A Community Model for Water in the Continental Earth System

Summary

Mathematical models are important tools to understand water in the continental Earth system, but none of the currently available models are adequately complete to tackle the diverse set of critical science questions facing society today. The large number of available models has limited collaborative development within the water sciences and limited interaction with sister disciplines, like the atmospheric and ocean sciences. This document outlines a plan to create a Community Water Model, which will be designed to provide the best available simulations of the hydrologic cycle and coupled water processes within the contiguous United States and adjacent watersheds. This will be accomplished by developing a modeling platform patterned after current multiphysics modeling software that has been optimized to solve problems related to water. The platform will provide the computational resources and access to data required to create a Community Water Model. The CWM will enable multiple workflow configurations to meet the wide range of objectives for which hydrologic models are constructed. It will be an open source code managed by the scientific community and it will grow in capabilities through contributions from the community. The Community Water Model can serve as the connection point for other disciplines that require prediction of continental water movement including general circulation, landform evolution, biogeochemical, vegetation, ecological and economic models.

Preface

The following document draws on ideas from the three CHyMP workshops organized by CUAHSI between March 2008 and March 2011. The CHyMP workshops recommended pursuing many topics to advance community modeling in water science, but the scope of the following document has been intentionally limited to a subset of those recommendations. The full recommendations from the workshops are in the workshop report (*Famiglietti et al*., 2011).

Introduction

Many scientists working on the movement and storage of water in surface, atmosphere and shallow subsurface systems use computer model simulations as an important tool in their investigations. Because of the vastly different dynamics of water in the atmosphere, on the surface and in the subsurface of the continent, separate models have been developed historically for each of these domains with the other domains treated as boundary conditions. Recent advances have enabled hydrologic models to be constructed that couple flow in the subsurface to flow on the ground surface, and some of them can also include coupling to water transport in the atmosphere. Furthermore, hydrologic models are routinely extended to consider coupled processes with vegetation, landforms, and chemical or mechanical interactions with rocks and biota.

These efforts, although promising, are currently fragmented across many research groups and across disciplines. This has resulted in redundancy and inefficiency, which have limited progress in water science. It has also hampered progress in atmospheric and ocean sciences, where simulators of weather, climate and ocean circulation are limited by the ability make use of the state-of-the-art simulations of continental water.

Many grand science challenges require fundamental understanding of water processes at the scale of a pore or plot, but this understanding must then be tested in other hydrogeologic settings and applied over large watersheds or continents. Most current simulators are unable to span the large range of spatial and temporal scales required to meet these demands. Other limitations are caused by complex interactions between coupled processes, which are essential to understanding critical phenomena. Coupling diverse processes across hydrologic domains that are dominated by different physics operating over a range of time scales is a challenge to current simulators.

Scientific progress could be accelerated by a new generation of simulator that integrates community knowledge and skills into an extensible *Community Water Model*. Such a model could serve as a benchmark against which to measure progress as new models are developed, and as a standard reference model for water movement that can be extended for use by all other disciplines where water plays a role.

A Community Water Model faces a substantial challenge in providing access to the necessary range of processes and scale, but there is precedent in this type of versatility with recent developments in multiphysics codes. We propose that the Community Water Model should be built on a cyberinfrastucture platform with the functionality of multiphysics codes, but with specialized capabilities developed to solve problems within the water science domain.

A Modeling Platform

One of the obstacles to achieving a Community Water Model is the diversity of processes that may be coupled over different scales. Water flow alone can be characterized by many differential equations (e.g. Richard's equation, depth averaged Navier Stokes, St. Venant, Kinematic wave, Diffusive wave, 3D incompressible Navier Stokes, Brinkman equation, among others), depending on the scale, the material through which flow is occurring, the degree of turbulence and other factors. An even broader range of equations are important when the many processes associated with water flow (e.g. mass transport, sediment transport, multi-phase flow, heat transport, poro-elasticity, interaction with vegetation or ecosystems, etc.) are considered.

There are many codes scattered across subdisciplines in the water sciences that solve these equations. One strategy to improve versatility is to use protocols like OpenMI or Common Component Architecture (CCA) used in Community Surface Dynamics Modeling System (CSDMS), where components can be extracted from different existing models and coupled together to create a new capability. Coupling existing models or model components by OpenMI or CSDMS leverages the significant library of legacy codes and is an important way to advance the science.

Hydrologic Multiphysics

An alternative approach is suggested by recent advances in software architecture designed from the start to solve problems where multiple physical or chemical processes can be coupled together in a variety of complex ways defined by the user. These so-called multiphysics codes (e.g. Comsol, CFD ACE+, Adina, Simulia, Ansys, Elmer, Algor) are capable of solving general problems in engineering and science, ranging from electromagnetics to acoustics, to fluid flow and transport. This approach solves multiple coupled equations in a single matrix and so for tightly coupled problems it can have significant advantages over componentized approaches.

Existing mulitphysics codes suggest that the broad computational versatility needed to implement a Community Water Model can be achieved. Multiphysics codes like Comsol use state-of-the-art numerical techniques that can be every bit as fast and robust as the numerical methods in specialized domain codes from water science. They have addressed and implemented full and segregated techniques for coupling equations representing different processes. Multiphysics codes use mesh generators that can create nested meshes over many orders of magnitude, providing a numerical foundation for addressing problems that require resolution over a large range of scales.

Despite their considerable strengths, current multiphysics codes also have several weaknesses. One of them is that multiphysics codes can perform poorly on some problems of particular importance to water science. This is understandable because current multiphysics codes are designed to solve a general set of problems, whereas some problems in water science require solving difficult, non-linear equations, like Richard's equation or the shallow water equations. Hydrologists have spent considerable effort to develop numerical techniques well suited to solving those problems, but their solution techniques are currently unavailable in general multiphysics codes. An additional weakness is that specialized techniques for configuring the spatial geometries required for analyses in water science (e.g. for representing topography or stratigraphy) are beyond the scope of general multiphysics codes. Another significant drawback is that the robust multiphysics software packages are sold commercially, so their source codes are proprietary and off limits to improvements from the community.

What is needed is a simulation platform with the versatility of current multiphysics codes, but with specialized, open-source capabilities developed to solve problems within the water science domain. We will call this a *hydrologic multiphysics* modeling platform. This platform would provide a single integrated tool designed with the versatility needed to analyze a range of important problems on multiple space and time scales.

Community Water Model

The objective of the Community Water Model (CWM) would be to provide the best available simulations of water-related processes anywhere in the U.S. and neighboring watersheds. Considerable progress has been made on this type of simulation, yet the best numerical solvers for specific problems and the appropriate process representation at any given scale remain open research questions. As a result, the Community Water Model needs to be built on computational infrastructure designed with the versatility to explore these questions. We suggest that the necessary infrastructure is a hydrologic multiphysics modeling platform.

An important aspect of the platform would be the ability to record workflow describing how a particular analysis was conducted. For example, workflow records would communicate how the geometry was constructed, how the mesh was generated and the solvers configured, as well as how the parameters were identified and myriad other details of the simulation. The results of different configurations would be evaluated by comparing them to a series of benchmark problems, and when a particularly powerful one was identified its workflow record would be available to replicate the configuration in other geographical areas or communicate the approach for other applications. Different workflow configurations describing water movement and its interaction with the earth system would be constructed to meet different objectives (such as flash flood prediction, nutrient cycling, or vegetation response to climate scenarios). Sharing these different workflow configurations will enable the community to explore alternate representations of processes, parameterizations, and model geometries to understand how these factors interact.

Ultimately, the CWM will be a collection of the best currently recognized platform workflows used to simulate water-related processes in the U.S. The CWM will evolve as understanding increases.

FUNCTIONALITY

The hydrologic multiphysics modeling platform would have the ability to:

- Simulate processes associated with flow and storage of all types of terrestrial water (ground water, vadose zone, streams, rivers, wetlands, lakes, snow/ice, etc.).
- Simulate processes over space scales from 10^{-1}m to 10^{6} m, and time scales from 10^{0}s to 10^{10}s.
- Simulate physical/chemical/biological processes involving water, including mass transport, biogeochemical reactions, erosion/deposition, porous media deformation, heat conduction and convection.
- Simulate interactions with ecosystems, including the natural and built environments.
- Simulate interaction of terrestrial water with atmosphere and oceans by coupling with existing atmospheric and ocean circulation models.
- Seamlessly access parameter data representing North America with the best currently available resolution.
- Implement a suite of mesh-based numerical methods (e.g. finite different, finite element, finite volume, lattice Boltzman) with the best available solvers to provide efficient solutions to diverse types of problems.
- Calibrate model parameters based on observations.
- Modify functions, coupling, processes, etc. for the user
- Enable access through a web browser.
- Document the model development workflow and make that information available to others including necessary documentation for publication.
- Provide transparent access to high-performance and high-throughput computing.
- Enable non-specialists to access state-of-the-art simulations of water
- Visualize results to facilitate interpretation.

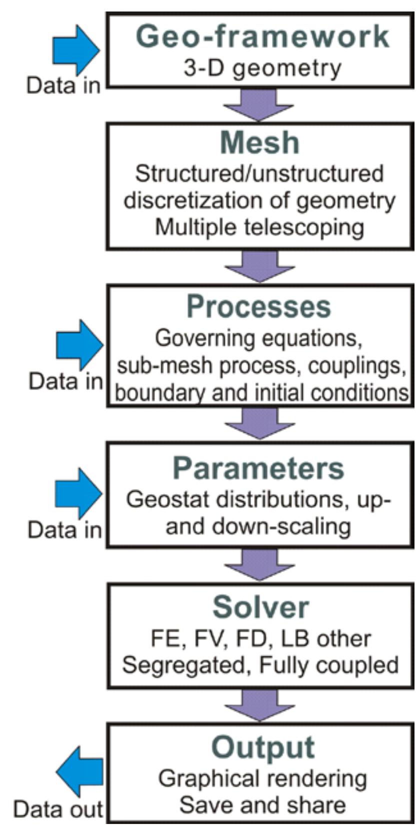

Figure 1. Major modules and basic workflow for developing a forward model using the platform.

THE PLATFORM ENVIRONMENT

The desired functionality is within reach by creating a cyberinfrastructure with six primary modules: a geometry builder, mesh generator, process selector, parameter database, numerical solver, and a module for rendering model output (Fig. 1). The six modules would be designed to work together in a graphical user interface. The platform would develop a forward analysis with workflow that implements a sequential application of the modules (Fig. 1). Estimation of parameter values can be improved by implementing an inverse module, which accesses the forward module along with external data, but is not shown in Figure 1.

Ready access to the data needed to calibrate and force simulations is as important to the success of the platform as access to advance numerical algorithms. The goal is for the platform to provide seamless access to the best available data for model development. Some data describing the continental environment are used by many classes of models. This includes standard datasets provided by federal agencies [e.g. topography (NED, ASTER), streams (NHDPlus), soil properties (STATSGO), weather (NAAR, RIST), ecosystems (MODIS, CDIAC), political boundaries and federally administered lands], datasets developed by the model calibration process, and new datasets that are compiled as part of this effort. Database development will be guided to support the processes included in the Community Water Model.

A spatial dataset describing hydrostratigraphy is among the most critical new datasets needed for water modeling. The integration of hydrostratigraphic datasets built around a data standard model (e.g. ArcHydro and ArcHydro Groundwater) will enable a rapid approach to formulating high-resolution simulations in the CWM.

IMPLEMENTATION

The modules, and the workflow that connects them (Fig. 1), would take advantage of proven modeling strategies to ensure rapid adoption and productive application by the community. Many modeling packages and platforms currently exist with modules similar to the ones outlined above, and the implementation strategy would start with an identification of the requirements for each module and an assessment of the capabilities of existing codes.

As an example, many of the modules outlined in Figure 1 are implemented in PIHM (Qu and Duffy, 2007, Kumar and Duffy, 2010) and this software is being integrated with data sources on the desktop (PIHM_GIS) and in WEB services, so it could serve as a good starting point. Physics modules currently used in PIHM would require modification, allowing new process equations and/or parameterizations. A programming strategy that uses object data structures will be important to making a flexible community code useful for a range of purposes and processes. A strong 3-D component to the Geo-framework module would be needed to represent subsurface geology, and this may best be accomplished by implementing the ArcHydroGW data model and using software techniques implemented in geologic models, like Subsurface Analyst. Another modification is the integration of additional numerical techniques. This could be accomplished by including finite difference and finite element equation generators and solvers that have been well tested by the community and that are currently using High Performance Computing (HPC). Parflow (Maxwell and Miller, 2005) and HydroGeoSphere (Sudicky et al. 2008) meet these criteria. Integrating the equation solving capabililities of PIHM, ParFlow and HydroGeoSphere on a single platform for creating geometries, integrating with data, generating meshes, and displaying results would go a long way to realizing an initial version of an extensible and flexible Community Water Model.

An initial implementation of the platform will grow with new contributions. One example would be the integration of carbon and nutrient cycling into the CWM, which would couple processes of water flow with equations describing fate and transport of compounds in ecosystems. This could involve solving the necessary equations within the platform, or coupling to an external code, such as RHESSys (Tague and Band, 2004). Domain experts for modeling carbon and nitrogen cycling would implement the best approach. Another extension is the ability to simulate 3D hydrodynamic processes in rivers, lakes and estuary settings. This capability would require including an additional set of equations and (possibly) a new solver to provide the turbulence closure that is needed in some surface water simulations. An initial implementation would make use of equation generators and solvers from tested codes, such as EFDC (Hamrick, 1996), or CWR-ELCOM (Hodges, 2000). Another approach is the hydrodynamic code PIHM_Hydro (Li and Duffy, 2011) which incorporates multiphysics strategies for sediment transport and bed evolution with GIS tools for domain decomposition and parameterization.

An important aspect of the platform is that the modules used to develop the geo-framework, obtain parameters, generate mesh and displace results for subsurface problems would be versatile enough to also set up problems involving surface water hydrodynamics or biogeochemical cycling. Moreover, numerical capabilities to do either full coupling, where all the equations are solved within a single matrix, or segregated coupling, which is accomplished by iteration between two separate solvers, would be implemented following similar approaches demonstrated in codes like Comsol or Fluent. Seamless integration with HPC for both equation solving and fast access to large datasets will be another important aspect.

While the strategy will be to implement full coupling of processes within the platform solvers, it will be impractical to represent all processes in this way. Some processes will be included by coupling the platform with existing simulators from other disciplines, like atmospheric and ocean circulation models, but also including landform evolution, biogeochemical, ecological, economic and other models. This will advance the capabilities of the platform, while also providing improved representation of continental hydrology to models from other disciplines.

Decision Points

The modeling platform to support the Community Water Model will be developed through a series of stages that are defined by decision points.

1. **Design** General architecture of the platform and coding approach identified, along with software methods used for modules. Proof-of-concept established through benchmarking of functional prototypes.
2. **Forward modeling** Basic functionality of the forward modeling workflow (Fig. 1) demonstrated with a minimum of two numerical methods and associated solvers. Data models for I/O established.
3. **Online access** Forward modeling capabilities (Fig. 1) available as an on-line resource. Transparent access to HPC.
4. **Community contributions** Demonstrate the ability of community to contribute code, data, and workflow. Implement community governance.

5. **Data integration** Linkages to data sets established and new datasets created. Initial continental hydrostratigraphy dataset established with mechanism for community contribution.
6. **Community Water Model** The ability to simulate water-related processes with the best currently available hydrology and resolution made available to on-line users.
7. **Inverse modeling** Additional functionality, including inverse modeling, up-scaling, and stochastic methods made available.

Community contributions

The platform would be a resource for the broad, Earth system science community, including researchers from across academia, government agencies, national labs, and the private sector. Initial design and construction would be done by a core development team of experts in domain modeling and software engineering. Subsequent developments that improve the modules or create new ones will evolve from contributions across the water science community and other disciplines.

In addition to code, the water science community will also contribute workflow describing how the platform was used to solve a particular problem. Workflow records identifying how the platform was configured to solve a problem will allow a particular approach to be documented and communicated to others (i.e., the workflow could be published).

Governance

The platform will be governed by representatives from the hydrologic and broader scientific community. An initial vision for governance includes a Scientific Steering Committee (SSC) and several Working Groups. The SSC would be responsible for overall coordination, while the Working Groups will evaluate and recommend modifications to the platform as well as coordinate input from the community. Computational innovations developed by the community would be evaluated and reviewed by the Working Groups and then integrated into the platform for beta testing and adoption.

Role of CUAHSI

CUAHSI will provide logistics to coordinate the large team of contributors involved in developing and governing the platform. High-level governance and oversight can be provided by the CUAHSI Informatics Standing Committee. The CUAHSI Data Center will play an important role in hosting and managing the on-line infrastructure, as well as the new datasets required for the platform to be an operational service. New computational hardware required to implement the platform would be operated through the CUAHSI Data Center.

REFERENCES

Famiglietti, J. S., L. C. Murdoch, V. Lakshmi, J. Arrigo and R. Hooper, 2011. Establishing a Framework for Community Modeling in Hydrologic Science. *Report from the 3rd Workshop on a Community Hydrologic Modeling Platform (CHyMP):* A Strategic and Implementation Plan, Beckman Center of the National Academies, University of California at Irvine. March 15-17, 2011.

Hamrick, J.M., 1992. Estuarine environmental impact assessment using a three-dimensional circulation and transport model. *Estuarine and Coastal Modeling, Proceedings of the 3rd*

International Conference, M.L. Spaulding et al, Eds., American Society of Civil Engineers, New York, 292-303.

Hodges, B.R., 2000. *Numerical Techniques in CWR ELCOM*, Centre for Water Research, Technical Report WP1422BH The University of Western Australia.

Li, S., and C. J. Duffy (2011), Fully coupled approach to modeling shallow water flow, sediment transport, and bed evolution in rivers, *Water Resour. Res.*, 47, W03508, doi:10.1029/2010WR009751

Maxwell, R.M. and N.L. Miller. 2005. Development of a coupled land surface and groundwater model. *Journal of Hydrometeorology*,6(3), 233-247, 2005.

Mukesh Kumar, Christopher J. Duffy, and Karen M. Salvage, A Second-Order Accurate, Finite Volume–Based, Integrated Hydrologic Modeling (FIHM) Framework for Simulation of Surface and Subsurface Flow. *Vadose Zone J.* 2009 8: 873-890

Qu, Y, and C. Duffy. 2005. A semidiscrete finite volume formulation for multiprocess watershed simulation. *Water Resources Research*. vol. 43, W08419, doi:10.1029/2006WR005752, 2007.

Sudicky, E.A., Jones, J.P., Park, Y.-J., Brookfield, A.E. and Colautti, D., 2008. Simulating complex flow and transport dynamics in an integrated surface-subsurface modelling framework, *Geosciences Journal*, 12(2), 16 pp., doi:10.1007/s12303-008-0013-x

Tague, C.L., L.E. Band, 2004. RHESSys: Regional Hydro-Ecologic Simulation System—An Object-Oriented Approach to Spatially Distributed Modeling of Carbon, Water, and Nutrient Cycling. *Earth Interactions*. 8: 1-42.