**yan-etal\_2024\_james**

**Ensemble-based Spatially Distributed CLM5 Hydrological Parameter Estimation for the Continental United States**

Hongxiang Yan1\*, Ning Sun1, Hisham Eldardiry1, Travis Thurber1, Patrick Reed2, Daniel Kennedy3, Sean Swenson3, and Jennie Rice1

1 Pacific Northwest National Laboratory, Richland, WA, USA

2 Department of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA

3 National Center for Atmospheric Research, Boulder, CO, USA

\* Correspondence: Hongxiang Yan, [hongxiang.yan@pnnl.gov](mailto:hongxiang.yan@pnnl.gov)

**Abstract**

One of the major challenges in large-domain hydrological modeling efforts lies in the estimation of spatially distributed hydrological parameters while simultaneously accounting for their associated uncertainties. Addressing this challenge is particularly difficult in ungauged locations. With growing societal demands for large-scale streamflow projections to inform water resource management and long-term planning, evaluating and constraining hydrological parameter uncertainty is increasingly vital. This study introduces a hybrid regionalization approach to enhance hydrological predictions of the Community Land Model version 5 (CLM5) across the Continental United States (CONUS), with a total of 50,629 1/8° grid cells. This hybrid method combines the strengths of two existing techniques: parameter regionalization and streamflow signature regionalization. It identifies ensemble behavioral parameters for each 1/8° grid cell across the CONUS domain, tailored to three distinct streamflow signatures focused on low flows, high flows, and annual water balance. Evaluating this hybrid method for 464 CAMELS (Catchment Attributes and Meteorology for Large-sample Studies) basins demonstrates a significant improvement in CLM5 hydrological predictions, even in challenging arid regions. In CONUS applications, the derived spatially distributed parameter sets capture both spatial continuity and variation of parameters, highlighting their heterogeneous nature within specific regions. Overall, this hybrid regionalization approach offers a promising solution to the complex task of improving hydrological modeling over large domains for important hydrological applications.

**Journal Reference**

Yan, H., Sun, N., Eldardiry, H., Thurber, T., Reed, P., Kennedy, D., Swenson, S., and Rice, J. (2024). Ensemble-based Spatially Distributed CLM5 Hydrological Parameter Estimation for the Continental United States. Submitted to *Journal of Advances in Modeling Earth Systems* – January 2024.

**CLM5 Model Reference**

CTSM Development Team. (2022). IMMM-SFA/im3-clm: v1.0.0 (v1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.6653705>

**Input Data Reference**

CAMELS dataset: <https://ral.ucar.edu/solutions/products/camels>

NLDAS-2 dataset: <https://disc.gsfc.nasa.gov/datasets?keywords=NLDAS>

GSCD datasets: <https://www.gloh2o.org/gscd/>

**Output Data Reference**

CLM5 CAMELS basin ensemble simulations: <https://data.msdlive.org/records/5rpkv-h8n12>

CLM5 CONUS ensemble simulations: XX1

Behavioral ensemble CLM5 parameters: <https://data.msdlive.org/records/41bw1-3q739>

**Code Reference**

Yan, H. (2024). Supporting code and files for Yan et al. 2024 – Journal of Advances in Modeling Earth Systems [Code]. Zenodo. XX2

**Reproduce My Experiment**

Clone the CLM5 repository (<https://doi.org/10.5281/zenodo.6653705>) to set up the CLM5 model, you will need to download the NLDAS-2 forcing data and convert them into netcdf format. You will also need to generate ensemble hydrological parameter value files using parameter values at <https://data.msdlive.org/records/41bw1-3q739>. Once you have finished all CLM5 ensemble runs, you can use the regression data files in the accompanying code repository (XX2) to constrain the behavioral ensemble parameters. The output data repository (XX1) already contains the ensemble output from the CLM5 model so you can skip rerunning the CLM5 model if you want to save time.

**Reproduce My Figures**

Use the accompanying code repository (XX2) to reproduce the main and supplementary figures used in this publication.

Figure\_3.m: Describe the regional mean daily FDC

Figure\_4.m: Describe the relative bias using 3 regionalization methods

Figure\_5.m: Describe the Qmean PDF of the behavioral parameter for 7 sites

Figure\_6.m: Describe the default and behavior parameters over the CONUS

Figure\_7.m: Describe the behavioral parameters for 7 sites

Figure\_8.m: Describe the 15 behavioral parameters for one site

Figure\_9.m: Describe the ensemble daily FDC for one site using one, two, three constrains

Figure\_10.m: Describe the Qmean prediction over the CONUS

Figure\_11.m: Describe the Q10 prediction over the CONUS

Figure\_12.m: Describe the Q90 prediction over the CONUS

Figure\_S1.m: Describe the Q10 PDF of the behavioral parameter for 7 sites

Figure\_S2.m: Describe the Q90 PDF of the behavioral parameter for 7 sites