Supplementary Information for Large-scale flood risk analysis of distributed Storm-water infrastructure serving 2 300 catchments in New York State

Omid Emamjomehzadeh[1*] and Omar Wani[1*]

[1*]Tandon School of Engineering, New York University, 6 Metro-tech Center, Brooklyn, New York, 11201, United States.

*Corresponding author(s). E-mail(s): omid.emamjomehzadeh@nyu.edu; omarwani@nyu.edu;

1 Terabytes of geospatial datasets

Table 1 summarizes the geospatial datasets employed in the analysis, including their names, references, and specific use case.

Table 1: Geospatial datasets used in this study and their usage

Name	Reference	Usage
Culvert inventory	[1]	Used to obtain culvert identification number, latitude, longitude, region, year built, number of spans, maximum span length, road type, design type, and material.
1-m digital elevation model (DEM)	[2]	Used for culverts' watershed delineation and morphological feature extraction (coverage shown in Fig. 1).
10-m DEM	[3]	Used for culverts' watershed delineation and morphological feature extraction where a 1-m DEM coverage, shown in Fig. 1, is unavailable.
Hydrologic unit code 12 (HUC 12)	[4]	Parallelization units for delineation and morphological-feature extraction algorithms.
Streets	[5]	Used to find road–stream crossings, and burn the DEM at those locations for a realistic stream network generation.
Railroads	[6]	Used to find railroads–stream crossings, and burn the DEM at those locations for a realistic stream network generation.
Streams	[7]	Used to find (rail)road–stream crossings, and burn the DEM at those locations for a realistic stream network generation.
Historical precipitation intensity duration frequency analysis	[8]	Used to get business-as-usual extreme precipitation depths for various durations and return periods.
Projected precipitation intensity duration frequency analysis	[9, 10]	Used to get projected extreme precipitation depths for various durations and return periods.
Historical land cover	[11]	Used to generate current curve number maps.
Projected land cover	[12]	Used to generate projected curve number maps.
Hydrologic soil group	[13]	Used to generate curve number maps.
Curve number map	[14]	Used to compare the curve number maps with another product.
USGS discharge gages: drainage area and time series	[15]	Used for the evaluation of the watershed delineation algorithm and peak-discharge algorithms' accuracy.

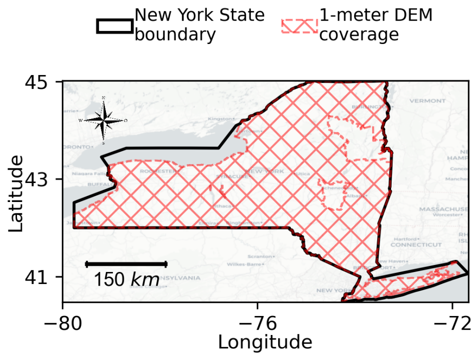

Fig. 1: Coverage of the 1-meter digital elevation model provided by *OpenTopography* in New York State.

2 Selecting USGS culvert-like stations for drainage area and peak runoff validation

To assess the accuracy of our watershed-delineation algorithm, we identified United States Geological Survey (USGS) streamflow-gaging stations in New York State with small drainage areas less than 2 mi^2, comparable to those of culverts. From this set, we selected stations within 20 m of a mapped stream and a road to approximate culvert-like settings, where small drainage areas coincide with road–stream intersections. The USGS and our scalable algorithm independently delineated watershed areas for the selected stations, and the two delineations were compared (Fig. 1, panel **b** of the main manuscript).

Long-term discharge records from culverts would be ideal to validate the hydrologic model's peak discharges, but such data are unavailable. As a proxy, we selected USGS streamflow-gaging stations in New York State with drainage areas less than 2 mi^2 and record lengths of more than 15 years. The latter criterion ensures that empirical estimates of the 2- and 5-year discharges—computed following the procedure in Appendix A of the main manuscript—have sufficiently narrow uncertainty for model validation. Five stations met these criteria (site numbers 01362342, 01422389, 04253294, 04253296, and 04253295) and were used as validation sites.

3 Selection process of confident batch of culverts

Several factors influence delineation accuracy and can introduce systematic bias into culverts' exceedance probabilities. We categorized the 8 062 culverts by delineation accuracy (Fig. 2). For benchmarking, delineations were also produced with the USGS *StreamStats* batch processing tool [16], which successfully delineated 6 351 culverts, compared with 6 465 delineated by our method. Culverts that our algorithm could not delineate were labeled "not analyzed," owing either to unsuccessful acquisition of high-resolution digital elevation models or to failure of rationality checks applied to watershed characteristics.

Among culverts delineated by both methods, 1 526 exhibited a normalized drainage-area difference more than 25% and were therefore classified as "large difference with StreamStats." To improve reliability, we excluded culverts in urbanized watersheds—defined as having more than 25% developed land cover in the National Land Cover Database (classes "developed–open space," "developed–low intensity," "developed–medium intensity," and "developed–high intensity")—because subsurface drainage not represented in our framework could bias results. This criterion eliminated 2 096 culverts ("urbanized"). We also removed 425 culverts with delineated drainage areas bigger than 10 km^2, as many had snapped to disproportionately large nearby streams ("too big"). After these filters, 2 418 culverts remained in the high-confidence set for subsequent analyses ("confident batch").

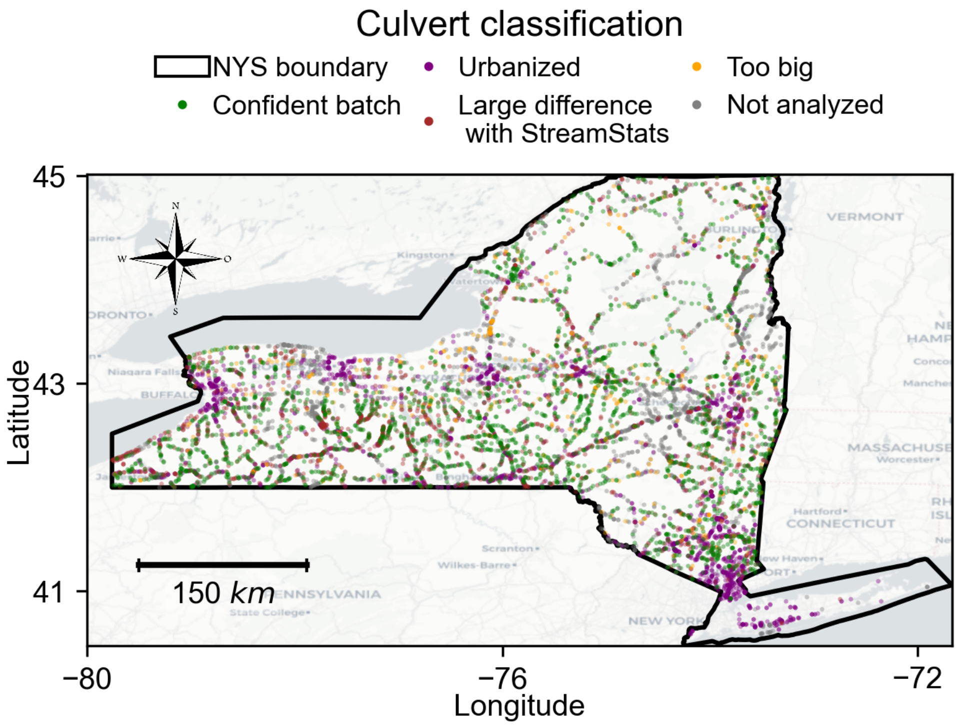

Fig. 2: Culverts and their color-dotted delineation accuracy category.

4 Maximum and allowable headwater

In the hydraulic-capacity calculation (Equation 4 in the main manuscript), the headwater, HW, is taken as either the maximum physically attainable headwater or the design-allowable headwater. The maximum headwater, $HW_{\max}$, is defined as $HW_{max} = D + Depth$, where D is the culvert rise and Depth is the fill thickness of the cover above the culvert crown. This represents the maximum potential headwater that can build up at the culvert inlet. According to the New York State Department of Transportation Highway Design Manual [17], design criteria limit the allowable headwater at the culvert upstream as a function of the culvert rise. Accordingly, the allowable headwater, HW_{Allow}, is computed using (1), with units in feet. The cumulative probability distribution functions of the maximum and allowable headwaters for the studied culverts are shown in Fig. 3.

$$HW_{\text{Allow}} = \begin{cases} \min\,(1.5\,D, HW_{max}), & 0 < D \leq 4 \\ \min\,(1.2\,D, HW_{max}), & 4 < D \leq 5 \\ \min\,(D, HW_{max}), & 5 < D \end{cases} \tag{1}$$

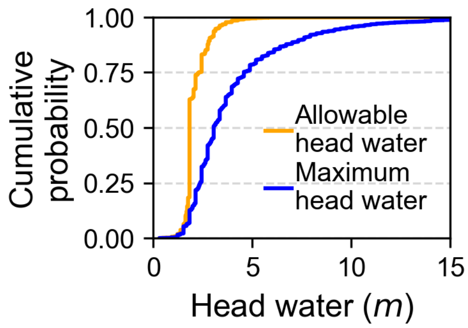

Fig. 3: Culverts and their color-dotted delineation accuracy category.

5 Uncertainty analysis

Whereas hydraulic analysis provides relatively constrained estimates of hydraulic capacity, the hydrologic model can yield a wider range of discharge values when its sensitive inputs or parameters are varied. In this section, the hydraulic capacity is treated as fixed, and the influence of uncertainties in the hydrologic model on peak discharges is examined. These uncertainties arise primarily from three sources: drainage area, precipitation extremes, and curve number (CN).

The primary source of uncertainty arises from the drainage area, which is affected by inaccuracies in culvert geolocation. To assess this impact, we developed four watershed delineation scenarios corresponding to maximum location errors of 10, 20, 40, and 80 m. The 10 m and 80 m scenarios represent the lower and upper bounds of the possible drainage areas, respectively. A second source of uncertainty is related to extreme precipitation estimates, with NOAA Atlas 14 providing lower-bound, expected-value, and upper-bound precipitation depths for specific durations and return periods, reflecting uncertainty in extreme precipitation statistics [8]. The third source of uncertainty pertains to the CN parameter. We derived the minimum and maximum CN values from both our analysis and the Google Earth Engine CN product [14], thus establishing a plausible range for this parameter.

We evaluated how input and parameter uncertainties map to peak discharge by constructing bounding estimates. The upper bound used the highest CN, the upper precipitation limit, and the largest watershed area (assuming an 80 m error); the lower bound used the lowest CN, the lower precipitation limit, and the smallest watershed area (assuming a 10 m error). Together, these bounds define an uncertainty envelope within which the peak discharge for each return period is expected to lie.

6 Discharge scaling by region

According to the New York State Department of Transportation Highway Design Manual, Albany, Bronx, Broome, Chenango, Columbia, Delaware, Dutchess, Essex, Franklin, Fulton, Greene, Hamilton, Herkimer, Kings, Madison, Montgomery, Nassau, New York, Oneida, Orange, Otsego, Putnam, Queens, Rensselaer, Richmond, Rockland, Saratoga, Schenectady, Schoharie, St. Lawrence, Suffolk, Sullivan, Ulster, Warren, Washington, and Westchester counties, the design discharge should be increased by 20%. Additionally, Allegany, Cattaraugus, Cayuga, Chautauqua, Chemung, Cortland, Erie, Genesee, Livingston, Monroe, Niagara, Ontario, Onondaga, Orleans, Schuyler, Seneca, Steuben, Tioga, Tompkins, Wayne, Wyoming, and Yates counties' design discharge should be increased by 10% [17]. For each return period, discharges were scaled by the recommended factors, and the results were compared with culvert capacities (Fig. 4, panel **b2**, main manuscript).

7 Hydraulic-capacity modeling (coefficient estimation)

The *HY-8* requires detailed information for each culvert—including inlet configuration, wingwall type, and other structural characteristics—to calculate the hydraulic coefficients a, b, c, d, e, and f following the *HY-8* user manual [18]. Such detailed information is not available in the large culvert dataset; instead, the design type of culverts is available in the dataset. Instead, this study employs three groups of predefined coefficients: **A**) circular culverts (design type codes 41 and 42), **B**) oval culverts (design type codes 25 and 26), and **C**) rectangular culverts (applied to all remaining cases where the design type does not fall into groups A or B). For each category (**A**, **B**, and **C**), the corresponding coefficient combinations are extracted from the *HY-8* user manual and summarized in Table 2 .

The hydraulic capacity of each culvert is calculated by solving Equation 4 for $Q_{capacity_n}$. Since it is a five-order polynomial expression, the root of this function can be found using numerical methods, which require an initial guess. This initial guess was set to 100 $\mathrm{ft^3 s^{-1}}$. The procedure iterates over all coefficient combinations associated with the culvert category (Table 2). The mean of the resulting values is taken as the culvert's expected hydraulic capacity, while the minimum and maximum values across the coefficient combinations define the parametric uncertainty bounds of this approach.

Table 2: Coefficients used to model hydraulic capacity for each group of culverts

Category/coefficient	a	b	c	d	e	f
A	0.187321	0.56771	-0.15654	0.044705	-0.00344	8.97E-05
	0.107137	0.757789	-0.36146	0.123393	-0.01606	0.000767
	0.167433	0.538595	-0.14937	0.039154	-0.00344	0.000116
	0.108786	0.662381	-0.2338	0.057959	-0.00558	0.000205
	0.114099	0.653562	-0.23362	0.059772	-0.00616	0.000243
	0.063343	0.766512	-0.3161	0.08767	-0.00984	0.000417
	0.08173	0.698353	-0.25368	0.065125	-0.0072	0.000312
	0.167287	0.558766	-0.15981	0.042007	-0.00369	0.000125
	0.087483	0.706578	-0.2533	0.0667	-0.00662	0.000251
	0.120659	0.630768	-0.21842	0.059182	-0.00599	0.000229
B	0.08905	0.71255	-0.27092	0.07925	-0.00798	0.00029
	0.12263	0.4825	-2.00E-05	-0.04287	0.01454	-0.00117
	0.14168	0.49323	-0.03235	-0.02098	0.00989	-0.00086
	0.09219	0.65732	-0.19423	0.04476	-0.00176	-0.00012
	0.0833	0.79514	-0.43408	0.16377	-0.02491	0.00141
	0.1062	0.7037	-0.3531	0.1374	-0.02076	0.00117
	0.23645	0.37198	-0.0401	0.03058	-0.00576	0.00045
	0.10212	0.72503	-0.34558	0.12454	-0.01676	0.00081
	0.11128	0.61058	-0.19494	0.05129	-0.00481	0.00017
	0.12346	0.50432	-0.13261	0.0402	-0.00448	0.00021
	0.09728	0.57515	-0.15977	0.04223	-0.00374	0.00012
	0.09455	0.61669	-0.22431	0.07407	-0.01002	0.00054
	0.16884	0.38783	-0.03679	0.01173	-0.00066	2.00E-05
	0.1301	0.43477	-0.07911	0.01764	-0.0011	2.00E-05
	0.09618	0.52593	-0.13504	0.03394	-0.00325	0.00013
C	0.122117	0.505435	-0.10856	0.020781	-0.00137	3.46E-05
	0.106759	0.455158	-0.08129	0.012156	-0.00068	1.48E-05
	0.166609	0.398935	-0.06404	0.011201	-0.00064	1.46E-05
	0.072493	0.507087	-0.11747	0.02217	-0.00149	3.80E-05
	0.144133	0.461363	-0.09215	0.020003	-0.00136	3.58E-05
	0.099563	0.441247	-0.07435	0.012732	-0.00076	1.77E-05

8 Automatic pipeline for DEM acquisition and watershed delineation

8.1 DEM acquisition

This study utilized a high-resolution 1-m DEM available through the U.S. Geological Survey's 3DEP program [2], accessed via the *OpenTopography* platform. The 1-meter

DEM provides the level of detail necessary for delineating the catchments of culverts, as it accurately represents the impact of roads on the surrounding watershed. However, in areas where the 1-meter DEM was unavailable, the analysis depended on the 10-m resolution DEM [3] to ensure better coverage of New York State.

Downloading and processing a state-wide 1-m resolution DEM is computationally infeasible due to the massive size of the data (on the terabit scale). To address this, the problem must be reformulated to leverage parallel computing. Specifically, the DEM for each HUC 12 can be downloaded separately, and the drainage area delineation for the culverts within each unit can be processed in parallel. Given that the catchment area for a culvert within a given HUC 12 is independent of the rest of the DEM, we can efficiently download the DEM for each HUC 12, perform watershed delineation, and extract the morphological features of the culverts within that unit independently.

The OpenTopography Application Programming Interface (API) is used to download 1-m DEM data for each HUC12 automatically; however, it imposes certain restrictions on access. Specifically, users are limited to a 300 number of queries every 24 hours, and each DEM download request is restricted to a maximum bounding box area of 250 km^2 for 1-meter resolution data. These limitations necessitate strategic planning when downloading large-scale DEM data to ensure compliance with *OpenTopography*'s access policies. When an HUC12 exceeded the bounding box limit, the script subdivided it into smaller sections (with 10% overlap) to comply with the platform's constraints. After downloading, the DEM tiles for each HUC12 were merged using *gdal.BuildVRT* [19] command and was then transformed to a GeoTIFF to form a complete elevation model for the HUC12. This iterative and modular approach ensured compliance with download restrictions and facilitated large-scale processing and management of several terabyte-scale elevation data required for the study.

8.2 Watershed delineation

For each HUC12 unit, the watershed delineation algorithm delineates culvert watersheds and extracts key morphological features required as inputs for the hydrologic model, including the watershed drainage area, longest flow path, and average catchment slope.

A key challenge addressed in the algorithm is the inaccuracy of culvert location data; recorded coordinates may be offset by several tens of meters. Even when locations are approximately correct, they are frequently mapped directly on roads. This introduces complications for watershed delineation, as roads act as artificial barriers to surface flow, causing points located on roads to generate unrealistically small catchments. Furthermore, digital elevation models do not represent underground drainage systems such as large culverts, making it difficult to correctly model stream networks passing through them. As a result, streams may appear to terminate at roads rather than pass through them, leading to fragmented or unrealistic stream networks and inaccurate catchment delineations.

To address these challenges, the workflow for each HUC12 begins by loading spatial datasets—streams, culverts, roads, and HUC12 boundaries—using *GeoPandas* [20]. A dedicated working directory is created for each HUC12, and the DEMs are preprocessed with *WhiteboxTools* [21]. Preprocessing includes filling depressions with the *wbt.fill_depressions* command and burning road–stream intersections into the DEM using *wbt.burn_streams_at_roads*. Afterwards, flow direction and flow accumulation rasters are generated with the *wbt.d8_pointer* and *wbt.d8_flow_accumulation* commands.

To address inaccuracies in the location of culverts, the script snaps the culvert point to the nearest highest-flow accumulated cell within predefined distances of 10, 20, 40, and 80 m of the original location using *wbt.snap_pour_points*. The code then delineates catchments using the *wbt.watershed* command, calculates the longest flow path using the *wbt.longest_flowpath* command, and finally calculates the average watershed slope using the *wbt.slope* command. Finally, all information is compiled into a structured data frame and saved as both an Excel file and a shape file for each HUC 12. This automated approach streamlines the complex task of watershed delineation and provides the morphologic features required for the hydrologic model.

Regarding the computation time, culverts' watershed delineation at the state-wide scale is a computationally intensive and complex task using a 1-m DEM. For example, in HUC12 '050100020104,' which contains seven culverts, delineating watersheds under four error scenarios of 10, 20, 40, and 80 m, requires approximately 1 215 seconds (about 20 minutes) using MSC v.1940 (64-bit, AMD64) on a Windows 11 system equipped with an Intel64 Family 6 Model 183 Stepping 1 processor (GenuineIntel) with 24 CPU cores. Extrapolated to all culverts, sequential processing would require more than 16 days. By batching computations into 5 groups and running them in parallel, the total time is reduced to 6 days. The choice of 5 batching groups reflects available computational resources; increasing the number of groups (e.g., to match the number of HUC12s, which is more than 1 000) would further reduce computational time.

References

[1] New York State Department of Transportation. NYSDOT Large Culverts Dataset. https://data.gis.ny.gov/datasets/9e038774ef034c7cae5374f3e23f7a67_1/about (2023). URL https://data.gis.ny.gov/datasets/9e038774ef034c7cae5374f3e23f7a67_1/about.

[2] U.S. Geological Survey. 3D Elevation Program (3DEP) - 1-Meter Digital Elevation Models (DEMs). https://opentopography.org/ (2021). URL https://opentopography.org/.

[3] United States Geological Survey. United States Geological Survey 3D Elevation Program 1/3 arc-second Digital Elevation Model. https://doi.org/10.5069/G98K778D (2021). URL https://doi.org/10.5069/G98K778D.

[4] New York State Department of State. Watershed Boundary Dataset (HUC 12) for New York State. https://opdgig.dos.ny.gov/datasets/NYSDOS::watershed-boundary-huc-12/about (2022). URL https://opdgig.dos.ny.gov/datasets/NYSDOS::watershed-boundary-huc-12/about.

[5] New York State Office of Information Technology Services. NYS Streets - Vector Street Layer for New York State. https://gis.ny.gov/streets-addresses (2023). URL https://data.gis.ny.gov/maps/7d5dff6bcf664e6f83abee9968fc7916/about.

[6] New York State Department of State. Railroad Dataset – Paths and Locations of Active and Abandoned Railroad Lines in New York State. https://opdgig.dos.ny.gov/datasets/NYSDOS::railroad/about (2022). URL https://opdgig.dos.ny.gov/datasets/NYSDOS::railroad/about.

[7] New York State Information Technology Services Geospatial Services. NYS Hydrography. https://gisservices.its.ny.gov/arcgis/rest/services/NYS_Hydrography/MapServer (2023). URL https://data.gis.ny.gov/maps/sharegisny::nys-hydrography/about?path=.

[8] Perica, S. *et al.* NOAA Atlas 14: Precipitation-Frequency Atlas of the United States, Volume 10, Version 3.0: Northeastern States. Tech. Rep. NOAA Atlas 14 Volume 10, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, Maryland (2015). URL https://www.weather.gov/media/owp/oh/hdsc/docs/Atlas14_Volume10.pdf.

[9] DeGaetano, A. T. & Castellano, C. M. Future projections of extreme precipitation intensity-duration-frequency curves for climate adaptation planning in New York State. *Climate Services* **5**, 23–35 (2017). URL https://www.sciencedirect.com/science/article/pii/S2405880716300516.

[10] Degaetano, A. T. & Castellano, C. M. Downscaled Projections of Extreme Rainfall in New York State Technical Document. Tech. Rep.

[11] Dewitz, J. National Land Cover Database (NLCD) 2021 Products (2023). URL https://doi.org/10.5066/P9JZ7AO3.

[12] Sohl, T. L. *et al.* Conterminous United States Land Cover Projections - 1992 to 2100 (2018). URL https://doi.org/10.5066/P95AK9HP.

[13] Esri. Hydrologic Soil Class Dataset. https://services1.arcgis.com/6jEyBm2OMIBxuM1E/arcgis/rest/ser (2022). URL https://services1.arcgis.com/6jEyBm2OMIBxuM1E/arcgis/rest/services/Hydrologic_Soil_Class_Public/FeatureServer.

[14] Jaafar, H. H., Ahmad, F. A. & El Beyrouthy, N. GCN250, new global gridded curve numbers for hydrologic modeling and design. *Scientific Data* **6**, 145 (2019). URL https://doi.org/10.1038/s41597-019-0155-x.

[15] Hodson, T. O., Hariharan, J. A., Black, S. & Horsburgh, J. S. dataretrieval (Python): A Python package for discovering and retrieving water data available from U.S. federal hydrologic web services. USGS software release (2023). URL https://doi.org/10.5066/P94I5TX3.

[16] U.S. Geological Survey. StreamStats (2019).

[17] NYSDOT. HIGHWAY DESIGN MANUAL, Chapter 8. Tech. Rep. (2021). URL https://www.dot.ny.gov/divisions/engineering/design/dqab/hdm/chapter-8.

[18] Christopher Smemoe. HY-8 User Manual (v7.5) HY-8 Culvert Analysis Program (2016). URL http://wmsdocs.aquaveo.com/HY-8_UserManual_v7.5.pdf.

[19] GDAL/OGR contributors. GDAL/OGR Geospatial Data Abstraction software Library (2025). URL https://gdal.org.

[20] Jordahl, K. *et al.* geopandas/geopandas: v0.8.1 (2020). URL https://doi.org/10.5281/zenodo.3946761.

[21] Lindsay, J. B. *The Whitebox Geospatial Analysis Tools project and open-access GIS*, 16–18 April (The University of Glasgow, 2014). URL https://doi.org/10.13140/RG.2.1.1010.8962.