2.B. Learning the phytoplankton ‘rules’ to predict water quality in lakes

Predicting freshwater quality in lakes is a global challenge. Many indicators of water quality (including transparency, chlorophyll, and cyanobacterial toxins) are governed by phytoplankton species composition and biomass. At a fundamental level, phytoplankton species composition and biomass are determined by ‘rules,’ or mechanistic relationships that link phytoplankton species performance to abiotic (light, nutrients) and biotic (zooplankton) factors. These rules control phytoplankton dynamics and are necessary for predicting water quality. However, due to the complexity of abiotic-biotic coupling, phytoplankton communities exhibit scale-dependent responses that are challenging to predict. Even with decades of monitoring data and increased modeling capacity, predicting the identity of dominant species remains problematic because the environment in which phytoplankton exist is dynamic, and processes governing phytoplankton occur at minute to seasonal scales. The urgency to make progress in this area is great, as phytoplankton communities and water quality are rapidly changing due to altered climate and land use.

Addressing these questions requires networks – networks of scientists accessing networks of data, and recreating virtual lake ecosystems using a network of computer resources. No two lakes are alike, and so a generalized modeling framework needs calibration to the local lake, which requires local knowledge and data as well as model tuning through parameter estimation. While recreating the physical environment in lakes is challenging, simulating the phytoplankton community with enough realism to understand and predict toxic species has proven to be extremely challenging. This is in fact so challenging that it leads researchers to question whether the ‘rules’ governing phytoplankton are well-enough understood.

In the PRAGMA lake expedition, scientists Hanson and Carey lead a team of limnologists who are using numerical simulation to examine the fundamental rules governing phytoplankton community dynamics. Similar to weather forecasts, lake numerical simulations recreate the physical and chemical environment in multiple dimensions, but with the added complexity of biology. For example, explosive growth of the toxin-producing alga, microcystis, requires just the right combination of physical environment, reduced resource competition from other algal species, and low abundance of zooplankton that would normally keep microcystic populations in check. However, the sequence of physical-biological events that lead to blooms of harmful species remains a mystery – in this case, to be solved *in silico*. Scientists seed a lake with a virtual phytoplankton community comprised of the physiological characteristics drawn from the established library of such traits. The community is allowed to develop and evolve in a virtual physical environment calibrated to observed data. Whether the evolved community does, or does not, match the observed may be a matter of chance, based on the traits initially selected. Thus, the procedure must be repeated thousands of times. Realizing this vision requires lots of computers, because each simulation can take hours to days to complete.

The researchers have traditionally used HTCondor (High Throughput Computing Condor), which manages the workload required for these simulations. However, one key challenge that has been difficult to overcome is lack of computational capacity: the software required for the researchers’ Windows-based simulation environments is not available in typical HTCondor pools. Rather than investing significant effort in development, porting, and testing their workflows to map to a different software environment, the researchers have focused on encapsulating their simulation environment in virtual machine (VM) images. Their preferred environment can be deployed, on-demand, on cloud infrastructures.

While VMs provide a basis for customizing software environments, another challenge remains: how to inter-connect VMs deployed across multiple institutions and possibly commercial clouds so that HTCondor and the simulations run seamlessly? The approach to address this problem is also to apply virtualization at the network layer, with the IP-over-P2P (IPOP) software-defined network developed at the University of Florida PRAGMA site.

The open-source IPOP overlay virtual network (www.ipop-project.org) allows the researchers to define and deploy their own virtual private network (VPN) that can span physical and virtual machines in their home institution, collaborating institutions, and commercial clouds. In a pilot implementation in the summer of 2013, students demonstrated the ability for Windows VMs configured with IPOP and deployed on Amazon’s EC2 cloud to automatically join a HTCondor pool that runs at the researchers’ lab. This is possible because IPOP provides an isolated virtual private address space, and self-configures overlay table routes in a manner that transparently deals with firewalls and network address translators (NATs).

How is it working? The power of networks – human, data, and computer – is paying off. Through the Global Lake Ecological Observatory Network (GLEON), Hanson, Carey and colleagues have amassed data from exemplar lakes, and the GLEON network is queuing up more. Numerical simulation software, provided by GLEON colleagues in Brazil, is being configured and calibrated to lakes. At the same time, GLEON scientists with experience in HTCondor are working with PRAGMA computer scientists from Florida to learn how overlay virtual networks can provide access to hundreds, and eventually thousands, of virtual machines to run simulations. The payoff may be a vastly improved capacity to understand and predict the diversity and dynamics of life in lakes, and thus better predict water quality in lakes.

Expedition Team: U Wisconsin: Paul Hanson, Virginia Tech: Cayelan Carey, U Florida: Renato Figueiredo

Expedition Students: U. Wisconsin: Grant Langlois, Luke Winslow; Virginia Tech: Jonathan Doubek U. Florida: Pierre St. Juste, Kyuho Jeong