Generalizing rapid flood predictions to unseen urban catchments with conditional generative adversarial networks

do Lago, Cesar A.F.; Giacomoni, Marcio H.; Bentivoglio, Roberto; Taormina, Riccardo; Gomes, Marcus N.; Mendiondo, Eduardo M.

DOI
10.1016/j.jhydrol.2023.129276

Publication date
2023

Document Version
Final published version

Published in
Journal of Hydrology

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

Journal of Hydrology 618 (2023) 129276

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Generalizing rapid flood predictions to unseen urban catchments with conditional generative adversarial networks

Cesar A.F. do Lago [a,b,*], Marcio H. Giacomoni [b], Roberto Bentivoglio [c], Riccardo Taormina [c], Marcus N. Gomes Junior [a,b], Eduardo M. Mendiondo [b]

[a] University of Texas at San Antonio - One UTSA Circle, San Antonio, TX 78249, USA
[b] University of Sao Paulo - Av. Trab. Sao Carlense, 400 - Parque Arnold Schimidt, Sao Carlos-SP, 13566-590, Brazil
[c] Delft University of Technology - Stevinweg 1, 2628 CN Delft, Netherlands

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor.

Keywords:
Rapid flood modeling
Deep learning
Generative adversarial networks
Rain on grid

ABSTRACT

Two-dimensional hydrodynamic models are computationally expensive. This drawback can limit their application to solving problems requiring real-time predictions or several simulation runs. Although the literature presented improvements in using Deep Learning as an alternative to hydrodynamic models, Artificial Neural Networks applications for flood prediction cannot satisfactorily predict floods for areas outside the training datasets with different boundary conditions. In this paper, we used a conditional generative adversarial network (cGAN) aiming to generalize flood predictions in catchments not included in the training process. The proposed method, called cGAN-Flood, uses two cGAN models to solve a rain-on-grid problem by first identifying wet cells and then estimating the water depths. The cGANs were trained using HEC-RAS outputs as ground truth. cGAN-Flood distributes a target flood volume (v_t) in a given catchment, which can be calculated via water balance from hydrological simulations. Our approach was trained on ten and tested on five urban catchments with distinct characteristics. The cGAN-Flood was compared to HEC-RAS for different rainfall magnitudes and surface roughness. We also compared our approach to the Weighted Cellular Automata 2D (WCA2D), a rapid flood model (RFM) used for rain-on-grid simulations. Our method successfully predicted water depths in the testing areas, showing that cGAN-Flood could generalize to different locations. However, cGAN-Flood tended to underestimate depths in channels in some areas for events with a small peak of precipitation intensity. cGAN-Flood was 50 and 250 times faster than WCA2D and HEC-RAS, respectively. Due to its computational efficiency and accuracy, we suggest that cGAN-Flood can be applied when fast simulations are necessary, and it can be a viable modeling solution for flood forecasts in large-scale watersheds.

1. Introduction

Floods are the costliest natural disaster for life losses and economic damages (Bulti and Abebe, 2020; Natarajan and Radhakrishnan, 2019). Climate change is likely to increase the frequency and intensity of extreme storm events, which can potentially magnify the impacts of flooding (Aich et al., 2016; Clavet-Gaumont et al., 2017; Yin et al., 2018; do Lago et al., 2021). Flood modeling is a fundamental tool for planning and designing flood mitigation in urban areas. Two-dimensional (2D) hydrodynamic approaches are the most effective models used for urban flood predictions (Cea and Costabile, 2022). However, they are computationally expensive, hampering their application in large-scale watersheds or when many runs are required (e.g., for Monte Carlo simulation and optimization problems). As an alternative, Rapid Flood Models (RFMs) have been developed based on simplified modeling approaches, such as Cellular Automata (Guidolin et al., 2016). More recently, Artificial Neural Networks models have also been trained to reduce the computation time of flood predictions (Kabir et al., 2020).

RFMs are fast flood models that decrease simulation time by implementing simplified hydraulic concepts rather than solving complex physical equations (Teng et al., 2017). These approaches usually aim to distribute a flood volume over a domain (e.g., Jamali et al. (2018, 2019), Ghimire et al. (2013), Guidolin et al. (2016)). While granting significant computational speedups, these simplifications reduce their prediction accuracy in areas characterized by high momentum and velocities. Moreover, RFMs may require low time steps in areas with high velocities and DEM resolution, resulting in longer computational runs (Guidolin et al., 2016).

Deep learning (DL) techniques have been used to predict floods with satisfactory results (Berkhahn et al., 2019; Dtissibe et al., 2020;

* Corresponding author at: University of Texas at San Antonio - One UTSA Circle, San Antonio, TX 78249, USA.
E-mail address: cesar.dolago@utsa.edu (C.A.F. do Lago).

https://doi.org/10.1016/j.jhydrol.2023.129276
Received 3 November 2022; Received in revised form 28 January 2023; Accepted 11 February 2023
Available online 16 February 2023

Kabir et al., 2020; Lowe et al., 2021). DL models use a multiple-layer structure, known as Artificial Neural Networks (ANN), to represent high levels of complexity and learn from high-dimensional data (Sit et al., 2020). Convolutional Neural Networks (CNN), in particular, have gained popularity in earth sciences due to their ability to process image data and extract features from multiple adjacent pixels (Kabir et al., 2020). CNNs have been applied successfully in flood prediction studies. Lowe et al. (2021), applied a U-Net based DL to model maximum water depths in an urban catchment using spatial inputs and rainfall characteristics. They trained their model on parts of the study area and could predict maximum water depths in unseen patches of the same domain. Guo et al. (2022) developed a similar DL model, trained with RFM outputs to predict maximum water depths and flow velocities in different catchments. Although the authors showed that the model could effectively generalize in unseen catchments, their approach does not generalize for different rainfall events.

Generative adversarial networks (GAN) (Goodfellow et al., 2014) can be used to improve CNN predictions. GANs include two convolutional-based DL structures, a generator ($\mathcal{G}$) and a discriminator (D), that compete during training. While D is trained to differentiate $\mathcal{G}$'s output from the actual data, $\mathcal{G}$ is trained to generate outputs to be classified as true data by the discriminator (Gonog and Zhou, 2019; Wang et al., 2017). This approach significantly increases the capacity to create more realistic results (Wang et al., 2017) and has been an effective method for generating high-resolution outputs (Cheng et al., 2020). Recently, Hofmann and Schüttrumpf (2021) applied a conditional generative adversarial network (cGAN), which is also CNN based, for predicting floods (floodGAN). Their approach uses spatially distributed rainfall as inputs to generate flood maps up to 10^6 times faster than physically based hydrodynamic modeling. However, the transferability of their method to other locations is a major challenge because floodGAN is domain-specific due to the lack of topographic information used as inputs.

Despite the significant advances in the state-of-art of DL, ensuring a meaningful physical connection between inputs and outputs is still a challenge (Willard et al., 2020; Swischuk et al., 2019). Merging DL with physically-based models is a promising solution to improve the accuracy of the predictions and the DL generalization capacity, as this combination can capture better the physical processes and the dynamics of an entire system (Willard et al., 2020; Palmitessa et al., 2022). However, building a hybrid physics-guided DL method with models requiring fixed input sizes, such as CNN, is difficult (Ma et al., 2021; Park et al., 2020). One alternative to the fixed input size issue of CNNs is to use multiple patches (smaller portions of the domain) to construct maps for scale-dependent problems (Martins et al., 2020). In the case of flood predictions, previous studies applied the boundary conditions (e.g., hydrograph or precipitation) for individual CNN patches (e.g., Guo et al. (2022), Lowe et al. (2021)). In these cases, the interaction between patches is lost, and a full assessment of the entire domain, such as maintaining the conservation of mass, becomes very complex. Therefore, the strategy of applying patches for flood predictions hampers the development of physics-guided DL models.

According to the literature, there is still a lack of an approach capable of generalizing flood predictions to different catchments and boundary conditions. The inability to generalize requires the retraining of the model for new areas or when local characteristics change. Training a DL model is a time-consuming process that demands extensive data. Training data is often obtained from hydrodynamic outputs and thus limits the advantage of such DL models over the traditional hydrodynamic approaches (Bentivoglio et al., 2022). In this paper, we aim to narrow this gap by developing a novel physics-guided DL method, named cGAN-Flood, based on a cGAN approach to predict flooding in unseen catchments and for different rainfall events. Our approach includes two generators that identify wet cells and calculate the water depths. The dual generator strategy enabled applying boundary conditions to the entire studied area, instead of to individual patches. Therefore, this method allowed us to enforce mass conservation for the whole catchment. We tested our approach in unseen catchments under varying rainfall conditions.

2. Methodology

The proposed deep learning modeling approach is based on the idea of distributing a target runoff volume (v_t) over a given catchment area to create a map with maximum flood depths ($\boldsymbol{D}$). The v_t can be calculated via water balance from hydrological simulations, which are traditionally faster than hydraulic models, while the flood volume distribution is performed with DL models.

2.1. Rapid flood modeling with cGAN-flood and the required input data

Rapid flood simulations with cGAN-Flood require the v_t, which is the volume to be distributed over the domain, and four spatial inputs: elevation, imperviousness, flow accumulation, and slope maps. These spatial input data have been chosen as they are often used to build hydrological models and are relevant for flood predictions (Lowe et al., 2021). cGAN-Flood is a dual generator approach that uses spatial characteristics to generate flood maps with a volume that matches v_t. A schematic representation of this hybrid physics-guided DL method is shown in Fig. 1. First, a generator ($\mathcal{G}_1$) is trained to rank which cells in the domain are likely to be flooded. The outputs of $\mathcal{G}_1$ are patches of the whole domain that are combined so that the full catchment is ranked (map $\boldsymbol{R}$). Then, a map with all wet cells ($\boldsymbol{W}$) is created by setting a threshold τ and considering all cells in $\boldsymbol{R}$ ranked above τ as wet. In other words, τ determines the magnitude of flooding, with lower values resulting in a larger number of wet cells. Finally, a second generator ($\mathcal{G}_2$) is trained to calculate the depths of wet cells in $\boldsymbol{W}$ to create $\boldsymbol{D}$ by combining multiple output patches of $\mathcal{G}_2$. The Pix2Pix method (Isola et al., 2017), which uses a conditional GAN (cGAN), was used to train the generators. In cGANs, the $\mathcal{G}$ and $\mathcal{D}$ are conditioned with additional information as inputs, instead of random noises alone like traditional GANs. In this paper, we conditioned the cGAN with local characteristics that affect flood predictions (details in Section 2.2).

The method requires finding a suitable τ, so that the total volume (v) in the final map $\boldsymbol{D}$ (which depends on $\boldsymbol{W}$) matches the volume v_t. This dual generator approach allows applying the boundary conditions (in this case, v_t) to the entire catchment and not only the area the patch where $\mathcal{G}$ is restricted to, unlike previous studies.

2.2. cGAN structure and maps construction

The Pix2Pix conditional GAN uses a generator ($\mathcal{G}$) with U-NET based ANN for image-to-image translation, i.e., the model uses noises and conditional images ($\boldsymbol{X}$) as inputs to $\mathcal{G}$ for generating an output image. Isola et al. (2017) observed that including noise as an input did not contribute to the generation of the final image as it usually happens with GANs. Therefore, we only considered images as inputs to cGAN-Flood. The training process uses a discriminator (D) that grades the output of $\mathcal{G}$ as a true or false image. $\mathcal{G}_1$ and $\mathcal{G}_2$ are trained separately, each one with a distinct discriminator (D_1 and D_2).

The $\mathcal{G}$ and D structures used in this paper are shown in Fig. 2. Each step of the downscale contains one 2D convolutional layer (conv2D) and a leaky ReLU activation function, while the step of the upscale process has one deconvolutional layer with a ReLU activation function. Finally, the generator's output layer uses hyperbolic tangent activation functions. D consists of two downscale blocks, followed by conv2D layers. Then a binary cross-entropy function is applied to the output matrix of D to classify the input image as "false" or "true". Elevation ($\boldsymbol{E}$), imperviousness ($\boldsymbol{I}$), flow accumulation ($\boldsymbol{F}$), and slope ($\boldsymbol{S}$) matrices for the whole catchment, with dimensions $m \times n$, are the required data for $\mathcal{G}_1$. The inputs for $\mathcal{G}_1(\boldsymbol{X_1})$ are four stacked submatrices (with the size of a patch) of $\boldsymbol{E}$, $\boldsymbol{I}$, $\boldsymbol{F}$, and $\boldsymbol{S}$, with dimensions $s \times s$, in which $s \leq m$ and $s \leq n$. The inputs to $\mathcal{G}_2(\boldsymbol{X_2})$ include a submatrix of the wet matrix $\boldsymbol{W}$. During the cGAN training, $\mathcal{G}_1(\boldsymbol{X_1})$ and $\mathcal{G}_2(\boldsymbol{X_2})$ are optimized to get as close as possible to ground-truth submatrixes $\boldsymbol{R}$ and $\boldsymbol{D}$, respectively. In our case, the ground-truth images are obtained from hydrodynamic

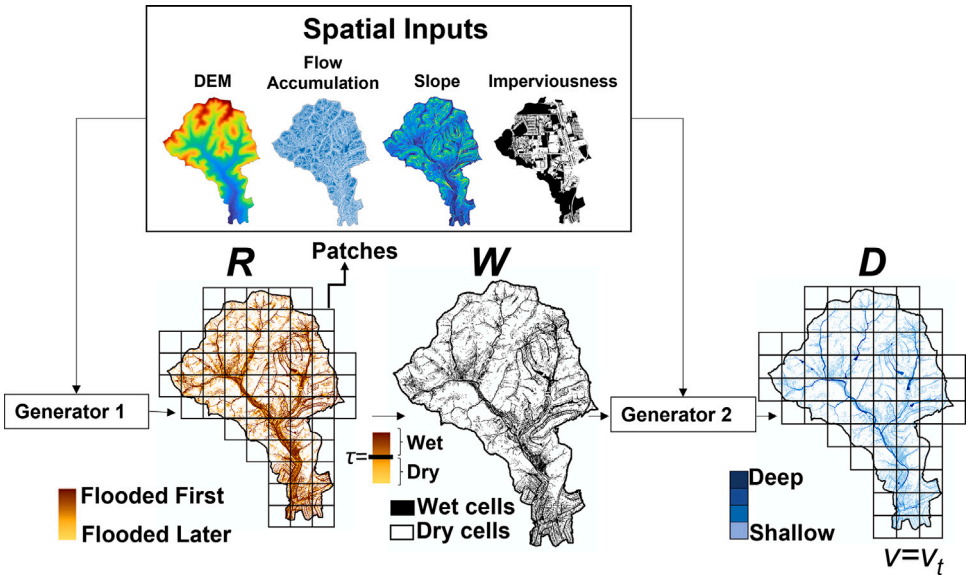

Fig. 1. Summary of the cGAN-Flood prediction process.

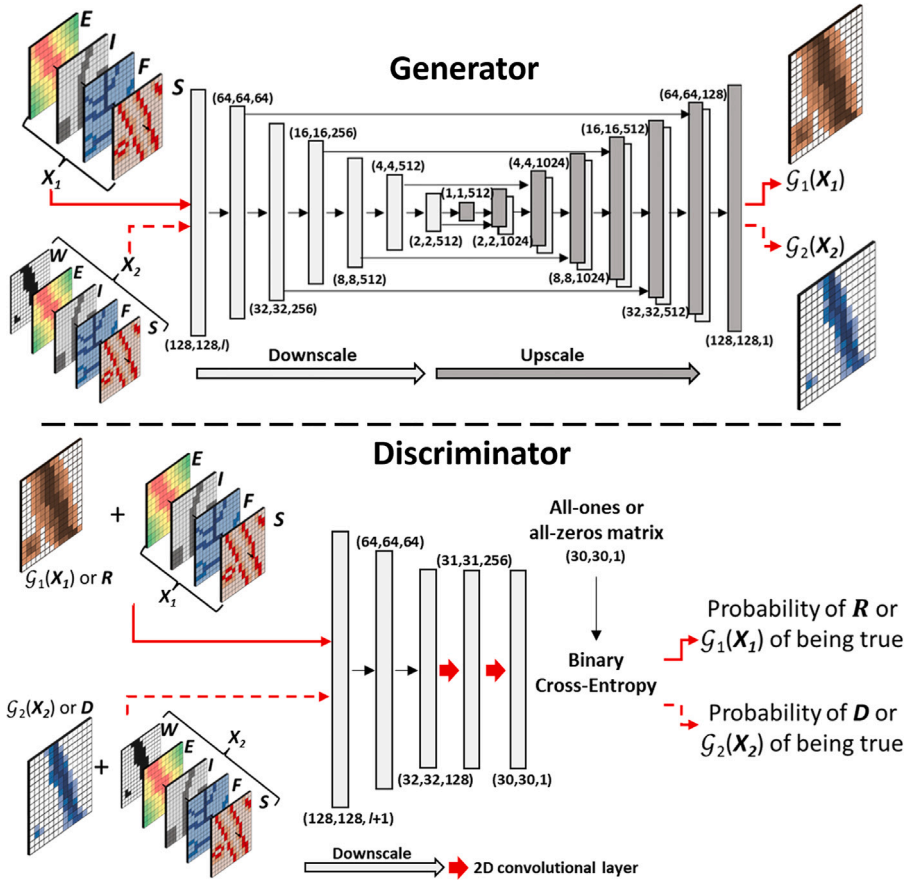

Fig. 2. Generators $\mathcal{G}_1$ and $\mathcal{G}_2$, and their respective discriminators, trained to reproduce $\boldsymbol{R}$ and $\boldsymbol{D}$ calculated from hydrodynamic modeling. The input to $\mathcal{G}_2$ includes $\boldsymbol{W}$, which is determined after thresholding $\mathcal{G}_1(\boldsymbol{X}_1)$.

models. The input of the Ds includes the same inputs of $\mathcal{G}$ plus the image to be classified, which can be either the generator's output ($\mathcal{G}_1(\boldsymbol{X_1})$ or $\mathcal{G}_2(\boldsymbol{X_2})$) or the ground-truth image. The ideal D returns a 0% probability of being a true image if fed with $\mathcal{G}(\boldsymbol{X})$, and a 100% if the input includes the ground-truth

The D of both cGANs are trained utilizing a loss function (Loss_D), which is the following:

$$\text{Loss}_D = L_{CE}(D(\boldsymbol{X}, \boldsymbol{T}), \boldsymbol{J}) + L_{CE}(D(\boldsymbol{X}, \mathcal{G}(\boldsymbol{X})), \boldsymbol{K}) \quad (1)$$

where $\boldsymbol{J}$ and $\boldsymbol{K}$ are label matrixes of ones and zeros, used to relate the true and false images, respectively. $\boldsymbol{T}$ represents the ground-truth

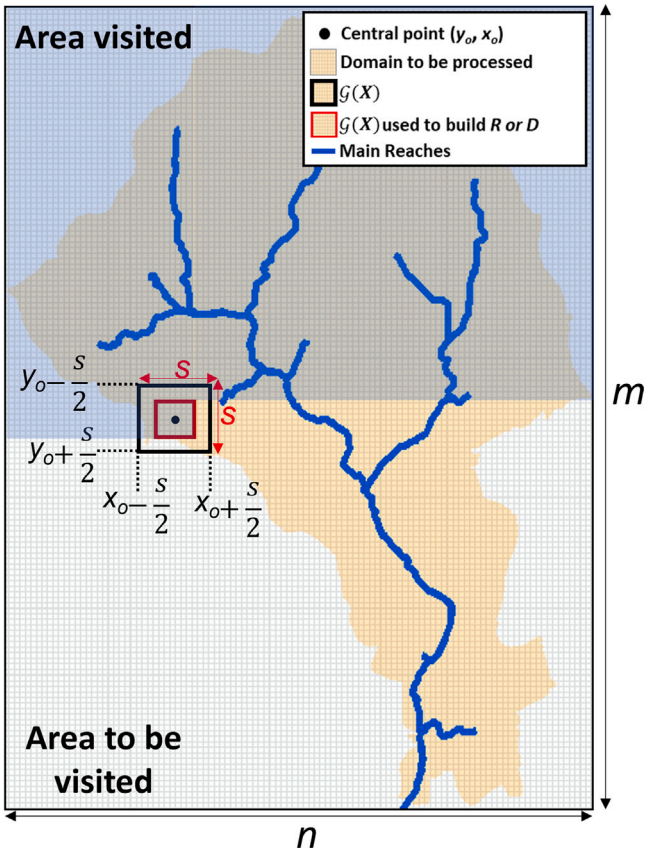

Fig. 3. Illustration of how $\mathcal{G}_1$ and $\mathcal{G}_2$ are used to build R and D maps.

image, either $\boldsymbol{R}$ or $\boldsymbol{D}$ calculated with a hydrodynamic model. L_{CE} stands for a binary cross-entropy function (Gneiting and Raftery, 2007). The $\mathcal{G}$ loss function ($\text{Loss}_{\mathcal{G}}$), used to train $\mathcal{G}_1$ and $\mathcal{G}_2$, is shown in Eq. (2). $\text{Loss}_{\mathcal{G}}$ combines the D grading with the least absolute deviation ($\mathcal{L}_1$) (Isola et al., 2017), shown in Eq. (3):

$$\text{Loss}_{\mathcal{G}} = L_{CE}(D(\boldsymbol{X}, \mathcal{G}(\boldsymbol{X})), \boldsymbol{J}) + \lambda \mathcal{L}_1 \tag{2}$$

$$\mathcal{L}_1 = \frac{1}{s^2} \sum_{i=0}^{s} \sum_{j=0}^{s} \left| \boldsymbol{T}^{i,j} - \mathcal{G}(\boldsymbol{X})^{i,j} \right| \tag{3}$$

where λ is constant.

Once the generators are trained, the construction of $\boldsymbol{R}$ and $\boldsymbol{D}$ is done by assigning $\mathcal{G}_1(\boldsymbol{X}_1)$ or $\mathcal{G}_2(\boldsymbol{X}_2)$ to the patch around a given coordinate (x_0, y_0). These coordinates slide through the domain until the whole catchment has been covered. However, preliminary analysis showed that the model uncertainties were more prominent at the borders, possibly because $\boldsymbol{E}$ contains the normalized elevations relative to the central cell (further details in the next section). Therefore, only the central portion of each patch is utilized, by excluding the 25% extremes, as depicted in Fig. 3. To discard points outside of the catchment area, we considered a mask matrix where a flood map is generated.

Since the maximum water depth significantly varies between different catchments, we considered a fixed maximum water depth threshold d_{max} to further improve the generalizability across catchments. Preliminary analyses showed that reducing the generator's predicted depth range improved the generalization potential significantly. Therefore, $\mathcal{G}_2$ computes only depths for shallow water ($d < d_{max}$) while simply identifying those considered deeper ($d \geq d_{max}$). Here, we used a $d_{max} = 0.3$ m, which is twice the depth considered a transition between concentrated shallow flows and open channel flow (USDA, 1986).

To estimate deeper water depths (≥0.3 m), we interpolated the water surface elevation (WSE) calculated at the frontier between the shallow and deep waters, assuming that the water level is linear within this region. We based this assumption on 1D hydrodynamic modeling (e.g., HEC-RAS 1D), according to which the water surface is leveled perpendicular to the flow direction with a single WSE for a given cross-section. In addition, 1D models assume that the water surface is linear between cross sections (Brunner, 2016).

Here, the final water depths are calculated by subtracting the interpolated WSE from the DEM. The whole process is shown in Fig. 4. This figure also shows that the WSE varies in space. The interpolation results of the WSE are directly dependent on the outputs of G2, which identifies the locations of deeper waters and dictates the values to be interpolated from the borders. It also should be noted that the interpolations tend to be perpendicular to the flow direction and flood propagation in channels, as higher water depths are more likely to be found along the thalweg, and the frontier between deep and shallow water tends to follow the river banks.

2.3. Data augmentation and normalization

Hydrodynamic simulations were performed to generate $\boldsymbol{R}$ and $\boldsymbol{D}$ maps used to train the generators. For a given domain under a rainfall with an increasing intensity over time, the ranking matrix $\boldsymbol{R}$ can be calculated as:

$$\boldsymbol{R} = \sum_{t=1}^{T} \boldsymbol{W}_t, \quad \boldsymbol{W}_t^{i,j} = \begin{cases} 1 \; if \; \boldsymbol{D}_t^{i,j} > 0 \\ 0 \; \text{elsewhere} \end{cases} \tag{4}$$

where $\boldsymbol{D}_t$ is the water depth map at time t, $\boldsymbol{W}_t$ is a binary map with unitary values for wet cells at time t, and T is the number of maps. The higher the $\boldsymbol{R}$ value for a given cell, the earliest it started to flood. In other words, cells with a high $\boldsymbol{R}$ value start to flood with small rainfall intensity, while the cells with lower $\boldsymbol{R}$ would only flood with more intense precipitations.

We collected the samples throughout the catchment by sliding a center point of coordinates (x_0, y_0) inside the catchment and selecting the patch of dimension $s \times s$ ($\mathcal{G}(\boldsymbol{X})$ as illustrated in Fig. 3). We employed data augmentation strategies to increase the generalizability of the proposed model. We augmented the number of training samples per catchment by multiplying the elevation maps $\boldsymbol{E}$ by a factor k_{tm}. A value lower than 1 generates flatter terrains and shallower channels, which are more likely to be flooded. On the other hand, a k_{tm} larger than 1 increases the slopes and deepens the channels, reducing the flood extension. In addition to the terrain modification, further data augmentation was performed by rotating and mirroring the patches. During the data collection, the central point slides $s/2$ plus a random number ranging from 0 to $s/2$, until the whole domain has been visited. This randomization helped to increase the variability of the training dataset, as the same points are not always selected for different $\boldsymbol{D}$ maps or k_{tm}.

Finally, we applied min–max normalization on the $\boldsymbol{E}$, $\boldsymbol{F}$, $\boldsymbol{S}$, $\boldsymbol{R}$, and $\boldsymbol{D}$ to facilitate the training process and improve its generalizability. The normalized values of these matrixes ranged from −1 to 1. The wet cell–matrix $\boldsymbol{W}$ and the imperviousness $\boldsymbol{I}$ (one and zero representing that a cell is fully impervious and pervious) were not normalized.

The maximum values used to normalize $\boldsymbol{R}$ and $\boldsymbol{D}$ were T (see Eq. (4)) and d_{max} (0.3 m), respectively. The $\boldsymbol{E}$ submatrix was normalized between −1 and 1, with the elevation relative to the central coordinate of the patch. This way, the central element of this input is always zero, while −1 and 1 represent the lowest and highest elevation cells. Fixating an origin helps the training process and facilitates the DL to recognize patterns (Gondhi and Kour, 2017).

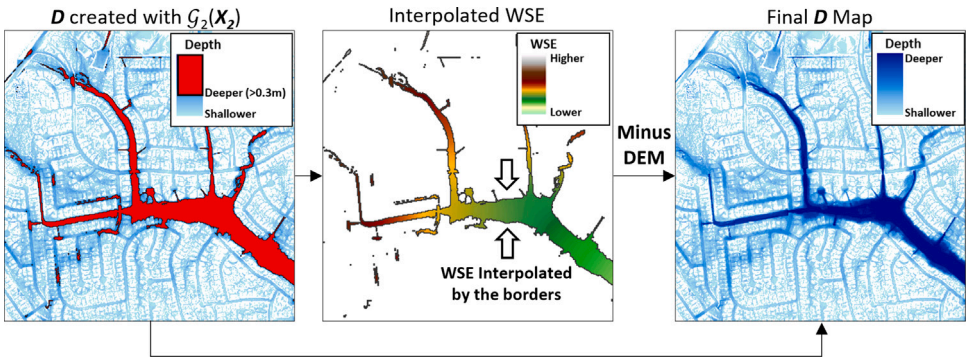

Fig. 4. Calculation of the water depth map $\boldsymbol{D}$ from the cGAN. The outputs of $\mathcal{G}_2$ which are predicted to be higher than d_{max} are interpolated assuming the water is leveled. The final water depth map D is then evaluated by subtracting the DEM from the interpolated WSE.

2.4. Estimating the flood magnitude with τ

One key component of our model is the ability to match the total flood runoff volume v_t, estimated by a hydrological model for a given catchment, thus providing a link with physical-based models and enforcing mass conservation. For this purpose, we iteratively modify the threshold τ, used in the determination of the wet cell map $\boldsymbol{W}$, until the relative error between the predicted volume v (Eq. (5)) and v_t is smaller than a pre-defined tolerance ξ_{tol}. The volume v and the relative error ε are calculated as:

$$v = A \sum_{ij} d^{ij} \tag{5}$$

$$\varepsilon = \frac{|v - v_t|}{v_t} \tag{6}$$

where A is the area of a single cell and d_{ij} is the water depth at row i and column j in $\boldsymbol{D}$. Decreasing the value of τ increases the flooded area and, consequentially, v. Therefore, the error, as a function of τ, is unimodal. This characteristic enables applying the Golden Search optimizing algorithm (Kiefer, 1953), an efficient method to single variable problems. Details of the Golden Search can be seen in Supplementary Material 1.

It is worth noting that our approach aims to distribute a v_t that corresponds to maximum flood depths in the entire domain, as shown in Eq. (6), which differs from instantaneous volume. The maximum depth is calculated with the depth peak of the whole rainfall event. Due to the flood wave propagation, the arrival time of maximum depth varies across cells. At the time of maximum instantaneous volume, not all cells in the domain present maximum water depth. Therefore, the maximum volume calculated with instantaneous depths is smaller than the volume calculated with a map of maximum depths.

In this paper, we assume that flood wave attenuation can be irrelevant at short distances. Therefore, the time of maximum flood depth of nearby cells is closely the same. In this case, the difference between maximum instantaneous volume and volume computed with maximum depths can be neglectable. The methods and assumptions for calculating v_t via hydrological modeling are depicted in Section 3.3.

2.5. Evaluation metrics

The root squared mean error (RSME) and the coefficient of determination (R^2) were used to evaluate the model performance. We also used the metrics hit rate (HR), false alarm rate (FAR), and critical success index (CSI) to evaluate the spatial performance of the final $\boldsymbol{D}$ map. HR, FAR, and CSI can be calculated with the following equations:

$$\mathrm{HR} = \frac{\mathrm{Hits}}{\mathrm{Misses}} \tag{7}$$

$$\mathrm{FAR} = \frac{\mathrm{Hits}}{\mathrm{Hits} + \mathrm{False\ Alarms}} \tag{8}$$

$$\mathrm{CSI} = \frac{\mathrm{Hits}}{\mathrm{Hits} + \mathrm{False\ Alarms} + \mathrm{Misses}} \tag{9}$$

where hits are the cells flooded in both models, false alarms are the cells flooded only with cGAN-Flood, and misses are the cells flooded only predicted by the hydrodynamic model. We considered a threshold of 0.01 m to distinguish dry and wet cells, a value used in other flood studies (Brown et al., 2007; Shi et al., 2021). Preliminary analysis (not shown) indicated that cells with flood depths between 0.01 m and 0.05 m accounted for an average of 8% of the total runoff volume in training areas. This is significant to evaluate cGAN-Flood as this method aims to match the total volume of $\boldsymbol{D}$ with v_t. In addition, accounting for shallow depths is relevant to understand the outputs of $\mathcal{G}_1$, as it computes wet cells for any depths above 0. Larger thresholds prevent the evaluation of the $\mathcal{G}_1$ performance in identifying the wet cells. To evaluate the performance of cGAN-Flood for cells with deeper flood levels, and the performance of $\mathcal{G}_2$, we also included $\mathrm{CSI}_{0.05}$ and $\mathrm{CSI}_{0.3}$ in our analysis to consider water depths higher than 0.05 m and 0.3 m, similarly to Lowe et al. (2021).

2.6. cGAN training and validation

Both training processes of $\mathcal{G}_1/D_1$ and $\mathcal{G}_2/D_2$ were performed with 200 epochs and a batch size of 128. The Adam optimizer (Kingma and Ba, 2014) was used with a learning rate of 0.005. The λ for $\mathcal{G}_1/D_1$ and $\mathcal{G}_2/D_2$ were 5 and 40, respectively, so that to the discriminator and MAE (in Eq. (2)) had approximately the same contribution to the total $\mathrm{Loss}_{\mathcal{G}}$ at approximately epoch 20, when the losses started to get stable. For each epoch, the parameters of $\mathcal{G}_1$ and $\mathcal{G}_2$ were saved so they could be used during validation.

For the model validation, we evaluated the capacity of $\mathcal{G}_1$ and $\mathcal{G}_2$ to generate full maps of the entire catchment, instead of computing the loss at each epoch during the training phase for each individual patches (Fig. 5). Step 1 evaluated the $\mathcal{G}_2$ and WSE interpolation (shown in Fig. 4) in reproducing the flood depths calculated with hydrodynamic models. During this validation, the $\boldsymbol{W}$ map, to be used as input to $\mathcal{G}_2$, was determined directly from hydrodynamic outputs to compare depths for the same wet cells. Then, $\boldsymbol{D}$ maps were generated with $\mathcal{G}_2$ parameters saved for each of the 200 epochs used for training. These maps were evaluated with $\mathrm{CSI}_{0.3}$, $\mathrm{CSI}_{0.05}$, and RSME metrics to identify at what epoch the parameters of $\mathcal{G}_2$ presented the most accurate flood depths map ($\boldsymbol{D}$). Since we used flood maps of the entire domain, we also validated the interaction between patches, instead of isolated $\mathcal{G}_2(\boldsymbol{X}_2)$. The interaction between patches is important because it can affect the WSE in deep areas, which is estimated via linear interpolation.

In Step 2, $\mathcal{G}_1$ was validated. Here, we generated $\boldsymbol{R}$ maps with $\mathcal{G}_1$ saved for all 200 epochs. The golden search single variable optimization method was performed for each of the 200 $\boldsymbol{R}$ maps to identify τ and what $\mathcal{G}_1$ model generates the most accurate $\boldsymbol{D}$. The validated $\mathcal{G}_2$ was

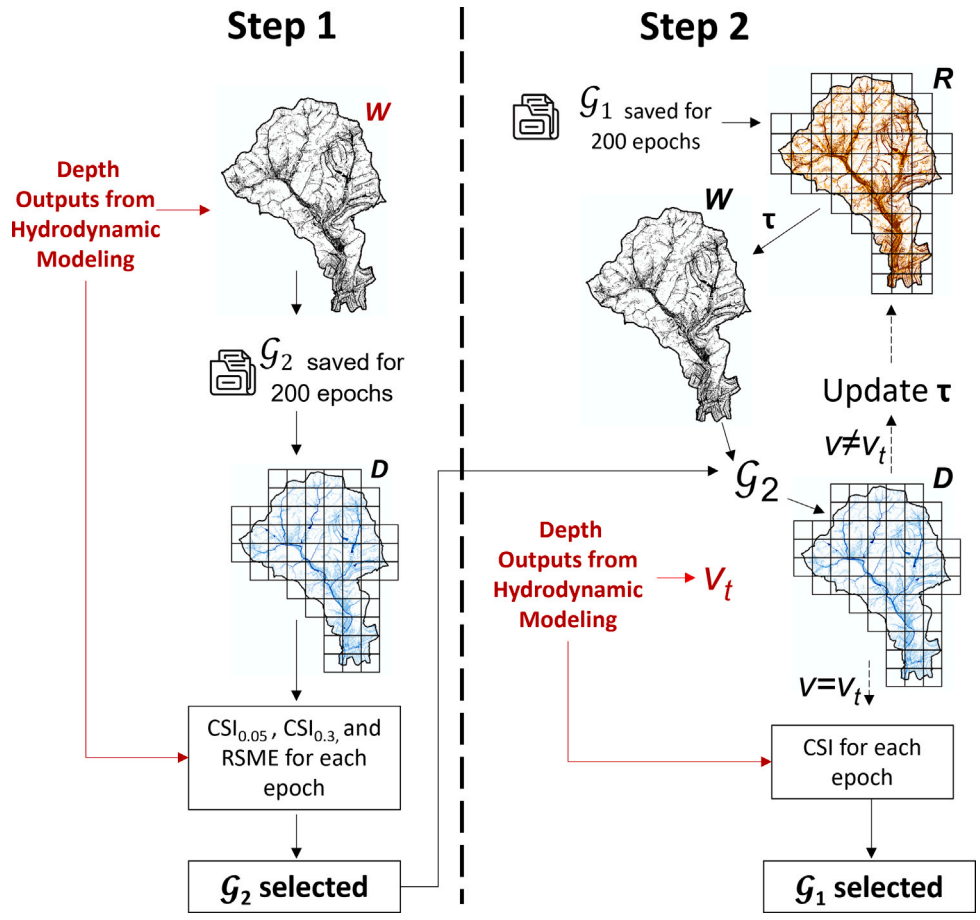

Fig. 5. Validation flowchart of $\mathcal{G}_2$ and $\mathcal{G}_1$. The parameters of both generators are saved for each epoch, which will later be used to build $\boldsymbol{D}$ and $\boldsymbol{R}$ for the entire catchment for comparison with hydrodynamic outputs.

selected and was used during the golden search for validating $\mathcal{G}_1$. Therefore, the validation of $\mathcal{G}_1$ can be interpreted as a validation of the entire cGAN-Flood method. Since $\boldsymbol{W}$ indicates the cells that are flooded for any depth above zero, the metric CSI was used to validate $\mathcal{G}_2$, which includes shallower depths.

3. Studied area and flood simulations

The training (TR), validation (VA) and testing (TE) datasets consisted of ten, two, and five sub-catchments in San Antonio, Texas (Fig. 6a). These sub-catchments are located in the Upper San Antonio and Leon Creek watersheds. The training sub-catchments are mostly clustered at the eastern portion of Leon Creek. Testing areas TE1, TE2, and TE3 are closer to the training areas, while VA1, VA2, TE4, and TE5 are outside the training areas cluster. The areas selected for testing and validation include catchments with different sizes, average cell slopes, and percentage of imperviousness, allowing cGAN-Flood to evaluate different spatial configurations. The characteristics of all areas are presented in Table 1.

HEC-RAS 2D models were created for each training and testing area with 3 m grid resolution, as recommended by Federal Emergency Management Agency, 2020, along with DEM and imperviousness maps of the same resolutions, both provided by the San Antonio River Authority (SARA) (E. Cavazos, personal communication, 2018). A required HEC-RAS 2D model parameter is the Manning's roughness coefficient for pervious and impervious cells. We selected n of 0.03 $m^{1/3}$ s^{-1} for impervious cells, a typical value that represents the roughness of concrete (Brunner, 2016). For pervious areas, we selected n = 0.1 $m^{1/3}$ s^{-1} which is in the recommended range for different pervious covers (e.g., bare field, grass, row crops, sparsely vegetated surfaces, and others) (Engman, 1986). The input rainfall in all catchments was assumed to be spatially uniform because the catchments are relatively small.

3.1. Training and validation

We used a synthetic linear rainfall hyetograph in the training areas (Fig. 6b). The rainfall intensity increases smoothly over time to minimize differences in water levels due to flood wave propagation. This rainfall has a 1-minute time step and a total duration of 100 min, with intensity varying from 7 to 700 mm/h, approximately twice the maximum rainfall intensity of the 100-year storm. We used this extreme precipitation intensity so that the flooded areas to be predicted are likely to be smaller than what the DL has seen during the training phase. Flood maps for every training area were acquired every 1 min to compute $\boldsymbol{R}$ with Eq. (4) and, therefore, T = 100. During validation of $\mathcal{G}_2$, $\boldsymbol{D}$ maps were collected every 5 min. This approach allows $\mathcal{G}_2$ to learn water depths from lighter and heavier rainfall intensity (t closer to 0 and T). We considered additional simulations obtained by terrain augmentation for the training dataset, as described in Section 2.2, using k_{tm} values of 0.8, 0.9, 1.1, and 1.2.

For validating both generators, we simulated maximum depths in the areas VA1 and VA2 using the 25- and 100-year storms with a duration of 3 h (Fig. 6c). The alternating blocks method was used to obtain the rainfall temporal distribution with a 5-minute resolution. We evaluated the $\boldsymbol{D}$ maps with the averages between the CSI and RSME metrics computed for both design storms and VA locations.

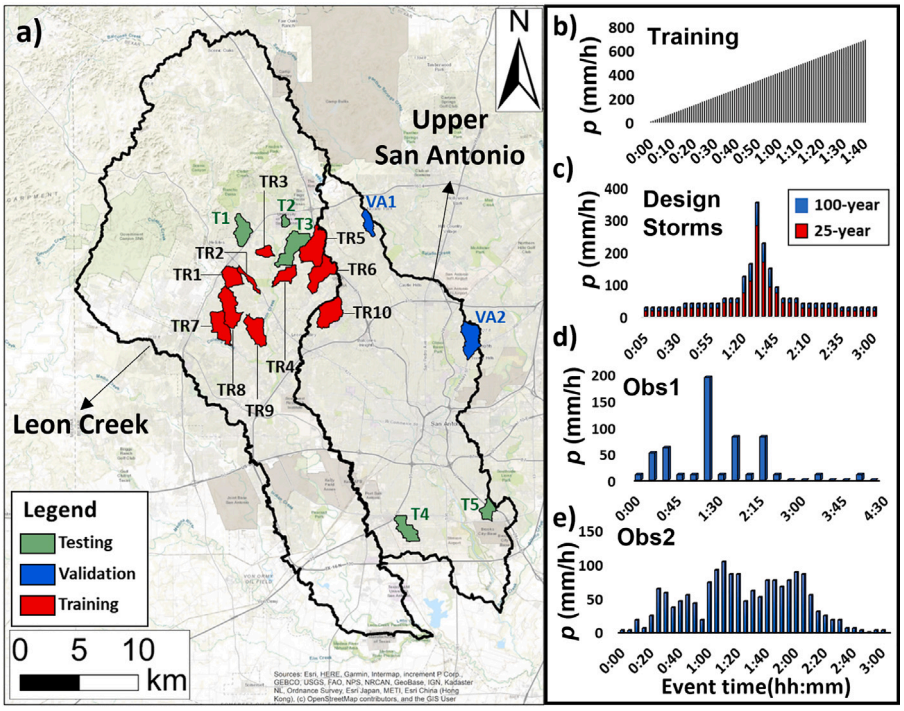

Fig. 6. Training, validation, and testing areas and rainfall conditions: (a) the different catchments, (b) the precipitation used in the training areas, (c) the design storms used for validation and testing, and two observed events (d) and (e) used for testing.

Table 1
Characteristics of the testing, validation, and training areas.

Location	Area (ha)	Number of cells ($\times 10^3$)	Average cell slope (°)	Impervious area (%)
TR1	226.9	252.1	17.5	44.7
TR2	84.8	94.2	14.7	43.1
TR3	99.8	110.9	14.7	46.9
TR4	187.6	208.5	10.9	46.8
TR5	463.8	515.3	13.3	58.9
TR6	371.5	412.8	14.5	36.1
TR7	309.6	344	12.7	48.4
TR8	383.6	426.3	14.7	46
TR9	307.3	341.4	15.4	38.4
TR10	378.1	420.1	17.1	57.2
VA1	134.6	149.5	13.1	29.9
VA2	341.9	379.8	13.7	48.9
TE1	258.9	287.7	14.7	33.4
TE2	54.4	60.4	13.8	63.8
TE3	406.5	451.6	12.5	43.2
TE4	251	277.8	8.2	45.5
TE5	146.9	163.2	12.2	54.1

3.2. Testing

We tested cGAN-Flood's capacity to reproduce HEC-RAS depth maps in the TE catchments, which were not included in the training or validation phases. Similarly to the validation, we calculated v_t from HEC-RAS flood maps (Eq. (5)) to compare maps with the same flood volume. Therefore, we excluded the uncertainties propagated from the hydrological model to analyze the errors that are solely produced by the cGAN-Flood method.

Four rainfall events were used, including the two design storms and two observed events (Fig. 6d and e). Observed event 1 (Obs1) and observed event 2 (Obs2) were measured at the San Antonio International Airport (NOAA Station ID: GHCND: USW00012921; Lat.: 29.54429°, Long.: −98.48395°) and in the TE2 catchment, respectively, with time resolutions of 15 and 5 min. Both observed events have a total of 130 mm in 3 h. Obs1 has a distinct rainfall peak, while the rainfall volume of Obs2 is more evenly distributed over time.

In addition to the different rainfall events, we also tested the model's capability to predict flood depths with different scenarios of Manning's n. These scenarios were simulated on HEC-RAS with Manning's n as twice (scenario M1) and a half (scenario M2) of the values used for training the cGAN (0.015 and 0.06 for the pervious, and 0.05 and 0.02 $m^{-1/3}$ s^{-1} for the impervious grids, respectively). We used the 25- and 100-year storms to evaluate the influence of surface roughness for different rainfall magnitudes. The surface roughness reflects how quickly the runoff flows out of the catchment and, consequently, the v_t to be distributed. For instance, a higher Manning's n results in a higher time of concentration and higher v_t for the same rainfall, due to a higher flood accumulation in the cells of the domain. The goal of this experiment is to evaluate how the changes in v_t, due to different surface roughness, could affect the performance of the cGAN-Flood model.

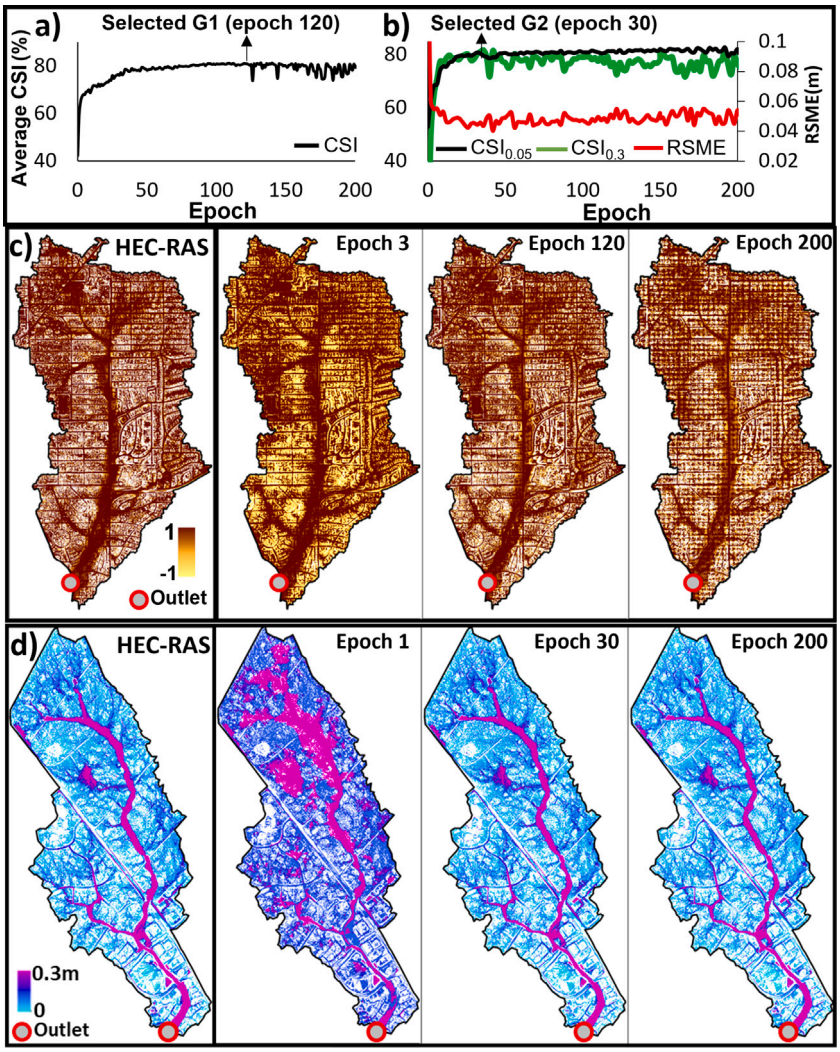

Fig. 7. Validation metrics of $\mathcal{G}_1$ and $\mathcal{G}_2$ (a) and examples R and D maps created for VA2 (b) and VA1 (c), for the 100-year storm, presented for initial epochs, the best epoch, and the last training epoch.

3.3. Hydrological modeling and cellular automata benchmark simulations

In this study, we used SWMM models that were previously developed by Zarezadeh (2017) for areas TE1 and TE2. The TE1 and TE2 models are subdivided into 34 and 163 sub-catchments, respectively. The computation of v_t was performed by summing the maximum instantaneous volume of each of the smaller components of the SWMM models separately, instead of a water balance applied to the whole catchment. This way, we reduce the error associated with the approximation that we assumed between the instantaneous maximum volume and the volume corresponding to maximum depth. Details of how v_t was calculated with SWMM can be seen in Supplementary Material 2. The longest segment in our SWMM models is approximately 770 m, corresponding to an open channel conduit link in area TE1. In supplementary material 3, we show that the effects of flood wave propagation could be neglected in that segment.

For these analyses, effective precipitation was inserted on HEC-RAS and WCA2D to account for infiltration. HEC-RAS 6.0 (Brunner, 2016) and newer versions have the option to compute infiltration for 2D simulations. However, we used effective precipitation as no infiltration was included in the HEC-RAS models used for training. Including infiltration would increase the cGAN-Flood complexity, as different infiltration capacities would have to be tested to account for different soil types. For comparison, WCA2D does not account for infiltration. The WCA2D only uses one global Manning's n for the whole domain, which was set as 0.025 and 0.055 $m^{-1/3}$ s^{-1} for areas TE1 and TE2, respectively. These values were obtained after calibration that ensured maximum $CSI_{0.03}$ for the 100-year storm when compared to HEC-RAS. We selected $CSI_{0.3}$, because 0.3 m was the minimum depth used to evaluate the WCA2D model, according to Guidolin et al. (2016). In addition, a 0.01s minimum time step was used, which showed no signs of instability. The two design storms and two observed events used for testing cGAN-Flood, were also modeled with WCA2D.

3.4. Hardware specifications

We performed all cGAN-Flood and WCA2D simulations on an NVIDIA GeForce RTX 2060 SUPER GPU (1650 MHz and 16 GB of RAM). HEC-RAS model does not support GPU; therefore, the hydrodynamic simulations were done using an i7 10700 CPU with 2.9 GHz.

4. Results and discussion

4.1. cGAN-Flood training and validation

The evolution of $\boldsymbol{R}$ and $\boldsymbol{D}$ maps created with $\mathcal{G}_1$ and $\mathcal{G}_2$ during the validation is illustrated in Fig. 7. The results show that the average CSI metric for $\boldsymbol{R}$ and $\boldsymbol{D}$ increased rapidly for the first epochs and then

Table 2
Performance metrics of cGAN-Flood for all testing areas.

Storm	Testing area	v_t (1000 m^3)	RSME (m)	R^2	HR (%)	FAR (%)	CSI (%)	$CSI_{0.05}$ (%)	$CSI_{0.3}$ (%)
25-year	TE1	285	0.09	0.90	94.4	14.8	81.1	72.7	69.0
	TE2	37	0.08	0.85	88.2	11.7	78.9	71.1	68.3
	TE3	486	0.13	0.89	93.2	15.3	79.8	72.3	68.0
	TE4	223	0.09	0.88	94.1	10.9	84.4	74.2	67.1
	TE5	119	0.17	0.84	83.9	11.9	75.3	72.1	67.9
100-year	TE1	339	0.10	0.90	94.6	13.3	82.6	74.6	70.9
	TE2	46	0.09	0.86	86.8	8.5	80.4	72.9	67.8
	TE3	595	0.14	0.87	93.5	13.4	81.7	74.4	67.5
	TE4	270	0.10	0.86	94.2	9.4	85.8	76.3	65.7
	TE5	143	0.16	0.78	84.3	10.3	76.9	73.7	69.3
Obs1	TE1	242	0.09	0.88	94.5	17.5	78.8	70.7	65.4
	TE2	34	0.08	0.84	89.0	14.8	77.1	69.3	67.1
	TE3	443	0.13	0.84	93.8	18.3	77.5	69.7	67.4
	TE4	201	0.09	0.84	94.0	12.4	83.0	73.3	66.9
	TE5	104	0.18	0.77	91.5	15.2	78.6	69.7	64.7
Obs2	TE1	196	0.10	0.82	92.7	23.1	72.5	63.9	60.0
	TE2	28	0.09	0.78	89.9	21.9	71.8	64.5	60.9
	TE3	412	0.17	0.72	94.8	28.3	69.0	58.8	61.0
	TE4	179	0.11	0.74	95.3	19.9	77.0	65.5	67.2
	TE5	82	0.18	0.69	92.4	24.4	71.2	61.1	58.7

stabilized (Fig. 7a and b). We selected the parameters saved at epoch 120 for $\mathcal{G}_1$, as CSI starts to get unstable beyond this point due to overfitting. Fig. 7c shows the $\boldsymbol{R}$ maps obtained by HEC-RAS simulations, and generated by $\mathcal{G}_2$ at Epochs 3, 120, and 200 in the VA2 area. In epoch 3, for example, $\mathcal{G}_1$ has already started to understand flooding features in VA2 and indicates the areas most susceptible to flooding. On the other hand, $\boldsymbol{R}$ shows signs of overfitting in epoch 200. In comparison to $\mathcal{G}_1$, $\mathcal{G}_2$ converged faster, as Fig. 7b shows. The highest $CSI_{0.3}$ was observed at epoch 30 and, beyond this point, RSME started to increase slowly. Fig. 7d illustrates $\mathcal{G}_2$ accuracy for water depths in comparison to HEC-RAS in epoch 30 for the 100-year storm.

4.2. cGAN-Flood testing

In this section, we present the results of the model performance in the testing areas. Here, cGAN-Flood distributes a v_t calculated from the HEC-RAS simulation. Therefore, the errors presented in this section are solely related to the cGAN-Flood methods and do not include uncertainties in flood volume propagated by the hydrological model. The model's performance was satisfactory for all storm events and testing areas, as shown in Table 2. Among the four tested rainfall storms, the cGAN-Flood obtained better performance for the 100-year storm as shown by all the goodness-of-fit metrics, with the exception of the RSME. This behavior occurred because the RSME is more sensitive to larger storms than the other metrics and the 100-year storm resulted in the highest flooding and water depths. R^2 values were above 0.7 for all testing cases, except in TE5 for Obs2. HR values range from 83.9 to 95.3% with relatively low variance across different events. The FAR ranged from 8.5 to 24.4%, with higher values (worst performances) found for the Obs2 event. The higher FAR shows that $\mathcal{G}_1$ tends to overestimate the number of wet cells for events with smaller rainfall intensity. CSI metric was above 70% for all catchments, with lower values for Obs2 event due to higher FAR. The $CSI_{0.05}$ and $CSI_{0.3}$ presented low values because errors from $\mathcal{G}_1$, which identified a larger number of wet cells in comparison to the ground truth, were propagated to $\mathcal{G}_2$ predictions.

Our results suggest that cGAN-Flood can generalize across different watersheds and rainfall forcings. For instance, cGAN-flood was capable of predicting floods for the Obs2 and 100-year storms in areas TE2 and TE3 (Fig. 8), regardless of the significant differences in the catchment area and rainfall magnitude. TE3 is approximately 7 times larger than TE2. In addition, cGAN-flood distributed a v_t more than 20 times larger in area TE3 for the 100-year storm, when compared to TE2 for the Obs2 event. The errors in depth are mainly within −0.1 and 0.1 m. In the area TE3, cGAN-Flood underestimated water depths in the main channel, with errors ranging from −0.3 to −0.1 m. Since cGAN-Flood matches volume with HEC-RAS in these analyses, there was a more significant frequency of errors up to 0.1 m to compensate for the underestimations in the main channels. In area TE2, cGAN-Flood overestimated depths in parts of the channel, but overall there was a balance between the positive and negative errors.

The results indicate that cGAN-Flood's performance for deeper water depths was lower than for shallow locations, as indicated by $CSI_{0.3}$ values. One reason is that the deeper flooding gets restricted to smaller portions of the domain. We noted that misses or false alarms significantly affected CSI, especially for the Obs2 event that generated smaller flooded areas due to a smaller rainfall peak. This pattern in performance is also observed with U-Flood (Lowe et al., 2021). Fig. 9 compares the flooded areas with $d > 0.3$ m in TE4 and TE5 for the 100-year storm and Obs2 events. In area TE4, both HEC-RAS and cGAN-Flood predicted many flooded streets for both events. The reason is that this area has a smaller average slope and tends to concentrate runoff more easily. On the other hand, flooding in TE5 is more restricted and cGAN-Flood tends to underestimate flooded areas. Despite these errors, cGAN-Flood was able to capture the flooded cells in this testing area.

4.3. Sensitivity of cGAN-Flood to different Manning' coefficients

The model's sensitivity was tested for different Manning's roughness values (Table 3). Here two scenarios were evaluated: M1 (lower Manning's n) and M2 (higher Manning's n). Both FAR and HR decrease for the M1 scenario when compared to Manning's n used for training. By increasing the surface roughness (M2 scenario), surface runoff flows more slowly and accumulates within the sub-catchments. As a result, the flooded area increases, which reduces the false alarms and increases the misses. With the increase in the number of wet cells in HEC-RAS outputs, cGAN-Flood had an accuracy improvement because some false alarms became hits. On the other hand, some of the dry cells become wet, which increases the number of misses. In the case of the M1 scenario, the number of flooded cells decreases with high runoff velocities. Therefore, both FAR and HR values tends to decrease.

Despite the decrease in HR for higher Manning's n, the overall performance of cGAN-Flood improved for larger flooded areas. The CSI, $CSI_{0.05}$, and $CSI_{0.3}$ increased for both 25- and 100-year storms in the M1 scenario compared to the original model. On the other hand, the performance of the cGAN-Flood decreased in most cases for M2. The worst performance was also observed for area TE5. The difference in flooded area predictions for lower and higher n values is illustrated in Fig. 10. The number of false alarms is significantly lower

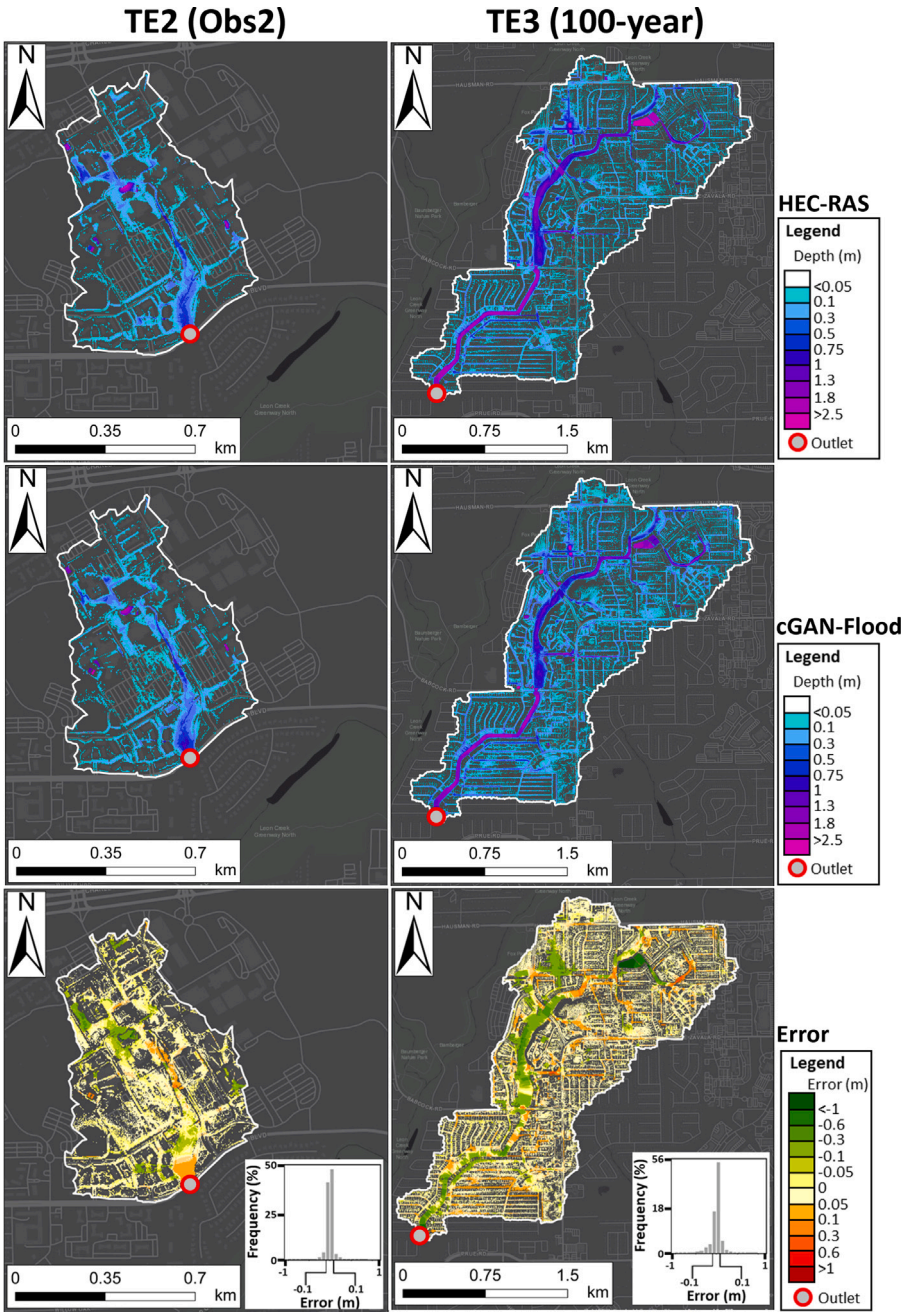

Fig. 8. Comparison of flood depths predicted with cGAN-Flood and HEC-RAS for areas TE2 and TE3.

for the M1 scenario (Fig. 10a, b) compared to M2 (10c and d). The large number of false alarm rates in M1 originated from $\boldsymbol{W}$ maps, as the cGAN-Flood also overestimates flooded areas for smaller depths. Therefore, the distribution of the flooded cells was miscalculated by $\mathcal{G}_1$ and propagated to $\mathcal{G}_2$ predictions. However, in general, the metrics show a satisfactory performance of the cGAN-Flood for different surface roughness scenarios, as shown in Table 3.

4.4. Coupling hydrological modeling with cGAN-Flood and comparison to WCA2D

In this section, we calculated v_t with a SWMM model to study the performance of cGAN-Flood when combined with a hydrological model. Table 4 reports the total flood volume v_t for the maximum depth map computed with HEC-RAS, WCA2D, and SWMM in areas TE1 and TE2. The results show that v_t was generally underestimated with SWMM, except for event Obs2 in TE1. The most significant error in volume was registered for Obs2 in TE2, which was 25.9% lower than HEC-RAS. One possible explanation is that the depression storage in the terrain is better captured with a fully-distributed model, which can be underestimated with the SWMM model.

The evaluation metrics for cGAN-Flood and WCA2D are shown in Table 5. Despite underestimating v_t, the volumes are similar to those computed with HEC-RAS and did not affect the overall performance of cGAN-Flood when calculating v_t from HEC-RAS maps (see Table 2). The exception was TE2 for Obs2, where the error in v_t was significant (see Table 4), affecting the prediction of the maximum flood. The $CSI_{0.3}$, in this case, was as low as 41.1%. Despite SWMM predicting a better v_t than WCA2D , the performance of cGAN-Flood was similar to the WCA2D model. The RSME values for cGAN-Flood in area TE1 were close to WCA2D and higher in area TE2. Overall, WCA2D overestimated depths for very shallow waters (between 0.01 and 0.05 m), which

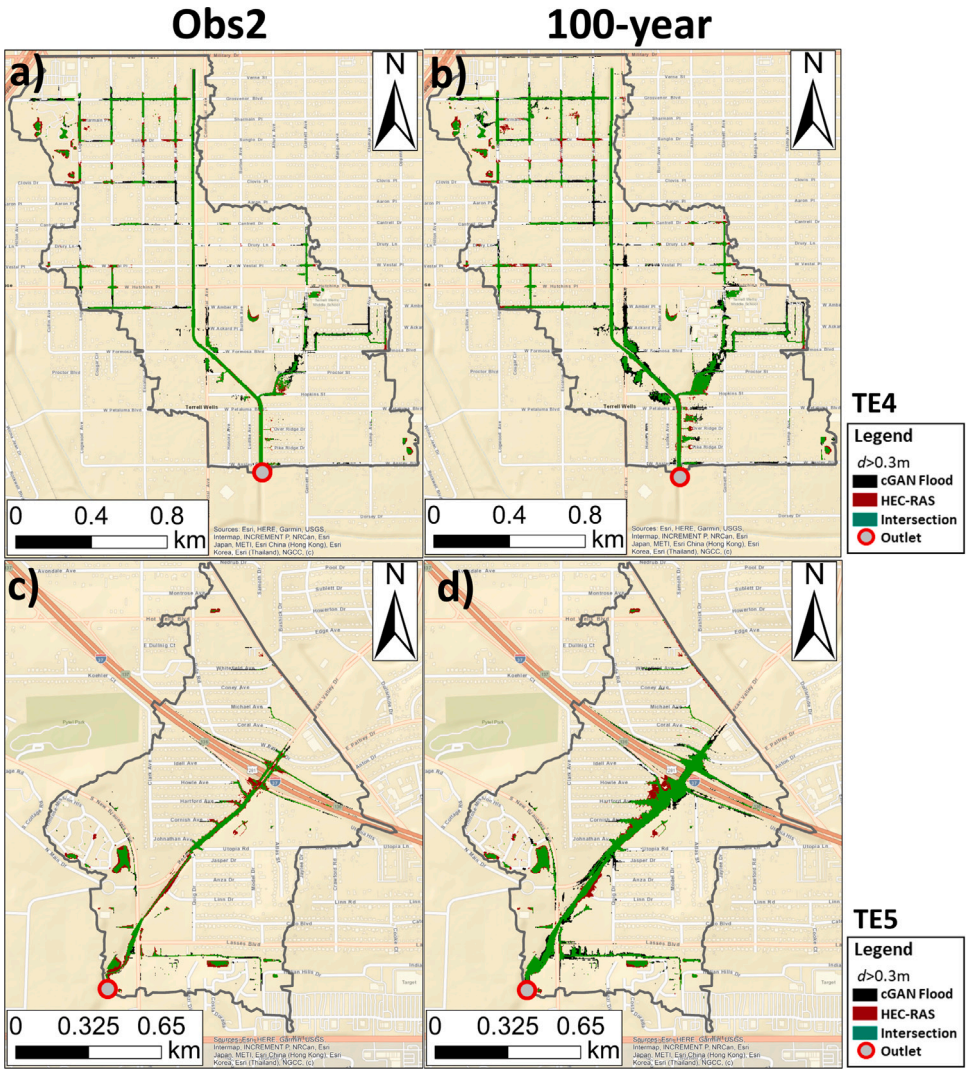

Fig. 9. Flooded areas with depth above 0.3 m for Obs2 (left) and 100-year (right) events for areas TE4 (a and b) and TE5 (c and d).

Table 3
Performance metrics of cGAN-Flood compared to HEC-RAS under different Manning's *n* scenarios.

Manning scenario	Storm	Testing area	v_t (1000 m^3)	RSME (m)	R^2	HR (%)	FAR (%)	CSI (%)	$CSI_{0.05}$ (%)	$CSI_{0.3}$ (%)
M1	25-year	TE1	301	0.087	0.90	92.2	11.4	82.5	74.7	70.7
		TE2	58	0.104	0.85	85.5	4.6	82.1	74.3	71.5
		TE3	593	0.119	0.89	92.3	8.5	85.0	79.0	69.5
		TE4	272	0.1	0.88	93.1	6.2	87.7	78.7	66.2
		TE5	154	0.152	0.84	90.1	7.3	84.1	77.9	69.1
	100-year	TE1	353	0.096	0.92	93.8	11.2	83.9	75.8	72.0
		TE2	58	0.104	0.85	85.5	4.6	82.1	74.3	71.5
		TE3	730	0.14	0.90	93.9	9.8	85.2	74.4	62.5
		TE4	332	0.089	0.93	92.9	5.2	88.4	80.2	66.1
		TE5	186	0.136	0.88	90.6	6.9	84.9	78.6	68.8
M2	25-year	TE1	288	0.093	0.90	93.3	15.2	80.0	72.4	68.0
		TE2	32	0.086	0.81	87.1	15.9	74.7	66.9	63.4
		TE3	412	0.135	0.79	94.4	20.2	76.2	67.8	65.2
		TE4	182	0.098	0.74	94.9	15.2	81.2	69.9	68.1
		TE5	97	0.175	0.70	91.1	17.9	76.0	67.2	61.6
	100-year	TE1	333	0.1	0.91	95.1	15.6	80.8	73.0	69.3
		TE2	38	0.085	0.84	87.8	14.0	76.9	67.9	66.0
		TE3	494	0.151	0.80	94.2	20.1	76.1	67.8	65.2
		TE4	212	0.093	0.83	94.7	14.2	81.9	71.4	67.9
		TE5	111	0.172	0.73	92.9	17.6	77.5	68.2	62.1

compromised the overall performance of models, as indicated by the CSI metric. However, $CSI_{0.05}$ was, in general, higher for WCA2D than for cGAN-Flood. The results show that cGAN-Flood performed better for very shallow depths (≤0.05 m), while WCA2D had better predictions

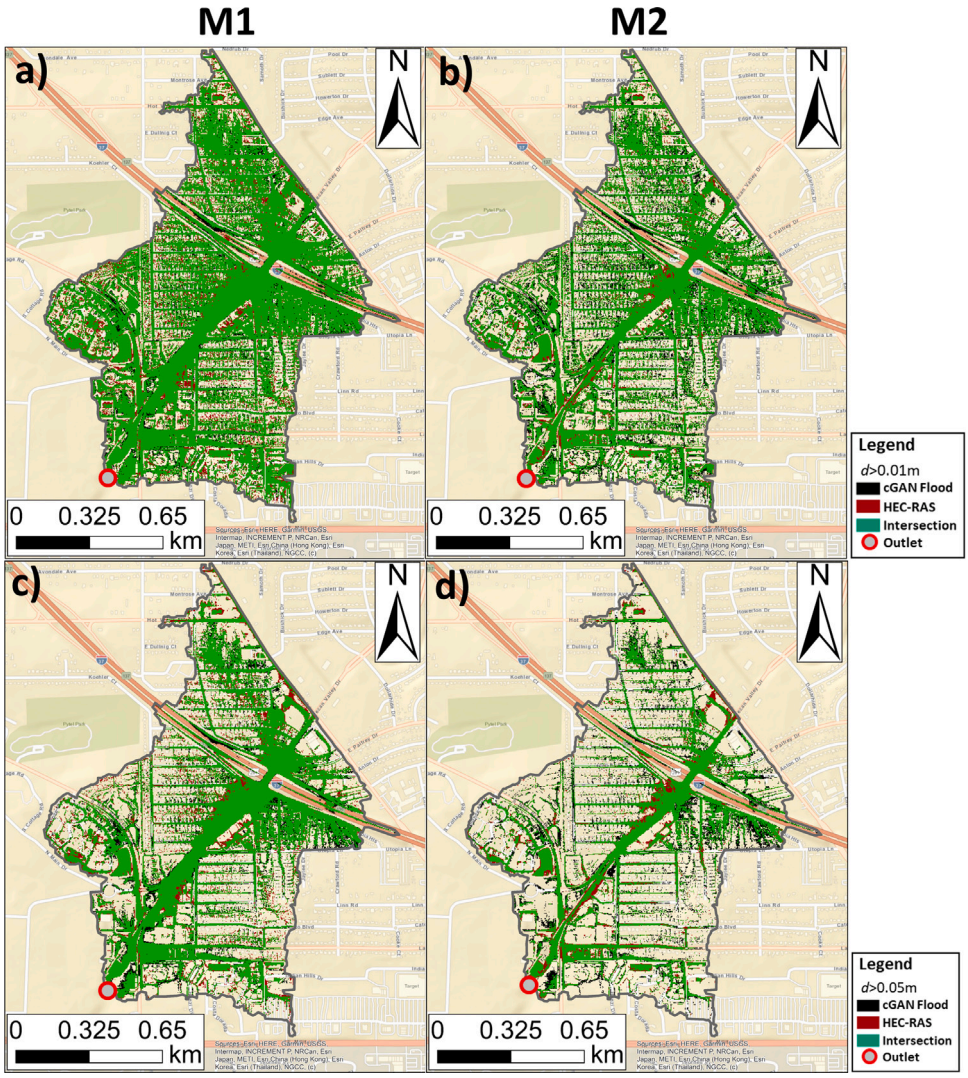

Fig. 10. Flooded areas with depth above 0.01 m and 0.05 m for Manning Scenario M1 (a and b) and M2 (c and d) in area TE5.

Table 4
Volume of maximum flooded depths maps calculated with HEC-RAS, SWMM and WCA2D.

Testing area	Storm	Volume HEC-RAS	v_t (SWMM)	Volume WCA2D
TE1	25-years	269	248 (−7.8%)	219 (−19%)
	100-years	329	308 (−6.3%)	273 (−17%)
	Obs1	234	227 (−2.9%)	197 (−16%)
	Obs2	182	198 (8.7%)	148 (−19%)
TE2	25-years	36	33 (−8.3%)	33 (−8%)
	100-years	44	41 (−6.8%)	43 (−2%)
	Obs1	32	29 (−9.3%)	32 (−2%)
	Obs2	27	20 (−25.9%)	26 (−4%)

for depths between 0.05 and 0.3 m. Concerning $CSI_{0.3}$, cGAN-Flood over-performed WCA2D for all rainfall events in area TE1 and under-performed in area TE2. For both models, $CSI_{0.05}$ is higher than $CSI_{0.3}$, which is expected as the considered flooded area is smaller for depths ≤0.05 m than ≤0.3 m. Lowe et al. (2021) also reported lower $CSI_{0.3}$ values than $CSI_{0.05}$.

Fig. 11 shows the maximum water depths computed for areas TE1 and TE2, for the 100-year and Obs2 rainfall events. Both models predicted well the depths and flooded area in TE1. However, WCA2D underestimated depths in the main channel. The reason is that WCA2D uses a single Manning's n for the entire domain, a limitation of this approach. The cGAN-Flood significantly underestimated the flooded area in TE2 for the Obs2 event. In this case, v_t calculated with SWMM was 25% lower than HEC-RAS, which resulted in a significantly lower flood area.

The cGAN-Flood coupled with the SWMM was significantly faster than HEC-RAS and WCA2D (Table 6). While WCA2D was on average 5 times faster than HEC-RAS, cGAN-Flood and SWMM combined were on average 250 times faster than HEC-RAS. The difference in simulation time is more significant for larger areas. For example, in area TE3 (largest training area), HEC-RAS simulations took an average of 2 h and 45 min between all four events, while cGAN-Flood (excluding SWMM) required an average of 30 s to complete all the simulations (330 times faster). The computational time of the cGAN-Flood method is more variable than HEC-RAS or WCA2D because it depends on the number of iterations of the golden search optimization method that estimates the

Table 5
Performance metrics of cGAN-Flood, combined with SWMM, compared to WCA2D.

	Testing area	Storm	RSME (m)	R^2	HR (%)	FAR (%)	CSI (%)	$CSI_{0.05}$ (%)	$CSI_{0.3}$ (%)
cGAN-Flood	TE1	100-year	0.1	0.88	93.4	12.5	82.4	74	70.9
		25-year	0.08	0.9	85.9	9.6	78.7	70.7	71.5
		Obs1	0.1	0.89	92.7	10	84	76	72.4
		Obs2	0.08	0.84	84.3	6.6	79.6	72.7	71.4
	TE2	100-year	0.09	0.82	94.5	15.6	80.5	71.7	69.7
		25-year	0.09	0.86	86.9	12.5	77.3	68.8	70
		Obs1	0.09	0.79	96.1	26.9	71	60.4	61.2
		Obs2	0.11	0.56	83.3	16.8	71.3	60.2	41.1
WCA2D	TE1	100-year	0.1	0.89	98.6	31.5	67.8	76.2	69.5
		25-year	0.09	0.9	98.3	33.6	65.7	75.7	70
		Obs1	0.09	0.91	97.6	33.7	65.3	75.3	69.8
		Obs2	0.09	0.86	95.2	35	62.9	73	66.7
	TE2	100-year	0.06	0.93	98.6	34.6	64.9	75.5	74.6
		25-year	0.05	0.93	98.6	39.6	59.8	76.9	73.9
		Obs1	0.05	0.94	98.1	41	58.3	75.9	71.6
		Obs2	0.05	0.94	96.6	42.2	56.7	77.7	69.6

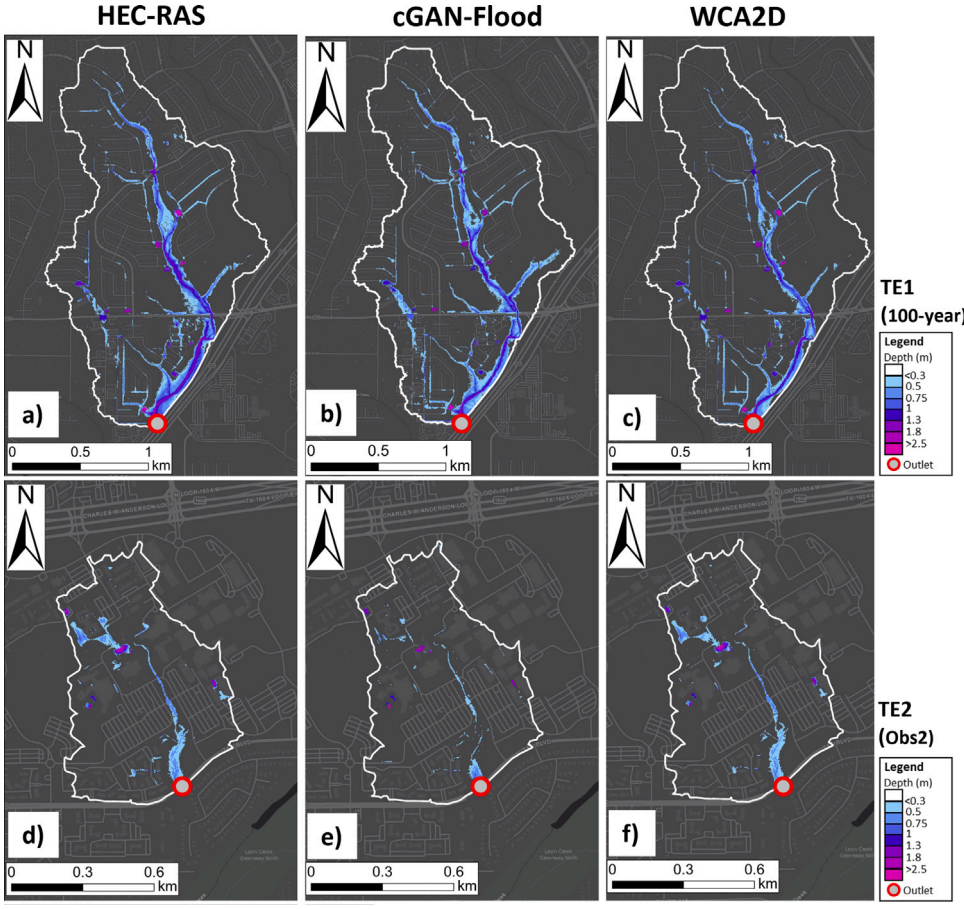

Fig. 11. cGAN-Flood + SWMM outputs compared to HEC-RAS and WCA2D for areas TE1 (100-year storm) and TE2 (Obs2 event).

parameter τ. For instance, cGAN-Flood found $\boldsymbol{D}$ after eight iterations for the 25-year storm and required only two iterations for the 100-year storm in area TE1. It is worth noting that v_t can be calculated with other hydrological models other than SWMM (e.g., HEC-HMS or GSSHA), which can affect the simulation speed and accuracy in computing v_t.

5. Advantages and limitations

Here we discuss the major advantages and some limitations of cGAN-Flood. The major advantage of cGAN-Flood is speed. The results show significantly faster simulations than hydrodynamic simulations while maintaining the overall good predictive performance of flood locations and water depths. Second, cGAN-Flood was able to generalize predictions in unseen catchments. The generalization can promote the application of DL in flood studies, as the necessity of retraining is reduced. The proposed cGAN-Flood model achieves results comparable with other RFMs but can significantly improve the computational time of large-scale flood simulations. Another advantage of cGAN-Flood compared to other deep learning methods found in the literature is the use of mass conservation, which gives the model a physical backbone, although it requires the use of a hydrological model to estimate the total runoff volumes.

Regarding limitations or deficiencies, the results indicate that cGAN-Flood tended to underestimate flood depths in channels for small

Table 6
Total simulation times with HEC-RAS, cGAN-Flood and WCA2D.

Testing area	Storm	HEC-RAS (s)	WCA2D (s)	cGAN-Flood (s)	cGAN-Flood with SWMM (s)
TE1	25-years	5357	1079	13.1	18.3
	100-years	5361	1204	7.1	13.5
	Obs1	5487	1175	16.6	22.5
	Obs2	5251	1144	16.4	21.9
TE2	25-years	1363	252	2.8	6.1
	100-years	1584	329	5.4	8.7
	Obs1	1469	290	3.8	7.1
	Obs2	1281	310	5.1	8.5

rainfall events. One limitation of our study is that infiltration and hydraulic structures (e.g., sewers, bridges, and culverts) were excluded from the HEC-RAS simulations used to generate the training dataset. Although the infiltration losses and drainage system can be simulated with SWMM and be reflected on v_t, the overland distribution of flooded volume might be affected because the location of flow and volumes (e.g., combined sewer overflow) is lost. The lack of infiltration losses and no options for the sewer system are also missing from other models, such as WCA2D. The catchments tested in this study rely mainly on open channels and overland flow. The model likely fails in urban catchments where underground sewer systems are significant. One solution might be applying cGAN-Flood in subdivided catchments in multiple simulations. In this case, v_t values could be calculated for each sub-catchment. and be used with cGAN-Flood to spread the flood volume, helping to identify vulnerable locations to flooding. More investigation is needed on how to incorporate drainage infrastructure in the model (e.g., sewers, culverts, bridges). Furthermore, subdividing the domain into smaller sub-catchments can also be necessary to reduce the error when calculating maximum depths in a large domain. The v_t can be underestimated if the instantaneous maximum runoff volume is calculated for large areas, as the effect of flood wave propagation can be significant. Finally, our model assumes a linear WSE for deeper waters, which can be less accurate than the outputs of 2D modeling with WCA2D and HEC-RAS. Brunner (2016) also highlights that linear assumptions of HEC-RAS 1D are a disadvantage compared to the HEC-RAS 2D.

One potential application of cGAN-Flood would be to improve the resolution of fully distributed hydrological and hydraulic models. The idea is to run fully distributed hydrological models (e.g., GHSSA) in a coarse spatial resolution and use cGAN-FLood to obtain high-resolution floodplains. One limiting aspect of cGAN-Flood is cell resolution. Currently, the cGAN-Flood model only works for a spatial resolution of 3 m. Although this is within the FEMA recommended scale for hydraulic studies of floodplains, cGAN-Flood cannot be used to generate flood maps on finer cells or in areas where high-resolution data do not exist. Predictions with different spatial resolutions would require building new hydrodynamic models to generate new data and re-train cGAN-Flood.

6. Conclusions and future model improvements

This paper developed the model cGAN-Flood, a rapid flood model based on conditional generative adversarial networks to estimate high-resolution maximum flood areas. This approach generates maximum flood depths by distributing a flood volume. cGAN-Flood uses two generators to first identify flooded cells and then to estimate water depths. This method was evaluated for catchments unseen to the generators during training or validation phases. Results showed that cGAN-Flood could generate accurate results for a variety of catchments, rainfall magnitudes, and surface roughness, indicating that the generators could satisfactorily generalize their predictions. Future model improvements could include infiltration, different Manning's *n*, and sewer systems. Adding these aspects could enable DL models to calculate flood depths directly from rainfall inputs, potentially eliminating the need to use a hydrological model before computing the flood volume to be distributed. These advancements, however, would require many additional simulations on catchments with different sewer system configurations, various infiltration capacities, and land cover types. We also suggest the development of new DL methods that can accurately generalize predictions of deeper water, which can replace the WSE interpolation approach adopted for cGAN-Flood and improve rapid flood modeling results".

CRediT authorship contribution statement

Cesar A.F. do Lago: Conceptualization, Methodology, Formal analysis, Software, Writing – original draft. **Marcio H. Giacomoni:** Conceptualization, Writing – review & editing, Supervision. **Roberto Bentivoglio:** Conceptualization, Writing – review & editing. **Riccardo Taormina:** Conceptualization, Writing – review & editing. **Marcus N. Gomes Junior:** Conceptualization, Writing – review & editing. **Eduardo M. Mendiondo:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was funded by the Transportation Consortium of South-Central States (Tran-SET) - University Transportation Center for Region 6 Project, USA #18HSTSA02 and by the Coordination for the Improvement of Higher Education Personnel (Capes), Brazil - Proex. The authors would like to thank the San Antonio River Authority (SARA) for providing us with the elevation and imperiousness data.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jhydrol.2023.129276.

References

Aich, V., Liersch, S., Vetter, T., Fournet, S., Andersson, J.C.M., Calmanti, S., van Weert, F.H.A., Hattermann, F.F., Paton, E.N., 2016. Flood projections within the Niger River Basin under future land use and climate change. Sci. Total Environ. 562, 666–677. http://dx.doi.org/10.1016/j.scitotenv.2016.04.021, URL: http://www.sciencedirect.com/science/article/pii/S0048969716307021.

Bentivoglio, R., Isufi, E., Jonkman, S.N., Taormina, R., 2022. Deep learning methods for flood mapping: a review of existing applications and future research directions. Hydrol. Earth Syst. Sci. 26 (16), 4345–4378.

Berkhahn, S., Fuchs, L., Neuweiler, I., 2019. An ensemble neural network model for real-time prediction of urban floods. J. Hydrol. 575, 743–754.

Brown, J.D., Spencer, T., Moeller, I., 2007. Modeling storm surge flooding of an urban area with particular reference to modeling uncertainties: A case study of Canvey Island, United Kingdom. Water Resour. Res. 43 (6).

Brunner, G., 2016. HEC-RAS River Analysis System, 2d Modeling User's Manual, Version 5.0. US Army Corps of Engineers, Hydrologic Engineering Center, Davis.

Bulti, D.T., Abebe, B.G., 2020. A review of flood modeling methods for urban pluvial flood application. Model. Earth Syst. Environ. 6 (3), 1293–1302.

Cea, L., Costabile, P., 2022. Flood risk in urban areas: Modelling, management and adaptation to climate change. a review. Hydrology 9 (3), 50.

Cheng, M., Fang, F., Pain, C.C., Navon, I., 2020. Data-driven modelling of nonlinear spatio-temporal fluid flows using a deep convolutional generative adversarial network. Comput. Methods Appl. Mech. Engrg. 365, 113000.

Clavet-Gaumont, J., Huard, D., Frigon, A., Koenig, K., Slota, P., Rousseau, A., Klein, I., Thiémonge, N., Houdré, F., Perdikaris, J., Turcotte, R., Lafleur, J., Larouche, B., 2017. Probable maximum flood in a changing climate: An overview for Canadian basins. J. Hydrol.: Reg. Stud. 13, 11–25. http://dx.doi.org/10.1016/j.ejrh.2017.07.003, URL: http://www.sciencedirect.com/science/article/pii/S2214581817301696.

do Lago, C.A.F., Hofheinz Giacomoni, M., Olivera, F., Mario Mendiondo, E., 2021. Assessing the impact of climate change on transportation infrastructure using the hydrologic-footprint-residence metric. J. Hydrol. Eng. 26 (5), 04021014.

Dtissibe, F.Y., Ari, A.A.A., Titouna, C., Thiare, O., Gueroui, A.M., 2020. Flood forecasting based on an artificial neural network scheme. Nat. Hazards 104 (2), 1211–1237.

Engman, E.T., 1986. Roughness coefficients for routing surface runoff. J. Irrig. Drain. Eng. 112 (1), 39–53.

Federal Emergency Management Agency, 2020. Guidance for flood risk analysis and mapping - flood depth and analysis rasters, Guidance Document 14.

Ghimire, B., Chen, A.S., Guidolin, M., Keedwell, E.C., Djordjević, S., Savić, D.A., 2013. Formulation of a fast 2D urban pluvial flood model using a cellular automata approach. J. Hydroinform. 15 (3), 676–686.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Amer. Statist. Assoc. 102 (477), 359–378.

Gondhi, N.K., Kour, E.N., 2017. A comparative analysis on various face recognition techniques. In: 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). pp. 8–13. http://dx.doi.org/10.1109/ICCONS.2017.8250626.

Gonog, L., Zhou, Y., 2019. A review: generative adversarial networks. In: 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, pp. 505–510.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Adv. Neural Inf. Process. Syst. 27.

Guidolin, M., Chen, A.S., Ghimire, B., Keedwell, E.C., Djordjević, S., Savić, D.A., 2016. A weighted cellular automata 2D inundation model for rapid flood analysis. Environ. Model. Softw. 84, 378–394.

Guo, Z., Moosavi, V., Leitão, J.P., 2022. Data-driven rapid flood prediction mapping with catchment generalizability. J. Hydrol. 609, 127726.

Hofmann, J., Schüttrumpf, H., 2021. floodGAN: Using deep adversarial learning to predict pluvial flooding in real time. Water 13 (16), 2255.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134.

Jamali, B., Bach, P.M., Cunningham, L., Deletic, A., 2019. A Cellular Automata fast flood evaluation (CA-fféé) model. Water Resour. Res. 55 (6), 4936–4953.

Jamali, B., Löwe, R., Bach, P.M., Urich, C., Arnbjerg-Nielsen, K., Deletic, A., 2018. A rapid urban flood inundation and damage assessment model. J. Hydrol. 564, 1085–1098.

Kabir, S., Patidar, S., Xia, X., Liang, Q., Neal, J., Pender, G., 2020. A deep convolutional neural network model for rapid prediction of fluvial flood inundation. J. Hydrol. 590, 125481.

Kiefer, J., 1953. Sequential minimax search for a maximum. Proc. Amer. Math. Soc. 4 (3), 502–506.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lowe, R., Böhm, J., Jensen, D.G., Leandro, J., Rasmussen, S.H., 2021. U-FLOOD–Topographic deep learning for predicting urban pluvial flood water depth. J. Hydrol. 603, 126898.

Ma, X., Guo, J., Sansom, A., McGuire, M., Kalaani, A., Chen, Q., Tang, S., Yang, Q., Fu, S., 2021. Spatial pyramid attention for deep convolutional neural networks. IEEE Trans. Multimed. 23, 3048–3058.

Martins, V.S., Kaleita, A.L., Gelder, B.K., da Silveira, H.L., Abe, C.A., 2020. Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution. ISPRS J. Photogramm. Remote Sens. 168, 56–73.

Natarajan, S., Radhakrishnan, N., 2019. Simulation of extreme event-based rainfall–runoff process of an urban catchment area using HEC-HMS. Model. Earth Syst. Environ. 5 (4), 1867–1881.

Palmitessa, R., Grum, M., Engsig-Karup, A.P., Löwe, R., 2022. Accelerating hydrodynamic simulations of urban drainage systems with physics-guided machine learning. arXiv preprint arXiv:2206.01538.

Park, J.-H., Choi, Y.-K., Kang, C., 2020. Fast cropping method for proper input size of convolutional neural networks in underwater photography. J. Soc. Inf. Disp. 28 (11), 872–881.

Shi, Z., Shen, Q., Tan, Q., Li, T., 2021. Development of integrated flooding early warning and rainfall runoff management platform for downtown area of shanghai. Sustainability 13 (20), 11250.

Sit, M., Demiray, B.Z., Xiang, Z., Ewing, G.J., Sermet, Y., Demir, I., 2020. A comprehensive review of deep learning applications in hydrology and water resources. Water Sci. Technol. 82 (12), 2635–2670.

Swischuk, R., Mainini, L., Peherstorfer, B., Willcox, K., 2019. Projection-based model reduction: Formulations for physics-based machine learning. Comput. & Fluids 179, 704–717.

Teng, J., Jakeman, A.J., Vaze, J., Croke, B.F., Dutta, D., Kim, S., 2017. Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. Environ. Model. Softw. 90, 201–216.

USDA, S., 1986. Urban Hydrology for Small Watersheds. Technical Release 55, pp. 2–6.

Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., Wang, F.-Y., 2017. Generative adversarial networks: introduction and outlook. IEEE/CAA J. Autom. Sin. 4 (4), 588–598.

Willard, J., Jia, X., Xu, S., Steinbach, M., Kumar, V., 2020. Integrating physics-based modeling with machine learning: A survey. pp. 1–34, 1, arXiv preprint arXiv:2003.04919.

Yin, J., Guo, S., He, S., Guo, J., Hong, X., Liu, Z., 2018. A copula-based analysis of projected climate changes to bivariate flood quantiles. J. Hydrol. 566, 23–42.

Zarezadeh, V., 2017. Coupling hydrologic and urbanization modeling: a multi-scale investigation to enhance urban water resource systems sustainability. The University of Texas at San Antonio.