Rethinking model reliability: A dual evaluation of the National Water Model using time series and extreme events

Md. Shahabul Alam[1], Savalan Naser Neisary[1], Briton Tebbey[1], Sifan A. Koriche[1], Ahmad Jan Khattak[2], Fred Ogden[3], Mukesh Kumar[1], Ryan Johnson[4],, **Steven Burian**[1]

[1]The University of Alabama
[2]Lynker
[3]NOAA OWP
[4]University of Utah

June 19, 2025

NHWC2025
BEYOND THE SAGUARO
Tucson, Arizona

CIROH
Cooperative Institute for Research to Operations in Hydrology

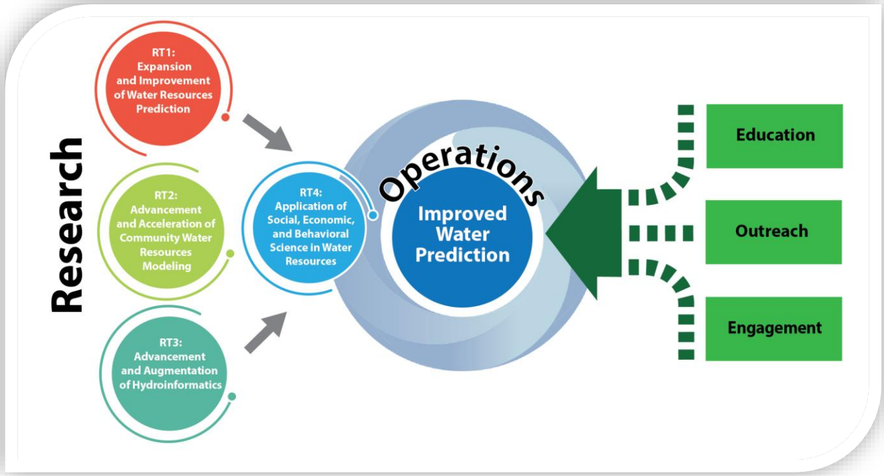
Introduction

- Extreme flood and drought predictions
  - Critical for water resource management and hazard mitigation
  - Traditional model evaluation often relies on long-term time series metrics – may overlook critical performance during extreme events
  - Event-based evaluation focuses on extreme flood and drought events
- We present a twofold evaluation of the National Water Model (NWM) version 3.0 by conducting the both
  - Traditional model evaluation approach
  - Event-based model evaluation approach

Cooperative Institute for Research to Operations in Hydrology (CIROH)

- Addressing flood / drought prediction challenges
- Advancing the Next Generation Water Resources Modeling Framework (NextGen)
- Accelerating research into operational hydrologic forecasting across the United States and its Territories

ciroh.ua.edu



CIROH

OWP OFFICE OF WATER PREDICTION

NOAA

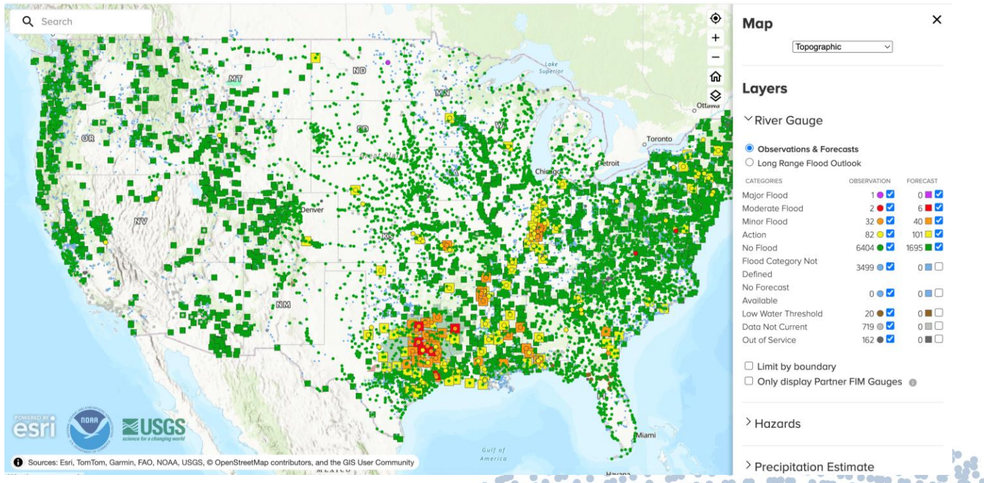
NWC NATIONAL WATER CENTER

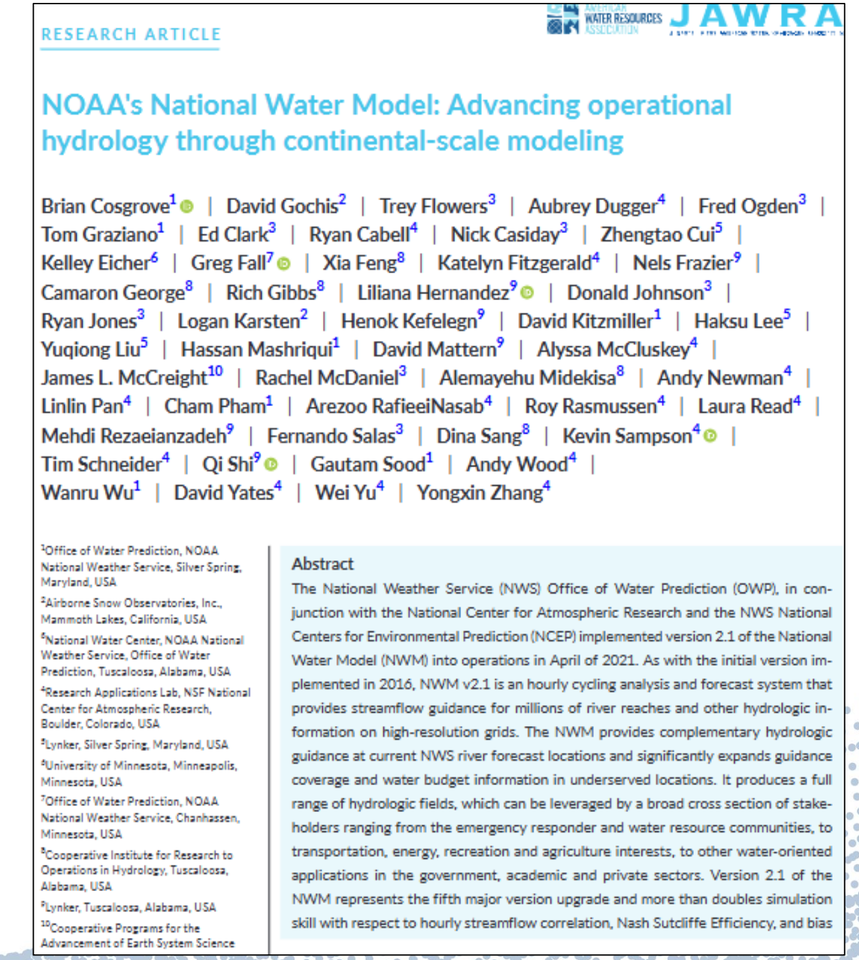
NWM and NextGen Framework

CIROH DocuHub

NextGen GitHub Repo

- NWM provides operational forecast streamflow guidance for currently underserved locations up to 3.4 million river miles
- NextGen is a continental-scale model-agnostic water resources modeling framework (flexible, consistent, interoperable)
- Generates high-resolution, spatially continuous estimates of major water cycle components
- Diverse model formulations (Conceptual, ML), open-source development (GitHub, unit testing, etc.), and friendly to domain scientists and engineers





RESEARCH ARTICLE

JAWRA

NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling

Brian Cosgrove[1] | David Gochis[2] | Trey Flowers[3] | Aubrey Dugger[4] | Fred Ogden[3] | Tom Graziano[1] | Ed Clark[3] | Ryan Cabell[4] | Nick Casiday[3] | Zhengtao Cui[5] | Kelley Eicher[6] | Greg Fall[7] | Xia Feng[8] | Katelyn Fitzgerald[4] | Nels Frazier[9] | Camaron George[8] | Rich Gibbs[8] | Liliana Hernandez[9] | Donald Johnson[3] | Ryan Jones[3] | Logan Karsten[2] | Henok Kefelegn[9] | David Kitzmiller[1] | Haksu Lee[5] | Yuqiong Liu[5] | Hassan Mashriqui[1] | David Mattern[9] | Alyssa McCluskey[4] | James L. McCreight[10] | Rachel McDaniel[3] | Alemayehu Midekisa[8] | Andy Newman[4] | Linlin Pan[4] | Cham Pham[1] | Arezoo RafieeiNasab[4] | Roy Rasmussen[4] | Laura Read[4] | Mehdi Rezaeianzadeh[9] | Fernando Salas[3] | Dina Sang[8] | Kevin Sampson[4] | Tim Schneider[4] | Qi Shi[9] | Gautam Sood[1] | Andy Wood[4] | Wanru Wu[1] | David Yates[4] | Wei Yu[4] | Yongxin Zhang[4]

[1]Office of Water Prediction, NOAA National Weather Service, Silver Spring, Maryland, USA
[2]Airborne Snow Observatories, Inc., Mammoth Lakes, California, USA
[3]National Water Center, NOAA National Weather Service, Office of Water Prediction, Tuscaloosa, Alabama, USA
[4]Research Applications Lab, NSF National Center for Atmospheric Research, Boulder, Colorado, USA
[5]Lynker, Silver Spring, Maryland, USA
[6]University of Minnesota, Minneapolis, Minnesota, USA
[7]Office of Water Prediction, NOAA National Weather Service, Chanhassen, Minnesota, USA
[8]Cooperative Institute for Research to Operations in Hydrology, Tuscaloosa, Alabama, USA
[9]Lynker, Tuscaloosa, Alabama, USA
[10]Cooperative Programs for the Advancement of Earth System Science

Abstract

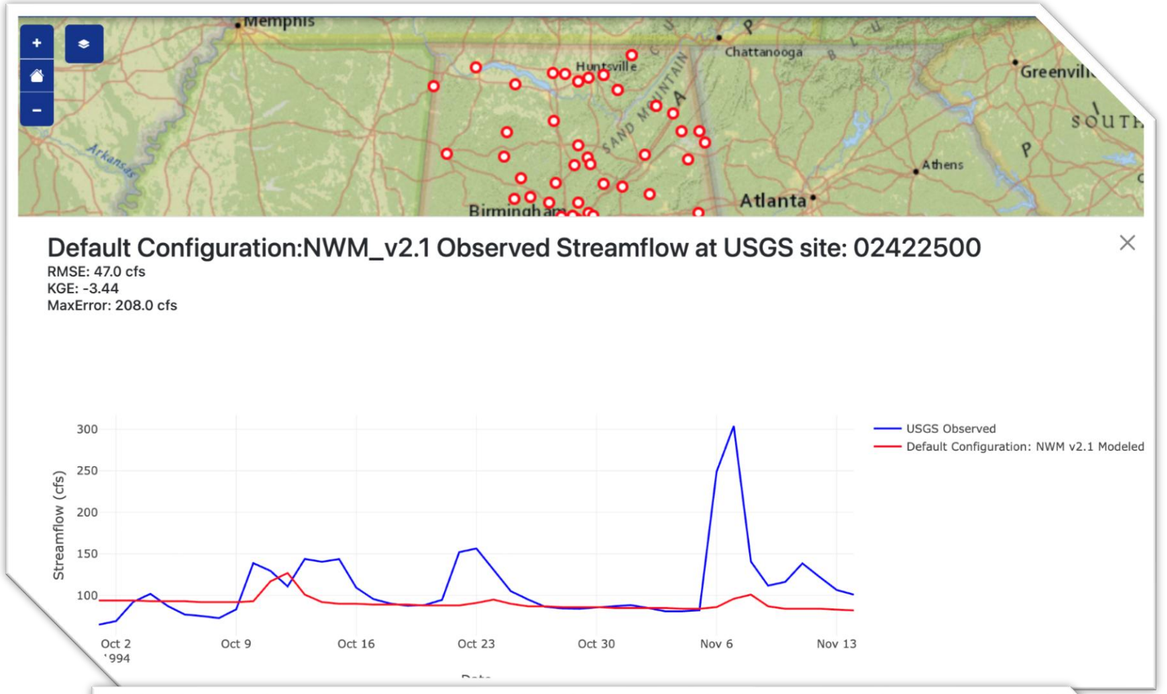
The National Weather Service (NWS) Office of Water Prediction (OWP), in conjunction with the National Center for Atmospheric Research and the NWS National Centers for Environmental Prediction (NCEP) implemented version 2.1 of the National Water Model (NWM) into operations in April of 2021. As with the initial version implemented in 2016, NWM v2.1 is an hourly cycling analysis and forecast system that provides streamflow guidance for millions of river reaches and other hydrologic information on high-resolution grids. The NWM provides complementary hydrologic guidance at current NWS river forecast locations and significantly expands guidance coverage and water budget information in underserved locations. It produces a full range of hydrologic fields, which can be leveraged by a broad cross section of stakeholders ranging from the emergency responder and water resource communities, to transportation, energy, recreation and agriculture interests, to other water-oriented applications in the government, academic and private sectors. Version 2.1 of the NWM represents the fifth major version upgrade and more than doubles simulation skill with respect to hourly streamflow correlation, Nash Sutcliffe Efficiency, and bias

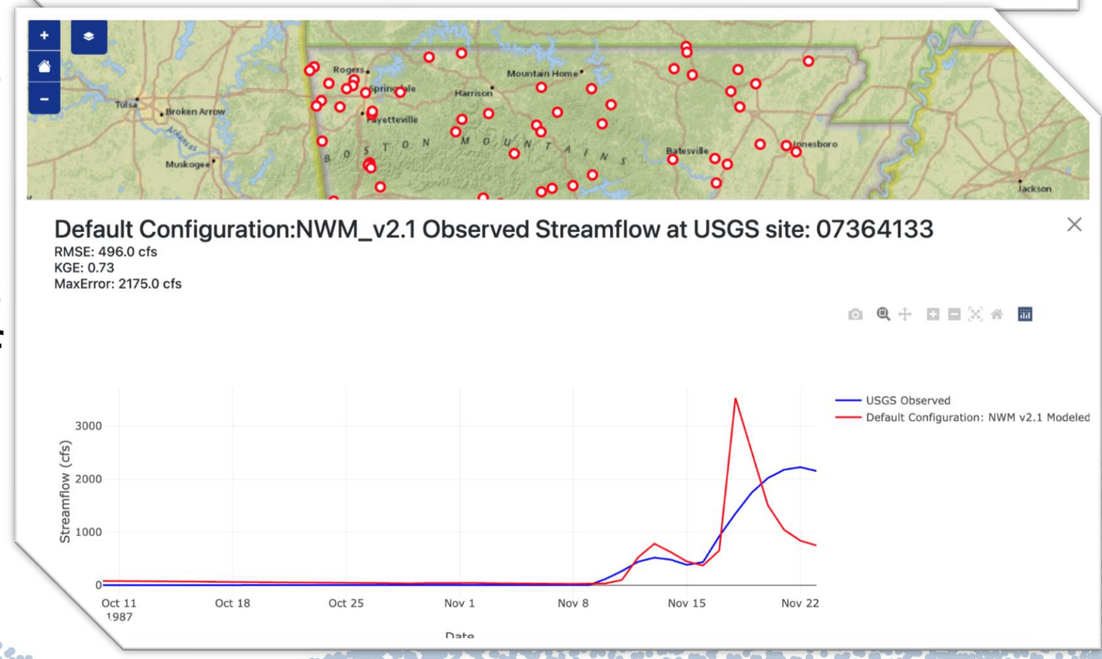
Study Objectives

Goal: Investigate NWM's reliability in predicting extreme events

Objectives:

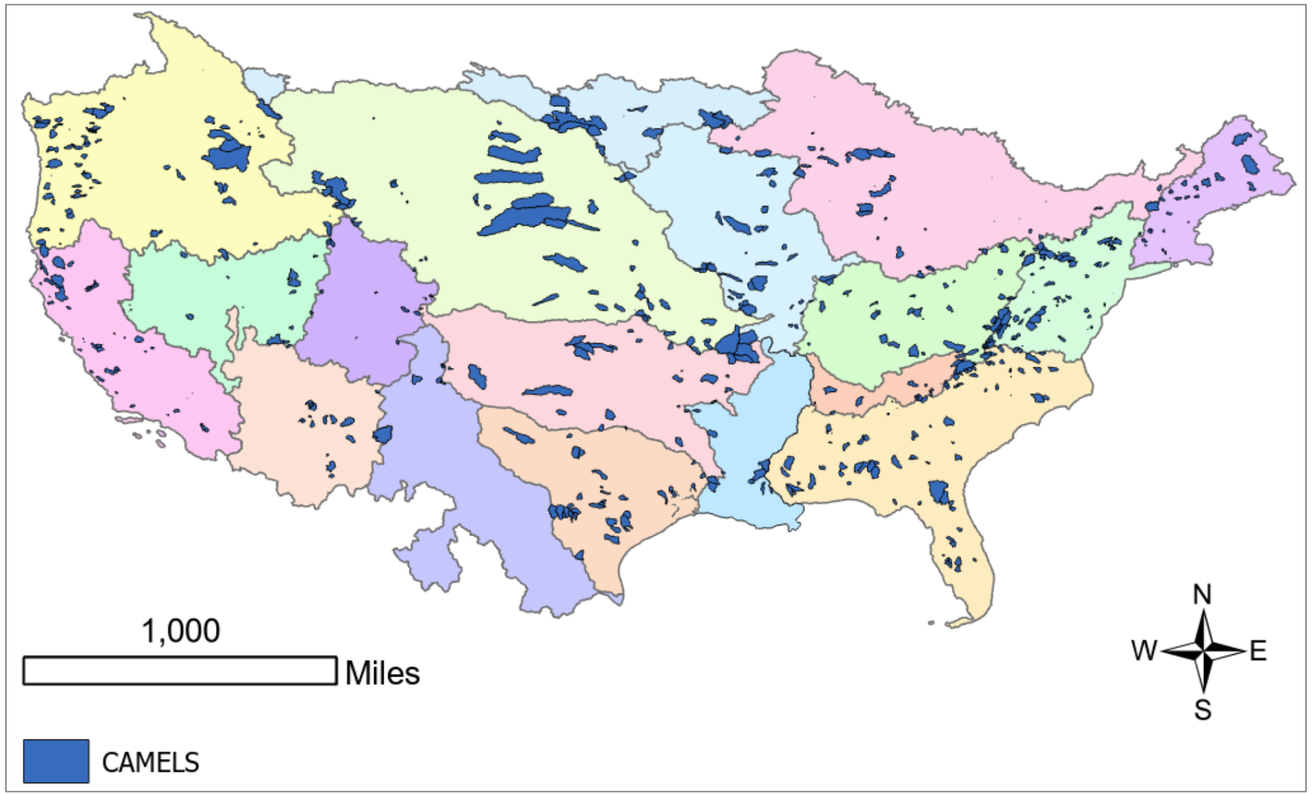
- Compute model performance metrics based on conventional time series analysis
- Compute model performance metrics based on flood and drought events of different return intervals



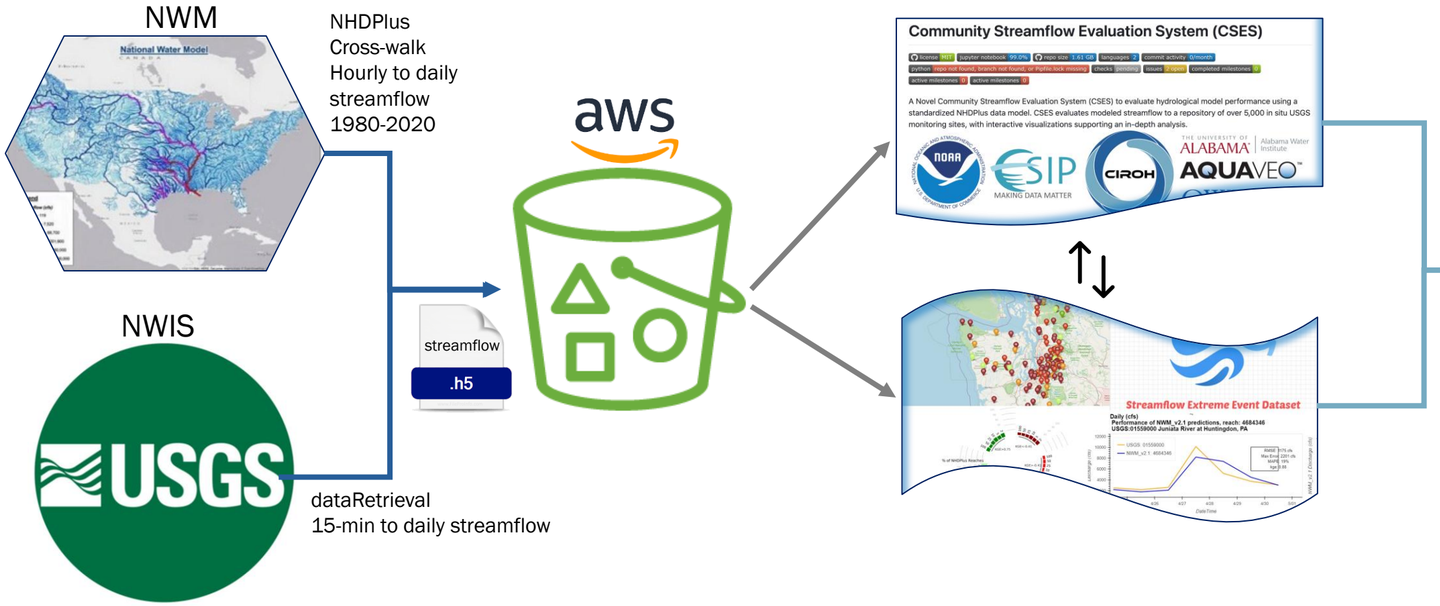


CAMELS Basins

- **How many basins?**
  - ✓ 671
  - ✓ We selected 634 that have streamflow data for both USGS and NWM
- **Size of the basins**
  - ✓ 4 to 25000 (approx.) sq. Km
- **Basin attributes**
  - ✓ Climate, Topography, Hydrologic Signatures, Soil, Vegetation, and Geology
- Why CAMELS basins?
  - ✓ Minimal human impacted – headwater basins
  - ✓ Included in the NWM calibration
  - ✓ Standard dataset



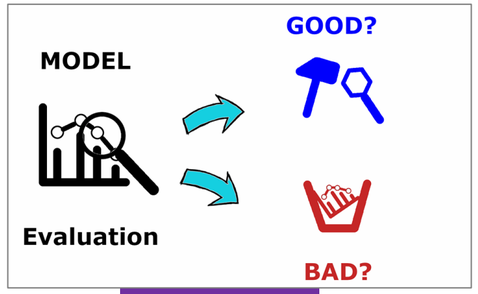
Methodology for model evaluation



CAMELS Dataset (Addor et al., 2017)

Total 671 Basins

634 Basins selected



NWM v3.0

Long time-series

Extreme events

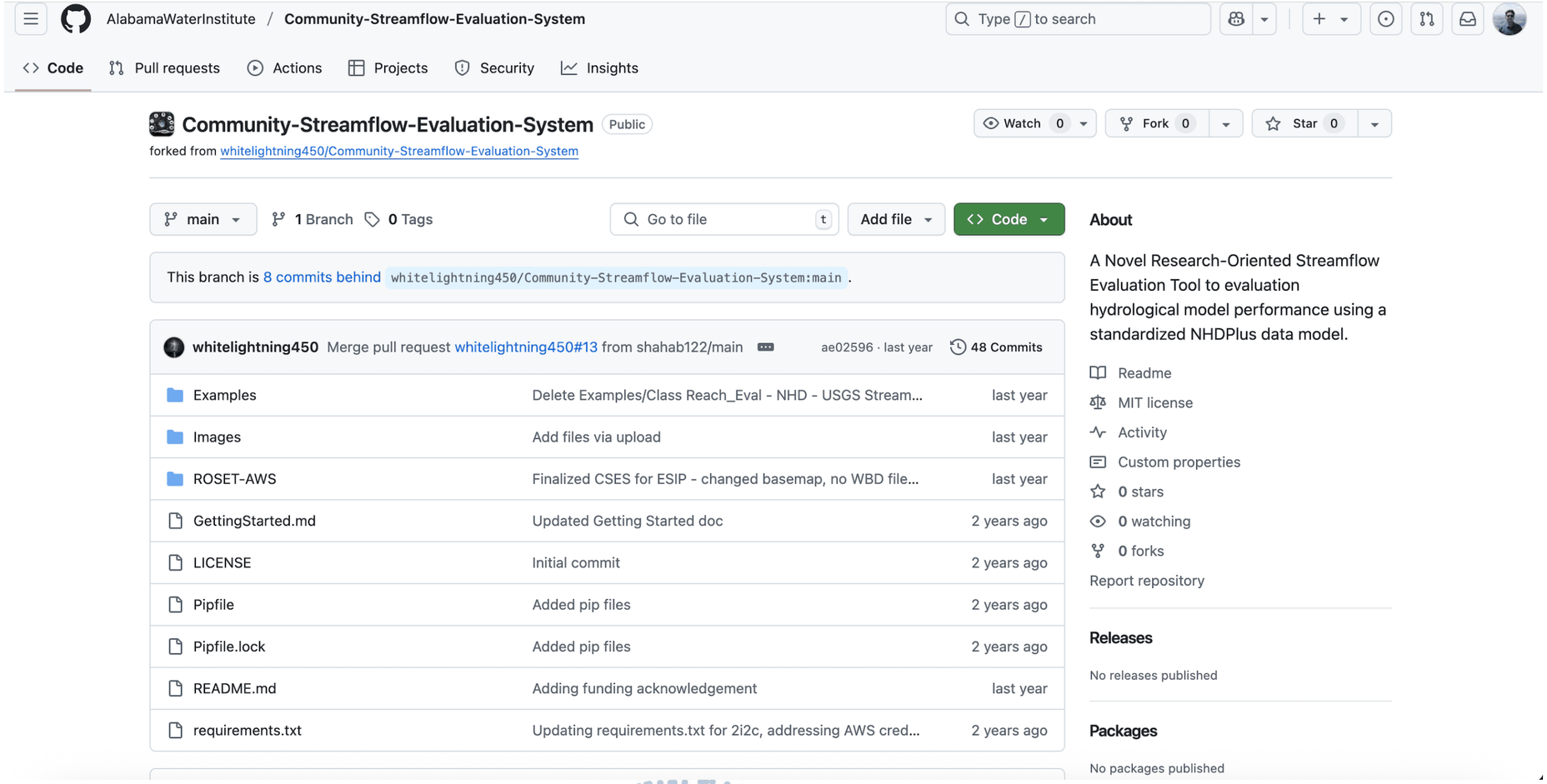
Flood: f(Annual Maximum Flow, T)
Drought: f(Annual Minimum Flow, T)

CIROH DocuHub

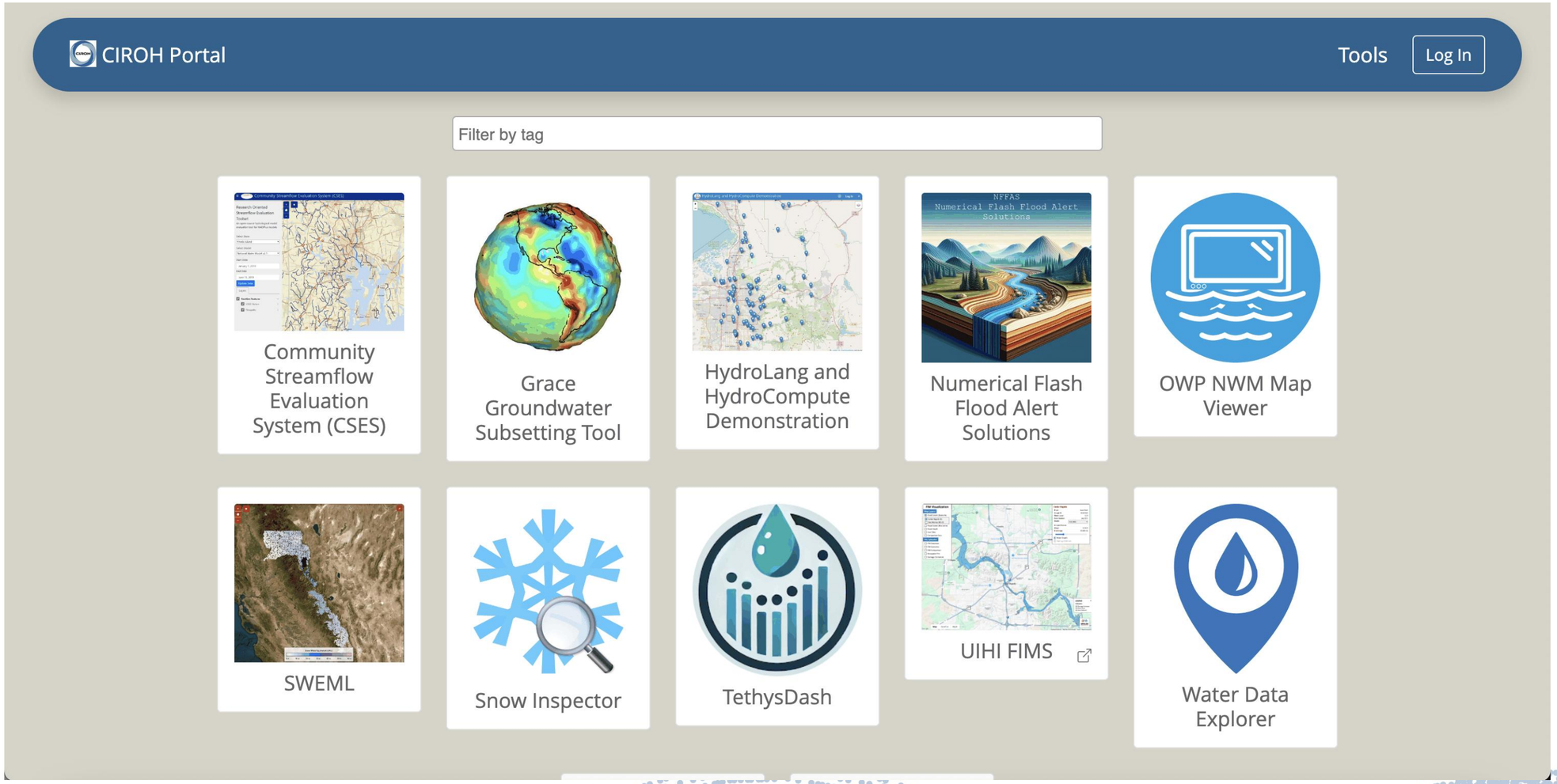
CSES GitHub

SEED-CSES GitHub

NWM evaluation using Python Tool



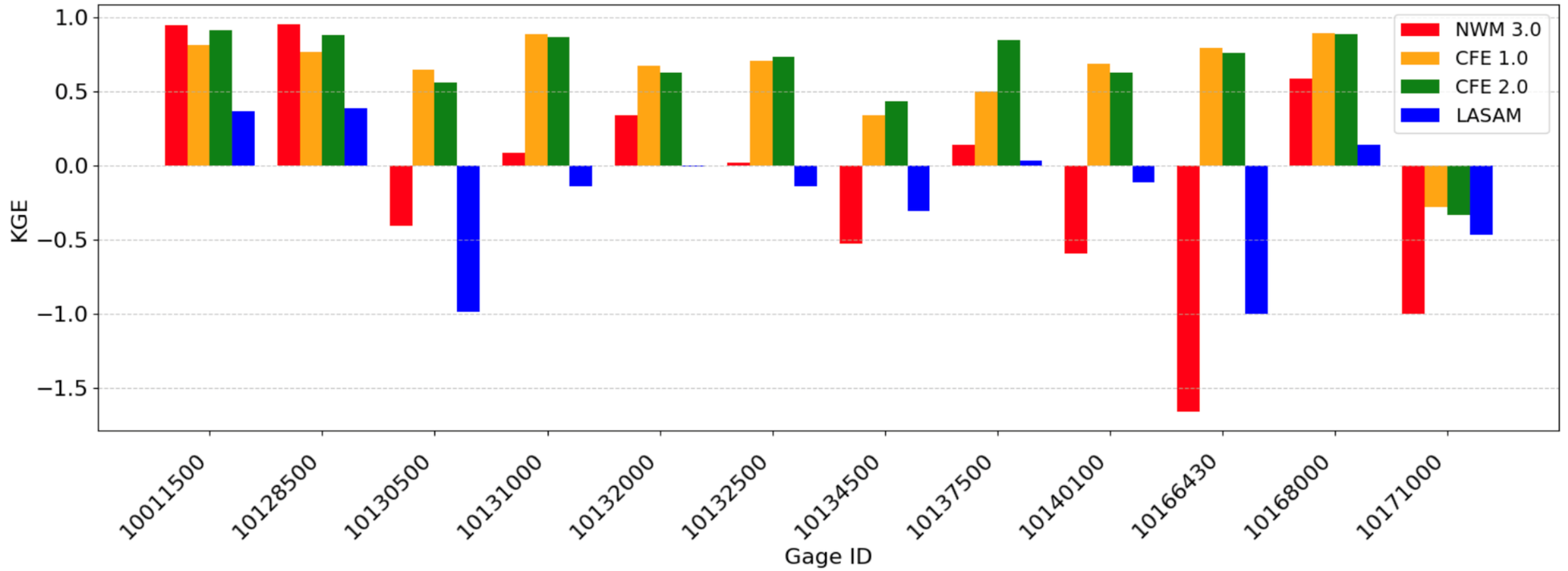
NWM evaluation using Tethys Tool



Results

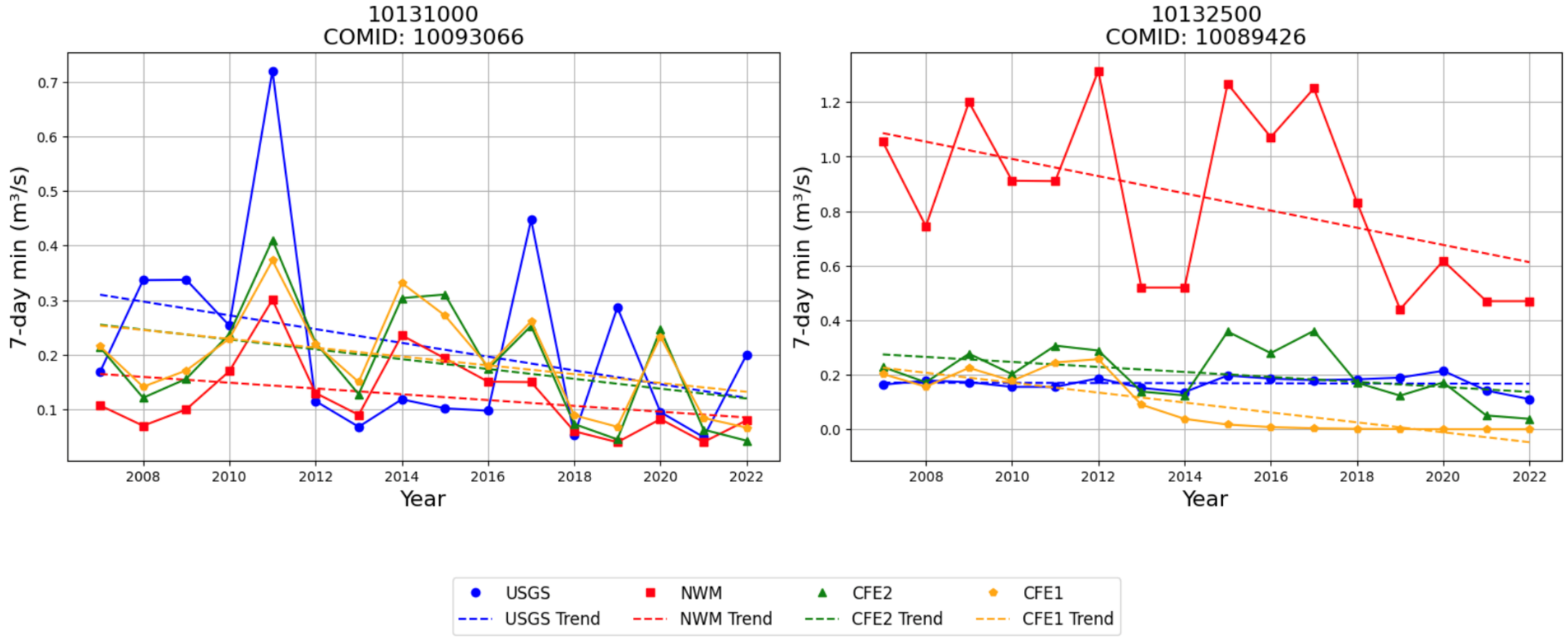
Model Performance – Traditional Approach

- CFE is either equivalent or better than NWM 3.0
- LASAM unexpectedly underperformed in an Arid/Semi-arid region



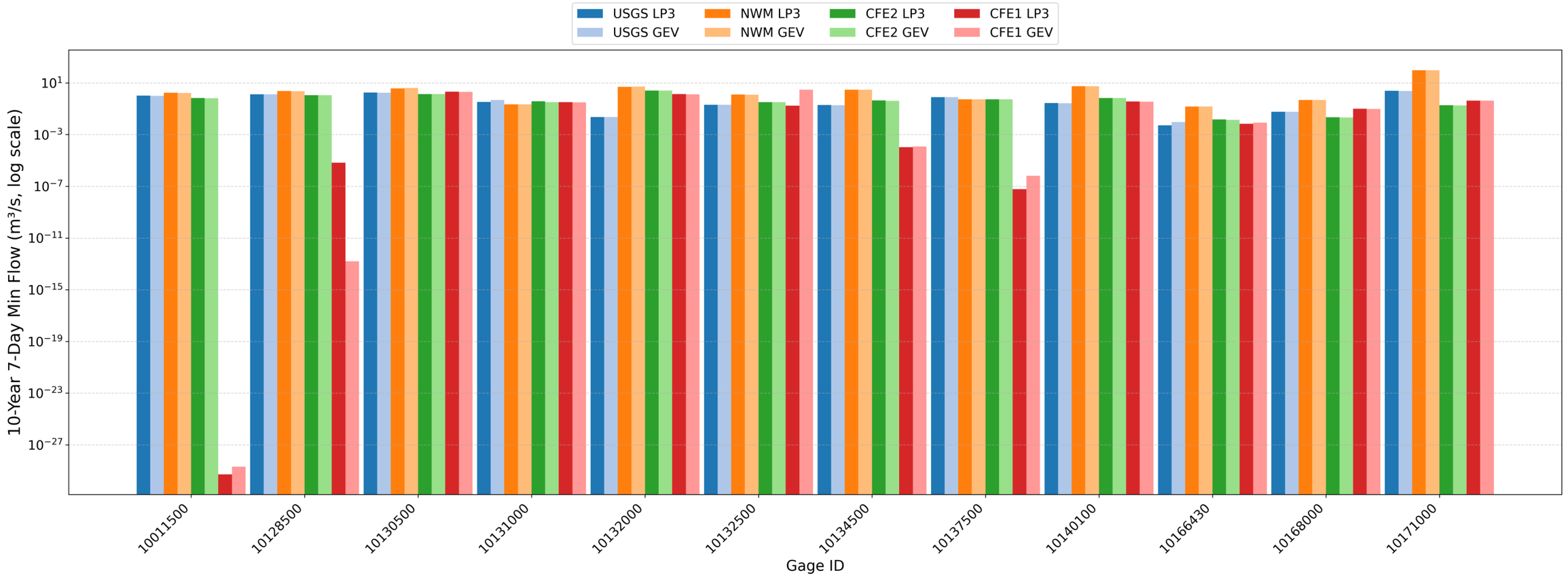
Model Performance – 7-day Flow Trends

- CFE 2.0 seems to show better performance in terms of trends and annual minimum 7-day values



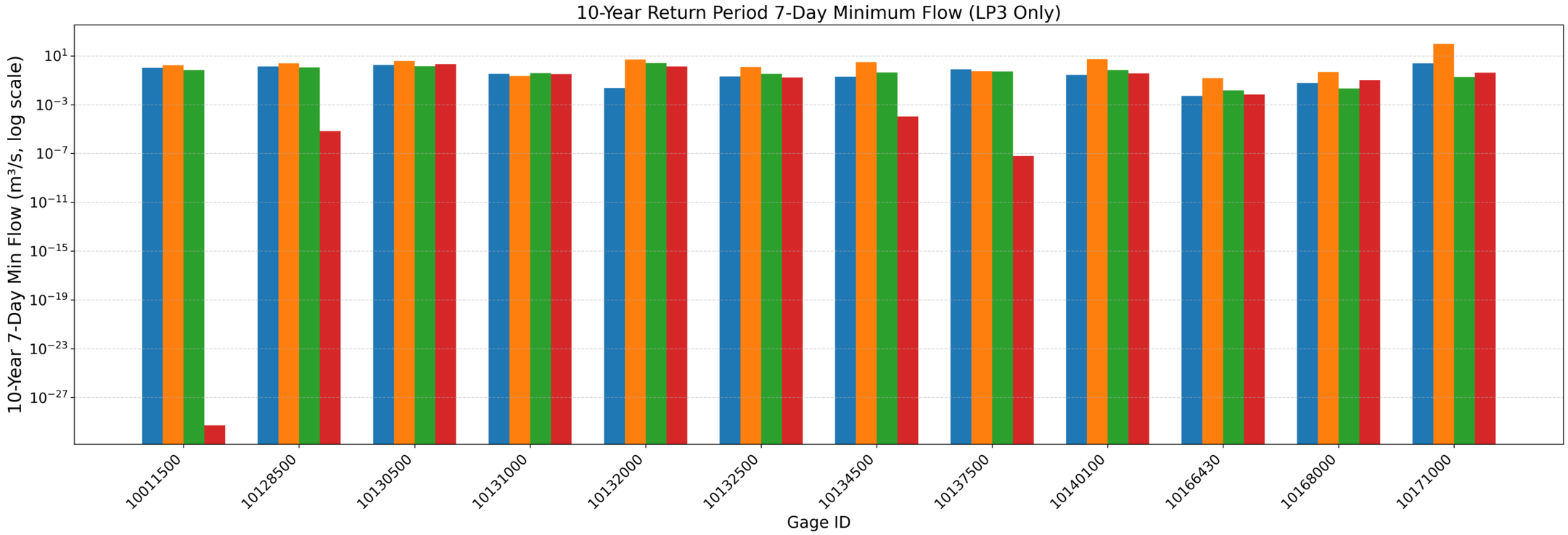
Model Performance – LP3 vs GEV

- Two distributions do not show any difference



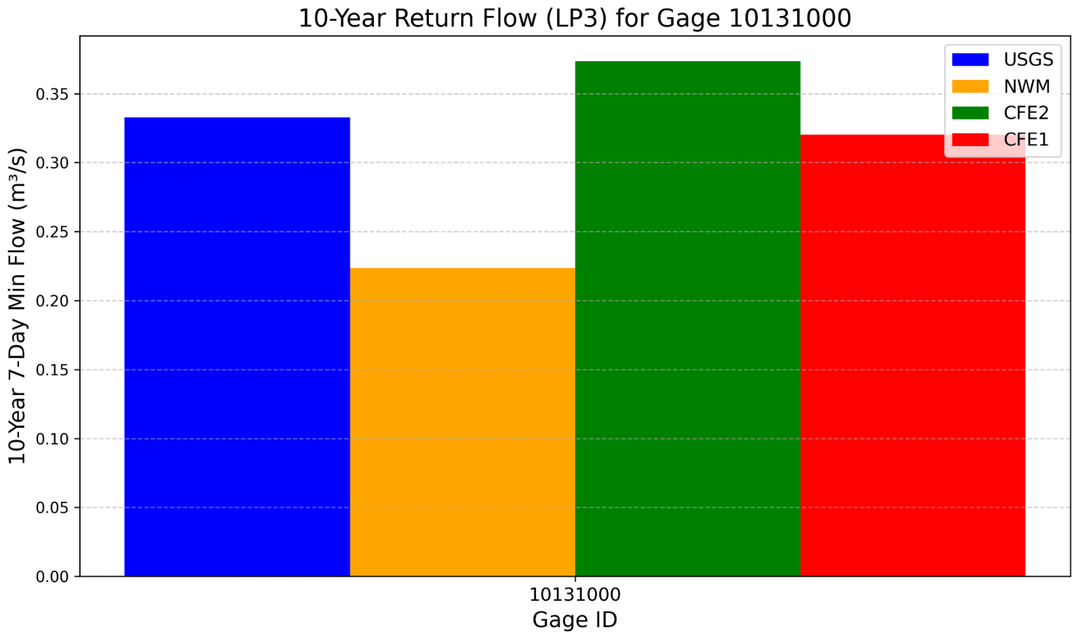
Model Performance – All Models and Basins

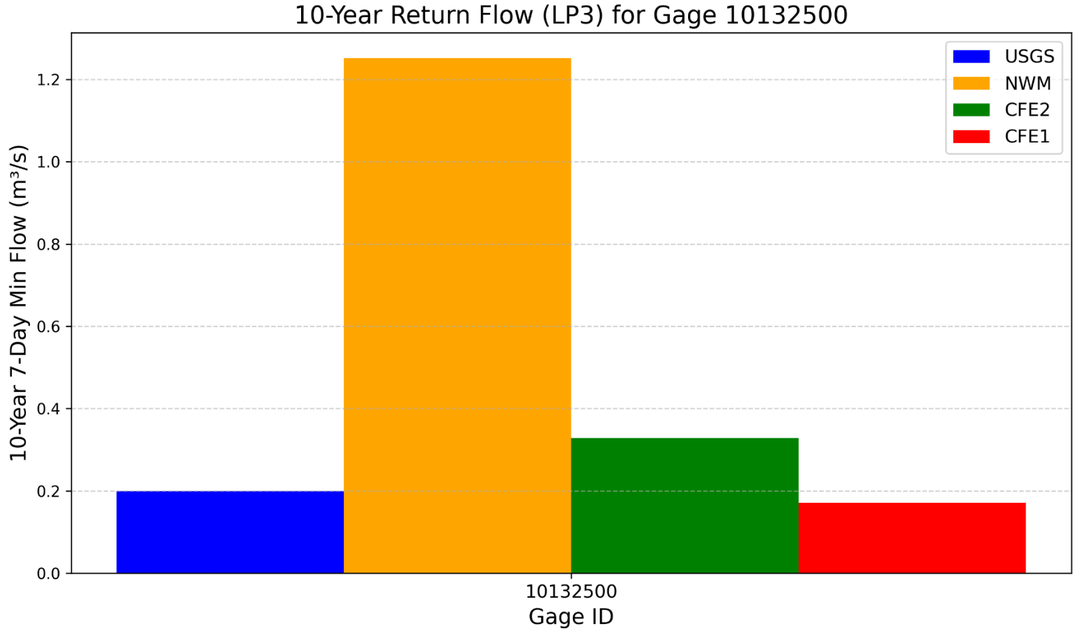
- CFE performs equal to or better than NWM3.0
- Little spatial variation in the 7Q10 values across the GSL basin



Model Performance – Two Selected Basins

- CFE can be used to estimate 7Q10 more accurately in the GSL basin





Key Findings and Conclusions

- CFE seems to be the best model for low-flow prediction
- Confirmed CFE 2.0 is an improvement over CFE 1.0 for low flows
- Accurate estimation of 7Q10 at the gauged basins
- Estimate of 7Q10 at ungauged basins may be reliable using the CFE-based streamflow simulations
- Accurate regionalization method could be an issue

Next Steps

- Expand to more arid/semi-arid basins
- Improve calibration in ungauged contexts: Deploy regionalization method
- Include uncertainty estimates
- Include more hydrological models (e.g., SUMMA, SAC-SMA, TOPMODEL)

CIROH
Cooperative Institute for Research to Operations in Hydrology

Thank You!

This research was supported by the Cooperative Institute for Research to Operations in Hydrology (CIROH)with funding under award NA22NWS4320003 from the NOAA Cooperative Institute Program. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the opinions of NOAA.