



Pegasus Users Group

MEETING



Building an Integrated Assessment Model with Pegasus

Patrick J Clemins

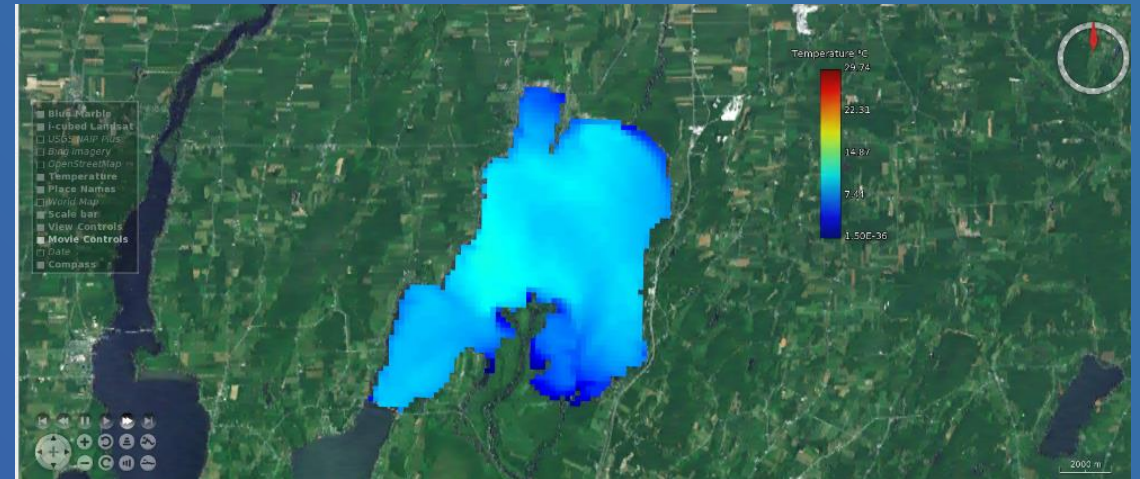
Manager, Cyberinfrastructure and Partnerships, Vermont EPSCoR
Research Affiliate, Dept of Computer Science, University of Vermont

January 23, 2021

Integrated Assessment Modeling

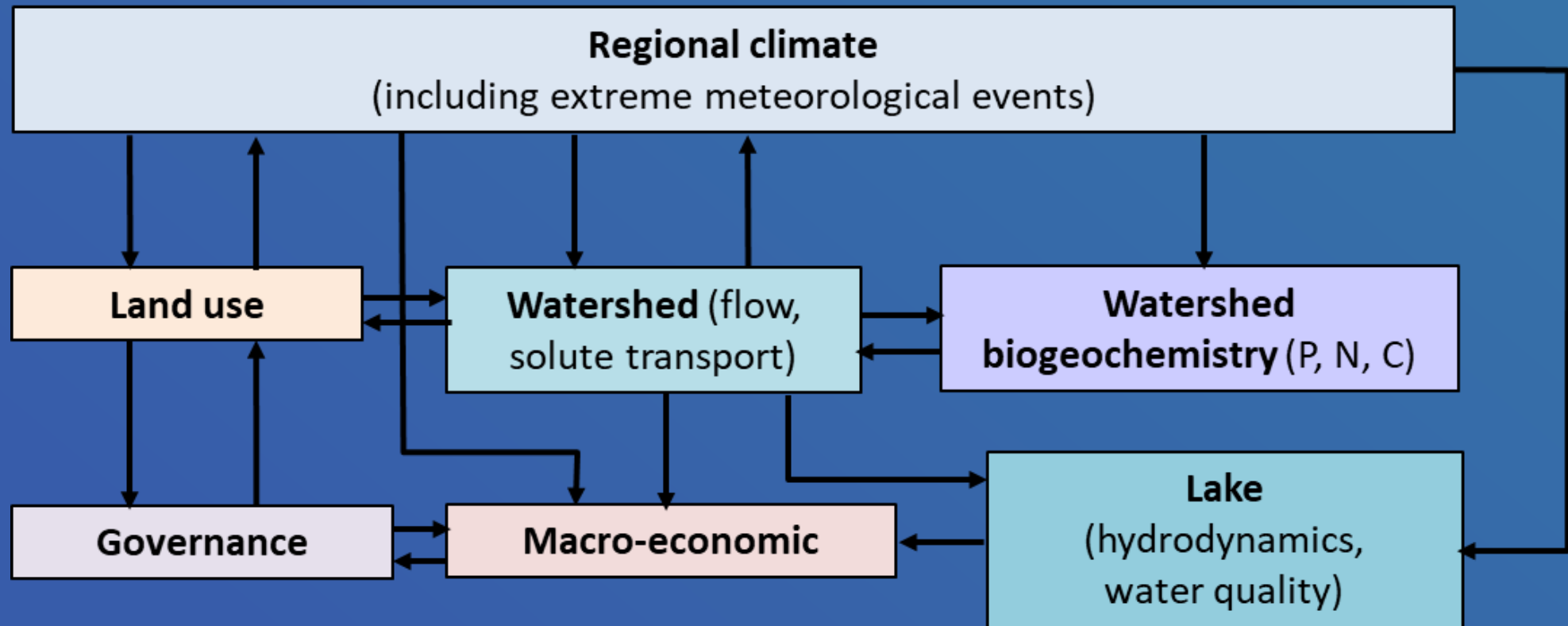


- Vermont EPSCoR's current project, Basin Resilience to Extreme Events (BREE), focuses on understanding the conditions that lead to Harmful Cyanobacteria Blooms (HCBs) in two swallow bays of Lake Champlain.
- IAMs project the impact of policy scenarios on socio-environmental systems (SEs), providing policy makers with data to make informed policy decisions.
- Some of the scenarios we explore encompass variability in climate change projections, land use change, agricultural best practices adoption, and the efficacy of those best practices.



Temperature in Missisquoi Bay, Lake Champlain

The Vermont EPSCoR IAM

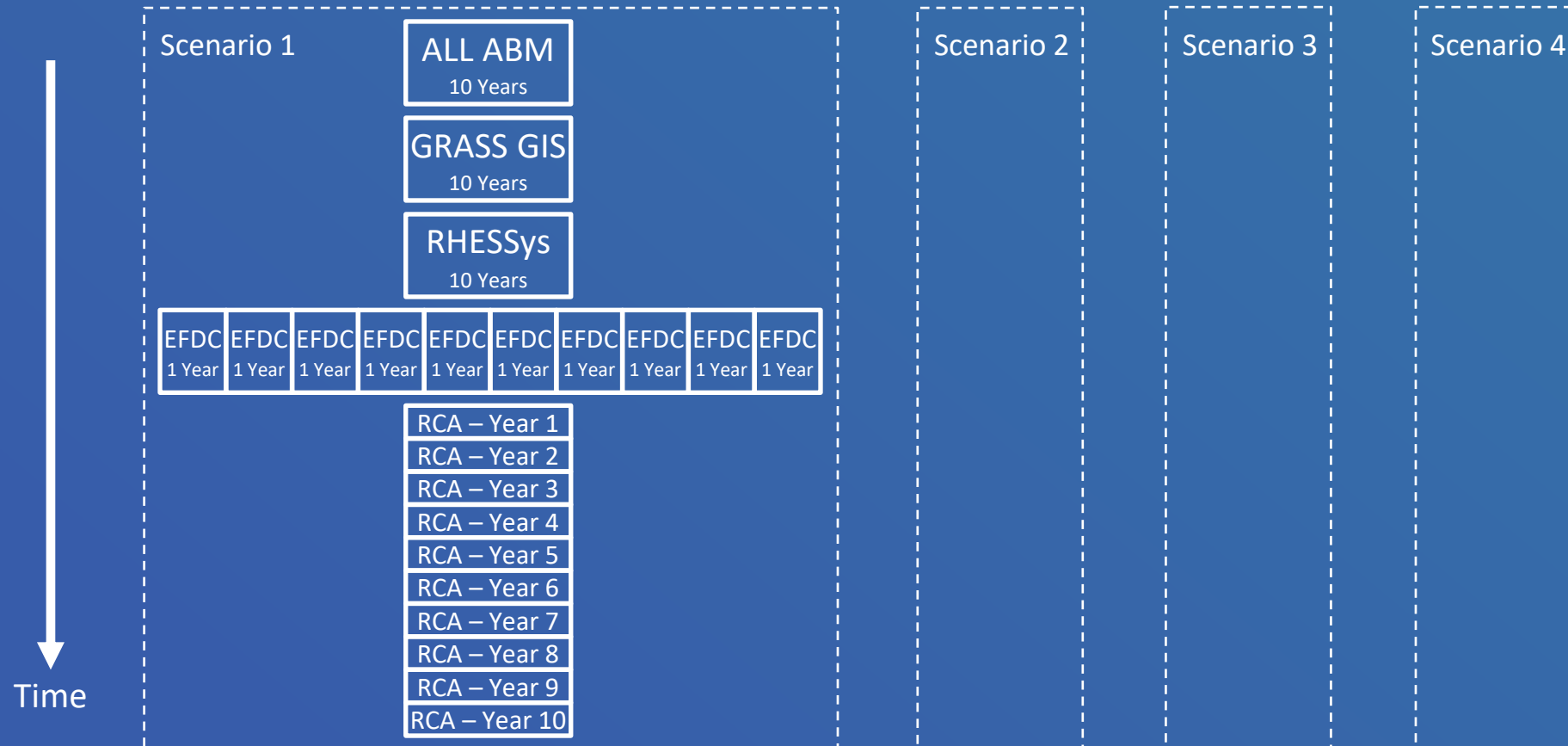


IAM Computational Challenges

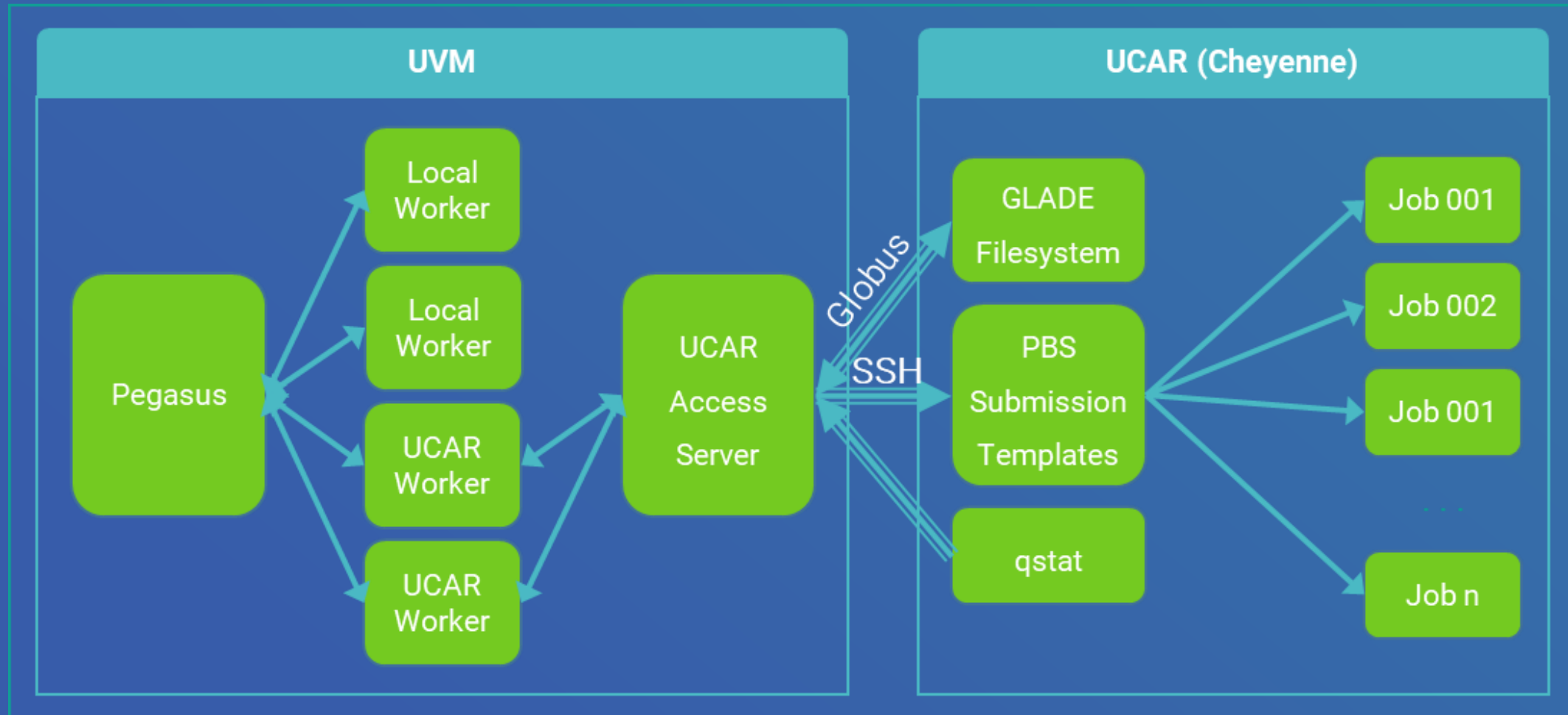


- The desire to use best in class community models for each IAM component means each model has different platform requirements
- Feedbacks
- Orchestrating data flows between models
- Managing the execution of multiple scenarios in parallel
- Multiple scenarios quickly outgrow the computational resources of a single workstation
 - 176 scenarios over 60 years:
 - ~ 10 days wall time, ~ 6.5 years CPU time
 - ~ 25,000 jobs, ~ 3.5 TB data generated

Feedbacks via Decadal Cascades



Heterogeneous Compute



Pegasus Feedback



- Workflows that require Pegasus also require significant computing expertise. Continue to focus on an audience with significant computing expertise.
 - The pegasus-service dashboard is useful for allowing our users to monitor their runs.
- Continue to improve support for HPC clusters running batch schedulers... or NOT?
 - BOSCO, PyGlidein, glite, etc. have all moved the bar since we started using Pegasus in 2014... but are essentially HTCondor extensions
 - What about a built in PBS / Slurm batch manager that collects jobs of similar runtimes, dynamically builds a submit script to maximize compute node usage, submits over SSH, and monitors jobs status.
 - And... include Globus file transfer support
- Post-workflow analysis / troubleshooting is still too difficult
 - Why aren't HTCondor .out and .err saved in Pegasus run directory?



Thank You!



This material is based upon work supported by the National Science Foundation under Grant No. OIA 1556770.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.