

Predictive value of current HEC-RAS 2D water levels for machine learning flood postprocessing in southeastern Louisiana

Zaid Alhusban

zalhus1@lsu.edu

Louisiana State University

Noujoud Nader

Louisiana State University

Carola Kaiser

Louisiana State University

Hartmut Kaiser

Louisiana State University

Research Article

Keywords: HEC-RAS 2D, machine learning, statistical postprocessing, structured cross-validation, water-surface elevation, flood risk assessment

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11492583/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Predictive value of current HEC-RAS 2D water levels for machine learning flood postprocessing in southeastern Louisiana

Zaid Alhusban*, Noujoud Nader, Carola Kaiser, Hartmut Kaiser

Center for Computation and Technology

Louisiana State University, Baton Rouge, Louisiana, USA

***Corresponding author: Zaid Alhusban**

Email: zalhus1@lsu.edu

Telephone: +1 302-252-1161

Abstract

Machine-learning postprocessing quantified information in current water-surface elevation (WSE) and its recent history from the Hydrologic Engineering Center's River Analysis System (HEC-RAS 2D). The analysis used 130 hydrographs and 5,134 paired 3-h records from eight southeastern Louisiana floods. Cross-validation excluded one complete flood from training per outer test fold. Event-macro hydrograph root-mean-square error (RMSE), which weighted floods equally, was 1.63 m for uncorrected HEC-RAS and 0.81 m for a model using forcing, time, location, and static descriptors without time-varying WSE. Adding current WSE reduced RMSE to 0.67 m and improved all eight floods. Adding recent history reduced RMSE to 0.62 m and improved seven floods. When each test gauge was excluded from training, models without WSE, with current WSE, and with current plus history scored 1.62, 0.96, and 0.87 m, respectively. The history model improved on uncorrected HEC-RAS at only 16 of 33 gauges, and the largest gains occurred at deeply incised lower-Amite mapped cells. WSE-informed hydrograph models improved high-water and centered errors but not exact 3-h changes, peak magnitude, or peak timing. A separate exploratory peak-magnitude model reduced mean absolute error from 0.48 to 0.44 m but did not estimate peak time. One-third of scheduled mapped outputs lacked supported WSE. Current WSE supplied the largest and most consistent increment, while recent history added a smaller and less consistent benefit. The analysis used completed floods and did not evaluate hydraulic-model validity or future-flood forecasts.

Keywords: HEC-RAS 2D; machine learning; statistical postprocessing; structured cross-validation; water-surface elevation; flood risk assessment

1. Introduction

Flooding is one of the most consequential natural hazards, affecting communities, infrastructure, and emergency operations. Accurate estimates of flood extent and water levels are critical for flood-risk assessment and timely decision-making. However, flood behavior is controlled by interacting rainfall, riverine, coastal, terrain, and hydraulic processes, making accurate simulation challenging, particularly during extreme events. Reliable flood-level estimates also support infrastructure planning and long-term risk reduction because predicted water depths, inundation extents, and flood durations inform exposure estimates, damage calculations, warning decisions, and flood-management planning (Merz et al. 2010; Teng et al. 2017). In low-lying coastal watersheds, intense rainfall can correspond with storm surge, elevated coastal water levels, or high river discharge. Dependence among these drivers increases compound flood potential along the Atlantic and Gulf coasts of the United States and in deltas and estuaries worldwide (Wahl et al. 2015; Ward et al. 2018). At an individual location, water level reflects the combined influence of rainfall, coastal and river boundary stages, precursor storage, terrain and channel geometry, flow resistance, hydraulic structures, connectivity, and routing through the river-floodplain system (Jafarzadegan et

al. 2023). Physics-based hydraulic models are therefore widely used to represent these interacting processes and estimate flood water levels across complex river–floodplain systems.

The Hydrologic Engineering Center's River Analysis System (HEC-RAS), developed by the U.S. Army Corps of Engineers, simulates river and floodplain flow (US Army Corps of Engineers, Hydrologic Engineering Center n.d.-b). In two-dimensional mode, HEC-RAS divides the landscape into computational cells and solves shallow-water equations for water depth, velocity, and water-surface elevation (WSE) through time. The solution represents storage, wetting and drying, hydraulic connections, and exchange across model boundaries. Accuracy depends on terrain, cell size, flow resistance, rainfall losses, initial water levels, boundary flows or stages, and numerical settings (US Army Corps of Engineers, Hydrologic Engineering Center n.d.-a, n.d.-b, n.d.-c). Even when hydraulic simulations contain systematic errors, their evolving WSE may encode useful information about flood state that is not available from forcing variables alone. Statistical postprocessing can use this simulated state to estimate the discrepancy between modeled and observed water levels.

Machine-learning (ML) postprocessing offers a potential approach for reducing systematic errors in hydrologic and hydrodynamic model outputs by learning relationships between model simulations and observations. Recent studies have similarly applied ML-based bias correction to physics-based storm-surge models using long short-term memory networks, graph-based models, and generative approaches (Giaremis et al. 2024; Nader et al. 2026a, 2026b). Giaremis et al. (2024) corrected storm tide errors with a recurrent network tested on Hurricane Ian. StormNet combined spatial relationships among gauges with temporal information to correct storm surge forecasts, while HURRI-GAN examined correction beyond gauge locations (Nader et al. 2026a, 2026b). Previous studies demonstrate that ML/statistical postprocessing can reduce hydrologic/hydrodynamic model errors, but performance varies with flow regime, model state, and forecast horizon. The hydraulic simulation remains unchanged while the correction model estimates a statistical relationship between model outputs and corresponding observations (Li et al. 2017; Papacharalampous and Tyralis 2022). A model may estimate observed water level directly or estimate the simulation residual and add the correction to the original output (Hammad et al. 2026). Correction models can add forcing, site attributes, current simulated WSE, and earlier hydraulic states while keeping the original simulation as an input. Current simulated WSE may be informative because it summarizes the model's accumulated response to forcing, storage, terrain, resistance, connectivity, and routing. Earlier simulated states may provide additional information about flood evolution and model memory. Postprocessing corrects the combined discrepancy between simulation and observation but does not identify or repair its physical cause. The possible improvement depends on the modeled variable, hydrologic state, forecast lead, available training data, and evaluation design (Li et al. 2017; Tedesco et al. 2024).

Prior postprocessing studies show that correction can be important but is not uniform across flow regimes or prediction settings. Driver-aware and state-dependent corrections have reduced streamflow errors across different flow conditions (Hammad et al. 2026; Neisary et al. 2025). For operational storm-surge forecasts, a neural-network correction reduced root-mean-square error (RMSE) by 36% at a 1-h lead but by 5% at 60 h (Tedesco et al. 2024). A neural-operator correction of coastal water-level guidance reported an overall RMSE of 0.13 m and similar error at stations excluded from training (Alipour et al. 2026). Other studies have used machine learning to supply hydrodynamic boundaries or estimate maximum flood depth (Huang et al. 2022; Pakdehi et al. 2024). These studies evaluated different outputs, including discharge, forecast water level, and maximum flood depth. Their different targets, lead times, station networks, and aggregation rules prevent direct numerical comparison with the RMSE calculated here for complete gauge hydrographs.

On the other hand, hydrodynamic surrogates address a different question: they learn to reproduce simulator fields or stages more quickly, rather than correcting them against observations (Haces-Garcia et al. 2025; Holmberg et al. 2025, 2026; Li et al. 2026). Their performance depends on the simulator used as the training target and does not reveal how much observation-correction skill comes from the hydraulic trajectory itself. Evaluation design matters because random record splits can place adjacent times from one flood in both training and testing, and familiar gauges can supply location-specific information that will not exist at a new location. Complete floods and locations must therefore be excluded from training when they are the intended transfer units (Acuña Espinoza et al. 2025; Bergmeir et al. 2018; Klemeš 1986; Meyer et al. 2018; Roberts et al. 2017). What remains unclear is how much

correction skill comes from current simulated WSE, how much comes from its recent history, and how much persists without gauge-specific training data.

The objective of this study is to quantify the incremental predictive information carried by HEC-RAS 2D WSE when postprocessing completed floods. The analysis uses eight historical floods observed at southeastern Louisiana gauges. Three research questions structure the analysis. First, *how much does the current simulated WSE, defined as the HEC-RAS value at the same time as the observation, add beyond forcing, time, gauge identity, and static site descriptors?* Second, *how much additional information is supplied by recent WSE lags, changes, and range?* Third, *do those increments persist when complete floods or mapped locations are excluded from training?* Prespecified sensitivity analyses separately examine observation reference, gauge-to-cell mapping, hydraulic support, and numerical quality. Comparing corrected and uncorrected water levels alone cannot determine whether improvement comes from the simulated hydraulic state or from persistent gauge effects. The matched models separate these sources of predictive information, and exclusion of complete floods and gauges tests how much remains without training data from the test flood or location.

To address these questions, the study systematically evaluates the incremental predictive value of HEC-RAS information, first considering the current simulated WSE and then incorporating information from its recent history. The same ML framework and evaluation procedure are maintained across these comparisons to isolate the contribution of each source of information. Because the eight floods were inspected before this analysis, the tests provide development evidence from completed floods rather than external validation.

The principal contributions are:

- Isolates the predictive information supplied by current HEC-RAS WSE beyond forcing, time, gauge identity, and static site descriptors.
- Quantifies the additional information supplied by recent WSE lags, changes, and range without presenting the selected history bundle as an optimized temporal model.
- Tests whether the information increments persist when complete floods and mapped locations are excluded from training and under observation-reference, mapping, support, and numerical-quality checks.

This paper is organized as follows. Section 2 presents the data and methods, Section 3 reports the results, Section 4 discusses their interpretation and limitations, and Section 5 summarizes the conclusions.

2. Data and methods

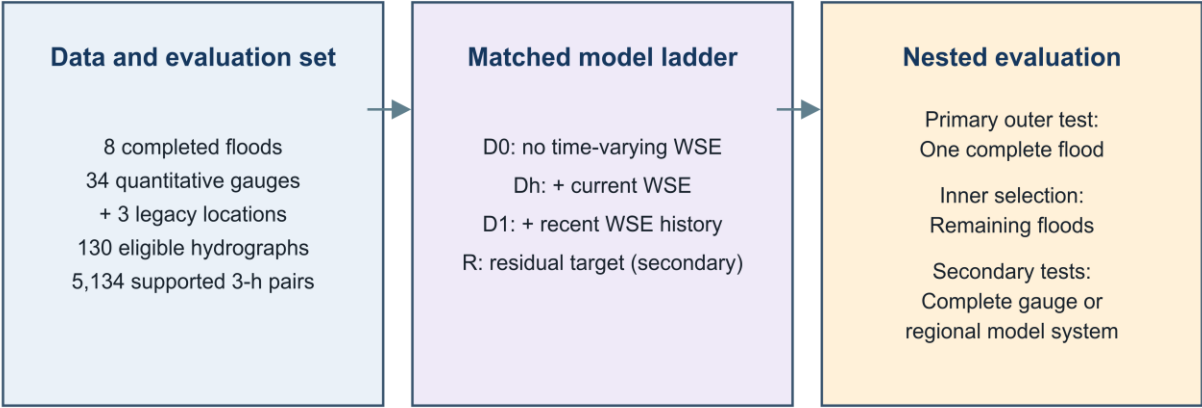
2.1 Overview of the study design

The study was designed to analyze the incremental predictive information provided by the simulated HEC-RAS hydraulic trajectory and to determine whether that information persists when predictions are made for floods or locations excluded from model training. For this purpose, four machine learning formulations were compared using matched evaluation records and a common learning framework, as described in Section 2.6.

The primary comparison sequentially added information from the HEC-RAS simulation: D0 excluded time-varying simulated WSE, Dh added the current simulated WSE, and D1 additionally incorporated recent WSE history. These formulations were evaluated using nested cross-validation in which one complete flood was excluded from training in each outer fold. This design allowed the contribution of current WSE and recent WSE history to be evaluated on the same test records. Additional analyses examined residual targeting, hydrograph and peak metrics, transfer to locations excluded from training, and sensitivity to key data and hydraulic-model assumptions. The analysis used eight completed floods from different years and flood types in southeastern Louisiana.

Figure 1 summarizes the overall analysis workflow. The eight historical floods were first matched with gauge observations and screened to define a common evaluation dataset. The matched ML sequence then progressively introduced current HEC-RAS WSE and recent WSE history while retaining the shared predictors and evaluation records. Finally, structured training exclusions were used to evaluate generalization across complete floods,

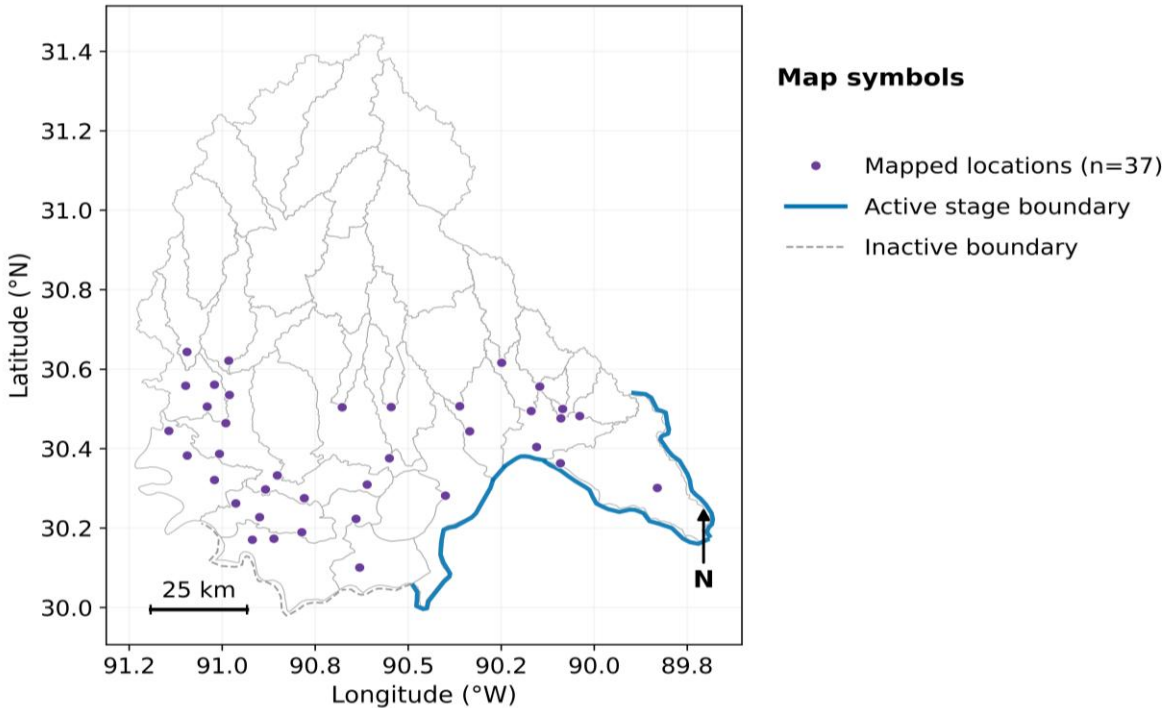
133 individual gauges, regional model systems, and joint flood–location combinations.



134
135 **Fig. 1** Study design and matched machine-learning comparisons. Data screening defines the common evaluation set, the model
136 sequence adds current water-surface elevation (WSE) and recent history, nested cross validation excludes complete floods from
137 training. Secondary tests exclude gauges, regional model systems, or joint flood location combinations

138 **2.2 Study area and flood events**

139 The study focuses on southeastern Louisiana, where the regional HEC-RAS 2D domain covers parts of the Amite,
140 Comite, Tickfaw, Tangipahoa, Tchefuncte, Natalbany, Maurepas, Lacombe, and Lake Pontchartrain systems in
141 southeastern Louisiana (Figure 2). The fixed geometry contains 35 two-dimensional flow areas. Its effective forcing
142 consists of distributed precipitation and five external stage boundaries; no explicit river inflow or baseflow is
143 connected to the active two-dimensional geometry.



144
145 **Fig. 2** Two-dimensional Hydrologic Engineering Center's River Analysis System (HEC-RAS) domain in southeastern
146 Louisiana, showing 35 flow-area boundaries, 37 mapped locations, and five active external-stage boundaries. Axes show
147 longitude and latitude; inactive boundary lines are retained for spatial context

148 Eight historical flood events were analyzed, representing different hydrometeorological conditions across the study
149 region (Table 1). Four events—Francine (2024), Gustav (2008), Ida (2021), and Isaac (2012)—were associated with

tropical systems, while August 2016, January 2013, March 2015, and March 2016 were dominated by rainfall within the model domain. The event set was defined before comparative machine-learning fitting, and no event was subsequently removed based on hydraulic or statistical performance.

Figure 2 shows the HEC-RAS 2D domain, mapped gauge locations, and external-stage boundaries. Table 1 summarizes the simulation period and external-stage forcing for each event, together with the corresponding evaluation data.

2.3 Simulated Data: HEC-RAS 2D model configuration and forcing

HEC-RAS 2D generated the simulated WSE series supplied to the postprocessing models. Hourly precipitation came from the Analysis of Record for Calibration version 1.1 and was area-weighted over the model polygons (Fall et al. 2023). Rainfall excess used the Soil Conservation Service Curve Number method, 2011 land cover, an initial-abstraction ratio of 0.20, and a fallback curve number of 80 for channel classes (Homer et al. 2015; US Department of Agriculture, Natural Resources Conservation Service 2004). The abstraction accounting restarted at each event, and no event-specific antecedent-moisture class was used.

External-stage boundary conditions varied among floods according to the available forcing data. Floods F2 (Francine), F3 (Gustav), F4 (Ida), and F5 (Isaac) used time-varying external stages derived from archived Advanced Circulation Model coastal simulations at fixed extraction points (Luettich and Westerink 2004). Floods F6 (January 2013), F7 (March 2015), and F8 (March 2016) used a constant 0.35 m ambient stage. Flood F1 (August 2016) used one processed coastal-average series at all five boundaries. The archives documented source files and time coverage but did not establish a common elevation datum or complete independence from earlier observational influence. All floods used the same initial-condition file.

Table 1. Eight completed flood events (F1–F8), simulation windows in Coordinated Universal Time (UTC), external-stage forcing, and evaluation data. Model-dry percentages refer to the complete 3-h schedule across all 37 mapped locations. Hydrograph and paired-record counts refer to the elevation-screened, model-supported primary cohort.

	Flood (type)	Simulation window, UTC	External-stage source	Model-dry schedule (%)	Eligible hydrographs	Paired records
F1	August 2016 (rainfall-dominated)	2016-08-11 06:00–2016-08-16 06:00	Processed coastal-average stage	13.1	22	779
F2	Francine 2024 (tropical)	2024-09-09 21:00–2024-09-14 07:00	Coastal-model stage; limited pre-window coverage	33.4	8	238
F3	Gustav 2008 (tropical)	2008-08-28 12:00–2008-09-04 06:00	Coastal-model stage	45.2	6	193
F4	Ida 2021 (tropical)	2021-08-27 12:00–2021-09-03 00:00	Coastal-model stage	20.0	32	1,332
F5	Isaac 2012 (tropical)	2012-08-23 12:00–2012-08-30 08:00	Coastal-model stage; incomplete recession	38.7	13	521
F6	January 2013 (rainfall-dominated)	2013-01-06 03:00–2013-01-13 12:00	Constant 0.35 m stage	14.4	23	1,060
F7	March 2015 (rainfall-dominated)	2015-03-06 03:00–2015-03-13 12:00	Constant 0.35 m stage	61.2	2	64
F8	March 2016 (rainfall-dominated)	2016-03-08 03:00–2016-03-15 12:00	Constant 0.35 m stage	32.2	24	947

The simulations were performed using HEC-RAS 6.6 with the Diffusion Wave equation set, a 15-s computational time step, and water-surface elevation output saved at 3-h intervals. The inherited configuration was not recalibrated for this analysis.

The hydraulic simulations have several limitations relevant to the interpretation of the machine-learning results. Complete earlier calibration provenance could not be recovered, and observed stages from August 2016 and Ida had informed pre-analysis configuration comparisons. In addition, the official known-issues notice for HEC-RAS versions 6.0–6.6 identifies a matrix-assembly parallelization bug that can cause non-reproducibility in unstable

models (US Army Corps of Engineers, Hydrologic Engineering Center 2024). Its effect on this panel cannot be determined because the exact executable build and binary hash were unavailable and an identical-input duplicate run was not retained. The hydraulic panel is therefore development material rather than independent model validation.

2.4 Observed Data: Water-level observations and gauge-to-model mapping

Observed water levels were obtained from U.S. Geological Survey (USGS) gauges within the model domain. USGS gage height is measured relative to a gauge-specific zero and is not automatically referenced to the North American Vertical Datum of 1988 (NAVD88) (US Geological Survey 2020, 2026). Flood-specific elevation transformations were reviewed from current metadata, historical reports, gauge manuscripts, annual peaks, and continuous-series documentation. Among the 239-gauge flood combinations with observations in the simulation windows, this review revised offsets for 23 combinations and excluded 29 with unresolved elevation references. The model terrain was intended to use NAVD88, but its realization, epoch, geoid, and complete transformation chain were not documented. All vertical comparisons therefore remain conditional on an incompletely documented model reference. A stricter analysis retained only 46 hydrographs and 1,848 records with directly documented observation offsets.

Each gauge location was mapped to the HEC-RAS computational mesh using model geometry only, without reference to observed water levels or prediction errors. Each location was assigned to its containing model area and nearest declared cell center using geometry alone. This reconstruction reproduced the archived area and cell index at all 37 mapped locations, including three legacy locations without quantitative observations. These 37 locations are spatial assignments, whereas the 23 revised offsets and 29 exclusions in the observation-reference audit are gauge flood combinations. Center distances ranged from 44.4 to 188.3 m. Polygon containment verifies the spatial index but not that a cell represents the same channel branch, levee side, structure, or hydraulic control as the gauge. Five same-area candidate cells were recorded at each location before performance evaluation. The complete models were refitted at each candidate rank on 4,490 common records, and a separate refit excluded the 17 locations in the lowest distance-proximity tier. No alternative cell was selected from observed agreement.

A separate proximity analysis classified mappings as high, moderate, or low according to prespecified distance criteria and repeated the analysis after excluding low-proximity locations. High proximity required designated- and fifth-candidate distances no greater than 75 and 200 m, respectively. Moderate proximity required distances no greater than 125 and 350 m without meeting the high rule; all other locations were low. Three locations met the high rule, 17 were moderate, and 17 were low; 19 of the 20 high or moderate locations contributed eligible hydrographs to the corresponding refit. These geometric tiers are neither probabilities nor certification of hydraulic equivalence. The observation-reference analysis was similarly bounded: 32 prespecified perturbation sets varied only documented observation offsets and repeated the fitting procedure while excluding each complete flood from training. The perturbation sets did not vary model terrain or boundary-stage reference.

2.5 Dataset construction and hydraulic support

The 34 gauges with quantitative observations across eight flood events defined 272 possible gauge–flood combinations. Of these, 239 contained observations within the corresponding simulation window. Gauge–flood hydrographs were retained for the primary evaluation after observation-reference screening, hydraulic-support screening, and application of minimum data-availability and variability criteria. Eligible hydrographs were required to contain at least 25 paired records and an observed WSE standard deviation greater than 0.05 m. These criteria resulted in 130 eligible hydrographs from 33 gauges and 5,134 paired 3-h records.

A paired record represents one observed–simulated WSE comparison at a retained 3-h output time, whereas a gauge–flood hydrograph comprises the complete retained series for one gauge during one flood. Serial observations within a hydrograph were therefore treated as repeated measurements within the same flood rather than as independent flood events.

A simulated WSE was classified as hydraulically supported only when it exceeded the mapped cell minimum by more than 0.03048 m (0.1 ft, the original model-unit threshold). The threshold increment is a study-specific

computational tolerance, not a physical wet/dry standard. Across 15,540 scheduled outputs, 5,136 (33.1%) failed that rule. After elevation screening, 2,016 of 9,586 exact observation-schedule records had observations above the same terrain threshold while the model was classified as dry. The mismatched records are support mismatches rather than validated wet/dry errors because the elevation reference and hydraulic correspondence remain conditional. The primary scores use only the 5,134 supported pairs; 5,128 were exact observation times and six used bounded linear interpolation. No observation was extrapolated.

2.6 Machine-learning postprocessing framework

Four machine-learning formulations were used to quantify the incremental predictive information provided by the simulated HEC-RAS hydraulic trajectory. For the three direct formulations, the prediction target, evaluation records, and learning framework were held constant while progressively adding information from the simulated WSE. The residual formulation used the same records and predictors as D1 but changed the training output. At gauge (s), flood (e), and time (t), (y) denotes WSE derived from the retained USGS gage-height record after the gauge–flood elevation transformation, (h) denotes simulated HEC-RAS WSE, (x) represents the shared non-dynamic predictors, and (T(h)) represents recent simulated WSE history.

The first formulation, D0, excluded time-varying simulated WSE and predicted observed WSE using only the shared predictors. The second formulation, Dh, added the current HEC-RAS WSE at the same time as the observation, thereby isolating the incremental information provided by the current simulated hydraulic state. The third formulation, D1, additionally incorporated recent simulated WSE history, including lags, changes, and a trailing range, to evaluate whether recent trajectory information provided predictive value beyond the current WSE. A fourth formulation, R, used the same predictors as D1 but predicted the HEC-RAS residual, (y-h), rather than observed WSE directly. D0–Dh–D1 constituted the primary matched comparison, while R was used as a secondary target comparison.

$$\hat{y}_{s,e,t}^{D0} = f_0(x_{s,e,t}). \quad (1)$$

$$\hat{y}_{s,e,t}^{Dh} = f_h(h_{s,e,t}, x_{s,e,t}). \quad (2)$$

$$\hat{y}_{s,e,t}^{D1} = f_1(h_{s,e,t}, T(h)_{s,e,t}, x_{s,e,t}). \quad (3)$$

$$\hat{y}_{s,e,t}^R = h_{s,e,t} + r(h_{s,e,t}, T(h)_{s,e,t}, x_{s,e,t}). \quad (4)$$

All four formulations shared rainfall, external stages, elapsed time, event phase, gauge and model-area identity, and static mapping descriptors. Rainfall predictors included the current excess rate, trailing 3–48 h totals, and cumulative excess. Boundary predictors contained the five current stages and their 6-h changes. Static predictors included mapped-cell minimum elevation, gauge-to-cell distance, event type, and whether the model area contained an active stage boundary. Event identity, observed-WSE derivatives, target residuals, and summaries from the test flood were prohibited.

Event phase was elapsed time divided by the prescribed event-window duration. Event phase did not use an observed-WSE landmark, but the predictor assumes that the event end is known because the analysis occurs after each flood is completed. The variable remained in the prespecified primary models to avoid switching after seeing results; a separately selected and fitted no-phase analysis tested whether the information ordering depended on it.

Table 2 defines the information added at each step. D0 is not completely independent of HEC-RAS because it retains model-derived static descriptors, model-area identity, the fixed mapping, and the common model-supported cohort. Dh adds the current simulated WSE. D1 adds exact 3- and 6-h lags, the corresponding changes, and the range over the current and four preceding supported values. Unsupported gaps can lengthen the clock span of the range, so it is not described as a fixed 12-h window. D1 is the prespecified history bundle rather than an optimized temporal model.

Table 2. Matched machine-learning formulations.

Code	Descriptive name	Model inputs added beyond the shared predictors	Training output	Role
D0	Model without time-varying WSE	None	Observed WSE	Primary baseline
Dh	Current-WSE model	Simulated WSE at the observation time	Observed WSE	Primary current-WSE increment
D1	Current-plus-history model	Current WSE, exact 3- and 6-h lags and changes, and trailing supported-value range	Observed WSE	Prespecified history increment
R	Residual model	Same hydraulic information as D1	Residual: observed minus simulated WSE	Secondary target comparison

Numerical imputation, scaling, and categorical encoding were fitted only on training data. The exact 3-h lag was missing in 2.43% of primary records, the 6-h lag in 4.85%, and the trailing range in 2.36%. Training-fold median imputation supplied one preprocessing architecture for nonlinear and linear comparisons. A separate check added missing-lag indicators. Unseen gauge or area categories in spatial tests received all-zero indicator columns rather than target-derived encodings.

The trajectory models used histogram gradient boosting with squared-error loss, learning rate 0.05, 300 iterations, no early stopping, and a fixed random seed (Pedregosa et al. 2011; Scikit-learn developers 2026). Four predeclared combinations of maximum leaf nodes, minimum samples per leaf, and L₂ regularization were (15, 30, 1), (31, 30, 1), (15, 60, 10), and (31, 60, 10). Every formulation was selected separately by inner complete-flood folds. A fixed-complexity sensitivity fitted the three direct models under each setting.

Training-only comparisons included gauge climatology, global and gauge-specific mean or median residual correction, regularized linear residual prediction, direct ridge models using the D0, Dh, and D1 information sets, and a gauge-aware slope-and-intercept model. Secondary feature-block models added only WSE lags, only changes, or only the trailing range to Dh. These comparisons tested whether a simpler correction or one component of recent history explained the main result.

2.7 Model training, evaluation, and sensitivity analyses

Model performance was evaluated in two stages: model settings were selected using only training floods, and predictions were then scored on the excluded flood or location. Preprocessing was repeated within each training split so that the test records contributed neither to input transformations nor to model selection.

Flood generalization. The primary evaluation used nested cross-validation that excluded one complete flood from training at a time. Each analyzed flood was treated as an event for weighting; event-macro RMSE gives every flood equal weight. Each outer fold designated one test flood and excluded all of its records from preprocessing, model-setting selection, and fitting. Inner folds applied the same exclusion to one of the seven training floods at a time. Records were weighted so that each flood contributed equally and each hydrograph within a flood contributed equally.

For hydrograph g ,

$$\text{RMSE}_g = \sqrt{\frac{1}{n_g} \sum_{i \in g} (p_i - y_i)^2}. \quad (5)$$

and the primary event-macro score was

299

$$\text{RMSE}_{\text{macro}} = \frac{1}{E} \sum_e \left[\frac{1}{G_e} \sum_{g \in e} \text{RMSE}_g \right]. \quad (6)$$

300 Here, i denotes a paired record, g a gauge flood hydrograph, and e a flood, p_i and y_i are the predicted and observed
 301 WSE. The count n_g is the number of paired records in hydrograph g , and G_e is the number of eligible hydrographs in
 302 flood e . E is the number of evaluated floods, equal to eight in the primary analysis. Equation (6) averages
 303 hydrograph RMSE within each flood and then weights the floods equally; it is not a pooled record level RMSE.

304 **Hydrograph and peak evaluation.** Complementary metrics were used to assess whether differences in overall WSE
 305 RMSE were also reflected in bias, high-water performance, hydrograph variation, short-interval changes, and flood
 306 peaks. These included pooled bias, median hydrograph Nash–Sutcliffe efficiency (NSE), RMSE within the top
 307 decile of observed WSE, centered RMSE, and exact 3-h first-difference RMSE (Nash and Sutcliffe 1970). Bias was
 308 pooled over the identical test records, whereas NSE was computed within each hydrograph and summarized by its
 309 median. The observed top decile was defined separately within each hydrograph, including ties at the 90th
 310 percentile. Centered RMSE was calculated after removing the mean error from each hydrograph. First differences
 311 used only adjacent paired records separated by exactly 3 h; hydrographs without such a pair were omitted only from
 312 that metric. Hydrograph-level metrics were averaged within floods and then equally across floods.

313 For hydrograph g , b_g is its mean prediction error and P_g contains the indices of adjacent paired records whose
 314 timestamps are exactly 3 hours apart. The two complementary RMSE measures are defined as follows:

$$\text{cRMSE}_g = \sqrt{\frac{1}{n_g} \sum_{i \in g} (p_i - y_i - b_g)^2}. \quad (7)$$

$$\text{dRMSE}_g = \sqrt{\frac{1}{|P_g|} \sum_{i \in P_g} [(p_i - p_{i-1}) - (y_i - y_{i-1})]^2}. \quad (8)$$

317 Peak performance was evaluated in two ways. First, peak magnitude and timing were derived from each continuous
 318 hydrograph prediction using the common supported timestamps, with the earliest time retained when maxima were
 319 tied. Peak magnitude and timing errors were averaged within floods and then equally across floods. Because the
 320 continuous models were selected using hydrograph RMSE rather than peak error, a second exploratory model was
 321 trained directly on one peak magnitude per hydrograph. Its information sequence added full-flood hydraulic-level
 322 summaries and then temporal-shape summaries to forcing, duration, and static predictors. A training-event mean
 323 peak-bias correction provided the simple baseline. The tree family and blend were selected within the same eight
 324 development folds; consequently, this peak-specific analysis was treated as exploratory and did not estimate peak
 325 timing.

326 **Location generalization.** Generalization to locations excluded from training was evaluated by repeating
 327 preprocessing, inner model selection, and fitting after removing all records from the test gauge or regional model
 328 system. Joint tests excluded both a complete flood and a location from training. The four location-based evaluation
 329 designs produced 33 gauge folds, six regional-system folds, 130 flood–gauge intersections, and 36 flood–system
 330 intersections. The six regional groups with contributing gauges were Amite, Comite, Natalbany, Tangipahoa,
 331 Tchefuncte–Bogue Falaya, and Tickfaw; these labels follow model-area names rather than independently surveyed
 332 watershed boundaries.

333 For the event-macro location scores, outer predictions were concatenated and hydrograph errors were averaged
 334 within floods before equal-flood averaging. Separate score-only summaries were calculated with equal weighting
 335 across gauges or regional systems using the accepted test predictions and therefore did not require model refitting or
 336 reselection. For the equal-gauge score, hydrograph RMSE was first averaged across the available floods for each
 337 gauge and the resulting 33 gauge scores were averaged equally. For the equal-system score, available gauge–flood
 338 hydrograph errors were averaged within each regional system and the six system scores were then averaged equally.

339 **Sensitivity analyses.** Prespecified sensitivity analyses examined whether the primary D0–Dh–D1 comparisons were
340 robust to uncertainty in observation reference, gauge-to-model mapping, hydraulic-model quality, and analytical
341 choices. Observation-reference analyses included a stricter subset restricted to directly documented offsets and
342 bounded perturbations of uncertain offsets. Mapping analyses considered the five prespecified candidate-cell ranks,
343 exclusion of low-proximity mappings, and exclusion of lower-Amite records. Hydraulic-quality analyses considered
344 local volume-accounting diagnostics, while analytical sensitivities examined event phase, trailing-range duration,
345 flood-edge effects, fixed model complexity, and a six-flood score excluding August 2016 and Ida. The analytical
346 repository retains the complete perturbation, fixed-complexity, hydraulic-diagnostic, and correction ledgers.
347 Observation-blind hydraulic repair runs were not used for statistical training or selected on the basis of observation
348 error.

349 OpenAI ChatGPT and Codex assisted with code and workflow drafting, quantitative consistency checks, prose
350 revision, figure and table editing, and critical review of manuscript structure and presentation. The tools supported
351 manuscript and workflow preparation; they did not provide independent scientific validation.

352 **3. Results**

353 **3.1 Complete-flood generalization and incremental WSE information**

354 The primary analysis evaluated the D0–Dh–D1 sequence after excluding each complete test flood from training to
355 quantify the incremental contribution of current simulated WSE and its recent history. The results are presented in
356 Table 3, Figure 3.

357 Uncorrected HEC-RAS had an event-macro hydrograph RMSE of 1.63 m (Table 3). The model without time-
358 varying WSE reduced it to 0.81 m. Adding current WSE reduced RMSE to 0.67 m, a further 0.14 m, and improved
359 all eight test floods. Adding recent WSE history reduced RMSE to 0.62 m, another 0.05 m, and improved seven
360 floods (Figure 3). In the specified sequence, current WSE supplied 74.1% of the 0.19 m reduction from D0 to D1;
361 this order-dependent percentage is descriptive rather than causal feature attribution.

362 The current-WSE flood effects ranged from 0.05 to 0.24 m of RMSE reduction and gave a two-sided sign-test p
363 value of 0.0078. The history effects ranged from a 0.13 m reduction to a 0.04 m increase and gave $p = 0.0703$.
364 March 2015, represented by only two hydrographs, was the adverse history case. Omitting any one flood effect left
365 both mean increments favorable. Giving every hydrograph equal weight instead of every flood also preserved the
366 ordering: uncorrected HEC-RAS, D0, Dh, D1, and R scored 1.35, 0.88, 0.72, 0.67, and 0.65 m. Current WSE
367 reduced RMSE in eight of eight floods, recent WSE history in seven of eight, and residual targeting in six of eight.

368 Descriptive event-cluster bootstrap intervals were $[-0.19, -0.10]$ m for Dh minus D0 and $[-0.08, -0.02]$ m for D1
369 minus Dh. The sign test measures directional consistency, whereas the bootstrap reflects the magnitudes of the eight
370 observed flood effects. Neither is treated as confirmatory inference for a population of future floods. The complete
371 event-level values remain available with the test predictions.

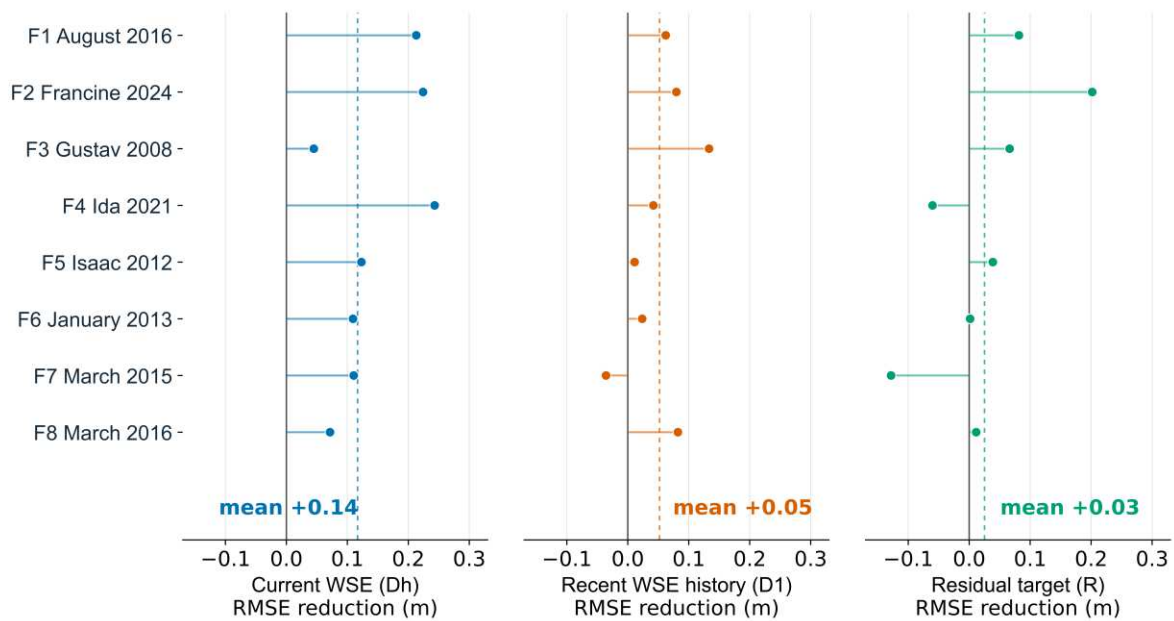


Fig. 3 Test flood root mean square error (RMSE) reductions after adding current water-surface elevation (WSE, D0 to Dh), recent WSE history (Dh to D1), or residual targeting (D1 to R). Positive values are the earlier model RMSE minus the later-model RMSE. Dashed lines show medians, and mean reductions are labeled within the panels. Flood labels F1–F8 correspond to Table 1

Residual targeting produced the lowest overall RMSE, 0.59 m, but its 0.03 m advantage over D1 varied among floods. The simpler comparisons were materially worse: gauge climatology scored 1.46 m, the gauge-aware slope-and-intercept model 0.90 m, and direct ridge models 1.26, 1.18, and 1.15 m as the D0, Dh, and D1 information sets were added. The main result is therefore not reproduced by a gauge mean, one constant correction, or one regularized linear transformation.

Recent-history performance depended on its representation. Lags only scored 0.66 m, changes only 0.60 m, and the trailing range only 0.66 m, compared with 0.67 m for current WSE and 0.62 m for the complete prespecified history bundle. A current value plus a lag and a current value plus the corresponding change contain equivalent information when both values are present. The changes-only advantage may therefore reflect missingness, finite-sample selection, or the response of the nonlinear learner to correlated predictors rather than a separate physical mechanism.

Table 3. Continuous-hydrograph performance on identical test records.

Formulation	Event-macro RMSE (m)	Median hydrograph NSE	Pooled bias (m)	Top-decile RMSE (m)	Centered RMSE (m)	3-h first-difference RMSE (m)
Uncorrected HEC-RAS	1.63	−0.20	−0.42	0.86	1.13	0.32
D0	0.81	0.18	−0.05	0.99	0.54	0.21
Dh	0.67	0.38	0.00	0.71	0.50	0.25
D1	0.62	0.47	−0.01	0.61	0.47	0.24
R	0.59	0.48	−0.03	0.60	0.45	0.24

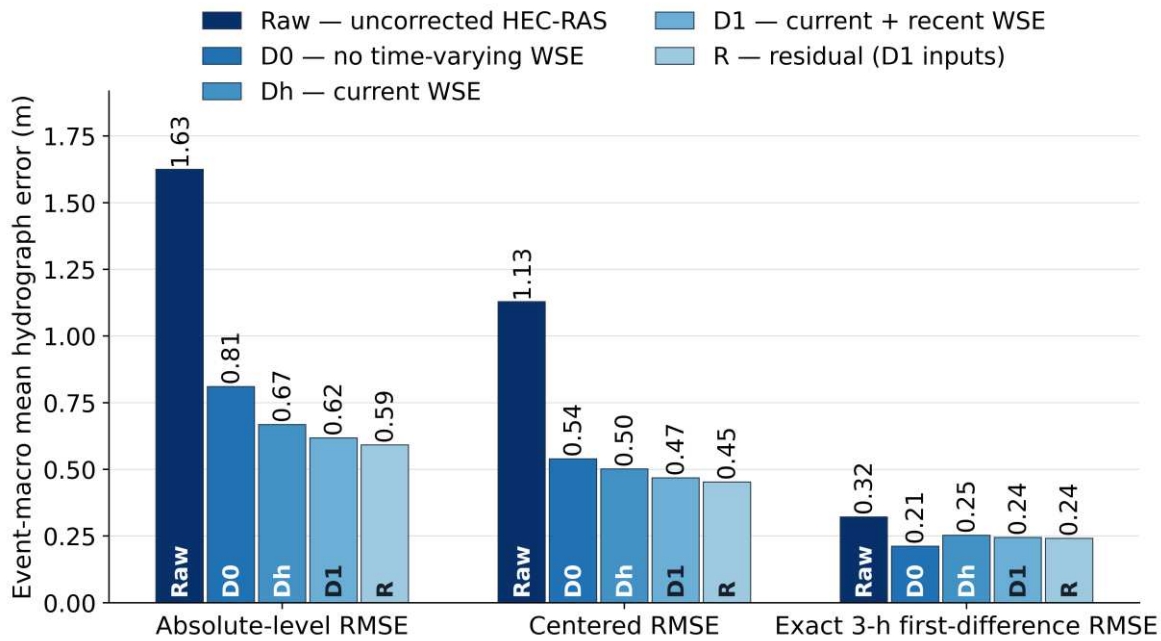
3.2 Hydrograph characteristics and peak performance

The endpoint analysis then checked whether the improvement in overall WSE error extended to bias, high water levels, centered hydrograph variation, short interval changes, and flood peaks. Across the identical test records, pooled bias changed from −0.42 m for uncorrected HEC-RAS to −0.05, 0.00, and −0.01 m for D0, Dh, and D1, while median hydrograph NSE increased from −0.20 to 0.18, 0.38, and 0.47 (Table 3). These metrics show that the improvement was not confined to RMSE.

395 The no-WSE model greatly improved overall RMSE but worsened error at the highest observed levels: its top-
 396 decile RMSE was 0.99 m versus 0.86 m for uncorrected HEC-RAS. Adding current WSE reduced that error to 0.71
 397 m, and recent history reduced it to 0.61 m. Thus, the largest high-water improvement came from adding current
 398 WSE.

399 Centered RMSE decreased from 0.54 m for D0 to 0.50 m for Dh and 0.47 m for D1 (Figure 4). One constant
 400 hydrograph offset does not explain the full WSE increment. The absolute-level reduction from D0 to Dh, however,
 401 was larger than the centered reduction, indicating that much of the improvement concerned flood-specific level
 402 rather than short-timescale variation.

403 Exact 3-h first-difference RMSE gave the opposite ranking. D0 was best at 0.21 m, compared with 0.25 m for Dh
 404 and 0.24 m for D1. The first-difference metric can favor smoother predictions and does not establish that D0
 405 reproduced timing correctly.



406 **Fig. 4** Hydrograph performance after excluding each complete test flood from training. Equal-flood mean hydrograph root-
 407 mean-square error (RMSE) is shown for absolute water-surface elevation (WSE), WSE centered within each hydrograph, and
 408 exact 3-h WSE differences
 409

410 Peak magnitude and timing derived directly from the predicted hydrographs showed limited improvement.
 411 Uncorrected HEC-RAS had a peak magnitude mean absolute error (MAE) of 0.50 m, compared with 0.94, 0.56,
 412 0.54, and 0.51 m for D0, Dh, D1, and R, respectively. Peak timing errors were 13–18 h and remained large after
 413 excluding hydrographs whose paired support observed peaks were within 12 h of a flood boundary.

414 The prespecified interior-peak subset contained 99 hydrographs from seven floods; Isaac supplied none.
 415 Uncorrected HEC-RAS still had the lowest trajectory-derived peak-magnitude error, 0.46 m versus 0.57–1.00 m for
 416 the statistical hydrographs, and timing errors remained 14.15–18.82 h. The poor timing result was therefore not
 417 explained only by peaks at the fixed flood boundaries.

418 The separate peak specific experiment used the same eight completed floods and flood based outer folds but
 419 answered a different question. It predicted one peak magnitude per hydrograph from full flood summaries rather
 420 than the continuous 3-h trajectory and did not estimate peak time. A training flood mean correction of the
 421 uncorrected peak scored 0.48 m. Adding hydraulic level summaries in a gradient boosting blend reduced error to
 422 0.45 m, and temporal shape summaries reduced it to 0.44 m. The final estimator improved six of eight floods. The
 423 estimator was selected from the same development folds and provided neither calibrated uncertainty nor a design
 424 bound.

3.3 Performance when complete gauges were excluded from training

Performance at locations excluded from training was assessed next. When every record from the test gauge was excluded from preprocessing, model selection, and fitting, the no-WSE model lost nearly all of its familiar gauge improvement and reached 1.62 m, the current WSE and history models scored 0.96 and 0.87 m (Table 4; Figure 5). D0 was also worse than uncorrected HEC-RAS when a regional system, a flood plus gauge, or a flood plus regional system was excluded from training. Current WSE restored a substantial average improvement in each design, and recent history added a smaller improvement.

When each regional system was excluded from training, D0, Dh, and D1 scored 1.85, 1.21, and 1.05 m. When both a flood and gauge were excluded from training, they scored 1.64, 0.99, and 0.93 m; excluding both a flood and regional system gave 1.88, 1.23, and 1.05 m. Uncorrected HEC-RAS remained 1.63 m because all spatial designs used the same primary records. These event-macro results weight floods equally, not locations.

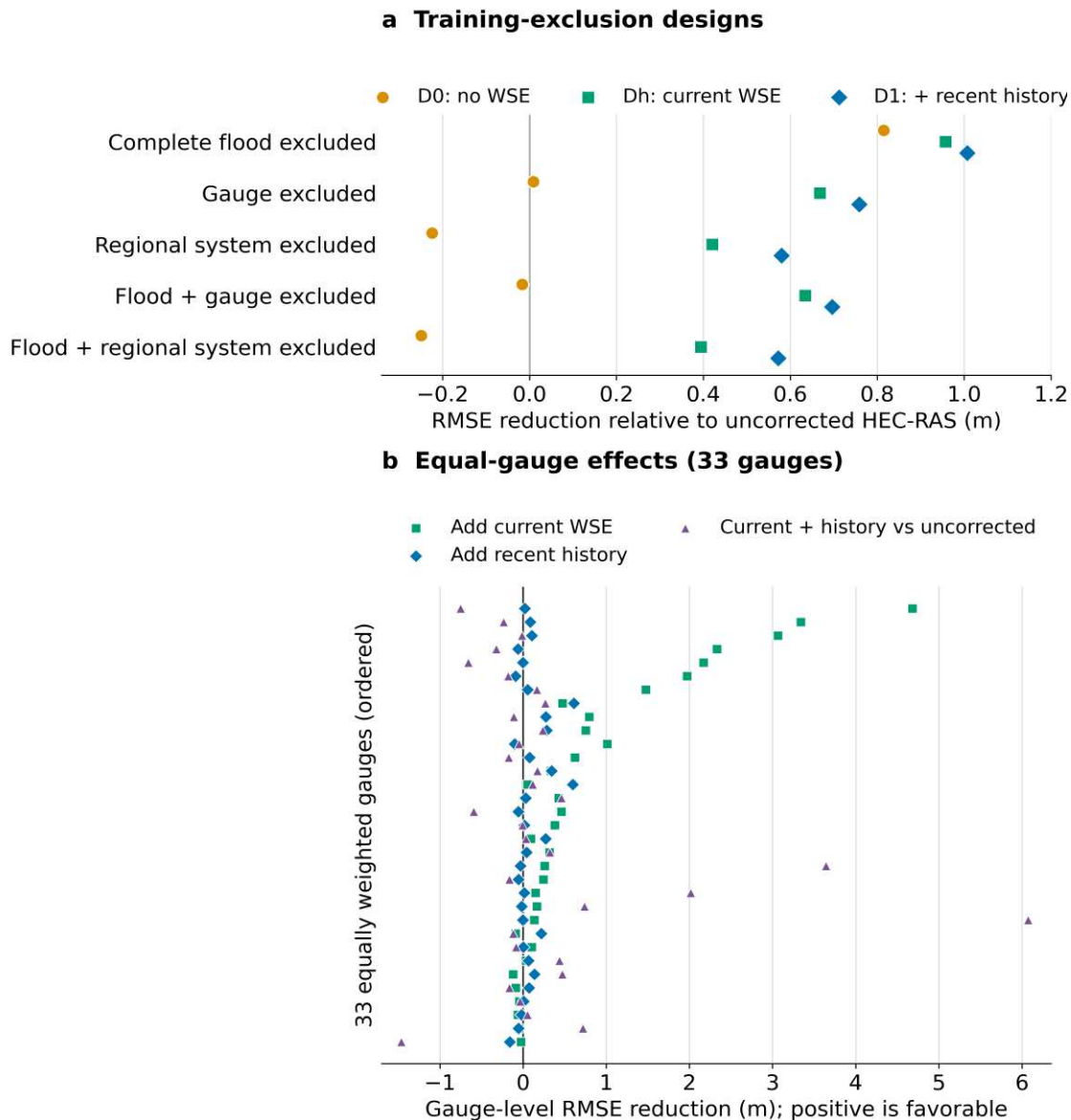


Fig. 5 Performance when floods and locations were excluded from training. (a) Root-mean-square error (RMSE) reduction relative to uncorrected Hydrologic Engineering Center's River Analysis System (HEC-RAS) output. (b) Equal-gauge effects from adding current water-surface elevation (WSE), adding recent WSE history, and comparing D1 with uncorrected output. Equal-gauge weighting preserved the mean D0–Dh–D1 ordering but revealed substantial heterogeneity. Uncorrected HEC-RAS, D0, Dh, and D1 scored 1.30, 1.83, 1.06, and 0.98 m. Dh improved D0 at 26 of 33 gauges,

and D1 improved Dh at 21. For the uncorrected-minus-D1 effect, D1 was better at only 16 gauges, the median was -0.01 m, and the mean was 0.33 m from the unrounded gauge scores. The mean remained positive when any one gauge effect was omitted and fell to 0.07 m after trimming 10% from each tail. The three largest gains occurred at deeply incised lower-Amite mapped cells. The transfer benefit was positive on average but concentrated and did not occur at most gauges.

The median D0-to-Dh gauge reduction was 0.32 m, with an interquartile range of 0.05 – 0.80 m. The median Dh-to-D1 reduction was 0.02 m, with an interquartile range of -0.02 – 0.11 m. Uncorrected minus D1 effects ranged from a 1.47 m deterioration to a 6.07 m improvement. This distribution explains why the mean and majority-gauge interpretations differ.

Equal-system weighting gave 1.24 , 1.79 , 1.23 , and 1.11 m for uncorrected HEC-RAS, D0, Dh, and D1. Both sequential additions lowered mean RMSE in five of six systems, while D1 improved on uncorrected HEC-RAS in three. These system groups follow model-area names and are not independent watershed delineations.

Table 4. Key flood-transfer, location-transfer, and robustness results (RMSE, m).

Evaluation design or analysis set	Scope	Uncorrected HEC-RAS	D0	Dh	D1
Complete flood excluded from training	130 hydrographs	1.63	0.81	0.67	0.62
Complete gauge excluded from training	33 folds	1.63	1.62	0.96	0.87
Regional model system excluded from training	6 folds	1.63	1.85	1.21	1.05
Flood and gauge jointly excluded from training	130 folds	1.63	1.64	0.99	0.93
Gauges weighted equally	Score-only summary	1.30	1.83	1.06	0.98
Systems weighted equally	Score-only summary	1.24	1.79	1.23	1.11
Observation offsets supported by direct documentation	46 hydrographs	1.35	1.11	0.74	0.74
Lower-Amite subdomain excluded	98 hydrographs	0.88	0.80	0.71	0.66
High or moderate mapping proximity	80 hydrographs	1.92	0.97	0.75	0.70
Maximum local accounting error $\leq 20\%$	107 hydrographs	1.34	0.87	0.78	0.75

3.4 Robustness, support, and physical limitations

Sensitivity analyses then checked whether the D0–Dh–D1 ordering changed with observation reference uncertainty, gauge to cell mapping, analysis choices, or hydraulic-model quality. The stricter observation-offset cohort showed the same asymmetry as the primary analysis. Current WSE reduced RMSE from 1.11 to 0.74 m; recent history changed the displayed value by less than 0.01 m. Across 32 bounded observation-offset perturbation refits, the current-WSE reduction ranged from 0.13 to 0.17 m and the history reduction from 0.02 to 0.07 m. Removing event phase, replacing the trailing range with a strict 12-h range, excluding 12 h from both flood boundaries, and fitting every fixed gradient-boosting complexity also preserved the D0–Dh–D1 ordering.

Removing March 2015 produced D0, Dh, and D1 scores of 0.85 , 0.71 , and 0.66 m; removing August 2016 produced 0.84 , 0.67 , and 0.64 m. A strict 12-h range gave 0.61 m for D1, while the no-event-phase refit gave 0.83 , 0.66 , and 0.61 m for D0, Dh, and D1. Adding missing-lag indicators left the two-decimal D1 score unchanged at 0.62 m.

The six-flood score excluding August 2016 and Ida from the test-flood average was 1.74 , 0.73 , 0.62 , and 0.57 m for uncorrected HEC-RAS, D0, Dh, and D1. Current WSE improved all six test floods and recent history improved five. The six-flood summary reduces dependence on the two floods known to have informed pre-analysis configuration comparisons, but complete earlier calibration provenance remains unavailable.

Mapping strongly affected absolute error. On 4,490 records supported at all five fixed candidate cells, uncorrected RMSE ranged from 0.96 to 1.81 m, whereas D0, Dh, and D1 ranged from 0.71–0.77, 0.58–0.62, and 0.56–0.59 m. Both sequential increments were favorable at every candidate rank, but no rank was declared hydraulically correct. At the Amite River near Highway 22, the designated cell minimum was –9.93 m and produced a first supported Isaac WSE of –9.85 m versus an observation of 0.46 m. A nearby positive-terrain candidate produced 0.62 m. The designated cell remained because alternative selection from observed agreement would violate the observation-independent mapping rule.

Excluding the lower-Amite subdomain reduced uncorrected HEC-RAS RMSE from 1.63 to 0.88 m and retained a D0-to-D1 reduction from 0.80 to 0.66 m. The proximity-screened refit retained 3,209 records from 80 hydrographs at 19 gauges and scored 0.97, 0.75, and 0.70 m for D0, Dh, and D1. The lower-Amite exclusion and proximity-screened refit preserve the information ordering while showing that absolute hydraulic error and the magnitude of statistical benefit depend strongly on mapping.

Statistical corrections were sometimes too large for unconstrained use. For D1, the median absolute adjustment was 0.52 m, the 95th percentile was 2.37 m, 2.7% exceeded 5 m, and the maximum was 11.33 m. About 10.2% of D1 predictions fell below the mapped cell minimum before constraints. Applying the terrain floor changed RMSE from 0.62 to 0.64 m. A training-derived 5 m correction cap increased it to 0.98 m, showing that large corrections contributed materially to the lower development error.

The primary analysis covered only model-supported records. Within the same 130 eligible hydrographs, adding 1,406 model-dry observations produced 6,540 records. A constrained hybrid used D1 where WSE was supported and a separately fitted D0 elsewhere; it scored 0.70 m and improved all eight floods. The hybrid did not cover excluded hydrographs or the complete 15,540-output schedule, and full-window peak-timing error remained 36.53 h. This larger timing error used the expanded full-window calculation, whereas the 13–18 h values used only common supported timestamps.

Whole-domain volume-accounting errors were below 5%, but maximum local errors ranged from 10.3% to 140.9%. Restricting the analysis to five floods with maximum local error no greater than 20% retained the model ordering, although the history increment was smaller (Table 4). Observation-blind repair runs did not meet the local criteria specified before those runs and never entered statistical training. For Francine, diagnostic runs at 5, 10, and 15 s did not establish stable trajectories under computation-step refinement. A separate August pilot tested saved-output spacing: common-time WSE from hourly and 3-h output runs differed by as much as 0.04 m, and one support classification changed. The pilot was stopped rather than extended or scored against observations.

4. Discussion

4.1 Information supplied by current simulated WSE

The largest and most consistent increment came from current HEC-RAS WSE. Current WSE lowered error in every test flood, improved high-water performance, and retained a substantial mean advantage when complete gauges were excluded from training. By contrast, the no-WSE model lost nearly all of its familiar-gauge improvement when the test gauge was excluded from training. Results at familiar gauges alone cannot identify the information supplied by the process model.

Earlier studies used different targets, lead times, station networks, and aggregation rules, so their reported errors are not directly comparable. National Water Model postprocessing reduced median streamflow RMSE by 25% across 30 gauges, while storm-surge correction gains weakened from 36% at a 1-h lead to 5% at 60 h (Neisary et al. 2025; Tedesco et al. 2024). A neural-operator correction of coastal guidance reported 0.13 m overall RMSE and 0.14 m at stations excluded from training (Alipour et al. 2026). In this study, the learner and test records remain fixed while current WSE and recent history are added in tests that exclude complete floods and locations from training.

Current WSE integrates the simulated effects of rainfall, boundary stage, storage, routing, and local geometry at the prediction time. Its value is not simply access to the boundary series because the same stages and changes were

516 supplied to every matched model. Simulated WSE may nevertheless encode those inputs through nonlinear routing
517 and storage, and the archived boundary stages are not proven independent of all observational influence.

518 D0 reduced ordinary level error but worsened top-decile RMSE relative to uncorrected HEC-RAS. Adding current
519 WSE reversed that deterioration. Gauge identity and forcing can approximate typical behavior at familiar locations,
520 but the evolving simulated state carries information about which flood and time are hydraulically high.

521 When each test gauge was excluded from training, D1 outperformed uncorrected HEC-RAS at 16 of 33 gauges, and
522 the trimmed mean was much smaller than the untrimmed mean. The largest benefits occurred at sites with extreme
523 mapping offsets. Simulated WSE retained predictive value without gauge-specific training data, but the benefit
524 varied widely among gauges.

525 **4.2 Recent history and peak performance were secondary**

526 Recent history added 0.05 m of RMSE reduction in the primary cohort and improved seven floods, but the
527 increment was less than 0.01 m in the subset with observation offsets supported by direct documentation. A
528 changes-only formulation also outperformed the complete prespecified history bundle. Recent WSE therefore
529 contains some useful information, but its contribution is smaller, representation-sensitive, and less stable than the
530 current-WSE increment.

531 The smaller history increment is plausible because current WSE already summarizes the preceding response of the
532 model. The common predictors also include rainfall history, boundary changes, and elapsed time. Exact lags and
533 changes can therefore be redundant at a 3-h sampling interval, particularly when intermittent model support creates
534 missing inputs. The results do not justify treating the complete history bundle as an optimal temporal representation.

535 Overall RMSE concealed an important trade-off: current WSE improved high-water and centered errors, but exact
536 3-h differences did not improve beyond the no-WSE model. Much of the current-WSE gain concerned flood-
537 specific level, although the centered improvement shows that one constant offset was not the whole explanation.
538 Incomplete vertical-reference documentation prevents assigning that level component to one physical mechanism.

539 D0 already includes gauge identity and mapped-cell minimum elevation, so it can learn a persistent offset at
540 familiar gauges. The remaining level error can vary among floods with storage, boundary routing, antecedent state,
541 and mapping behavior. Incomplete reference information prevents separating these factors from datum error.

542 Peak magnitude also depends on the training target. Maxima extracted from models trained for continuous
543 hydrograph RMSE did not beat uncorrected HEC-RAS. The separate peak-specific model did reduce magnitude
544 error to 0.44 m, but it was selected from the same eight development floods and did not estimate peak time. The
545 peak-specific result supports a focused magnitude finding from completed floods, not operational timing or a
546 calibrated design level.

547 **4.3 Hydraulic support, mapping, and evaluation priorities**

548 The statistical results do not repair the hydraulic model. The configuration omits explicit river inflow and baseflow,
549 uses simplified boundary stages for several floods, shares one initial state, and has incomplete calibration
550 provenance. Local volume errors reached 140.9%, the Francine computation-step diagnostics did not establish
551 stable trajectories, and the HEC-RAS 6.6 repeatability concern remains unresolved. Lower postprocessed RMSE
552 does not validate volumes, connectivity, inundation, or timing (Beven and Binley 1992; Wagener et al. 2001).

553 The model terrain is intended to represent NAVD88 but its realization is incompletely documented, and geometry-
554 based containment does not establish hydraulic equivalence between a gauge and cell. Candidate-cell tests
555 preserved the model ordering but changed uncorrected error substantially. A specialist review of channel branch,
556 levee side, structures, and hydraulic control is required before a mapping can be considered operationally
557 defensible.

558 The largest gains when gauges were excluded from training occurred near deeply incised lower-Amite cells. At
559 these locations, simulated WSE helped estimate the state-dependent size of the fixed mapping offset. The near-zero
560 median D1-versus-uncorrected effect indicates that this benefit was not typical across gauges.

561 The support rule also selects the better-represented part of the schedule. One-third of scheduled mapped outputs
562 lacked supported WSE, and the primary hydrographs required at least 25 supported pairs. The 0.62 m result is
563 therefore conditional on hydraulic support. Statistical fallback expanded coverage only within hydrographs already
564 eligible from their observed record and did not correct simulated wetting or arrival time.

565 Some of the reported error reduction depends on corrections above 5 m and predictions below mapped terrain
566 before constraints. Floors and caps reduce these violations but do not enforce conservation or momentum. These
567 large corrections and physical violations limit operational use.

568 These findings suggest a practical evaluation sequence for process-model postprocessing: compare against a
569 matched model without time-varying process output, exclude complete floods and locations from training, and then
570 examine support, mapping, elevation reference, and physical violations. For this configuration, improving forcing,
571 local volume balance, elevation documentation, and gauge-to-cell mapping should precede further expansion of
572 recent-history predictors. A genuinely new flood and hydraulically reviewed mappings are needed for the next
573 evaluation.

574 **5. Conclusions**

575 Current simulated WSE provided the most consistent information gain across completed floods and remained useful
576 on average when gauge-specific training data were unavailable. Recent history added a smaller and representation-
577 dependent increment. The hydrograph models improved absolute levels and high-water performance, whereas peak
578 magnitude improved only in the separate exploratory peak-specific analysis and peak timing remained poor. The
579 average gauge-exclusion benefit was concentrated at lower-Amite sites with severe mapping offsets and was not
580 typical across gauges. All findings remain conditional on supported WSE, the fixed mapping, incomplete elevation
581 references, and the imperfect hydraulic configuration. Future work should prioritize forcing, local volume balance,
582 hydraulically reviewed gauge-to-cell assignments, and evaluation on a genuinely new flood before considering
583 operational use.

584 **Declarations**

585 **Competing interests**

586 On behalf of all authors, the corresponding author states that there is no conflict of interest.

587 **Funding**

588 The authors received no funding for this work.

589

590 References

- 591 Acuña Espinoza E, Loritz R, Kratzert F, Klotz D, Gauch M, Álvarez Chaves M, Ehret U (2025) Analyzing the generalization
592 capabilities of a hybrid hydrological model for extrapolation to extreme events. *Hydrology and Earth System Sciences*
593 29:1277–1294. <https://doi.org/10.5194/hess-29-1277-2025>
- 594 Alipour A, Daneshvar F, Mani S, Seroka G, Shi L, Yang Z, Funakoshi Y, Britzolakis G, Myers EP, Azizzadenesheli K,
595 Moghimi S (2026) Neural operators for continuous bias correction of water level forecast guidance. *Journal of Geophysical*
596 *Research: Machine Learning and Computation* 3(1):e2025JH000853. <https://doi.org/10.1029/2025JH000853>
- 597 Bergmeir C, Hyndman RJ, Koo B (2018) A note on the validity of cross-validation for evaluating autoregressive time series
598 prediction. *Computational Statistics & Data Analysis* 120:70–83. <https://doi.org/10.1016/j.csda.2017.11.003>
- 599 Beven K, Binley A (1992) The future of distributed models: model calibration and uncertainty prediction. *Hydrological*
600 *Processes* 6(3):279–298. <https://doi.org/10.1002/hyp.3360060305>
- 601 Fall G, Kitzmiller D, Pavlovic S, Zhang Z, Patrick N, St. Laurent M, Trypaluk C, Wu W, Miller D (2023) The Office of Water
602 Prediction's Analysis of Record for Calibration, version 1.1: dataset description and precipitation evaluation. *Journal of the*
603 *American Water Resources Association* 59(6):1246–1272. <https://doi.org/10.1111/1752-1688.13143>
- 604 Giaremis S, Nader N, Dawson C, Kaiser C, Nikidis E, Kaiser H (2024) Storm surge modeling in the AI era: Using LSTM-based
605 machine learning for enhancing forecasting accuracy. *Coastal Engineering* 191:104532.
606 <https://doi.org/10.1016/j.coastaleng.2024.104532>
- 607 Haces-Garcia F, Ross N, Glennie CL, Rifai HS, Hoskere V, Ekhtari N (2025) Rapid 2D hydrodynamic flood modeling using
608 deep learning surrogates. *Journal of Hydrology* 651:132561. <https://doi.org/10.1016/j.jhydrol.2024.132561>
- 609 Hammad M, Mehrotra R, Sharma A (2026) Can state-dependant residual modeling improve legacy hydrological model
610 simulations? *Water Resources Research* 62(7):e2026WR044285. <https://doi.org/10.1029/2026WR044285>
- 611 Holmberg E, Pokhrel P, Ioup E, Pathak K, Sloan S, Niles K, Ratcliff J, Flanagan M, Guetl C, Simeonov J, Abdelguerfi M (2025)
612 Accelerating HEC-RAS: A recurrent neural operator for rapid river forecasting. In: *COCE 2025: The Second International*
613 *Conference on Technologies for Marine and Coastal Ecosystems*, pp 38–45. IARIA.
614 https://www.thinkmind.org/library/COCE/COCE_2025/coce_2025_1_70_30035.html Accessed 21 Oct 2026
- 615 Holmberg E, Ioup E, Ferdaus MM, Abdelguerfi M, Simeonov J (2026) Learned response-field inertia operator for HEC-RAS
616 2D water-surface elevation prediction. *arXiv preprint*. arXiv:2606.06385. <https://doi.org/10.48550/arXiv.2606.06385>
- 617 Homer CG, Dewitz JA, Yang L, Jin S, Danielson P, Xian GZ, Coulston J, Herold ND, Wickham JD, Megown K (2015)
618 Completion of the 2011 National Land Cover Database for the conterminous United States—representing a decade of land-
619 cover change information. *Photogrammetric Engineering & Remote Sensing* 81(5):345–354.
620 <https://pubs.usgs.gov/publication/70146301> Accessed 21 Oct 2026
- 621 Huang S, Xia J, Wang Y, Wang W, Zeng S, She D, Wang G (2022) Coupling machine learning into hydrodynamic models to
622 improve river modeling with complex boundary conditions. *Water Resources Research* 58(10):e2022WR032183.
623 <https://doi.org/10.1029/2022WR032183>
- 624 Jafarzadegan K, Moradkhani H, Pappenberger F, Moftakhari H, Bates P, Abbaszadeh P, Marsooli R, Ferreira C, Cloke HL,
625 Ogden F, Duan Q (2023) Recent advances and new frontiers in riverine and coastal flood modeling. *Reviews of Geophysics*
626 61(2):e2022RG000788. <https://doi.org/10.1029/2022RG000788>
- 627 Klemes V (1986) Operational testing of hydrological simulation models. *Hydrological Sciences Journal* 31(1):13–24.
628 <https://doi.org/10.1080/02626668609491024>
- 629 Li W, Duan Q, Miao C, Ye A, Gong W, Di Z (2017) A review on statistical postprocessing methods for hydrometeorological
630 ensemble forecasting. *WIREs Water* 4(6):e1246. <https://doi.org/10.1002/wat2.1246>
- 631 Li Y, Ye Q, Zhang S, Liu T, Zeng L (2026) Physics-informed operator learning for rapid hydrograph-to-field prediction on an
632 unstructured-mesh flood benchmark. *Journal of Hydrology* 678:136017. <https://doi.org/10.1016/j.jhydrol.2026.136017>
- 633 Luettich RA Jr, Westerink JJ (2004) *Formulation and numerical implementation of the 2D/3D ADCIRC finite element model,*
634 *version 44.XX*. https://adcirc.org/wp-content/uploads/sites/2255/2018/11/adcirc_theory_2004_12_08.pdf Accessed 21 Oct
635 2026
- 636 Merz B, Kreibich H, Schwarze R, Thieken A (2010) Review article 'Assessment of economic flood damage.' *Natural Hazards*
637 *and Earth System Sciences* 10:1697–1724. <https://doi.org/10.5194/nhess-10-1697-2010>
- 638 Meyer H, Reudenbach C, Hengl T, Taturji M, Nauss T (2018) Improving performance of spatio-temporal machine learning
639 models using forward feature selection and target-oriented validation. *Environmental Modelling & Software* 101:1–9.
640 <https://doi.org/10.1016/j.envsoft.2017.12.001>
- 641 Nader N, Majed H, Giaremis S, El Osta R, Dawson C, Kaiser C, Kaiser H (2026a) HURRI-GAN: A Novel Approach for
642 Hurricane Bias-Correction Beyond Gauge Stations using Generative Adversarial Networks. *arXiv preprint* 2603.06649.
643 <https://doi.org/10.48550/arXiv.2603.06649>
- 644 Nader N, Giaremis S, Dawson C, Kaiser C, Mohammadiporshokoh K, Kaiser H (2026b) StormNet: Improving storm surge
645 predictions with a GNN-based spatio-temporal offset forecasting model. *arXiv preprint* 2604.20688.
646 <https://doi.org/10.48550/arXiv.2604.20688>
- 647 Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models, part I: a discussion of principles. *Journal of*
648 *Hydrology* 10(3):282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)

- Neisary SN, Johnson RC, Alam MS, Burian SJ (2025) A post-processing machine-learning framework for bias-correcting National Water Model outputs by accounting for dominant streamflow drivers. *Environmental Modelling & Software* 190:106459. <https://doi.org/10.1016/j.envsoft.2025.106459>
- Pakdehi M, Ahmadisharaf E, Nazari B, Cho E (2024) Transferability of machine-learning-based modeling frameworks across flood events for hindcasting maximum river water depths in coastal watersheds. *Natural Hazards and Earth System Sciences* 24:3537–3559. <https://doi.org/10.5194/nhess-24-3537-2024>
- Papacharalampous GA, Tyralis H (2022) A review of machine learning concepts and methods for addressing challenges in probabilistic hydrological post-processing and forecasting. *Frontiers in Water* 4:961954. <https://doi.org/10.3389/frwa.2022.961954>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É (2011) Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830. <https://jmlr.org/papers/v12/pedregosa11a.html> Accessed 21 Oct 2026
- Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, Hauenstein S, Lahoz-Monfort JJ, Schröder B, Thuiller W, Warton DI, Wintle BA, Hartig F, Dormann CF (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40:913–929. <https://doi.org/10.1111/ecog.02881>
- Scikit-learn developers (2026) *HistGradientBoostingRegressor*, *scikit-learn 1.9.0 documentation*. <https://scikit-learn.org/1.9/modules/generated/sklearn.ensemble.HistGradientBoostingRegressor.html> Accessed 21 Oct 2026
- Tedesco P, Rabault J, Sætra ML, Kristensen NM, Aarnes OJ, Breivik Ø, Mauritzen C, Sætra Ø (2024) Bias correction of operational storm surge forecasts using neural networks. *Ocean Modelling* 188:102334. <https://doi.org/10.1016/j.ocemod.2024.102334>
- Teng J, Jakeman AJ, Vaze J, Croke BFW, Dutta D, Kim S (2017) Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software* 90:201–216. <https://doi.org/10.1016/j.envsoft.2017.01.006>
- US Army Corps of Engineers, Hydrologic Engineering Center (2024) *HEC-RAS known issues, version 6.6*. <https://www.hec.usace.army.mil/confluence/rasdocs/raski/6.6> Accessed 21 Oct 2026
- US Army Corps of Engineers, Hydrologic Engineering Center (n.d.-a) *Computation options and tolerances*. <https://www.hec.usace.army.mil/confluence/rasdocs/r2dum/6.6/running-a-model-with-2d-flow-areas/2d-computation-options-and-tolerances> Accessed 21 Oct 2026
- US Army Corps of Engineers, Hydrologic Engineering Center (n.d.-b) *HEC-RAS 2D User's Manual, version 6.6*. https://www.hec.usace.army.mil/software/hecras/documentation/HEC-RAS_2D_Users_Manual_v6.6.pdf Accessed 21 Oct 2026
- US Army Corps of Engineers, Hydrologic Engineering Center (n.d.-c) *HEC-RAS 6.6 Documentation*. <https://www.hec.usace.army.mil/software/hecras/documentation.aspx> Accessed 21 Oct 2026
- US Department of Agriculture, Natural Resources Conservation Service (2004) *National Engineering Handbook, Part 630 Hydrology, Chapter 10: Estimation of direct runoff from storm rainfall*. <https://directives.nrcs.usda.gov/sites/default/files/201712930608/7300.pdf> Accessed 21 Oct 2026
- US Geological Survey (2020) *WMA 2020.03: Guidance on the display and handling of continuous water-surface elevation data for stations with an accurate geodetic vertical datum*. <https://water.usgs.gov/water-resources/memos/memo.php?id=4449> Accessed 21 Oct 2026
- US Geological Survey (2026) *Continuous values*. <https://api.waterdata.usgs.gov/ogcapi/v0/collections/continuous> Accessed 21 Oct 2026
- Wagener T, Boyle DP, Lees MJ, Wheeler HS, Gupta HV, Sorooshian S (2001) A framework for development and application of hydrological models. *Hydrology and Earth System Sciences* 5(1):13–26. <https://doi.org/10.5194/hess-5-13-2001>
- Wahl T, Jain S, Bender J, Meyers SD, Luther ME (2015) Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nature Climate Change* 5:1093–1097. <https://doi.org/10.1038/nclimate2736>
- Ward PJ, Couasnon A, Eilander D, Haigh ID, Hendry A, Muis S, Veldkamp TIE, Winsemius HC, Wahl T (2018) Dependence between high sea-level and high river discharge increases flood hazard in global deltas and estuaries. *Environmental Research Letters* 13:084012. <https://doi.org/10.1088/1748-9326/aad400>