Emulating Hydrodynamic Models with Neural Operators for Rapid Storm Assessment *

Emma L. McDaniel†
Human Resilience Technology Group
MIT Lincoln Laboratory
emma.mcdaniel@ll.mit.edu

Jeffrey Liu†
Human Resilience Technology Group
MIT Lincoln Laboratory
jeffrey.liu@ll.mit.edu

Chad Council
Human Resilience Technology Group
MIT Lincoln Laboratory
chad.council@ll.mit.edu

Jenny Rowlett
Human Resilience Technology Group
MIT Lincoln Laboratory
jinmei.rowlett@ll.mit.edu

Teresa Fazio
Human Resilience Technology Group
MIT Lincoln Laboratory
teresa.fazio@ll.mit.edu

Julia Hopkins
Human Resilience Technology Group
MIT Lincoln Laboratory
julia.hopkins@ll.mit.edu

ABSTRACT

Timely, accurate, and trustworthy flood models for rapid decision making during and after extreme storms do not exist in an operational capacity, as flood models often compromise between physical accuracy and speed. Physics-based flood models that explicitly simulate mass and momentum balances can be accurate, but are prohibitively computationally expensive. Stochastic models built from synthesizing sparse amounts of existing flood data are timely, but do not model the underlying hydrodynamics of a flood, limiting simulations to water levels alone. Capturing the full impact of a flood disaster requires models that include strong currents that can harm people, flood momentum that damages infrastructure, and debris flows that can block water outflows and evacuation routes. To address this operational gap, we are developing a near-real-time "nowcast" flood disaster simulation fusing physics-based models with machine learning. As a part of this effort, we are creating training datasets using physics-based hydrodynamic models verified against recent historical flood events in the United States. These data will inform a timely flood model that simulates more than water levels: flow properties such as velocity can improve damage and risk calculations, as well as situational awareness of regions that could be hazardous during a flood. With this nowcast, we aim to build the timely, trustworthy foundation of tools that decision-makers can use to prepare for, respond to, and recover from flood disasters.

Keywords

Flood Modeling, Flood Nowcasting, Neural Surrogates, Physics Emulators, Graph Neural Network, Neural Operator, Climate Resilience, Flood Risk, Decision Support, Floods,Work-in-progress

†corresponding and co-first authors

WiP Paper – Natural Hazards in a Changing World
Proceedings of the 23rd ISCRAM Conference – The Hague, the Netherlands June 2026
Caroline Rizza, Apoorva Chauhan, Amy Matser, Joyce Kox, Willem Treurniet and Jeroen Wolbers eds.

INTRODUCTION

Flood modeling provides critical support for decision-making throughout the flood disaster lifecycle (Hammond et al. 2015). For mitigation and preparedness, models can enable proactive identification of infrastructure vulnerabilities, optimization of evacuation networks, and prioritization of floodplain management policies and physical infrastructure upgrades before a storm event. During active disasters, real-time model outputs can support emergency managers by providing dynamic hazard assessments that can be used for resource allocation and infrastructure risk evaluation as events unfold. In the recovery phase, impact maps derived from simulated water inundation, water velocity, and sediment transport can assist in identifying severely damaged areas as well as supporting damage-assessment teams, search and rescue operations, and reconstruction efforts.

Despite the importance of flood modeling, existing approaches face significant limitations in operational environments. Physics-based simulations offer high physical fidelity by solving governing hydrodynamic equations to represent water depth, flow velocity, momentum, and debris transport. However, the computational demands for complex simulations result in runtimes of hours to days, making these models impractical for real-time disaster response. Physics-based models appropriate for simulating 2D overland flood impacts include HEC-RAS (Brunner 2020) and LISFLOOD-FP (Joint Research Center 2020), which illustrate two different physics-based approaches to flood modeling. HEC-RAS solves the shallow water Reynold's Averaged Navier Stokes (RANS) equations, which balances mass, energy, and momentum throughout the domain. LISFLOOD applies a simpler set of equations, relying primarily on mass balance to distribute flood waters. Both approaches, with the correct forcing and topographical inputs, have been shown to provide reasonable estimates of flood extent (Dimitriadis et al. 2016; Horritt and Bates 2002), though HEC-RAS directly simulates the momentum that impacts infrastructure and debris transport. Physics-based models provide high physical accuracy but suffer from computational inefficiency, particularly for ensemble simulations or high-resolution analyses.

Alternatively, data-driven, empirical models are built on historical observations of flood impacts, measurements which are often sparse and limited to water depth as this can be measured post-event (e.g. using water marks) (Marvi 2020). The United States Geological Survey (USGS) maintains a network of 8,705 stream gages measuring streamflow, and an additional 3,460 measuring water depth (United States Geological Survey 2026). This provides a reasonable snapshot of riverine processes during flood events, but does not extend to overland flow where water encounters varying topography and landcover. While these statistical models can have runtimes of minutes, the lack of overland flow observations in their training prevents them from capturing critical aspects of flood damage such as water velocity and sediment transport.

Observation- or remote sensing-derived flood extent products can also provide valuable inputs to the emergency manager in the post-flood time line, but such products can be expensive, slow to acquire, and inconsistently available, making it challenging to develop response or recovery processes around them. Machine learning approaches trained on flood observations from gages or flood extent from opportunistic remote-sensing tend to inherit similar limitations, preventing direct inference of essential flood impact characteristics like water velocity. Balancing speed and physical fidelity remains a central challenge in flood-risk research, particularly for operational applications that span the full disaster life cycle. This represents a critical gap because water velocity, not just depth, is the primary driver of infrastructure damage, debris transport, and contamination pathways during flood events.

The standard practice of treating flood depth as a proxy for flood damage relies on a simplified assumption that higher water levels during a storm event translate to faster water and thus greater damage potential (Marvi 2020; Redondo-Tilano et al. 2025; Romali et al. 2025). The correlation between depth and damage depends on post-storm damage assessments corresponding to flood depth, if known, and can be useful for first-order damage predictions. Simulating the flood property directly responsible for damage, flow velocity, enables more accurate risk and damage analysis by quantifying infrastructure damage and debris transport mechanisms. Operationally, this capability would support disaster response and emergency management through real-time flood velocity forecasting and critical infrastructure risk assessment, directly informing evacuation routing, response resource allocation, and recovery prioritization.

Recent policy frameworks explicitly identify the need for technologies that support practical resilience building: the UN Climate Technology Centre and Network's Practitioner's Guide identifies real-time monitoring with forecasts, early warning systems, barrier placement, and hazard mapping as essential needs (CTCN 2017), while the Global Assessment Report on Disaster Risk Reduction 2019 emphasizes a shift toward impact-focused modeling that avoids reliance on historical data (McGlade et al. 2019). Therefore, improving speed for flood modeling in operational contexts while ensuring physical fidelity in water velocity prediction is essential for advancing decision-making in this realm.

This work-in-progress introduces ML-HYDRAS (Machine Learning-HYdrodynamic Driven Rapid Assessment of Storms), a near-real-time flood model designed to bridge the gap between physical fidelity and operational

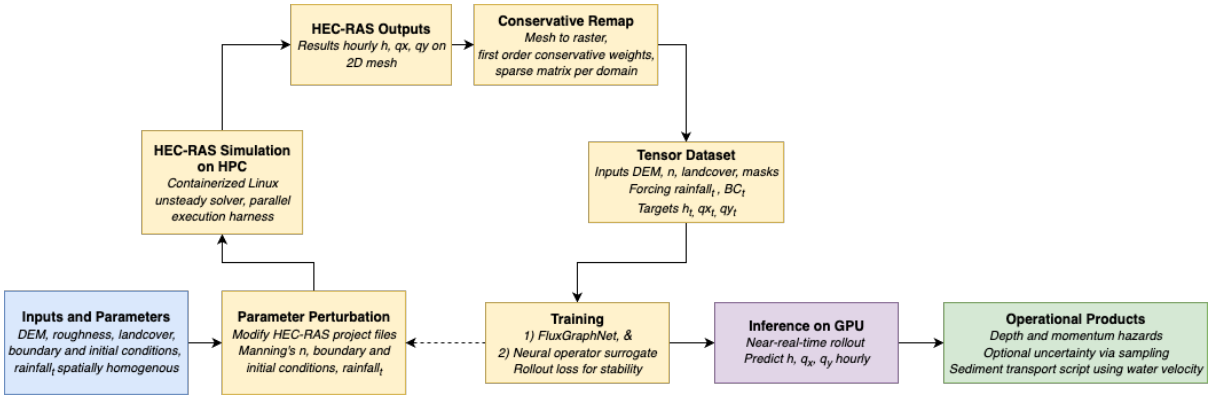

Figure 1. System overview of ML-HYDRAS: HEC-RAS simulations are perturbed and executed at scale on HPC. Two parallel neural network architectural pathways are explored: (1) FluxGraphNet, and (2) hybrid FNO-CNN surrogate. Both approaches predict hourly depth and discharge fields with optional uncertainty estimation.

speed. ML-HYDRAS combines physics-based principles with machine learning through two architectural efforts trained on synthetic datasets generated from HEC-RAS hydrodynamic simulations, preserving physical consistency while achieving inference times measured in minutes rather than hours. The first effort, FluxGraphNet, extends NVIDIA's HydroGraphNet physics-constrained graph neural network (GNN) approach to operate natively on unstructured HEC-RAS meshes by incorporating edge-level normal depth-integrated discharge (q_n) alongside node-level water depth (Taghizadeh et al. 2025). The second effort develops a hybrid Fourier Neural Operator (FNO) with Convolutional Neural Network (CNN) surrogate architecture that operates on conservatively remapped raster representations, enabling global spectral mixing while addressing non-periodic boundary conditions through local convolutional blocks. This raster-based approach addresses scalability challenges for larger domains where GNN memory requirements may become prohibitive, drawing inspiration from recent atmospheric forecasting models that apply FNO architectures to global weather prediction (Kovachki et al. 2023; Bonev et al. 2025). While recent efforts have demonstrated physics-informed neural networks for flood depth prediction (Feng et al. 2025; Yin et al. 2025), ML-HYDRAS extends this capability to depth-integrated momentum fields, enabling velocity-based hazard assessment essential for operational decision-making.

Both approaches endeavor to be geographically-agnostic, suitable for a wide range of geographic areas and complexities, while preserving physical fidelity and achieving rapid runtimes that output forecasts of multiple evolution states across the flooding event, enabling applications across preparedness, response, and recovery phases. In this paper, we document the following contributions:

- A containerized HEC-RAS data-generation pipeline for HPC.
- An extension of HydroGraphNet, FluxGraphNet, which incorporates edge-level normal depth-integrated discharge.
- A hybrid FNO–CNN surrogate architecture tailored to non-periodic hydrodynamic domains on raster grids.
- Training strategies emphasizing rollout stability and physical plausibility for both architectures.

METHODOLOGY

Hydrologic Model

Many physics-based models simulating the shallow water Navier Stokes equations exist (e.g. HydroMT, LISFLOOD-FP) and would be appropriate for this research. HEC-RAS (Hydraulic Engineer Center River Analysis System) was chosen owing to its physics infrastructure, numerical stability, extensive validation in the lab and field, as well as current use in operational projects by the United States Army Corp of Engineers (Brunner et al. 2020). HEC-RAS solves for 2D unsteady flow by conserving mass and momentum across all cells in the prescribed grid. Mass conservation includes horizontal flux through a grid cell, infiltration (water absorbed by the ground), and evapotranspiration. Momentum conservation includes contributions from hydrostatic pressure, wind shear stress, bed shear stress, and diffusion as an approximation for turbulent motion. The hydrostatic pressure and turbulence simplification are both hallmarks of the shallow water approximation, which assumes variations in

vertical momentum are negligible. Since we aim to emulate depth-integrated momentum fields, training on a solver that explicitly resolves velocity is critical.

The model requires detailed topography as the base for the computational grid, and boundary conditions prescribed where water is likely to flow in or out of the domain. Multiple processes such as evapotranspiration, infiltration, and precipitation can be included or excluded in the simulation depending on user preference. Precipitation is prescribed at discrete locations corresponding to spatially aggregated radar-derived rainfall rates over partitioned subregions; HEC-RAS then performs spatial interpolation across the computational domain. Evapotranspiration and infiltration can spatially vary depending on landcover values (including soil qualities) and are prescribed as constant rates in time.

Topography, landcover, and boundary flow rates were acquired from USGS databases of 1 meter digital elevation models and 30 meter landcover maps (U.S. Geological Survey 2019; U.S. Geological Survey 2024). The computational grid was drawn such that most edges bordered areas of high relief, allowing water to flow into the domain rather than out. Precipitation values for a given storm were derived from nearby radar stations and fed into the model as a time-varying, spatially-varying hourly rate.

The training data will be built from this initial model schema, modifying the boundary conditions (flow rates, precipitation) and the initial conditions (topography, landcover). At present, we initially focus on simulating two high relief areas recently subject to flood disasters: the Winooski River Watershed in the state of Vermont and the watershed surrounding the city of Asheville in North Carolina. Vermont has been subject to extensive flooding the last three summers (Whittle and Casey 2025) and as such there is information about flood extent and damage that can be used to validate the physics-based model. Asheville was flooded during Hurricane Helene in 2024 causing damage, landslides, and noted transport of contamination (PFAS, wastewater) in the freshwater system (Walker-Franklin et al. 2025; Scheip et al. 2026). Though these two flood-prone, mountainous areas are the initial focus, we plan to extend to modeling additional domains to provide additional diversity to training data and to test generalizability across held-out domains.

Training Dataset

To construct the training dataset, we plan on running many HEC-RAS simulations across different initial and boundary conditions on various geographic domains. To assist with the generation of this data, we have created a containerized runtime for the Linux HEC-RAS unsteady solver binary. This container easily allows us to run many simulations in parallel using the Lincoln Laboratory Supercomputing Center high performance computing (HPC) cluster (Reuther et al. 2018). We mount each run's project folder in each container via a bind mount to the host HPC filesystem. As the Linux HEC-RAS solvers lack a GUI, the project needs to be initialized using the Windows version of HEC-RAS.

To construct a training dataset that spans a physically meaningful range of flood behaviors, we generate synthetic design storms and systematically perturb hydrologic and hydraulic parameters across controlled ranges. Our objective is to sample a structured parameter space representative of realistic storm and watershed conditions, rather than relying solely on a small number of historical realizations.

We employ synthetic design storms parameterized by total accumulation, peak intensity, duration, and temporal distribution (e.g., front-loaded, symmetric, or back-loaded hyetographs). For the Vermont study domain, peak rainfall intensities are varied across a representative range of moderate to extreme events (e.g., approximately 10–100 mm/hr), with storm durations spanning multi-hour to multi-day events. Boundary discharge conditions are perturbed in a manner consistent with storm magnitude, ranging from nominal baseflow conditions to high-flow scenarios informed by regional flood-frequency analyses. This sampling strategy exposes the model to both overbank flooding and higher-energy channel-driven regimes.

Hydraulic roughness (Manning's n) coefficients are varied within literature-supported ranges for dominant landcover classes (e.g., ±20%–40% around nominal calibrated values), enabling the surrogate to learn sensitivity to surface resistance and channel friction. Small-magnitude elevation perturbations (e.g., decimeter-scale adjustments) are introduced to reflect digital elevation model uncertainty and to encourage robustness to minor topographic variation. Parameter perturbations are applied independently in most simulations, with a subset of runs including correlated high-intensity rainfall and inflow scenarios to better represent compound extremes.

At approximately 100 ft (~30 m) raster resolution, a single domain realization produces on the order of 10^5 spatial degrees of freedom per hourly time step. Using the containerized HEC-RAS workflow deployed on HPC infrastructure, we conservatively estimate that on the order of 10^3 simulations can be generated within practical computational budgets. Each simulation produces multi-day hourly time series of depth and discharge fields, yielding

millions of spatiotemporal training samples after conservative remapping to the raster grid (for the FNO-CNN pathway) or direct extraction of mesh-native outputs (for FluxGraphNet).

This synthetic design-storm framework enables controlled coverage of the governing parameter space while maintaining physical consistency through the underlying hydrodynamic solver. The resulting dataset is intended to approximate the functional mapping from static terrain features and time-varying forcing to depth-integrated discharge fields across a broad range of plausible flood conditions, supporting generalization beyond any single observed event.

Architecture Pathway 1: FluxGraphNet

FluxGraphNet is a graph neural network architecture for flood modeling that operates natively on unstructured HEC-RAS meshes, extending NVIDIA's HydroGraphNet (Taghizadeh et al. 2025). The architecture uses the standard MeshGraphNet-style encoder-processor-decoder (Pfaff et al. 2020). The novel extension is that FluxGraphNet predicts both node-level water depth and edge-level normal depth-integrated discharge (q_n), the depth-integrated discharge perpendicular to each mesh face. This design is physically motivated: in HEC-RAS's finite-volume formulation, face-normal depth-integrated discharge is the natural inter-cell quantity that lives on mesh faces. By treating mesh cells as graph nodes and mesh faces as graph edges, FluxGraphNet preserves the structure of the underlying finite-volume formulation. The architecture modification adds q_n to the edge encoder inputs and introduces a second decoder head on edge embeddings to predict Δq_n. A physics-informed loss term encourages mass and momentum conservation across the graph.

Architecture Pathway 2: Hybrid FNO-CNN on Raster Grids

HEC-RAS solves the two-dimensional shallow-water equations on an unstructured finite-volume mesh. For the purpose of conditioning a neural operator, a rectangular grid is preferred, as it is a natural representation as tensors for deep learning frameworks. We thus resample the mesh into a grid. It is important that the resampling scheme be conservative of mass and flux, and thus we cannot use a simple interpolation or nearest-neighbor resampling scheme. We plan to explore various grid resolutions and their impact on performance and stability.

To address this, we use a weighted-average first-order conservative remap of the mesh onto a rectangular grid using a weighted average of the mesh cell values that intersect with a given grid cell inspired by (P. W. Jones 1999).

Let Ω denote the spatial domain, $\{\mathcal{K}_k\}$ the set of mesh cells, and $\{C_i\}$ the set of raster cells.

Depth (Mass) Resampling Let h_k denote the cell-averaged water depth (m) in mesh cell $\mathcal{K}_k$, and let $|\mathcal{K}_k|$ denote its area. We assume a first-order (piecewise constant) representation within each mesh cell:

$$h(x, y) \approx h_k \quad \text{for } (x, y) \in \mathcal{K}_k.$$

For each raster cell C_i with area $|C_i|$, we define its depth as the area-weighted average of intersecting mesh cells:

$$h_i = \frac{1}{|C_i|} \sum_k h_k \, |\mathcal{K}_k \cap C_i|.$$

Here $|\mathcal{K}_k \cap C_i|$ denotes the area of intersection between mesh cell $\mathcal{K}_k$ and raster cell C_i.

This scheme preserves total water volume:

$$\sum_i h_i |C_i| = \sum_k h_k |\mathcal{K}_k|,$$

up to geometric discretization error. Since h is an intensive variable (m), area-weighted averaging ensures conservation of the extensive quantity $h\,dA$ (volume).

Face Flux Resampling Let $\mathbf{q}_k = (q_{x,k}, q_{y,k})$ denote the depth-integrated discharge (m^2/s) in mesh cell $\mathcal{K}_k$, where

$$\mathbf{q}_k = h_k \mathbf{u}_k,$$

and $\mathbf{u}_k$ is the depth-averaged velocity (m/s). We again assume a first-order representation:

$$\mathbf{q}(x, y) \approx \mathbf{q}_k \quad \text{for } (x, y) \in \mathcal{K}_k.$$

Let f denote a raster cell face (a line segment), with unit outward normal $\mathbf{n}_f$. The physically consistent volumetric flux (m^3/s) across f is

$$F_f = \int_f \mathbf{q} \cdot \mathbf{n}_f \, ds.$$

Under the piecewise-constant assumption, this becomes

$$F_f = \sum_k (\mathbf{q}_k \cdot \mathbf{n}_f) \, \ell_{kf},$$

where ℓ_{kf} is the length (m) of raster face f within mesh cell $\mathcal{K}_k$:

$$\ell_{kf} = \text{length}\,(f \cap \mathcal{K}_k)\,.$$

Dimensional consistency follows from

$$(m^2/s) \times (m) = m^3/s.$$

This construction yields a discrete flux field on the raster grid that is consistent with the finite-volume form of the shallow-water mass balance:

$$\frac{d}{dt} \int_{C_i} h \, dA = - \int_{\partial C_i} \mathbf{q} \cdot \mathbf{n} \, ds + \int_{C_i} R \, dA,$$

where R denotes rainfall input.

By reconstructing face-normal depth-integrated discharges directly rather than averaging cell-centered discharge vectors, we preserve the divergence structure required for discrete mass conservation. The geometric weights $|\mathcal{K}_k \cap C_i|$ and ℓ_{kf} depend only on mesh and raster geometry and are therefore precomputed and stored as sparse operators for each domain.

Compressed Operator Learning for 2D Shallow-Water Dynamics

Our surrogate approximates the hourly time-evolution operator of a 2D shallow-water system. Let G denote static spatial inputs (e.g., DEM, derived slope features, roughness, landcover, and boundary masks), let $F(t)$ denote time-varying forcings (rainfall and boundary conditions), and let the hydrodynamic state on the raster be $S(t) = \{h(t), q_x(t), q_y(t)\}$. We adopt an autoregressive residual formulation,

$$S(t+1) = S(t) + \Delta S_\theta \big(S(t), G, F(t)\big),$$

which empirically improves stability over multi-step rollout compared to direct state prediction.

Spatial Encoder and Latent Compression

The raw raster inputs may range from 500×500 to 1000×1000 cells at approximately 100 ft resolution. To make operator learning computationally tractable, we first compress the spatial inputs into a lower-resolution latent representation. The encoder maps the multi-channel raster inputs into a latent tensor

$$Z(t) \in \mathbb{R}^{H' \times W' \times C},$$

where $H', W' < H, W$ and C is the latent channel dimension.

Rather than exploring a wide range of backbone families, we focus on a convolutional encoder-decoder architecture with downsampling (e.g., U-Net–style or ConvNeXt-style blocks). Convolutional encoders provide strong locality bias and translation equivariance, which align with the predominantly local transport structure of shallow-water dynamics. Downsampling reduces the spatial resolution by a fixed factor (e.g., 2×–4×), allowing the operator core to act on a compressed spatial grid while retaining large-scale hydraulic structure.

All static fields, dynamic state variables, rainfall inputs, and boundary indicator masks are concatenated as input channels. The encoder jointly embeds heterogeneous continuous and categorical inputs into a shared latent space suitable for operator learning.

Hybrid Neural Operator Core in Latent Space

The operator core O_θ acts on the compressed latent representation $Z(t)$. By operating in latent space rather than on the full-resolution raster, we significantly reduce the memory and computational cost of global spectral mixing.

We model the time-evolution operator using a hybrid design combining global spectral convolution (Fourier Neural Operator, FNO (Li et al. 2021)) with local convolutional residual blocks. Conceptually, this architecture approximates the nonlinear evolution operator of the shallow-water PDE over one time step.

Fourier Neural Operator Block. Given $Z \in \mathbb{R}^{H' \times W' \times C}$, an FNO layer computes

$$\mathrm{FNO}(Z) = \mathcal{F}^{-1}\Big(\hat{K}_\theta \odot \mathcal{F}(Z)\Big),$$

where $\mathcal{F}$ denotes the 2D Fourier transform applied channel-wise, $\hat{K}_\theta$ is a learned complex-valued spectral filter truncated to a fixed number of modes, and $\odot$ denotes element-wise multiplication in Fourier space. Truncation of high-frequency modes controls computational cost and regularizes the operator.

Local Convolutional Residual Block. To improve representation of localized nonlinearities (e.g., wetting/drying fronts, sharp gradients, channelized flow), we incorporate local convolutional updates:

$$\mathrm{Conv}(Z) = \phi\big(\mathrm{Conv}_{3\times3}(\phi(\mathrm{Conv}_{3\times3}(Z)))\big),$$

where ϕ denotes a nonlinearity (e.g., GELU). We use a two-layer Conv–nonlinearity–Conv block inspired by VGG (Simonyan and Zisserman 2015) to introduce local nonlinearity and expand the effective receptive field while maintaining a stable, compact residual update suitable for autoregressive shallow-water dynamics. These blocks introduce strong local inductive bias and mitigate spectral smoothing effects.

Hybrid Residual Update. Each hybrid layer combines global and local components:

$$Z^{(\ell+1)} = Z^{(\ell)} + \alpha\,\mathrm{FNO}\big(Z^{(\ell)}\big) + \beta\,\mathrm{Conv}\big(Z^{(\ell)}\big),$$

for $\ell = 1, \ldots, L$, where α and β are learnable scalars. Stacking L such layers yields

$$Z'(t) = O_\theta\big(Z(t)\big).$$

Operating in compressed latent space reduces memory requirements and alleviates the computational burden associated with spectral transforms on large rasters, while preserving basin-scale coupling through global mixing.

Boundary and Non-Periodicity Considerations

Standard FNO layers implicitly assume periodic boundary conditions, which can induce wrap-around artifacts in non-periodic hydrodynamic domains. To mitigate this, we: 1) Encode boundary indicator masks and boundary forcing values explicitly as raster channels prior to encoding; 2) Use hybrid local convolutional blocks to reinforce physically local transport behavior; 3) Evaluate non-periodic padding strategies (e.g., zero or reflective padding) in the encoder to reduce edge artifacts.

These design choices reduce unphysical cross-domain interactions while retaining the benefits of spectral mixing.

Decoder and Output Parameterization

A convolutional decoder upsamples the latent output $Z'(t)$ back to raster resolution and predicts the residual update

$$\Delta S(t) = \{\Delta h(t), \Delta q_x(t), \Delta q_y(t)\}.$$

Depth non-negativity is enforced through projection (e.g., clamping or softplus-based parameterization) to prevent unphysical negative water depths. Wet/dry masking is applied during training to avoid over-penalizing momentum errors in near-dry regions.

Training for Stability and Physical Plausibility

To improve multi-day stability, we train using multi-step rollout loss over short horizons rather than single-step supervision alone. In addition to state RMSE for $\{h, q_x, q_y\}$, we include auxiliary terms: 1) A penalty on negative depth values, 2) A global mass-balance consistency loss comparing domain-integrated depth change to reconstructed boundary fluxes and rainfall input, 3) Optional weighting of momentum errors by depth to reduce noise amplification in shallow cells.

These constraints are intended to improve physical plausibility and reduce error accumulation during autoregressive rollout, while maintaining computational efficiency suitable for near-real-time inference.

DISCUSSION

The outputs of ML-HYDRAS are designed to support future integration into operational decision-support systems for flood disasters. By providing both water depth and depth-integrated discharge (from which velocity can be derived), the framework produces physically interpretable fields that extend beyond inundation alone. These outputs can inform hazard characterization, infrastructure risk assessment, and impact modeling across preparedness, response, and recovery phases.

The dual-pathway effort enables comparative evaluation of two fundamentally different approaches to surrogate modeling, each presenting distinct trade-offs between physical consistency and generalizability. FluxGraphNet operates directly on the native unstructured mesh, preserving the finite-volume structure of HEC-RAS and avoiding remapping errors, while the FNO-CNN pathway leverages the computational efficiency and established training infrastructure of raster-based neural operators. Both approaches predict physically consistent depth and discharge fields, but differ fundamentally in their treatment of spatial discretization and boundary conditions.

The FNO-CNN approach prioritizes generalizability through its raster representation, which provides domain independence. Once trained, the model operates on a standardized spatial data structure that can represent any geographic region at the target resolution, enabling straightforward deployment to new regions without significant modification. However, this flexibility introduces remapping errors during translation from the unstructured mesh to a regular grid. Conversely, FluxGraphNet eliminates translation errors by preserving edge-based flux formulation and maintaining alignment with the underlying mesh. However, this native representation likely will tie the graph structure to specific domain geometry, where each watershed's unique mesh structure may complicate transfer learning. By developing both approaches, we aim to empirically assess these trade-offs and ultimately, inform deployment decisions based on operational requirements.

Within the United States, one potential pathway for operational relevance is compatibility with the National Weather Service (NWS) water resources decision-support ecosystem (e.g., water.noaa.gov), which supports the agency's Weather-Ready Nation initiative (National Weather Service 2011; Frolov et al. 2025). Current NWS flood mapping products primarily emphasize inundation extent (National Weather Service 2025). A surrogate model capable of rapidly producing both depth and discharge fields could complement these products by enabling momentum-informed hazard indicators, such as high-velocity corridors or depth–velocity thresholds associated with vehicle instability and structural damage. We emphasize that integration would require formal evaluation, validation, and alignment with agency workflows; here we identify compatibility and potential utility rather than deployment.

Similarly, ML-HYDRAS outputs are structurally compatible with downstream impact models such as FEMA's Hazus framework, which currently estimates losses using depth–damage relationships (Scawthorn et al. 2006). Incorporating velocity-informed metrics may enable refinement of damage curves and derivation of secondary products, including debris-transport potential or infrastructure failure thresholds. These extensions would require empirical validation but illustrate how depth-integrated discharge fields expand the decision-relevant feature space beyond static inundation maps.

A key advantage of a neural surrogate relative to a full hydrodynamic solver is computational efficiency. Once trained, inference can be performed orders of magnitude faster than in physics-based simulations, enabling rapid scenario evaluation and ensemble forecasting. This efficiency directly supports uncertainty quantification through Monte Carlo methods. For example, uncertainty in rainfall intensity, boundary inflow, roughness coefficients, or topographic perturbations can be propagated through the surrogate by sampling across plausible parameter distributions. Producing tens to hundreds of realizations becomes computationally feasible, allowing estimation of probabilistic inundation maps, velocity exceedance probabilities, and confidence intervals on derived risk metrics. Such probabilistic outputs are often more actionable for decision-makers than single deterministic forecasts.

We note several limitations of the current phase of work. First, the training dataset is generated from a limited number of observed and synthetic storms and a physics-based solver within two specific watersheds, which may limit generalization across hydroclimatic regimes. Secondly, active urban drainage systems and controlled infrastructure are not yet modeled. These constraints define the current scope and motivate future extensions toward multi-watershed training, and integration with additional hydrodynamic components.

The operational value of ML-HYDRAS lies in its ability to resolve the disconnect between decision-maker requirements and the capabilities of current tools within actionable timeframes. Emergency managers preparing for flooding events need to evaluate dozens of scenarios to prepare for diverse potential outcomes, a task requiring ensemble forecasts that physics-based models cannot deliver at scale. During active flooding, responders could use hourly-updated velocity maps to assess risks to critical infrastructure that may cause cascading failures that static inundation maps cannot provide. In recovery, damage assessment teams need spatially-explicit estimates of peak flow forces to validate insurance claims and prioritize reconstruction investments.

By producing momentum-conserving predictions at inference speeds measured in minutes rather than hours, ML-HYDRAS enables these workflows to transition from aspirational to operational. An emergency manager can evaluate 100 synthetic storm scenarios overnight to identify the 15 road segments most likely to become impassable, informing targeted upgrades. A responder can ingest real-time radar during an active event to produce 6-hour velocity forecasts, triggering pre-emptive bridge closures before critical thresholds are reached. A damage assessment team can retrospectively simulate an observed event to generate building-level hydrodynamic load estimates to better understand the impacts of the event.

These applications share a common requirement: physically-grounded predictions of both inundation extent and flow momentum, delivered at speeds compatible with operational decision cycles. Ultimately, the measure of success for ML-HYDRAS is not predictive accuracy alone, but whether its outputs enable more informed decisions under the time and uncertainty constraints of disasters. By providing rapid access to velocity fields that capture the physical mechanisms of flood damage, the framework aims to support more resilient communities through improved preparedness, more effective response, and more informed recovery.

GENERATIVE AI DISCLOSURE

This manuscript was prepared with the assistance of generative artificial intelligence (GenAI) tools (ChatGPT). These tools were used to support drafting, editing for clarity, and formatting in LaTeX. All technical content, methodological decisions, mathematical formulations, and scientific claims were developed and verified by the authors, who take full responsibility for the accuracy and integrity of the work.

REFERENCES

Bonev, B., Kurth, T., Mahesh, A., Bisson, M., Kossaifi, J., Kashinath, K., Anandkumar, A., Collins, W. D., Pritchard, M. S., and Keller, A. (2025). "Fourcastnet 3: A geometric approach to probabilistic machine-learning weather forecasting at scale". In: *arXiv preprint arXiv:2507.12144*.

Brunner, G. (2020). *HEC-RAS river analysis system. Hydraulic reference manual: Version 6.0.* United States Army Corp of Engineers.

Brunner, G., Sanchez, A., Molls, T., and Parr, D. (2020). *HEC-RAS Verification and Validation Tests*. Tech. rep.

CTCN (2017). *Climate Change Adaptation Technologies for Water: A Practitioner's Guide to Adaptation Technologies for Increased Water Sector Resilience*. URL: https://www.ctc-n.org/sites/default/files/resources/water_adaptation_technologies_0.pdf.

Dimitriadis, P., Tegos, A., Oikonomou, A., Pagana, V., Koukouvinos, A., Mamassis, N., Koutsoyiannis, D., and Efstratiadis, A. (2016). "Comparative evaluation of 1D and quasi-2D hydraulic models based on benchmark and real-world applications for uncertainty assessment in flood mapping". In: *Journal of Hydrology* 534, pp. 478–492.

Feng, D., Tan, Z., Lin, Z., Xu, D., Yu, C.-W., and He, Q. (2025). "A comparative study of physics-informed and data-driven neural networks for compound flood simulation at river-ocean interfaces: A case study of Hurricane Irene". In: *Journal of Geophysical Research: Machine Learning and Computation* 2.4, e2025JH000758.

Frolov, S., Garrett, K., Jankov, I., Kleist, D., Stewart, J. Q., and Ten Hoeve, J. (2025). "Integration of Emerging Data-Driven Models into the NOAA Research-to-Operations Pipeline for Numerical Weather Prediction". In: *Bulletin of the American Meteorological Society* 106.2, E430–E437.

Hammond, M., Chen, A., Djordjević, S., Butler, D., and Mark, O. (2015). "Urban flood impact assessment: A state-of-the-art review". In: *Urban Water Journal* 12.1. Ed. by J. Leandro, pp. 14–29.

Horritt, M. and Bates, P. (2002). "Evaluation of 1D and 2D numerical models for predicting river flood inundation". In: *Journal of Hydrology* 268.1, pp. 87–99.

Joint Research Center (2020). *LISFLOOD - a distributed hydrological rainfall-runoff model: Model Documentation*. European Commission.

Jones, P. W. (1999). "First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates". In: *Monthly Weather Review* 127.9, pp. 2204–2210.

Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., and Anandkumar, A. (2023). "Neural operator: Learning maps between function spaces with applications to pdes". In: *Journal of Machine Learning Research* 24.89, pp. 1–97.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. (2021). *Fourier Neural Operator for Parametric Partial Differential Equations*. arXiv: 2010.08895 [cs.LG].

Marvi, M. (2020). "A review of flood damage analysis for a building structure and contents". In: *Natural Hazards* 102, pp. 967–995.

McGlade, J., Bankoff, G., Abrahams, J., Cooper-Knock, S., Cotecchia, F., Desanker, P., Erian, W., Gencer, E., Gibson, L., Girgin, S., et al. (2019). *Global assessment report on disaster risk reduction 2019*. UN Office for Disaster Risk Reduction.

National Weather Service (2011). *NOAA'S NATIONAL WEATHER SERVICE STRATEGIC PLAN: Building a Weather-Ready Nation*. URL: https://www.weather.gov/media/wrn/strategic_plan.pdf.

National Weather Service (2025). *NOAA's transformative Flood Inundation Mapping expands to 60% of U.S.* URL: https://www.noaa.gov/news-release/noaas-transformative-flood-inundation-mapping-expands-to-60-of-us.

Pfaff, T., Fortunato, M., Sanchez-Gonzalez, A., and Battaglia, P. (2020). "Learning mesh-based simulation with graph networks". In: *International Conference on Learning Representations*.

Redondo-Tilano, S. A., Boucher, M.-A., and Lacey, J. (2025). "Emerging strategies for addressing flood-damage modeling issues: A review". In: *International Journal of Disaster Risk Reduction* 116, p. 105058.

Reuther, A., Kepner, J., Byun, C., Samsi, S., Arcand, W., Bestor, D., Bergeron, B., Gadepally, V., Houle, M., Hubbell, M., et al. (2018). "Interactive supercomputing on 40,000 cores for machine learning and data analysis". In: *2018 IEEE High Performance extreme Computing Conference (HPEC)*. IEEE, pp. 1–6.

Romali, N., Sulong, S., and Kawasaki, A. (2025). "A Systematic Review of Flood Damage Assessment: Insight for the Data-Scarce Regions". In: *Water Resources Management* 39, pp. 4707–4734.

Scawthorn, C., Flores, P., Blais, N., Seligson, H., Tate, E., Chang, S., Mifflin, E., Thomas, W., Murphy, J., Jones, C., et al. (2006). "HAZUS-MH flood loss estimation methodology. II. Damage and loss assessment". In: *Natural Hazards Review* 7.2, pp. 72–81.

Scheip, C., Lang, K., Prince, P., Wegmann, K., Wooten, R., Reynolds, A., and Harris, D.-A. (2026). "A storm-sculpted landscape—Observations from post-Helene lidar in the Hickory Nut Gorge, North Carolina". In: *Earth Surface Processes and Landforms* 51.1, e70228. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/esp.70228.

Simonyan, K. and Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv: 1409.1556 [cs.CV].

Taghizadeh, M., Zandsalimi, Z., Nabian, M. A., Shafiee-Jood, M., and Alemazkoor, N. (2025). "Interpretable physics-informed graph neural networks for flood forecasting". In: *Computer-Aided Civil and Infrastructure Engineering* 40.18, pp. 2629–2649.

U.S. Geological Survey (2019). *The National Map - New data delivery homepage, advanced viewer, lidar visualization: U.S. Geological Survey Fact Sheet 2019-3032*.

U.S. Geological Survey (2024). *Annual NLCD Collection 1 Science Products (ver. 1.1, June 2025): U.S. Geological Survey data release*.

United States Geological Survey (2026). *Water Data for the Nation*. URL: https://waterdata.usgs.gov/.

Walker-Franklin, I., Blake, S., Thorp, E., and Tuberty, S. (2025). "Preliminary Identification of PFAS and Other Emerging Contaminants in the French Broad River, NC Post-Hurricane Helene". In: *Toxics* 13.11.

Whittle, P. and Casey, M. (2025). "Why does Vermont keep flooding? It's complicated, but experts warn it could become the norm". In: *AP News*.

Yin, Z., Shi, J., Bian, L., Campbell, W. H., Zanje, S. R., Hu, B., and Leon, A. S. (2025). "Physics-informed neural network approach for solving the one-dimensional unsteady shallow-water equations in riverine systems". In: *Journal of Hydraulic Engineering* 151.1, p. 04024060.