UrbanFloodBench Full Dataset README

Full event simulation outputs for coupled 1D-2D urban flood benchmark models

Version: 1.0

Important dataset-scope note. This repository is intended to host the full UrbanFloodBench simulation dataset. Unlike the Kaggle competition release, this full repository provides complete event simulation results for all four models and does not apply the competition-specific temporal truncation or warm-up/ending cuts used for the leaderboard task. The data should be interpreted as HEC-RAS reference simulation outputs from controlled synthetic-rainfall experiments, not as calibrated observations of real flood events.

1. Overview

UrbanFloodBench is a controlled benchmark dataset for coupled 1D-2D urban flood simulation emulation. The dataset contains four rain-on-grid HEC-RAS hydrodynamic models. In each model, the subsurface drainage system is represented as a 1D network and overland flow is represented on a 2D computational grid. The dataset is designed to support reproducible research on data-driven and surrogate modelling methods for coupled urban flood systems. For scientific context, users should refer to the accompanying paper by Lim et al. (2026).

Hydrodynamic modelling software: HEC-RAS River Analysis System, version 6.7 Beta 4a (USACE HEC, 2025a).

Dataset type: Synthetic design-storm HEC-RAS reference simulations; complete model outputs are provided at 5-minute intervals.

Spatial domains: 1D drainage network nodes/links, 2D surface grid cells/links, and 1D-2D coupling connections.

Rainfall forcing: Spatially uniform rain-on-grid input with temporally varying synthetic rainfall patterns.

2. How to cite this dataset

Please cite the accompanying paper when using this dataset:

- Lim, Jia Yu and Herath, Herath Mudiyanselage Viraj Vidura and Rasnayaka, Sanka and Marshall, Lucy and Zou, Hui and Saha, Abhishek and Hossain, Jobayer and Tong, Jing Yen and Lee, Ing Zhen and Tsukada, Yukiya and Gandhi, Shrey and Matsumaro, Katsutoshi and Raj, Sajay and Mazur, Artyom and Khlopotnykh, Andrey and Cheshenko, Dmytro, UrbanFloodBench: Bridging AI and Hydrology through Benchmarking of Coupled 1D-2D Urban Flood Surrogate Models. Available at SSRN: https://ssrn.com/abstract=6900753 or http://dx.doi.org/10.2139/ssrn.6900753

If referring to the Kaggle competition release, please also cite:

- Jia Yu Lim, Sanka Rasnayaka, Viraj Vidura Herath Herath Mudiyanselage, Lucy Marshall, Hui Zou, and Abhishek Saha. UrbanFloodBench: Flood Modelling. https://kaggle.com/competitions/urban-flood-modelling, 2026. Kaggle.

3. Dataset scope and limitations

- **Controlled HEC-RAS simulations:** The dataset contains outputs from controlled hydrodynamic simulations rather than observed flood events.

- **Synthetic rainfall:** Rainfall inputs are synthetic design storms with spatially uniform rain-on-grid forcing and temporally varying rainfall patterns.
- **Simplified models:** The first three models are adapted from HEC-RAS example models, while the fourth is a real-world-inspired Coogee, NSW model. Hydraulic controls such as pumps and levees were excluded to simplify system dynamics.
- **No event calibration:** The models were not calibrated to reproduce observed events and should not be interpreted as representing actual conditions at their named locations.
- **Full-event release:** This repository provides full event simulation outputs without the competition-specific temporal truncation used in the Kaggle leaderboard dataset.
- **Units:** Models 1-3 use US customary units, while Model 4 uses SI units. Users should handle units explicitly when combining models (users can refer to the unit conversion scripts given in the online repository if needed https://github.com/Pyrexiaa/urbanfloodbench-2026).

4. Recommended repository organization

Each model is organised into three main data groups: dynamic files, static files, and geospatial shapefiles. Due to individual file size limitations, Model 03 and Model 04 simulations are divided into several parts.

```
UrbanFloodBench Dataset

  1.  README.pdf

  2.  Model_01.zip/

    Dynamic_files/

    Static_files/

    Shapefiles/

  3.  Model_02.zip/

    Dynamic_files/

    Static_files/

    Shapefiles/

  4.  Model_03_Part_1.zip/

    Dynamic_files (Part 1)/

    Static_files/

    Shapefiles/

  5.  Model_03_Part_2.zip/

    Dynamic_files (Part 2)/

  6.  Model_03_Part_3.zip/

    Dynamic_files (Part 3)/

  7.  Model_03_Part_4.zip/

    Dynamic_files (Part 4)/
```

8. Model_03_Part_5.zip/

   Dynamic_files (Part 5)/

9. Model_03_Part_6.zip/

   Dynamic_files (Part 6)/

10. Model_03_Part_7.zip/

    Dynamic_files (Part 7)/

11. Model_03_Part_8.zip/

    Dynamic_files (Part 8)/

12. Model_03_Part_9.zip/

    Dynamic_files (Part 9)/

13. Model_03_Part_10.zip/

    Dynamic_files (Part 10)/

14. Model_04_Part_1.zip/

    Dynamic_files (Part 1)/

    Static_files/

    Shapefiles/

15. Model_04_Part_2.zip/

    Dynamic_files (Part 2)/

- **Dynamic files:** Time-varying simulation outputs and rainfall forcing recorded at 5-minute intervals.
- **Static files:** Node and edge attributes describing the 1D drainage network and 2D surface grid.
- **Shapefiles/geospatial files:** Spatial topology and geometry files for 1D nodes, 2D nodes/cells, 1D links, 2D links, and 1D-2D coupling connections.

5. Simulation and rainfall generation

- **Hydrodynamic setup:** All four models are rain-on-grid, coupled 1D-2D HEC-RAS hydrodynamic systems. The 1D domain represents subsurface drainage flow and the 2D domain represents overland surface flow.
- **Governing equations:** The governing equations for both the 1D and 2D domains are based on the shallow water equations.
- **Time stepping:** The 2D computations use fixed time steps: 0.5 s for Models 01 and 04, 1 s for Model 02, and 4 s for Model 03. Adaptive time stepping is used for the 1D domain, with target Courant number 0.90.
- **Output interval:** Hydrodynamic outputs are recorded at 5-minute intervals.
- **Design storms:** For each model, approximately 100 synthetic design rainfall events were generated using return periods of 5-year, 20-year, and 100-year and durations of 3 h, 12 h, and 24 h (please note the simulation times are longer than rainfall durations with warm up and after event run time).
- **Rainfall intensity sources:** IDF curves for Models 01-03 were obtained from the NOAA Precipitation Frequency Data Server (PFDS). IDF curves for Model 04 were obtained from the Australian Bureau of Meteorology Design Rainfall Data System.

- **Temporal patterns:** Synthetic temporal patterns were obtained from the Australian Rainfall and Runoff Data Hub.
- **Ghost cells:** HEC-RAS ghost cells at domain boundaries were removed from the dataset, except for cells corresponding to true 2D boundary conditions.

6. Model summary

Model	Name	Source	Units and events	Area and terrain	Hydraulic domains	Downstream boundary
01	Small Neighborhood Drainage - Beaver Lake	HEC-RAS example model; simplest benchmark configuration	US customary; 97 events	Area: 0.21 km2; terrain: 293-360 ft	1D: 17 nodes, including 1 boundary node; 2D: 3,704 internal cells + 12 boundary nodes	2D constant stage 0 ft; 1D normal depth
02	Small City Urban Drainage - Davis, CA	HEC-RAS example model; pump structures removed	US customary; 99 events	Area: 6.53 km2; terrain: 35-58 ft	1D: 198 nodes, including 1 boundary node; 2D: 4,299 cells	Outflow through 1D boundary node; normal depth
03	Metro Area Drainage (Skeletonized) - New Orleans	HEC-RAS example model; pump structures removed	US customary; 97 events	Area: 63.2 km2; terrain: -21 to 22 ft	1D: 65 nodes, including 1 boundary node; 2D: 19,711 cell nodes + 13 boundary nodes	2D constant stage -18 ft; 1D normal depth
04	Coogee, NSW, Australia	Real-world-inspired model; 1D network adapted from Wu et al. (2023); topography from ELVIS	SI; 85 events	Area: 3.56 km2; terrain: -0.5 to 84 m	1D: 293 nodes, all internal; 2D: 8,801 cells + 104 boundary nodes	2D sea boundary; constant stage 1 m

7. Model descriptions

Model 01: Small Neighborhood Drainage - Beaver Lake

This is the simplest configuration in the benchmark. It combines a small 1D pipe network with a 2D surface domain draining toward a downstream lake. It is intended as a baseline case for initial testing and validation of surrogate methods.

Model 02: Small City Urban Drainage - Davis, CA

This medium-scale urban catchment includes a larger interconnected 1D drainage network coupled to a 2D surface domain. Pump structures from the original example were removed to simplify the hydraulic configuration.

Model 03: Metro Area Drainage (Skeletonized) - New Orleans

This is the largest spatial domain in the benchmark. It contains a substantial 2D surface grid coupled with a moderately sized 1D drainage network. Pump structures were removed to simplify the system.

Model 04: Coogee, NSW, Australia

This real-world-inspired catchment is based on a Coogee, NSW drainage network adapted from Wu et al. (2023), with topography derived from the ELVIS elevation dataset. It has steep terrain and a dense 1D drainage network discharging toward the ocean.

Please refer to the accompanying paper for more details.

8. Data variables and units

The dataset is structured in a node-and-link graph format that reflects the coupled HEC-RAS topology. Static variables describe geometry, topography, roughness, and connectivity. Dynamic variables describe time-varying rainfall and hydraulic outputs. Signed dynamic quantities follow the from-node to to-node direction convention for each link type. Models 01-03 use US customary units and Model 04 uses SI units.

Domain	Type	Feature	US units	SI units	Description
1D node	Static	Position (x, y)	ft	m	Planimetric coordinates
1D node	Static	Depth	ft	m	Surface-to-invert distance
1D node	Static	Invert elevation	ft	m	Lowest internal node elevation
1D node	Static	Surface elevation	ft	m	Ground elevation at node
1D node	Static	Base area	ft2	m2	Node storage cross-section
1D edge	Static	Length	ft	m	Physical pipe length
1D edge	Static	Relative position (x, y)	ft	m	Downstream minus upstream offset
1D edge	Static	Diameter	ft	m	Internal or hydraulic diameter
1D edge	Static	Shape	-	-	Cross-section shape, e.g., circular
1D edge	Static	Roughness	s ft^-1/3	s m^-1/3	Manning's n
1D edge	Static	Slope	-	-	Longitudinal bed slope
1D node	Dynamic	Water level	ft	m	Junction water surface elevation
1D node	Dynamic	Inlet flow (1D-2D)	ft3/s	m3/s	Signed coupling exchange flow
1D edge	Dynamic	Flow; velocity	ft3/s; ft/s	m3/s; m/s	Signed pipe discharge and velocity
2D node	Static	Position (x, y)	ft	m	Cell centroid coordinates
2D node	Static	Area	ft2	m2	Surface cell area
2D node	Static	Roughness	s ft^-1/3	s m^-1/3	Manning's n
2D node	Static	Minimum elevation	ft	m	Minimum terrain elevation in cell
2D node	Static	Centroid elevation	ft	m	DEM elevation at cell centroid
2D node	Static	Aspect	degrees	degrees	Steepest slope direction
2D node	Static	Curvature	ft^-1	m^-1	Surface concavity/convexity
2D node	Static	Flow accumulation	-	-	Upstream contributing area index
2D edge	Static	Relative position (x, y)	ft	m	Adjacent centroid offset
2D edge	Static	Face length	ft	m	Shared cell boundary length
2D edge	Static	Length	ft	m	Centroid-to-centroid distance
2D edge	Static	Slope	-	-	Local terrain slope
2D node	Dynamic	Rainfall depth	in	mm	Applied rainfall per time step
2D node	Dynamic	Water level	ft	m	Inundation water surface elevation
2D node	Dynamic	Water volume	ft3	m3	Stored water volume per cell
2D edge	Dynamic	Flow; velocity	ft3/s; ft/s	m3/s; m/s	Signed inter-cell flow and velocity
All	Geospatial	Shapefiles	-	-	Spatial topology and terrain reference

9. File group descriptions

9.1 Dynamic files

- Contain time-varying simulation outputs for each event at 5-minute intervals.
- Include 1D and 2D water levels, 2D water volumes, 1D/2D/1D-2D discharge, flow velocity, and rainfall depth where applicable.
- Dynamic files are the primary source for reconstructing simulated hydraulic states through time.
- This full repository provides complete event trajectories and does not apply the temporal truncation used in the Kaggle competition dataset.

9.2 Static files

- Contain 1D node, 1D edge, 2D node, and 2D edge attributes.
- Static 1D node attributes include node coordinates, depth, invert elevation, surface elevation, and base area.
- Static 1D edge attributes include pipe length, position offsets, diameter, shape, roughness, and slope.
- Static 2D node attributes include cell coordinates, area, roughness, minimum elevation, centroid elevation, aspect, curvature, and flow accumulation.
- Static 2D edge attributes include relative position, face length, centroid-to-centroid length, and local terrain slope.

9.3 Shapefiles and geospatial files

- Include spatial reference files for 1D nodes, 2D nodes/cells, 1D links, 2D links, and 1D-2D coupling connections.
- These files encode the spatial topology and connectivity of the coupled hydraulic model.
- They are provided for visualisation, spatial analysis, feature engineering, and graph construction.

10. Relationship to the Kaggle competition data

The Kaggle competition release and this full repository serve different purposes. The Kaggle competition dataset contained training inputs and held-out test inputs for Models 01 and 02, with test target values withheld during the competition. A separate reproducibility release contains the complete competition targets for Models 01 and 02. This full repository is intended to provide complete simulation outputs for all four models and all events, including Models 03 and 04, without competition-specific temporal truncation.

Release	Models included	Main contents	Purpose
Competition dataset	Models 01-02	Training inputs/targets and held-out test inputs without target data	Kaggle competition use
Competition reproducibility dataset	Models 01-02	Training data plus held-out test target data	Reproduce the competition evaluation
Full benchmark dataset	Models 01-04	Complete event simulation outputs for all models	Broader benchmarking, cross-model analysis, and future reproducibility

11. Suggested use and interpretation

- **Use as HEC-RAS reference simulations:** Treat the outputs as reference solutions generated by HEC-RAS under controlled synthetic rainfall scenarios.
- **Do not treat as observations:** The simulations are not calibrated to observed events and should not be described as observed ground truth.
- **Handle units carefully:** Models 01-03 use US customary units and Model 04 uses SI units.
- **Use full events for new experiments:** Because this repository contains complete event trajectories, users can define their own train/validation/test splits, warm-up handling, event truncation, and target variables.
- **Refer to the paper:** The accompanying paper provides the scientific motivation, benchmark design, and competition context. This README focuses on the dataset itself.

12. Contact

Jia Yu Lim	lim.jia.yu.desmond@gmail.com
Viraj Vidura Herath	viraj.herath@sydney.edu.au
Sanka Rasnayaka	sanka@nus.edu.sg

References

1. Australian Bureau of Meteorology, 2026. Revised Intensity-Frequency-Duration (IFD) Design Rainfall Data System. URL: https://www.bom.gov.au/water/designRainfalls/revised-ifd/. Accessed: 2026-10-25.
2. Babister, M., Trim, A., Testoni, I., Retallick, M., 2016. The Australian Rainfall & Runoff Datahub, 37th Hydrology and Water Resources Symposium, Queenstown, New Zealand.
3. Geoscience Australia, 2021. ELVIS - Elevation and Depth - Foundation Spatial Data. URL: https://elevation.fsdf.org.au/. Accessed: 2024-09-04.
4. Jia Yu Lim, Sanka Rasnayaka, Viraj Vidura Herath Herath Mudiyanselage, Lucy Marshall, Hui Zou, and Abhishek Saha. UrbanFloodBench: Flood Modelling. https://kaggle.com/competitions/urban-flood-modelling, 2026. Kaggle.
5. Lim, Jia Yu and Herath, Herath Mudiyanselage Viraj Vidura and Rasnayaka, Sanka and Marshall, Lucy and Zou, Hui and Saha, Abhishek and Hossain, Jobayer and Tong, Jing Yen and Lee, Ing Zhen and Tsukada, Yukiya and Gandhi, Shrey and Matsumaro, Katsutoshi and Raj, Sajay and Mazur, Artyom and Khlopotnykh, Andrey and Cheshenko, Dmytro., 2026. UrbanFloodBench: Bridging AI and Hydrology through Benchmarking of Coupled 1D-2D Urban Flood Surrogate Models. URL: http://dx.doi.org/10.2139/ssrn.6900753.
6. National Oceanic and Atmospheric Administration, 2026. Precipitation Frequency Data Server (PFDS). URL: https://hdsc.nws.noaa.gov/pfds/. Accessed: 2025-10-01.
7. US Army Corps of Engineers, Hydrologic Engineering Center, 2025a. HEC-RAS River Analysis System, version 6.7 Beta 4a. URL: https://www.hec.usace.army.mil/software/hec-ras/download.aspx. Accessed: 2025-09-20.
8. US Army Corps of Engineers, Hydrologic Engineering Center, 2025b. Metro Area Drainage (Skeletonized) - New Orleans. URL: https://www.hec.usace.army.mil/confluence/rasdocs/hgt/latest/sample-datasets/metro-area-drainage-skelentonzied-new-orleans. Accessed: 2026-04-13.
9. US Army Corps of Engineers, Hydrologic Engineering Center, 2025c. Small City Urban Drainage - Davis, CA. URL: https://www.hec.usace.army.mil/confluence/rasdocs/hgt/latest/sample-datasets/small-city-urban-drainage-davis-ca. Accessed: 2026-04-13.
10. US Army Corps of Engineers, Hydrologic Engineering Center, 2025d. Small Neighborhood Drainage - Beaver Lake. URL: https://www.hec.usace.army.mil/confluence/rasdocs/hgt/latest/sample-datasets/small-neighborhood-drainage-beaver-lake. Accessed: 2026-04-13.
11. Wu, W., Jamali, B., Zhang, K., Marshall, L., Deletic, A., 2023. Water sensitive urban design (WSUD) spatial prioritisation through global sensitivity analysis for effective urban pluvial flood mitigation. Water Research 235, 119888. doi:https://doi.org/10.1016/j.watres.2023.119888.