Dataset for training and testing in *"Physics-informed Graph Neural Networks for Operational Flood Modeling"*

Authors: Carlo Malapad Acosta, Herath Mudiyanselage Viraj Vidura Herath, Jia Yu Lim, Abhishek Saha, Sanka Rasnayaka, Lucy Marshall

Contact: viraj.herath@sydney.edu.au

1. Summary

This dataset was used in the paper *"Physics-informed Graph Neural Networks for Operational Flood Modeling"* [1]. This is comprised of dataset summary files for train and test splitting, shape files for the mesh structure, a DEM file for elevation-related features and HEC-RAS simulation files for simulated flood events. The target catchment for this study was taken from a section of the Wollombi River watershed located in New South Wales, Australia. An unstructured mesh and 56 flow-dominant flood events were generated using the HEC-RAS hydraulic modelling software developed by the US Army Corps of Engineers [2]. The DEM for the target catchment was obtained from the ELVIS – Elevation and Depth – Foundation Spatial Data portal [3]. Further methodological details and guidance on dataset usage can be found in the associated publication.

2. Files

The main folder of the dataset contains this README.pdf file, and dataset summary files (train.csv and test.csv). The dataset summary files are used to determine which flood events should be used for training and testing.

MAIN FOLDER:

File Name	Description	File Type
README.pdf	Overview of dataset	PDF
train.csv	Dummy train dataset summary file	CSV
test.csv	Dummy test dataset summary file	CSV

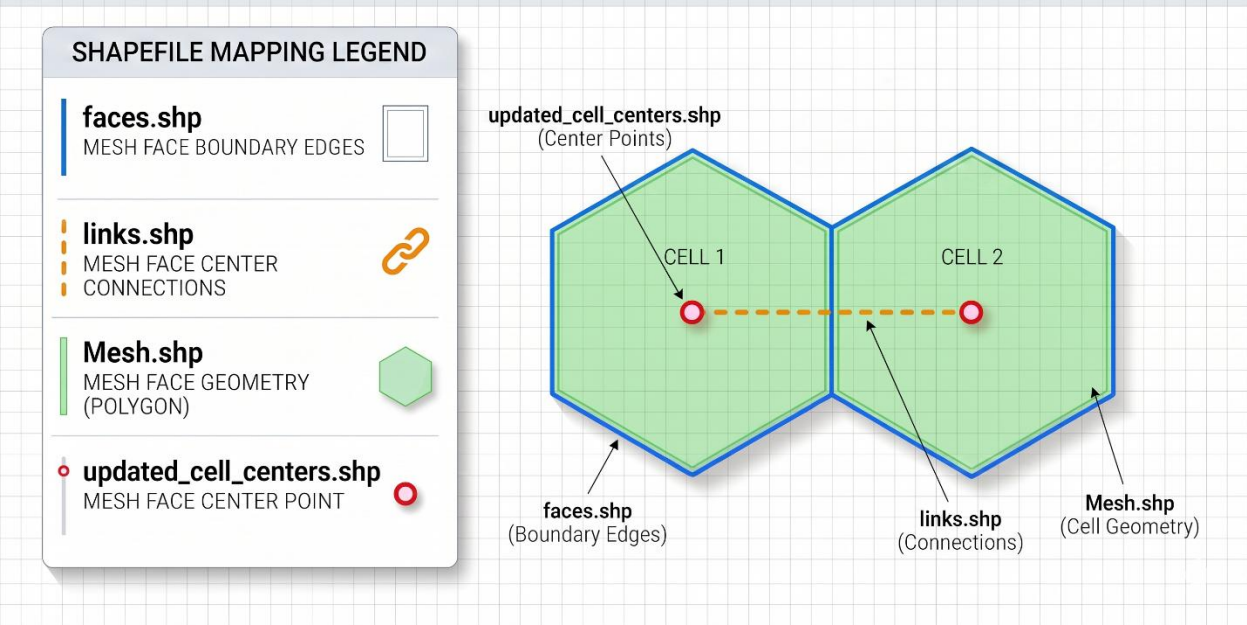
Aside from these files, it contains two subdirectories: GEOMETRY and HDF_FILES.

GEOMETRY: This folder contains static information such as the DEM and mesh structure. The mesh is created by discretizing the Wollombi catchment through the HEC-RAS hydraulic modeling software [2]. Following the ESRI Shapefile format [4], files describing the mesh are comprised of a collection of 5 file types: SHP, SHX, dBASE, PRJ and CPG. The mesh created in this study has 1129 mesh cells and 2473 mesh cell center links.

File Name	Description	File Type
DEM.tif	Pixel-wise DEM of Wollombi catchment	TIF
faces.shp	Mesh face boundary edges (main shape file)	SHP

faces.shx	Mesh face boundary edges (shape index file)	SHX
faces.dbf	Mesh face boundary edges (attribute table)	dBASE
faces.prg	Mesh face boundary edges (projection file)	PRJ
faces.cpg	Mesh face boundary edges (codepage file file)	CPG
links.shp	Mesh face center connection (main shape file)	SHP
links.shx	Mesh face center connection (shape index file)	SHX
links.dbf	Mesh face center connection (attribute table)	dBASE
links.prg	Mesh face center connection (projection file)	PRJ
links.cpg	Mesh face center connection (codepage file file)	CPG
Mesh.shp	Mesh face geometry (main shape file)	SHP
Mesh.shx	Mesh face geometry (shape index file)	SHX
Mesh.dbf	Mesh face geometry (attribute table)	dBASE
Mesh.prg	Mesh face geometry (projection file)	PRJ
updated_cell_centers.shp	Mesh face centers (main shape file)	SHP
updated_cell_centers.shx	Mesh face centers (shape index file)	SHX
updated_cell_centers.dbf	Mesh face centers (attribute table)	dBASE
updated_cell_centers.prg	Mesh face centers (projection file)	PRJ
updated_cell_centers.cpg	Mesh face centers (codepage file file)	CPG

Below is a diagram visualizing the corresponding structural property described by each shapefile.



HDF_FILES: This folder contains the simulated flood events using the HEC-RAS hydraulic modelling software [2]. The simulation was performed on the mesh described in the GEOMETRY folder, with one edge for inflow boundary conditions and two edges for outflow boundary conditions. A total of 56 flood events were generated, each with 2593 timesteps at a 5-minute interval.

File Name	Description	File Type
Model_01.p22.hdf	HEC-RAS simulation (Run ID 17)	HDF
Model_01.p23.hdf	HEC-RAS simulation (Run ID 18)	HDF
Model_01.p24.hdf	HEC-RAS simulation (Run ID 19)	HDF
Model_01.p25.hdf	HEC-RAS simulation (Run ID 20)	HDF
Model_01.p26.hdf	HEC-RAS simulation (Run ID 21)	HDF
Model_01.p27.hdf	HEC-RAS simulation (Run ID 22)	HDF
Model_01.p28.hdf	HEC-RAS simulation (Run ID 23)	HDF
Model_01.p29.hdf	HEC-RAS simulation (Run ID 24)	HDF
Model_01.p30.hdf	HEC-RAS simulation (Run ID 25)	HDF
Model_01.p31.hdf	HEC-RAS simulation (Run ID 26)	HDF
Model_01.p32.hdf	HEC-RAS simulation (Run ID 27)	HDF
Model_01.p33.hdf	HEC-RAS simulation (Run ID 28)	HDF
Model_01.p34.hdf	HEC-RAS simulation (Run ID 29)	HDF
Model_01.p35.hdf	HEC-RAS simulation (Run ID 30)	HDF
Model_01.p36.hdf	HEC-RAS simulation (Run ID 31)	HDF
Model_01.p37.hdf	HEC-RAS simulation (Run ID 32)	HDF
Model_01.p38.hdf	HEC-RAS simulation (Run ID 33)	HDF
Model_01.p39.hdf	HEC-RAS simulation (Run ID 34)	HDF
Model_01.p40.hdf	HEC-RAS simulation (Run ID 35)	HDF
Model_01.p41.hdf	HEC-RAS simulation (Run ID 36)	HDF
Model_01.p42.hdf	HEC-RAS simulation (Run ID 37)	HDF
Model_01.p43.hdf	HEC-RAS simulation (Run ID 38)	HDF
Model_01.p44.hdf	HEC-RAS simulation (Run ID 39)	HDF
Model_01.p45.hdf	HEC-RAS simulation (Run ID 40)	HDF
Model_01.p46.hdf	HEC-RAS simulation (Run ID 41)	HDF
Model_01.p47.hdf	HEC-RAS simulation (Run ID 42)	HDF
Model_01.p48.hdf	HEC-RAS simulation (Run ID 43)	HDF
Model_01.p49.hdf	HEC-RAS simulation (Run ID 44)	HDF
Model_01.p50.hdf	HEC-RAS simulation (Run ID 45)	HDF
Model_01.p51.hdf	HEC-RAS simulation (Run ID 46)	HDF
Model_01.p52.hdf	HEC-RAS simulation (Run ID 47)	HDF
Model_01.p53.hdf	HEC-RAS simulation (Run ID 48)	HDF
Model_01.p54.hdf	HEC-RAS simulation (Run ID 49)	HDF
Model_01.p55.hdf	HEC-RAS simulation (Run ID 50)	HDF
Model_01.p56.hdf	HEC-RAS simulation (Run ID 51)	HDF
Model_01.p57.hdf	HEC-RAS simulation (Run ID 52)	HDF
Model_01.p58.hdf	HEC-RAS simulation (Run ID 53)	HDF
Model_01.p59.hdf	HEC-RAS simulation (Run ID 54)	HDF
Model_01.p60.hdf	HEC-RAS simulation (Run ID 55)	HDF
Model_01.p61.hdf	HEC-RAS simulation (Run ID 56)	HDF
Model_01.p62.hdf	HEC-RAS simulation (Run ID 57)	HDF
Model_01.p63.hdf	HEC-RAS simulation (Run ID 58)	HDF
Model_01.p64.hdf	HEC-RAS simulation (Run ID 59)	HDF
Model_01.p65.hdf	HEC-RAS simulation (Run ID 60)	HDF

Model_01.p66.hdf	HEC-RAS simulation (Run ID 61)	HDF
Model_01.p67.hdf	HEC-RAS simulation (Run ID 62)	HDF
Model_01.p68.hdf	HEC-RAS simulation (Run ID 63)	HDF
Model_01.p69.hdf	HEC-RAS simulation (Run ID 64)	HDF
Model_01.p70.hdf	HEC-RAS simulation (Run ID 65)	HDF
Model_01.p71.hdf	HEC-RAS simulation (Run ID 66)	HDF
Model_01.p72.hdf	HEC-RAS simulation (Run ID 67)	HDF
Model_01.p73.hdf	HEC-RAS simulation (Run ID 68)	HDF
Model_01.p74.hdf	HEC-RAS simulation (Run ID 69)	HDF
Model_01.p75.hdf	HEC-RAS simulation (Run ID 70)	HDF
Model_01.p76.hdf	HEC-RAS simulation (Run ID 71)	HDF
Model_01.p77.hdf	HEC-RAS simulation (Run ID 72)	HDF

3. Units

The table below describes the units and data source of important features used in the dataset.

Feature	**Data source**	**Unit**
Elevation	DEM	m
Water Volume	HEC-RAS simulation	m^3
Rainfall	HEC-RAS simulation	mm
Water Face Flow	HEC-RAS simulation	m^3/s

4. Usage

This dataset can be used with the code repository for this paper. To do so, place all of the files in the main folder under the data/datasets/raw folder. This is the default dataset location but other paths may be used by defining them in your config.yaml file.

References

[1] Acosta, C.M., Herath, H.M.V.V., Lim, J.Y., Saha, A., Rasnayaka, S., Marshall, L., 2025. Physics-informed graph neural networks for operational flood modeling. URL: https://arxiv.org/abs/2512.23964, arXiv:2512.23964.

[2] Gary W. Brunner. HEC-RAS Hydraulic Reference Manual. Hydrologic Engineering Center, Institute for Water Resources, U.S. Army Corps of Engineers, 2025.

[3] Geoscience Australia, 2021. ELVIS – Elevation and Depth – Foundation Spatial Data. Commonwealth of Australia. Available at: https://elevation.fsdf.org.au/ (accessed 4 September 2024).

[4] Environmental Systems Research Institute, 1998. ESRI Shapefile Technical Description. ESRI White Paper J-7855. Environmental Systems Research Institute, Inc. Redlands, CA, USA. URL: https://www.esri.com/content/dam/esrisites/sitecore-archive/Files/Pdfs/library/whitepapers/pdfs/shapefile.pdf