

High Performance Computing for Hydrological Sciences

Venkat Lakshmi

(vlakshmi@stanford.edu)

Abstract

The purpose of this short document is to demonstrate the readiness of hydrological sciences to use high performance computing and utilization such facilities for the advancement of Earth sciences and environmental prediction.

Introduction

High performance computing has been long used in the disciplines of atmospheric and ocean sciences due to their dependence on fine spatial grids and small time steps for integration. Historically, the land surface interaction with the atmosphere, biosphere and ocean had been represented by lumped, or spatially-aggregated, models. Lumped models do not have large computational requirements. However, with the growing recognition that the spatially distributed feedback from the land surface is important to weather events and the climate system, representation of the land surface is done with increasingly complex (and physically complete) models. As a result, the spatial grids and the temporal resolutions have become finer and thereby warrant computers with far greater computational and storage capacity than those used in the past.

In addition to coupled atmosphere-biosphere-ocean-land surface problems, there are a host of additional hydrological problems that require larger computer resources. These are – coupled groundwater-surface water-land surface mesoscale modeling; land-atmosphere-biosphere interaction modeling, calibration and uncertainty analysis of hydrologic systems; ensemble hydrological forecasting; land information systems and continental water dynamics.

In this document we attempt to briefly explain some of the science and the computational requirements associated with each of the above listed hydrological problems. Whereas we recognize that this list of problems is by no means exhaustive, it is a reasonable starting point for the utilization of high performance computing in hydrology.

We organize the document as follows: We start with the coupled surface water-subsurface and groundwater system which has a very heavy computational requirement for solving coupled atmosphere-land surface-subsurface equations. This is an area of obvious importance to future societal demands; surface water withdrawals impact the groundwater table and also impact the land-atmosphere fluxes. The land surface-atmosphere-biosphere modeling is an area of impact for weather forecasting where the influence of the land surface needs to be taken into account. The short time steps for integration and fine spatial resolution modeling is the direction of the future in this area of research. The third area of study concerns uncertainty and calibration of the basic hydrologic unit – the watershed. The first two areas of study deal with larger spatial areas

and mostly water quantity (and not quality), this third area deals with smaller watershed hydrological fluxes of water and nutrients. Fourthly there is an example of coupling water balance and water quality modeling across a series of spatial scales. Then, we deal with ensemble hydrological forecasting with Kalman filtering techniques followed by land information system that sets a framework for global land surface hydrological modeling. Lastly, all of this is integrated by continental water dynamics which is a continental scale routing of water fluxes through overland flow followed by stream and river routing.

Several authors have contributed to this document. The diversity of the scientific questions as well as accounting of the computational requirements conveys to the reader the necessity and urgency of high performance computing in hydrological sciences.

Coupled Groundwater-Surfacewater-Landsurface-Mesoscale Modeling

(Reed Maxwell, maxwell5@llnl.gov, Stefan Kollet, kollet2@llnl.gov and Tina Chow, tinakc@berkeley.edu)

ParFlow watershed model:

The core of the integrated watershed management modeling framework consists of ParFlow (Ashby and Falgout, 1996; Jones and Woodward, 2001), a parallel, variably saturated groundwater flow code that is especially suitable for large scale problems. ParFlow's development has been on-going for more than 10 years and resulted in some of the most advanced numerical solvers and multigrid preconditioners for massively parallel computer environments that are available today. Many of the numerical tools developed within the ParFlow platform were turned into libraries that are now distributed and maintained at LLNL. An additional major advantage of ParFlow is the use of a sophisticated octree-space partitioning algorithm to depict complex structures in three-space, such as topography, different hydrologic facies, and watershed boundaries. All these components implemented into ParFlow enable large scale, high resolution watershed simulations. We are not aware of any computational platform that is able to provide similar performance. ParFlow simulates the three-dimensional variably saturated subsurface flow in strongly heterogeneous porous media in three spatial dimensions. Thus, lateral flow and redistribution of soil moisture are explicitly accounted for in the simulations. Advanced boundary conditions, such as a free-surface overland flow, afford the simulation of hillslope runoff and channel routing in a truly integrated fashion (Kollet and Maxwell, 2006). Topographic information needed as input is derived from widely available digital elevation models at varying resolution using simple utility programs. Distributed surface roughness can be applied to honor different land cover types in the watershed.

Recently, the ParFlow platform was extended to incorporate physical processes that are related to the energy and mass balance at the land surface. This has been done by integrating a land surface model, namely the Common Land Model (CLM; Dai et al, 2003), into ParFlow (Maxwell and Miller, 2005). The new coupled model is able to quantify the physical feedbacks that occur between the land surface such as plant interception, uptake and evapotranspiration and the subsurface. The rationale for this approach is the oversimplification of the upper boundary condition common in groundwater flow models that cannot account for surface processes such as evapotranspiration, which is an important transient sink term in the water budget. On the other hand, land surface models, widely used in watershed simulation activities, oversimplify the subsurface by only accounting for the shallow root zone and neglecting horizontal distribution of soil moisture and groundwater, which is especially important in long term simulations and accurate representation of the base flow component of hydrographs. CLM simulates all the major components of the energy and mass balance at the land surface: evaporation from canopy and the ground surface; transpiration from plants; ground heat flux; freeze-thaw processes; sensible heat fluxes; etc. Spatially distributed land cover and soil cover are explicitly honored. Atmospheric data, such as temperature, precipitation, solar radiation, humidity, and barometric pressure, are used to force the model and can be applied in a distributed fashion as well. By coupling ParFlow

with CLM, we arrive at a watershed modeling tool that uses atmospheric forcing to drive the simulations; simulates the mass flow from the subsurface to the surface in three spatial dimensions including deep groundwater flow; and accounts for major surface processes. In this modeling approach, the most important outcome is that feedbacks, which have been previously ignored, are now honored explicitly. For example, three-dimensional convergence of groundwater flow at discharge zones, i.e. river valleys, results in higher soil moisture values, which directly impact the ground heat fluxes and evapotranspiration values. Changes in these components are fed back into the model immediately by impacting the soil moisture distribution along the convergence zones, i.e. river corridors. Common modeling approaches are not able to quantify these types of interactions, which constitutes a major deficiency.

The ParFlow-CLM platform is currently applied to the Little Washita watershed, a 600 square kilometer watershed located in Southwestern Oklahoma which is part of the DOE Atmospheric Radiation Monitoring (ARM-CART) facility. This application, which is summarized here, illustrates the major aspects of the coupled model. The major input include the aforementioned atmospheric forcing, land and soil cover information, topography, and a geologic model consisting of the different hydrofacies of the subsurface. In order to obtain a realistic initialization of the major state variables, such as soil moisture and temperature, the model is run repeatedly over one year using the same forcing until a dynamic equilibrium is reached (commonly referred to as model spinup). The results of this model spinup can then be used to perform pathway analyses of contaminants and pathogens or to perform scenario calculations under varying atmospheric forcing and land cover conditions. Figure 1 below shows the soil moisture distribution in the Little Washita watershed in July after the spinup is completed. The non-uniform moisture distribution is a result of the three-dimensional subsurface hydrodynamics and the non-uniform soil and land cover distributions that were derived from USGS data sets. Note, groundwater converges and discharges in the valley, forming the Little Washita River.

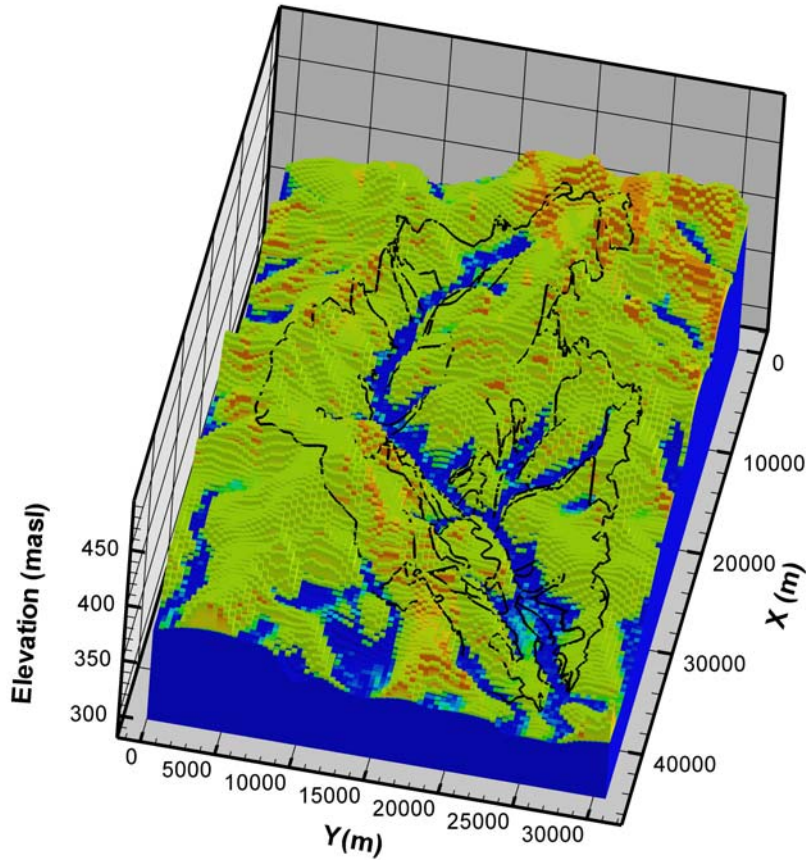


Figure1 Three dimensional snapshot of the simulated saturation in the Little Washita Watershed looking in the westward direction. The saturation values range from blue (fully saturated) to red (residual saturation) and the watershed outline has been overlain at the ground surface.

Coupled groundwater-atmosphere model:

The final component of our integrated watershed modeling framework involves coupling to a mesoscale atmospheric model. Mesoscale models currently rely on an integrated land-surface model to provide fluxes of heat, momentum, and moisture from the land surface to the atmosphere. While improvements have been made by tuning land-surface models for a variety of test cases, the models are limited to vertical transport in the soil column. They are thus unable to capture lateral variations in soil moisture and limited in their ability to provide spatial variability in predicted land surface fluxes. Current mesoscale atmospheric models are therefore not provided with realistic forcing at the surface because land-use models cannot represent runoff and subsurface lateral transport that is present when terrain or moisture gradients exist. This can lead to serious errors in model predictions during periods when thermal forcing dominates the diurnal development of the boundary layer.

Our integrated watershed model framework couples ParFlow with a mesoscale atmospheric model (ARPS) to investigate the effects of soil moisture heterogeneity on boundary layer processes. ARPS (Advanced Regional Prediction System) is a mesoscale atmospheric model developed for storm-scale applications. The coupled model, known as PF.ARPS, is fully parallel. At each time step, ParFlow provides ARPS with soil moisture information that includes the effects of ponding, runoff, and seepage that occur. In turn, ARPS, through its land-surface model, provides ParFlow with precipitation and evapotranspiration. This leads to a fully coupled model which can represent spatial variations in land-surface forcing driven by 3D atmospheric and subsurface components.

Our test case is the Little Washita watershed in Oklahoma, which has been the subject of numerous studies and provides a unique source of both subsurface, surface, and atmospheric data for validation. Results indicate sensitivity of convective storm generation and precipitation events to soil moisture fields. Soil moisture fields from PF.ARPS provide much more realistic spatial variability and can keep the shallow soil wetter than ARPS because of groundwater replenishment. Studies are underway to determine the quantitative impact of antecedent soil moisture fields on atmospheric convection and precipitation processes.

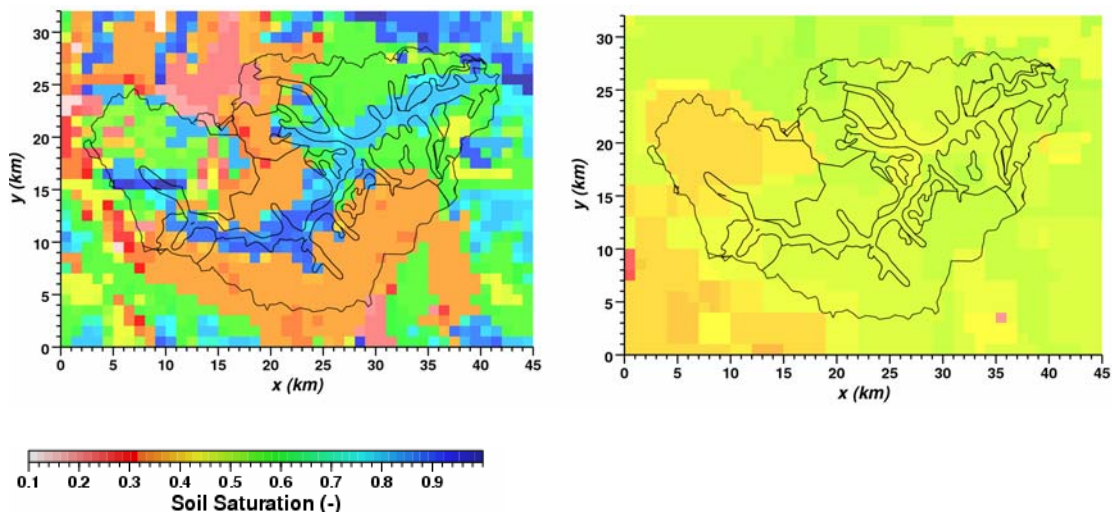


Figure 2 Soil moisture fields from PF.ARPS (left) and ARPS (right). Watershed outline for the Little Washita River is shown in black contours. PF.ARPS provides realistic spatial variability. ARPS uses initial data from regional reanalysis forcing (NARR) and does not allow lateral transport.

Please refer to Appendix I for a listing of computational details for various modeling scenarios.

References

- Ashby, S.F. and R.D. Falgout (1996). A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations. *Nuclear Science and Engineering*, 124(1):145-59.
- Chow, F.K., Kollet, S.J., Maxwell, R.M., and Q. Duan. 2006. Effects of soil moisture heterogeneity on boundary layer flow with coupled groundwater, land-surface, and mesoscale atmospheric modeling. Paper 5.6. *17th Symposium on Boundary Layers and Turbulence*, American Meteorological Society, 6 pages.
- Dai, Y, X Zeng, RE Dickinson, I Baker, BG Bonan, MG Bolsilovich, SA Denning, PA Dirmeyer, PR Houser, G Niu, KW Olsen, CA Schlosser and Z-L Yang (2003). The common land model. *Bulletin American Meteorological Society*, 1013-1023.
- Jones J.E. and C.S. Woodward (2001). Newton-krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Advances in Water Resources*, 24:763-774.
- Kollet, S.J. and Maxwell, R.M., Integrated surface-groundwater modeling: A free surface overland flow boundary condition in a parallel groundwater flow model, *Advances in Water Resources* (29), 2006 pp 945-958
- Maxwell, R.M. and N.L. Miller (2005) Development of a coupled land surface and groundwater model. *Journal of Hydrometeorology*, 6(3) 233-247.

Continental-Scale High Resolution Land Data Assimilation System for Coupled Land-Surface/Atmospheric Models

(Fei Chen, feichen@ucar.edu, Kevin Manning, kmanning@ucar.edu, Venkat Lakshmi, vlakshmi@stanford.edu, Eric Small, Eric.Small@colorado.edu, Dev Niyogi, dniyogi@purdue.edu; Peggy LeMone, LeMone@ucar.edu, Stan Trier, Trier@ucar.edu)

Although the important role of soil moisture in deep-convection development has been recognized, it remains the most difficult variable to obtain, because there is no routine high-resolution soil moisture observation at the continental scale. The soil moisture fields from the 4-dimensional data assimilation (4DDA) systems of coupled land/atmosphere models suffer substantial errors and drift owing to model precipitation, temperature, and radiation biases. An alternative strategy is to utilize precipitation observed by gauge/radar, satellite-observed surface solar insolation, and meteorological analysis to drive an off-line simulation of a LSM, so that the evolution of soil moisture does not suffer from the model biases in surface forcing fields. Hence, we developed a High-Resolution Land Data Assimilation System (HRLDAS, Chen et al., 2006). HRLDAS simulates the entire surface energy and water budgets including surface latent and sensible heat fluxes, soil moisture and temperature, surface runoff, and water table recharge. Figure 3 shows an example of HRLDAS application

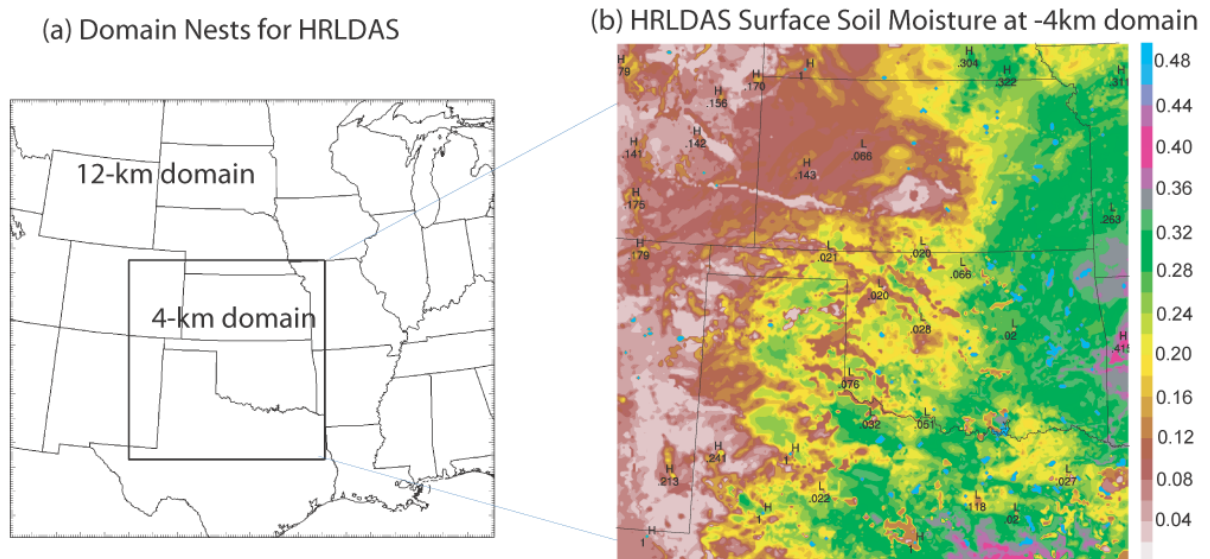


Figure 3 Nested grids used for the HRLDAS IHOP-2002 experiment. (a) 12-km outer grid and 4-km inner grid; (b) HRLDAS surface volumetric soil moisture (contours in $0.02 \text{ m}^3 \text{ m}^{-3}$ intervals starting from $0.02 \text{ m}^3 \text{ m}^{-3}$) valid at 1200 UTC 29 May 2002.

Understanding the feedback between land-surface variability and precipitation has been a central issue, because of its potential benefit in improving weather, hydrologic, and climate predictability. In a study focusing on heavy precipitation associated with a dryline in the south central US, Trier et al. (2004) found that fine-scale ($L \sim 10 \text{ km}$) boundary-layer circulations that directly trigger deep convection are confined within a mesoscale region containing a deeper and more unstable Planetary Boundary Layer

(PBL), and that this region is a result of a surface sensible heat-flux maximum over dry soils. They utilized the soil moisture fields from HRLDAS and from the National Centers for Environmental Prediction (NCEP) Eta Model Data Assimilation System (EDAS) to initialize the MM5 model. The simulation initialized with EDAS soil moisture did not initiate deep convection along the dryline in Texas (see Figure 4) as shown in the satellite image because of subtle differences in soil moisture and in the subsequent evolution of the boundary layer. By contrast, more accurate fine-scale soil moisture and temperature conditions significantly improved the 12-hour forecast of location and amount of this heavy precipitation event.

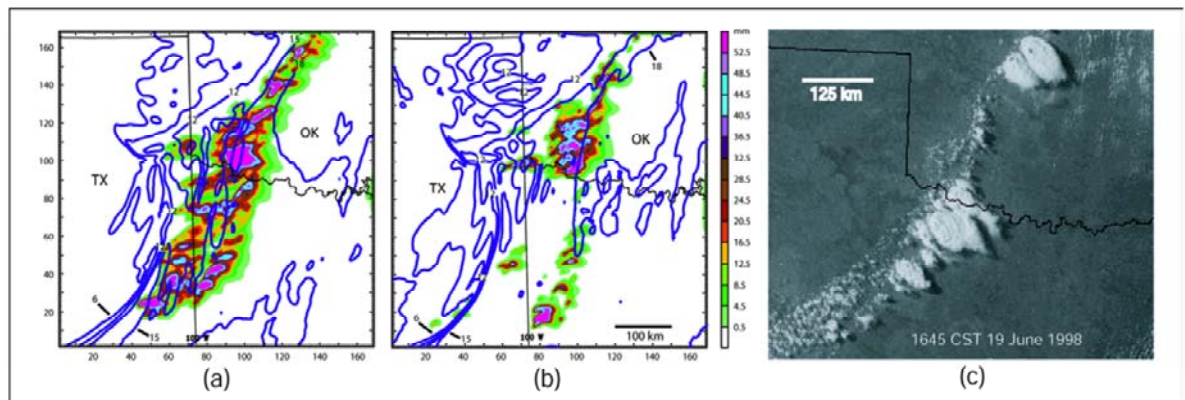


Figure 4 Simulated 3h precipitation from 1500 to 1800 CST 19 June 1998, color coded at 4 mm intervals starting at 0.5 mm, and 1500 CST 19 June surface water vapor mixing ratio in blue contours, with 3 g kg⁻¹ intervals starting at 6 gkg⁻¹, for the MM5 3-km domain: (a) MM5 simulation initialized with HRLDAS soil moisture, (b) MM5 simulation initialized with NCEP EDAS soil moisture, (c) GOES-8 visible satellite imagery for 1645 CST 19 June 1998.

References

Chen, F. K.W. Manning, M.A. LeMone, S.B. Trier, J. G. Alfieri, R. Roberts, M. Tewari, D. Niyogi, T. W. Horst, S. P. Oncley, J. B. Basara, and P. D. Blanken, 2006: Description and Evaluation of the Characteristics of the NCAR High-Resolution Land Data Assimilation System. *J. Appl. Meteor. and Climate*, in press.

Trier, S., F. Chen, and K. Manning, 2004: A Study of convection initiation in a mesoscale model using high-resolution land surface initiation conditions. *Mon. Wea. Rev.*, 132, 2954-2976.

Calibration and Uncertainty Analysis for Hydrologic Systems (Chris Shoemaker, cas12@cornell.edu)

There exists a role for HPC in developing high quality, data- and process-based models of specific areas that give us hydrologic knowledge and understanding and also help us quantify the uncertainty associated with model predictions. Some model parameters need to be calibrated to insure site-specific characteristics are incorporated. Model forecasts give valuable information, but there is an increasing recognition that the uncertainty associated with these forecasts needs to be quantified. A thorough calibration can require thousands of simulation runs depending on the number of parameters being calibrated and the procedure used for calibration. An uncertainty analysis using the well known methods GLUE (e.g. Beven and Freer, 2001) or MCMC (Kuczera and Parent, 1998) can easily require over 10,000 simulations. Hence HPC is necessary for calibration and uncertainty analysis even if the individual model can be simulated in a reasonable amount of time (e.g. 10 minutes) on a PC.

Most of the previous applications of automatic calibration and uncertainty methods for watersheds have been to rainfall- runoff models that take only seconds per simulation (e.g. Duan et al., 1992). However, with access to HPC, we can apply automatic calibration uncertainty analysis to watershed models that are spatially distributed and include water quality information. If a spatially distributed watershed model takes 10 minutes to run over the data period, the CPU time required is about 2 months to do 10,000 simulations in serial. This is certainly the kind of problem that should be done using HPC.

The water quality/biogeochemistry aspect typically is more uncertain in its response to a specific precipitation event than is flow. Hence calibration and uncertainty are considerably more difficult, meaning an increase both in the simulation time and in the number of simulations required to do uncertainty analysis when biogeochemistry and water quality are included in addition to hydrologic flow.

Groundwater calibration has been done using derivative-based methods like PEST (e.g. Doherty and Johnston, 2003) for unimodal problems and less efficient heuristics like genetic algorithms for multimodal problems. However, for general nonlinear groundwater problems there is no guarantee that the calibration problem is unimodal and so the global optimization methods have to be used. There has been a lot of investigation of risk (e.g. Buenos et al. 2006) and uncertainty analysis in groundwater, but the computationally fast methods rely upon linear relationships. For a general nonlinear groundwater problem, for example one involving reactive chemistry and microbiology, calibration and uncertainty can be highly computationally demanding (Mugunthan et. al. 2005, Mugunthan and Shoemaker, 2006). And of course for hydrometeorological models that are already computationally demanding for just a single simulation there is even in more need of HPC for model calibration and uncertainty analysis.

There is currently considerable research focuses on computational methods to reduce the computational requirement for calibration and uncertainty analysis (e.g. Regis and Shoemaker, 2006b, Vrugt et al. 2003, Tolson and Shoemaker, 2006, Mugunthan and

Shoemaker, 2005,2006, Blizniouk et al. submitted). However, even using newer methods, the time requirements to do calibration and uncertainty analysis are going to be large if it is done thoroughly with spatially distributed simulation models that have some description of processes affecting water quality and biogeochemistry as well as flow. Some of these newer methods have already been developed for parallel/distributed processing by HPC (e.g. Regis and Shoemaker, 2006a).

Hence HPC is important for all hydrological systems for which a nonlinear spatially distributed model is necessary if the model is to be calibrated and especially if an uncertainty analysis is desired.

References

Benekos, I.D., C.A. Shoemaker, J.R. Stedinger, "Probabilistic Risk and Uncertainty Analysis for Bioremediation of Four Chlorinated Ethenes in Groundwater", **Stoch. Environ. Res. Ris. Assess.** DOI 10.1007/s00477-006-0071-4, 2006a

Beven, K. and J. Freer "Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology," **Journal of Hydrology**, 249(1-4):11-29, 2001

Blizniouk, N., D. Ruppert, C. Shoemaker, R. Regis, S. Wild, P. Mugunthan, "Bayesian Calibration of Computationally Expensive Models Using Optimization and Radial Basis Function Approximation," submitted paper

Doherty, J. and J.M. Johnston, "Methodologies for Calibration and Predictive Analysis of a Watershed Model," **Jn. Of the Amer. Water. Res. Association**, 2003

Duan, Q. S. Sorooshian, and V. Gupta, "Effective and efficient global optimization method for conceptual rainfall-runoff models," **Water Resources Research**, 28(4), 1015-1031, 1992

Kuczera, G. and E. Parent, "Monte Carlo Assessment of Parameter Uncertainty in Conceptual Catchment Models: the Metropolis Algorithm," **Journal of Hydrology**, 211 (1-4): 69-85.

Mugunthan, P., C.A. Shoemaker, R. G. Regis "Comparison of Function Approximation, Heuristic and Derivative-based Methods for Automatic Calibration of Computationally Expensive Groundwater Bioremediation Models," **Water Resources Research** Vol. 41, W11427, doi: 10.1029/2005WR004134, Dec. 2005

Mugunthan, P., C.A. Shoemaker, "Assessing the Impacts of Parameter Uncertainty for Computationally Expensive Groundwater Models," **Water Resources Research**, *in press October 2006*

Regis, Rommel and C. Shoemaker, "Parallel Radial Basis Function Methods for the Global Optimization of Expensive Functions," **European Journal of Operations Research**, *in press 2006a*

Regis, R.G., C.A. Shoemaker, "A Stochastic Radial Basis Function Method for the Global Optimization of Expensive Functions", **INFORMS Journal of Computing**, *in press 2006b*

Tolson, B. and C.A. Shoemaker, "The Dynamically Dimensioned Search Algorithm for Computationally Efficient Automatic Calibration of Environmental Simulation Models," **Water Resources Research**, *in press 2006a*

Tolson, B. and C.A. Shoemaker, "Cannonsville Watershed SWAT2000 Model Development, Calibration and Validation," **Journal of Hydrology**, *in review 2006b*

Vrugt, J.A., H.V. Gupta, W. Boten, S.Sorooshian, "A Shuffled Complex Evolution Metropolis Algorithm for Optimization and Uncertainty Assessment of Hydrologic Model Parameter," **Water Resources Research**, 39, 8, 1201, doi:10.1029/2002WR001642,2003

Coupled Modeling of Hydroclimate and Nutrient Cycling: Applications to Water Quality and Water Balance Across Scales

C.M. Oldenburg¹ and N.L. Miller^{1,2} (nlmiller@Berkeley.edu), F. Maggi², C. Gu², W.J. Riley¹, T. Xu¹, L. Pan¹, C.I. Steefel¹, J. Jin¹, Y. Rubin^{2,3}, and G.M. Hornberger^{2,4}

¹Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA

²The Berkeley Water Center, Berkeley, CA

³Civil and Environmental Engineering Dept., University of California, Berkeley

⁴Department of Environmental Sciences, University of Virginia, Charlottesville

We are developing a coupled modeling capability for atmospheric, land surface, hydrologic, and biogeochemical processes for prediction and interpretation of the effects of climate change. This new biogeochemical transport modeling capability is currently focused on answering the question: *What are the effects of regional climate change in California on the dynamics of the nitrogen cycle in the Central Valley (CV)?* This problem is being addressed through the coupled modeling of climate, infiltration, vegetation dynamics, microbial processes, chemical reactions, and chemical transport within the near-surface, vadose, and saturated zones. The expected outcome of simulation studies with the new code is a better understanding of the nitrogen cycle and how it will change under various future climate, land-use, and hydrologic-change scenarios. In addition, this work will address problems associated with upscaling of detailed model processes, modeling hydrobiological processes at watershed to regional scales, and using local data to constrain larger-scale models (e.g., 1-km resolution, San Joaquin Basin domain).

Our approach to developing this simulation capability is based on existing established codes such as the LBNL subsurface simulators TOUGH2 (Pruess et al. 1999) and TOUGHREACT (Xu et al. 2004). Recently, Pan et al (2007) has coupled the National Center for Atmospheric Research (NCAR) Community Land Model version 3 (CLM3) with TOUGH2 (CLMT2), which considers water, carbon, and energy, but not reactive (biogeochemical) transport. On the biogeochemical side, prior developments of TOUGHREACT (Xu et al. 2004; Xu 2007) and CRUNCH (Steefel 2001) have generated the foundations for reactive biogeochemical transport. Additionally, Riley and Matson (2000) developed, tested, and applied a nitrogen cycling, land-surface model in agricultural settings. Here we are further developing the biogeochemical transport capabilities of TOUGHREACT with a focus on nitrogen to treat biogeochemical networks and nutrient cycling by including time-dependent biomass growth and death, multiple reaction networks, and advective and diffusive transport in aqueous and gas phases. This new code will be coupled to CLM3 (e.g., Pan et al., 2007) to produce a code capable of simulating climate effects on the nitrogen cycle. The familiarity and credibility of the models we are building upon (TOUGH2, TOUGHREACT, and CLM3) will encourage early adoption of this coupled modeling framework by outside research groups.

We have formulated reaction networks for nitrogen based on literature surveys, and implemented these into TOUGHREACT to create a unique coupled hydro-biogeochemistry simulation capability. Current efforts in the project are focused on verifying nitrogen cycle dynamics by comparison to field and laboratory data. Examples of comparisons of TOUGHREACT simulations against field data for a case of the application of anhydrous ammonia to a tomato field in Sacramento, CA (Venterea, 2000a, b) are shown in Fig. 1 a-d. The results suggest that the fundamental nitrogen dynamics processes are being simulated well with apparent differences likely arising from heterogeneity and variability such as rainfall events not included in the simplified model.

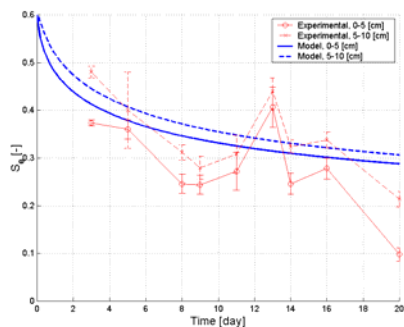


Figure 1a. Comparison between experimental and TOUGHREACT results of soil water saturation.

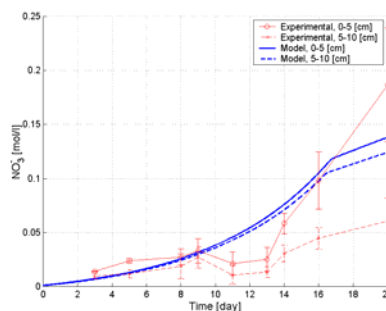


Figure 1c. Comparison between experimental and TOUGHREACT results for Nitrate (NO_3^-) in the aqueous phase.

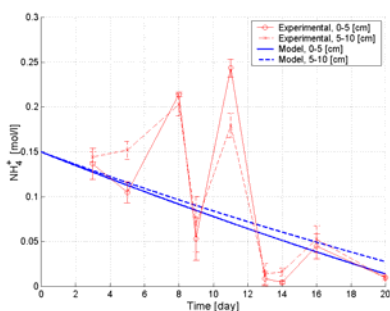


Figure 1b. Comparison between experimental and TOUGHREACT results of Ammonium (NH_4^+) in the aqueous phase.

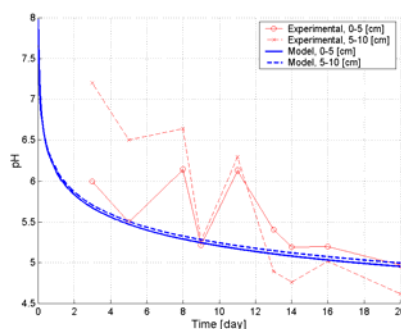


Figure 1d. Comparison between experimental and TOUGHREACT results of pH.

With the development of massively parallel computing facilities, high-resolution climate modeling, with sophisticated model physics and biogeochemical processes have become feasible. In anticipation of the coupling of the biogeochemical process model involving nitrogen to climate simulations, we are performing long-term integrations with state-of-the-art regional climate models on our multiprocessor computing facilities to quantify hydroclimate change and variability. Specifically, we have coupled the Penn State /NCAR fifth-generation Mesoscale Model (MM5) to CLM3 (Jin and Miller 2005). CLM3 physically describes the mass and heat transfer within the snowpack using five snow layers that include liquid water and solid ice. Interactions among the snow, soil, and vegetation are a function of the CLM3 mass and energy equations. A sophisticated

surface albedo scheme is chosen to improve the surface energy balance simulations. Introduction of a maximum of eight sub-cells within each MM5-CLM3 cell strengthens the description of land-surface heterogeneity. The vegetation is dynamically generated under adequate soil and atmospheric conditions favoring vegetation respiration and photosynthesis processes.

Compared with the original version of MM5, the coupled MM5-CLM3 significantly improves snow, surface temperature, and precipitation simulations. Figure 2a shows the time series of snow water equivalent (SWE) averaged over 50 SNOTEL stations in the Columbia River basin. The SWEs produced by MM5-CLM3 are in very good agreement with the observations due to the sophisticated snow physics and related model processes (Jin et al. 2002), while the SWEs from the original MM5 with simple snow processes are

greatly underestimated. The 50-station averaged temperatures from MM5-CLM3 (Figure 2b) are also consistent with observations, but the original MM5 produces cold biases during the early simulation period due to the large amount of energy consumed by the faster snow melt. The precipitation simulations from both versions of MM5 agree well with observations (Figure 2c), but our minor improvement within MM5-CLM3 reduces the over-estimated precipitation in the original version of MM5. We have recently coupled the NCAR Weather Research and Forecasting version 2.2 (WRF) with CLM and have begun 20-year WRF-CLM3, along with WRF-Noah integrations on our local computer cluster.

Concurrent to this activity, we have coupled CLM3 to TOUGH2 and refer to this new code as CLMT2 (Pan et al. 2007). This has been fully evaluated and compared with observations and simulations from a similar code, CLM-ParFlow (Maxwell and Miller 2005). Figure 3a and b indicate the significant improvement in dynamically simulating the water-table and evapotranspiration variations. This result is a very important step in examining the sensitivity of climate on groundwater, a necessary step toward understanding the role of nutrient cycling.

Plans for ongoing work include continued testing and validation of the new capability of TOUGHREACT and coupling WRF-CLM3 to TOUGHREACT. Once completed, we will use the coupled WRF-CLM3-TOUGHREACT model to investigate questions including (1) effects of climate change/variability on nitrogen cycling; (2) level of detail needed in process models for given questions and scales; (3) sensitivity of results to various properties and parameters. Applications of the coupled TOUGHREACT-CLM3 code are expected to be numerically intensive, but amenable to parallelization.

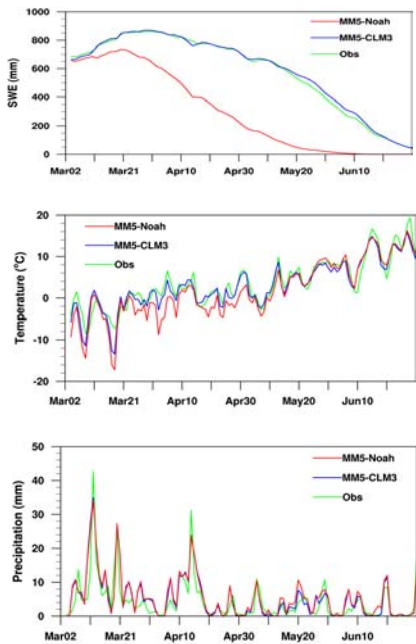


Figure 2. Comparison of observations with simulations of MM5-Noah, and MM5-CLM3 for (1) snow water equivalent (mm), (2) temperature (°C), and (3) precipitation (mm).

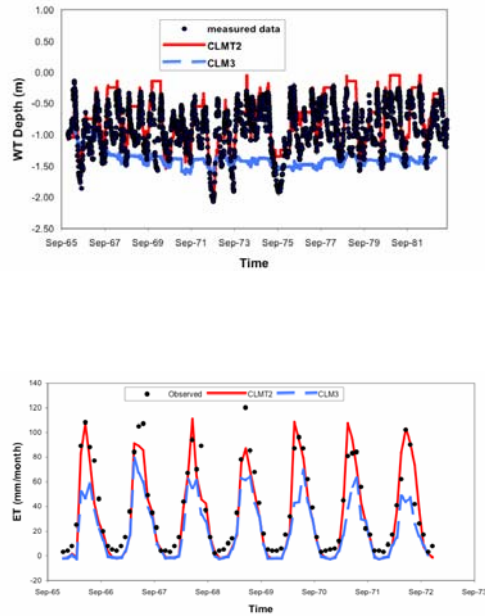


Figure 3. The CLMT2, CLM2 and observed water table depth and evapotranspiration.

References

- Bonan, G.B., K.W. Oleson, M. Vertenstein, S. Levis, X.B. Zeng, Y.J. Dai, R.E. Dickinson, and Z.L. Yang, 2002: The land surface climatology of the community land model coupled to the NCAR community climate model, *Journal of Climate*, 15, 3123-3149.
- Jin, J. and N.L. Miller, 2007: Coupling of MM5 with CLM3 to improve regional snow and vegetation simulations, *Monthly Weather Review*, In Press.
- Maxwell, R. M. and N.L. Miller, 2005: Development of a coupled surface and groundwater model. *J. Hydrometeorology*,
- Pan, L., J. Jin, N. Miller, Y.-S. Wu, and G. Bodvarsson, 2007: Modeling hydraulic responses to meteorological forcing: From canopy to aquifer, *Vadose Zone Journal*, In press.
- Pruess, K., C.M. Oldenburg, and G.J. Moridis, 1999: TOUGH2 User's Guide Version 2, LBNL-43134, E. O. Lawrence Berkeley National Laboratory, Berkeley, CA.
- Riley, W. J., and P. A. Matson, 2000: NLOSS: A mechanistic model of denitrified N₂O and N₂ evolution from soil, *Soil Science*, 165, 237-249.
- Steeffel, C.I., 2001: GIMRT v. 1.2, Software for multicomponent, multidimensional reactive transport User's Guide, UCRL-MA-143182, Lawrence Livermore National Laboratory, Livermore CA.
- Venterea R.T. and Rolston E.E., 2000a, Mechanistic modeling of nitrite accumulation and nitrogen oxide gas emission during nitrification. *Journal of Environmental Quality* 29(6), 1741-1751.
- Venterea R.T. and Rolston E.E., 2000b: Nitric and nitrous oxide emission following fertilizer application to agricultural soil: Biotic and abiotic mechanisms and kinetics. *Journal of Geophysical Research* 105(D12), 117-129.
- Xu T., Sonnenthal E.L, Spycher N.L., and Pruess K., 2004: TOUGHREACT User's Guide: A Simulation Program for Nonisothermal Multiphase Reactive Geochemical Transport in Variably Saturated Geologic Media. LBNL-55460, 187p.
- Xu, T., Incorporation of Aqueous Reaction Kinetics and Biodegradation into TOUGHREACT: Application of a Multi-region Model to Hydrobiogeochemical Transport of Denitrification and Sulfate Reduction, *Vadose Zone Journal*, In press, 2007.

Ensemble Hydrologic Prediction

Hamid Moradkhani (hamidm@cecs.pdx.edu), Soroosh Sorooshian (soroosh@uci.edu)

Advanced Hydrologic Prediction Service (AHPS) is an initiative of National Oceanic and Atmospheric Administration (NOAA) which has been designed to support the infusion of new science and technology into the provision of information and products in water prediction and its associated uncertainties within the NWSRFS. Despite significant advances in development of AHPS modules, there still exists some limitations and research needs for further improvement of this system (NRC, 2006). As such ensemble Streamflow Prediction (ESP) system is one of the key elements of AHPS, which despite its plausible approach to develop a suite of probabilistic forecasts, requires the inclusion of more advanced techniques to address all sources of uncertainties in prediction products. Besides the input and output error sources that are considered in the current ESP procedure, the uncertainties in model parameter estimates and the status of model state variables prior to forecast can also result in substantial error in river forecasting. The recent developments within the Monte Carlo framework using the sequential Bayesian method (ensemble filtering), now make it possible to properly account for various uncertainties in data, parameters, state variables and model structure.

Hydrologic data assimilation problems using ensemble filtering are, however, often faced with the high dimensionality of state vectors owing to spatial discretization needed to represent underlying hydrologic processes over high resolution computational grids. In practice, this generally demands processing of massive datasets with computationally demanding distributed models. Therefore, the implementation of sequential Bayesian method and ensemble filtering technique to extract the time-varying posterior distribution of states (e.g. soil moisture) [Margulis et al., 2002] and to a much higher computationally demanding extent, combined state-parameter [Moradkhani et al. 2005; Vrugt et al. 2006] would justify the need for high performance computational capacity.

Margulis, S. A., and D. McLaughlin, D. Entekhabi, and S. Dunne, (2002), Land data assimilation and estimation of soil moisture using measurements from the southern great plains 1997 Field experiment, *Water Resour. Res.*, 38, 1299, doi:10.1029/2001WR001114.

Moradkhani H., K.-L. Hsu, H. Gupta, S. Sorooshian (2005b), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resour. Res.*, 41, W05012, doi:10.1029/2004WR003604.

National Research Council (2006), *Toward a New Advanced Hydrologic Prediction Service (AHPS)*, National Academies Press.

Vrugt, J. A., H. V. Gupta, B. O. Nuallain and W. Bouten (2006), Real time data assimilation for operational ensemble streamflow forecasting, *Journal of Hydrometeorology*, 7, 548-565.

Land Information System (LIS)

(Christa Peters-Lidard, christa.peters@nasa.gov, Sujay Kumar, sujay@hsb.gsfc.nasa.gov)

Land Information System (LIS) is a comprehensive land surface modeling and data assimilation system developed by the scientists at the NASA Goddard Space Flight Center (GSFC), that integrates the use of community land surface models, hydrologic observations, data assimilation techniques, and required computing tools to facilitate land surface modeling at regional, continental and global scales as fine as 1km. Modeling the global land surface as two-dimensional regions or cells at 1km leads to approximately 150 million land points, that require approximately 600 Gbytes of output per day, several terabytes of memory for simulation and many months of computational time. LIS, by taking advantage of the advanced software engineering and use of high performance computing demonstrated the global application of the system at 1km resolution in near real time, on a variety of computer platforms. The large data throughput from the land surface models was addressed by a high availability parallel I/O based on the Graphical Analysis and Display System (GrADS)-Distributed Oceanographic Data System (DODS) servers with dynamic load balancing and distributed data storage and highly scalable data replication with peer-to-peer technologies. Figure 5 shows a comparison of the normalized timing and simulation throughput for LIS for a global land surface simulation at 1km resolution. It can be observed that the use of adequate computing resources is required for achieving a reasonable computational throughput.

LIS has been coupled to regional climate models such as the Weather Research and Forecasting (WRF) and cloud resolving models such as the Goddard Cumulus Ensemble (GCE) model. These coupled systems have greatly enhanced our ability to study coupled energy and water cycles, and to further understand the requirements to better predict interactions between clouds, radiation and the land surface. However, these coupled systems that include the interaction of complex components increase the computational requirements significantly. Application of the LIS-WRF system at continental scales requires approximately 100GBytes of memory and around 128 processors to achieve close to near real-time performance. Clearly, a multiprocessor environment with adequate computational resources is essential for the effective application of the coupled system.

A key capability in LIS is the support for data assimilation that combines land surface processes with satellite and ancillary observations for the determination of optimal land surface states. The data assimilation structures in LIS provides a computing infrastructure to directly ingest remotely sensed high resolution observations of land surface conditions to produce accurate, spatially and temporally consistent fields of land surface states, with reduced associated error. The data assimilation system in LIS uses an Ensemble Kalman Filter (EnKF) approach to generate assimilated or analysis products. The EnKF is a monte-carlo approach to the nonlinear filtering problem and is based on the approximation of the conditional probability densities of interest by a finite number of randomly generated model trajectories. The EnKF is particularly well suited to the nonlinear and intermittent character of land surface processes. To effectively simulate the background-error covariances, a large enough ensemble size need to be maintained in the data assimilation process, which linearly increases the computational resource

requirements. Even with a modest ensemble size of 10, the computational estimates in Figure 5 would increase tenfold in a data assimilation setting. Therefore, to effectively assimilate high resolution remote sensed observations at large spatial scales using EnKF requires a scalable parallel computing architecture.

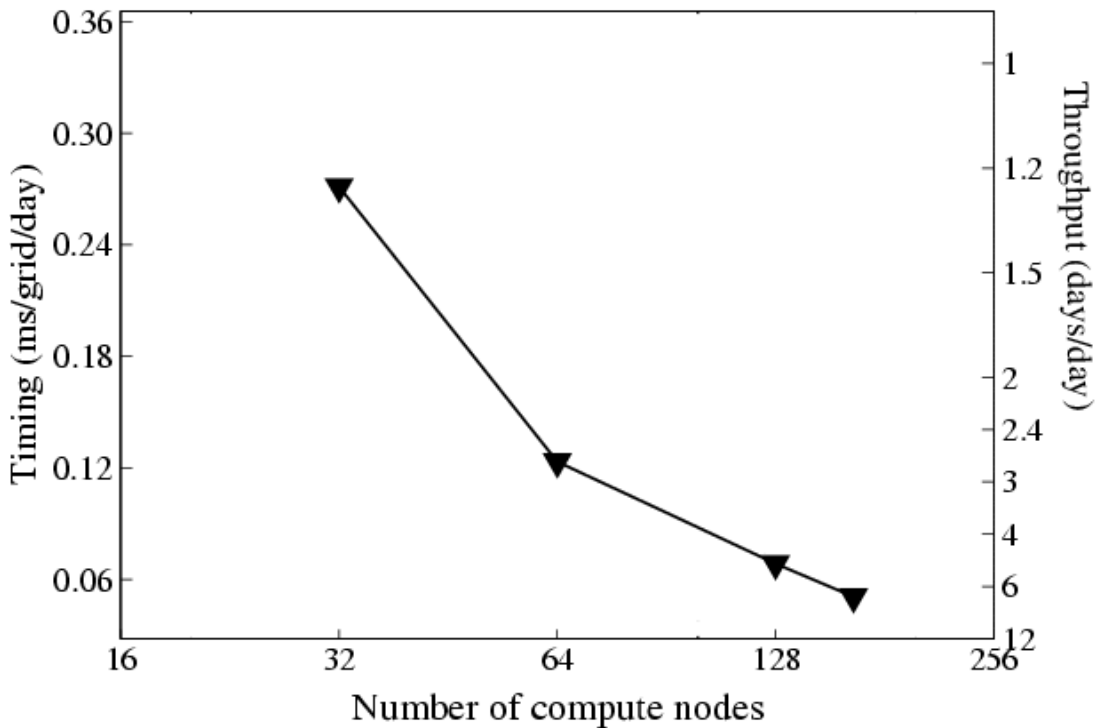


Figure 5: Comparison of land surface simulation computational performance (using Community Land Model (CLM) in LIS) for global executions at 1km spatial resolution in a multiprocessing environment. Using 192 processors, a throughput of approximately 6 simulated days per day was achieved (On a custom-built distributed memory Linux cluster with 192 nodes of AMD XP processors of 1.53GHz and fast Ethernet connections).

References

Kumar, S. V. and Peters-Lidard, C. D., and others, Land information System: An interoperable framework for high resolution land surface modeling, *Environmental Modeling and Software*, (21), 2006, pp. 1402-1415

Continental Water Dynamics

(David Gochis, gochis@rap.ucar.edu, Dennis Lettenmaier, dennisl@u.washington.edu, David Maidment, maidment@mail.utexas.edu, Efi-Foufoula Georgiou, efi@umn.edu)

As a part of continental water dynamics for the United States, a relevant project would be to simultaneously simulate the flow in all the rivers, streams and waterbodies of the nation that appear on 1:100K scale maps. This activity is facilitated by the recent publication of NHDPlus, a 1:100K scale production of the National Hydrography and National Elevation datasets. More information on NHDPlus is available at - <http://www.horizon-systems.com/nhdplus/>. Systems like this offer the possibility for hydrologic prediction at lead times from days (floods) to months and years (seasonal to interannual prediction; drought), and have been exploited already in a hydrologic forecast system for the western U.S. (Wood and Lettenmaier, 2006). Advancing these concepts through incorporation or, for instance, ensemble prediction methods and data assimilation demands high performance computing.

A small prototype of this system has been developed that links a hydrologically-enhanced version of the community Noah land surface model (Gochis and Chen, 2003) with the NHDPlus terrain and stream topology for a watershed in Ohio. HPC will enable a significant expansion of this initial work for hydrological studies over the CONUS that capitalizes on emerging hydrographic data structures (e.g. NHDPlus).

A logical extension of this work will center on the use of data assimilation to examine the potential ingest of approximately 6000 stream gage records that are available approximately every 4 hours from USGS. While computationally expansive, the use of hydrologic data assimilation to constrain the state-space of rainfall-runoff and water resources models is currently recognized as a potentially valuable enhancement to our environmental monitoring and prediction systems.

A national hydrological information system is to be deployed in conjunction with the 1:100K NHDPlus dataset, which is now complete for the nation, and onto which the USGS has snapped 18,000 streamgage locations and checked them for accuracy in the District Offices. At present 14,271 of these stations have physical measurement data from individual stream gagings loaded on the NWIS web site (<http://nwis.waterdata.usgs.gov/usa/nwis/measurements>). These data include discharge, surface width, cross-sectional area, stage height and mean velocity, for each individual gauging station. From the NHDPlus it is feasible to get stream slope, upstream drainage area, proportion of the drainage area in different land uses (though not soils or geological context), and a set of indices related to stream order. All of the available NHDPlus vector datasets have been assembled along with and the 16 or 17 of the 18 regions in the continental US have now been completed by the Environmental Protection Agency.

Please refer to Appendix II for a detailed description of computational details for various modeling scenarios.

References

Gochis, D.J. and F. Chen, 2003: Hydrological enhancements to the Community Noah land surface model: Technical Description. NCAR Science and Technical Note, NCAR/TN-454+STR, September, 2003.

Wood, A.W. and D.P. Lettenmaier, 2006, [A testbed for new seasonal hydrologic forecasting approaches in the western U.S.](#), in press, Bulletin of the American Meteorological Society.

Appendix I

CPU usage and Simulation times for a coupled surface water, groundwater land surface problem

Region	Surface Nodes	Surface Resolution	Depth	Vertical Resolution	Total Unknowns	Time Step	Run Time	Number of CPUs	CPU	Wall Clock
Central Oklahoma	128X88	1km	20m	0.5m	450,560	1 hour	1 yr	32	3.02GHz	60 h
Little Washita	128X88	350m	196m	0.5m	4,415,488	1h	1 yr	170	2.6GHz	100h

The central OK problem at 1km resolution, 128x88km laterally and 20m (40 cells at 0.5m resolution) of subsurface (totaling 450k compute nodes) takes about 60h (wall clock) to run one year at hourly timesteps (and generates about 200GB of output including the subsurface) using 32p of a locally-managed linux cluster (Intel Xeon, 3.02Ghz cpu).

A finer resolution problem encompassing a much greater depth (over 100m of subsurface) for the Little Washita Watershed totaling 4.5M cells runs on 170 processors on a LLNL institutional machines (MCR) and a year takes around 100h (wall clock). Model scaling is very good (around 80%) up to 1,000p (the largest number we have run on).

A problem that encompasses the Continental US outlined above (5,500km x 3,500km x 8, 154MN) is only 341 times larger than the central OK problem and would take roughly 20,000 hours on 32p (852d) or 2,046 (83d) hours on 320p or roughly 654h (27d) on 1000p. This does not account for difference in processor speed for newer machines, the clusters we run on are a couple years old, but overall, while this is a huge problem, it is very much attainable using High Performance Computing.

Appendix II

Computational Requirements for High Resolution Explicit Hydrological Modeling of the Coterminous U.S. (CONUS)

- 1) **General Motivation:** Physical factors influencing the infiltration and horizontal redistribution of moisture across the landscape occur on spatial scales far finer than those currently represented in coupled land-atmosphere models. Physically-based, watershed modeling techniques have been developed which explicitly represent the influence of terrain on the movement of water but such models are computationally intensive and are typically executed over very limited domains. To dramatically advance the representation of surface hydrological processes in coupled land-atmosphere models methods, a new paradigm for integrating process-based distributed hydrological models with high resolution atmospheric models is needed. Such landscape-scale approaches to surface hydrological modeling are essential for advancing the representation of pathways storages and fluxes of energy, water and biogeochemical constituents across highly heterogeneous land forms. These new approaches will demand the use of highly-paralleled computational platforms in order to make such computations feasible. This document aims to provide an initial, *order of magnitude estimate* of the computational demand of performing high resolution, distributed hydrological simulations over a truly continental domain.

- 2) **Scope of Simulation:** This simulation example has been developed to address the issue of land data assimilation as a quintessential step in the initialization of a coupled land-atmosphere prediction system. We assume the following model setup:
 - Model horizontal grid resolution: 1 km x 1 km
 - Model grid dimensions: 5,500 x 3,500 (to encompass the CONUS)
 - Number of vertical layers: 8
 - Multiple for representing sub-grid topography: 10 (i.e. sub-grid has a grid increment of 100m)
 - Model run time: 1 year
 - Model time step: 1 hour

This configuration will permit simulation of a 1 year spin-up period of an uncoupled land surface model, driven by observed meteorological forcing for the CONUS domain. Results from this simulation could be used as initial conditions for a coupled land-atmosphere prediction or could be directly analyzed.

- 3) **Estimated Performance Analysis:** The results of the estimated performance analysis given below. The analyses are based on existing model simulations using an identical configuration but over a much smaller domain. The performance from the actual small domain simulation has been scaled to the hypothetical domain defined in the section above. No adjustment for memory scaling and I/O processing has been incorporated. A 10% processor efficiency

factor has been used in the calculation of estimated floating point operations (flop).

Estimated model runtime of CONUS domain on one 2.4GHz Intel processor: 45,606 wall clock hrs/sim.yr (5.2 yrs/sim.yr)

Estimated number of total floating point operations assuming 1 floating point operation per CPU cycle and 10% efficiency of the CPU: 39.4 PFLOP.

Estimated required storage: 28.1 Tb for one simulation year, hourly output. This estimated storage includes model state and flux variables for the full domain at 1km grid spacing, all model levels, including several fields on the model subgrid with 100m grid spacing.