Land Model / Biogeochemistry Working Group Meeting 2023

Improving the hydrological performance of CTSM through parameter optimization and large-sample watershed modeling

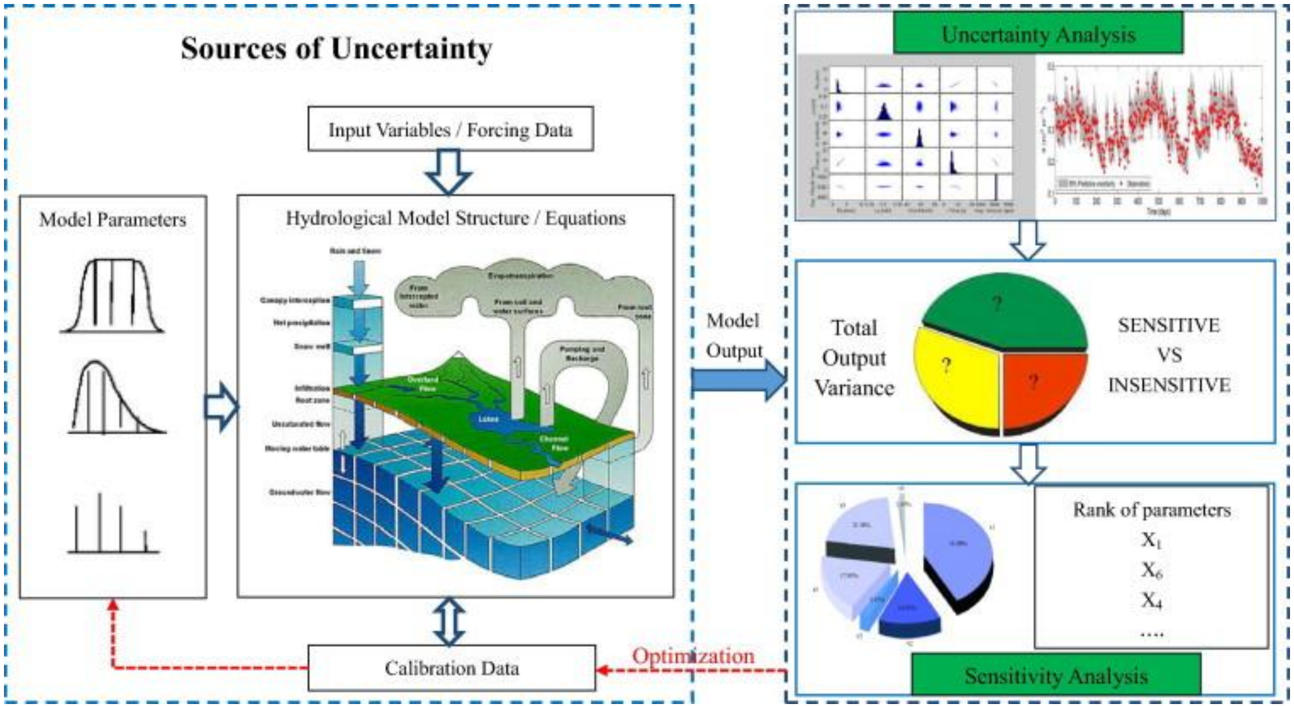
Guoqiang Tang, Andy Wood, Sean Swenson
CGD/TSS, NCAR

Feb 6, 2023

NCAR
UCAR

Using a model to represent real-world hydrology

- Model representations require choices of model structure and physics (parameterizations) and depend on specification of inputs: forcings and parameter values.
- These modeling choices and input specifications are inherently uncertain … *a long-standing challenge*



Song et al, 2015

Different communities approach this challenge differently

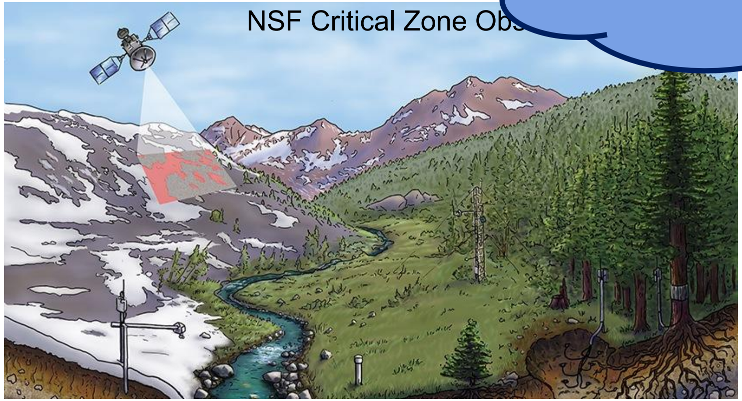
The **earth system modeling** community advances model *fidelity* through:

- improving observational analyses to better estimate inputs (forcings, parameters, structure)
- using observational studies to refine parameterizations (physics)
- increasing complexity of process representation
- **philosophy**: *we can eventually re*
everything through better obser
improvement

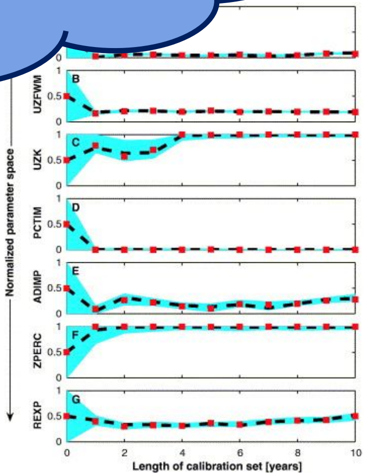
The **applied hydrological modeling** community advances model *performance* through:

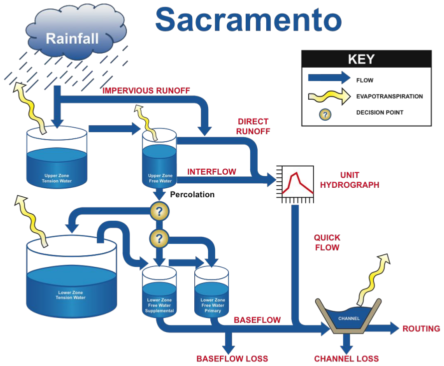
- improving observational analyses to better estimate inputs (forcings, parameters)
- optimization of model parameters and structure given the existing parameterizations
- rovement of physics -- but only as needed
- *we will never represent everything but*
timize what we don't know

This project brings an applied hydrology mindset to the ESM development context



https://www.nsf.gov/news/newsmedia/ENV-discoveries/CZO-discovery-series.jsp





Applied ESM-based modeling seeks both realism *AND* performance

A number of new water security related projects are exploring the use of CTSM as a process/physics advance over more common 'applied-hydrology' models

- climate change studies – land modeling uncertainty is a key component (Lehner et al., 2019)
- flood, drought, and hydrologic prediction applications – supporting water management agency missions

This presentation describes initial work supporting a climate change project sponsored by USACE*

Overarching goal

- develop land models that can represent current hydrology (performance) as well as climate change impacts on hydrology (fidelity) in both coupled and offline context

Immediate goal

- develop CTSM configurations and parameter sets that perform well for hydrology – and with robust climate-hydrology sensitivities

First steps

- use common parameter estimation approaches from applied hydrological modeling for CTSM
- develop a large-sample small-watershed CTSM implementation for investigating parameter estimation and configuration strategies (US-focused, for now)

**US Army Corps of Engineers (USACE) – Climate Preparedness and Resilience Program*

Hydrologic model parameter estimation

- A decades-old practice in applied hydrology with many algorithms and much theory (geo-informatics)
- Now there are multiple available multi-method packages for parameter sensitivity assessment and optimization … as well as individual researcher's methods – e.g.:

Mathematical and statistical methods to help scientists and engineers assess and improve the accuracy of computational models

DAKOTA

Explore and predict with confidence

https://dakota.sandia.gov/

OSTRICH - Optimization Software Toolkit

v17

OSTRICH, developed by L. Shawn Matott, is a model-independent multi-algorithm paralell-friendly optimization and parameter estimation tool that implements numerous model-independent optmization and calibration (parameter estimation) algorithms,

http://www.civil.uwaterloo.ca/envmodelling/Ostrich.html

MO-ASMO

Water Resources Research

Research Article | Open Access

Multiobjective adaptive surrogate modeling-based optimization for parameter estimation of large, complex geophysical models

Wei Gong, Qingyun Duan, Jianduo Li, Chen Wang, Zhenhua Di, Aizhong Ye, Chiyuan Miao, Yongjiu Dai

https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2015WR018230

Welcome to SPOTPY

SPOTPY

A Statistical Parameter Optimization Tool for Python

https://spotpy.readthedocs.io/en/latest/

PEST

Model Independent Parameter Estimation & Uncertainty Analysis

https://pesthomepage.org/

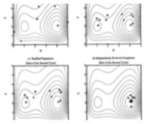
e.g., MCMC

Review | Open Access | Published: 29 May 2020

A survey of Monte Carlo methods for parameter estimation

David Luengo, Luca Martino, Mónica Bugallo, Víctor Elvira & Simo Särkkä

EURASIP Journal on Advances in Signal Processing **2020**, Article number: 25 (2020) | Cite this article



Shuffled Complex Evolution (SCE-UA) Method

Version 1.0.0.0 (420 KB) by Qingyun Duan

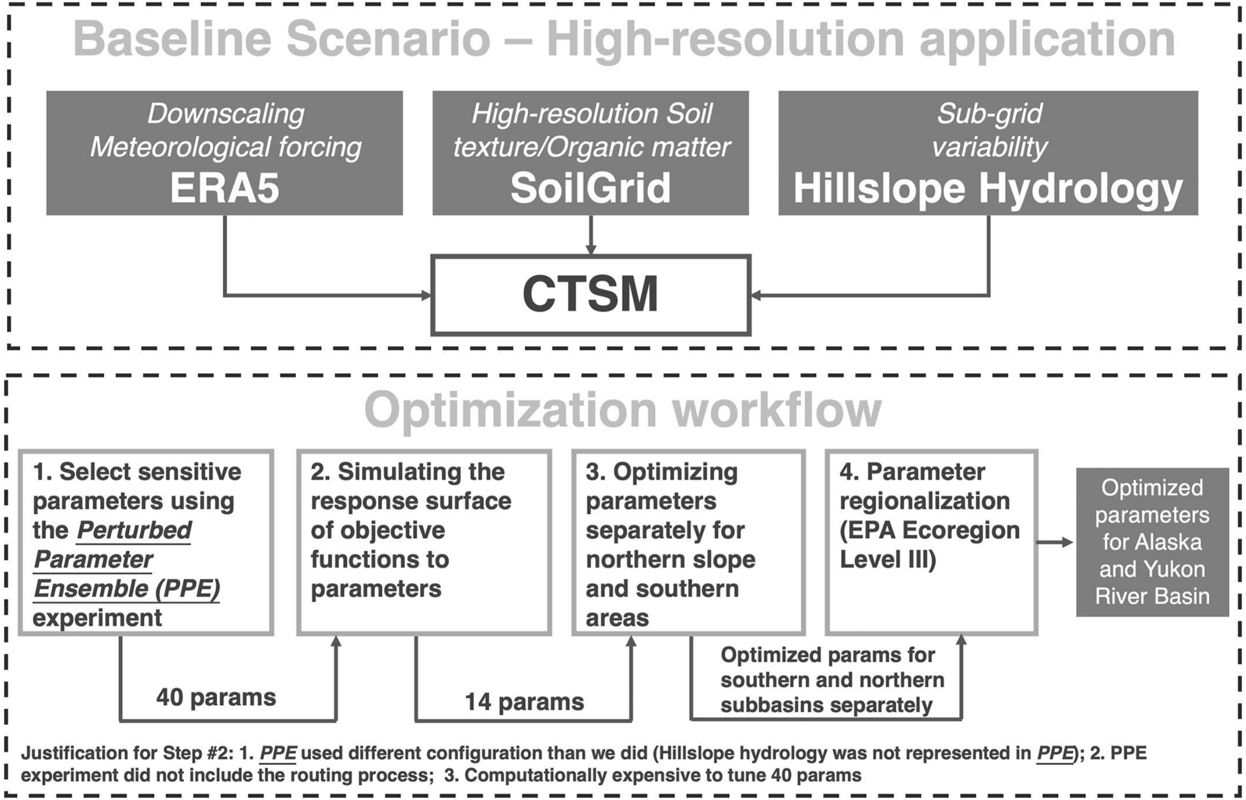
An efficient and robuse global optimization method.

Duan et al, WRR, 1992

Prior research with CTSM parameter optimization

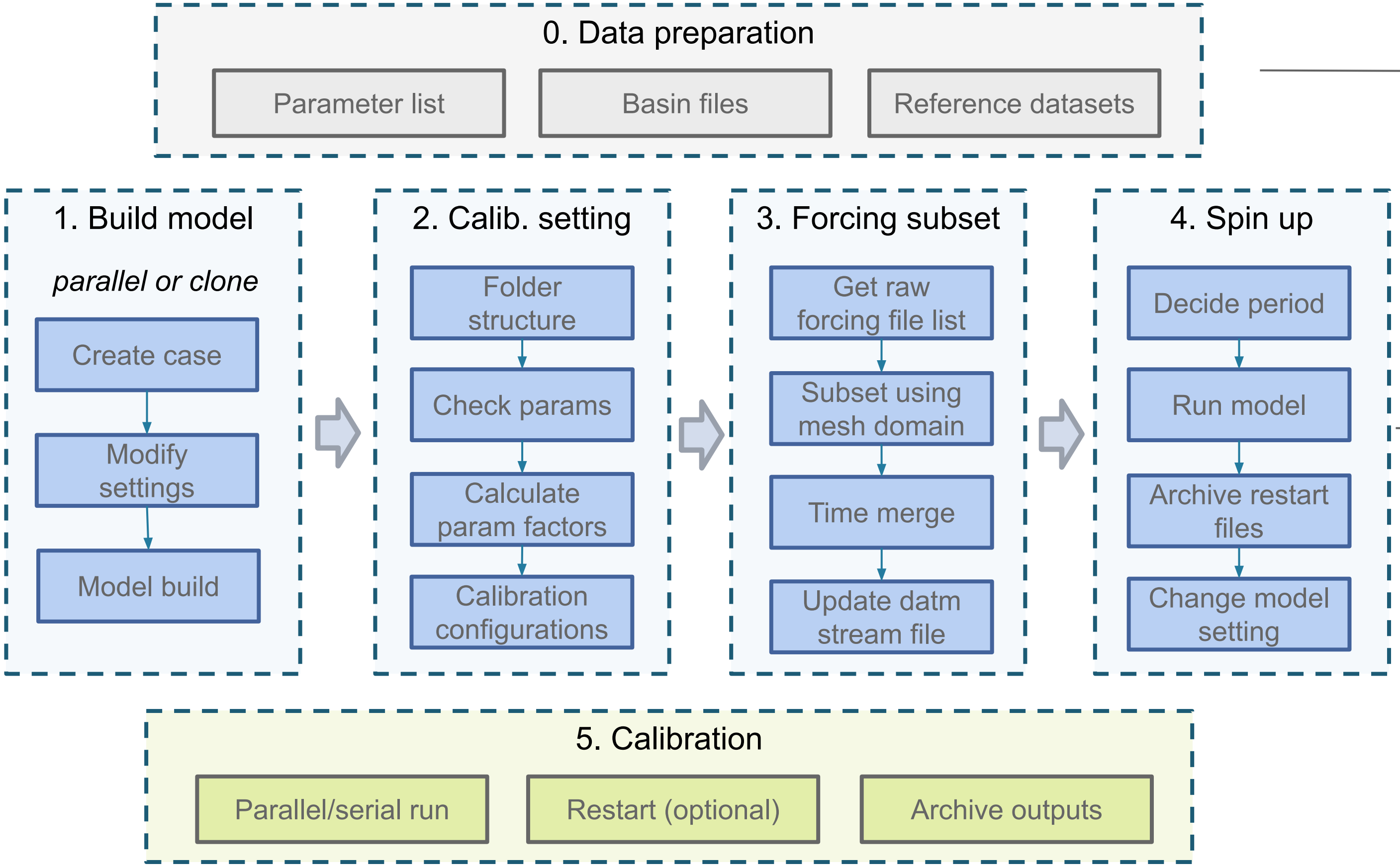
- Calibration of the CTSM model in Alaska and the Yukon River Basins
- Used MO-ASMO algorithm (surrogate modeling)
- The mean skill[1] of daily streamflow increased from 0.43 to 0.63
- *See **Y. Cheng** presentation in this session for details*

[1]skill ~ Kling-Gupta Efficiency (KGE) score



Cheng et al., 2023

A CTSM Parameter Optimization Framework



- Dataset dependent

- Python
- Configuration file
- Automatic workflow

All data and settings can be generated by submitting one job.

A CTSM Parameter Optimization Framework

0. Data preparation

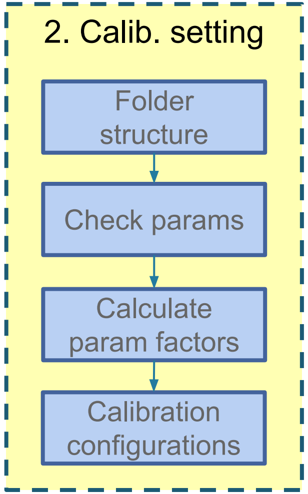
Parameter list | Basin files | Reference datasets

The parameters are selected based on PPE parameter table and manual identification of key hydrological processes. The list is not parameter specific.

1	Parameter	Default	Lower	Upper	Source	Method	Type	Binding
2	vcmaxha	72000	20000	250000	Param	Multiplicative	Hydrology	None
3	om_frac_sf	1	0.25	2	Param	Multiplicative	Hydrology	None
4	slopebeta	-3	-10	-0.5	Param	Multiplicative	Hydrology	None
5	fff	0.5	0.01	10	Param	Multiplicative	Hydrology	None
6	e_ice	6	1	8	Param	Multiplicative	Hydrology	None
7	liq_canopy_storage_scalar	0.1	0.025	4	Param	Multiplicative	Hydrology	None
8	baseflow_scalar	Default	0.0005	0.1	Namelist	Multiplicative	Hydrology	None
9	FMAX	Default	0.2	0.8	Surfdata	Multiplicative	Hydrology	None
10	hksat_sf	Default	0.9	9	Param	Multiplicative	Hydrology	None
11	krmax	1.22E-09	5.83E-11	6.90E-09	Param	Multiplicative	Plant hydrau	None
12	d_max	15	5	100	Param	Multiplicative	Stomatal resi	None
13	frac_sat_soil_dsl_init	0.8	0.25	2	Param	Multiplicative	Stomatal resi	None
14	cv	0.01	0.0025	0.04	Param	Multiplicative	Stomatal resi	None
15	a_coef	0.13	0.05	0.15	Param	Multiplicative	Stomatal resi	None
16	upplim_destruct_metamorph	175	10	500	Namelist	Multiplicative	Snow Proces	None
17	n_melt_coef	200	25	600	Param	Multiplicative	Snow Proces	None
18	medlynintercept	100	1	20000	Param	Multiplicative	Stomatal resi	None
19	precip_repartition_nonglc_all_rain_t	2	0	4	Namelist	Additive	Hydrology	precip_repart

- Parameters in parameter netcdf, surface data netcdf, and namelist text files are suppoted
- Multiplicative and additive factors are supported
- Binding parameters will use the same factors.
- Default and Type are optional.

A CTSM Parameter Optimization Framework



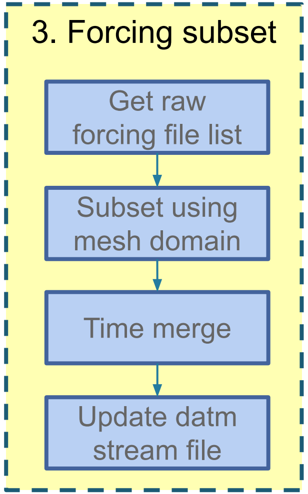
The Optimization Software Toolkit for Research Involving Computational Heuristics (**OSTRICH**) is a model-independent program that automates the processes of model calibration and design optimization without requiring the user to write any additional software.

Global search algorithms implemented within OSTRICH

Acronym	Algorithm	# Objectives	Serial?*	Parallel?	Warm Start?	Pre-Emption?	Parameter Correction?	List of Initial Parameters?	Math and Stats?	Line Search?	Reference or Contact Information
APPSO	Asynchronous Parallel Particle Swarm Optimization	1		■	■	■	■	■			(Venter and Sobieszczanski-Sobieski, 2006)
BEERS	Balanced Exploration-Exploitation Random Search	1	■		■						lsmatott@buffalo.edu
BGA	Binary-coded Genetic Algorithm	1	■	■	■		■	■			(Yoon and Shoemaker, 1999)
CSA	Combinatorial Simulated Annealing	1	■		■						(Kirkpatrick *et al.*, 1983)
DDDS	Discrete DDS	1	■		■	■					(Tolson *et al.*, 2009)
DDS	Dynamically Dimensioned Search	1	■		■	■	■				(Tolson and Shoemaker, 2007)
PDDS	Asynchronous Parallel DDS	1		■	■	■	■	■			(Tolson *et al.*, 2014)
PSO	Particle Swarm Optimization	1	■	■	■	■	■	■			(Beielstein *et al.*, 2002; Kennedy *et al.*, 2001; Kennedy and Eberhart, 1995)
RGA	Real-coded Genetic Algorithm	1	■	■	■		■	■			(Yoon and Shoemaker, 2001)
SA	Simulated Annealing	1	■	■	■		■				(Dougherty and Marryott, 1991; Marryott *et al.*, 1993)
SCE	Shuffled Complex Evolution	1	■		■		■				(Duan *et al.*, 1993; Duan *et al.*, 1992)
SMPLR	Sampling Algorithm (Big Bang - Big Crunch)	1	■		■			■			(Erol and Eksin, 2006)
VSA	Vanderbilt-Louie Simulated Annealing	1	■	■	■		■				(Vanderbilt and Louie, 1984)

https://usbr.github.io/ostrich Matott et al., 2011, 2012

A CTSM Parameter Optimization Framework



ctsmforc.NLDAS2.0.125d.v1.Prec.1980-01.nc
ctsmforc.NLDAS2.0.125d.v1.Prec.1980-02.nc
ctsmforc.NLDAS2.0.125d.v1.Prec.1980-03.nc
…
ctsmforc.NLDAS2.0.125d.v1.Prec.2018-12.nc

Raw NLDAS forcing

subset_ctsmforc.NLDAS2.0.125d.v1.Prec.1980-01.nc
subset_ctsmforc.NLDAS2.0.125d.v1.Prec.1980-02.nc
subset_ctsmforc.NLDAS2.0.125d.v1.Prec.1980-03.nc
…
subset_ctsmforc.NLDAS2.0.125d.v1.Prec.2018-12.nc

Subsetting
- Effectively reduce time cost for regional studies

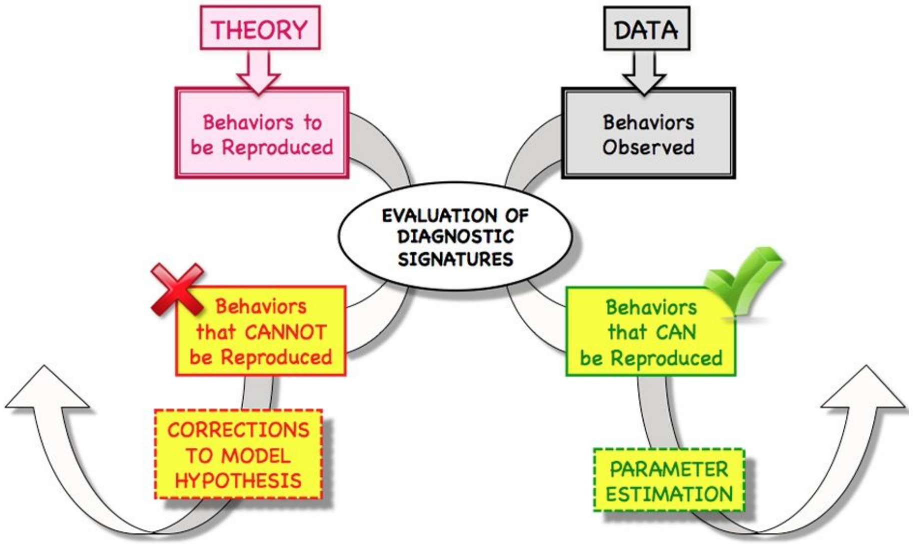
subset_ctsmforc.NLDAS2.0.125d.v1.Prec.1980-1984.nc
subset_ctsmforc.NLDAS2.0.125d.v1.Prec.1985-1989.nc
subset_ctsmforc.NLDAS2.0.125d.v1.Prec.1990-1994.nc
…
subset_ctsmforc.NLDAS2.0.125d.v1.Prec.2015-2018.nc

Time merging (month to X-years)
- Easier file management
- Avoid excess file numbers in some systems

Using large-sample watershed modeling to support learning and development

Benefits of large-sample watershed modeling

- **Improved accuracy**: Broad understanding of the model's performance, limitations and variability
- **Statistical robustness**: Increase the statistical robustness of the simulation and calibration results
- **Regional variations**: To identify and account for regional variations in model parameters and to test the generalizability of the model across different basins.
- **Improved understanding**: Reveal important relationships and dependencies between the model parameters, leading to a deeper understanding of the underlying hydrological processes.
- **Better representation**: A better representation of the diversity and variability of natural systems, enabling the assessment of the impacts of changes in a more comprehensive manner.

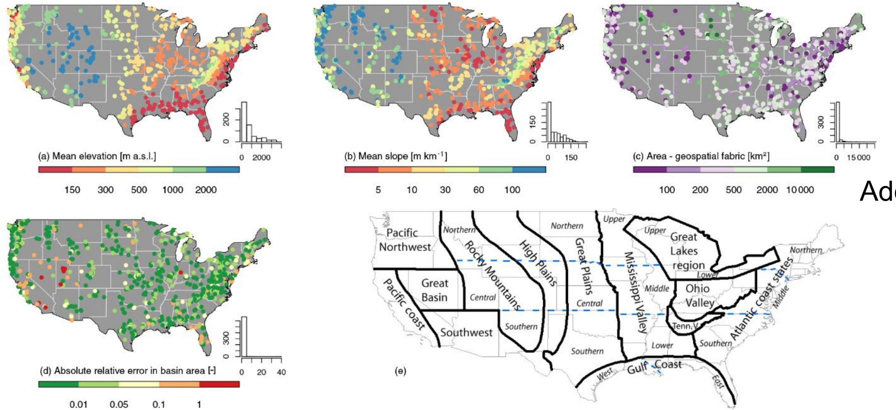


Gupta et al., Large-sample hydrology: a need to balance depth with breadth. HESS. 2014

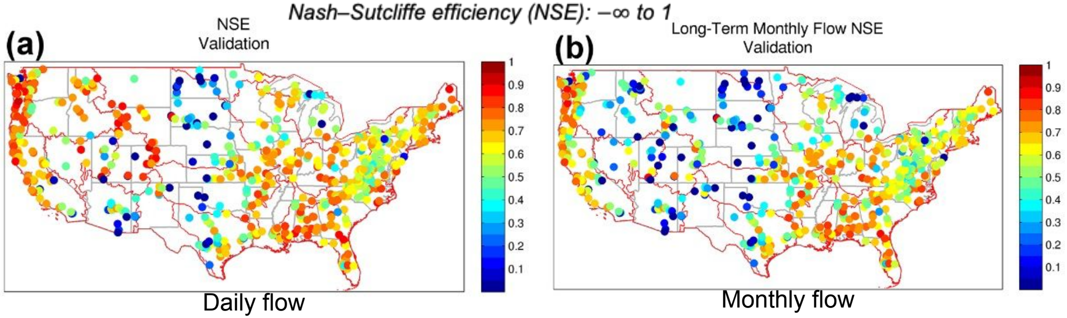
Large-sample watershed modeling

CAMELS (Catchment Attributes and Meteorology for Large-sample Studies

- A comprehensive set of catchment attributes, meteorological variables, streamflow observations, and model results for 671 US catchments
- Widely used in hydrology research to develop and evaluate hydrological models, variability and predictability
- Has been a central dataset in the global rise of machine learning in hydrology
- Has been extended in many countries by independent efforts
- Was originally developed in NCAR RAL to study streamflow predictability and model complexity

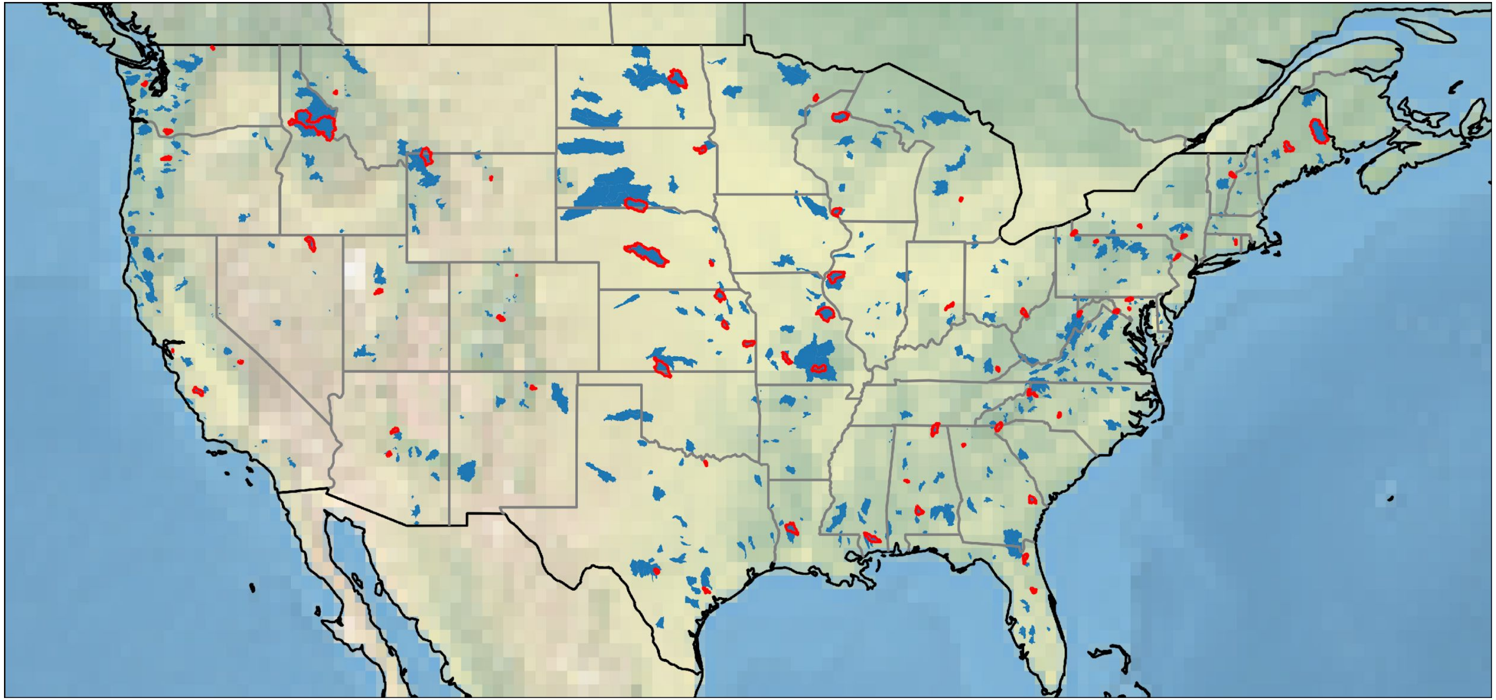


Addor et al., 2017



Newman et al., 2015

Large-sample watershed modeling using CAMELS



Blue: 671 CAMELS basins

Red: 10% randomly selected basins for this presentation.

- Each basin is simplified as a mesh grid to facilitate large-sample modeling.
- For nested basins (i.e., upstream VS downstream), the split strategy is adopted to subtract upstream basins from downstream basins because mesh grids cannot overlap.
- All the 671 basins of CAMELS will be used in the final experiment.

Large-sample watershed modeling using CAMELS

Modified Kling-Gupta Efficiency (KGE')
$-\infty$ *(worst) to 1* (best)

$$KGE' = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}$$

Correlation coefficient

Bias ratio

Variability ratio

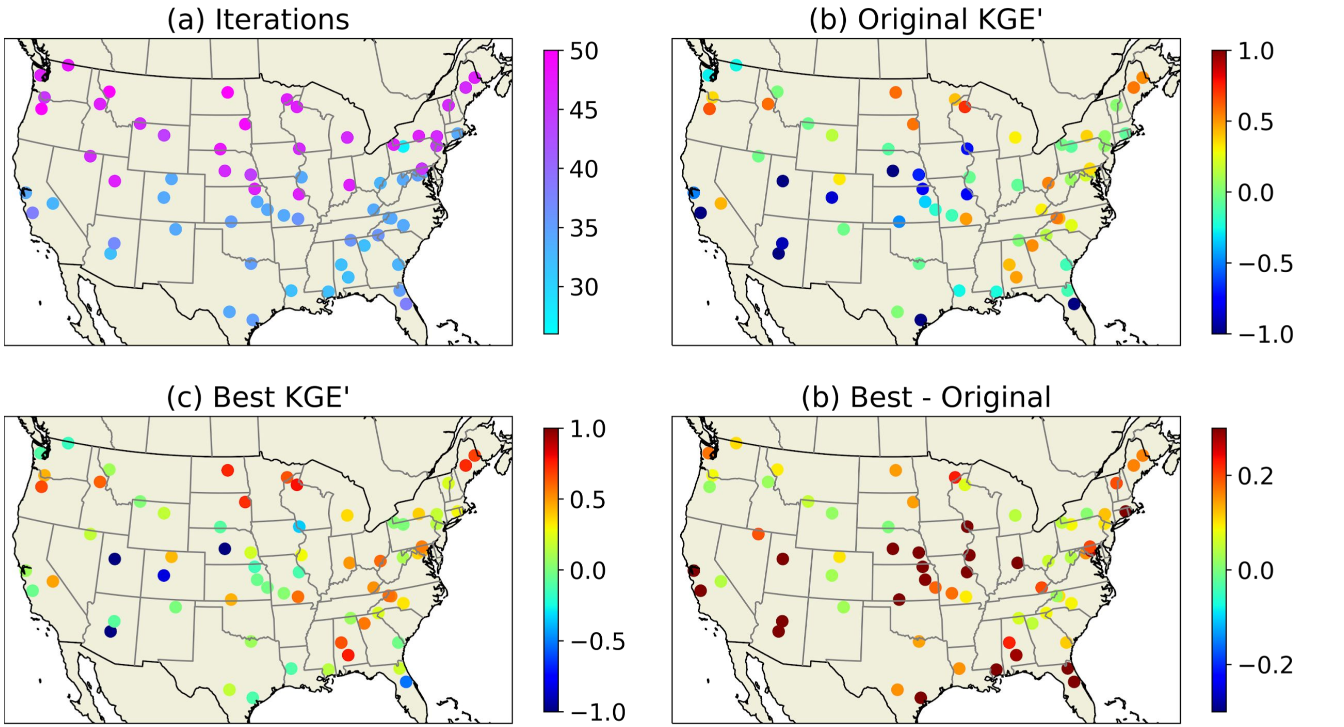
For the calibration period:

Computation

- 1 CPU and12 hours are allocated to each basin
- ~40 trials per basin, while normally hundreds of trials are needed to achieve ideal calibration

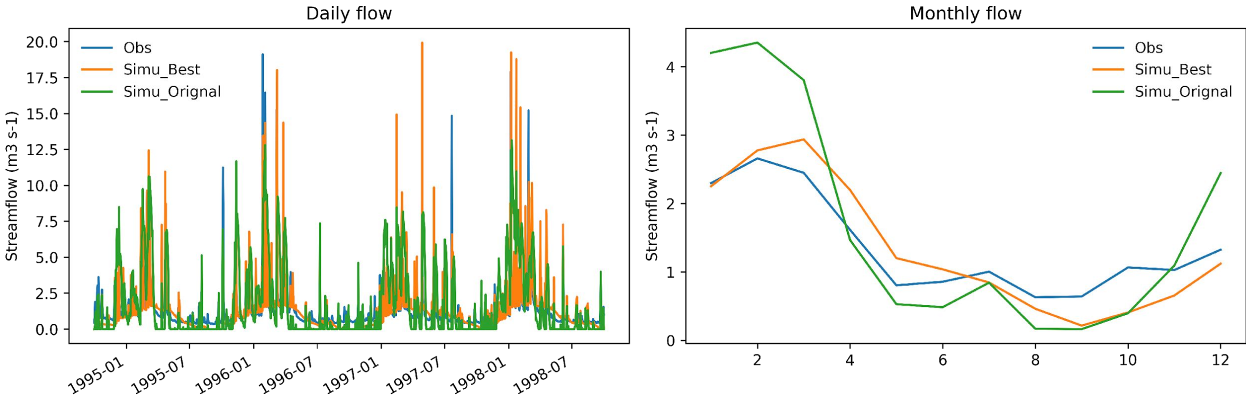
Results

- KGE' increases in 66 out of 67 basins after calibration.
- The median KGE' increases from -0.01 to 0.17 after calibration.
- The median/mean of "Best - Original" KGE' is 0.15/0.53.

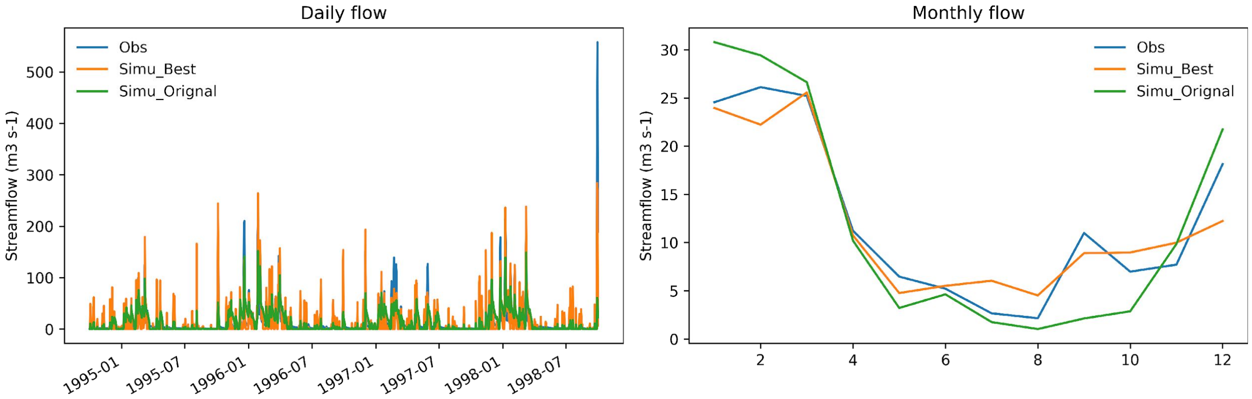


Large-sample watershed modeling using CAMELS

Example-1: 02465493
KGE’: 0.43 -> 0.64



Example-2: 02427250
KGE’: 0.48 -> 0.76



Summary and Next Steps

A streamlined CTSM calibration workflow and hydrology 'testbed':

- This new CTSM calibration capability development supports a larger project to assess the robustness of different hydrological model configurations for projecting forced responses to climate change.
- The CAMELS-CTSM implementation offers a useful and efficient testbed for evaluating alternative CTSM model configurations and development choices.
- The parameter estimation workflow will enhance the local performance of the CTSM hydrology component and yield insights into regional to continental parameter estimation strategies.

Next steps:

- Future calibration development efforts include:
  - improving parallel computation
  - multi objective calibration
  - further parameter refinement
  - distributed domains
  - the use of river routing
  - regionalization to uncalibrated basins
- We will also assess different structural configuration options and the hillslope parameterization

Thank you!

guoqiang@ucar.edu