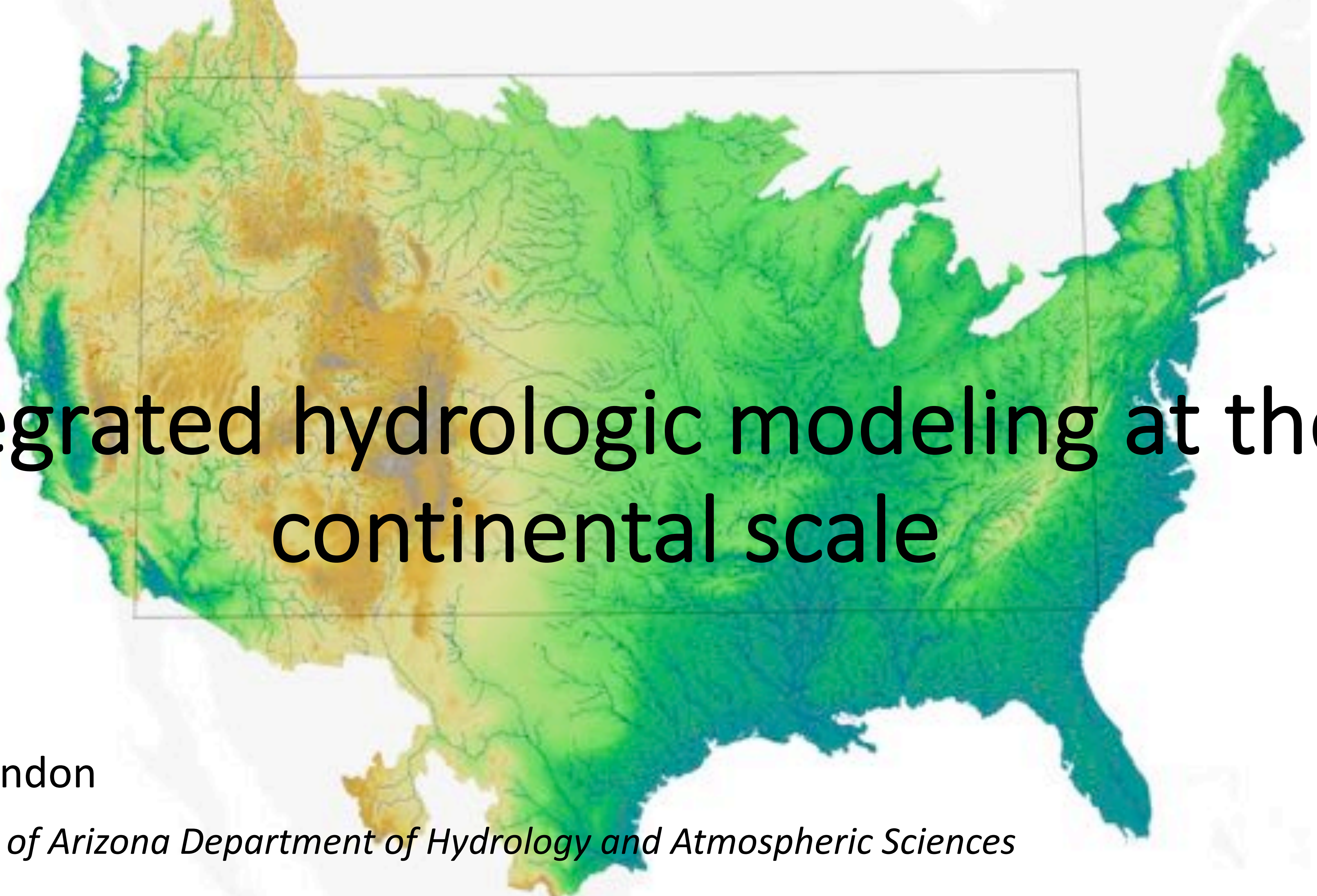
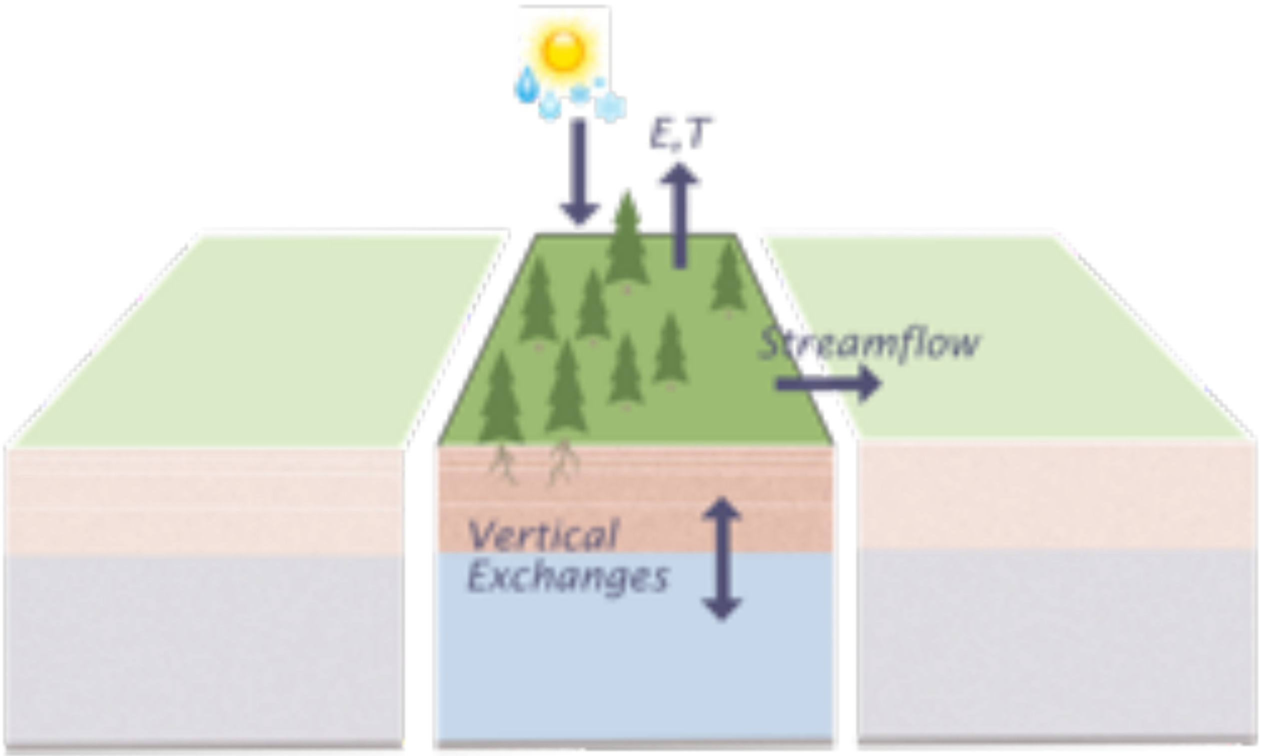


Integrated hydrologic modeling at the continental scale

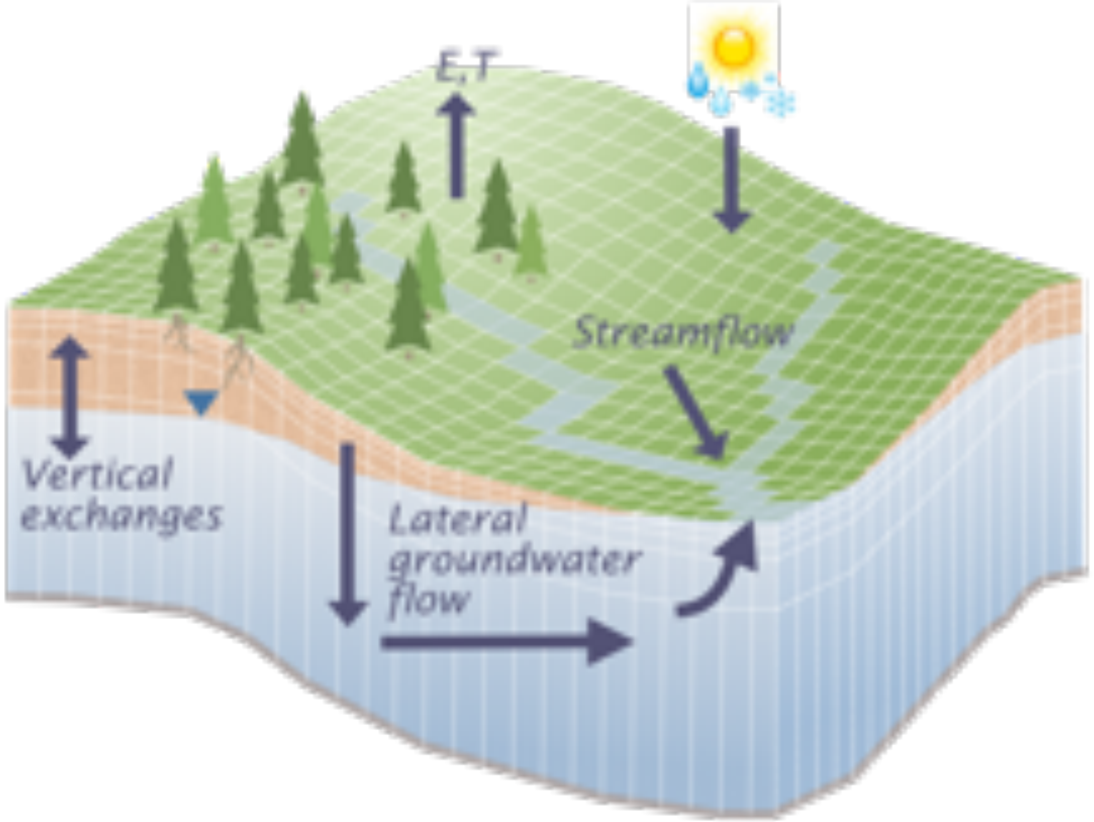
Laura Condon

University of Arizona Department of Hydrology and Atmospheric Sciences

Integrated hydrologic models can be used to explore groundwater surface water interactions



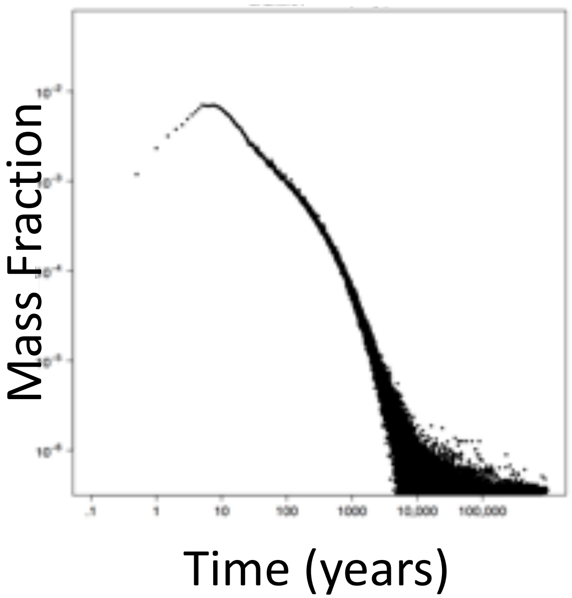
Traditional land surface models **simplify** processes but are **efficient**



Integrated hydrologic models are **comprehensive** but **expensive**

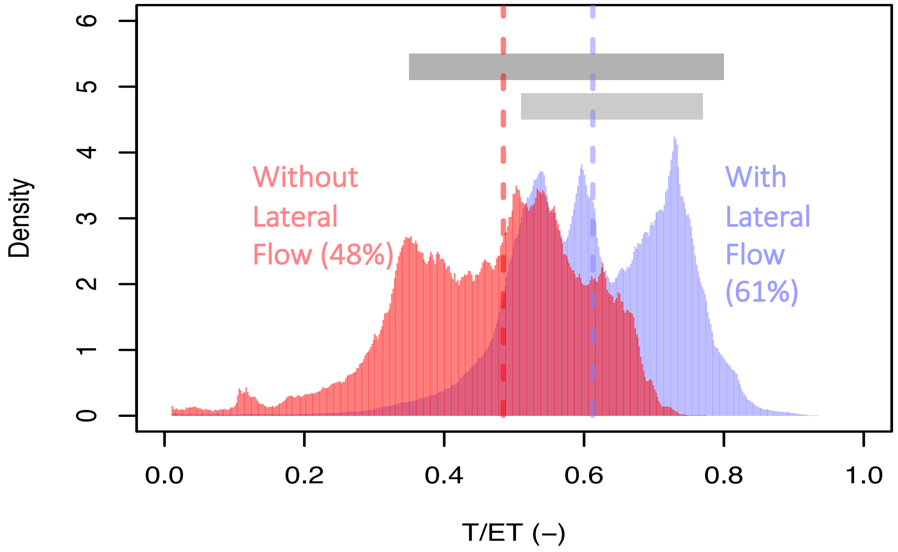
Integrated models as a tool to characterize the role of groundwater in hydrologic systems

Groundwater residence time distributions



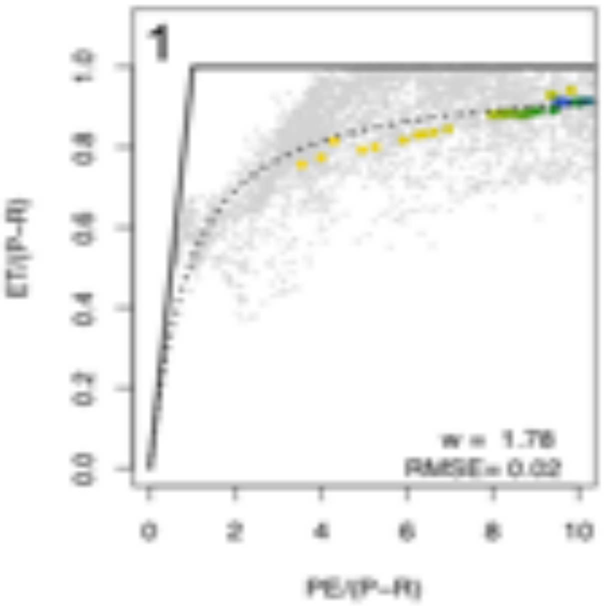
Maxwell, et al. (2016). "The imprint of climate and geology on the residence times of groundwater." Geophysical Research Letters **43**(2): 701-708.

Contributions of lateral subsurface flow to transpiration partitioning

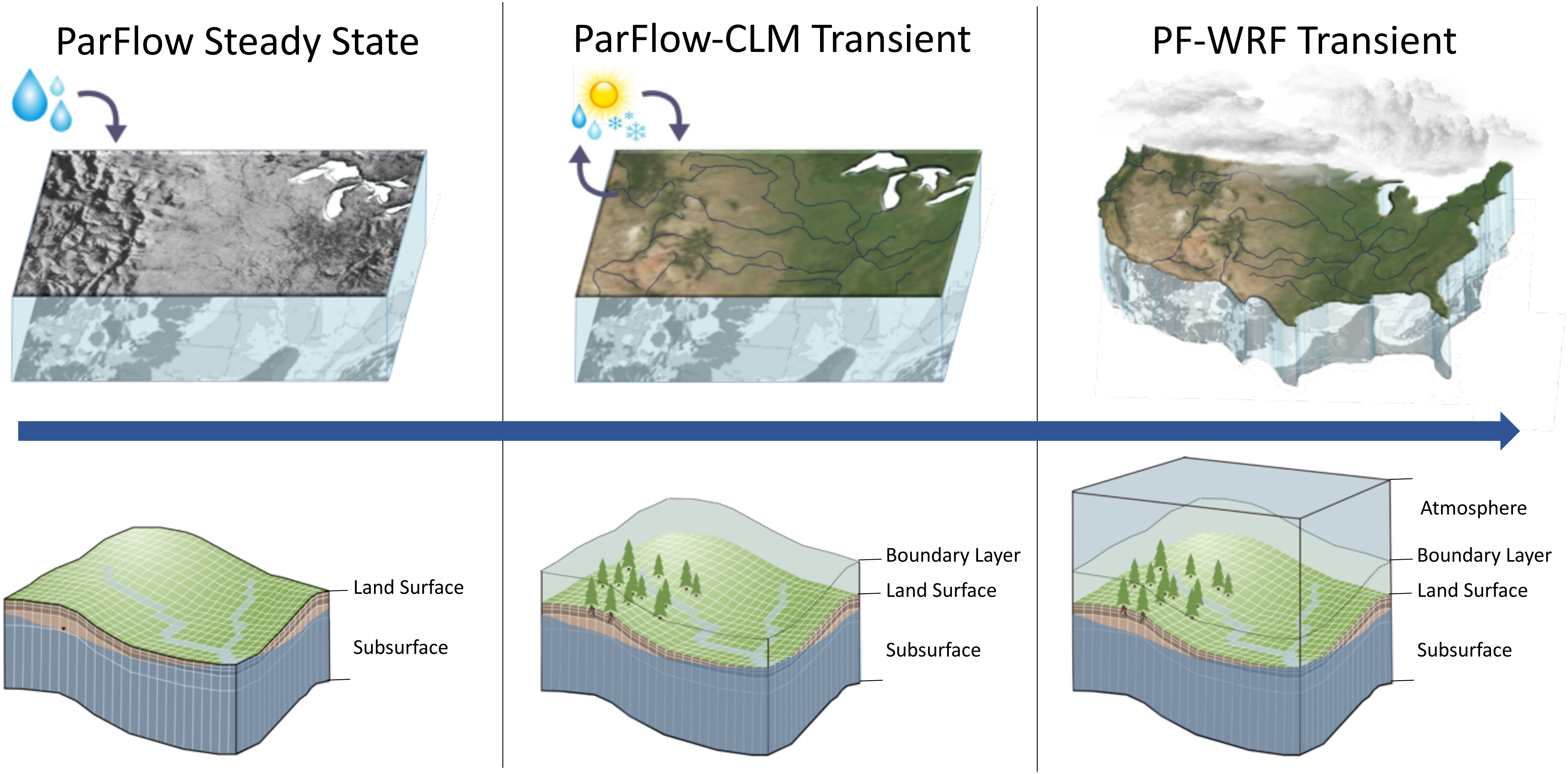


Maxwell and Condon (2016). "Connections between groundwater flow and transpiration partitioning." Science 353(6297): 377-380.

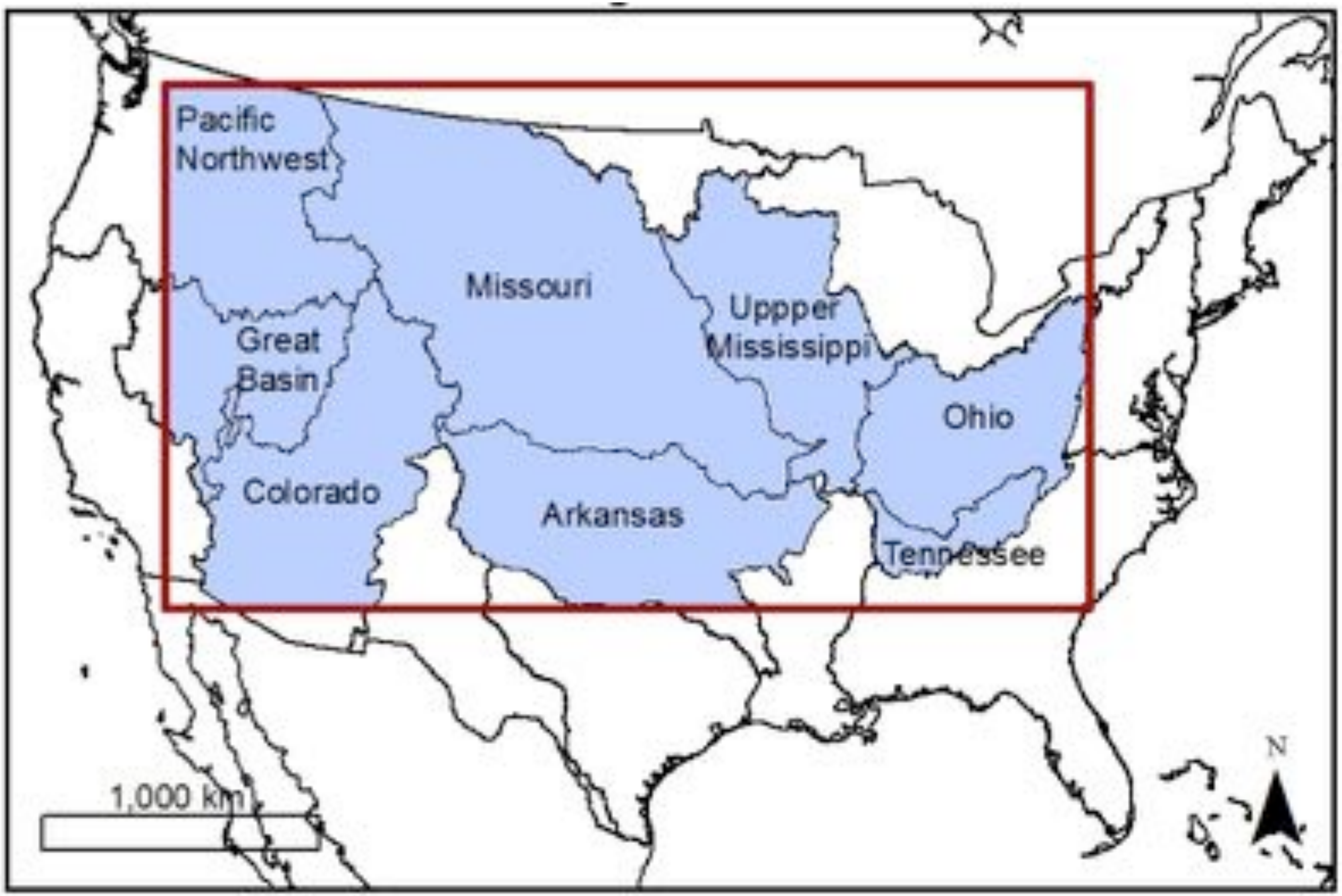
The role of storage changes in surface water partitioning



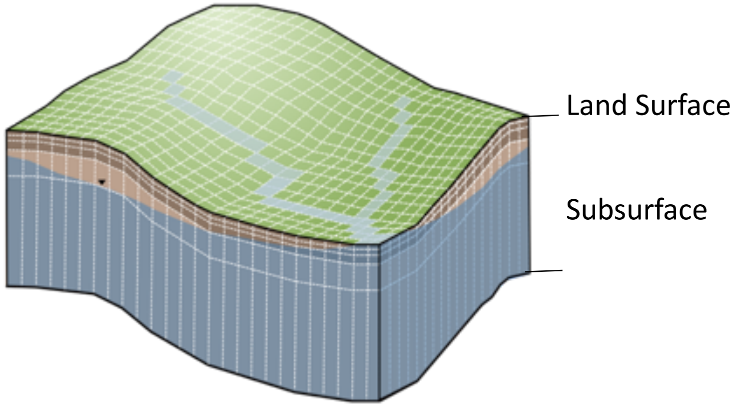
Condon, and Maxwell (2017). "Systematic shifts in Budyko relationships caused by groundwater storage changes." Hydrol. Earth Syst. Sci. **21**(2): 1117-1135.

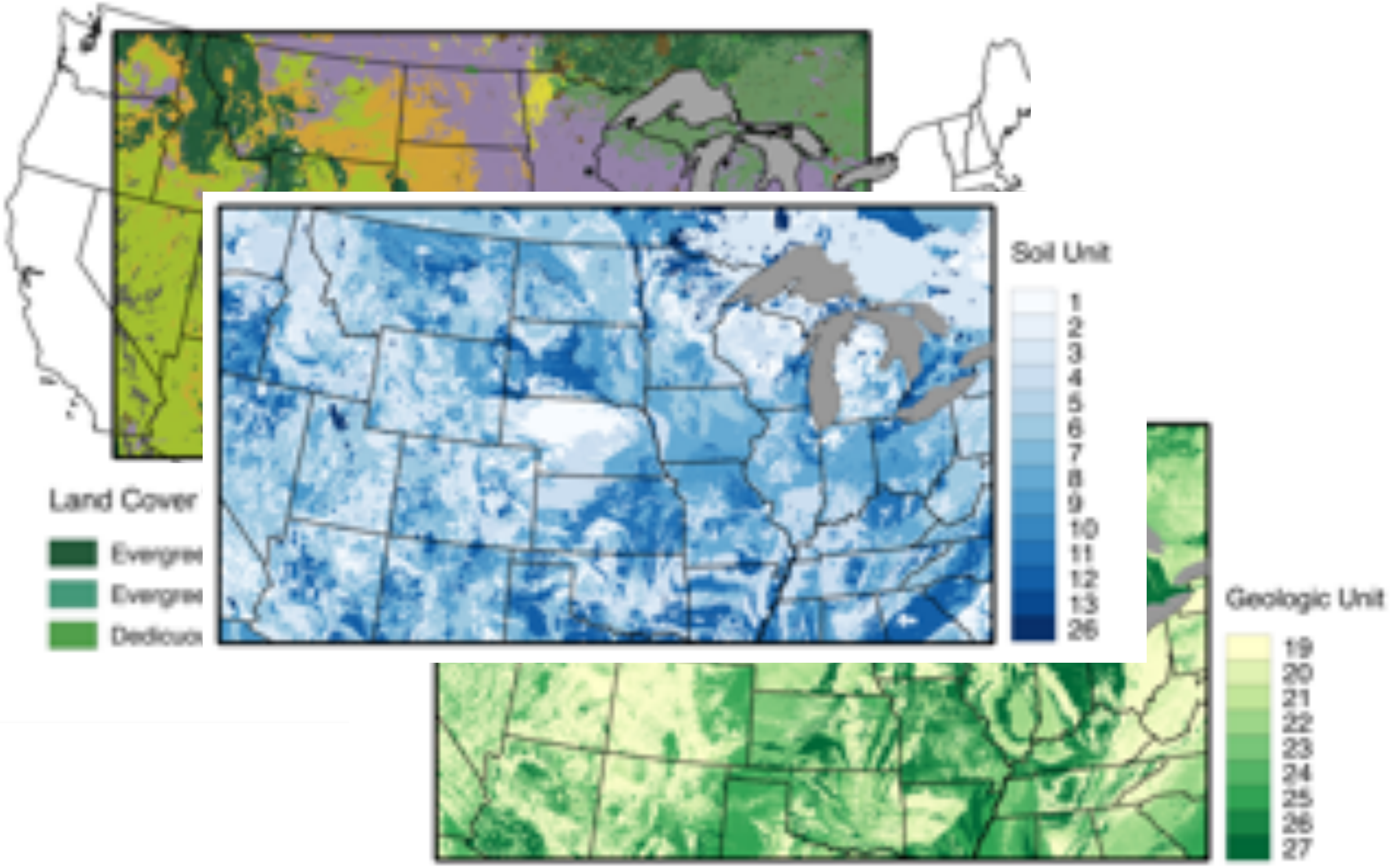


ParFlow Steady State
ParFlow-CLM Transient
PF-WRF Transient
Land Surface
Subsurface
Boundary Layer
Land Surface
Subsurface
Atmosphere
Boundary Layer
Land Surface
Subsurface



Hyper-resolution ($1km^2$) ParFlow-model covering ~6,000,000 km^2 of the continental US





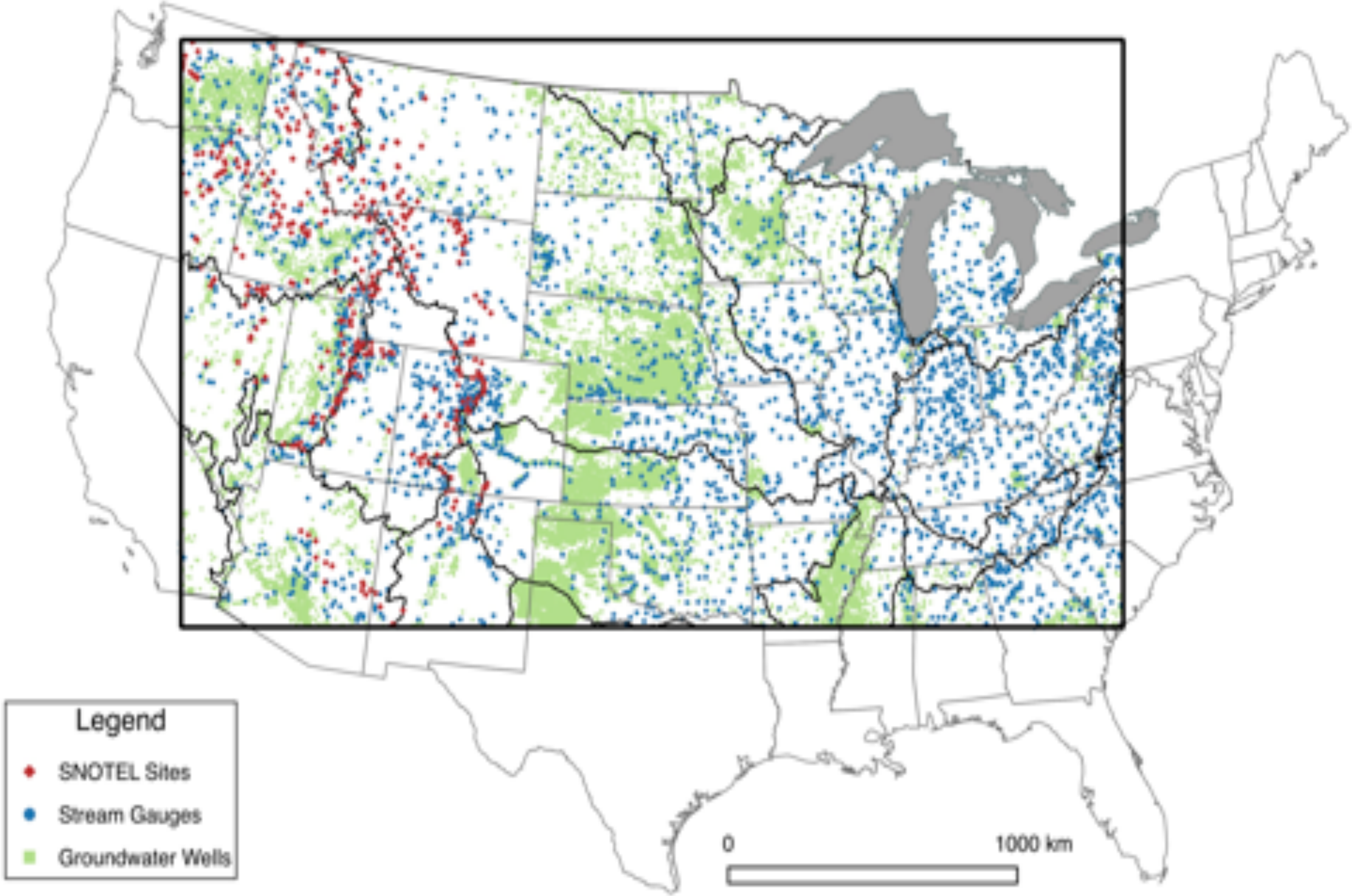
Model Inputs: From national datasets but with customizations

- Topograph
- land cover
- soil properties
- hydraulic conductivity
- atmospheric forcings

This is a data and computationally intensive problem

- 500,000 core hours
- 2,000 cores
- 1 week wall clock
- 3TB input
- 13TB total output
- 200GB summarized output

Model Evaluation

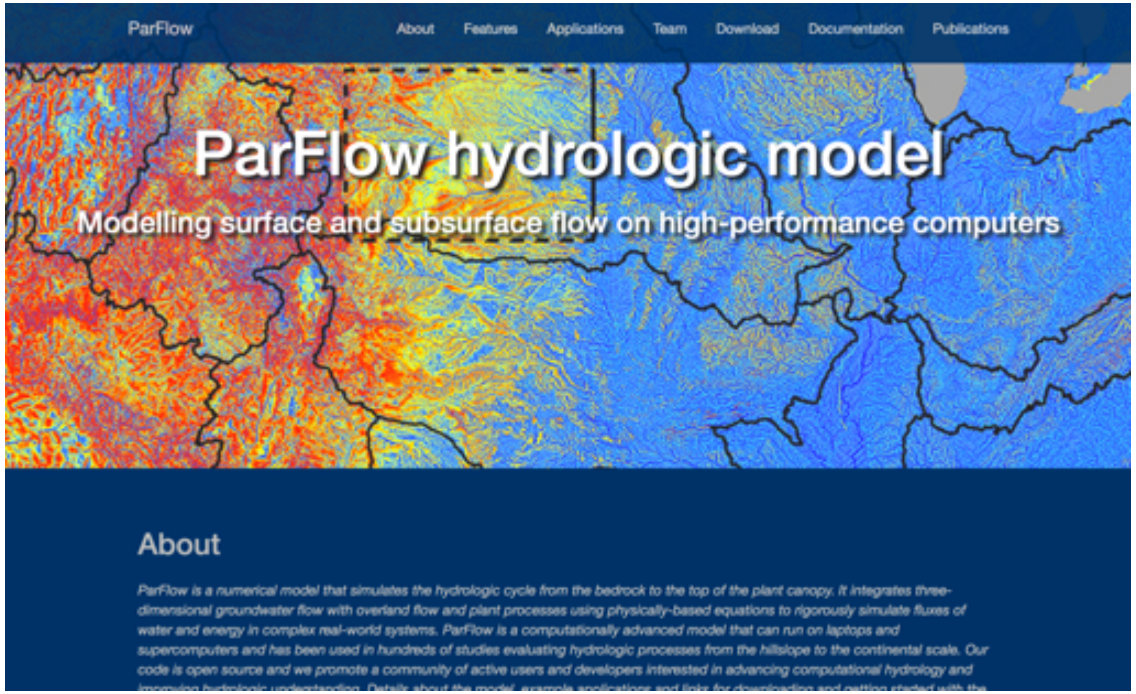


~1.2 million observations available for a one-year simulation

- 378 SNOTEL Stations
- 3,050 USGS gages
- 29,385 USGS Wells
- Varying temporal resolution

The ParFlow project has a diverse user community and distributed development

- 200 users from academia, research and industry with a range of **backgrounds, HPC** and **modeling** experience
- **Ten developers** from academia and national laboratories worldwide
- Range of priorities with no **centralized funding**
- Open source, community development with **best software practices**



https://www.parflow.org

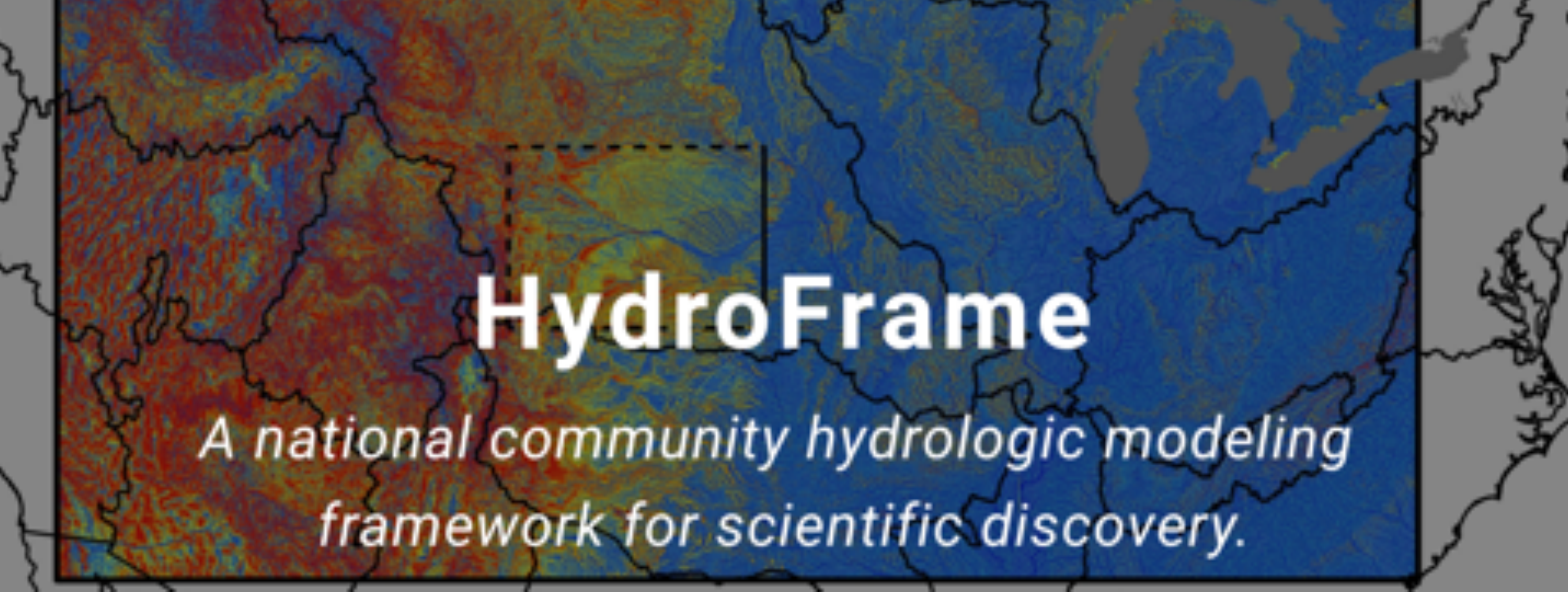
IDEAS-Watersheds

Interoperable Design of Extreme-scale Application Software (IDEAS)

U.S. DEPARTMENT OF ENERGY

Office of Science

Argonne NATIONAL LABORATORY · BERKELEY LAB · Lawrence Livermore National Laboratory · Los Alamos NATIONAL LABORATORY EST.1943 · OAK RIDGE National Laboratory · Pacific Northwest NATIONAL LABORATORY · ARIZONA · NCAR · SCHOOL OF MINES 1874 COLORADO



HydroFrame

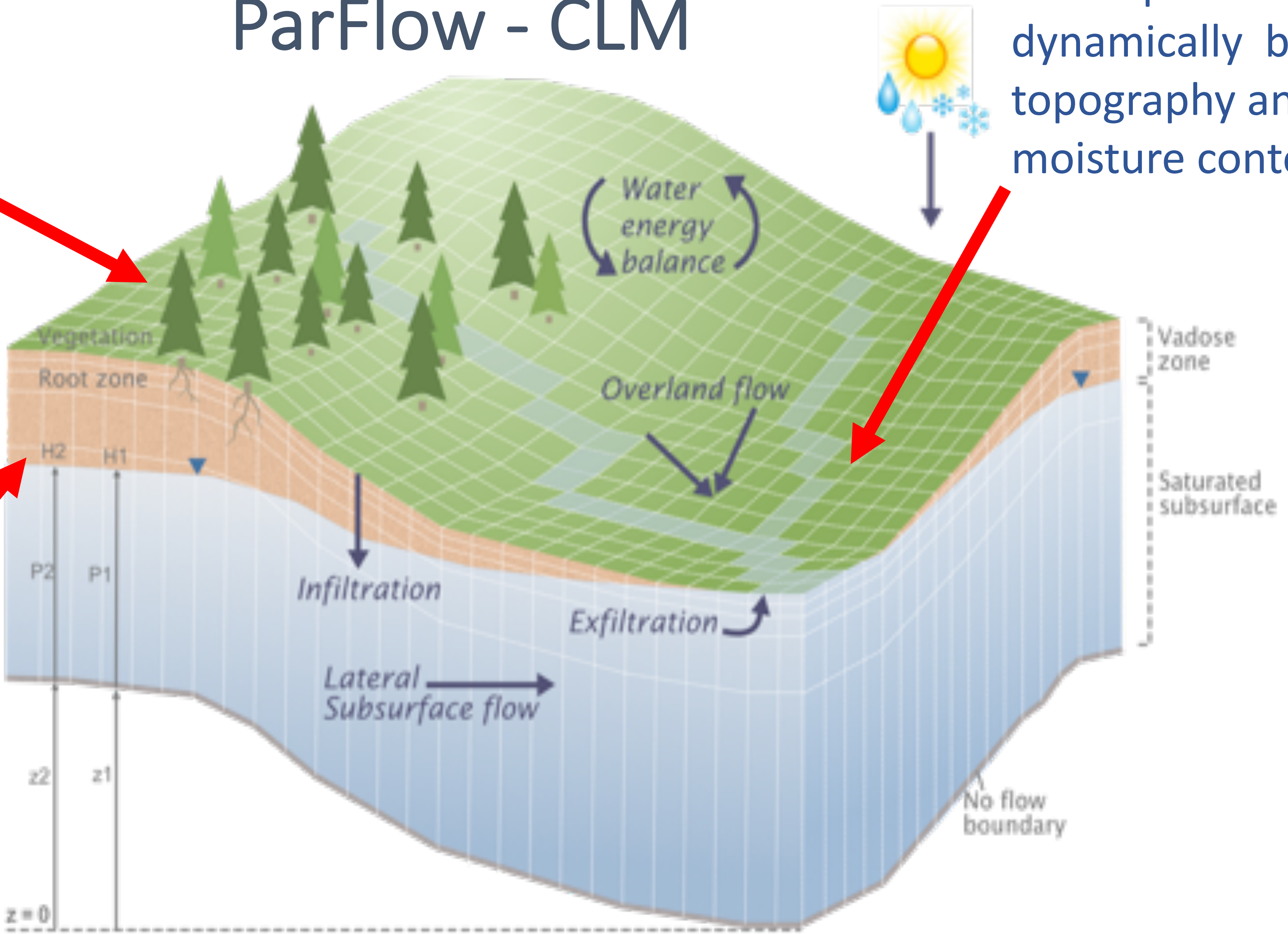
A national community hydrologic modeling framework for scientific discovery.

NSF
COLORADO SCHOOL OF MINES
THE UNIVERSITY OF ARIZONA
BOISE STATE
University of Colorado Boulder
NCAR
CUAHSI
SDSC SAN DIEGO SUPERCOMPUTER CENTER
UtahStateUniversity

ParFlow - CLM

Land energy fluxes controlled by soil moisture and atmospheric variables

Surface water bodies develop and evolve dynamically based on topography and moisture content



Variably saturated flow is solved throughout the subsurface (i.e. no a priori specification of vadose zone)

Simulation / Experiment Descriptors		Simulation / Experiment Descriptor Classes			
		SAVE MORE OUTPUT		SAVE LESS OUTPUT	
Descriptor	**Descriptor definition**	**Class 1**	**Class 2**	**Class 3**	**Theme**
Model/Code Availability/Ease of use	How accessible is this particular version of the model/code? How often does the code version change? Ease of software installation, setup, etc. IP barriers, embargo periods for new model development?	Difficult to acquire & manage	Model is shareable, but specific changes were implemented that make it unique.	Validated version of a highly accessible model was used/. Easy to install and run on many environments	Accessibility
Platform/System Availability	How specialized the platform needed is (particular hardware, compilers, source code needed)	Requires resources that are more difficult to get access to. Could be scale of resources or type. E.g. general desktop computing vs specific HPC.		Does not require special hardware resources to run	Accessibility
Human Effort	Person-hours required to reproduce dataset	Significant time & expertise required to replicate simulation. Likely will require contact with & guidance from original data producer(s).		Trivial effort required to replicate simulation for most end users.	Accessibility
Simulation Inputs	How much effort is it to get and manage all the inputs used by the simulation?	If inputs are difficult to acquire & manage, retaining output lowers burden for others who might want to re-run model or use outputs.		Easy to acquire & manage	Accessibility

Simulation / Experiment Descriptors		Simulation / Experiment Descriptor Classes			
		SAVE MORE OUTPUT		SAVE LESS OUTPUT	
Descriptor	**Descriptor definition**	**Class 1**	**Class 2**	**Class 3**	**Theme**
Computational Cost	The economic cost (combination of run time and computer access costs) of completing the simulations	High computational cost and can only be produced with specialized platforms	Moderate computational cost, but access to needed platforms straightforward	Small computational cost with no special platform needs	Cost
Storage needs/costs	The volume of output that is actually generated by the model experiment or simulation.	Would be inexpensive to save the complete model output		Expensive storage can put a cap on how much data are saved	Cost
Data transfer cost	Limitations on transferring data	If you can use subsetting tools to reduce transfer cost		No subsetting available, would need to transfer in large chunks	Cost
Archiving/Curation Cost	The economic cost of archiving the simulations - who will pay for it now and in the future? And for how long? Is there the availability of a budget, storage space, repo, etc. Willingness and means to curate, maintain, and migrate as needed, now and into the future. This includes the availability of a suitable repository within budget	If willingness and means exist, keeping more output is appropriate. Good organzation and control reduces human resource cost.	If willingness, but fewer means. (Potentially keeping a documented workflow, notebooks and code, and subsets of data)	If no willingness and means, there is less value in keeping data.	Cost
Feature Reproducibility	The ability to reproduce specific (atmospheric) features (of given scale) within an acceptable statistical range of error.	Would be difficult to reproduce due to nonlinearity of phenomena being studied	Would be difficult to reproduce some feature details, but general findings are robust	No issues with reproducibility (could be due to study subject or to model packaging, e.g. containerization)	Reproducibility

Questions?

lecondon@arizona.edu