RiTHyMs v1.1

River Temperature Time Series for Hydrothermal Modeling and Analysis

RiTHyMs v1.1 is a reproducible large-sample Long Short-Term Memory (LSTM) workflow for estimating daily stream temperature at unregulated river reaches across the conterminous United States (CONUS). It builds on RiTHyMs v1.0, which demonstrated that nationally available meteorological, hydrologic, and geospatial inputs can support transferable daily stream-temperature prediction at monitored and ungauged reaches.

RiTHyMs supports aquatic-ecosystem assessment, water-quality analysis, hydrothermal and energy-water studies, and process-based river or reservoir modeling. It is applicable where upstream regulation and other local anthropogenic thermal inputs are absent or explicitly represented outside the model.

Version 1.1 updates

RiTHyMs v1.1 improves national coverage, reproducibility, and representation of thermally relevant upstream conditions.

- Extends the training record to water years 1980-2016 and evaluates independent 2017-2019 predictions.
- Expands assigned out-of-fold test coverage from 300 sites in v1.0 to 469 sites in v1.1.
- Uses fractional Daymet grid-cell coverage and flow-network-distance weighting for atmospheric forcing aggregation, reducing dilution of near-reach conditions in large heterogeneous watersheds.
- Adds wind, PRMS hydrologic states, flow, snowmelt, daylength, shortwave energy, and engineered thermal predictors.
- Adds watershed, riparian land-cover, hydrogeologic, and glacier-related static features.
- Uses six-cluster, ten-fold spatial cross-validation, a 365-day LSTM input sequence, training-only normalization, masked missing observations, deterministic execution, and fixed 100-epoch training.

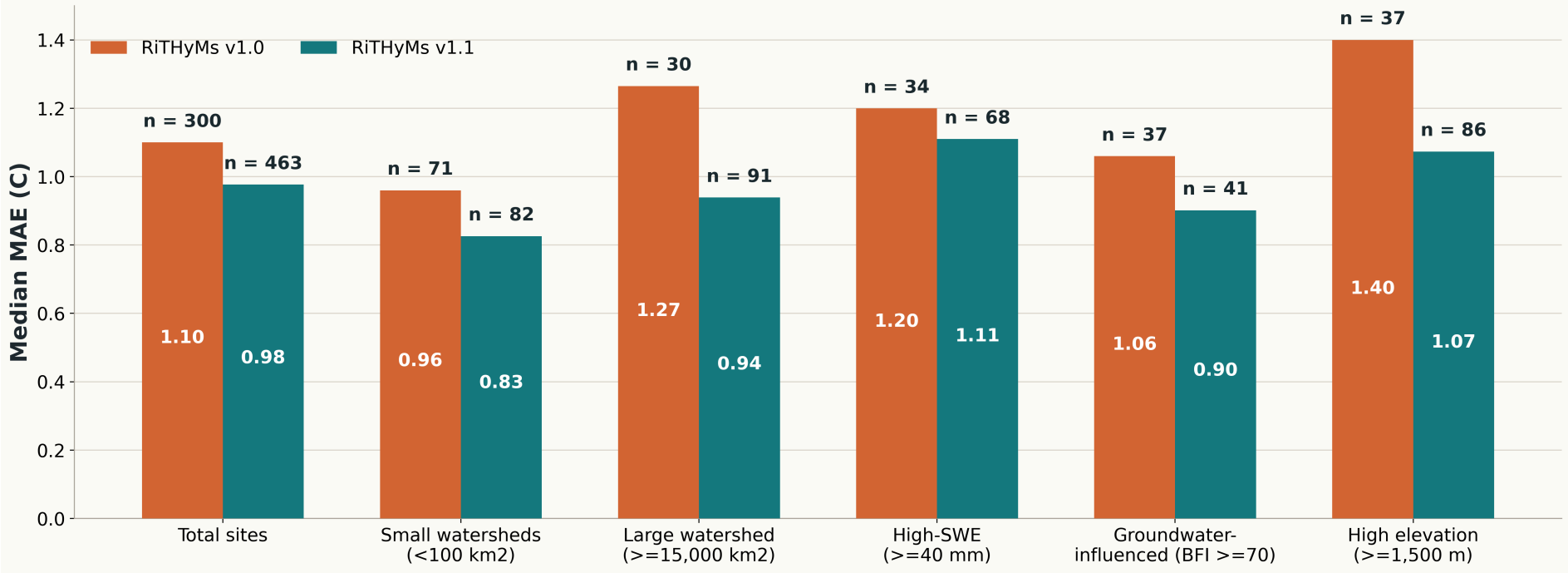
The v1.1 evaluation includes 463 scored out-of-fold test sites, with median MAE of approximately 0.98 C and median NSE of approximately 0.97.

RiTHyMs v1.1 method updates

Component	RiTHyMs v1.0	RiTHyMs v1.1
Test coverage	300 out-of-fold test sites	463 out-of-fold test sites
Training period	2010-2016	1980-2016
Upstream aggregation	Mean over thermally relevant upstream area	Flow-network-distance-weighted upstream mean
Dynamic predictors	7 daily predictors	18 daily predictors: Adds baseflow, SWE-melt, and engineered thermal features
Static features	13 static features	20 static features: Adds glacier fraction and riparian land-cover features

Out-of-fold performance across hydrologic and landscape settings

Comparison of all out-of-fold test sites; lower median MAE indicates better performance.



Reproduce the workflow

1. Clone the repository

```
git clone https://code.ornl.gov/u96/rithyms-v1.1.git
```

Required input files are available through HydroSource. Download RiTHyMs_v1-1_inputs.zip and extract it at the repository root, preserving its inputs/ directory structure, before running the workflow.

2. Create the pinned environment

Install uv, then run:

```
uv sync --locked
```

The project requires Python 3.12. The locked environment uses the CUDA 12.8 PyTorch wheel. A compatible NVIDIA driver and GPU are recommended for practical training time, but CPU execution is supported.

3. Run the full pipeline

For an NVIDIA GPU:

```
uv run python scripts/0_run_reproducible_distance_weighted_pipeline.py --gpu 0
```

For CPU execution:

```
uv run python scripts/0_run_reproducible_distance_weighted_pipeline.py --gpu -1
```

The pipeline performs the following steps:

1. Delineates each outlet's dissolved upstream watershed and calculates static attributes.
2. Downloads and aggregates daily Daymet and Dayflow forcings to create distance-weighted model inputs.
3. Creates six hydroclimatic clusters and ten spatial train/validation/test folds.
4. Trains and evaluates one independent LSTM model per fold.

Generated watershed, attribute, time-series, and fold files are written under `inputs/` and `folds/`. Timestamped cross-validation outputs, fold weights, metrics, and logs are written under `runs/`.

Repository layout

Path	Contents
`inputs/rithyms_v11_latest_run_site_inventory.csv`	Model outlets, gage IDs, locations, data coverage, and train/test eligibility.
`inputs/nhdplusv2/`	NHDPlusV2 flowline and catchment source data.
`inputs/nlcd/`	StreamCat static-descriptor table.
`inputs/glacier/`	Glacier, perennial snowfield, and buried-ice inventory.
`inputs/reference/`	Dayflow COMID-to-HUC8 linkage.
`inputs/stream_temperature/`	Screened daily mean stream-temperature observations.
`config_files/`	Frozen configuration file for each cross-validation fold.
`scripts/`	Watershed, feature, fold, and LSTM workflow scripts.

Scope and limitations

RiTHyMs estimates natural or near-natural stream-temperature conditions. It should not be interpreted as a direct model of releases from dams, hydropower facilities, canals, springs, tidal reaches, cooling-water discharges, or other local thermal modifications unless those processes are represented separately. Model performance can vary where observations are sparse or where thermal controls are not represented by the nationally available inputs.

Data sources

RiTHyMs v1.1 uses NHDPlusV2 hydrography, EPA StreamCat descriptors, Daymet gridded meteorological forcing, Dayflow modeled streamflow, glacier inventory data, and screened multi-agency stream-temperature observations. Please follow the citation, license, and redistribution requirements of each original source when publishing results or redistributing source data.

Citation

For the underlying RiTHyMs framework, cite:

> Siddik, M. A. B., Turner, S., Gomez-Velez, J., Tanim, A., and Kao, S.-C. (2026). Advancing stream temperature prediction with a generalizable large-sample framework across CONUS river reaches. *Journal of Hydrology*, 666, 134780. https://doi.org/10.1016/j.jhydrol.2025.134780

Also cite the RiTHyMs v1.1 data release:

> Md Abu Bakar Siddik, Sean Turner and Shih-Chieh Kao. 2026. RiTHyMs: Foundational Dataset for Developing Large-Sample Stream Temperature Models in the Conterminous United States, Version 1.1. HydroSource. Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA. https://doi.org/10.21951/RiTHyMs_V1/17239901

Contact and license

For questions, open a GitLab issue or contact: AB SIddik - siddikm@ornl.gov.