A TeraGrid Science Gateway for Studying Hydrological Extremes using a Land Data Assimilation System

A TeraGrid enabled science gateway to study surface hydrology is under development at Purdue University. Such a gateway is needed for class room and graduate research as well as for developing a multidisciplinary framework that integrates environmental observatories and tools for drought/flood mapping and assessment. The gateway is expected to be a resource for developing class room activities in which students can create if-then scenarios and get familiar with issues such as the interlinking between land use change and regional hydrological changes, the complexities in understanding the onset or demise of a drought, the linkage between regional air and water quality and surface hydrology, and the integration of multi-scale datasets from surface and remote sensed satellite data into a modeling and synthesis framework. Building on this initial focus, the overall objective is to make this gateway an online tool for regional policy makers and researchers to integrate their data or simulate and visualize their case studies for hypothesis tests or scenario development.

The computational framework for such a gateway builds on a Land Data Assimilation System (LDAS) involving the Community Noah Land Surface Model (LSM). The data assimilation system simulates spatial and temporal changes in land-state variables, such as soil moisture and soil temperature and surface evaporation/transpiration and heating. The model fields are often used as boundary conditions for weather forecast and environmental models such as the coupled Weather Research Forecasting (WRF) model. The LDAS framework provides the means for integrating the heterogeneous and discrete surface and remote sensed datasets as they become available into a gridded field. The initial conditions are then used in conjunction with offline meteorology to develop gridded surface and subsurface fields. Currently there are several challenges for users to perform research using the LDAS model. The simulation process is computationally demanding. For a nested domain over upper great lakes states including Indiana a 12km grid spacing based outer domain is set up with 290x248 grid points with a 4 km spatial resolution inner domain with 462x387 grid points over Indiana. Typically this system has to be run for several seasons to generate results that can be used for developing the water balance and related surface hydrological studies. Such a run could take about 48 hours of clock time to complete the simulation process with a single CPU personal computer. A second challenge is the if-then scenarios that are of interest often need manipulation of software and input files which is typically possible only for an advanced user. This restricts its use only to trained researchers.

To overcome these challenges, the TeraGrid science gateway now integrates data management, LDAS simulation, data processing, and visualization capabilities in a single web based system. The gateway hides the complexity of the model setup and eliminates the need for users to learn the technical details of TeraGrid. The portal was developed using the GridSphere portal framework. It provides a wizard type interface for users to configure the parameters of a LDAS simulation, submit it to the Steele cluster at Purdue, monitor the job status, and fetch the output. It also provides a number of visualization functionalities (such as contour maps, animations, and time series plots) using the NCAR Command Language (NCL) to help users extract meaningful information from the simulation results. Behind the scene, the portal sets up all the required software packages in a community software area at Purdue TeraGrid site. It uses a shared community account which makes it easier for portal users to access TeraGrid computation resources. Simulation for each LDAS domain is submitted as a DAGMAN job, which consists of two PBS jobs for data consolidation and Noah LSM simulation as needed by LDAS simulation.

This portal provides new research opportunities by making available the LDAS model and the TeraGrid resources to a wide user community. For example, the portal has been integrated with the C4E4 (Cyberinfrastructure for End-to-End Environmental Exploration) framework which is an NSF funded cyberinfrastructure for environmental observatories (CEO:P) project to study environmental impacts over regional watershed areas using heterogeneous data sets and modeling tools. The LDAS fields can be used as the input of the Soil and Water Assessment Tools (SWAT) model which is part of the C4E4 system and also enabled by TeraGrid resources at Purdue. The portal is also being used to study drought onsets and analyses in a NSF INTEROP project for a regional drought information network called DRInet and a student learning environment as part of a NSF CAREER project.