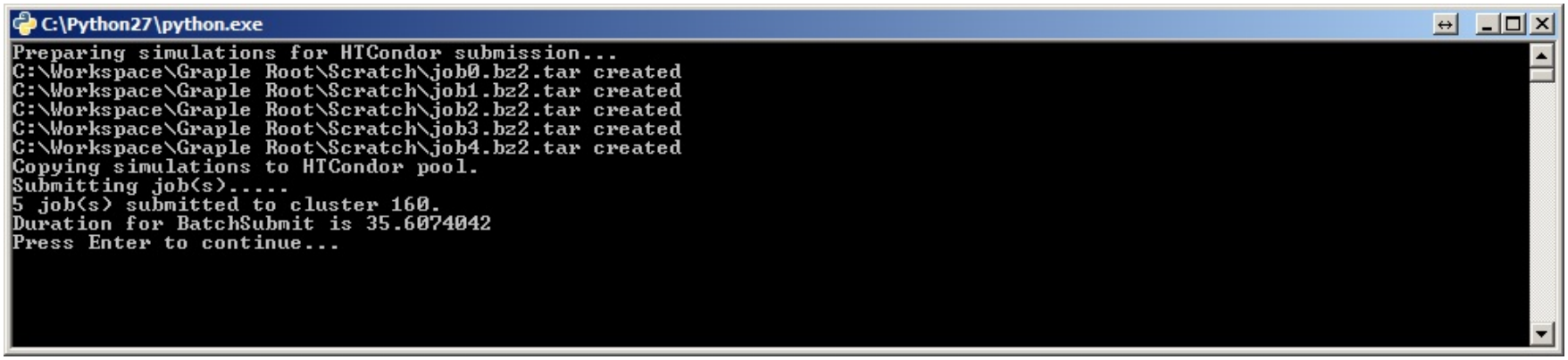
GLM simulation directories



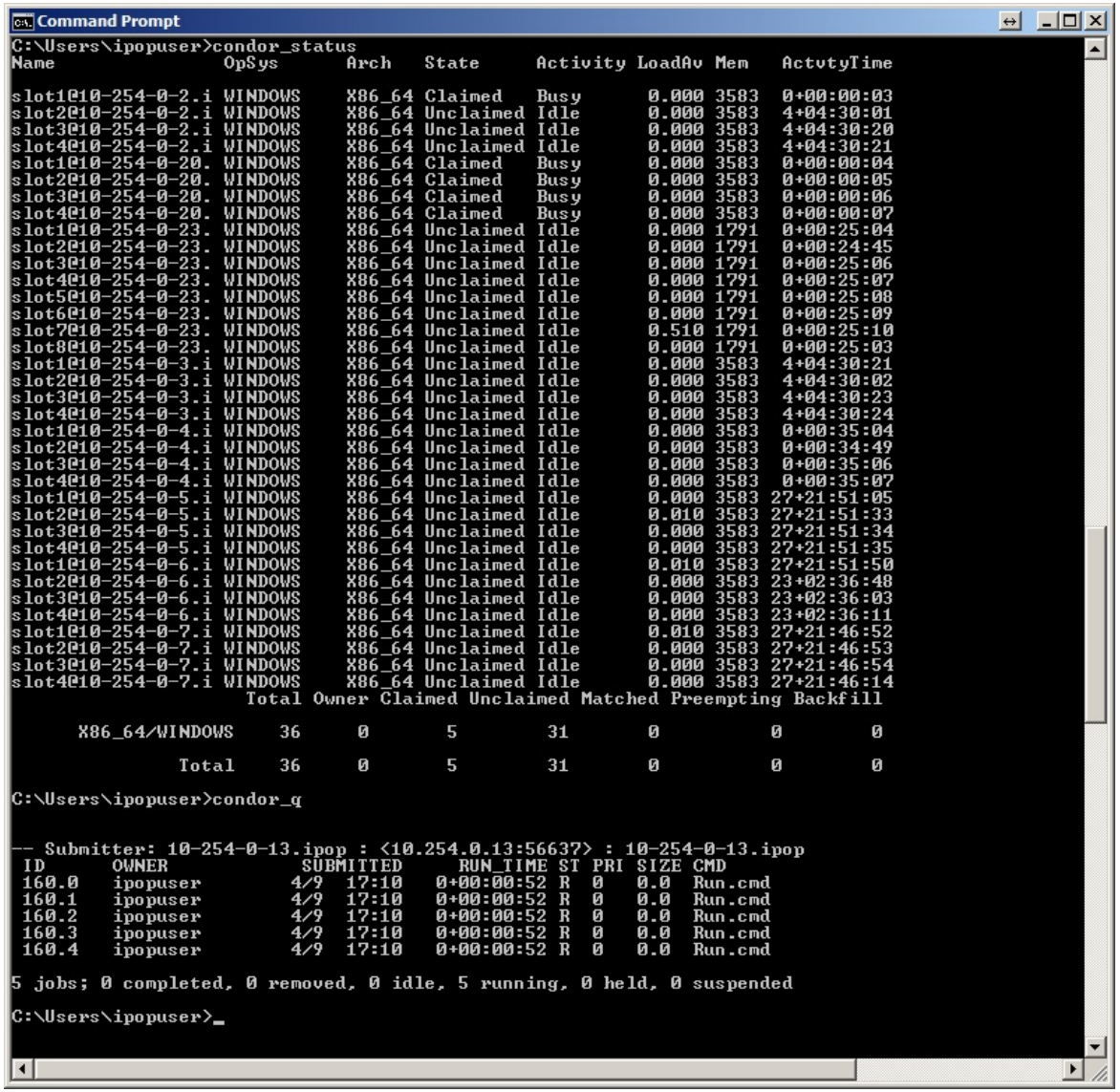
Submit jobs


Graple Root
Computer ▸ Local Disk (C:) ▸ Workspace ▸ Graple Root ▸
Search Graple Root
Organize
Open
New folder
Libraries
Documents
Music
Pictures
Videos
Computer
Local Disk (C:)
61c15c1c9c5ea62dd4fc
condor
ipop-14.07.01-x86_win7
PerfLogs
Program Files
Program Files (x86)
ProgramData
Python27
temp
Users
Windows
Workspace
Graple Root
GLM
Graple
Scripts
Sims
Network
Name
Date modified
Type
Size
GLM 12/22/2014 7:54 PM File folder
Graple 3/11/2015 7:00 PM File folder
Scripts 12/22/2014 8:03 PM File folder
Sims 4/9/2015 4:18 PM File folder
CreateWorkingFolders.py 2/2/2015 4:09 PM Python File 1 KB
graple.log 2/2/2015 6:52 PM Text Document 0 KB
ProcessGrapleBatchOutputs.py 1/22/2015 6:39 PM Python File 1 KB
SubmitGrapleBatch.py 1/22/2015 6:39 PM Python File 1 KB
Type: Python File
Size: 238 bytes
Date modified: 1/22/2015 6:39 PM
SubmitGrapleBatch.py Date modified: 1/22/2015 6:39 PM Date created: 1/6/2015 4:09 PM
Python File Size: 238 bytes

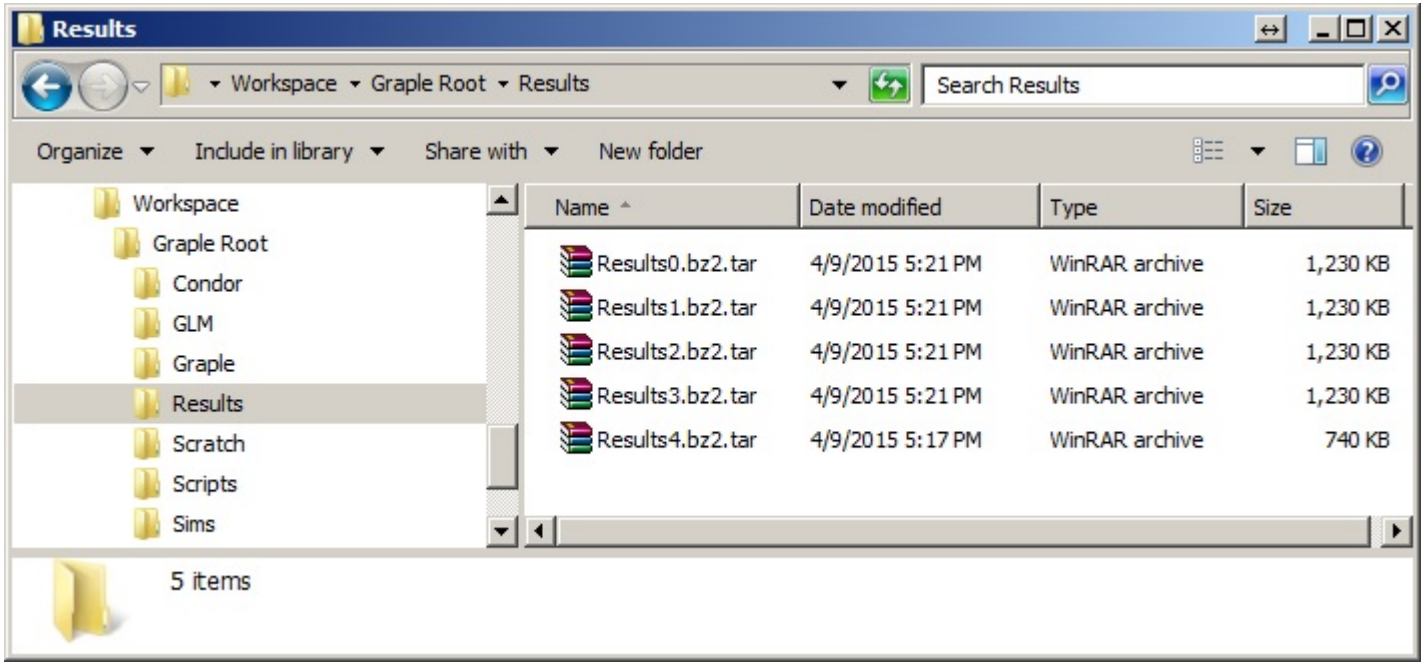
Jobs submitted



Jobs running



Results

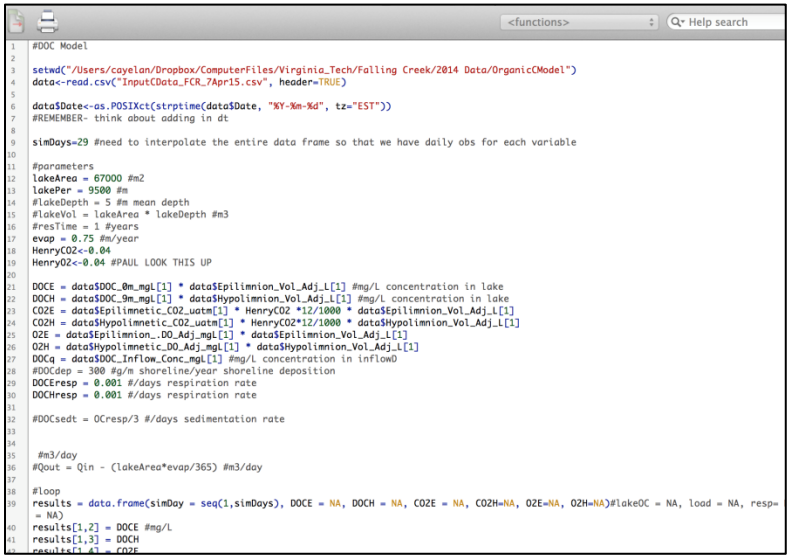

Results
Workspace ▸ Grapple Root ▸ Results
Search Results
Organize
Include in library
Share with
New folder
Workspace
Grapple Root
Condor
GLM
Grapple
Results
Scratch
Scripts
Sims
Name
Date modified
Type
Size
Results0.bz2.tar 4/9/2015 5:21 PM WinRAR archive 1,230 KB
Results1.bz2.tar 4/9/2015 5:21 PM WinRAR archive 1,230 KB
Results2.bz2.tar 4/9/2015 5:21 PM WinRAR archive 1,230 KB
Results3.bz2.tar 4/9/2015 5:21 PM WinRAR archive 1,230 KB
Results4.bz2.tar 4/9/2015 5:17 PM WinRAR archive 740 KB
5 items

Meta-Goals

- Exploit lake sensor network observations (GLEON) with overlay network technology (PRAGMA) to expand our modeling capacity =(GRAPLE)
- **GRAPLE** = **G**LEON **R**esearch **A**nd **P**RAGMA **L**ake **E**xpedition: integrating disparate technology, knowledge, and researchers to advance lake science
- Our ultimate long-term goal: expand the lake expedition to include cohorts of graduate and undergraduate students through targeted training in both computer science and limnology

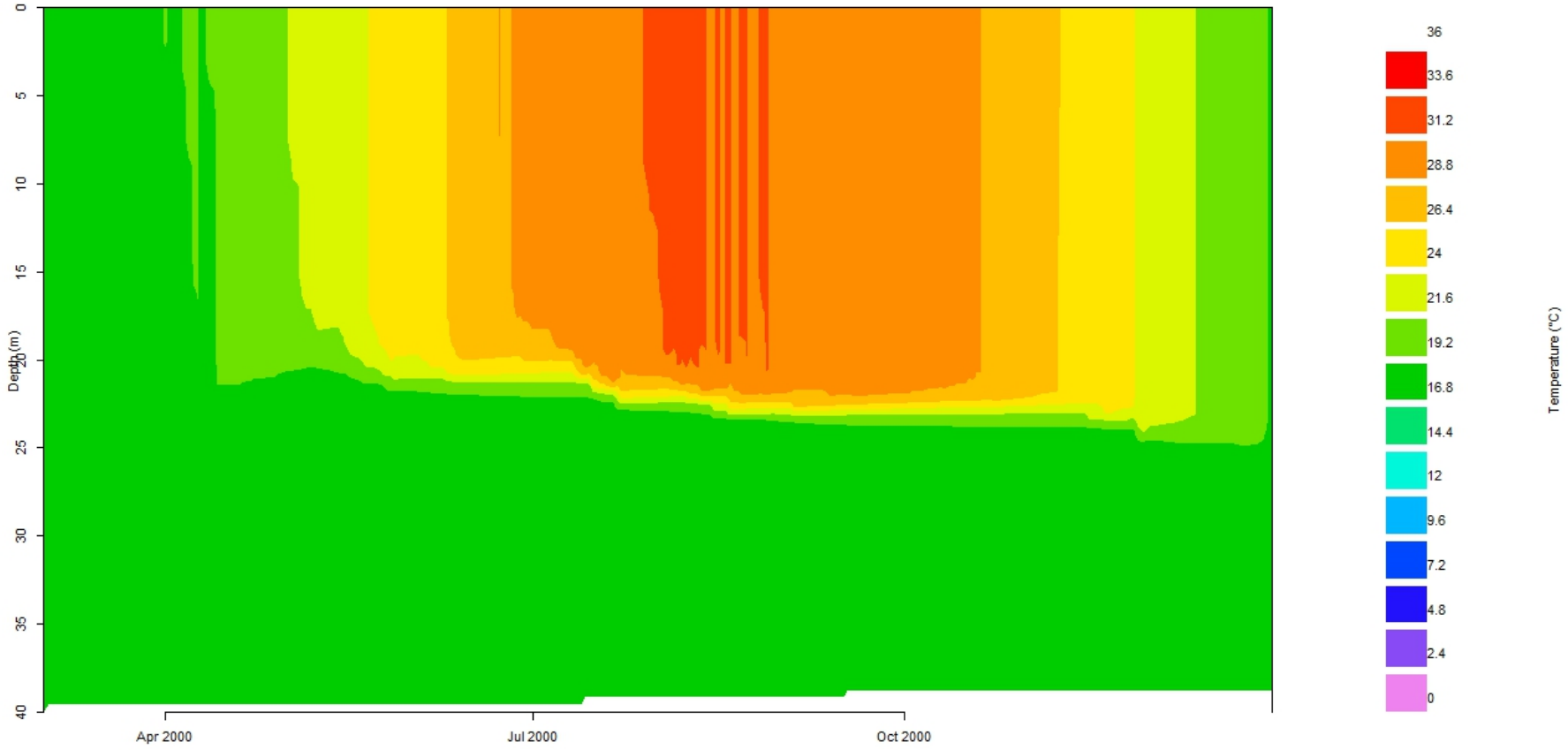
Expanding GRAPLE globally

- We have developed simple GRAPLE R scripts as 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*

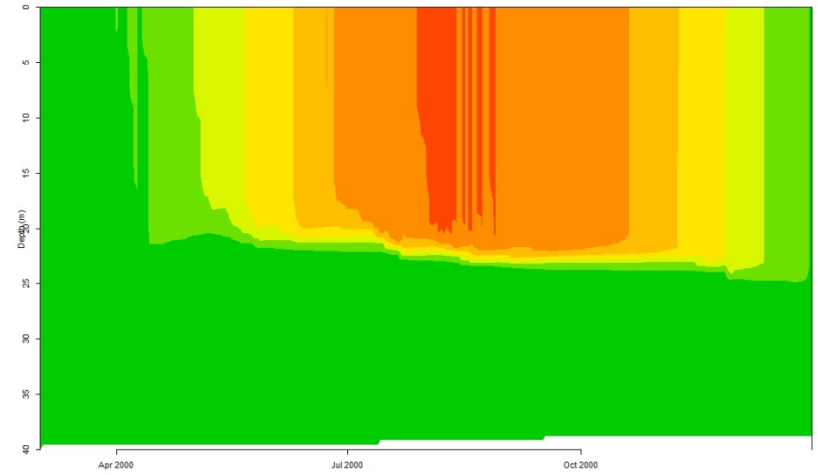


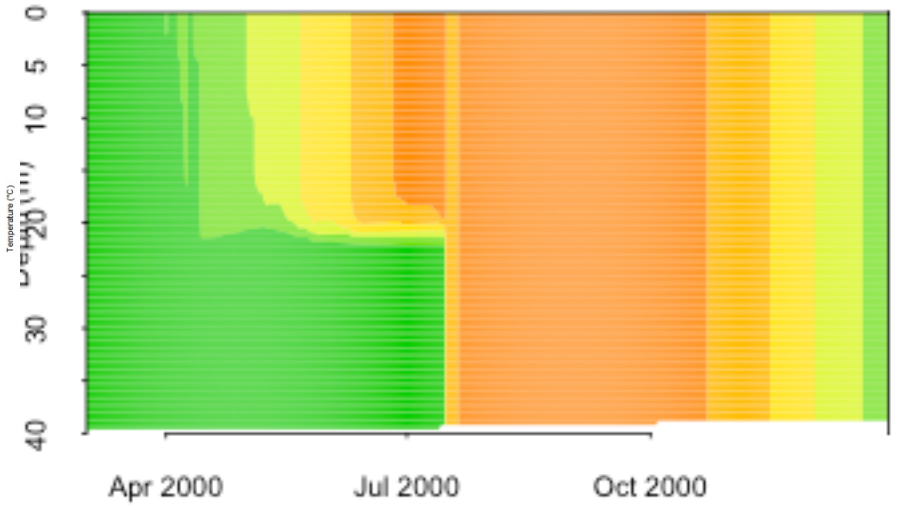
```
#DOC Model

setwd("/Users/cayelan/Dropbox/ComputerFiles/Virginia_Tech/Falling Creek/2014 Data/OrganicCModel")
data<-read.csv("InputCData_FCR_7Apr15.csv", header=TRUE)

data$Date<-as.POSIXct(strptime(data$Date, "%Y-%m-%d", tz="EST"))
#REMEMBER- think about adding in dt

simDays=29 #need to interpolate the entire data frame so that we have daily obs for each variable

#parameters
lakeArea = 67000 #m2
lakePer = 9500 #m
#lakeDepth = 5 #m mean depth
#lakeVol = lakeArea * lakeDepth #m3
#resTime = 1 #years
evap = 0.75 #m/year
HenryCO2<-0.04
HenryO2<-0.04 #PAUL LOOK THIS UP

DOCE = data$DOC_0m_mgL[1] * data$Epilimnion_Vol_Adj_L[1] #mg/L concentration in lake
DOCH = data$DOC_9m_mgL[1] * data$Hypolimnion_Vol_Adj_L[1] #mg/L concentration in lake
CO2E = data$Epilimnetic_CO2_uatm[1] * HenryCO2 *12/1000 * data$Epilimnion_Vol_Adj_L[1]
CO2H = data$Hypolimnetic_CO2_uatm[1] * HenryCO2*12/1000 * data$Hypolimnion_Vol_Adj_L[1]
O2E = data$Epilimnion_.DO_Adj_mgL[1] * data$Epilimnion_Vol_Adj_L[1]
O2H = data$Hypolimnetic_DO_Adj_mgL[1] * data$Hypolimnion_Vol_Adj_L[1]
DOCq = data$DOC_Inflow_Conc_mgL[1] #mg/L concentration in inflowD
#DOCdep = 300 #g/m shoreline/year shoreline deposition
DOCEresp = 0.001 #/days respiration rate
DOCHresp = 0.001 #/days respiration rate

#DOCsedt = OCresp/3 #/days sedimentation rate


 #m3/day
#Qout = Qin - (lakeArea*evap/365) #m3/day

#loop
results = data.frame(simDay = seq(1,simDays), DOCE = NA, DOCH = NA, CO2E = NA, CO2H=NA, O2E=NA, O2H=NA)#lakeOC = NA, load = NA, resp=
 = NA)
results[1,2] = DOCE #mg/L
results[1,3] = DOCH
```

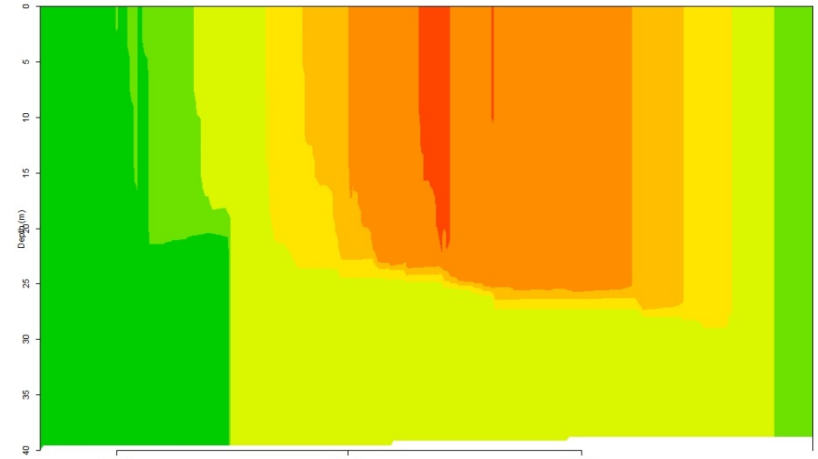
Baseline scenario: results from April 2015

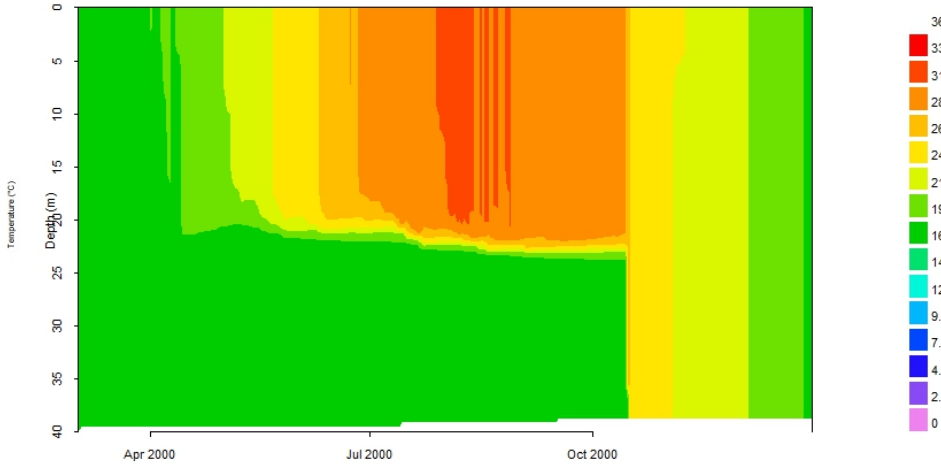


Manipulating climate *in silico* as a teaching tool across disciplines









Next steps

- The first cohort of students at Virginia Tech are being formally assessed to track their responses to using GLM and overlay technology
- This module will be disseminated to 8 universities for testing in the fall as part of the U.S. NSF-sponsored Project EDDIE 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
- All GRAPLE materials will be available open-source for GLEONites and students to access

- Modules
  - Ice Phenology
- Meetings
- Research
- About
  - Goals
  - Data Sets
  - Implementation
  - Evaluation
- Contact
- Forms
- Project EDDIE Home
- CeMaST Home

Project EDDIE: Ice Phenology Module

Project Summary

Lakes are changing worldwide due to altered climate. Many lakes that were historically frozen in the winter are now experiencing fewer days of ice cover and earlier ice-off dates. In this module, students will explore long-term ice-off datasets from several lakes and use linear regression to make predictions about ice-off dates in the future. Project EDDIE modules are designed with an A-B-C structure to make them flexible and adaptable to a range of student levels and course structures.

Acknowledgment

We recognize that many users might wish to modify this activity. Please include text in each file that you use that acknowledges the original development of this module. We suggest this text, which is currently included in each file:

> *"This module was initially developed by Carey, C.C., J.L. Klug, and D.C. Richardson. 1 April 2015. Project EDDIE: Lake Ice Phenology. Project EDDIE Module 1, Version 1. http://cemast.illinoisstate.edu/data-for-students/modules/ice-phenology.shtml. Module development was supported by NSF DEB 1245707."*

Citation

Please cite this module as: "Carey, C.C., J.L. Klug, and D.C. Richardson. 1 April 2015. Project EDDIE: Lake Ice Phenology. Project EDDIE Module 1, Version 1. http://cemast.illinoisstate.edu/data-for-students/modules/ice-phenology.shtml."

Class Time

This entire module can be completed in one 2-3 hour lab period or two 50 minute lecture periods for introductory or intermediate level students. Activities A and B could be completed with upper level students in one 50-60 minute lecture period. Students will need 1-2 hours outside of class to prepare for the exercise and complete the homework activities.

Files

- Student Handout [docx]
- Student Datasets [xls]
- Instructor's Manual [docx]
- Instructor's PowerPoint [pptx]
- Instructor's Homework Part C Answer Key [xlsx]

Project EDDIE is supported by NSF DEB 1245707.

projectEDDIE.org

Expanding GRAPLE beyond this initial expedition

- Our work to date has demonstrated that CS-limnology synergies can provide new tools for tackling local to global water issues
- We are poised to lead a new wave of technology-enabled, multi-disciplinary research that will transform the science we can accomplish
  - Make distributed computing accessible to domain scientists by simplifying the steps needed to participate in the overlay network
  - We have identified two GLEON working groups that will be trained through these scripts to enable their research
  - Grow the next generation of users by training students how to use GLM & overlay networks by teaching these scripts in classrooms worldwide

Acknowledgments

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