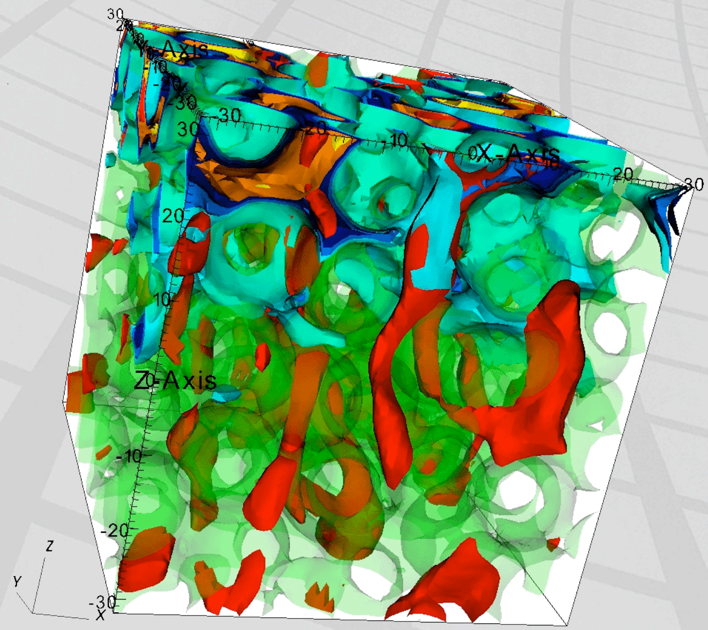
Environmental Science Case Study: Subsurface Reactive Transport Modeling



Timothy D. Scheibe

Staff Scientist
Pacific Northwest National Laboratory

Large Scale Production Computing Requirements for Biological and Environmental Research

May 7-8, 2009
Rockville, Maryland

Pacific Northwest
NATIONAL LABORATORY

BER Mission Needs:

- DOE/BER long-term measure for **Environmental Remediation:**

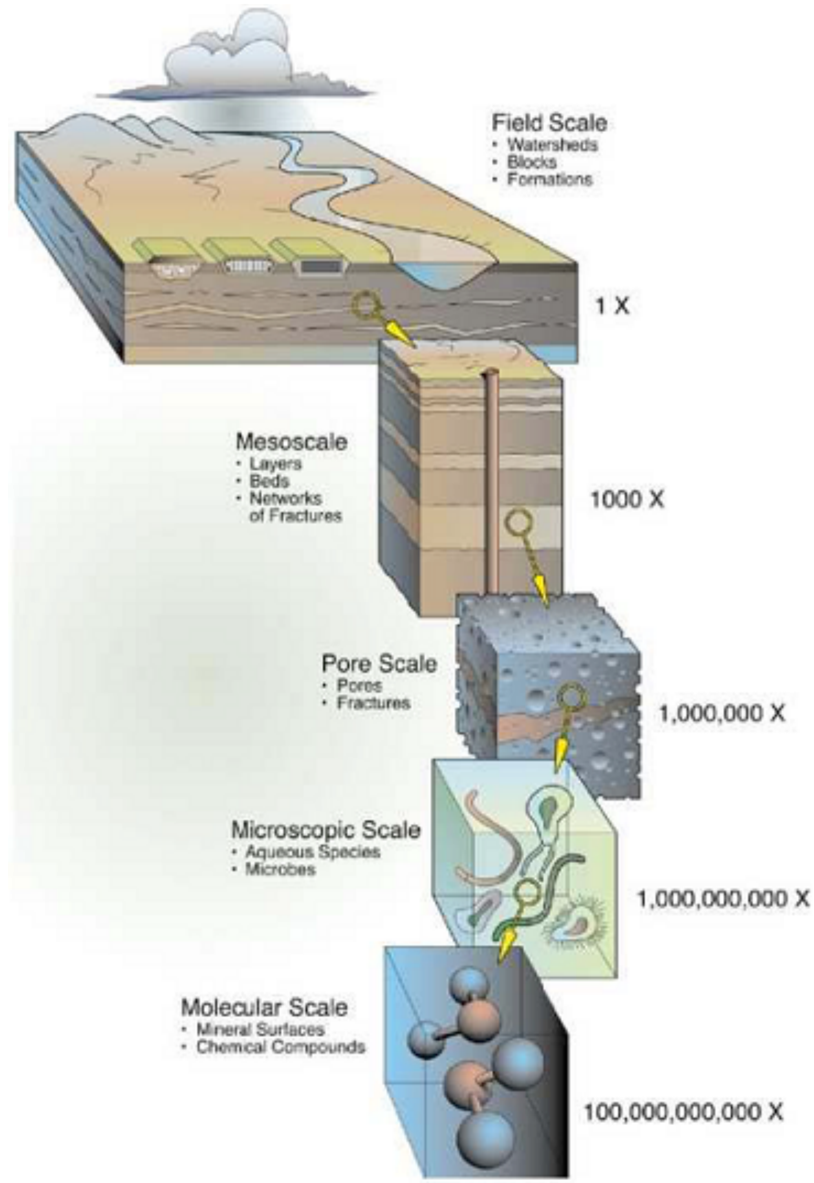
“Provide sufficient scientific understanding such that DOE sites would be able to incorporate coupled physical, chemical and biological processes into decision making for environmental remediation and long-term stewardship.”

Pacific Northwest NATIONAL LABORATORY

Three Critical Science Needs

- **Critical issue #1: Impacts of spatial heterogeneity in the subsurface**
  - Anomalous field-scale transport
  - Discrepancy between field-scale and laboratory parameters
  - Creation and large-scale effects of microenvironments
  - Significantly impacts risk assessment and design of remedial actions → DOE cost and effectiveness

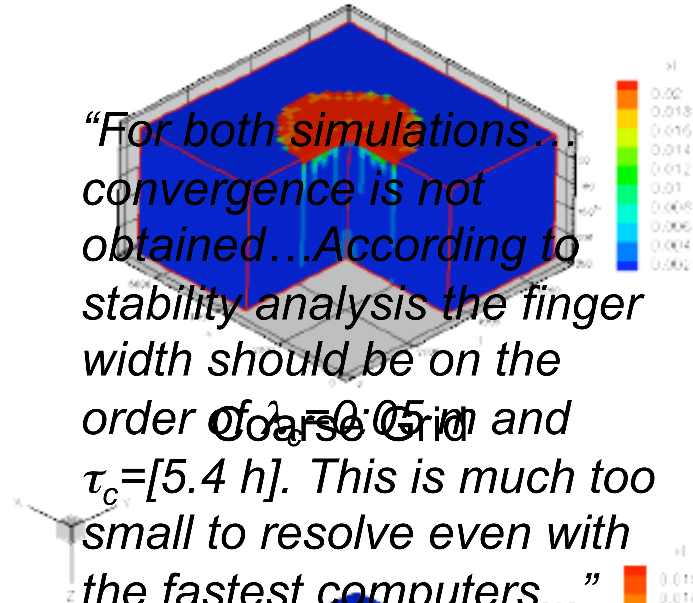
- <u>Need for high-performance computing:</u>
  - High spatial resolution to capture effects of multi-scale heterogeneity



Three Critical Science Needs

- **Critical issue #2: Coupled processes (multi-phase, multi-domain)**
  - Water, air/gas, non-aqueous phase liquids (oils, solvents), supercritical fluids (CO_2)
  - Mineral precipitation / biofilm formation → coupling between transport, flow and reaction
  - Highly localized processes

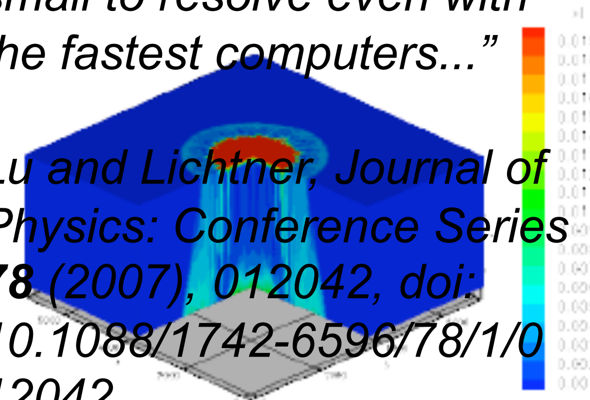
- <u>Need for high-performance computing:</u>
  - High spatial resolution for localized processes
  - Multicomponent chemistry
  - Multiphase fluids (non-linear physics)
  - Multidomain physics



Coarse Grid

"For both simulations… convergence is not obtained…According to stability analysis the finger width should be on the order of λ_c=0.05 m and τ_c=[5.4 h]. This is much too small to resolve even with the fastest computers..."

Lu and Lichtner, Journal of Physics: Conference Series ***78*** *(2007), 012042, doi: 10.1088/1742-6596/78/1/012042*

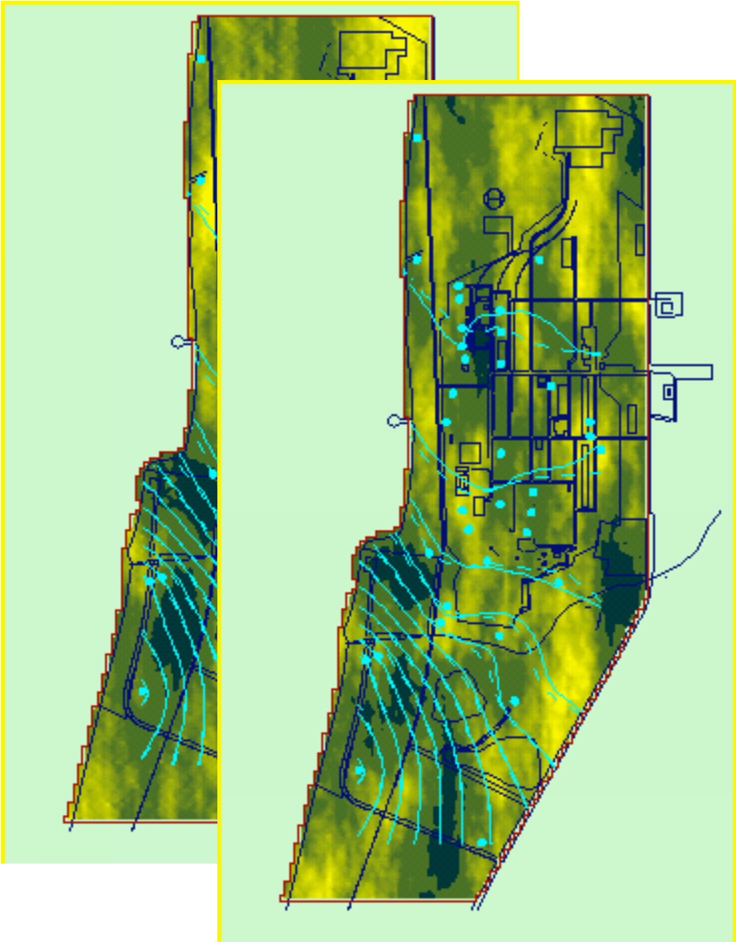


Fine Grid

Three Critical Science Needs

- **Critical issue #3: Uncertainty Quantification**
  - Model parameters are poorly known because of lack of characterization data → predictive uncertainty
  - Parameter estimation (inverse modeling) is non-unique
  - Integrate diverse data types (geophysical, hydrologic, geologic)

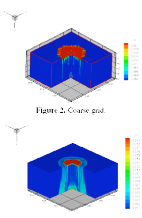
- <u>Need for high-performance computing:</u>
  - Multiple realizations in Monte Carlo stochastic simulation
  - Parameter inversion with thousands to millions of parameters

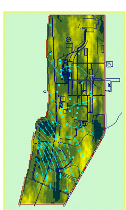


Doherty, J. Groundwater model calibration using pilot points and regularisation, *Ground Water* **41** (2): 170-177, 2003.

Pacific Northwest NATIONAL LABORATORY

Computational Requirements





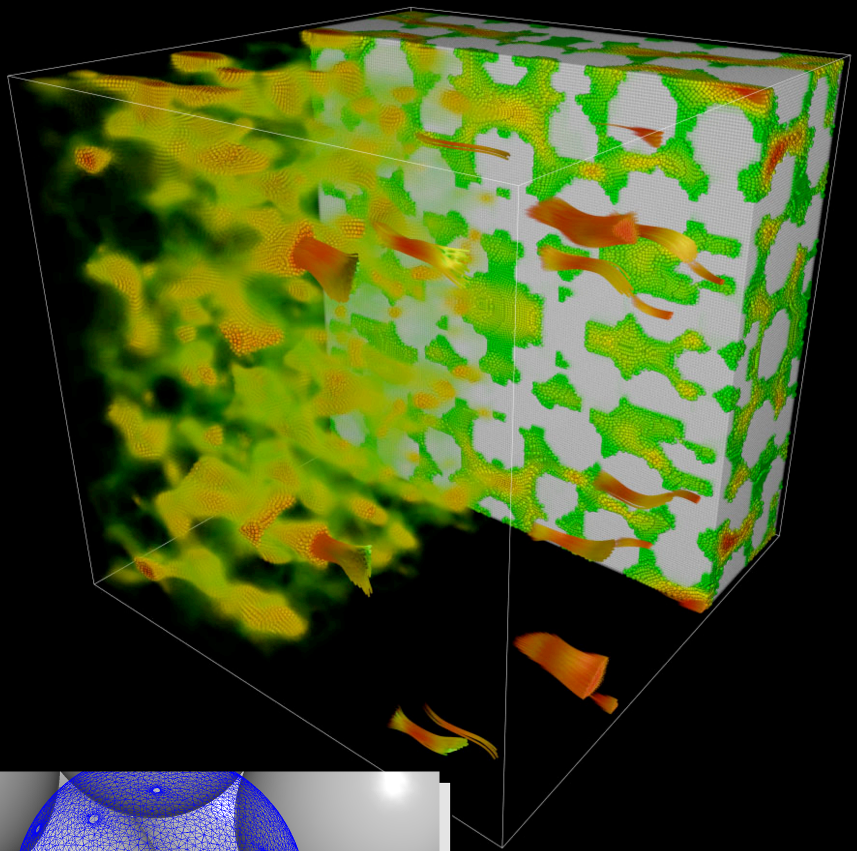
Science Issue	Problem Size	Computational Scale
Spatial heterogeneity +	<u>10 M grid cells</u>	Terascale
Coupled processes +	x 3 phases x 10 doms x 20 comps = <u>1 B DOF</u>	Petascale
Uncertainty quantification / inverse modeling	x 1000 realizations = <u>1 T DOF</u>	Exascale?

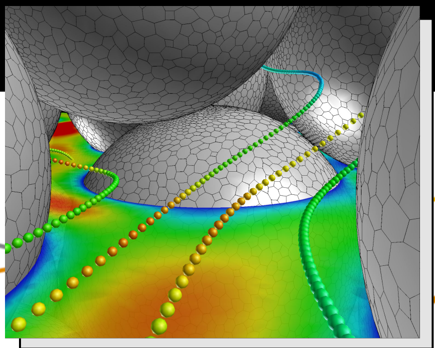
Focus Areas for Computational Research:

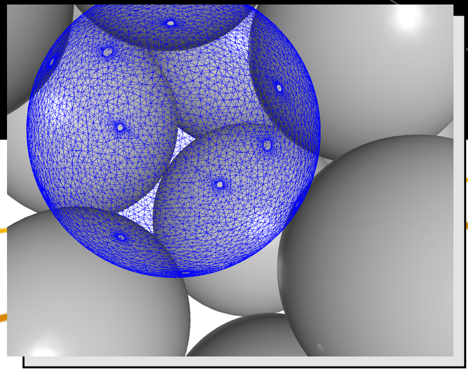
- Pore-Scale Simulation: Incorporate and understand fundamental biogeochemical processes
  - Multicomponent, multidomain, multiphase
- Scale Integration: Use fundamental-scale models to inform larger-scale simulations
  - Upscaling
  - Multiscale Hybrid Modeling
- Field-Scale Simulation: Apply high-end computational tools to simulate complex processes at field sites
  - Data integration / parameter estimation
  - High-resolution simulation
  - Uncertainty quantification

Pore-Scale Simulation

- Microbiological and geochemical processes are controlled by local (pore) environments
- Pore-scale simulation serves as the foundation for fundamentally-sound models at larger scales
- Grid- and particle-based methods implemented on parallel computers



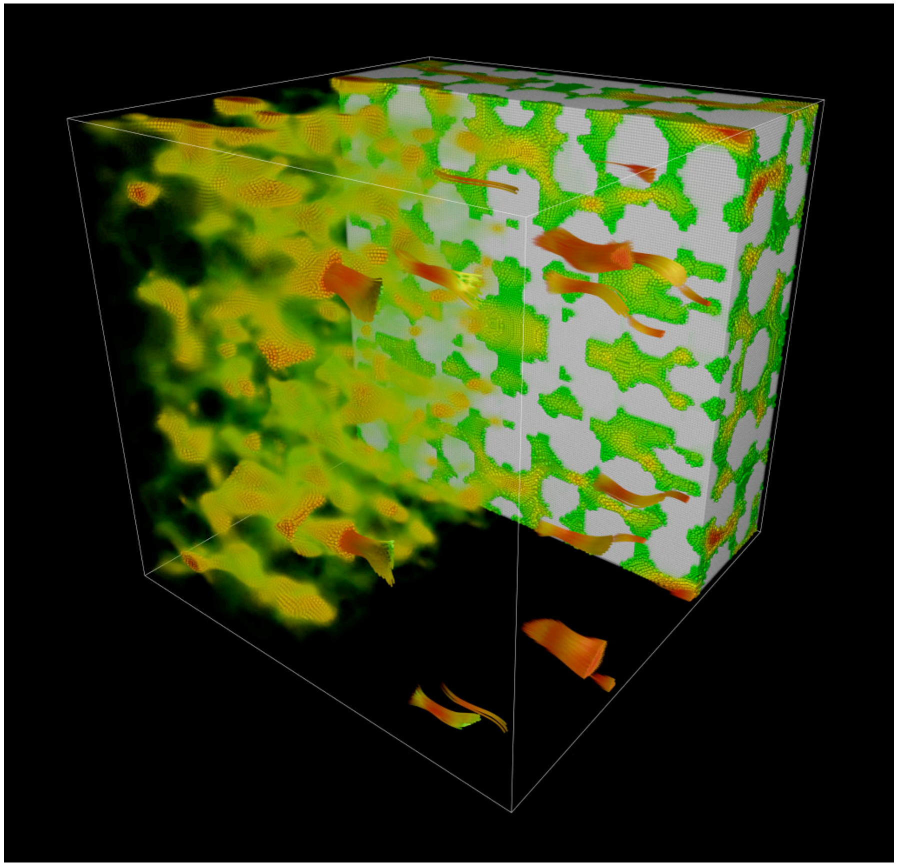




Pacific Northwest
NATIONAL LABORATORY

Pore-Scale Simulation - SPH

- 3D Parallel Code – Smoothed Particle Hydrodynamics
  - SPH is a particle-based (Lagrangean) method first developed for astrophysical applications
  - Our 3D code is built using Common Component Architecture and Global Arrays
  - Simulation shown used seven million computational particles
  - Runs on EMSL “chinook” and NERSC “franklin” supercomputers



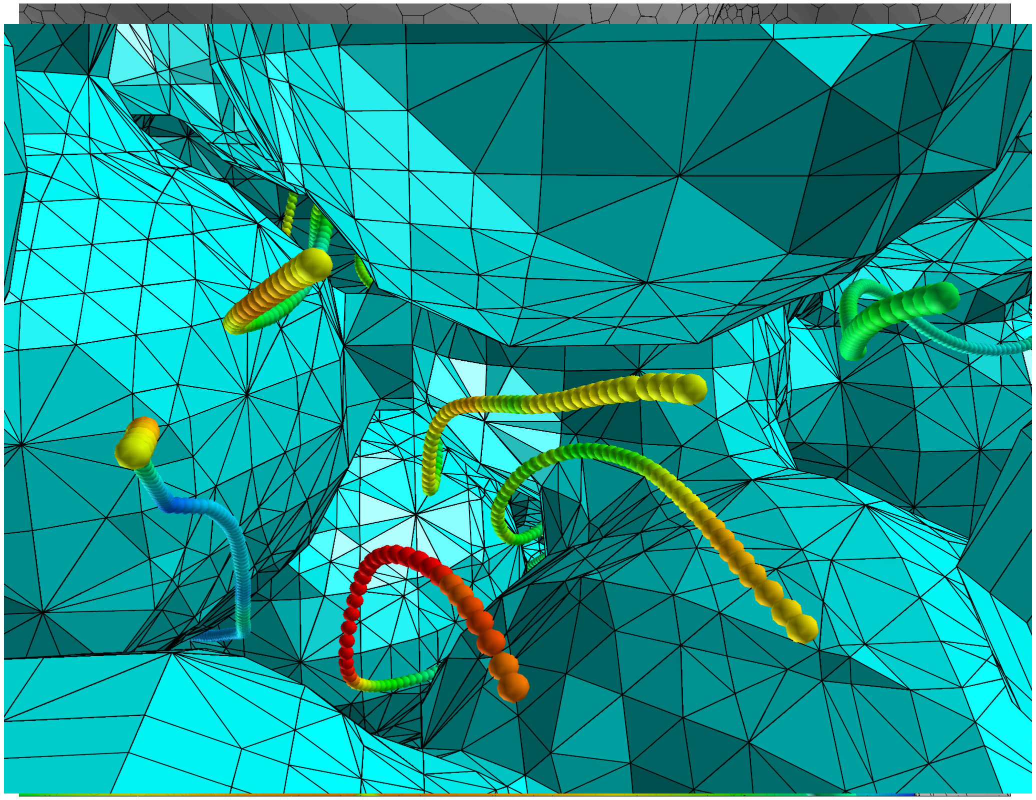
Visualization by Kwan-Liu Ma and Chad Jones of UC Davis, Institute for Ultra-Scale Visualization

Palmer et al., "A Component Based Framework for Smoothed Particle Hydrodynamics Simulations of Reactive Fluid Flow in Porous Media," submitted

Pacific Northwest NATIONAL LABORATORY

Pore-Scale Simulation - CFD

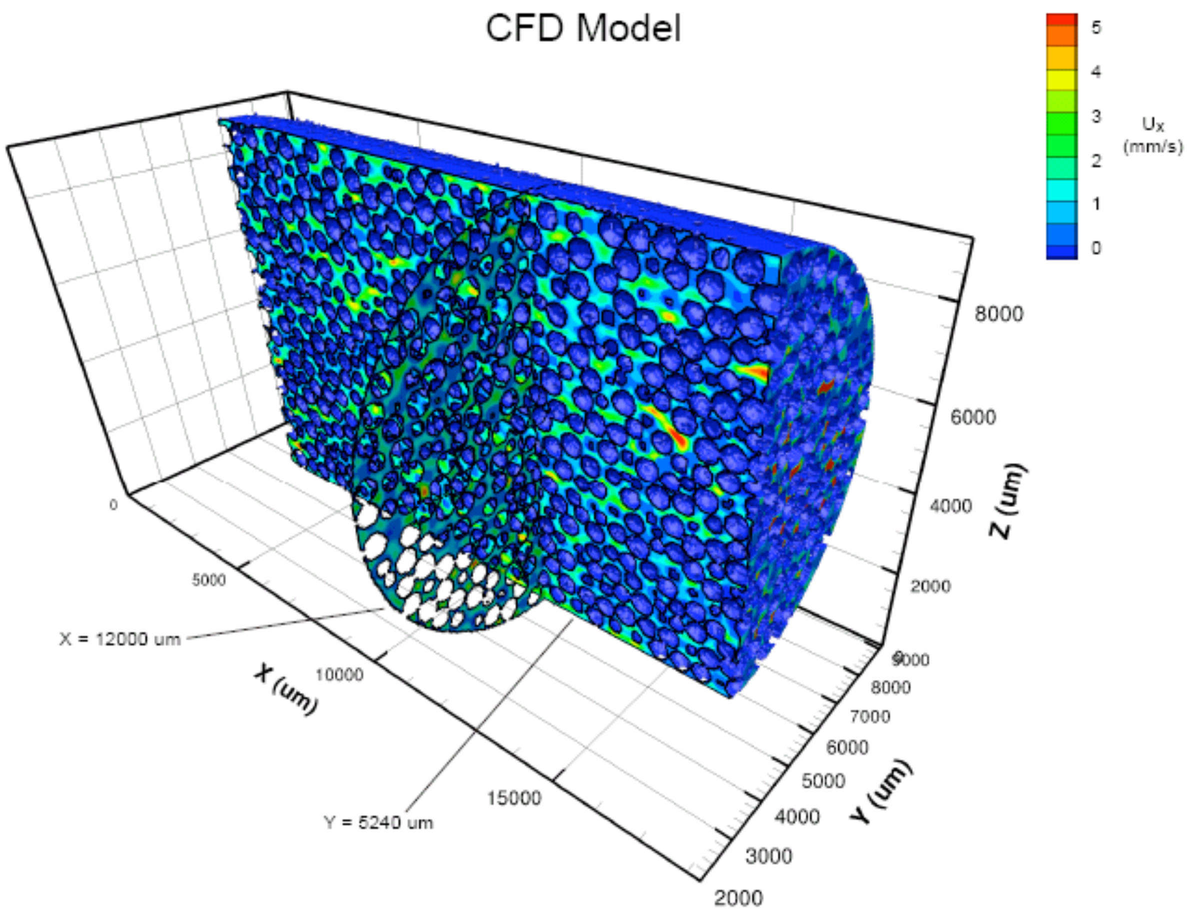
- 3D Parallel Code – Computational Fluid Dynamics
  - Geometry modeling and mesh generation for arbitrary pore geometry
  - PNNL Code – TE^2THYS, runs on Altix cluster, Chinook, and Franklin



Visualization by John Serkowski, PNNL

Relating Pore- and Continuum-Scale Models

- Obvious problem: It is impractical to simulate engineering problems of interest with pore-scale resolution

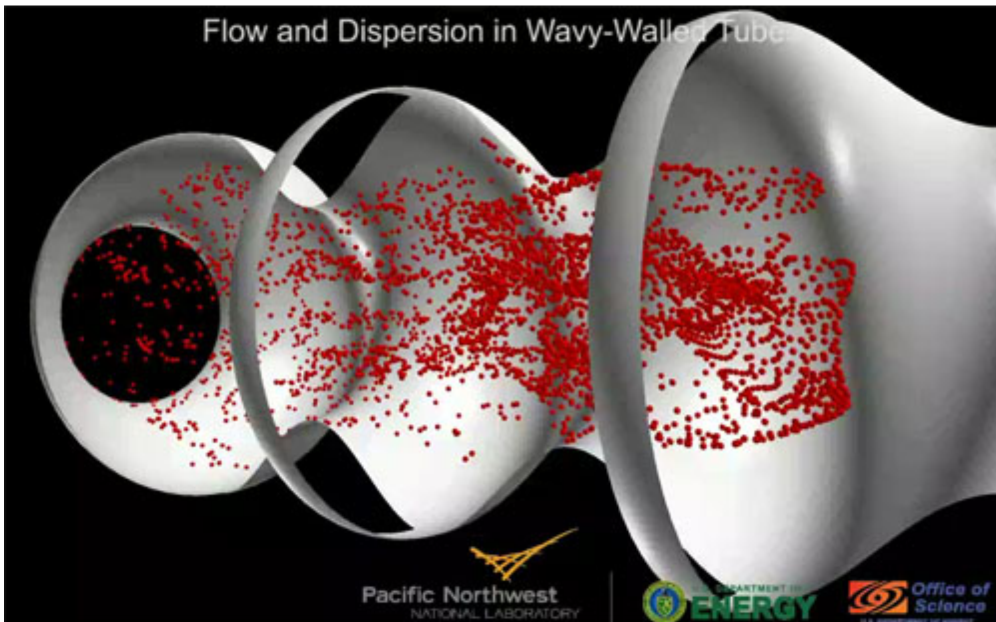


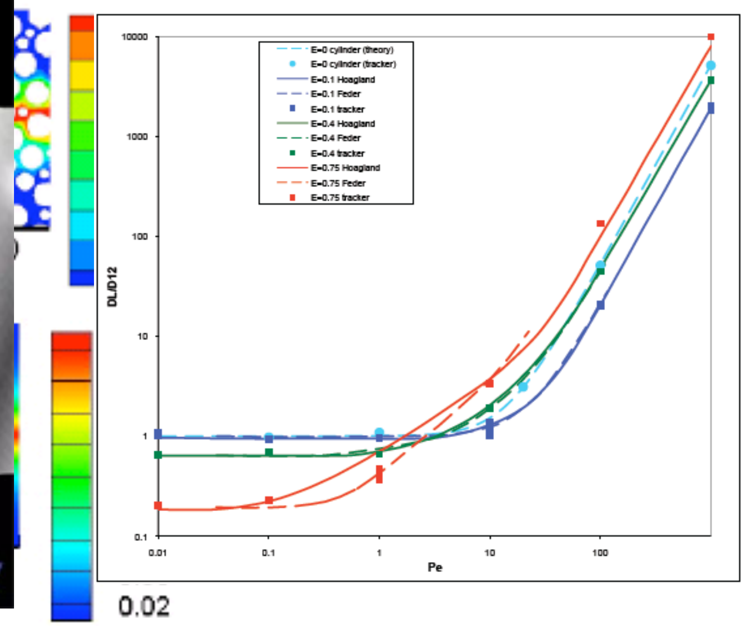
D = 10 mm

L = 15 mm

Upscaling

- Use detailed (pore-scale) information to define continuum-scale models (equations) and parameters
- Example: Dispersion



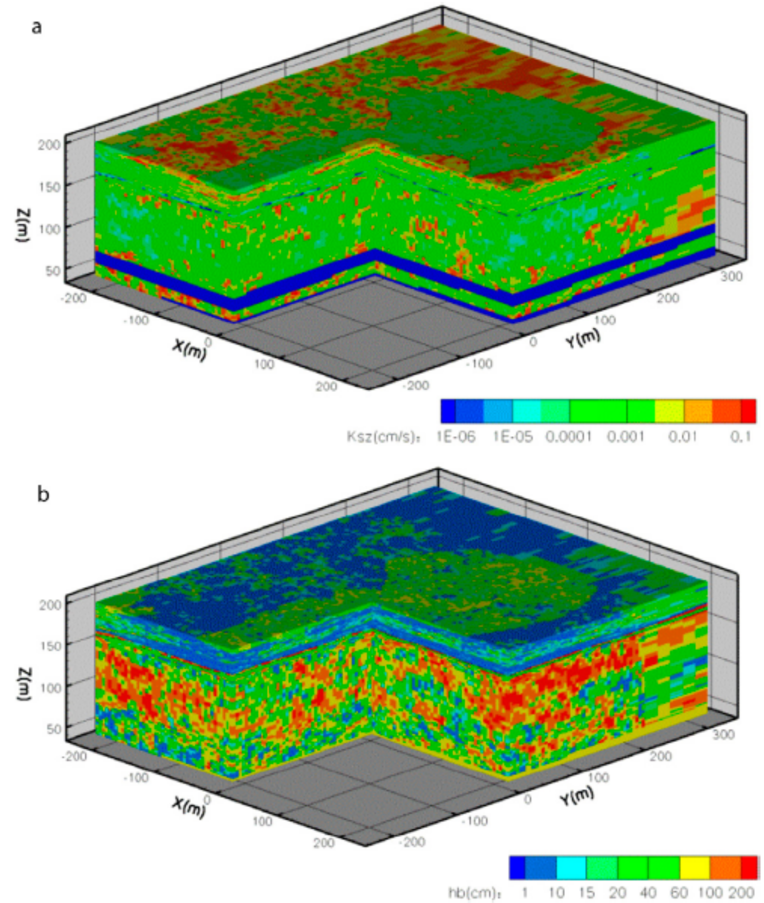


- Other upscaled processes/parameters: diffusive mass transfer, biofilm dynamics, mixing-controlled reactions, surface sorption

Pacific Northwest
NATIONAL LABORATORY

Continuum-Scale Simulation - STOMP

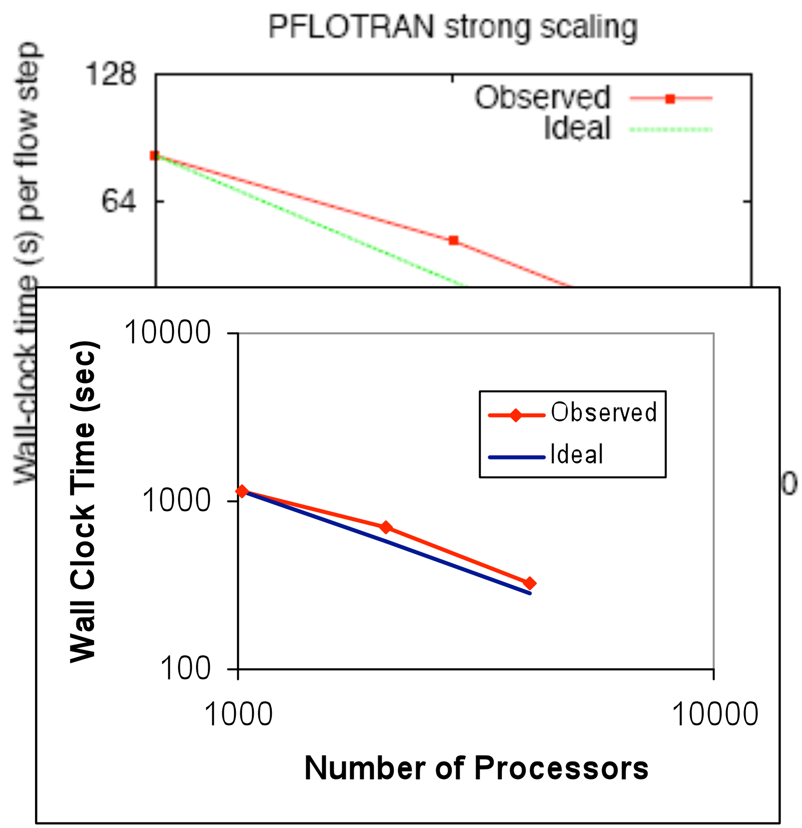
- 3D Parallel Code – Subsurface Transport Over Multiple Phases (STOMP)
  - Current version used for Hanford Site applications has limited scalability
  - Code is being redesigned for
    - enhanced flexibility
      - modularity using CCA
      - alternative grid components using ITAPS
    - scalability
      - Global Arrays
      - TOPS solvers



White et al., *Vadose Zone Journal*, 7:654-666, 2008.

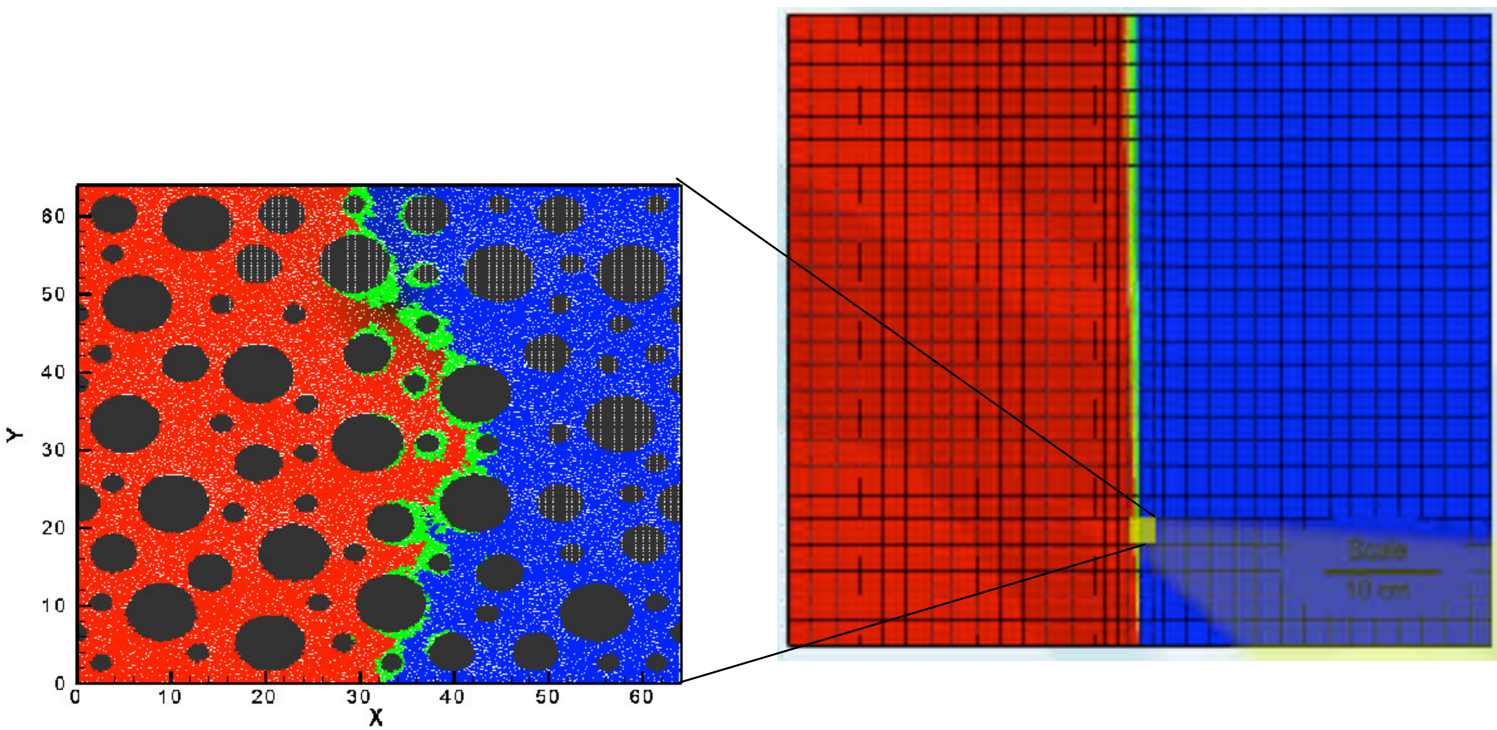
Continuum-Scale Simulation

- 3D Parallel Code – PFLOTRAN
  - Development under SciDAC Science Application (Peter Lichtner, LANL, PI)
  - Based on PETSc framework (solvers and data structures); provides links to other software packages.
  - Relative parallel efficiency of 79% at 12,000 cores (strong-scaling study of 500 million nodes – ORNL Jaguar).
- Proof-of-principle run on a one billion node (4096×2048×128 = 1,073,741,824 nodes) problem. Timings for a single step run on 1024 to 4096 cores are shown.



Hybrid Multiscale

- Couples pore-scale and continuum-scale models in a single simulation

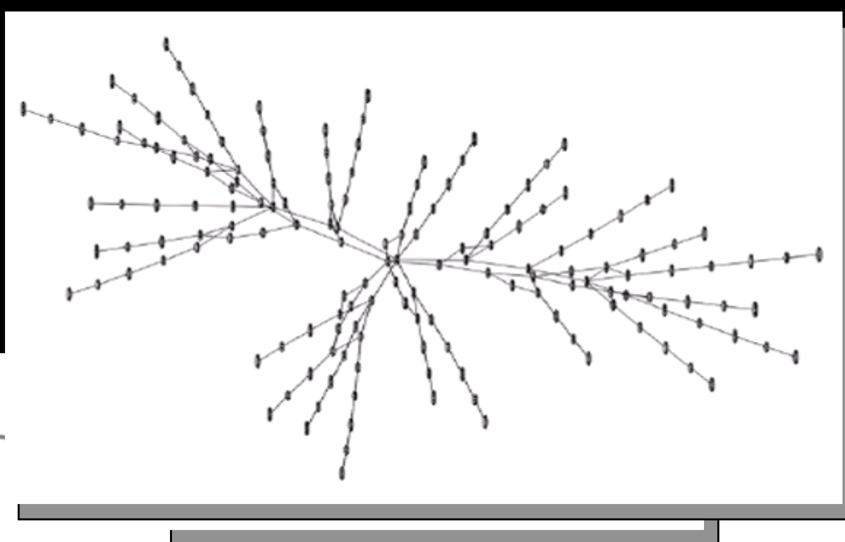


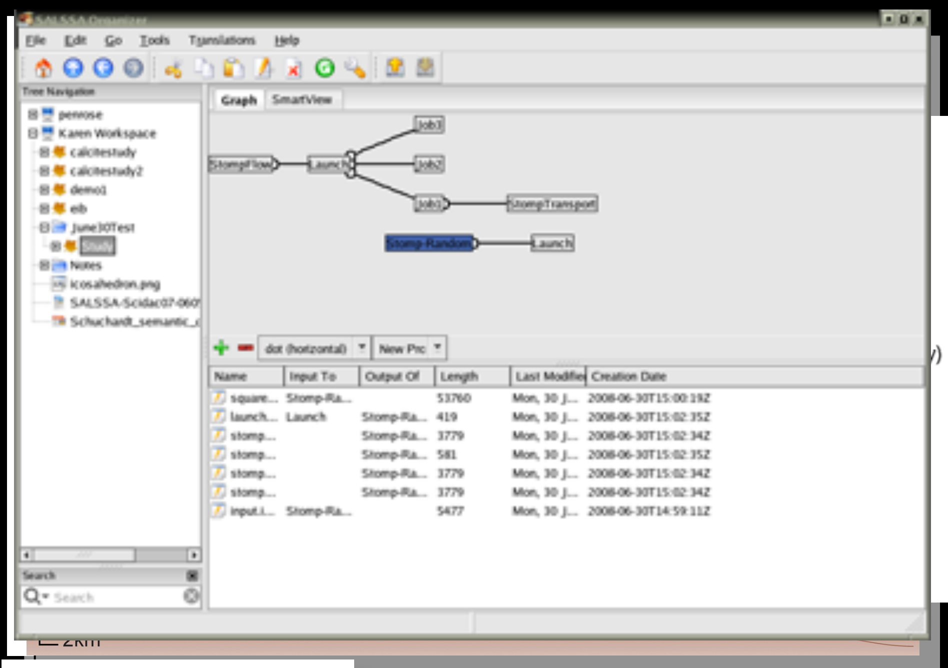
Tartakovsky, A., G. Redden, P. C. Lichtner, T. D. Scheibe, and P. Meakin (2008), Mixing-Induced Precipitation: Experimental Study and Multi-Scale Numerical Analysis, *Water Resources Research*, 44, W06S04, doi: 10.1029/2006WR005725, 2008.

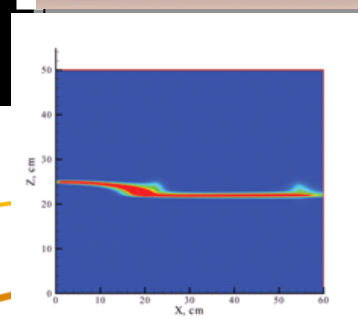
Pacific Northwest
NATIONAL LABORATORY

Field-Scale Applications

- CO2 Sequestration
- Fate, transport, and remediation of metals and radionuclides
- Large-scale heterogeneity modeling
- Data and scientific workflow management







y of the Australian CO_2CRC

Pacific Northwest NATIONAL LABORATORY

1. Projects Overview

- Hybrid Multiscale Modeling (SciDAC)
  - PIs: Tim Scheibe, Bruce Palmer, Karen Schuchardt (PNNL); Daniel Tartakovsky (UCSD); Paul Meakin, George Redden (INL); Scott Brooks (ORNL)
  - FY 2007-2010
  - Couple pore- and continuum-scale models of porous media flow and reactive transport
- PFlotran (SciDAC)
  - PIs: Peter Lichtner (LANL); Glenn Hammond (PNNL); Al Valocchi (UIUC); Richard Mills (ORNL)
  - FY 2007-2011
  - Develop highly-scalable continuum-scale model of porous media flow and reactive transport
- PNNL Science Focus Area / Hanford Integrated Field Research Challenge (ERSP)
  - Integrate molecular to field scale information to predict uranium transport at the Hanford Site 300 Area

2. Current HPC Requirements
(see slide notes)

- Architectures
  - Cray XT4 (NERSC)
  - HP Opteron cluster with Infiniband Network (EMSL)
- Compute/memory load
  - Memory load is not tracked but is not severe
  - 300K processor hours at EMSL, 750 processor hours at NERSC
- Data read/written
  - 1 – 10 GByte per snapshot
- Necessary software, services or infrastructure
  - Petsc and other solver libraries
  - High performance parallel IO libraries
  - Data archive
  - Common Component Architecture (CCA)
  - Global Arrays
  - Support for shared libraries (CCA)
- Current primary codes and their methods or algorithms
  - Smoothed particle hydrodynamics (SPH): Lagrangian algorithm, explicit timestepping
  - TE^2THYS: Finite Volume, particle tracking
  - STOMP, PFLOTRAN: Eulerian algorithms, non-linear implicit solution schemes (phase switching)
- Known limitations/obstacles/bottlenecks
  - Non-scaling solvers (anything less than order N or maybe NlnNis going to kill us)
  - Low performing parallel IO libraries. Performance is substantially less than theoretical peak, even for relatively ideal cases (large contiguous writes)

Pacific Northwest
NATIONAL LABORATORY

3. HPC Usage and Methods for the Next 3-5 Years

(see slide notes)

- Upcoming changes to codes/methods/approaches
  - SPH: running larger problems with concurrent increase in IO and data storage
  - Darcy scale continuum: Not clear, currently under investigation
- Changes to Compute/memory load
  - Expected to reach 10 million processor hours
- Changes to Data read/written
  - Move up to 100-1000 GBytes per snapshot
- Changes to necessary software, services or infrastructure
  - Archive needs to be incorporated into job execution services (e.g. write data to disk and then move it to archive before job execution completes) to eliminate possible flooding of disk
  - Visualization and other analysis tools need to execute in parallel to support very large data sets
- Anticipated limitations/obstacles/bottlenecks on 10K-1000K PE system.
  - No programming models
  - Unable to identify sufficient parallelism
  - Excessive synchronization (?)

4. Summary

- What new science results might be afforded by improvements in NERSC computing hardware, software and services?
  - Detailed simulations of pore scale flow that can be used as the basis for developing more sophisticated upscaling approaches for coarser simulations
  - High resolution, multicomponent, multiphysics simulations of groundwater sites extending over extensive time periods (100s-1000s years) for analyzing contaminated DOE sites
- Recommendations on NERSC architecture, system configuration and the associated service requirements needed for your science
  - Resources dedicated to software (libraries) and operating systems must be commensurate with resources dedicated to hardware
  - High performance parallel IO libraries (capable of delivering a significant fraction of theoretical bandwidth for large contiguous writes)
  - Programming models for multicore architectures
  - Profiling tools that support detailed examination of the behavior of code segments and can organize results in human-understandable form. It will probably become increasingly necessary to do debugging and profiling on very large processor counts

For More Information…

- Tim Scheibe – Pacific Northwest National Laboratory
  - Tim.scheibe@pnl.gov
  - 509-372-6065
- Websites:
  - http://subsurface.pnl.gov/
  - http://software.lanl.gov/pflotran/
  - http://ifchanford.pnl.gov/