Navigating the Clouds on the Horizon

A Vision for Reproducible Hydrologic Modeling in the Cloud

Alejandro (Lejo) Flores
Department of Geosciences, Boise State University

The Big Picture of this Talk

Takeaways for this talk:

- Cloud is increasingly important in computational Earth and environmental science
- Many hydrologic modeling workflows and user cases fit well in a cloud computing model
- Open and reproducible computing tools can bring computational research and decision-making closer

BUT

- There’s an urgent need for effective training and onboarding

NSF Mid-Career Advancement Program

From solicitation: “envision **new insights on existing problems** or identify new but related problems **previously inaccessible without new methodology or expertise** from other fields are encouraged.”

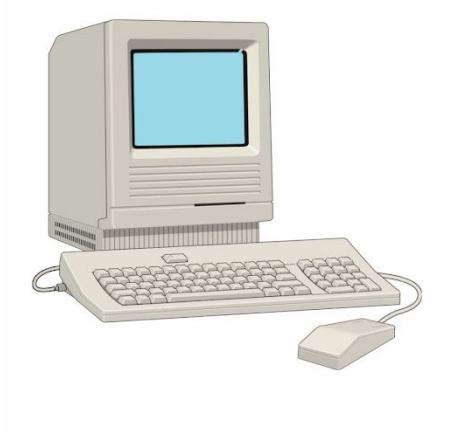
Award Abstract # 2121108
MCA: Navigating the Clouds on the Horizon: Research and Education for Cloud-enabled Computational Hydrology in the Data Revolution

NSF Org:	EAR Division Of Earth Sciences
Awardee:	BOISE STATE UNIVERSITY
Initial Amendment Date:	June 22, 2021
Latest Amendment Date:	June 22, 2021
Award Number:	2121108
Award Instrument:	Standard Grant

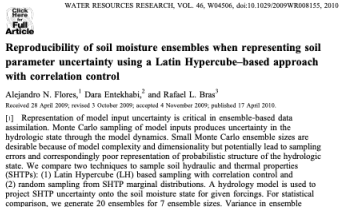
Legacy Modeling Workflow

Acquire some input data

Run a model



Write a paper



WATER RESOURCES RESEARCH, VOL. 46, W04506, doi:10.1029/2009WR008155, 2010

Click Here for Full Article

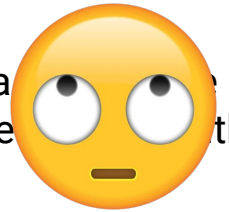
Reproducibility of soil moisture ensembles when representing soil parameter uncertainty using a Latin Hypercube–based approach with correlation control

Alejandro N. Flores,[1] Dara Entekhabi,[2] and Rafael L. Bras[3]

Received 28 April 2009; revised 3 October 2009; accepted 4 November 2009; published 17 April 2010.

[1] Representation of model input uncertainty is critical in ensemble-based data assimilation. Monte Carlo sampling of model inputs produces uncertainty in the hydrologic state through the model dynamics. Small Monte Carlo ensemble sizes are desirable because of model complexity and dimensionality but potentially lead to sampling errors and correspondingly poor representation of probabilistic structure of the hydrologic state. We compare two techniques to sample soil hydraulic and thermal properties (SHTPs): (1) Latin Hypercube (LH) based sampling with correlation control and (2) random sampling from SHTP marginal distributions. A hydrology model is used to project SHTP uncertainty onto the soil moisture state for given forcings. For statistical comparison, we generate 20 ensembles for 7 ensemble sizes. Variance in ensemble

“Data on
reque thors…”



Building Blocks of a New Model

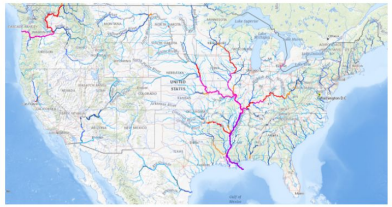
The Future is Open!

CUAHSI
Universities Allied for Water Research

HOME SUBSET API

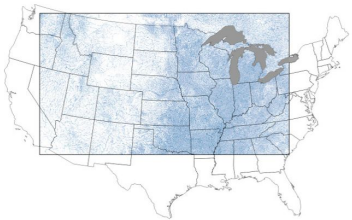
Welcome to the CUAHSI Domain Subsetter!

The purpose of this application is to introduce a collaborative effort for preparing, publishing, and sharing subsets of the National Water Model input data at watershed scales. With a combination of modern cyberinfrastructure techniques and state-of-the-science modeling tools, researchers will have access to subsets of National Water Model information that would otherwise require extensive computational resources. This work will provide the foundation onto which similar efforts can be applied to other large-scale model simulations and input data.



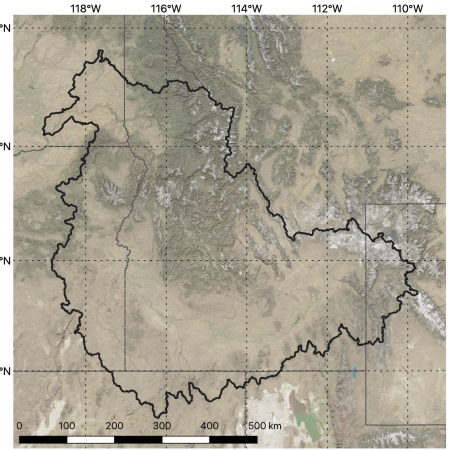
National Water Model

The National Water Model is a hydrologic modeling framework that simulates observed and forecast streamflow over the entire continental United States. It's

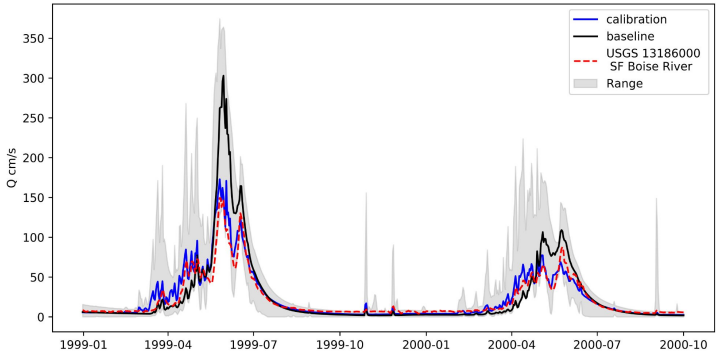


ParFlow-CONUS

ParFlow is a parallel, integrated hydrology model that simulates spatially distributed surface and subsurface



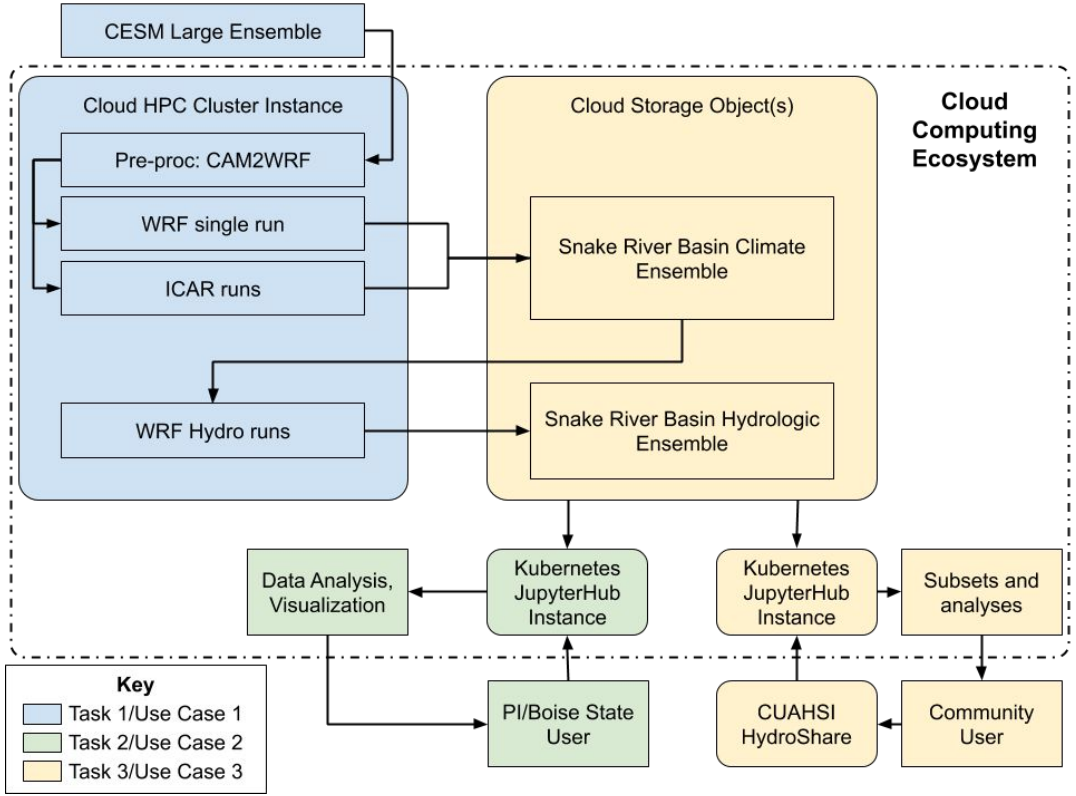
WRF-Hydro
MODELING SYSTEM



Centering Objectives on User Experiences

Computational Hydrology Use Cases	Research Objectives	Educational Objectives
User 1: Hydrologic Modeler Motivating question: How do I replicate and effectively use an HPC environment in the cloud?	Use cloud computing to produce a regional hydroclimatic variability and change dataset using existing data and models	Design, build, and use a computing environment in the cloud to run a hydrologic model
User 2: Water Data Scientist Motivating question: How can I use the cloud to analyze a large dataset so I don't have to download it all?	Develop interactive computing environments that allow for analysis and visualization of the regional dataset	Perform analyses on large spatiotemporal datasets using distributed cloud computing
User 3: Big Data Generator Motivating question: How do I ensure the community can discover, access, and use portions of the dataset of interest to them?	Subset and organize regional dataset storage to enable users to explore, analyze, download and use the data relevant to their need/use	Store and share a dataset in a cloud-native format and demonstrate its use

Learning While Doing...



Adding Value to Stakeholders

Managers and decision makers that have knowledge, information, and data gaps to support decisions & policy

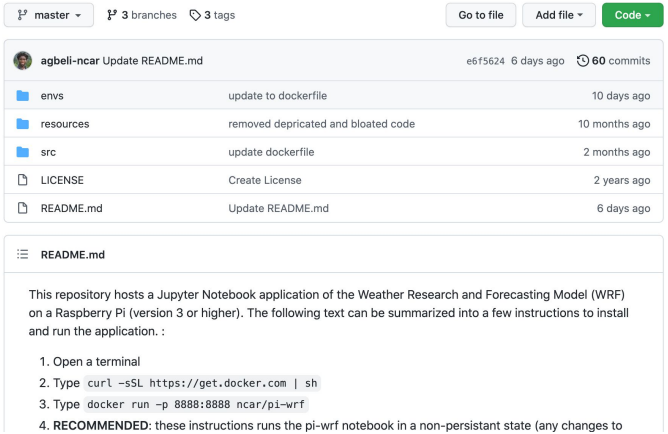
Basic and applied researchers that bring local context through additional modeling, analysis, interpretation

Originators of data (e.g., CMIP), model and algorithm developers

Beyond This Project

- How do we integrate these training opportunities into our existing curricula? What are innovative ways to share training across institutions?
- How to we extend training and workforce development to partners in management agencies to ensure that the data we produce meets their needs?

Broadening Participation



NAMPA SCHOOL DISTRICT

WRF

Thank You!

Support provided by National Science Foundation Mid-Career Advancement award EAR-2121108