Received: 1 October 2024 | Accepted: 7 April 2025

DOI: 10.1111/mice.13490

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING WILEY

RESEARCH ARTICLE

High-resolution flood probability mapping using generative machine learning with large-scale synthetic precipitation and inundation data

Lipai Huang[1] | **Federico Antolini**[2] | **Ali Mostafavi**[1,2] | **Russell Blessing**[2] | **Matthew Garcia**[3] | **Samuel D. Brody**[2]

[1]Urban Resilience.AI Lab, Zachry Department of Civil and Environmental Engineering, Texas A&M University, Texas, USA

[2]Institute for a Disaster Resilient Texas, Texas A&M University, Texas, USA

[3]Civil and Environmental Engineering, Rice University, Houston, Texas, USA

Correspondence
Lipai Huang, Urban Resilience.AI Lab, Zachry Department of Civil and Environmental Engineering, Texas A&M University, TX, USA.
Email: lipai.huang@tamu.edu

Funding information
National Science Foundation, Grant/Award Number: 1832662; Texas A&M X-Grant Presidential Excellence Fund

Abstract

High-resolution flood probability maps are instrumental for assessing flood risk but are often limited by the availability of historical data. Additionally, producing simulated data needed for creating probabilistic flood maps using physics-based models involves significant computation and time effort, which inhibit its feasibility. To address this gap, this study introduces Precipitation-Flood Depth Generative Pipeline, a novel methodology that leverages generative machine learning to generate large-scale synthetic inundation data to produce probabilistic flood maps. With a focus on Harris County, Texas, Precipitation-Flood Depth Generative Pipeline begins with training a cell-wise depth estimator using a number of precipitation-flood events model with a physics-based model. This cell-wise depth estimator, which emphasizes precipitation-based features, outperforms universal models. Subsequently, the conditional generative adversarial network (CTGAN) is used to conditionally generate synthetic precipitation point cloud, which are filtered using strategic thresholds to align with realistic precipitation patterns. Hence, a precipitation feature pool is constructed for each cell, enabling strategic sampling and the generation of synthetic precipitation events. After generating 10,000 synthetic events, flood probability maps are created for various inundation depths. Validation using similarity and correlation metrics confirms the accuracy of the synthetic depth distributions. The Precipitation-Flood Depth Generative Pipeline provides a scalable solution to generate synthetic flood depth data needed for high-resolution flood probability maps, which can enhance flood mitigation planning.

1 | INTRODUCTION

Flood hazards pose a significant threat to cities and communities globally, resulting in extensive physical damage and substantial economic costs due to material losses and human casualties, particularly in densely populated regions. In the United States, flood-related insurance claims have consistently averaged around $40 billion annually over the past four decades (Yildirim et al., 2022). With the projected increase in the frequency and severity of flood events across various US regions (Musselman et al., 2018), this financial burden is expected not only to persist but also to escalate. An accurate assessment of flood risk is crucial to plan for efficient response and recovery of potentially affected communities. Flood risk characterization traditionally involves the creation of floodplain maps, which delineate areas typically prone to flooding based on historical data and hydrologic simulations. However, floodplain maps often have limitations due to their reliance on limited historical data, which may not accurately reflect current and future conditions. Also, development of floodplains using physics-based hydraulic and hydrologic (H&H) models is computationally expensive. In addition, these maps usually depict flood extent for a few specific flooding frequency, limiting their utility in risk assessments and communication. The gaps in traditional floodplain mapping underscore the need to incorporate probability into flood maps beyond the traditional 100-year and 500-year frequencies. In the context of this study, probabilistic flood maps quantify flood risk from a distribution of rainfall events, without explicitly considering the frequency of each of those events. That is, probabilistic flood maps estimate the likelihood that a rainfall event causes a certain inundation in a given place (e.g., 1 ft) from a large number of rainfall events, without considering the probability of occurrence of each event. For example, if 350 out of 1000 rainfall events cause at least 1-ft inundation in a certain place, a probabilistic flood map will associate a likelihood of 0.35 to the 1-ft inundation of that place for rainfall events with the range of characteristics considered. This quantified likelihood provides a measure of 1-ft inundation risk in the area given the characteristics of rainfall events. In this context, this probability is different from the annual likelihood typically used, which is based on a statistical analysis of historical storms and the extrapolation of flood depth and/or extent values for a few return periods (e.g., 10-, 100-, 500-year storm). By offering a spatially continuous probabilistic measure of flood risks, maps provide quantitative insights that can help policymakers, planners, and communities in the mitigation and preparedness phases with tailored, data-driven strategies that lessen the impacts of potential flooding (Nofal et al., 2024).

The creation of flood probability maps hinges on simulating a large number of flood inundation scenarios based on which the likelihood of a range of flood depths across a region can be estimated. In fact, one of the major barriers to creation of flood probability maps is the lack of numerous flood scenarios. On one hand, the number of historical flood events in a given region is not large enough to build a robust statistical distribution of rainfall events and their associated flood depths. Also, depending on the rate of land use change, historical data may not be adequate to estimate flood risk under current conditions. On the other hand, generating a large number of simulated flood scenarios using H&H models is computationally expensive and would not be feasible. An alternative to historical data is represented by synthetic high-resolution flood scenarios generated using generative models.

The use of generative models and machine learning has gained significant attention in flood risk analysis. Recent studies have adopted generative adversarial networks (GANs) to generate synthetic precipitation data integrated with hydrologic models such as soil and water assessment tool (SWAT) for flood frequency analysis (Ji et al., 2024), and to support real-time urban flood forecasting through event-based modeling (Piadeh et al., 2023). Deep learning architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been applied for flash flood prediction and integration with hydrodynamic models for flood mapping and risk assessment (Naeini & Snaiki, 2024). Hybrid models such as Fast Gated recurrent neural networks (FastGRNN) combined with fully convolutional networks (FCN) have shown strong performance in predicting flood status from multivariate sensor networks (Dong et al., 2021), and ensemble machine learning methods like Random Forest and Extreme Gradient Boosting (XGBoost) have been widely used to generate flood susceptibility maps from spatial conditioning factors (Lyu & Yin, 2023; Zhu, Lai, et al., 2024). Other studies have leveraged long short-term memory (LSTM) networks and conditional GANs to improve flood prediction accuracy in urban environments (Do Lago et al., 2023; McSpadden et al., 2024), while Ho et al. (2025) applied a vision-language segmentation model to estimate building lowest floor elevation (LFE) from street imagery for enhanced flood exposure mapping. System-level modeling efforts have included Bayesian network frameworks to simulate cascading failures in flood control infrastructure (Dong et al., 2020), stochastic programming for resilient planning under extreme weather scenarios (Zhang & Alipour, 2023), and structural monitoring of flood-critical assets such as dams through intelligent segmentation and multisource machine learning approaches (Yao et al., 2024; Zhu, Niu, et al., 2024). Generative methods have also been explored for infrastructure

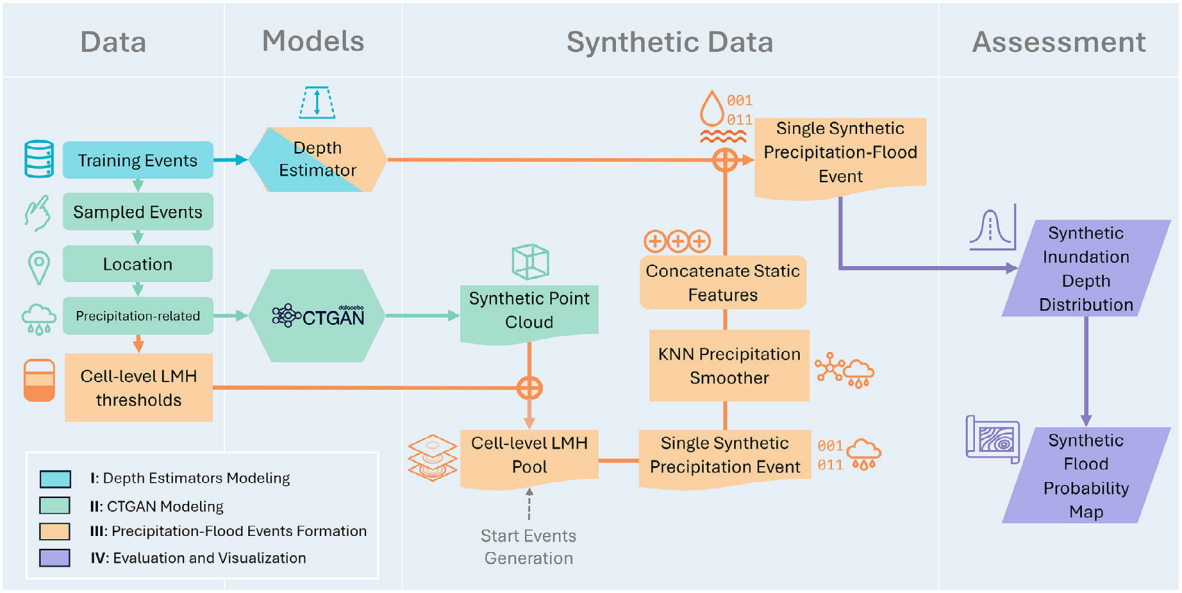

FIGURE 1 The workflow consists of four main steps. Step I involves developing and selecting the optimal depth estimator among universal and cell-wise models based on training precipitation-flood events. Step II focuses on event sampling, preprocessing point-level features, and training a constraint-aware conditional generative adversarial network (CTGAN) to generate a synthetic point cloud through conditional sampling. In Step III, a precipitation feature-level pool is created for each cell mesh using thresholds derived from sampled training event features, enabling strategic sampling to generate synthetic precipitation event features. The trained depth estimator from Step I then processes K-nearest neighbor (KNN)-smoothed synthetic precipitation features along with static features to predict synthetic flood depths. Finally, Step IV iterates Step III to generate thousands of synthetic precipitation events, forming a synthetic depth distribution and ultimately constructing a synthetic flood probability map.

monitoring, including conditional GAN-based models for detecting bridge damage via time-frequency decomposition (Lee et al., 2025). Additionally, large-scale machine learning systems such as the telescopic broad Bayesian learning framework have been introduced for streaming data adaptation in civil infrastructure contexts (Yuen & Kuok, 2025). Integrated models that assess transportation resilience under flood disruptions have further expanded machine learning's role in flood-aware infrastructure planning (Yin et al., 2023). Despite these advancements, few existing approaches focus on generating synthetic flood data tailored for probabilistic flood mapping.

To address these challenges, we propose the Precipitation-Flood Depth Generative Pipeline, a novel methodology that harnesses generative machine learning to synthesize large-scale precipitation data and predict corresponding flood inundation depths through strategic sampling and modeling. The main workflow is illustrated in Figure 1. Precipitation-Flood Depth Generative Pipeline starts with the training of a depth estimator using a number of H&H model-generated flood events, incorporating precipitation-based, spatial, and region-specific features. Following this, a Conditional GAN (CTGAN) with constraints is employed to generate synthetic rainfall precipitation events. Strategic thresholds are established to filter these synthetic records, ensuring close alignment with true precipitation patterns. For each location, synthetic events are smoothed using a K-nearest neighbors (KNN) algorithm and processed through the trained depth estimator to derive synthetic depth distributions. By iterating this procedure across multiple synthetic rainfall events, we construct flood probability maps for different combinations of rainfall characteristics (duration, peak precipitation intensity, and cumulative precipitation). These maps are validated using similarity and correlation metrics to confirm the correspondence of synthetic depth distributions to training data.

The novel contributions of the model presented in this paper are threefold:

(1) We introduce a novel methodology that utilizes a surrogate machine learning pipeline to estimate flood probabilities using synthetic precipitation-flood events generated by CTGAN and depth estimator.
(2) The cell-wise machine learning modeling, introduced in Section 2.3, presents an innovative approach to feature engineering, yielding enhanced prediction performance compared to global models.
(3) We develop an all-to-one event sampling algorithm designed to strategically improve the quality of synthetic records while preserving nonlinearity, ensuring a more realistic simulation of flood events.

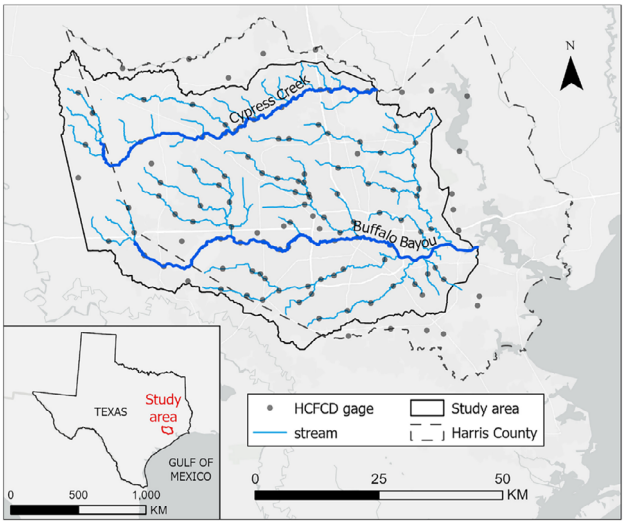

FIGURE 2 Study area and Harris County Flood Control District gauge distribution. Map generated using ArcGIS Pro 3.0.0 (https://pro.arcgis.com/).

Precipitation-Flood Depth Generative Pipeline provides a scalable solution for generating high-resolution flood probability maps, to support to flood preparedness and mitigation efforts. The following section explains the components of the Precipitation-Flood Depth Generative Pipeline and the associated data sets.

2 | MATERIAL AND METHODS

2.1 | Study region and data

2.1.1 | Study region

As shown in Figure 2, Harris County, Texas, serves as the test bed for developing the Precipitation-Flood Depth Generative Pipeline. The primary watersheds in this area are Cypress Creek and Buffalo Bayou. Harris County, located at the heart of the Greater Houston Metropolitan Statistical Area, covers 1778 square miles (4605 square kilometers) and has seen its population grow to over 4.5 million in the past decade. The county's flat topography, with elevations ranging from −40 ft (−12.19 m) to 300 ft (91.44 m), coupled with its largely developed land (over two-thirds), along with 20% pasture and cultivated lands (Dewitz, 2024), contributes to significant flood risk. The Cypress Creek watershed in the northern part and the Buffalo Bayou system in the central and southern areas flow into the San Jacinto River and the Ship Channel, eventually draining into the Gulf of Mexico. Dense urban development, poor natural drainage, limited soil infiltration, and the county's subtropical climate further exacerbate its vulnerability to chronic flooding.

2.1.2 | Data

In creating Precipitation-Flood Depth Generative Pipeline, there is a need for an initial set of rainfall events and related flood inundation scenarios. The lack of sufficient historical precipitation flood inundation data was addressed using 592 precipitation events processed with a physical-based model (Garcia, 2023) to obtain corresponding flood scenarios. Specifically, Garcia (2023) considered storm events occurred in Harris County in the period 2014–2022, and applied Rasterized Time-series Resampling Method (RTRM) to build 592 synthetic, realistic storms varying by spatial extent, duration, and rainfall intensity. Garcia et al. (2023) modeled those events in HEC-RAS 2D and produced as many flood scenarios. The study region in HEC-RAS 2D was modeled using a 1200 ft × 1200 ft mesh grid, with additional refinements applied near major rivers and tributaries to enhance both computational accuracy and stability. This final mesh consists of 26,301 individual cells. Manning's roughness coefficient and imperviousness values were sourced from the 2019 National Land Cover Data set (Dewitz, 2024), while soil infiltration values were obtained from the Gridded Soil Survey Geographic database (Soil Survey Staff, 2023). Model validation involved analyzing data from Harris County Flood Control District gauges (Harris County Flood Control District, 2023) during five significant flood events from 2016 to 2020, including the 2016 Tax Day flood and Hurricane Harvey in 2017. This validation process, which calibrated the HEC-RAS 2D model for global minimum error across all gauges, confirmed the model's reliability and accuracy for simulating diverse flood scenarios (Garcia et al., 2023). We used the 592 rainfall events as the initial set from which to generate further synthetic rainfall events, as described in Section 2.4 and Section 2.5. The flood depth scenarios became the input for the depth estimator, a surrogate machine learning model for determining cell-wise maximum flood depth, described in Section 2.3.

2.2 | Preprocessing

We employed Voronoi tessellation (Konrad, 2021) to partition the training data set, assigning each point within the study area a unique polygonal region to ensure there is no overlap among cells. Additionally, we segmented the 592 rainfall events used in the HEC-RAS 2D flood scenarios into three classes by inundation depth level, to ensure consistent stratification for subsequent data set splitting and resampling. Specifically, for the training set, events are classified with mean depths of ≤ 2 inches, between > 2 and ≤ 6 inches, and > 6 inches in a ratio of 2:4:1, respectively. The validation set consists of 20% of the

TABLE 1 Features used in depth estimator.

Feature	Level	Unit	Description
Cumulative precipitation	All	Inch	Event total precipitation
Peak precipitation	All	Inch	Event peak precipitation
Duration	All	Hour	Effective precipitation duration
Channel	Universal	Binary (1 or 0)	Identification of channel cell
Terrain elevation	Universal	Feet	Geographical height of each cell
Nine heavy cumulative precipitation ratios (HCPRs; one for each watershed)	Cell wise	Scalar (range 0 to 1)	Proportion of watershed area experiencing heavy cumulative precipitation
Nine heavy peak precipitation ratios (HPPRs; one for each watershed)	Cell wise	Scalar (range 0 to 1)	Proportion of watershed area experiencing heavy peak precipitation

training set volume, with the remaining data designated as the test set. This approach optimizes the separation of events to ensure relatively balanced representation across the training, validation, and testing phases, facilitating more effective modeling.

2.3 | Depth estimator model

For the depth estimator component of Precipitation-Flood Depth Generative Pipeline, we chose a tree-based machine learning model trained on simulated scenarios from the HEC-RAS 2D model. Compared to traditional machine learning models, tree-based models have shown superior performance due to their ability to manage complex, nonlinear relationships and interactions within the data. XGBoost (Chen & Guestrin, 2016; Shehadeh et al., 2021), in particular, excels with its gradient boosting framework, which iteratively improves model accuracy by minimizing residual errors from previous iterations. Additionally, deep learning models in regression, such as the Transformer (Castangia et al., 2023; Vaswani et al., 2017), utilize attention mechanisms to capture long-range dependencies and complex feature interactions, providing substantial benefits in predictive performance. In this study, we employed the XGBoost regressor and the regression Transformer as universal depth estimators, training them on physics-based model-generated events that include both spatial features and precipitation-based features, as shown in Table 1. This integration aims to enhance the accuracy and robustness of maximum flood inundation estimations. Furthermore, we compared these universal models with a cell-wise model, MaxFloodCast V2 (Lee et al., 2024), an XGBoost-based depth estimator aggregating 26,301 machine learning models, one for each of the 26,301 cells of the mesh. In MaxFloodCast V2, each cell model is trained independently of the others, without considering precipitation or water levels in neighboring cells. However, spatial dependencies exist in the natural water flow. For example, the water level of cells located in a channel is influenced by the water draining from upstream areas. To account for this spatial influence while maintaining cell-wise model independence, we introduced a heavy cumulative precipitation ratio (HCPR) and a heavy peak precipitation ratio (HPPR). Heavy precipitation is the hourly peak intensity or cumulative rain volume above a certain threshold established for a study area. HCPR and HPPR are defined at the watershed level as in Equation (1) and Equation (2).

$$\mathrm{R}_i^c = \frac{\sum_j AC_{i,j} \times h_{i,j}^c}{AW_i}, c \in \{\mathfrak{c}, \mathfrak{p}\} \tag{1}$$

$$h_{i,j}^c = \begin{cases} 1, & \text{if } p_{i,j}^c > 2 \text{ in} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where c represents HCPR as $\mathfrak{c}$ and HPPR as $\mathfrak{p}$, with R denoting the respective heavy precipitation ratio. AW_i refers to the total area of watershed i, while $AC_{i,j}$ represents the area of cell j within watershed i. The binary variable $h_{i,j}$ identifies whether cell j in watershed i experiences heavy precipitation, and $p_{i,j}$ denotes the corresponding precipitation condition for cell j within the same watershed. HCPR and HPPR are defined at the watershed scale, and are uniform across all cells within the same watershed during a rainfall event. These ratios can differ considerably across watersheds and across events, and offer substantial explanatory potential as surrogates for precipitation patterns.

We assessed the predictive performance of the depth estimators (in Section 3.1) using root mean squared error (RMSE) and R^2 scores (Chicco et al., 2021) to evaluate both accuracy and model explainability. RMSE measures the average magnitude of the prediction errors, providing insight into the estimator's precision, while R^2 indicates the proportion of variance in the observed data that is predictable from the input features, reflecting the model's

explanatory power. In addition, we conducted a comprehensive analysis of the aggregated performance of these models on both channel and nonchannel cells. This distinction is crucial due to the varying hydrological and ground texture characteristics inherent to these environments, which influence the uncertainty and behavior of flood dynamics.

2.4 | Rainfall data augmentation via CTGAN

Given the limited physics-based model simulated precipitation-flood events, to generate relatively robust synthetic rainfall events in our study, we strategically sampled 315 events as explained in Section 2.2 and then employed CTGAN from Synthetic Data Vault (SDV) (Xu et al., 2019; Patki et al., 2016) for tabular data augmentation. CTGAN extends the GAN framework to handle structured data, particularly tabular data sets, making it highly effective for generating synthetic data in scenarios with small, imbalanced data sets. One of CTGAN's strengths is its ability to support the specification of constraints and conditions during the data generation process, enabling the preservation of specific relationships or attributes present in the original data set.

We converted the training cell-level data into point-level data using centroids to facilitate generation. Consequently, the input attributes for CTGAN included latitude, longitude, cumulative precipitation, peak precipitation, and precipitation duration. To enhance the quality of the synthetic records, we constrained conditional sampling to ensure that synthetic points satisfied the following conditions: (1) Sampled points are inside the study area (Figure 2); (2) for a sampled point, the cumulative precipitation is always greater than the peak precipitation value; (3) for a sampled point, the peak precipitation is greater than the average precipitation intensity, calculated by dividing the cumulative precipitation by the event duration.

To determine the optimal hyperparameters for the generator and discriminator learning rates, as well as the stoppage epochs, we conducted a comprehensive hyperparameter grid search to explore the optimal learning rates for the generator and discriminator, ensuring both networks improved steadily without overpowering each other. Additionally, selecting appropriate stoppage epochs enabled us to terminate the training process at the most accurate point, preventing overfitting or underfitting. The optimal checkpoint was selected based on the corresponding synthetic data set achieving the highest average marginal distribution, as evaluated using the Kolmogorov–Smirnov statistic (KS statistic) (Finner et al., 2018). The KS statistic K for a particular feature x_i between the training data set and the synthetic data set is defined as follows:

$$K_{i,n,m} = \sup_{x_i} |F_{t,n}(x_i) - F_{s,m}(x_i)| \tag{3}$$

where $F_{t,n}$ and $F_{s,m}$ represent the empirical distribution functions of the training data set and the synthetic data set generated by CTGAN, respectively. The function $\sup_{x_i}$ denotes the supremum across the domain of feature x_i. The overall quality score for all features, excluding location-based features, in the synthetic data set is computed using the following formula:

$$Score = \frac{\sum_{i=1}^{N}(1 - K_{i,n,m})}{N} \tag{4}$$

where N is the number of features. A higher score indicates a higher quality of the synthetic data, signifying a closer approximation to the training data distribution. The outcome of this step is a comprehensive synthetic point cloud containing 10,000,000 synthetic rainfall records generated by CTGAN, representing various precipitation durations. This diverse point cloud presents a significant challenge in processing and formulating individual precipitation events under varying conditions.

TABLE 2 Definition of low, medium, and high (LMH) thresholds for each precipitation-based feature within a cell mesh.

Class	Range
Low	$[0\,,\ \mu_i - \theta_1 \cdot \sigma_i]$
Medium	$(\mu_i - \theta_1 \cdot \sigma_i\,,\ \mu_i + \theta_2 \cdot \sigma_i]$
High	$(\mu_i + \theta_2 \cdot \sigma_i\,,\ +\infty)$

2.5 | Rainfall events generation

In the synthetic point data set, which comprises collections of precipitation records, we have predefined thresholds for cumulative precipitation, peak intensity, and duration, respectively—categorized as low, medium, and high (LMH)—for each cell mesh based on training data. These categories are established based on the conditions shown in Table 2, where μ_i is the mean of a precipitation-based feature x_i from training data set, and σ_i refers to its standard deviation. θ_1 and θ_2 are two positive constants. Consequently, for each cell mesh, we defined low (L), medium (M), and high (H) thresholds based on the distribution of precipitation-related features—cumulative precipitation, peak precipitation, and duration—derived from the training data set. These thresholds were then applied

to the synthetic point data set generated as described in Section 2.4. For each cell, we constructed a corresponding precipitation feature-level pool comprising all possible LMH combinations of cumulative precipitation, peak precipitation, and duration, resulting in a total of 27 combinations (i.e., LLL, LLM,…, HHH.). To ensure consistency and representativeness, we systematically iterated over all synthetic point records, retaining indices of synthetic points that fell within the cell boundary while preserving the original distribution of the pool.

Establishing the synthetic indices pool for each cell enabled to generate global precipitation events with various precipitation distributions across the study area. To generate an event, a specific duration was sampled randomly from the aggregated pool and set as the global rainfall duration, reducing the possible cumulative–peak–duration combinations from 27 to 9 (e.g., if the input synthetic duration is H, then the potential set of combination is LLH, LMH, LHH, MLH, MMH, MHH, HLH, HMH, HHH). Subsequently, a synthetic record was randomly sampled for each cell from its own pool. The sampling was repeated for all the cells of the mesh. This strategic sampling allowed to create single synthetic precipitation events from a large point cloud, transitioning from point-level to cell-level representation in accord with the original distribution of precipitation characteristics in the study area. To enhance the approximation of training precipitation events, we applied the KNN method as a smoother (Zhou et al., 2021) to cumulative and peak precipitation features, respectively. Additionally, we computed HCPR and HPPR from these smoothed values. Following this cell-level processing, we utilized the depth estimator with optimal predictive performance as identified in Section 2.3 to generate inundation depths resulting from each synthetic precipitation event. The depth estimator was applied to the thousands of generated rainfall events and produced a synthetic inundation depth distribution in each cell.

To assess the event-generation performance, we evaluated the statistical similarity between the inundation depth distribution of the training set of rainfall events and the inundation depth distribution of the synthetic set. We implemented interpolation using the formula in Equation (5) for the synthetic set depth distribution $\mathbf{d_s}$ to ensure the same dimension as the training set depth distribution $\mathbf{d_t}$, facilitating the evaluation. Let $\mathbf{d_t} = [d_t(1), d_t(2), \ldots, d_t(m)]$ be the training set depth vector with m records, and $\mathbf{d_s} = [d_s(1), d_s(2), \ldots, d_s(n)]$ be the synthetic set depth vector with n synthetic records. We define an interpolation function f such that:

$$f(i) = d_s(i) \quad \text{for} \quad i = 1, 2, \ldots, n$$

The interpolated values $d'_s(j)$ for $j = 1, 2, \ldots, m$ are given by:

$$d'_s(j) = f\left(\frac{j-1}{m-1} \cdot (n-1) + 1\right) \tag{5}$$

This process ensures that $\mathbf{d_t}$ and $\mathbf{d'_s}$ are of the same length, enabling valid comparisons using multiple statistical measures. Cosine similarity (Vijaymeena and Kavitha, 2016) assesses their directional similarity. Pearson correlation (Freedman et al., 2007) evaluates the linear relationship between the two distributions. Finally, Kullback–Leibler (KL) divergence (Hershey & Olsen, 2007) quantifies how one probability distribution diverges from the other. The corresponding formulas can be written as:

$$\text{Cosine Similarity}(\mathbf{d_t}, \mathbf{d'_s}) = \frac{\mathbf{d_t} \cdot \mathbf{d'_s}}{\|\mathbf{d_t}\| \|\mathbf{d'_s}\|} \tag{6}$$

$$\text{Corr}(\mathbf{d_t}, \mathbf{d'_s}) = \frac{\sum_{i=1}^{m} (d_t(i) - \bar{d}_t)(d'_s(i) - \bar{d}'_s)}{\sqrt{\sum_{i=1}^{m} (d_t(i) - \bar{d}_t)^2 \sum_{i=1}^{m} (d'_s(i) - \bar{d}'_s)^2}} \tag{7}$$

$$D_{KL}(\mathbf{p} \parallel \mathbf{q}) = \sum_{i=1}^{m} p(i) \log \frac{p(i)}{q(i)} \tag{8}$$

$$p(i) = \frac{d_t(i)}{\sum_{k=1}^{m} d_t(k)} \tag{9}$$

$$q(i) = \frac{d'_s(i)}{\sum_{k=1}^{m} d'_s(k)} \tag{10}$$

where $\mathbf{p}$ and $\mathbf{q}$ are the normalized versions of $\mathbf{d_t}$ and $\mathbf{d'_s}$, respectively, as represented in Equations (9) and (10). Besides the averaged synthetic depth distribution across the study region, we further explore the performance by aggregating the results into channel and nonchannel levels, as described in Section 2.3. This detailed analysis allows to assess how well the synthetic data capture the unique characteristics of different regions in the study area.

3 | RESULTS

3.1 | Selection of depth estimator

To understand the advancements of tree-based models and transformers in regression problems, specifically in the context of inundation depth estimation, we compared the predictive performance of two universal models, XGBoost Regressor and Regression Transformer, with one cell-wise model, MaxFloodCast V2. MaxFloodCast V2 utilizes an XGBoost-based architecture and incorporates heavy precipitation ratio features. We trained the Regression Transformer using an encoder with four layers, a model

TABLE 3 Depth estimator comparison on test data set.

Model	Level	Overal RMSE	Overal R^2	Channel RMSE	Channel R^2	Nonchannel RMSE	Nonchannel R^2	Time
Regression Transformer (Vaswani et al., 2017)	Universal	2.6634	0.6537	4.1264	0.7278	2.4133	0.4928	5.8 s
XGBoost Regressor (Chen & Guestrin, 2016)	Universal	2.6916	0.6463	4.2755	0.7078	2.4153	0.4920	4.1 s
MaxFloodCast V2 (Lee et al., 2024)	Cell-wise	0.6996	0.9189	1.7280	0.9107	0.5680	0.9200	10.6 s

dimensionality of 128, a feed-forward layer size of 2048, and eight attention heads with a 25% dropout rate on eight NVIDIA RTX A100 GPUs. The XGBoost Regressor was trained on an NVIDIA RTX A6000, while MaxFloodCast V2 was trained in parallel on an AMD EPYC 7702P 64-Core Processor, sharing the same hyperparameter settings as the XGBoost Regressor. The model's objective function aimed to minimize RMSE, utilizing a learning rate of 0.01. The XGBoost algorithm generated 1000 trees with a maximum depth of 5, applying L1 regularization to mitigate overfitting and employing a subsample ratio of 0.3 to introduce nonlinearity and improve model robustness. Following the sampling strategy described in Section 2.2, we configured 252 precipitation-flood events for training, 63 events for validation, and the remaining events as the test set. The comparison results are presented in Table 3, which assesses the overall, channel, and nonchannel RMSE, as well as the R^2 score.

The comparison of results reveals significant differences in the predictive performance between the universal models and the cell-wise model. The Regression Transformer and XGBoost Regressor both show substantial errors, with overall RMSE scores above 2 ft, indicating they struggle to accurately predict inundation depths. Their performance is particularly poor in channel cells, where RMSE values exceed 4 ft. In contrast, MaxFloodCast V2 demonstrates superior performance with an overall RMSE of 0.6996 ft and an R^2 value of 0.9189, significantly outperforming the universal models. It achieves lower RMSE in both channel (1.7280 ft) and nonchannel cells (0.5680 ft), highlighting its accuracy and reliability. The cell-wise approach of MaxFloodCast V2, which tailors the model to the specific characteristics of each cell, allows it to better capture local variations in precipitation and terrain features, resulting in more accurate predictions. Also, the incorporation of heavy precipitation ratio features enhances its ability to capture the influence of runoff from upstream areas on channel cell depths. The data set used for training, generated by the physics-based model, is both limited and imbalanced, comprising 90 training events. This limitation affects the universal models more significantly, as they are less adept at handling such imbalances compared to the specialized cell-wise approach of MaxFloodCast V2. Overall, while the cell-wise model demands higher computational resources and longer processing times than universal models, its advantages outweigh these costs in the context of this pipeline, which focuses on region-specific flood depth estimation. Given its superior performance in capturing complex hydrological dynamics, the computational overhead is a minor trade-off, making MaxFloodCast V2 the optimal depth estimator for the Precipitation-Flood Depth Generative Pipeline.

3.2 | CTGAN-generated precipitation records quality review

Among the initial 592 precipitation-flood events, we strategically sampled 90 events following the 2:4:1 ratio described in Section 2.2 and configured 36 hyperparameter sets for our grid search. This involved combining six pairs of learning rate settings with six distinct stoppage epochs, increasing from 50 epochs to 300 epochs in increments of 50, and modeling on NVIDIA's RTX A6000 GPU. This meticulous grid search focused on generating records under the condition duration ≥ 1, targeting the potential simulation of global precipitation events. As shown in Figure 3I, we identified the optimal checkpoint with a generator learning rate of 10^{-2}, a discriminator learning rate of 10^{-3}, and early stopping at 250 epochs. Notably, the generator learning rate is set to be 10 times higher than that of the discriminator to prevent the generator from being overpowered in the early stages before it can effectively learn the underlying data patterns. Ultimately, both the generator and discriminator converged before reaching the 300-epoch limit. The optimal settings enabled a direct comparison of cumulative precipitation, peak precipitation, and duration between the training data set and the 10,000,000 synthetic data points generated by CTGAN, as depicted in Figure 3 II. This configuration produced an average marginal distribution score of 0.802, demonstrating a high fidelity of the synthetic data, particularly for cumulative and peak precipitation when compared to the original data set. As shown in Figure 3 II(c), we

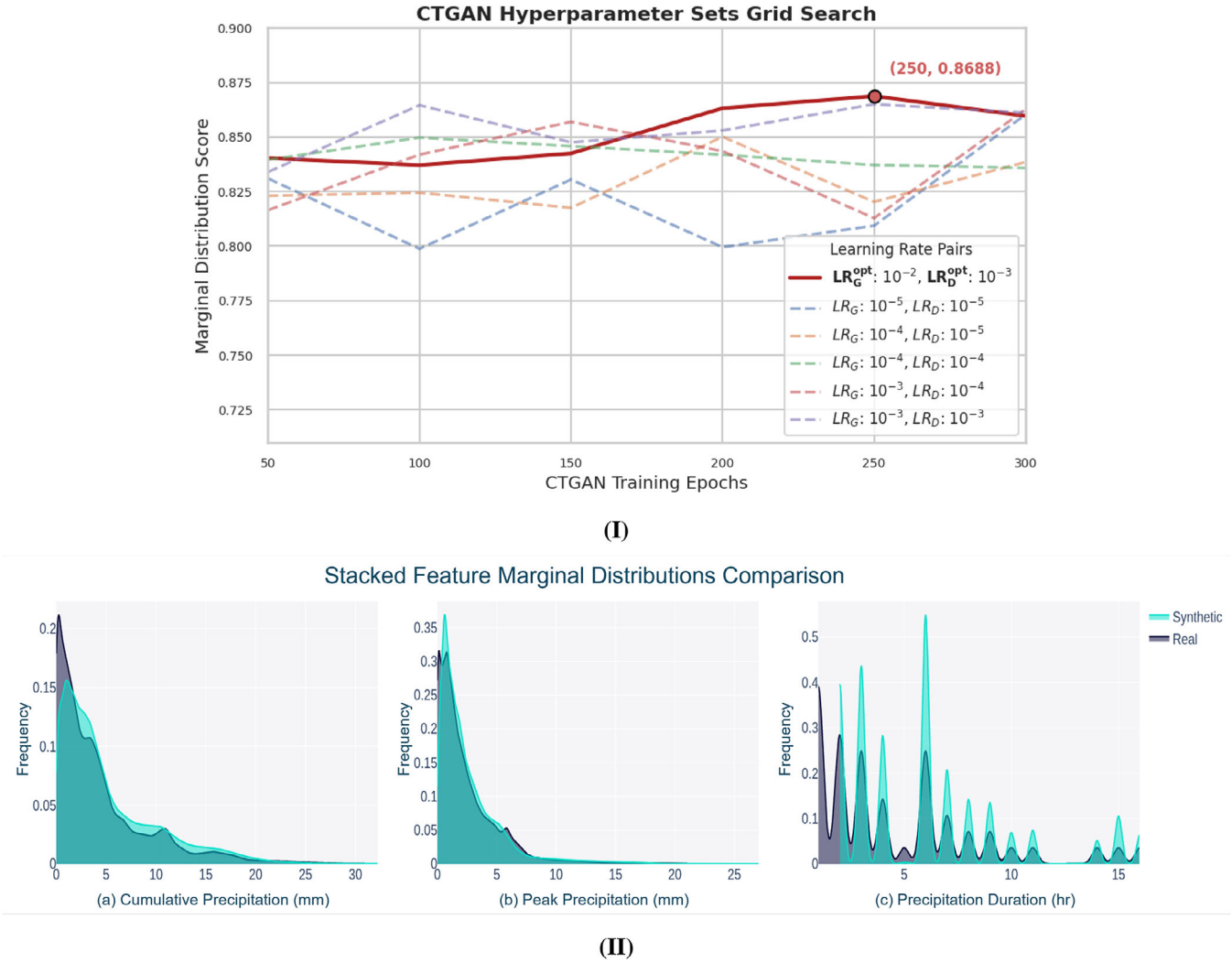

FIGURE 3 (I) Grid search results for conditional generative adversarial network (CTGAN) hyperparameters. The optimal hyperparameter set was selected based on the highest average marginal distribution scores across all synthetic precipitation-based features generated by the best-performing CTGAN configurations. The key hyperparameters explored in the grid search include the generator and discriminator learning rates, as well as the number of training epochs. (II) Stacked marginal distribution comparison of three synthetic features between training data and synthetic data: (a) cumulative precipitation, (b) peak precipitation, and (c) duration. The distribution of the training data set is represented in gray, while the synthetic distribution is depicted in light blue. Distributions were generated using Synthetic Data Vault (SDV) (Patki et al., 2016).

enforced a constraint in CTGAN that limits duration ≥ 1, since our focus is primarily on global precipitation event generation. Additionally, the CTGAN model maintains the flexibility to generate local precipitation events by allowing cumulative precipitation to drop to zero. Moving forward, the challenge remains to construct multiple precipitation-depth events from the synthetic point data set that closely correlate with training events.

3.3 | Synthetic rainfall event assessment

Using the strategic filtering and sampling methods outlined in Section 2.5, we generated 10,000 synthetic rainfall events through parallel computing. For each event, the sampled global duration level guided and constrained the selection of precipitation features from each cell's tailored pool. Leveraging the 27-element precipitation-level combination pool described in Section 2.5, the cell-level sampling preserved regular precipitation trends while maintaining the probability of extreme scenarios. As the number of simulated events increased, the overall synthetic precipitation patterns closely aligned with those in the training data set. Due to the independence and nonlinearity of point data in CTGAN-generated records, cumulative and peak precipitation maps often exhibited a "pepper-salt" appearance. To mitigate this, we applied KNN smoothing, selecting the value of K based on the synthetic event's duration follow-

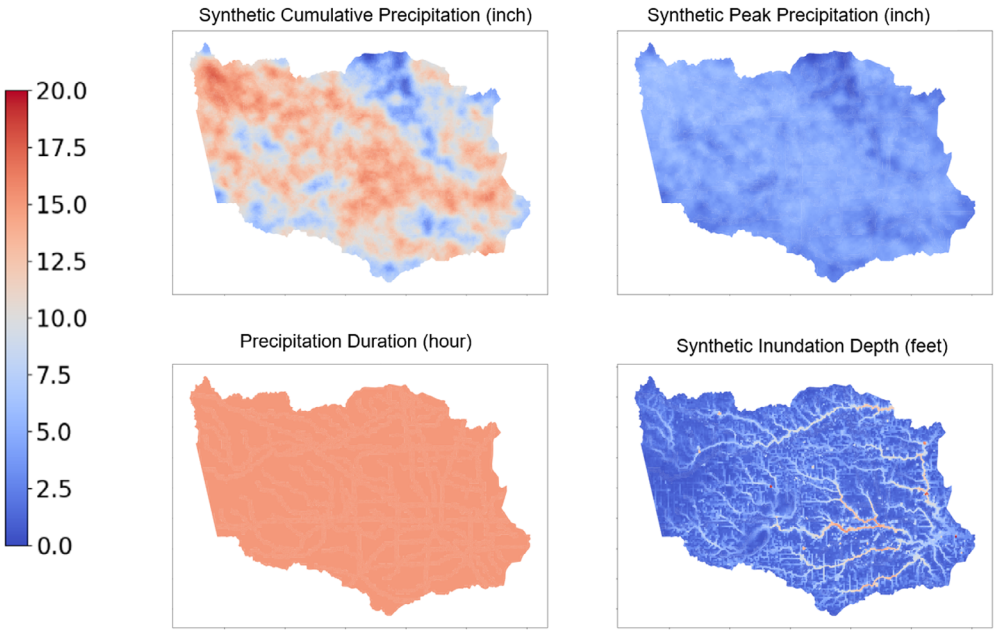

FIGURE 4 Synthetic flood event with 15 h global precipitation. All the attribute maps are in the same scale and they share the same color bar with different units: inch, inch, hour, and feet, respectively. The synthetic cumulative precipitation and peak precipitation are processed by 50-NN smoother. Maps generated by *Geopandas* Python package.

ing that the longer duration correspond to larger K values, and vice versa. From these smoothed precipitation-based features, we derived HCPR and HPPR, resulting in 21 input features for the depth estimator, MaxFloodCast V2, which was then used to generate synthetic inundation depths and complete the precipitation-flood event. The event generation process averaged 10 min for 100 rainfall events in parallel, utilizing 100 GB of memory on an AMD EPYC 7702P 64-core processor.

Figure 4 illustrates a synthetic precipitation-flood event with a global duration of 15 h, comprising four subplots: synthetic cumulative precipitation, synthetic peak precipitation, precipitation duration, and synthetic inundation depth. In Harris County, this 15-h storm event would have a return period of approximately 200 years (Perica et al., 2018). Cumulative and peak precipitation maps show some spatial variability, as expected for a relatively long but still sub-daily rainfall. The inundation depth map highlights flood-prone regions, particularly along river channels and low-lying areas. This result is consistent with the expectation that a 200-year storm would cause multiple floodings at the local level in the area.

3.4 | Synthetic rainfall distribution assessment

We generated 10,000 synthetic precipitation-flood events, each cell having a comprehensive synthetic depth distribution for direct comparison with the training data set. To assess the performance of the synthetic events quantitatively, we utilized four key metrics—RMSE, cosine similarity, correlation, and KL divergence—comparing them against the limited number of sampled training events. To ensure a fair comparison, the synthetic events were downsampled to match the size of the training data. This downsampling process was repeated 50 times to minimize variability introduced by randomness. A detailed statistical comparison between the normalized versions of the training and the synthetic data sets is provided in Table 4, with the corresponding visualizations shown in Figure 5.

The sampled synthetic flood depth distribution demonstrates comparable overall performance, as evidenced by the RMSE and cosine similarity metrics, indicating that the generated events align closely with the training data and accurately capture key hydrological characteristics and trends. Notably, the mean RMSE for nonchannel cells is 2.10 ft, substantially lower than the 5.99 ft recorded for channel cells. This disparity suggests that the synthetic data more effectively represent regions with simpler hydrodynamics, while the higher RMSE in channel cells reflects the inherent difficulty in modeling real-world depth distributions in areas characterized by dynamic flow patterns and greater variability. Nevertheless, the synthetic data perform well even in these more complex regions, achieving a mean cosine similarity of 0.85 for channel cells, highlighting the model's ability to capture the overall structure and trends of the real data.

TABLE 4 Statistical comparison of sampled training depth distribution and downsampled synthetic depth distribution for key metrics across overall, channel, and nonchannel cells.

Metric	RMSE (ft)	Cosine similarity	Correlation	KL divergence
Overall				
Count	26301	26301	26301	26301
Mean	2.5813	0.8537	0.5621	0.8332
Std	2.0925	0.0553	0.0996	0.4610
Min	0.3248	0.3523	0.0010	0.0236
25%	1.1264	0.8452	0.5205	0.5308
50%	1.8231	0.8627	0.5588	0.7692
75%	3.3417	0.8911	0.6034	1.0227
Max	26.982	0.9923	0.8812	6.2701
Channel				
Count	2985	2985	2985	2985
Mean	5.9978	0.8461	0.5433	0.7831
Std	2.7955	0.0489	0.0998	0.4392
Min	0.6084	0.4735	0.0015	0.0905
25%	3.9428	0.8276	0.5151	0.4697
50%	5.4801	0.8662	0.5513	0.7598
75%	7.7058	0.8785	0.6121	1.0152
Max	27.015	0.9578	0.8015	5.1892
Nonchannel				
Count	23316	23316	23316	23316
Mean	2.0992	0.8696	0.5525	0.8224
Std	1.4678	0.0562	0.1013	0.4532
Min	0.3221	0.3511	0.0010	0.0240
25%	1.0912	0.8415	0.5240	0.5521
50%	1.6098	0.8772	0.5644	0.7559
75%	2.7623	0.9028	0.6183	1.0194
Max	16.940	0.9815	0.8897	6.2550

In nonchannel cells, the model excels with a mean cosine similarity of 0.87 and a relatively low KL divergence of 0.82. These metrics suggest that the synthetic data closely match the probability distributions of the real data in these regions, where simpler hydrological behavior makes depth distributions easier to replicate. In contrast, the KL divergence for channel cells, though still low at 0.78, indicates some deviation between the synthetic and real data, which is expected given the complexity and variability of flood behavior in these more dynamic areas.

The correlation between the synthetic and real data sets further reinforces the model's effectiveness. Nonchannel cells exhibit a mean correlation of 0.55, while channel cells show a slightly lower but still robust correlation of 0.54. This highlights the model's ability to capture linear relationships across both cell types, though the more intricate hydrodynamic processes in channel cells present additional challenges. Hydrologically, these findings are particularly significant. Channel cells, which are more prone to flooding due to their dynamic flow patterns, underscore the model's strength in capturing intricate interactions even under complex conditions. The model's ability to generalize across nonchannel cells, reflected in their lower RMSE and KL divergence values, is equally noteworthy, showing the model's potential to explore a variety of flood scenarios without overfitting to specific events.

Geographically, the distinction between channel and nonchannel cells emphasizes the importance of local topographical and hydrological features in flood modeling. The synthetic data perform better in regions with clear hydrodynamic patterns, while still maintaining a reasonable alignment in flat areas where riverine and flash flooding may compound.

Overall, these results highlight the strong potential of the synthetic data set for reliable flood depth estimation across diverse regions, especially in flood-prone channel areas where accurate modeling is crucial for effective risk assessment and mitigation. The close alignment

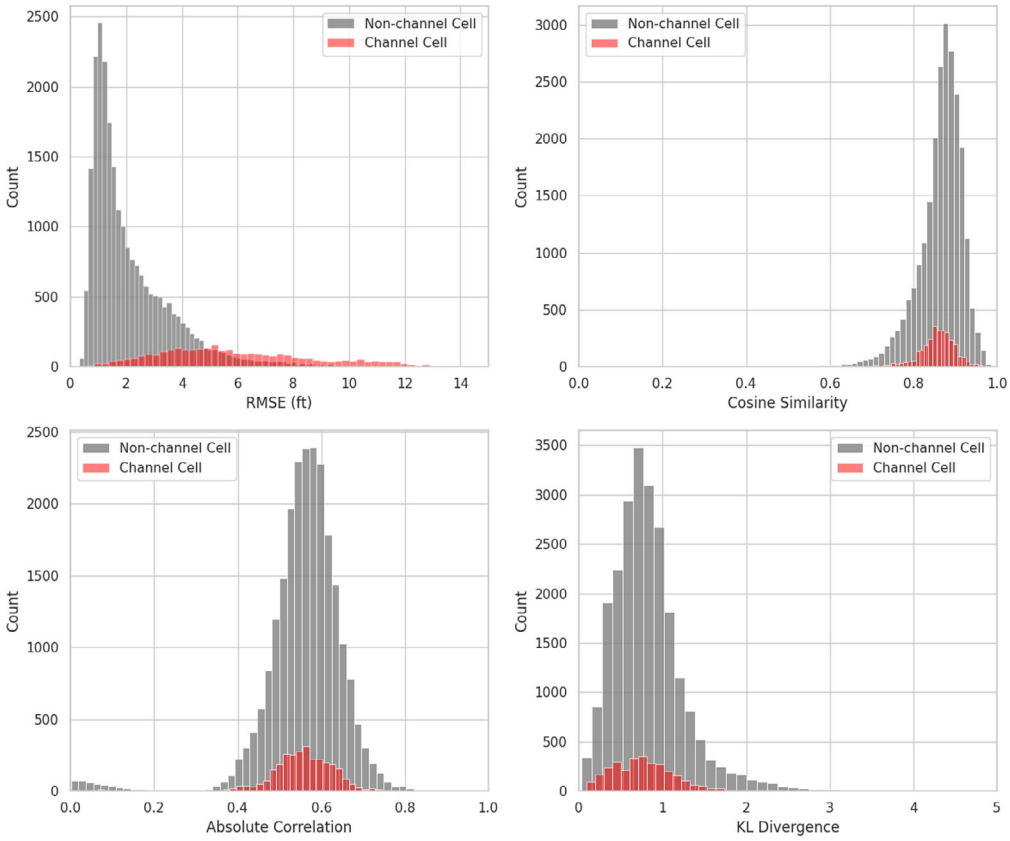

FIGURE 5 Histograms of flood assessment metrics: Comparison of metrics between channel cells and nonchannel cells, based on the differences between the sampled training depth distribution and the downsampled synthetic depth distribution.

between the synthetic and training data across multiple metrics underscores the robustness of the approach, demonstrating its ability to replicate complex flood patterns while maintaining generalizability across varied hydrological landscapes. These findings affirm the method's suitability for predictive flood modeling and underscore its potential as a powerful tool for future flood risk assessment and management efforts.

3.5 | Flood probability maps

Flood depths generated by a large number of synthetic precipitation-flood events represent a foundational data set from which one can estimate flood probability. Probability is calculated cell-wise as the proportion of synthetic events that cause a certain value of flood depth or higher. Using the 10,000 synthetic events simulated in Section 3.3, we built synthetic flood probability maps for the following inundation depths: 1 ft, 2 ft, 4 ft, and 6 ft (Figure 6). In Figure 6.IV, for example, a high $P(\text{depth} \geq 6ft)$ in a cell means that for a generic flood event, there is a high probability that flood level will equal or exceed 6 ft in that cell. In other words, the probability value serves as a quantification of risk for a certain flooding depth. This interpretation should be distinct from annual flood probability, which is traditionally used in flood mapping. As shown in Figure 6, at 1 ft and 2 ft depths, flood probabilities greater than 0.5 (high probability) are predominantly observed in the northeastern, southeastern, and central regions of the study area, while lower probabilities are seen in the western and southern areas. As the depth threshold increases to 4 ft and 6 ft, the high-probability zones become more concentrated along rivers/channels and low-lying regions.

4 | DISCUSSION

In this study, we presented an innovative application methodology, Precipitation-Flood Depth Generative Pipeline, for generating high-resolution flood probability maps using synthetic precipitation-flood events. Our approach leveraged advanced machine learning techniques, including CTGAN (Xu et al., 2019) as precipitation-based features generator and XGBoost-based MaxFlood-

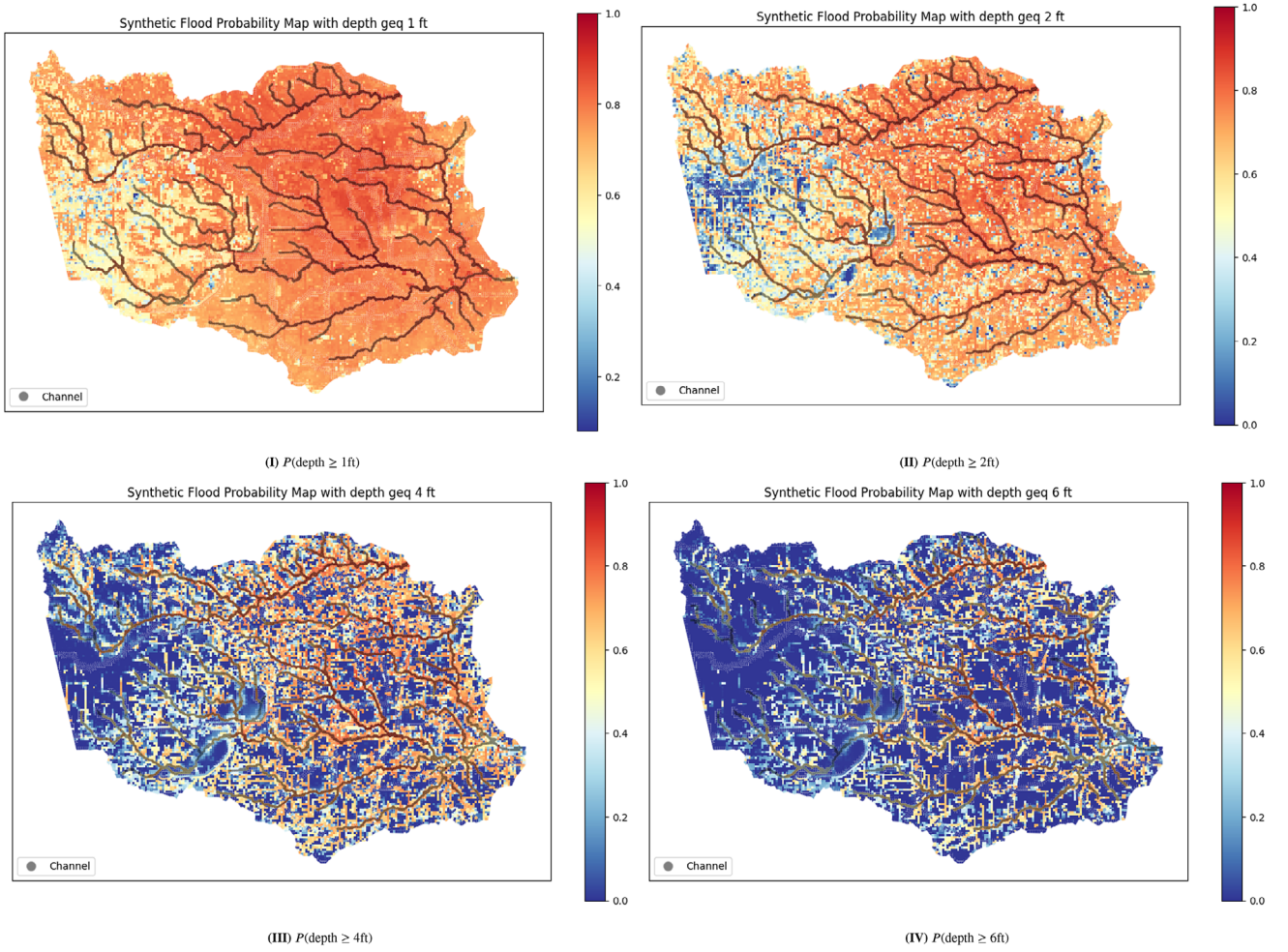

FIGURE 6 Synthetic flood probability maps based on four inundation depth criteria. The color gradient from blue to red represents probabilities ranging from 0 to 1 for a global precipitation event, with blue indicating lower probabilities and red indicating higher probabilities. Channel cells are highlighted in gray. Maps generated by *Geopandas* Python package.

Cast (Lee et al., 2024) as synthetic depth estimator, to overcome the limitations of traditional flood modeling that rely on historical data and records, not always available. To address the limitations in feature dimensionality found in existing CTGAN tools and the challenge of generating individual precipitation-flood events from a massive synthetic point cloud, we designed a robust and systematic workflow. This includes cell-level pool creation, strategic sampling, noise smoothing, depth synthesis, and the eventual formation of individual events. Each step in the process is carefully structured to avoid overfitting and ensure that the generated synthetic events accurately reflect real-world flood scenarios while maintaining computational efficiency and high fidelity to hydrological patterns. By generating 10,000 synthetic precipitation-flood events, we created a data set that captures a wide spectrum of potential flood scenarios, increasing the robustness of flood risk assessments. The RMSE values highlight a notable difference between channel and nonchannel cells, reflecting the challenges in accurately modeling flood depth distribution in regions with dynamic flows and greater variability. These findings demonstrate the model's ability to capture key trends and nonlinear relationships, particularly along channels and in floodplain areas, which are more susceptible to flooding. Additionally, the synthetic data's capacity to generalize beyond the training events highlights its potential for producing possible flood scenarios, which is critical for accurate risk assessment and mitigation planning.

4.1 | Spatial nature of weather and flooding

Despite the advancements introduced with the Precipitation-Flood Depth Generative Pipeline, this approach has some limitations in generating local precipitation events. While CTGAN can generate records

with no precipitation, these are sparse, and we observed a tendency to generate storms covering most of the study area. In reality, short storms (0.5–2 h), whether of high or low intensity, should affect a portion of the study area, that is, the cells that receive rain and the cells downstream. The localized pattern of short storms is not well captured by the CTGAN generator. The prevalence of large-extent storms in the synthetic data set influenced flood depth distributions at the cell level and the final maps. As a result, flood probability for low depths (1 and 2 ft) is likely overestimated outside of channels, given the spatial uniformity of high probabilities (> 75%) observable even on elevated road cells and at watershed divides 6. Traditional flood probability maps for Harris County show the extent of the 100-year and 500-year floodplain, but do not provide information on associated flood depths at the largest scale, complicating a direct comparison with the synthetic flood probability maps in Figure 6. For the higher depths considered (4 and 6 ft), synthetic flood probabilities have more distinct values between low-lying and high-lying areas, and align more closely to the latest 100-year and 500-year floodplain extents (TWDB, 2024).

Synthetic flood probability overestimation can be explained because the three storm characteristics, that is, duration, cumulative rainfall, and peak intensity, are generated at the cell level independently, without considering the spatial autocorrelation existing in an actual storm. Future work will focus on additional characteristics that capture the spatial nature of precipitation events. These include the spatial autocorrelation of storm events of different duration, intensity, and extent, the directions along which storms typically move, and other spatio-temporal properties defined at the event-scale rather than cell-wise, for example, rotation and velocity.

4.2 | Computational efficiency

The computational complexity of the pipeline is overall significant. The cell-wise depth estimator, MaxFloodCast, costs higher computational resource than the universal models tested (Section 3.1), taking approximately 10 min and 100 GB CPU memory to generate 100 precipitation-flood events. This performance is however competitive over HEC-RAS 2D, for which Lee et al. (2024) found that the runtime for a single simulation was around three minutes on average. The obvious advantage of MaxFloodCast is the parallelization across cells, which is not possible in HEC-RAS 2D. Future work should improve the computational efficiency of the pipeline, especially in presence of additional relationships of spatial and hydrological nature that would enhance precipitation event generation.

4.3 | Applications and future research

The Precipitation-Flood Depth Generative Pipeline was applied in Harris County due to the availability of an initial data set of storms for training, as well as of a calibrated physics-based model that provided the data for training the depth-estimator. The pipeline can be applied in any study area as long as a set of historical storms and resulting flood depth levels are available. However, different model parameterization might be necessary to adapt to region-specific meteorological and hydrological conditions, especially for the storm generator component. Future research should determine what the appropriate scale is for the pipeline application, and if the scale difference between weather analysis (regional scale) and flood depth analysis (watershed scale) represents an issue for implementing the pipeline and producing accurate flood probability maps. For example, weather radar measurements tend to underestimate precipitation in small catchments (Sokol et al., 2021), especially those that exhibit a flashy response (Ochoa-Rodriguez et al., 2015). Consequently, the storm generator trained with historical radar data in these areas would be biased toward low-intensity and smoothed storms and underestimate flood risk. In absence of a local network of rain gauges that helps with radar error mitigation, precipitation analysis should be expanded to a larger scale and include nearby watersheds, to ensure that regional storm patterns are accurately captured and represented in the synthetic storm data set.

Flood probability maps visualize flood risk as a function of water depth. In regions with a flat ground surface that are subject to both riverine and flash flood mechanisms, maps can support depth-specific flood management strategies. For example, they can leverage LFE data sets to assess flood risk at the building level; to prioritize structural mitigation interventions (e.g., foundation raising) based on inundation probability at a certain water height threshold; or to identify roads with a high probability to become impassable due to a 2-ft inundation. The nonbinary nature of maps further allows for comparative flood risk assessments across different inundation levels, enabling more nuanced and targeted flood mitigation strategies across diverse landscapes.

Finally, this study is an example of synthetic data implementation in disaster resilience models. Real and synthetic data support the development of robust AI models for various hazard-related applications (Liu et al., 2025). For example, AI-generated storms can be used to augment flood scenarios and loss data in addition to the available historical data, increasing awareness on vulnerable assets and assisting decision-makers on preparedness and mitigation actions in peace time. Further, AI models play a

pivotal role in hybrid models for weather and flood forecasting, where real-time measurements from gauges and radar are integrated with synthetic data (Slater et al., 2023). Despite the development of hybrid forecasting models in many regions of the world, their effect on disaster operations is not observable because of their novelty and limited adoption (Panahi et al., 2021). Therefore, future studies should investigate how to leverage GAN and other generative ML methods for augmenting data from past storm and flood events, but also explore how these methods and their products can be effectively and reliably integrated with decisions and operations before, during, and after a flood event.

5 | CONCLUSIONS

In this study, we proposed the Precipitation-Flood Depth Generative Pipeline, a novel machine learning framework for generating synthetic precipitation-flood events to construct high-resolution probabilistic flood maps. By integrating a CTGAN for synthetic precipitation generation with a specialized cell-wise depth estimator (MaxFlood-Cast V2), our approach addresses some of the limitations in traditional flood mapping methods, including data scarcity and computational costs associated with physics-based simulations. The results show the capability of the pipeline to produce flood probability maps that align with hydrological intuition and known flood-prone areas, highlighting its potential as a scalable tool for flood risk assessment.

Despite its advantages, certain limitations remain, such as potential inconsistencies in extreme precipitation cases and challenges in fully capturing spatial dependencies. Future research will focus on enhancing the integration of generative models with physics-based simulations, improving event-generation mechanisms to better capture localized precipitations, and further benchmarking against alternative deep learning models. Additionally, extending this framework to different geographical regions and incorporating real-time flood forecasting applications could expand its scope in disaster resilience and urban planning.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support provided by the National Science Foundation under the CRISP 2.0 Type 2 grant (grant number 1832662) and the Texas A&M X-Grant Presidential Excellence Fund. The perspectives, results, conclusions, and suggestions presented in this research are entirely those of the authors and do not necessarily represent the viewpoints of the funding organizations.

DATA AVAILABILITY STATEMENT

The code and data that support the findings of this study are available from the corresponding author upon request.

REFERENCES

Castangia, M., Grajales, L. M. M., Aliberti, A., Rossi, C., Macii, A., Macii, E., & Patti, E. (2023). Transformer neural networks for interpretable flood forecasting. *Environmental Modelling & Software*, *160*, 105581. https://doi.org/10.1016/j.envsoft.2022.105581

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In: *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PEERJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Dewitz, J. (2024). *National land cover database (NLCD) 2019 products (ver. 3.0, february 2024)*. US Geological Survey. https://doi.org/10.5066/P9KZCM54

Do Lago, C. A., Giacomoni, M. H., Bentivoglio, R., Taormina, R., Junior, M. N. G., & Mendiondo, E. M. (2023). Generalizing rapid flood predictions to unseen urban catchments with conditional generative adversarial networks. *Journal of Hydrology*, *618*, 129276. https://doi.org/10.1016/j.jhydrol.2023.129276

Dong, S., Yu, T., Farahmand, H., & Mostafavi, A. (2020). Bayesian modeling of flood control networks for failure cascade characterization and vulnerability assessment. *Computer-Aided Civil and Infrastructure Engineering*, *35*(7), 668–684. https://doi.org/10.1111/mice.12527

Dong, S., Yu, T., Farahmand, H., & Mostafavi, A. (2021). A hybrid deep learning model for predictive flood warning and situation awareness using channel network sensors data. *Computer-Aided Civil and Infrastructure Engineering*, *36*(4), 402–420. https://doi.org/10.1111/mice.12629

Finner, H., & Gontscharuk, V. (2018). Two-sample Kolmogorov–Smirnov-type tests revisited: Old and new tests in terms of local levels. *The Annals of Statistics*, *46*(6A). https://doi.org/10.1214/17-aos1647

Freedman, D. R., Pisani, R., & Purves, R. (2007). Statistics. (4th ed. 415-424, 488-495, 523-540). W. W. Norton & Company, New York.

Garcia, M., Juan, A., Doss-Gollin, J., & Bedient, P. (2023). Leveraging mesh modularization to lower the computational cost of localized updates to regional 2D hydrodynamic model outputs. *Engineering Applications of Computational Fluid Mechanics*, *17*(1), 2225584. https://doi.org/10.1080/19942060.2023.2225584

Garcia, M. S. (2023). Novel urban floodplain modeling methods for applications in coupling surrogate machine learning methods (Doctoral dissertation, Rice University). https://hdl.handle.net/1911/115079

Harris County Flood Control District. (2023). *Harris county flood warning system*. https://www.harriscountyfws.org/

Hershey, J. R., & Olsen, P. A. (2007). Approximating the Kullback Leibler divergence between Gaussian mixture models. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSSP '07* (*Vol. 4*, pp. IV-317–IV-320). IEEE Computer Society. https://doi.org/10.1109/ICASSP.2007.366913

Ho, Y.-H., Li, L., & Mostafavi, A. (2025). Integrated vision language and foundation model for automated estimation of building lowest floor elevation. *Computer-Aided Civil and Infrastructure Engineering*, *40*(1), 75–90. https://doi.org/10.1111/mice.13310

Ji, H. K., Mirzaei, M., Lai, S. H., Dehghani, A., & Dehghani, A. (2024). Implementing generative adversarial network (GAN) as a data-driven multi-site stochastic weather generator for flood frequency estimation. *Environmental Modelling & Software*, *172*, 105896. https://doi.org/10.1016/j.envsoft.2023.105896

Konrad, M. (2021). Geovoronoi (Version 0.3.0) [Computer software]. https://doi.org/10.5281/ZENODO.4531338

Lee, C.-C., Huang, L., Antolini, F., Garcia, M., Juan, A., Brody, S. D., & Mostafavi, A. (2024). Predicting peak inundation depths with a physics informed machine learning model. *Scientific Reports*, *14*(1), 14826. https://doi.org/10.1038/s41598-024-65570-8

Lee, J. S., Park, J., Kim, H. M., & Kim, R. E. (2025). Damage detection for railway bridges using time-frequency decomposition and conditional generative model. *Computer-Aided Civil and Infrastructure Engineering*, *40*(7), 959–977. https://doi.org/10.1111/mice.13372

Liu, Z., Coleman, N., Patrascu, F. I., Yin, K., Li, X., & Mostafavi, A. (2025). Artificial intelligence for flood risk management: A comprehensive state-of-the-art review and future directions. *International Journal of Disaster Risk Reduction*, *117*, 105110. https://doi.org/10.1016/j.ijdrr.2024.105110

Lyu, H.-M., & Yin, Z.-Y. (2023). Flood susceptibility prediction using tree-based machine learning models in the GBA. *Sustainable Cities and Society*, *97*, 104744. https://doi.org/10.1016/j.scs.2023.104744

McSpadden, D., Goldenberg, S., Roy, B., Schram, M., Goodall, J. L., & Richter, H. (2024). A comparison of machine learning surrogate models of street-scale flooding in Norfolk, Virginia. *Machine Learning with Applications*, *15*, 100518. https://doi.org/10.1016/j.mlwa.2023.100518

Musselman, K. N., Lehner, F., Ikeda, K., Clark, M. P., Prein, A. F., Liu, C., Barlage, M., & Rasmussen, R. (2018). Projected increases and shifts in rain-on-snow flood risk over western North America. *Nature Climate Change*, *8*(9), 808–812. https://doi.org/10.1038/s41558-018-0236-4

Naeini, S. S., & Snaiki, R. (2024). A physics-informed machine learning model for time-dependent wave runup prediction. *Ocean Engineering*, *295*, 116986. https://doi.org/10.1016/j.oceaneng.2024.116986

Nofal, O., Rosenheim, N., Kameshwar, S., Patil, J., Zhou, X., van de Lindt, J. W., Duenas-Osorio, L., Cha, E. J., Endrami, A., Sutley, E., Cutler, H., Lu, T., Wang, C., & Jeon, H. (2024). Community-level post-hazard functionality methodology for buildings exposed to floods. *Computer-Aided Civil and Infrastructure Engineering*, *39*(8), 1099–1122. https://doi.org/10.1111/mice.13135

Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., Kroll, S., Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P., & ten Veldhuis, M.-C. (2015). Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. *Journal of Hydrology*, *531*, 389–407. https://doi.org/10.1016/j.jhydrol.2015.05.035

Panahi, M., Jaafari, A., Shirzadi, A., Shahabi, H., Rahmati, O., Omidvar, E., Lee, S., & Bui, D. T. (2021). Deep learning neural networks for spatially explicit prediction of flash flood probability. *Geoscience Frontiers*, *12*(3), 101076. https://doi.org/10.1016/j.gsf.2020.09.007

Patki, N., Wedge, R., & Veeramachaneni, K. (2016). The synthetic data vault. In *2016 IEEE 3rd International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE Computer Society, 399–410. https://doi.org/10.1109/DSAA.2016.49

Perica, S., Pavlovic, S., St Laurent, M., Trypaluk, C., Unruh, D., & Wilhite, O. (2018). NOAA Atlas 14 Volume 11 Version 2, *Precipitation-Frequency Atlas of the United States, Texas*. NOAA, National Weather Service, Silver Spring, MD. https://doi.org/10.25923/1ceg-5094

Piadeh, F., Behzadian, K., Chen, A. S., Campos, L. C., Rizzuto, J. P., & Kapelan, Z. (2023). Event-based decision support algorithm for real-time flood forecasting in urban drainage systems using machine learning modelling. *Environmental Modelling & Software*, *167*, 105772. https://doi.org/10.1016/j.envsoft.2023.105772

Shehadeh, A., Alshboul, O., Al Mamlook, R. E., & Hamedat, O. (2021). Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, lightgbm, and XGBoost regression. *Automation in Construction*, *129*, 103827. https://doi.org/10.1016/j.autcon.2021.103827

Slater, L. J., Arnal, L., Boucher, M.-A., Chang, A. Y.-Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A., & Zappa, M. (2023). Hybrid forecasting: Blending climate predictions with AI models. *Hydrology and Earth System Sciences*, *27*(9), 1865–1889. https://doi.org/10.5194/hess-27-1865-2023

Sokol, Z., Szturc, J., Orellana-Alvear, J., Popova, J., Jurczyk, A., & Célleri, R. (2021). The role of weather radar in rainfall estimation and its application in meteorological and hydrological modelling—A review. *Remote Sensing*, *13*(3), 351. https://doi.org/10.3390/rs13030351

Soil Survey Staff. *Gridded Soil Survey Geographic (gSSURGO) Database for Texas. United States Department of Agriculture, Natural Resources Conservation Service*. Available online at https://gdg.sc.egov.usda.gov/. December 1, 2024 (2023 official release).

TWDB (2024). *Flood quilt (2024)*. https://twdb-flood-planning-resources-twdb.hub.arcgis.com/pages/flood-quilt-2024

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon et al. (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30, pp. 6000–6010). Curran Associates, Inc.

Vijaymeena, M., & Kavitha, K. (2016). A survey on similarity measures in text mining. *Machine Learning and Applications: An International Journal*, *3*(2), 19–28. https://doi.org/10.5121/mlaij.2016.3103

Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 32, pp. 7335–7345). Curran Associates, Inc.

Yao, K., Wen, Z., Shao, C., Yang, J., & Su, H. (2024). A multisource data-driven monitoring model for assessing concrete dam behavior. *Computer-Aided Civil and Infrastructure Engineering*, *39*(23), 3595–3609. https://doi.org/10.1111/mice.13232

Yildirim, E., Just, C., & Demir, I. (2022). Flood risk assessment and quantification at the community and property level in the state of

Iowa. *International Journal of Disaster Risk Reduction*, *77*, 103106. https://doi.org/10.1016/j.ijdrr.2022.103106

Yin, K., Wu, J., Wang, W., Lee, D.-H., & Wei, Y. (2023). An integrated resilience assessment model of urban transportation network: A case study of 40 cities in China. *Transportation Research Part A: Policy and Practice*, *173*, 103687. https://doi.org/10.1016/j.tra.2023.103687

Yuen, K.-V., & Kuok, S.-C. (2025). Telescopic broad Bayesian learning for big data stream. *Computer-Aided Civil and Infrastructure Engineering*, *40*(1), 33–53. https://doi.org/10.1111/mice.13305

Zhang, N., & Alipour, A. (2023). A stochastic programming approach to enhance the resilience of infrastructure under weather-related risk. *Computer-Aided Civil and Infrastructure Engineering*, *38*(4), 411–432. https://doi.org/10.1111/mice.12843

Zhou, Y., Huang, L., Zhou, T., & Shao, L. (2021). Many-to-one distribution learning and K-nearest neighbor smoothing for thoracic disease identification. Proceedings of the AAAI Conference on Artificial Intelligence, *35*(1), 768–776. https://doi.org/10.1609/aaai.v35i1.16158

Zhu, K., Lai, C., Wang, Z., Zeng, Z., Mao, Z., & Chen, X. (2024). A novel framework for feature simplification and selection in flood susceptibility assessment based on machine learning. *Journal of Hydrology: Regional Studies*, *52*, 101739. https://doi.org/10.1016/j.ejrh.2024.101739

Zhu, Y., Niu, X., & Tian, J. (2024). A machine vision-based intelligent segmentation method for dam underwater cracks using swarm optimization algorithm and deep learning. *Computer-Aided Civil and Infrastructure Engineering*, *40*(10), 1405–1417. https://doi.org/10.1111/mice.13343

How to cite this article: Huang, L., Antolini, F., Mostafavi, A., Blessing, R., Garcia, M., & Brody, S. D. (2025). High-resolution flood probability mapping using generative machine learning with large-scale synthetic precipitation and inundation data. *Computer-Aided Civil and Infrastructure Engineering*, 1–17. https://doi.org/10.1111/mice.13490