U.S. DEPARTMENT OF ENERGY | Office of Science

IDEAS–Watersheds

June 2021

Accelerating watershed science through a community-driven software ecosystem

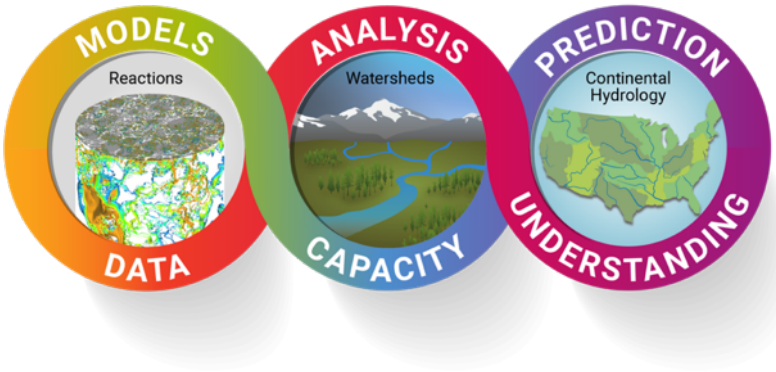
FY21 ANNUAL REPORT FOR July 1, 2020 – June 30, 2021

Principal Investigator

J. David Moulton
Los Alamos National Laboratory
Phone: 505-665-4712 | E-mail: moulton@lanl.gov

Co-Investigators & Leads

Scott Painter, Oak Ridge National Laboratory
Xingyuan Chen, Pacific Northwest National Laboratory
Sergi Molins, Lawrence Berkeley National Laboratory
Laura Condon, University of Arizona
Reed Maxwell, Princeton University
Dave Gochis, NCAR
Steve Smith, Lawrence Livermore National Laboratory
Hai Ah Nam, Los Alamos National Laboratory



DOE Office of Science Program Manager

Paul Bayer, Environmental System Science Program, Climate and Environmental Sciences Division, BER

LA-UR-21-31652

Los Alamos NATIONAL LABORATORY EST. 1943 | Lawrence Livermore National Laboratory | BERKELEY LAB | OAK RIDGE National Laboratory | Pacific Northwest NATIONAL LABORATORY

PRINCETON UNIVERSITY | ARIZONA | NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

TABLE OF CONTENTS

1 IDEAS-WATERSHEDS OVERVIEW

Watersheds play a critical role in our water supply infrastructure and require sustainable management in a changing environment. Sustainable management of watershed systems and their interaction with the built environment rely on understanding the hydrologic and biogeochemical processes that control watershed system dynamics and water availability and quality. The overarching objective of the U.S. Department of Energy's (DOE's) Environmental System Science (ESS) program is to advance a robust, predictive understanding of how watersheds function and respond to perturbations as integrated hydrobiogeochemical systems. ESS supports a network of watershed testbeds within the United States where national laboratories and university partners work in interdisciplinary teams to advance watershed system science for energy.

The ***Interoperable Design of Extreme-scale Application Software*** **(IDEAS)** project was initiated in 2014 to help ESS and other scientists in the Office of Biological and Environmental Research (BER) increase software development productivity—a key aspect of overall scientific productivity—through a new interdisciplinary and agile approach to creating sustainable, reliable, high-performance scientific software. ***IDEAS-Watersheds***, in particular, focuses on development and demonstration of critical modeling capabilities needed to further advance a process-rich computational capability to support watershed hydrobiogeochemical system science, eventually at river-basin scales – including improved representations of biogeochemical processes and their hydrological controls in the headwaters and stream/river corridors of watersheds. The IDEAS-Watersheds project supports a community collaboration model centered around an "ecosystem" of interoperable components that could be used to build or extend scientific applications. This collaboration model recognizes that existing scientific capabilities currently reside in a small number of existing integrated codes. While each of these integrated codes has a significant range of capability, none contains the full range of scientific modeling capabilities needed for a comprehensive watershed system perspective. It is thus advantageous to use codes in combination, taking representations of different processes from different codes.

2 IDEAS-WATERSHEDS OBJECTIVES

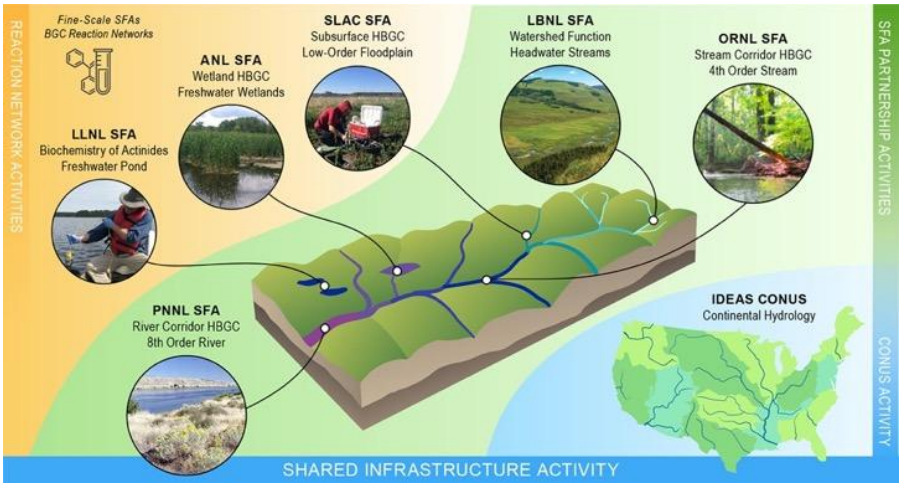


Figure 1. IDEAS-Watersheds is organized around six Research Activities. These activities address the needs identified in our partnerships along three cornerstones of ESS research -- Reaction Networks, Watershed Hydrobiogeochemistry, and Continental Hydrology-- while enabling us to develop shared solutions to move information and capability across the project and ESS as a whole.

The project is organized around six Research Activities (**Figure 1**), to address important scientific challenges and advance software development methodologies and engagement in the growing community-driven software ecosystem.

IDEAS-Watersheds Research Activities and Scientific Objectives

Partnership Activities with SBR's SFAs

OBJECTIVE: Better understand and represent how biogeochemical processes in spatially limited metabolically active zones interact with steady and unsteady hydrologic exchange flows to control reach-to-watershed-scale export of nutrients and inorganic contaminants.

- **Watershed Function SFA Partnership: East River Use Case.** Perturbations to mountainous watersheds (e.g., floods, drought, early snowmelt) impact the downstream delivery of water, nutrients, carbon, and metals. Currently, no single model can capture all relevant processes across this domain at fine resolution. IDEAS-Watersheds Partnership with the Watershed Function SFA at LBNL aims to develop a multiscale modeling framework that will allow us to consider processes at different resolutions within the watershed, including both the software tools and workflows required to enable this framework.
- **Critical Interfaces SFA Partnership: East Fork Poplar Creek Use Case.** Metabolically active transient storage zones (MATSZs), slow-flowing biogeochemical hotspots adjacent to flowing stream channels, are responsible for a significant portion of carbon, nutrient, and trace metal processing, thus affecting stream biogeochemistry and, ultimately, downstream water quality. IDEAS-Watersheds Partnership with the Critical Interfaces SFA at ORNL is developing a stream corridor modeling framework that allows laboratory-derived understanding of biogeochemical processes occurring in MATSZs to be combined with reach-scale observations. The partnership supports the over-arching strategy of a multiscale river network modeling system to represent how MATSZs influence downstream water quality by processing carbon, nutrients, and trace metals.
- **River Corridor SFA Partnership: Columbia River Use Case.** A key element of our proposed work is understanding impacts of disturbances, with emphasis on wildfires and modified precipitation regimes, both prevalent disturbances that impact river corridor hydrobiogeochemistry. The IDEAS-Watersheds Partnership with the River Corridor SFA at PNNL aims to achieve the SFA's goal in translating the fundamental process understanding of river corridor hydrobiogeochemistry under baseline and disturbed conditions into predictive, interoperable models across watersheds. The IDEAS-Watersheds software ecosystem is leveraged to couple watershed hydrologic, biogeochemical and land-surface processes for integrated watershed modeling to route the water and other chemical constituents to river corridors.

CONUS Activity

OBJECTIVE: Advance the modeling platform to better provide regional- and continental-scale hydrologic drivers that can bridge between individual sites and provide larger-scale context.

Simulating integrated flow over continental scales at so-called hyper-resolution is an identified grand challenge in computational hydrology. To adequately capture feedback between deeper subsurface flow, the land energy budget, and the lower atmosphere, explicit connections need to be made between these systems in large-scale models. The IDEAS-Watersheds CONUS Activity supports the development of an integrated hydrologic modeling platform of CONUS using ParFlow-CLM. The CONUS model, shown in the map of CONUS 2.0, bridges across our study areas and provides a scaling framework from the reach scale up to watershed and regional systems.

Reaction Network Activity

OBJECTIVE: Enhance the capabilities of geochemistry reaction modeling tools by leveraging fundamental genomic and molecular advances.

Activities will focus on two major aspects: (1) develop new approaches and workflows to incorporate new data and knowledge into the reaction models, including the derivation of reaction networks from metagenomics data and reaction parameters; and (2) implement reaction models that reflect the new understanding needed in the codes. Collaborations with other IDEAS Research Activities and partnerships will be fostered when shared interests are identified. For example, the ANL SFA shares the same interests with the PNNL SFA in generating reaction networks using KBase tools. This new connection to KBase highlights our growing collaboration with DOE-BER user facilities (EMSL), the knowledge base (KBase), and data archives and services (ESS-DIVE), to accelerate development of new reaction network models.

Shared Infrastructure Activity

OBJECTIVE: Coordinate the development of common workflow tools and software interfaces to support interoperability and reduce duplication of effort, and ensure long-term sustainability of the software ecosystem.

Activities address workflow tools for transferring information across scales, customized meshing, and model-data integration, and interfaces to couple multiple process-based codes and support code interoperability. The IDEAS project has a more intentional design philosophy to interoperability by advancing formal interfaces to key component capability to significantly reduce duplication of effort and improve scientific productivity for computational environmental scientists.

3 PROGRAM STRUCTURE

To execute this interdisciplinary project, the team utilizes a matrixed organizational structure with research teams focused around the Research Activities and Cornerstone Leads (**Appendix C**) coordinating among the activities to ensure effective integration across teams to effectively coordinate across multiple labs and universities shown in **Figure 2**.

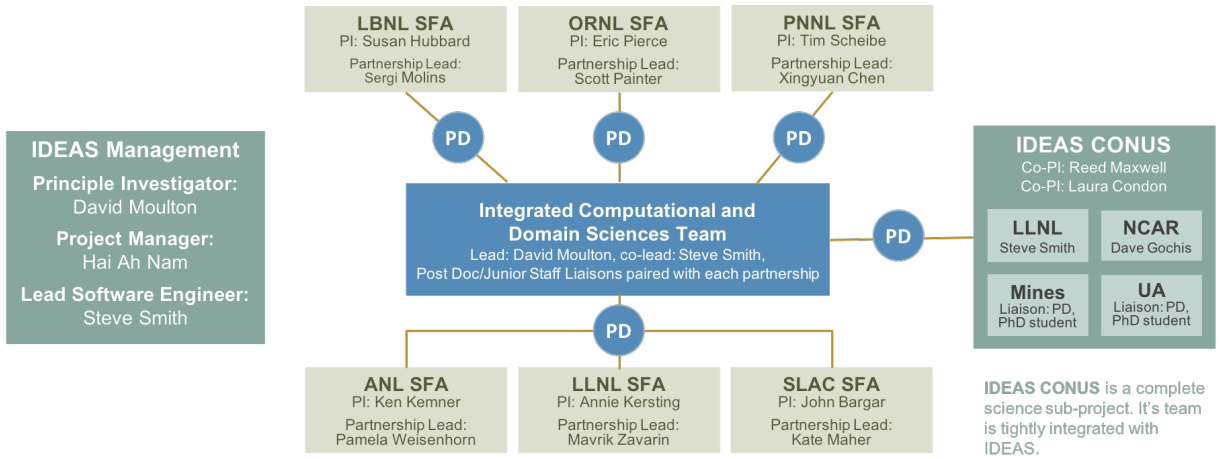


Figure 2. The Integrated Computational and Domain Sciences team. This team is composed of jointly supported postdocs and junior staff that act as liaisons with their respective partnership use cases and integrators across the project. The IDEAS CONUS Postdoctoral liaisons are represented by the PD icon. It provides a focal point for training and support of the broader community. Each Partnership has a lead to facilitate collaboration.

Postdoctoral Appointments & Early Career Staff

IDEAS-Watersheds emphasizes Training, Community Building, and Outreach to funded and project-affiliated postdoctoral researchers, early career members and students from a variety of backgrounds required to address the diverse needs of this project.

Each SFA Partnership supports one-half of a postdoctoral researcher and/or junior staff member, with IDEAS-Watersheds providing the remaining support. This co-funding model is designed to ensure integration with the SFAs and facilitate training of early career researchers.

Table 2. FY21 cohort of postdocs (PD), early career (EC) staff members and students (GRA) funded by IDEAS-Watersheds

Area of Impact	Name	ORG	Contributions to IDEAS-Watersheds
4.1.1 Watershed Function SFA Partnership	Ilhan Ozgen (PD)	LBNL	Expertise in computational hydrology. Developing multi scale simulation tools.
	Zexuan Xu (EC)	LBNL	Broad expertise in hydrological modeling ranging from hillslope-, to watershed-, to global scales. Driving the application of reactive transport for integrated hydrology
4.1.2 Critical Interfaces SFA Partnership	Saubhagya Rathore (PD)	ORNL	Expertise in stochastic hydrology. Developing a Bayesian framework in python to estimate parameters in the ADELS model
	Ahmad Jan (EC)***	ORNL	Developing multiscale modeling capabilities in ATS, here implementing the ADELS Lagrangian subgrid streambed hyporheic exchange model.
4.1.3 River Corridor SFA Partnership	Peishi Jiang (PD)	PNNL	Developing data assimilation techniques for parameter estimation for subsurface flow and transport problems. He has linked DART with PFLOTRAN using Jupyter Notebook.
	Pin Shuai (EC)	PNNL	Developing testing and expanding the Watershed Workflow for ATS modeling
	Xuehang Song (EC)	PNNL	Developing particle tracking for unstructured grids
4.1.4 Continental Modeling Platform and Simulations	Hoang Tran (PD)	CSM	Developing CONUS 2.0 models and a long-term reanalysis of the Upper Colorado Basin to connect CONUS work with the LBL-SFA.
	Mary Michael Forrester (GRA, PhD)*	CSM	Ran a multi-year simulation of CONUS 1.0 and rigorously compared it to observations and data products.
	Jun Zhang (PD)	UA	Expertise in atmospheric modeling. Developed the topographic inputs for the CONUS2.0 ParFlow model and working on model development for the ICoM project.

	Jen Stayeart (GRA, MS)	UA	Masters student studying Hydrology. Building a national database of historical reservoir operations for the CONUS simulations.
	Staff Scientist Software Eng. (0.25)**	LLNL	Will assist development teams across the software ecosystem in implementing best practices, and will lead some training sessions.
4.1.5 Reaction Network Activities	Postdoctoral Researcher (0.5)**	LANL	Will assist in developing 2D and 3D integrated hydrology/reactive transport simulations in support of the fine-scale SFA activities
	Dipankar Dwivedi (EC)***	LBNL	Supports development of reaction networks for the simulation of East River reactive transport.
4.1.6 Shared Infrastructure	Daniel Livingston (EC)	LANL	Pursuing his Masters in computer science while working on meshing activities, leads TINerator.
	Svetlana Tokareva (EC)	LANL	Improving interoperability of Amanzi and ATS kernels, and is expanding the overland flow capabilities.
	Angelic Arzola Roig (Summer UGS)	LANL	Developing IDEAS-Watersheds Software Ecosystem web site on ideas-watersheds.github.io,
* Finished PhD at CSM, Spring 2020; replacement identified ** Searching for candidates			***Began as a PD for IDEAS-Classic, now early career staff

Each CONUS Activity site lead (Princeton, UA) partially supports a postdoctoral researcher and graduate student funded through IDEAS-Watersheds. LLNL will host a software engineering focused staff scientist on CONUS funded through IDEAS-Watershed. LANL partially supports two postdoctoral researchers focused on Fine-Scale Activities and Shared Infrastructure funded through IDEAS-Watershed.

Additional students, early career staff, and postdocs not directly funded through IDEAS-Watersheds are encouraged to participate in activities. For example, postdoctoral researchers from the other SFAs attend regularly scheduled activities (Tristan Babey from SLAC and Pamela Weisenhorn from ANL). Also, postdocs from IDEAS-Classic remain integrated with the project as early career staff member.

4 PERFORMANCE MILESTONES AND METRICS

The key deliverables for Y2 of the IDEAS-Watersheds project include:

Watershed Function SFA Partnership: East River Use Case	**Develop approaches to simulate flow, transport, and biogeochemical processes at the East River Watershed and sub-catchments using a multiscale approach, including:** • Developing criteria for mesh refinement and their impact on aggregate watershed behavior, • Development and application of integrated reactive transport capabilities to predict chemical exports - specifically, nitrogen exports- in mountainous watersheds linked to hydrologic dynamics adjusting model resolution where needed.
Critical Interfaces SFA Partnership: East Fork Poplar Creek Use Case	**Develop process-rich modeling framework that enables laboratory-scale experiments to be linked with reach-scale field observations to perform:** • Reach-scale reactive transport demonstration simulations and • River-basin–scale demonstration simulations.
River Corridor SFA Partnership: Columbia River Use Case	**Develop and demonstrate workflows to enable studies of hydrobiogeochemistry in the river corridor:** • Developing uncertainty quantification and ensemble Kalman filter-based data assimilation workflows using Jupyter Notebooks. • Develop KBase Apps that enable incorporation of Omics information into reactive transport modeling. • Develop python application for particle tracking as a post-processing step on the velocity fields generated by flow and transport models using unstructured grids (e.g., PFLOTRAN, ATS).
CONUS Activity	**Develop a high-resolution integrated hydrologic modeling platform that aligns with the NWM and is interoperable with other components of the software ecosystem to be a resource for our community, including:** • National steady state groundwater initialization, supported by development of national hydrologic modeling input data sets, and workflows for creating watershed test domains. • Expansion of the ParFlow python interface and related workflows, particle tracking and GPU support.
Reaction Network Activity	**Develop reaction networks to describe biogeochemical transformations.** • Support development and maintenance of reaction capabilities in the software ecosystem.
Shared Infrastructure Activity	**Coordinate the development of that shared infrastructure to reduce duplication of effort and ensure long-term sustainability of the capability.** • Build code-agnostic workflow tools to facilitate watershed hydrobiogeochemical simulations and analyses across multiple resolutions and scales. • Develop domain specific interfaces to enable sharing established capabilities as components. • Create a sustainable software ecosystem that enhances the productivity of the community.

Progress toward these deliverables is described in the following sections.

4.1 Review of Scientific Progress

4.1.1 Watershed Function SFA Partnership: East River Use Case

Contributors (LBNL): Ilhan Ozgen-Xian, Zexuan Xu, Dipankar Dwivedi, Sergi Molins

Progress has continued in the two main tasks in this Partnership, the development of a multiresolution approach that refines in the areas of interest for the simulation of integrated hydrology and reactive transport in watersheds, and the development, testing and application of reactive transport capabilities in the context of integrated hydrology to quantitatively test models of nutrient fluxes in watersheds, including nitrogen export.

The wavelet-based mesh refinement approach developed in FY20 (Ozgen-Xian et al., 2020) has been applied to generate the domain in the simulations of surface-surface hydrology of the Lower Triangle Region. These simulations, focusing on integrated hydrology initially, are an ongoing task in the SFA. For these simulations, we have evaluated different criteria for refinement and currently are using a priori-computed flow metrics (Tarboton, 1997, Water Resour. Res.) which result in refinement in the areas in and around the streams (Figure 3). A python implementation of the wavelet-based mesh refinement approach was incorporated in TINerator in Nov 2020 (Ozgen-Xian, PR 221, 2020), becoming part of the shared infrastructure. In the Lower Triangle region simulations, we are also using PriorityFlow, developed in the Continental Modeling Platform and Simulations (CONUS) component of the project, as the tool for topographic pre-processing of the digital elevation model (DEM) data. The data-to-simulation workflow is as follows: topographic pre-processing with PriorityFlow, mesh generation with TINerator using a wavelet-based approach using a single mesh refinement criterion, mapping of (heterogeneous) geological properties and model parameters to the mesh and simulation.

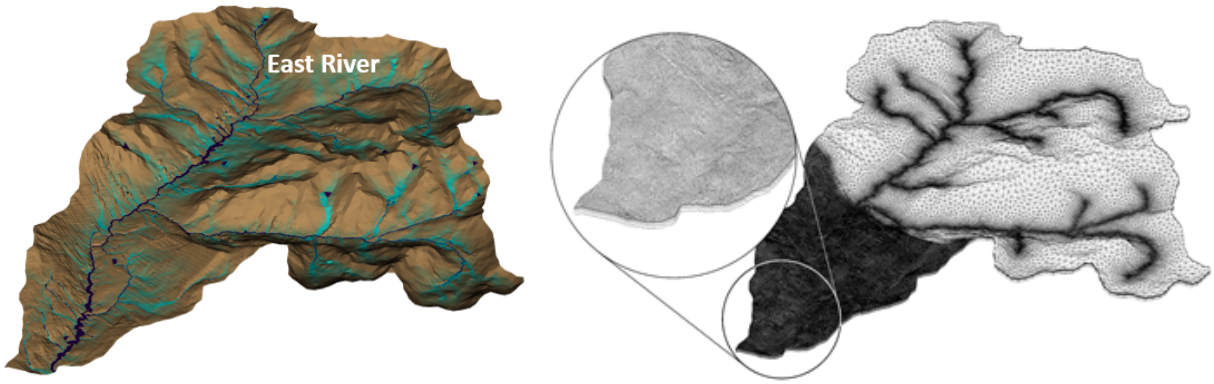


Figure 3. Simulated surface water depth and multiresolution mesh used in the simulation of integrated hydrology processes at the East River [Ozgen-Xian et al., ongoing].

We have completed developing the new reactive transport capabilities in ATS. We have used the simulations of integrated hydrology and reactive transport in the Copper Creek catchment to drive the development and to perform thorough testing of the capabilities. A particular focus in this FY21 has been in capturing accurately solute mass balance across the surface-subsurface interface in the presence of rain water and snow melt water sources. Because water may be absent from the surface domain (dry conditions), the mass of solutes in rain water or snow melt may contribute directly to the subsurface domain. A novel scheme for the conservation of the mass of solutes has been implemented that is applicable to any of the possible evolutions of the surface water (dry, dry-to-wet, wet, wet-to-dry). Two manuscripts are in late stages of preparation and will be submitted in July 2021. One describes the Copper Creek application, in partnership with Watershed Function SFA work, and focuses on unraveling concentration-discharge relationships in mountainous watersheds (Xu et al., 2021, in prep., Figure 4) The other describes the

development of the reactive transport model in the context of integrated hydrology and focuses on the implementation details including the novel mass conservation scheme for the surface solute sources and the interoperable computational approach to code development (Molins et al., 2021, in prep.- see highlights, in section 4.2).

FY22 Plans

With the current progress in the two main tasks, in FY22 we will bring together both new capabilities, mesh refinement and integrated reactive transport, to bear on the Priority Areas that guide the Watershed Function SFA in the FY20-22 period, namely to predict how snow dynamics impact aggregated watershed water and nitrogen exports. The planned activities in FY22 are as follows:

- Completion of Lower Triangle integrated hydrology simulations
- Development of the reactive transport model for the Lower Triangle and simulation building on the integrated hydrology results with a focus on processes affecting nitrogen export
- Expand Lower Triangle Region simulations to incorporate upstream catchments

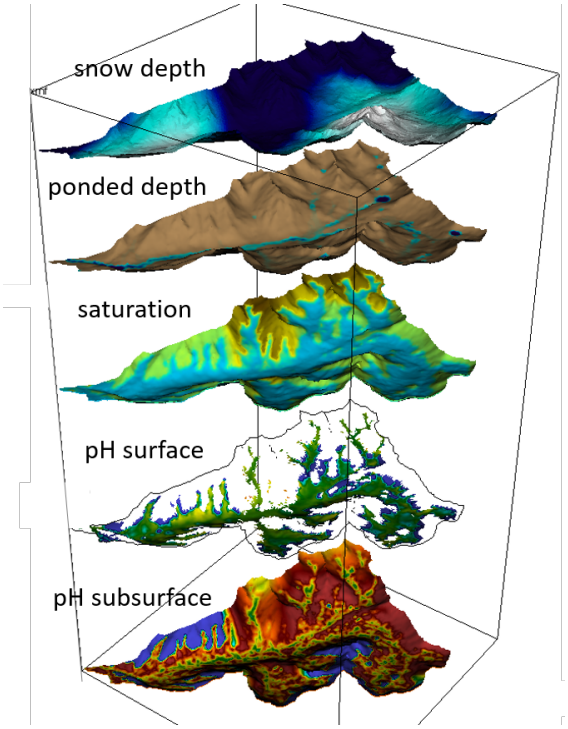


Figure 4. Simulated reactive transport in the Copper Creek catchment including key variables: snow depth, water depth, subsurface saturation, pH in the surface water and pH in the subsurface water.

4.1.2 Critical Interfaces SFA Partnership: East Fork Poplar Creek Use Case

Contributors (ORNL): Ahmad Jan, Saubhagya Rathore, Ethan T. Coon, Scott L. Painter

Significant progress was made toward our broad goal of a new modeling capability for reactive transport in river corridors that is tractable at river basin scales and informed by laboratory and field observations. Central to our strategy is our multiscale modeling capability ADELS (Advection Dispersion Equations with Lagrangian Subgrid) model [Painter, 2018; 2021] that extends highly successful residence-time models to accommodate nonlinear multicomponent reactions. The approach moves biogeochemical process representation to subgrid models that each represent an ensemble of hyporheic-zone flowpaths, thus allowing biogeochemical processes to be represented in great detail at their native spatial scales without averaging over important fine-scale variability in redox states that occurs within sediments and periphyton biofilms.

Notably, the ADELS model implementation in Amanzi-ATS was extended from conservative tracers to multicomponent reactive transport and extensively tested [Jan et al. 2021]. Reactive transport simulations in East Fork Poplar Creek (EFPC) and the Portage River Basin were then used to demonstrate the new capability (see Section 4.2, highlights).

Additional analysis [Painter, 2021] of the ADELS model structure and comparison to an alternative multiscale model based on multiple transient storage zones (multirate transient storage model, mTSM) showed that the two model structures are mathematically equivalent for non-reacting tracers. Despite that equivalence, the two models predict very different results for reactive transport. In particular, each transient storage zone in the mTSM model is directly coupled to the flowing channel and assumed to be well mixed,

which suppresses the formation of geochemically important anaerobic zones, thus limiting its potential as a general-purpose model for reactive transport in stream corridors. The ADELS model, by contrast, fully supports the formation of steep geochemical gradients and redox zonation in the hyporheic zone, which are commonly observed. The comparison also emphasizes that nonreacting tracers alone are unable to constrain models for hyporheic exchange and need to be augmented by additional observations like reactive tracer tests.

Methods for estimating model parameters from results of stream tracer tests were developed, implemented in python workflows, and evaluated [Rathore et al. 2021a] using a public dataset from tracer tests in the Hammer Stream in West Sussex, UK. For that work, we used Bayesian inference and the Markov Chain Monte Carlo (MCMC) method, which provides estimates of uncertainty in the various model parameters. Representative results are shown in Figure 5. The hyporheic travel time distributions were estimated without assumptions about the distribution shape and for the first time, simultaneously with channel properties. Moreover, we demonstrated that simultaneous analysis of observations from multiple locations constrains parameters better than analyzing those observations individually. The methods and workflow tools developed represent a powerful new capability for analyzing future tracer tests to estimate model parameters for EFPC and other streams.

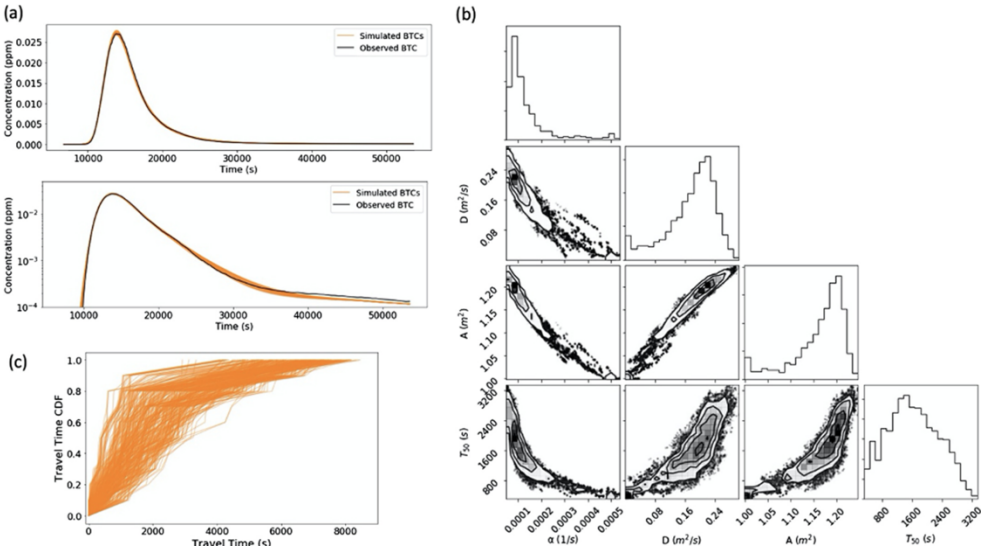


Figure 5. (a) predictive uncertainty plot from an ensemble of 320 estimated parameter sets; (b) estimated joint distributions of hyporheic exchange coefficient, dispersion coefficient, channel area, median travel time; and (c) ensemble of 320 estimated travel time CDFs. CDF, Cumulative Distribution Function [Rathore et al. 2021a].

Based on our experience with Bayesian parameter estimation applied to the stream tracer tests, a novel Bayesian joint-fitting strategy was developed (Figure 6) [Rathore et al. 2021b] and used to reinterpret results from the sorption and methylation experiments on EFPC sediments [Schwartz et al. 2021]. Specifically, we applied Bayesian inference and MCMC simultaneously to multiple sorption and methylation experiments instead of the traditional approach of analyzing subprocess models individually. Joint fitting more rigorously propagates uncertainties between different subprocess models and also allows information that is shared between datasets to be more fully used. Bayesian inference and MCMC also produces the full joint distribution of parameters, thus providing global uncertainty estimates and facilitating the detection of overparameterization manifesting as null spaces in the parameter space. The identification of null spaces guided the simplification of certain subprocess models: fast kinetic sorption was replaced by equilibrium sorption and Monod demethylation was replaced by first-order demethylation based on the analysis. The proposed scheme will benefit parameter estimation for other complex biogeochemical systems characterized through multiple experiments.

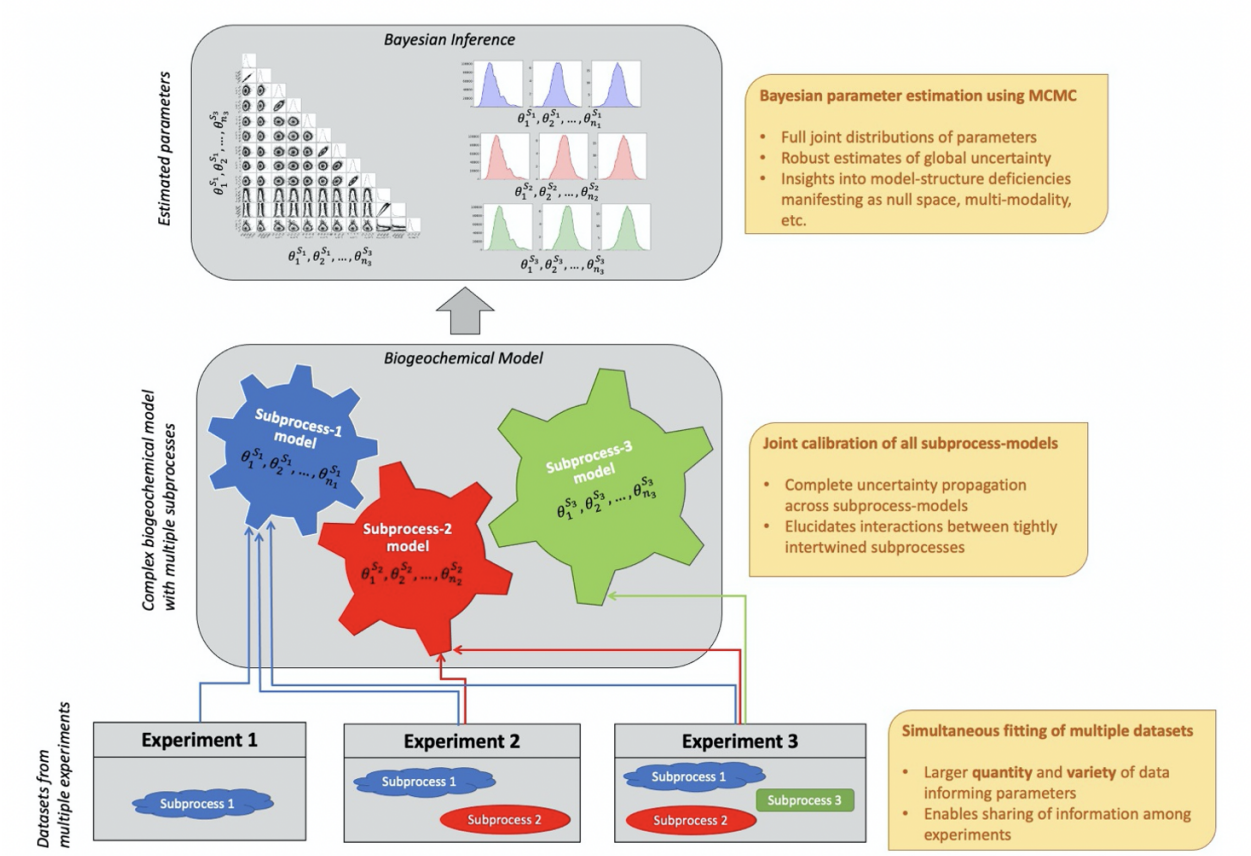


Figure 6. Schematics of Bayesian joint-fitting workflow for calibrating complex biogeochemical models. The approach offers multiple advantages over sequential fitting of experiments on sorption and reaction kinetics including efficient utilization of datasets, full uncertainty propagation, robust estimation, and detection of parameterization deficiencies [Rathore et al. 2021b].

FY22 Plans

- The ADELS framework will be extended to accommodate unsteady discharge. That capability will then be used to analyze tracer tests affected by unsteady flow.
- We will work with the developers of the NEXSS model to couple it with our workflow for building stream network models in ATS.
- Premodeling of planned tracer tests in EFPC will be undertaken using the ADELS model in ATS.

4.1.3 River Corridor SFA Partnership: Columbia River Use Case

Contributors (PNNL): Xingyuan Chen, Hyun-Seob Song (University of Nebraska-Lincoln), Kewei Chen, Peishi Jiang, Xuehang Song, Joon Yong Lee, and Pin Shuai

Particle-tracking simulation is a popular technique in groundwater modeling to identify the source of groundwater containments, estimate travel time and age of groundwater. We developed a generic particle tracking code using output from PFLOTRAN as the use case for initial development. A novel velocity interpolation approach was used to reconstruct the velocity field at any point in the domain by using the flux output of groundwater simulators. The velocity reconstruction is based on the mesh used in the groundwater flow simulation, thus avoiding over smoothing of velocity fields across the water-sediment interface. We then implemented random walk particle tracking based on a standard form of the dispersion tensor to account for dispersion and diffusion processes in porous media. The particle tracking code is

written in Python using the JSON format for model configuration and can be launched with command line arguments. Alternatively, users can also configure and execute the particle tracking code through a Jupyter Notebook. A set of Jupyter Notebook templates were developed for users to build their own particle tracking models. The particle tracking code also includes a post-processing module that can generate XDMF and HDF5 datasets for visualization using ParaView or VisIt. The implementation of the particle tracking code was tested in synthetic cases designed to simulate solute transport under known boundary constraints. The random walk particle tracking code for unstructured finite volume groundwater simulators provided a user-friendly way to conduct groundwater pathline simulation in the increasing trend of adopting unstructured meshes in subsurface flow modeling. The source code of the particle tracking codes is available at https://gitlab.pnnl.gov/sbrsfa/ideas/particle_tracking_usg.

To develop computational tools and modeling pipelines that facilitate omics data integration with reactive transport models, our team focused on developing a KBase Software Development Kit (SDK), including Apps as well as standalone software tools. First, we demonstrated the effectiveness of our metagenome-to-reactive transport modeling workflow (termed Metagenome/MAG Integration into Ecosystem Models or MAGIEC) through the 2020 Multiscale Microbial Dynamics Modeling Summer School co-organized by EMSL and PNNL SBR SFA, and held in July 2020 at PNNL. All the KBase narratives that were used in the EMSL Summer School are publicly available at https://narrative.kbase.us/#/orgs/emsl2020ss. Second, we developed KBase Narratives and SDK Apps (required for demonstrating the MAGIEC workflow) that automate (1) construction of metagenome metabolic networks, (2) integration of high-resolution metabolomics data into metabolic networks, and (3) development of dynamic substrate-explicit models based on thermodynamic theory (See Figure 7).

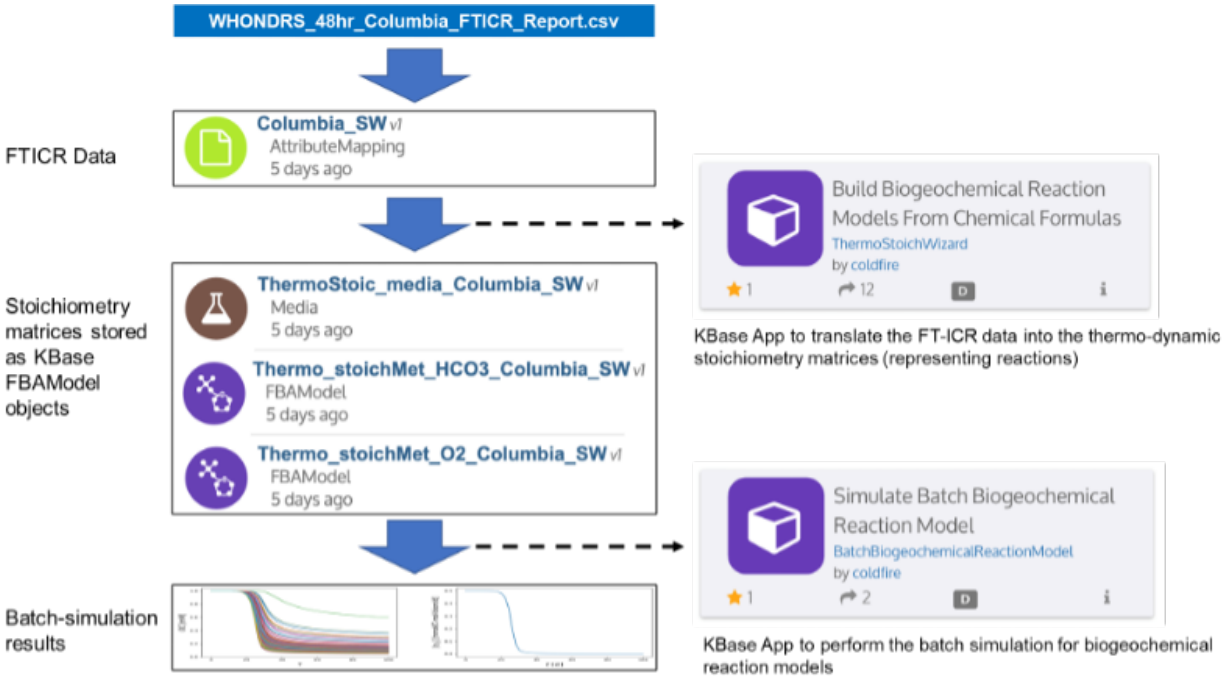


Figure 7: Illustration of the workflow to build dynamic substrate-explicit models based on thermodynamic theory using KBase SDK Apps. This workflow consists of two KBase Apps: 1) the 'Build Stoichiometric Reaction Models from Chemical Formulas' App to generate thermodynamic properties and stoichiometric reactions from molecular formulas of compounds and 2) the 'Simulate Batch Biogeochemical Reaction Model' App to simulate biogeochemical models in a batch reactor. A WHONDRS FTICR-MS data file (generated from Formularity and R-code; available at Environmental Systems Science Data Infrastructure for a Virtual Ecosystem) is imported as an input data for this workflow. Each App provides hyperlinks to download output files and a KBase Report to display the outputs with visualization.

The demonstration was also extended to (4) coupling metabolic networks with reactive transport models using the neural network-based reduced-order modeling approach. We shared software tools (KBase Apps available at https://narrative.kbase.us/#catalog/modules/ThermoStoichWizard and https://narrative.kbase.us/#catalog/modules/BatchBiogeochemicalReactionModel, and R-codes available at https://github.com/hyunseobsong/lambda) required for implementing the work on (3) through recent publication (Song et al., 2020); additional manuscripts on (1), (2), and (4) are in preparation. Lastly, we developed a conceptually new machine learning method that enables classification of high-resolution metabolomics data into subgroups with distinct biogeochemical characteristics, which is an important step to develop condition- or site-specific biogeochemical and reactive transport models. Our algorithm holds novelty in feature engineering (i.e., defining inputs to machine learning models) based on the eigenmode concept so that the classification efficiency is maximized. We successfully tested the performance of the developed method using Fourier-transform ion cyclotron resonance – mass spectrometry (FTICR-MS) data from river corridor samples collected through WHONDRS 48 Hour Diel Cycling Study. We initiated the development of an initial version of software tools for application to general data classification problems, the results of which will be shared in public in the near future.

FY22 Plans

As a result of the triennial review the River Corridor SFA went through in May 2020, the modeling effort of the SFA continues its focus on quantifying the aggregated effects of river corridor hydrologic exchange flows (HEFs), dissolved organic carbon (DOM) chemistry, and microbial activity on biogeochemical cycling, water quality, and contaminant mobility, with a change in geographic domain from the Hanford reach to the Yakima River Basin. A new research element on wildfire impacts on river corridor biogeochemical processes has also been added. The watershed- and basin-scale river corridor models (RCMs) will incorporate hydrologic connectivity and molecular information on microbiome structure (i.e., species composition and distribution of enzyme-encoding genes), microbial expression, and metabolomes to capture distinct water quality signatures in connection to variations in land use, hydrogeology, climate, and disturbances. The IDEAS-Watersheds partnership will facilitate the development of hydrologic and biogeochemical watershed models through particle tracking, tracer simulations, and interfaces to link integrated hydrologic models with available biogeochemistry engines. To support these new science directions, our FY22 effort will focus on the following:

- Watershed model intercomparison (ATS, SWAT, NWM, PRMS, GSFlow, RHESSys) using American river Watershed within the Yakima River Basin as a testbed.
- Developing river corridor model workflow in Python leveraging the Watershed Workflow tools and the Python version of NEXSS.
- Developing a river corridor model intercomparison between the Eulerian and Lagrangian approaches.

4.1.4 Continental Modeling Platform and Simulations (CONUS)

Contributors (UA, CSM, Princeton, LLNL): Laura Condon, Reed Maxwell, Steve Smith

This year significant progress was made in the development of the CONUS2.0 model domain and the national modeling framework. Last year we focused on the surface development constructing a grid and land surface datasets that align with the national water model domain. This year we continued these efforts working to finalize all of the model input datasets. Postdoc Dr. Jun Zhang under the supervision of Co-PI Condon completed a national topographic dataset for hydrologic simulation (see Figure 8). This dataset is documented in a manuscript in press in Earth System Science Data and the dataset and processing tools are published with a static DOI (Zhang et al., In Press). Additionally, under the supervision of Co-PI Maxwell, Dr. Hoang Tran, a post doc, and Jackson Swilley, a graduate student, are leading the efforts to develop and evaluate national subsurface parameterizations. This work builds off of a new national aquifer

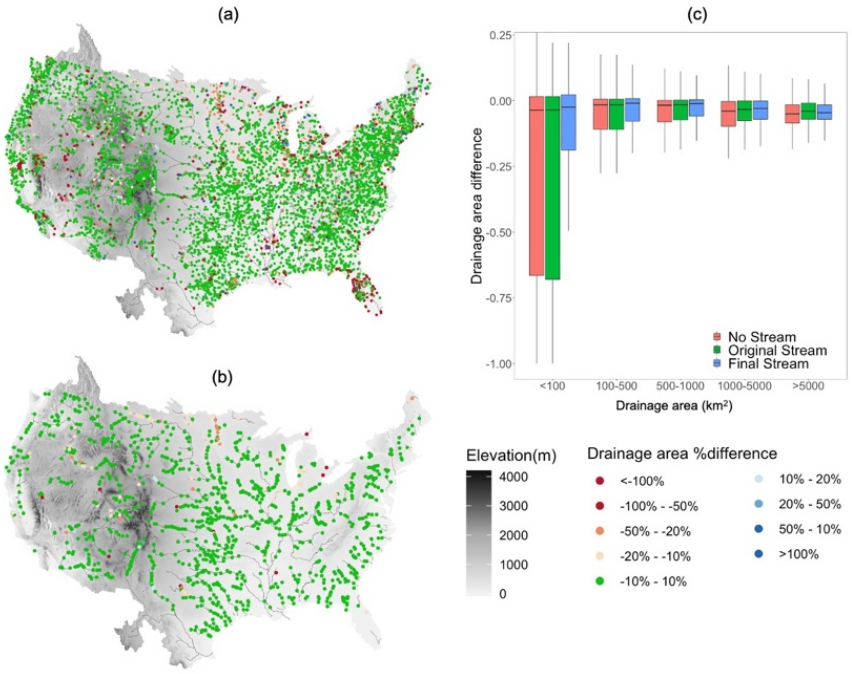


Figure 8: Drainage area comparison between processed with final order stream network and USGS gage measurements. (a) all gages for comparison; (b) gages with drainage area over 5000 km². (c) the boxplot of drainage area difference with different stream networks grouped by drainage area size. (Figure reproduced from Zhang et al., In Press)

parameterization that was developed by former IDEAS post doc Inge DeGraaf (DeGraaf et al. 2020). Our team assembled a large ensemble of potential subsurface configurations using different combinations of existing datasets and analytical solutions. The final CONUS2.0 subsurface extends significantly deeper than the CONUS1.0 domain (500m versus 100m) and has 15 layers (CONUS1.0 had 5). The added depth and layers in the CONUS2.0 subsurface will allow us to better resolve intermediate groundwater depths and pumping interactions and will improve simulation of confined systems. Finally, graduate student Jen Steyaert under the supervision of Co-PI Condon assembled a national database of historical reservoir operations for more than 600 larger reservoirs across the US. This work is currently in review in Scientific Data. The ResOpsUS dataset is the first of its kind and provides unprecedented ability to evaluate actual historical operations. In collaboration with Nathalie Voisin and Sean Turner at PNNL we have used this dataset to develop a national set of historical operating policies that can be used in simulations.

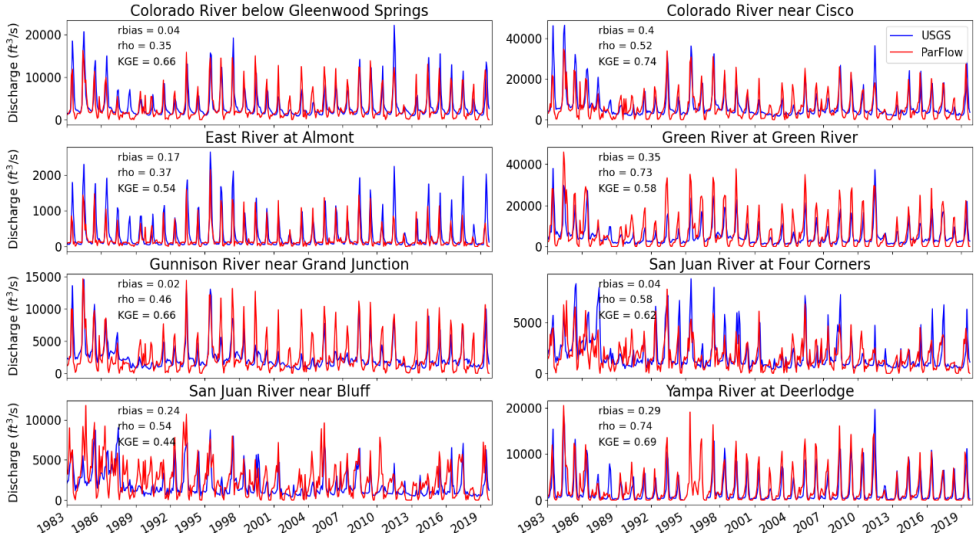


Figure 9: Plots of simulated and observed streamflows for some example gauges in the upper Colorado

In addition to input development, we have also completed extensive model simulations this year. For the first part of the year, we focused our efforts on watershed domains. Postdoc Hoang Tran completed a 20-year retrospective simulation of the Upper Colorado River basin based of the original CONUS model (Figure 9). This dataset is in review in Scientific Data. We also developed testbeds in the Upper Colorado River Basin, the Delaware/Susquehanna and the Washita. Our teams at the University of Arizona and Princeton used these watersheds (which encompass a range of topographic, hydrogeologic and climatic settings) to test out potential subsurface configurations. We ran hundreds of tests to identify the ideal national parameter set. Using what we learned from the watershed testbeds the team finalized our first national parameterization set for the CONUS2.0 model. Currently postdoc Tran is leading the spin up of the CONUS2.0 model using this dataset. Additionally, the tests for the Delaware were used to assemble a ParFlow model for the ICoM project (post-doc Zhang is also funded on this project).

CONUS2.0 development is happening in parallel with model developments to improve workflows and simulation capabilities. This year we publicly released two versions of ParFlow. We moved the continuous integration system from TravisCI to GitHub Actions and created a ParFlow Google Group to provide a more visible and easier to access user forum. We closed 30 issue reports for questions, bug and feature improvements. 30 pull requests were reviewed and merged. The ParFlow community generated 14 of the pull requests and 16 originated from the core ParFlow team supported by IDEAS. Examples of community contributions include improvements to the existing OpenMP and Cuda support, adding Kokkos support, and NetCDF enhancements. Examples of improvements funded by IDEAS-Watersheds include updating and refactoring the HYPRE interface to improve maintainability and bug/improvements in the boundary conditions discovered when setting up the CONUS model. Through a DOE IDEAS–Watersheds sponsored SBIR project we have been collaborating with the software company KitWare to continue develop improved metadata standards and Python tools for interacting with ParFlow models. Postdoc Chen Yang under the supervision of Co-PI Maxwell developed a GPU compatible version of the EcoSlim particle tracking code and that runs on multiple GPUs using both OpenMP and MPI. Speedups of greater than 100X have been realized on single GPU implementations. Results of this work are published in Computers and Geoscience (Yang et al., 2021). Dr. Yang has also explored load balancing and domain decomposition schemes for the multiple GPU configuration (primarily with MPI). Results of this work are currently in review.

FY22 Plans

- Our team will continue working on the CONUS2.0 model development focusing on (1) groundwater initialization, (2) model validation with observations, and (3) development of anthropogenic inputs
- We plan to continue our subdomain work in the Upper Colorado and the ICoM domain using them as testbeds for more rigorous model evaluation and testing. Additionally, these test domains will be used to prototype our sub-setting workflows.
- We will continue ParFlow development focusing on model documentation, expansion of the python interface and related workflows, and GPU implementations.

4.1.5 Fine-Scale Activities: Reaction Networks

Contributors: Sergi Molins, Dipankar Dwivedi

In the IDEAS Watersheds annual meeting held October 7-8, 2020, a group discussion prioritized the need for developing and benchmarking capabilities for gas transport in the unsaturated zone. As a result, we have re-scoped the activities in this cornerstone from the planned work on mineral reactions in CrunchFlow to work on gas reactions and transport. The main task currently underway in the fine-scale activities is to develop a benchmark problem for gas diffusion linked to floodplain environments. This involves the development of a problem that simulates variable saturated flow (e.g., using Richards equation) and reactive transport that includes gas equilibrium reactions in addition to aqueous and mineral reactions. Two codes

in the IDEAS Watersheds software ecosystems are being used to develop this benchmark: CrunchFlow and PFLOTRAN. In parallel, the Alquimia interface is being upgraded to explicitly consider gas species. Currently, the gas reaction capabilities of CrunchFlow have been implemented in the Alquimia interface but not yet for PFLOTRAN (https://github.com/smolins/alquimia-dev/tree/new-gases). When this upgrade is incorporated in Amanzi-ATS, we will be able to validate the new capabilities against the newly developed benchmark.

The benchmarking of gas transport and reactions is an example of the central role of the geochemical interface Alquimia in the project's approach to validate new capabilities and make them available as part of the software ecosystem. In this FY, Alquimia has enabled activities throughout the project. Specifically:

- We have completed the implementation of reactive transport capabilities in ATS via the use of Alquimia. This has enabled both the fully-resolved integrated reactive transport by the LBNL Partnership (Section 4.1.1) and the multiscale reactive transport in river corridors by the ORNL Partnership (Section 4.1.2).
- We have implemented the Alquimia interface in Parflow, merge to the master branch is pending.

A manuscript on Alquimia has been in preparation for some time, but now with Parflow supporting geochemistry in the subsurface through the Alquima interface, the manuscript is being expanded to further highlight its potential.

FY22 Plans

- Finish and submit the Alquimia manuscript. Obtain a DOI from DOE Code (osti.gov).
- Merge Alquimia branch of Parflow into master branch.
- Continue to upgrade Alquimia to incorporate gas reactions from PFLOTRAN.
- Complete gas benchmark problem and use it to test new Amanzi-ATS gas transport capabilities.
- Based on discussions to be held in the planned annual meeting (Fall 2021), prioritize cornerstone activities, which may include resuming work on implementation of a more flexible to capturing mineral reactions in geochemical codes and treatment of colloids in reactive transport modeling.

4.1.6 Shared Infrastructure to Advance the Software Ecosystem

The Shared Infrastructure activities integrate the team and development across the other research activities, and hence, draw contributions from the entire team while providing opportunities for each team to leverage these capabilities.

Multiscale Workflows: Implicit in all the use case activities is an underlying workflow that executes the necessary sequences of steps to produce results and analyses. We are striving to coordinate and guide these activities to help support the development of shared workflows that are more code and model agnostic. We continued to improve the TINerator mesh generation toolset, integrating the CONUS led PriorityFlow pre-processing capability and the wavelet-based mesh refinement (see Section 4.1.1), and improving interaction with GIS tools. TINerator is being used at East River, the River Corridor, and with external collaborators to create unstructured meshes for ATS simulations. Watershed Workflow, developed originally through an LDRD project at ORNL led by Ethan Coon is primarily supported by ExaSheds now, and is increasingly used in the ORNL and PNNL SFA partnerships. This workflow tool excels at gathering data from a range of other agencies to accelerate the development of integrated hydrology models. In addition, a new feature for meshing stream networks has been added to support the multiscale reactive transport modeling (Section 4.1.2). We are working to explore interoperability of components across Watershed Workflow and TINerator, as we strive to create more uniform workflows for users of different models. Finally, the successful development of KBase Apps to support a workflow connecting -omics to reactive transport models in PFLTORAN was described in Section 4.1.3.

To address the needs of model data integration and model archiving we participated in the recent ESS-DIVE workshop, particularly looking at the API for automated uploads, and to review the guidance for the model data (input/output) submission. During this workshop we identified the need for much more detailed guidance on the accompanying README file. We plan to work with the ExaSheds project and our SFA partners to help develop this guidance as well as some use case examples.

To record these workflows in a reproducible form, both for journal submission and for transferability to new problems across the community, we continue to adopt Python and R scripting within Jupyter notebooks. We held two modeling seminars where team members shared their experiences and best practices recommendations for Jupyter notebooks. These sessions were very well attended and we are planning to develop best practices guidance with examples for the team, which we'll share through our website once they are mature enough.

Interfaces: We have continued to enhance the Alquimia interface. We improved the recovery from a failed geochemistry reaction step, and support for gas phase transport is under development (Section 4.1.5). Community interest in the Alquimia interface has grown, with independent work by Dr. Ben Sulman (supported through DOE Early Career Research and NGEE Arctic) enabling the treatment of immobile species, and using it in the coupling of ELM and PFLOTRAN. We plan to work with Dr. Sulman to integrate this capability in a future release. Development of the PHREEQc interface has been delayed due to departure of key staff from the ORNL SFA, and will be prioritized when we can identify a strong use case and an interested user. Interest in a similar interface for Land Models remains high, but the problem is complex. We made good progress on the ATS to PF-CLM interface development and testing, and expect this to be formally released later this year. This represents the fine-scale hydrology coupling to Land Model interface concept, while the Land Model to fine-scale hydrology coupling is now being explored in the COMPASS projects. In addition, coupling to FATES has been explored in both contexts as well. Since there is significant interest in the community, we anticipate working with the CI working groups to hold a short virtual workshop on this topic this fall.

Sustainable Software Ecosystem: We have made good progress developing a new web site for the Ideas-Watersheds project and software ecosystem (ideas-watersheds.github.io). After exploring a variety of tools, we have settled on using the Jekyll templating system that is integrated with GitHub. We believe this will be easier to maintain and makes it easier to integrate automated tools (e.g., tools used at software.llnl.gov) for collecting content and metadata about the codes in the software ecosystem. In addition, we have decided to make a more formal distinction between codes that we are actively developing (Primary Codes), and those that we use “as is” from other projects (Third Party Codes). Building on the policies developed for the Extreme-scale Software Development Kit (xSDK) and the Extreme-scale Scientific Software Stack (E4S) we will develop policies for Primary Codes and Third Party Codes that we will consider members of our software ecosystem. Since primary codes are under active development in our community we will include in these policies registration with DOE Code, and requirements for additional metadata (such as documentation, tutorials, project affiliations, demos), which will be displayed on our software ecosystem web site. We will work with Tim Scheibe to ensure the software ecosystem resources that we develop on GitHub are coordinated with, and complementary to, the development of the CD-MII platform. Finally, we worked with ESS PMs to improve the software release form that helps clarify the purpose and review status of new or updated open-source software in order to accelerate PMs approval of the release.

To provide user friendly deployments of the software ecosystem we are exploring both Spack and Docker. Spack is the scientific package manager for the Exascale Computing Project (ECP) and both xSDK and E4S require Spack builds of their member packages. We have continued to improve our Spack builds and anticipate formally submitting Spack package files for our codes over the next few months. In addition, we have been steadily improving our Docker builds of many of our primary codes (e.g., ParFlow, and Amanzi-ATS), with increasing use of these containers in tutorials and first-time users.

FY22 Plans

- Explore interoperability of components of Watershed Workflow and TINerator, and consider integration of other workflow tools (KBase, DART). Reach out to CD-MII to connect with their workflow tools and development.
- Develop guidance / best practices for uploading model data (input/output) to ESS-DIVE
- Develop policies for membership of Primary and Third Party Codes in the software ecosystem.
- Develop both Spack and Docker deployments of the primary codes in the IDEAS-Watersheds Software Ecosystem (explore registering this collection as an SDK in E4S or xSDK).

4.2 Select Research Highlights

In FY21, a total of 16 manuscripts were published and 12 presentations provided by IDEAS-Watersheds. Both ParFlow and ATS had significant stable code releases for the community this past year. See Appendix A for a full list of publications, presentations and code releases. In this section we highlight four notable accomplishments.

A NEW APPROACH TO SIMULATE MULTICOMPONENT SOLUTE TRANSPORT AND GEOCHEMICAL REACTIONS IN INTEGRATED HYDROLOGY PROBLEMS

New Science: Investigation of hydrological systems has become integrated, with consideration of water fluxes in both the overland and the subsurface compartments in spatially distributed models. Only recently, however, reactive transport models have been developed for integrated hydrology problems. Here we developed a new model for the simulation of multicomponent solute transport and geochemical reactions coupled to surface-subsurface flow, where reactive transport is considered in both the overland and the subsurface domains. A novel scheme for the conservation of the mass of solutes has been implemented that is applicable to any of the possible evolutions of the surface water (dry, dry-to-wet, wet, wet-to-dry) (Figure ZZ). Geochemical reactions have been incorporated using the principle of interoperable development via the Alquimia interface (Figure 10).

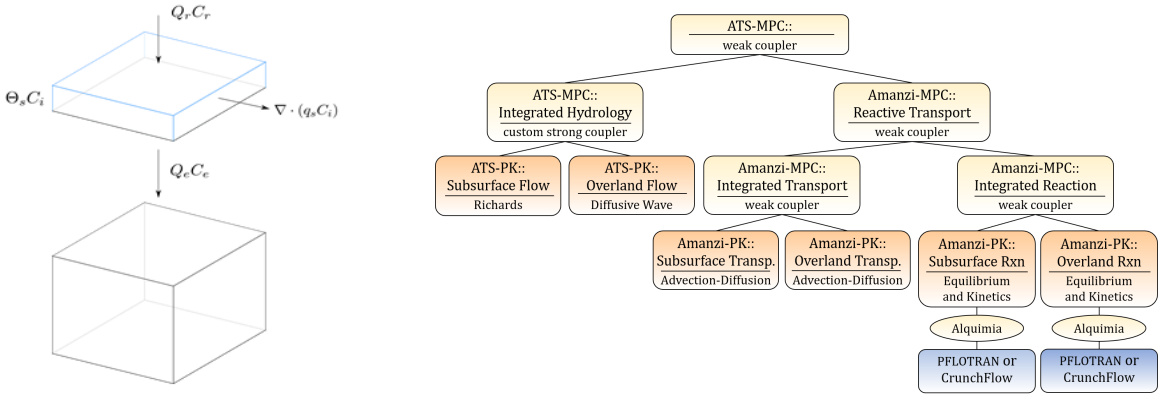


Figure 10: Mass balance between surface and subsurface (left), Process kernel (PK) tree for integrated hydrology coupled with reactive transport using Alquimia to support different reaction networks in the surface and subsurface (right).

Significance: The new model has been implemented in the Advanced Terrestrial Simulator and has already enabled the investigation of concentration-discharge relationships at the Copper Creek catchment of the East River. This capability will enable investigation of water quality in watershed subsystems (floodplains, meadows, hillslopes) as well as the watershed as a whole.

Citation: Molins, Svyatskiy, Xu, Coon, and Moulton (2021) A New Approach to Simulate Multicomponent Solute Transport and Geochemical Reactions in Integrated Hydrology Problems (Draft with target submission date: July 2021).

MULTISCALE MODEL FOR NETWORK-SCALE REACTIVE TRANSPORT IN RIVER CORRIDORS IMPLEMENTED IN ATS

New Science: The implementation of ADELS in ATS was extended from conservative tracers to multicomponent reactive transport. The implementation uses topologically defined meshes to represent stream/river networks and associates a hyporheic subgrid model with each channel grid cell. The Alquimia interface is used to access the general geochemical reaction capability within the PFLOTRAN software. Multiple subgrid models with different reaction models may be specified for each channel grid cell to represent, for example, metabolically active and inactive transient storage. The meshing infrastructure required for the multiple computational meshes of the multiscale model is an extension of earlier work from the IDEAS-Classic project. The implementation was verified against independent solutions. Mesh convergence studies using denitrification as an example revealed that semi-distributed conceptualizations like that used for reactive transport in the SWAT model can produce significant spatial discretization error (Figure 11, left panel). Additional simulations using a subbasin of the Portage River Basin reveal complex spatial patterns in denitrification including spatially localized hotspots (Figure 11, right panel) that are difficult to represent in earlier simulation tools.

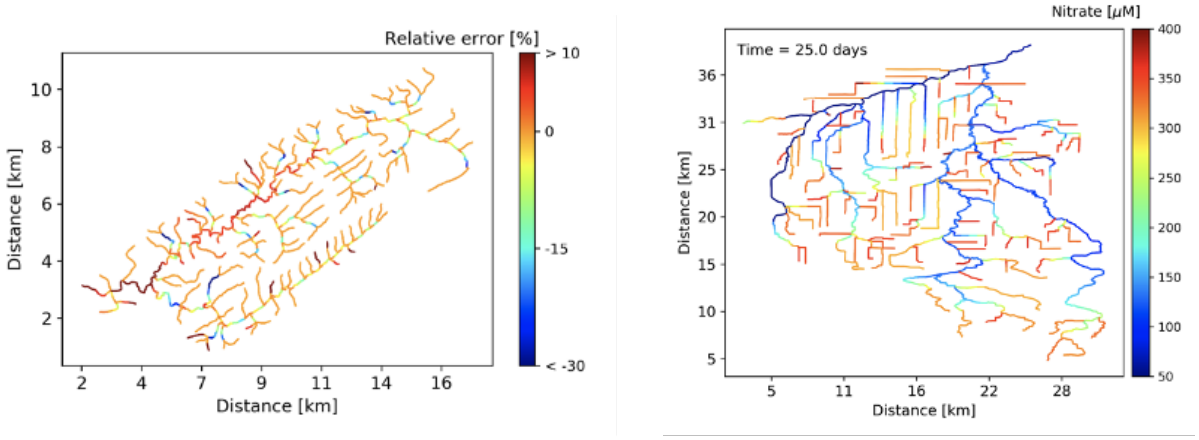


Figure 11. Left: Difference in calculated denitrification between coarse and fine resolution representations of the stream network in EFPC. The coarse representation, which used one channel grid cell per channel segment as in semi-distributed models like SWAT, resulted in significant spatial discretization error. Right: Calculated nitrate concentration in a nitrate loading experiment in a sub-basin of the Portage River Basin.

Significance: This work makes it possible, for the first time, to directly represent the important effects of fine-scale geochemical gradients in the hyporheic zone in river network-scale models. The new capability enables a new generation of models for nutrient, carbon, and metal processing in stream and river networks.

Citation: Jan, A., E. T. Coon, and S. L. Painter (2021), Toward more mechanistic representations of biogeochemical processes in stream and river networks: Implementation and demonstration of a multiscale model. In review

DART-PFLOTRAN: AN ENSEMBLE-BASED DATA ASSIMILATION SYSTEM FOR ESTIMATING SUBSURFACE FLOW AND TRANSPORT MODEL PARAMETERS

New Science: Iterative Ensemble-based Data Assimilation (EDA) methods, such as the ensemble smoother with multiple data assimilation (ES-MDA), have been developed to alleviate the accuracy deterioration caused by the nonlinear relationship in the evolution of model states or between model parameters and model states. We developed a generic EDA software framework by linking two open-source community tools: Data Assimilation Research Testbed (DART) as the core assimilation engine and PFLOTRAN as the subsurface flow and transport model, named DART-PFLOTRAN (Figure 12). We implemented DART-PFLOTRAN in a combination of Python, C-shell, and Fortran. We also provided a Jupyter notebook

template as an alternative to python scripting for users to configure the data assimilation options, set up forward simulation models, and eventually execute the sequential ES-MDA workflow using the C-shell script. We then verified the implementation of ES-MDA in DART-PFLOTRAN using two synthetic cases designed to estimate static permeability and dynamic exchange fluxes across the riverbed from continuous temperature measurements. Both cases yielded accurate estimations of the parameters compared to their synthetic truth.

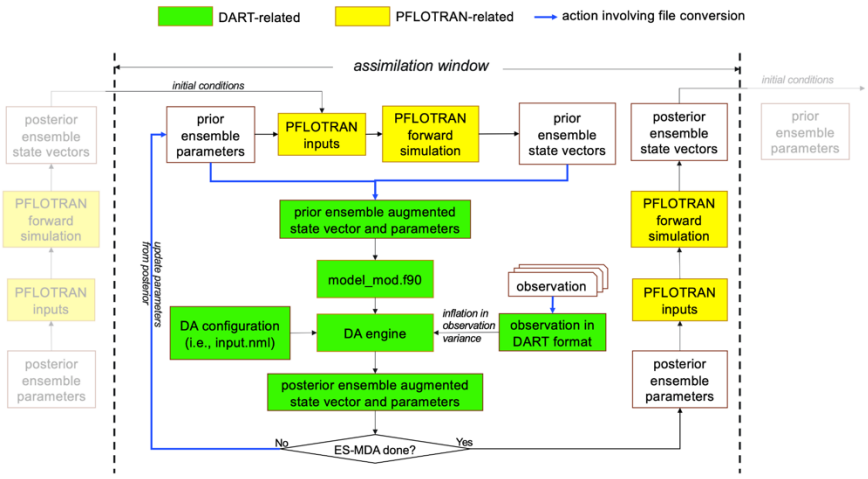


Figure 12: Design of the DART-PFLOTRAN software framework in one assimilation window. Green and yellow boxes are files associated with DART and PFLOTRAN, respectively; and blue arrows represent actions involving file format conversion.

Significance: The user-friendly software framework for performing EDA with computationally intensive forward models and heterogeneous observational data will significantly increase scientific productivity. With the added flexibility in subsetting observation data in space and time, DART-PFLOTRAN is poised for large-scale applications with complex and highly heterogeneous terrain. The workflow developed to link DART and PFLOTRAN can also be extended to link DART with other similar simulators such as the Advanced Terrestrial Simulator and ParFlow, which will greatly accelerate the integration of multi-scale and multi-type observations above and below ground with watershed models to improve the predictability of a wide variety of real systems.

Code availability: The source code of DART-PFLOTRAN is available at https://gitlab.pnnl.gov/sbrsfa/dart-pflotran.

Citation: Peishi Jiang, Xingyuan Chen, Kewei Chen, Jeffrey Anderson, Nancy Collins, and Mohamad EL Gharamti (2021). "DART-PFLOTRAN: An Ensemble-based Data Assimilation System for Estimating Subsurface Flow and Transport Model Parameters. Environmental Modelling & Software", 105074.

NOVEL DATASETS FOR NATIONAL HYDROLOGIC MODELING

New Science: Our team has developed a series of improved national datasets including topography, hydrostratigraphy, Manning's roughness and human reservoir operations. The topographic dataset (highlighted Figure 13) uses a novel hybrid priority flow approach developed to ensure realistic hydrologic drainage patterns that align with river network mappings. The national dataset was evaluated for drainage properties and stream network drainage areas across the US. Additionally, we have developed a

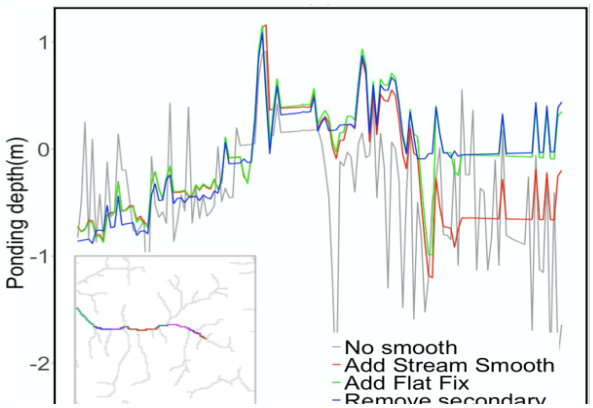


Figure 13. Ponding depths along a sample stream reach (shown on map) illustrating the impact of topographic smoothing operations on simulated flows.

subsurface dataset that includes hydrogeologic properties extending hundreds of meters below the surface. This dataset combines existing global, and national datasets to provide an integrated 3D dataset that includes hydrologic properties in addition to geologic classifications.

Significance: This dataset is the only national topography product processed specifically for gridded hydrologic modeling. It addresses the challenges of slope inconsistencies and drainage pattern mismatches existing in other products. The subsurface dataset is the only 3D hydraulic conductivity product for the US.

Citation: Zhang, J, LE Condon, H Tran and RM Maxwell. A National topographic dataset for hydrological modeling over the contiguous United States, Earth System Science data, In Press.

4.3 Training, Community Building, and Outreach

The commitment of the IDEAS project to the BER and broader community is evident in the wide range of outreach activities it has led and supported. Our community-building activities are aimed at raising awareness of the fundamental role software plays in overall scientific productivity, and engaging the community in discussions of how software can be better developed, used, and sustained.

Key Objectives:

1. Training in modern software development methodologies, workflows, and modeling.
2. Connecting the SFAs and Cornerstone Activities.
3. Growing the watershed science community.

Project Communication & Team Building Activities

- Virtual All Hands Meeting: Status Updates, Activity Brainstorming, and Team Building, Zoom 9:00AM-1:00PM Pacific Time each day, October 7-8, 2020.
- IDEAS-Watersheds Bi-Weekly Email Digest
- IDEAS-Watersheds Slack Channel
- IDEAS-Watersheds mailing lists (team, leads)
- IDEAS-Watersheds Leads Bi-Weekly Meeting
- Partnership - LBNL weekly meeting (includes team members from other labs)
- Partnership - PNNL weekly meeting (includes team members from other labs)

Training Events

- ***Sept 2019 – present – Software Best Practices Seminar***
  Steve Smith is leading a bi-weekly webinar series on Software Best Practices topics. Based on team discussions we selected topics of highest interest with focus on the unique challenges of scientific and HPC DevOps practices. Workflow topics included Juptyer Notebooks and using the Unix command line efficiently. Continuous integration was discussed in several sessions by projects using both Travis CI and Google Actions. Projects shared strategies for addressing performance portability on accelerators. Software deployment was discussed in sessions on DOE Code presented by Sara Studwell (OSTI) and on Spack presented by Todd Gamblin (LLNL).
- ***Sept 2019 – present – Modeling Case Studies Bi-Weekly Seminar***
  David Moulton is leading a bi-weekly webinar series on Modeling Case Studies. This series is designed to raise awareness of the challenges and approaches to mechanistic modeling of complex environmental systems. Modeling itself is often overlooked, here we share experiences and current work across the project, and strive to stimulate new collaborations on synergistic topics. For example, we had a series of presentations on approaches to multiscale reactive transport modeling, which led to an invitation to Jesus Gomez-Velez to present his NEXSS model, and now a new collaboration between ORNL, PNNL and Jesus's research group. Other topic we have covered include, DEM processing, mesh generation, reactive transport in integrated hydrology models, and particle tracking algorithms. Upcoming topics include land models, and colloid facilitated reactive transport.

Outreach

Outreach activities that benefit the broader community led by project members in FY21 include:

- [E. Coon, A. Jan, D. Moulton, S. Molins, S. Painter, Z. Xu] ATS (virtual) short course Sept. 1-2, 2021.
    - First ATS training course, jointly supported by ExaSheds and IDEAS-Watersheds
- [R. Maxwell, L. Condon] ParFlow Shortcourse Series
    - Joint US-Germany ParFlow short course rescheduled due to Covid (Summer 2022).
- Planning ESS Cyberinfrastructure Virtual Annual Meeting (target September 2021). Moulton leads the working groups Executive Committee, and is co-organizing the meeting along with Xingyuan Chen from IDEAS-Watersheds. He is also leading development of a White Paper on the short- and long-term vision for the CI working groups and CI in ESS.
- [R. Maxwell] A Better Understanding of Water Availability in the U.S. Through Community Tools, a talk at Princeton's High Meadows Environmental Institute, https://youtu.be/R5cokxumg9U, May 4, 2021.
- [L. Condon] Water Beneath our Feet, a talk on groundwater in University of Arizona College of Science lecture series, https://youtu.be/-70OZRRDTfQ, April 22, 2021.
- [D. Moulton, E. Coon] Continuous Improvement of Software Engineering, briefing to USGS Integrated Modeling and Prediction on Research Software Engineering, February 26, 2021.
- [D. Moulton] IDEAS-Watersheds: Accelerating watershed science through a community-driven software ecosystem, Interagency Coastal IHTM Working Group, February 16, 2021.
- [R. Maxwell] 2020 Darcy Lecturer. The Henry Darcy Distinguished Lecture Series in Groundwater Science chooses one outstanding groundwater professional to share their work both nationally and internationally on a year-long lecture tour. His talks highlight the CONUS model development and the challenges and opportunities of computational hydrology. This has been an amazing opportunity to highlight some of the work of the IDEAS project and to bring attention to all of the work happening at the intersection of hydrology and computer science.
- [K. Chen and H.-S. Song] demonstrated the effectiveness of our metagenome-to-reactive transport modeling workflow through the 2020 Multiscale Microbial Dynamics Modeling Summer School co-organized by EMSL and PNNL River Corridor SFA, July 6-10 2020

4.4 Data & Software Management Activities

Due to the partnership structure of IDEAS-Watersheds, simulation data generated on the three SFA use cases will be managed through SFAs, complying with their specific data management plans.

ESS-DIVE will ultimately serve as the final repository for archiving, searching, and dissemination of ESS data products. Our team is working with the ESS-DIVE team to improve their initial guidance for archiving model data packages by enhancing the descriptions of the README content, as well as creating example data sets for submission, archiving and sharing.

IDEAS-CONUS generates large output (many terabytes per simulation) due to the combination of large spatial scales and high resolution. All published CONUS simulations are currently stored with DOIs on a server at UArizona that is publicly accessible through the Cyverse cyber infrastructure project. Large simulations are stored at the Princeton HydroData center maintained by coPI Maxwell. Project members all have access to this server both through Princeton as well as through the HydroFrame interface.

The IDEAS-Watersheds Software Ecosystem is comprised of the following entities: Primary Codes and Primary Workflow Tools, in which our team has an active development role, as well as Third-Party Codes and Third-Party Workflow Tools, which our team are uses. We note that Primary Codes include domain science components and libraries (e.g., Alquimia). In addition, this software ecosystem depends on other Third-Party libraries for lower-level infrastructure, frameworks and mathematical libraries. To help manage the software ecosystem, IDEAS-Watersheds has purchased GitHub Teams to orchestrate various efforts and provide a central location to store data products: https://github.com/IDEAS-Watersheds.

For FY22, we plan to track software activities such as code releases, software downloads, and community activities.

5 STAFFING AND BUDGET SUMMARY

Budget Summary

Institution	Year 1	Year 2	Year 3	Total
Los Alamos National Laboratory & Subcontracts				
Internal LANL	$700K	$700K	$700K	$2,100K
Colorado School of Mines (CSM)	$100K	$100K	$100K	$300K
University of Arizona (UA)	$100K	$100K	$100K	$300K
National Center for Atmospheric Research (NCAR)	$100K	$100K	$100K	$300K
Lawrence Berkeley National Laboratory	$400K	$400K	$400K	$1,200K
Lawrence Livermore National Laboratory	$250K	$250K	$250K	$750K
Oak Ridge National Laboratory	$300K	$300K	$300K	$900K
Pacific Northwest National Laboratory	$300K	$300K	$300K	$900K
Total	$2,250K	$2,250K	$2,250K	**$6,750K**

Subcontracts: All subcontracts were awarded in FY20 to UA, CSM and NCAR. PI Maxwell moved to Princeton University, starting Sept 2020 but transitioning slowly to remain at both CSM and Princeton until September 2021. The subcontract funding for CSM will move to Princeton for Y2 and Y3.

Budget Element	FY21 Budget ($k)	Y2 Effort	3 YR Total
Management (reporting, website, advisory, committee, project support)	205	0.33	**1 FTE**
4.1 Watershed Function SFA Partnership: East River Use Case	220	0.70	2.1 FTE
4.2 Critical Interfaces SFA Partnership: East Fork Poplar Creek Use Case	300	0.87	2.6 FTE
4.3 River Corridor SFA Partnership: Columbia River Use Case	284	1.06	3.2 FTE
4.4 Continental Modeling Platform and Simulations	420	2.45	7.35 FTE
4.5 Fine-Scale Activities: Reaction Networks	307	0.92	2.8 FTE
4.6 Shared Infrastructure to Advance Software Ecosystem	372	1.10	3.2 FTE
5.4 Training, Community Building, Outreach	142	0.20	0.6 FTE
Totals	*2,250*	7.63	22.85 FTE

Personnel Actions: Due to the budget reduction for CSM and UA reflected in the Budget Summary, IDEAS-Watersheds will only be able to partially support a PD and/or GRA (not fully support both as originally planned). For FY20, both PD and GRA were funded through carryover funds. For FY21, Jun Zhang (UA) will be funded through ICoM and Jen Steyart will start her PhD funded through this project. Mary Forrester (CSM) finished her PhD in Spring 2020 and was not funded this past year. Jackson Swilley (Mines) began working with IDEAS-Watersheds in Summer 2020 but TA'd ¾ of that year, Anna Ryken (Mines) was funded by the LBL SFA, graduated in Spring 2021, and worked only a small fraction on IDEAS. Jackson Swilley and PD Hoang Tran move to Princeton with PI Maxwell and were funded by other projects (NSF) though contributed heavily to IDEAS deliverables. PD Chen Yang was funded at Princeton under IDEAS at 40% time. Dr. Yang worked on GPU parallelism for EcoSLIM.

Hai Ah Nam recently moved from LANL to LBNL, but is unable to continue in this role. We are actively seeking a new project coordinator. Joe Beismen (PD) left LANL and we are seeking a post doc in support of integrated hydrology with reactive transport for the Reaction Network scale SFA sites.

6 REFERENCES CITED

Özgen-Xian, I., Kesserwani, G., Caviedes-Voullième, D., Molins, S., Xu, Z., Dwivedi, D., et al. (2020). Wavelet-based local mesh refinement for rainfall–runoff simulations. Journal of Hydroinformatics. https://doi.org/10.2166/hydro.2020.198

Özgen-Xian, I. (2020). Wavelet-based mesh refinement, Pull request for TINerator merged by Daniel Livingston, https://github.com/lanl/LaGriT/pull/211, November 2020.

Tarboton, D. G. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resources Research, 33(2), 309–319. https://doi.org/10.1029/96WR03137

Painter, S. L. (2021), On the Representation of Hyporheic Exchange in Models for Reactive Transport in Stream and River Corridors, Frontiers in Water, 2(69), doi: 10.3389/frwa.2020.595538.

Rathore, S. S., A. Jan, E. T. Coon, and S. L. Painter (2021), On the Reliability of Parameter Inferences in a Multiscale Model for Transport in Stream Corridors, Water Resour Res, 57(5), e2020WR028908, doi: https://doi.org/10.1029/2020WR028908.

Rathore, S. S., G. E. Schwartz, S. C. Brooks, and S. L. Painter (2021), Model Selection and Joint Estimation of Biogeochemical Model Parameters from Multiple Experiments: A Bayesian Approach Applied to Mercury Methylation. In review

Jan, A., E. T. Coon, and S. L. Painter (2021), Toward more mechanistic representations of biogeochemical processes in stream and river networks: Implementation and demonstration of a multiscale model. In review

Schwartz, G.E., Muller, K.A., Rathore, S.S., Wilpiszeski, R.L., Carrell, A.A., Cregger, M.A., Elias, D.A., Podar, M., Painter, S.L., Brooks, S.C., 2021. Incorporating Variable Sediment Microbial Activity into Methylmercury Production Kinetics Modeling. Science of The Total Environment Under Review.

Song, Hyun-Seob, James C. Stegen, Emily B. Graham, Joon-Yong Lee, Vanessa A. Garayburu-Caruso, William C. Nelson, Xingyuan Chen, J. David Moulton, and Timothy D. Scheibe. "Representing organic matter thermodynamics in biogeochemical reactions via substrate-explicit modeling." *Frontiers in microbiology* 11 (2020): 2625.

de Graaf, I, Condon L. E. and Maxwell R. M., Hyper-resolution continental-scale 3D-aquifer parameterization for groundwater modelling, Water Resources Research, 56, e2019WR026004, https://doi.org/10.1029/2019WR026004, 2020.

Yang, C., et al., *Accelerating the Lagrangian particle tracking of residence time distributions and source water mixing towards large scales*. Computers & Geosciences, 2021. 151: p. 104760.

Zhang, J, LE Condon, H Tran and RM Maxwell. A National topographic dataset for hydrological modeling over the contiguous United States, Earth System Science data, In Press.

Appendix A: **Publications**

Watershed Function SFA Partnership: East River Use Case
PUBLICATIONS

1. Özgen-Xian, I., Navas-Montilla, A., Dwivedi, D., & Molins, S. (2021). Hyperbolic Reformulation Approach to Enable Efficient Simulation of Groundwater Flow and Reactive Transport. Environmental Engineering Science, 38(3), 181–191. https://doi.org/10.1089/ees.2020.0363
2. I. Özgen-Xian, G. Kesserwani, D. Caviedes-Voullième, S. Molins, Z. Xu, D. Dwivedi, J. D. Moulton, C. I. Steefel, "Wavelet-based local mesh refinement for rainfall-runoff simulations" Journal of Hydroinformatics, **in press** (2020).

PRESENTATIONS

3. S. Molins, D. Svyatskiy, E. Coon, Z. Xu, and D. Moulton (2021) (Invited) Modeling Reactive Transport across the Surface-Subsurface Interface, SIAM Geosciences 2021
4. Özgen-Xian, I., Caviedes-Voullieme, D., Crompton, O., Morales-Hernandez, M., Dwivedi, D., Xu, Z., Molins S., Steefel, C.I. (2020). Overland flow connectivity states in Lower Triangle Region, East River, Colorado, AGU Fall Meeting 2020, H215–0012.
5. Molins, S., Svyatskiy, D., Xu, Z., & Moulton, J. D. (2020). A reactive transport model for integrated surface-subsurface hydrology, AGU Fall Meeting 2020, H112–0019.
6. Brodie, E., King, E., Marschmann, G., Karaoz, U., Molins, S., & Steefel, C. I. (2020). Exploring The Relationship Between Biological Complexity, Spatial Heterogeneity and Process Rates In A Genome-informed Reactive Transport Model, AGU Fall Meeting 2020, H112–0019.B078–0010.

Critical Interfaces SFA Partnership: East Fork Poplar Creek Use Case
PUBLICATIONS

7. S. L. Painter (2021) On the Representation of Hyporheic Exchange in Models for Reactive Transport in Stream and River Corridors. Frontiers in Water, 2:595538, https://doi.org/10.3389/frwa.2020.595538
8. Rathore, S. S., A. Jan, E. T. Coon, and S. L. Painter (2021). On the reliability of parameter inferences in a multiscale model for transport in stream corridors. Water Resources Research, 57, https://doi.org/10.1029/2020WR028908

PRESENTATIONS

9. Rathore, S. et al. (2020) Improving parameter inferences in a multiscale model for transport in stream corridors. AGU Fall Meeting abstract H174-09, San Francisco, CA.

River Corridor SFA Partnership: Columbia River Use Case
PUBLICATIONS

10. Jiang P., X. Chen, K. Chen, J. Anderson, N. Collins, T. Hoar, and M. Gharamti. 2021. "DART-PFLOTRAN: An Ensemble-based Data Assimilation System for Estimating Subsurface Flow and Transport Model Parameters." Environmental Modeling & Software. Accepted
11. Wu R., X. Chen, G.E. Hammond, G. Bisht, X. Song, M. Huang, and G. Niu, et al. Coupling Surface Flow with High-performance Subsurface Reactive Flow and Transport Code PFLOTRAN. Environmental Modelling & Software, 10.1016/j.envsoft.2021.104959
12. Song H., J.C. Stegen, E.B. Graham, J. Lee, V.A. Garayburu-Caruso, W.C. Nelson, and X. Chen, et al. 2020. "Representing Organic Matter Thermodynamics in Biogeochemical Reactions via Substrate-Explicit Modeling." Frontiers in Microbiology, 10.3389/fmicb.2020.531756

PRESENTATIONS

13. Hyun-Seob Song , Christopher Henry , Joon-Yong Lee , William C Nelson , Janaka Edirisinghe, Jianqiu Zheng , John D Moulton , Xingyuan Chen , Emily Bonnell Graham , James Stegen and Timothy D Scheibe (2020). Machine Learning-based Metabolic Network Modeling for Omics-integrated Biogeochemical and Reactive Transport Simulations. AGU Fall Meeing, Online, USA.
14. Jiang P., X. Chen, K. Chen, J. Anderson, N. Collins, T. Hoar, and M. Gharamti (2020). Using Information Flow for Whole System Understanding from Component Dynamics." AGU Fall Meeting, Online, United States.

Continental Modeling Platform and Simulations
PUBLICATIONS

15. Tijerina, D., L. Condon, K. FitzGerald, A. Dugger, M. Forrester, W. Yu, D. Gochis, R. M. Maxwell, (2021, in press) CHIP - Continental Hydrologic Intercomparison Project: a framework for large-scale hydrology model comparisons. Bull. Amer. Meteor. Soc.
16. Zhang, J., L.E. Condon, H. Tran and R.M. Maxwell. (2021, in press) A national topographic dataset for hydrologic modeling over the contiguous United States, Earth Systems Science Data.
17. O'Neill, M. M. F., Tijerina, D. T., Condon, L. E., and Maxwell, R. M. (2020)Assessment of the ParFlow-CLM CONUS 1.0 integrated hydrologic model: Evaluation of hyper-resolution water balance components across the contiguous United States, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-364.
18. Chen X., R.M. Lee, D. Dwivedi, K. Son, Y. Fang, X. Zhang, and E.B. Graham, et al (2020). "Integrating Field Observations and Process-based Modeling to Predict Watershed Water Quality under Environmental Perturbations." Journal of Hydrology, 125762. Doi: 10.1016/j.jhydrol.2020.125762
19. DeGraaf, I, LE Condon and RM Maxwell (2020). Hyper-resolution continental-scale 3D-aquifer parameterization for groundwater modelling, Water Resources Research,56, e2019WR026004, https://doi.org/10.1029/2019WR026004
20. Kuffour, B.N.O., Engdahl, N.B., Woodward, C.S., Condon, L.E., Kollet, S.J. and and Maxwell, R.M. (2020) Simulating coupled surface-subsurface flows with ParFlow v3.5.0: capabilities, applications, and ongoing development of an open-source, massively parallel, integrated hydrologic model. Geoscientific Model Development, 13:1373-1397, https://doi.org/10.5194/gmd-13-1373-2020.
21. Condon, LE, AL Atchley and RM Maxwell (2020). Evapotranspiration depletes groundwater under warming over the contiguous United States. Nature Communications, 11(873).
22. Forrester, M.M. and Maxwell, R.M. Impact of lateral groundwater flow and subsurface lower boundary conditions on atmospheric boundary layer development over complex terrain. Journal of Hydrometeorology, in press, doi:10.1175/JHM-D-19-0029.1, 2020
23. Forrester, M.M., Condon, L.E., Tijerina, D., Maxwell, R.M., Continental-scale modeling of terrestrial water and energy at hyper resolution: Performance analysis of the coupled land surface- hydrology model ParFlow-CLM configured over the contiguous United States (CONUS version 1.0), in prep, submission targeted for Geoscientific Model Development.

PRESENTATIONS

24. Zhang, J., Condon, L.E., Moulton, J.D., Zhang, Y. and Bisht, G., 2020, December. Comparison

between multiple physical-based hydrologic models in modeling groundwater-surface water interactions in a coastal watershed. American Geophysical Union Fall Meeting, San Francisco, CA and Virtual, December 2020. Enlightening Poster presentation.

25. Steyaert, J., Condon, L. E., Turner, S.D.,Voisin, N., A Retrospective Analysis of Reservoir Operations: What Actually Happens During a Drought?, Talk, El Dia del Agua y del Atmosfera Tucson, AZ, March 31st, 2021
26. Steyaert, J., Condon, L. E., Turner, S.D., A Retrospective Analysis of Reservoir Operations: What Actually Happens During a Drought?, Talk American Geophysical Union Fall Meeting, Abstract: H155-05, San Francisco, CA and Virtual, December 2020.
27. Condon, L.E. "Hydrologic Modeling at the Continental Scale: Exploring the Role of Data Science in Discovery and Innovation", Women in Data Science Tucson, Virtual Conference, April 2021,

Shared Infrastructure
PRESENTATIONS

28. Moulton, J. D., A Team of Teams Approach to Sustainable Software Ecosystems, presentation in SIAM CSE21 Minisymposium: Building Sustainable Software Communities and Sustainable Software https://doi.org/10.6084/m9.figshare.14192531.v2, March 2, 2021.

CODE RELEASES - Primary Codes

Advanced Terrestrial Simulator (Amanzi-ATS)
Leads: Ethan Coon (ATS) and David Moulton (Amanzi)
Version: 1.1
URL: https://github.com/amanzi/ats/releases/tag/ats-1.1.0
Updates include reliable integrated (surface and subsurface) transport, and reactive transport through the Alquimia geochemical interface library, allowing access to PFLOTRAN and Crunch reaction networks. Also includes improved support for multi-scale models of in-stream hyporheic exchange, support for the effects of salinity on liquid density, and preliminary support for sediment transport.

Alquimia
Lead: Sergi Molins
Version: 1.0.8
URL: https://github.com/LBL-EESA/alquimia-dev/releases/tag/v1.0.8
Addresses failed steps in the PFLOTRAN geochemistry engine, and a memory management issue in CrunchFlow.

ParFlow
Leads: Steve Smith, Reed Maxwell, Laura Condon
Version: 3.9.0
URL: https://github.com/parflow/parflow/releases/tag/v3.9.0
DOI: 10.5281/zenodo.4957977
Added support for building under Spack and package for ParFlow added to Spack releases

Version: 3.8.0
URL: https://github.com/parflow/parflow/releases/tag/v3.8.0
DOI: 10.5281/zenodo.4816885
Improved Python interface, added Kokkos support, XSDK 0.5 version support, added compression support for NetCDF, bugfixes for PFMGOctree solver

Appendix B: **Codes in the Software Ecosystem**

The IDEAS-Watersheds Software Ecosystem is comprised of Primary Codes and Primary Workflow Tools, in which our team has an active development role, as well as Third-Party Codes and Third-Party Workflow Tools, which our team are users. We note that Primary Codes include domain science components and libraries (e.g., Alquimia).

Primary Codes

B.1 Alquimia

Alquimia is a biogeochemistry Application Program Interface and wrapper library that provides unified access to biogeochemistry capabilities from mature geochemical codes, allowing any subsurface flow and transport simulator to access a range of functionality. Alquimia is not an implementation of a biogeochemistry reaction library and does not do any geochemical calculations. It provides unified data structures and subroutine signatures so that existing geochemical codes perform these calculations. Currently, Alquimia provides access to the geochemical codes PFLOTRAN and CrunchFlow and can be used for the simulation of aqueous complexation reactions, radioactive decay, ion exchange, surface complexation, and mineral dissolution-precipitation.

B.2 Amanzi

Amanzi provides a flexible and extensible flow and reactive transport simulation capability for environmental applications. It includes general polyhedral mesh infrastructure, which leverages MSTK, advanced discretizations of process models, including traditional finite volume schemes, mimetic finite differences, and nonlinear finite volumes. In addition, it provides advanced nonlinear solvers, such as Nonlinear Krylov Acceleration (NKA) and Anderson Acceleration, and leverages Trilinos-ML and Hypre Algebraic Multigrid for scalable solvers. The multiphysics framework, Arcos, provides a flexibility for hierarchical weak and strong coupling of processes with subcycling. Geochemistry support is provided through the Alquimia interface library and can use the geochemistry engine from PFLOTRAN or CrunchFlow. The code is parallel and makes leverages Trilinos and PETSc. of open-source parallel frameworks such as Trilinos, PETSc. Amanzi is used to model contaminant migration at various DOE waste sites (e.g., Hanford, and Savannah River), and is generally applicable to groundwater contaminant migration under partially saturated, nonisothermal conditions and its interaction with surface water. Amanzi is jointly developed by LANL, LBNL, and PNNL, and ORNL as an open source project under the three-clause BSD license.

B.3 ATS

The Advanced Terrestrial Simulator is a code for solving ecosystem-based, integrated, distributed hydrology. It builds on the multi-physics framework and toolsets (mesh infrastructure, discretizations, solvers) provided by Amanzi and is a key driver of development of the flexible multiphysics framework Acros. Capabilities are largely based on solving various forms of Richards equation coupled to a surface flow equation, along with the needed sources and sinks for ecosystem and climate models. This can (but need not) include thermal processes (especially ice for frozen soils), evapotranspiration, albedo-driven surface energy balances, snow, biogeochemistry, plant dynamics, deformation, transport, and much more. It is unique in its thermal integrated hydrology capabilities, which includes thermal energy with freeze/thaw processes in both the surface and subsurface water, and its reactive transport capabilities, which are also coupled in both surface and subsurface water. ATS is jointly developed by LANL, LBNL, and ORNL as an open source project under the three-clause BSD license.

B.4 CrunchFlow

CrunchFlow is an open-source software package for simulating reactive flow and transport applied to a variety of problems in the Earth and environmental sciences that can include an arbitrary number of reactions, including mineral dissolution/precipitation, ion exchange, surface complexation, and microbially mediated reactions (Steefel et al, 2015). Two approaches are available at runtime for coupling reactions and transport: (1) a global-implicit approach that solves transport and reactions simultaneously, and (2) an operator-splitting approach based on a sequential non-iterative solve.

B.5 EcoSLIM

EcoSLIM is a Lagrangian particle-tracking model designed to operate with integrated hydrologic models.(Reed M. Maxwell et al., 2018) Particles are initialized in the subsurface or added at the surface via precipitation fluxes. Spatially gridded velocity, pressure, and saturation fields are used to move particles through the domain and particles can exit via overland flow or evapotranspiration. In addition to steady-state simulations, EcoSLIM can also use transient model outputs to develop time-varying residence-time distributions. The incorporation of precipitation sources from the land-surface model also allows for the direct evaluation of source water contribution (e.g., rain versus snow). EcoSLIM has been demonstrated using outputs from the ParFlow-CLM hydrologic model; however, the model can easily be combined with other hydrologic tools that generate similar gridded outputs. In proof-of-concept simulations, EcoSLIM has been used to explore spatial variability in plant water sources (i.e., old vs. young water), connections between plan rooting depth and ET residence-time distributions, and seasonal variability in residence-time distributions and source water partitioning.(Reed M. Maxwell et al., 2018)

B.6 ParFlow

The integrated watershed model ParFlow solves saturated and variably saturated flow in three dimensions using either an orthogonal (Jones & Woodward, 2001; Kollet & Maxwell, 2006) or a terrain-following, semi-structured mesh that enables fine vertical resolution near the land surface and deep (~1 km) confined and unconfined aquifers. ParFlow models dynamic surface and subsurface flow solving the simplified shallow water equations implicitly coupled to Richards' equation; this allows for two-way exchanges and intermittency in stream and river flow (Kollet & Maxwell, 2006). ParFlow uses robust linear (Ashby & Falgout, 1996) and nonlinear (Jones & Woodward, 2001; Osei-Kuffuor et al., 2014) solution techniques and exhibits efficient parallel scaling to large processor counts, more than 450K cores (Kollet et al., 2010), enabling very large-extent simulations with fine spatial resolution.

In addition, ParFlow is coupled to the land-surface model CLM (Ferguson et al., 2016; Jefferson & Maxwell, 2015; Jefferson et al., 2017; Reed M. Maxwell & Miller, 2005), providing a comprehensive representation of vegetation, snow, and land-atmosphere water and energy fluxes. CLM is unique in the land surface modeling community because it is a module that is called from within ParFlow. This modeling framework has been shown to represent the observed range of temporal scales and non-stationary behavior (R. M. Maxwell et al., 2015) making it appropriate for the proposed work. A 1 km lateral resolution ParFlow-CLM model of CONUS has been developed and used to evaluate large-scale controls on groundwater configuration and connections between lateral groundwater flow and land-surface partitioning (Laura E. Condon & Maxwell, 2015; L. E. Condon & Maxwell, 2017; Reed M. Maxwell & Condon, 2016; R. M. Maxwell et al., 2015). ParFlow-CLM is currently distributed with ParFlow through the GitHub repo. In the IDEAS project the CLM land model interface will be generalized and also be made available as a separate library that other codes in the eco-system can connect use still within the ParFlow GitHub repo.

B.7 PFLOTRAN

PFLOTRAN is an open-source, state-of-the-art massively parallel subsurface flow and reactive transport code. PFLOTRAN solves a system of generally nonlinear partial differential equations describing multiphase, multicomponent, and multiscale reactive flow and transport in porous materials. The code is designed to run on massively parallel computing architectures as well as workstations and laptops. PFLOTRAN is written in object-oriented, free-formatted FORTRAN 2003. The reactive transport equations can be solved using either a fully implicit Newton-Raphson algorithm or the less robust operator-splitting method. Geochemical capabilities include aqueous complexation, sorption, mineral precipitation and dissolution, and microbially mediated reactions.

Primary Workflow Tools

B.8 TINerator

TINerator is a Python module for the easy creation of unstructured 3D and 2.5D meshes from GIS data sources. Digital Elevation Maps (DEMs) can be quickly turned into unstructured triangulated meshes, and then further refined by the import of flowline shapefiles or automatically through watershed delineation. Advanced layering and attribute functions then allow for a complex subsurface to be defined. TINerator comes with a host of 2D and 3D visualization functions, allowing the user to view the status of the mesh at every step in the workflow. In addition, there are geometrical tools for removing triangles outside of a polygon, generating quality analytics on the mesh, adding cell- and node-based attributes to a mesh, and much more.

Third-Party Codes

B.9 OpenFOAM

Open source Computational Fluid Dynamics (CFD) software. It has an extensive range of capabilities to solve complex fluid flows involving turbulence, heat transfer and chemical reactions. It has established a large user community across most areas of engineering and Science.

Key Reference: http://openfoam.com/

B.10 SWAT

Soil & Water Assessment Tool. A watershed model developed for the USDA Agricultural Research Service. SWAT can be used to predict the impact of land management practices on water, sediment and agricultural chemical yields in large complex watersheds. It can deal with varying soils, land use and management conditions over long periods of time.

Third-Party Workflow Tools

B.11 LaGriT/PyLaGriT

LaGriT is a software tool for generating, editing and optimizing multi-material unstructured finite element grids; it also maintains the geometric integrity of complex input volumes, surfaces, and geologic data and produces an optimal grid (Delaunay, Voronoi) elements. The data structures used in the code are compact and powerful and expandable to include hybrid meshes (tet, hex, prism, pyramid, quadrilateral, triangle, line), however the main algorithms are for triangle and tetrahedral meshes. The LaGriT tools

are used in many projects including ASCEM meshing for Amanzi, Discrete Fracture Networks (DFN), Arctic Permafrost, and Subsurface Flow and Transport models using FEHM and PFLOTRAN. PyLaGriT provides a python interface to LaGriT capabilities, making it easier to incorporate mesh generation in modeling workflows.

B.12 Watershed Workflow

Watershed Workflow is a python-based, open source chain of tools for generating meshes and other data inputs for hyper-resolution hydrology, anywhere in the (conterminous + Alaska) US. Hyper-resolution hydrologic models have huge data requirements, thanks to their large extent (full river basins) and very high resolution (often ~10-100 meters). Furthermore, most process-rich models of integrated, distributed hydrology at this scale require meshes that understand both surface land cover and subsurface structure. Typical data needs for simulations such as these include: Watershed delineation (what is your domain?); Hydrography data (river network geometry, hydrographs for model evaluation); A digital elevation model (DEM) for surface topography; Surface land use / land cover; Subsurface soil types and properties; Meterological data, and more.

Appendix C: **Project Lead Roles and Responsibilities**

Table 1. IDEAS-Watersheds project management team

Team Member	Role	Area of Responsibility
David Moulton (LANL)	Project PI	Overall project, direct communication with SBR program managers. Co-lead the Integrated Domain and Computational Science team.
Hai Ah Nam (LANL)	Project Coordinator	Project management, tracking of deliverables, sub-contract development, collecting highlights, and development of cloud resources and workflows for team collaboration.
Steve Smith (LLNL)	Software Lead	Development of software methodologies best practices resources, guidance and training for the team. Co-lead the Integrated Domain and Computational Science team.
RESEARCH ACTIVITY LEADS		
Sergi Molins (LBNL)	Watershed Function SFA Partnership	Lead the Partnership Activities with the SBR SFAs at home organization. Each of the Partnership leads is currently supported on the associated SFA (typically 0.2–0.6 FTE) and will receive support from IDEAS-Watersheds.
Scott Painter (ORNL)	Critical Interfaces SFA Partnership	
Xingyuan Chen (PNNL)	River Corridor SFA Partnership	
Reed Maxwell (Princeton) **Laura Condon (UA)** **Dave Gochis (NCAR)**	CONUS Activity	Develop the CONUS2.0 domain, national simulations, data management, and workflow tools for model sub-setting and small-scale simulation.
Sergi Molins (LBNL)	Fine-Scale Partnership Activity	Lead and coordinate enhancements of biogeochemical reaction modeling capabilities with the ANL, SLAC, and LLNL SFAs.
David Moulton (LANL)	Shared Infrastructure Activity Lead	Lead and coordinate synergistic development to ensure new capabilities in Multiscale Workflows, Interfaces, and the software ecosystem can be used across all partnerships and CONUS.
David Moulton (LANL)	Training, Community Building, Outreach	Lead, develop and coordinate outreach activities with the broader community. Key elements of training in HPC will be led Nam and in software engineering methodologies by Smith.
CORNERSTONE LEADS		
Sergi Molins (LBNL) **Xingyuan Chen (PNNL)**	Reaction Networks Cornerstone Co-Leads	Coordinate and track development and enhancement of capabilities that support the simulation of biogeochemical reaction networks throughout the project. Chen focuses on metagenomics-to-reactions workflows. Molins focuses on other reaction network development.
Scott Painter (ORNL)	Watershed Hydrobiogeochemistry Cornerstone Lead	Coordinate with each of the Partnership teams to ensure they have the resources and integrated support they need for development and use of the new or enhanced capabilities, including support for model input and development, simulation campaigns, and analysis of results.
Laura Condon (UA) **Reed Maxwell (Princeton)**	Continental Hydrology Cornerstone Lead	Coordinate between the CONUS Activity and each Partnership Activity to ensure that the CONUS Activity is delivering products and capability of value for the SFA Partnerships. Condon focuses on workflow development and model sub-setting tools. Maxwell coordinates model coupling connections across the software ecosystem.