Physically-based dimensionless features for pluvial flood mapping with machine learning

Mark S. Bartlett\textsuperscript{1}, Jared Van Blitterswyk\textsuperscript{2}, Martha Farella\textsuperscript{3}, Jinshu Li\textsuperscript{1}, Curtis Smith\textsuperscript{5}, Anthony J. Parolari\textsuperscript{6,7}, Lalitha Krishnamoorthy\textsuperscript{8}, Assaad Mrad\textsuperscript{9}

\textsuperscript{1}The Water Institute, Baton Rouge, LA, USA
\textsuperscript{2}Stantec, Ottawa, ON, Canada
\textsuperscript{3}Stantec, Flagstaff, AZ, USA
\textsuperscript{4}Stantec, Pasadena, CA, USA
\textsuperscript{5}Stantec, New York, NY, USA
\textsuperscript{6}Stantec, Lombard, IL, USA
\textsuperscript{7}Department of Civil, Construction, and Environmental Engineering, Marquette University, Milwaukee, WI, USA
\textsuperscript{8}Stantec, Austin, TX, USA
\textsuperscript{9}Stantec, Raleigh, NC, USA

Key Points:

\begin{itemize}
\item Buckingham $\Pi$ theorem is used to create a dimensionless formulation of pluvial flood mapping.
\item A logistic regression machine learning model uses the dimensionless features to rapidly map pluvial flooding across geographies.
\item Dimensionless features outperform dimensional features, especially when the ML model is trained and tested in different regions.
\end{itemize}

Corresponding author: Mark S. Bartlett, Mark.Bartlett@gmail.com
Abstract

Rapid delineation of flash flood extents is critical to mobilize emergency resources and to manage evacuations, thereby saving lives and property. Machine learning (ML) approaches enable rapid flood delineation with reduced computational demand compared to conventional high-resolution, 2D flood models. However, existing ML approaches are limited by a lack of generalization to never-before-seen conditions. Here, we propose a framework to improve ML model generalization based on dimensionless, multi-scale features that capture the similarity of the flooding process across regions. The dimensionless features are constrained with the Buckingham II theorem and used with a logistic regression model for a probabilistic determination of flood risk. The features were calculated at different scales by varying accumulation thresholds for stream delineation. The modeled flood maps compared well with the results of 2D hydraulic models that are the basis of the Federal Emergency Management Agency (FEMA) flood hazard maps. Dimensionless features outperformed dimensional features, with some of the largest gains (in the AUC) occurring when the model was trained in one region and tested in another. Dimensionless and multi-scale features in ML flood modeling have the potential to improve generalization, enabling mapping in unmapped areas and across a broader spectrum of landscapes, climates, and events.

1 Introduction

Floods cause dozens of deaths and billions of dollars in economic losses annually in the United States alone (Cornwall, 2021; Ashley & Ashley, 2008). Urbanization and increases in heavy rainfall due to climate change have already increased the frequency and severity of floods, which are anticipated to worsen with further climate change (Wasko et al., 2021). To adequately mitigate flood impacts, rapid flood forecasting is needed, but this is still difficult to achieve in practice. Real-time flood forecasting at high resolution is partly prohibited by the high computational demand of high-resolution models (Ivanov et al., 2021).

Strategies to address the computational demand of high-resolution hydraulic models include inundation mapping, by filling digital terrain models (DTMs), and data-driven, or Machine Learning (ML), methods. Inundation mapping with high-resolution terrain data, such as the Height Above Nearest Drainage (HAND) procedure, is a promising approach to rapid flood mapping that can provide actionable information to first responders. The HAND procedure utilizes Manning’s equation to calculate flood stages based on a flow forecast and reach-averaged value for cross-section area, wetted perimeter, and flow surface area (Zheng, Tarboton, et al., 2018; Scriven et al., 2021; Zheng, Maidment, et al., 2018; Garousi-Nejad et al., 2019). Although these methods are fast, their limitations include a lack of bathymetry and infrastructure data and worse performance in smaller, low-order streams and small drainage pathways such as streets (Hocini et al., 2021; Johnson et al., 2019; Garousi-Nejad et al., 2019). A key challenge in assessing pluvial flooding is accurately resolving flood risks from both small, low-order streams and larger, high-order streams. A location may be flooded by nearby small drainage pathways (e.g., streets) or by larger streams that inundate these smaller drainage pathways.

An alternative to flood inundation mapping to improve the speed of flood forecasting is to emulate the hydraulic models using ML. Recently, ML approaches to flood mapping have received significant attention because of their speed and cost-effectiveness compared with previous approaches (Ivanov et al., 2021; Bentivoglio et al., 2022). ML flood mapping models are trained against either satellite data, in-situ measurements, or hydrodynamic model output, and use a wide variety of input features. Typical input features include landscape and geomorphic characteristics, including elevation, curvature, topographic wetness index, slope, river density, drainage distance to the nearest river, HAND, and soil type (Bentivoglio et al., 2022; Bui et al., 2018). Such geomorphic and
landscape features have been used as inputs to several linear binary classifiers (Nachappa et al., 2020; Hosseiny et al., 2020; Rahmati et al., 2020; Giovannettone et al., 2018; Collins et al., 2022; Samela et al., 2016; Manfreda, Samela, et al., 2014; Manfreda, Nardi, et al., 2014; Tavares da Costa et al., 2020). Features are generally selected based on data availability and hypotheses regarding their control of the physical processes governing runoff generation and flood development.

One challenge facing the widespread adoption of ML flood models is their generalization to storms or locations outside of the training sample. Methods to improve model generalization can further alleviate constraints by limiting the data volume and computational time needed to apply models across large and diverse areas (Cache et al., 2024; Pakdehi et al., 2023; Wagenaar et al., 2018; Seleem et al., 2022). This challenge mirrors that of the over-parameterization of spatially-explicit, process-based models (Jakeman & Hornberger, 1993; Beven & Freer, 2001; Beven, 2006). One approach to address model over-parameterization and generalization is to transform the model into a reduced-dimension space that retains essential features of the underlying physics across catchments. For example, Hu et al. (2019) apply linear decomposition to the input features and train the model in the reduced-order space.

Alternatively, the Buckingham Π theorem may be used to reformulate the governing equations of the physical process into a reduced-order, non-dimensional context that inserts engineering and science metadata into the ML process (Rudolph et al., 1998; Porporato, 2022; Oppenheimer et al., 2023; Gunaratnam et al., 2003). Importantly, relative to linear decomposition, the Buckingham Π approach constrains the number of dimensionless features based on the underlying physics. Such an approach improves generalization, particularly when extrapolating across scales (Oppenheimer et al., 2023). To evaluate the ability of this approach to improve ML flood model generalization, we apply the Buckingham Π theorem to the underlying mass and momentum balance equations and develop a set of dimensionless hydraulic and hydrologic indices. In turn, these indices are used as inputs to a logistic regression model that predicts flooding across different geographies.

To address pluvial flooding from both small, low-order streams and larger, high-order streams, we define these dimensionless hydraulic and hydrologic indices relative to a point, as well as in relation to two different drainage pathway delineations: local and non-local. The local delineation captures the small drainage pathways (e.g., streets) and low-order streams by using a low threshold for flow accumulation to determine stream origins. In contrast, the non-local delineation maps only higher-order streams by applying a much larger flow accumulation threshold. The logistic regression model then uses these dimensionless indices from the point scale and both the local and non-local delineations to predict flood risks across the various scales of drainage pathways.

The overall ML approach includes definition of the dimensionless input features, estimation of the input features, and the model training and testing (Figure 1). In Section 2, we use the Buckingham Π theorem to derive dimensionless indices to capture the similarity of the flood process based on descriptions for both flow hydraulics and flood hydrology. We then discuss how the Buckingham Π theorem-defined flood response is subsumed by a logistic regression ML model. In Section 3.1, we discuss the steps used to process the dimensionless indices from raw data and models used to approximate both flow and channel properties. To capture the flood process in detail, these indices are defined at different scales: 1) at a point, 2) for smaller local flow paths (e.g., streets), and 3) for larger streams and rivers. Section 3.2 describes the ML methodology, including label data, model training, and performance assessment. In Section 4, we compare the performance of the proposed dimensionless features with the traditional dimensional features. This comparison is conducted with HUC 12 catchments in 2 distinct regions, Chicago and New Jersey. Finally we discuss the broad applicability of dimensionless indices in ML to improve generalization, interpretability, scalability, and efficiency.
Figure 1. The ML process is applied to selected HUC 12 watersheds across the United States. For each selected HUC 12 watershed, a) the raw data is retrieved, b) the raw data is processed into derivative data, c) the flood hydrology and extents are downloaded, d) the derivative data and flood hydrology are used to construct a series of geomorphological instantaneous unit hydrographs to relate the peak (maximum) flow to contributing area, e) the flow and derivative data are combined into dimensionless indices that capture the hydraulic and hydrologic similarity of the flooding process, f) the dimensionless data is related to the flood extents by an ML logistic regression model, and g) the logistic regression model produces flood extents, flood probability, and the error in comparison to the original training data.
Table 1. Variable and parameter definitions

| Symbol | Units | Description |
|--------|-------|-------------|
| A      | \([L^2]\) | Channel cross section area |
| \(A_s\) | \([L^2]\) | Surface area of flooding |
| \(a_c\) | \([L^2]\) | Contributing area |
| \(a_{cr}\) | \([L^2]\) | Contributing area producing runoff |
| \(q_t\) | \([L]\) | Contributing area per unit contour length |
| \(D_f\) | \([\text{--}]\) | Budyko dryness index |
| \(f_{th}(t)\) | \([1/T]\) | Instantaneous unit hydrograph (IUH), i.e., travel time distribution |
| \(g\) | \([L/T^2]\) | Gravitational acceleration |
| \(h\) | \([L]\) | Water depth (in channel) |
| \(h_{\text{max}}\) | \([L]\) | Maximum water depth (in channel) during a storm event |
| \(K\) | \([L^3/T]\) | Conveyance |
| \(K_v\) | \([L/T]\) | Conveyance velocity |
| \(k_s\) | \([L/T]\) | Saturated hydraulic conductivity |
| \(l_c\) | \([L]\) | Channel length along the channel centerline |
| \(n\) | \([T/L^{1/3}]\) | Manning’s roughness coefficient |
| \(Q\) | \([L^3/T]\) | Flow rate in channel |
| \(Q_{\text{max}}\) | \([L^3/T]\) | Peak flow rate in channel |
| \(q_d\) | \([L^2/T]\) | Groundwater flux |
| \(q_r\) | \([L^2/T]\) | Recharge rate per unit contour length |
| \(q_o\) | \([L^2/T]\) | Subsurface flow per unit contour length |
| \(r\) | \([L/T]\) | Recharge rate per unit area |
| \(R_h\) | \([L]\) | Hydraulic radius |
| \(S\) | \([\text{--}]\) | Topographic slope at a point |
| \(S_0\) | \([\text{--}]\) | Channel bed slope |
| \(S_f\) | \([\text{--}]\) | Friction slope |
| \(t\) | \([T]\) | Time |
| \(u\) | \([L/T]\) | Flow velocity |
| \(x\) | \([L]\) | Distance along the flow centerline |
| \(y_d\) | \([L]\) | Horizontal distance from nearest drainage |
| \(z_w\) | \([L]\) | Distance to channel water |
| \(z_d\) | \([L]\) | Vertical distance above nearest drainage |
| \(\pi_1, \ldots, \pi_n\) | \([\text{--}]\) | \(\Pi\)-theorem terms related to hydraulics |
| \(\Pi_1, \ldots, \Pi_n\) | \([\text{--}]\) | \(\Pi\)-theorem terms related to hydrology |

\(a\) Variables in the text with an overline bar indicate a reach-averaged value, e.g., Zheng, Tarboton, et al. (2018).

2 Theory

Flooding is likely to occur in areas that are 1) adjacent to rivers, streams, and localized flow paths and 2) persistently wet (e.g., wetland areas or areas of flow convergence). The first case captures the storm event dynamics where the intensity of precipitation causes flow path expansion, while the second case captures the long-term controls of topography and climate on patterns of saturation. For each case, we consider the fundamental mathematical description of the flooding process, isolate the governing variables, and utilize the Buckingham \(\Pi\) Theorem to create dimensionless indices that relate the similarity of the flooding process across divergent conditions.

2.1 Storm Event Hydrology and Hydraulics

For flooding by flow path expansion, the flow hydraulics are represented by the Saint–Venant equations describing conservation of mass (i.e., continuity) and momentum as follows (Chow, 1959; Moussa & Bocquillon, 2000; Erçan et al., 2014):
Figure 2. The governing parameters for the flow hydraulics, where at a point, flooding occurs when the distance to water is less than or equal to zero.

\[
\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0 \tag{1}
\]

\[
\frac{\partial u_c}{\partial t} + u_c \frac{\partial u_c}{\partial x} + g \frac{\partial h}{\partial x} + g (S_f - S_0) = 0 \tag{2}
\]

where \( t \) is time, \( x \) is the spatial coordinate along the channel centerline, \( h \) is the water depth, \( Q \) is the flow rate, \( A \) is the channel cross-sectional area, \( u_c \) is the depth averaged velocity in the \( x \) direction, \( g \) is gravitational acceleration, \( S_0 \) is the ground slope of the channel, and \( S_f \) is the friction slope (Fig. 2). The Saint-Venant equations apply to a 1D topological description of the watershed flow path network. Note that the flow, \( Q \), could account for infiltration and exfiltration along the channel length, and in steady state, Eqs. (1) and (2) are the basis of the standard step method of HEC-RAS 1D. The coupled set of equations (1) and (2) is closed by assuming a function for the friction slope (Yang et al., 2019; Ercan et al., 2014), i.e.,

\[
S_f = \frac{Q|Q|}{K^2} \tag{3}
\]

where \( K = \frac{1}{n} R_h^{2/3} A \) is the conveyance, for which \( R_h \) is the hydraulic radius, and \( n \) is Manning’s roughness coefficient.

For a flood event, equations (1) and (2) determine the dynamics of water flow and the hydraulic pressure and depths, where \( h_{\text{max}} \) is the maximum depth at a point along the 1D topological description of the channel. We assume that any given point is flooded when \( h_{\text{max}} \) exceeds the elevation above the nearest drainage, \( z_d \) (Fig. 2). That is, a point is flooded when the distance to water is negative, \( z_w = h_{\text{max}} - z_d \leq 0 \). We assume \( z_w \) depends on:

1. \( Q \ [L^3/T] \), the flow rate, for the flood magnitude constraint,
2. \( z_d \ [L] \), the height above the nearest drainage, for constraining the number of flooded points,
3. \( K_v \ [L/T] \), the velocity of conveyance (i.e., \( K/A \)) of flow representing the channel capacity constraint,
4. \( A \ [L^2] \), the channel cross-sectional area,
5. \( gS_0 \ [L/T^2] \), the slope multiplied by gravity, representing the external force driving the flow downslope.

The velocity, \( u_c \), and the friction slope, \( S_f \), are not considered since they are derived directly from the listed variables. Furthermore, we do not consider time because flooding is based on the maximum flow and height of the water level, \( Q_{\text{max}} \) and \( h_{\text{max}} \), respectively.
Hence, the depth of the water, $z_w$, is considered the maximum level of water during a flood event.

Based on the Saint-Venant equations and the assumed channel geometry, the distance to water is a function of 5 variables, that is, $z_w = f(Q, K_v, z_d, A, g S_0)$, and two dimensions of length [L] and time [T]. Following the Buckingham Π theorem, the required number of π groups is $6 - 2 = 4$, where we have 1 dependent variable and 5 independent variables. The π groups based on the two repeating variables, $K_v$ and $A$, are

\[
\pi_1 = \frac{z_w}{\sqrt{A}}, \\
\pi_2 = \frac{Q}{K_v A}, \\
\pi_3 = \frac{z_d}{\sqrt{A}}, \\
\pi_4 = \frac{g S_0 \sqrt{A}}{K_v^2},
\]

where the groupings were selected by inspection. This process yields the following dimensionless π group relationship:

\[
\pi_1 = f(\pi_2, \pi_3, \pi_4),
\]

where $\pi_1$, the distance to the flood surface, $z_w$, normalized by the square root of the channel cross sectional area, $\sqrt{A}$, is a function of the other terms, $\pi_2$, $\pi_3$, and $\pi_4$. The term $\pi_2$ represents the flow to flow capacity ratio, $\pi_3$ accounts for how the geometry of the channel (e.g., expansion and contraction of the channel) impacts flooding, while $\pi_4$ represents the flow velocity constrained by the gravitational force over the flow velocity constrained by the frictional force. Note that $\pi_3$ does not exist without reference to a stream or river to define the height above the nearest drainage, $z_d$.

### 2.2 Long-term Hydrology

To identify flooding in typically wet areas with saturated soil, we consider the assumptions of TOPModel (Beven, 2012) where the dynamics of the water table depth from the surface are considered as a succession of steady states that balance the long-term watershed recharge with the outflow from the saturated soil layer. Accordingly, the long term recharge (per unit contour length), $q_r$, is given by

\[
q_r = a_t r,
\]

where $a_t$ is the contributing area (per unit contour length), and $r$ is the uniform recharge rate (per unit area). Note that $a_t$ is equal to the contributing area, $a_i$. In steady state, $q_r$ is balanced by the outflow governed by Darcy’s Law with the hydraulic gradient, $dh/dx$, approximated by the local topographic slope, $S$ (Beven, 2012), and a hydraulic conductivity that decreases exponentially with the soil depth, i.e.,

\[
k_s \int_{z_g}^{\infty} e^{-\frac{z}{z_d}} dz = k_s z_d e^{-\frac{z}{z_d}},
\]

where $k_s$ is the saturated hydraulic conductivity at the surface that decreases exponentially with a decay factor of $1/z_d$ (Ducharne, 2009), $z_g$ is the depth to groundwater, and
\[ q_o = S k_z z_d e^{-\frac{z_g}{z_d}}, \]

where \( q_o \) is the (per unit contour length) outflow.

In steady state, the inflow of Eq. (9) equals the outflow of Eq. (11) and, after rearranging terms, we retrieve the relationship between two dimensionless terms,

\[ \frac{z_g}{z_d} = -\ln \left[ \frac{a_l r}{z_d k_z S} \right], \]

where the depth to groundwater, \( z_g \), normalized by the height above drainage, \( z_d \), is equivalent to the natural log of the recharge inflow normalized by the outflow of Darcy’s Law with the hydraulic gradient subsumed by the topographic slope, \( S \), over a depth equal to the height above nearest drainage (per unit contour length) (Fig. 3).

Based on Eq. (12), the distance to the surface of the groundwater, \( z_g = f(z_d, q_r, q_d) \), depends on

1. \( z_d \) [L], the height above drainage, for constraining the number of flooded points,

Figure 3. The watershed hydrology for calculating wet saturated areas (prone to flooding) is based on a) the TOPModel assumption of a spatially variable inflow rate derived from the contributing area (per unit contour length), \( a_l \), multiplied by the watershed recharge rate, \( r \), here derived from b) a partition of rainfall to evapotranspiration, \( \langle E/R \rangle \), given by the Budyko curve, that is balanced by c) a water table that parallels the surface slope with soil water lateral transmission, \( q_o \), that decreases exponentially with depth.
2. \( q_r \, [L^2/T] \), the moisture recharge flux into a point per unit contour length,
3. \( q_d \, [L/T] \), the groundwater flux based on Darcy’s Law, i.e., \( k_s S \), assuming the hydraulic gradient, \( dh/dx \), is represented by the topographic surface slope, \( S \).

Therefore, the system consists of 4 variables and 2 dimensions, so the Buckingham Π theorem yields two Π groups \((4 - 2 = 2)\). From Eq. (12), these Π groups are

\[
\Pi_1 = \frac{z_d}{z_d}, \\
\Pi_2 = \frac{a_l r}{z_d k_s S},
\]

where we have substituted \( q_r \) with \( a_l r \) from Eq. (9) and \( q_d \) with \( k_s S \) following the assumption of TOPModel. These Π groups provide for the following dimensionless relationship:

\[
\Pi_1 = \ln(\Pi_2),
\]

where the natural log function is inferred from inspection of Eq. (12). The term \( \ln(\Pi_2) \) is similar to the topographic wetness index (Beven & Kirkby, 1979); but, differently, the argument of the natural log function, \( \Pi_2 \), is dimensionless and reflects the ratio of water accumulation to water outflow from groundwater (per unit contour length). In \( \ln(\Pi_2) \), the addition of \( z_d \) is similar to the TWI that accounts for the landscape position relative to water (groundwater and surface water), which was found to better represent field observations (Meles et al., 2020).

### 2.3 From Buckingham Π theorem to ML

Following the Buckingham Π theorem, flooding is determined by the dimensionless groups \( \pi_1 \) or \( \Pi_1 \) for flooding from expansion of the channel flow path and in saturated areas (i.e., lakes and depressions), respectively. In both cases, when \( \pi_1 \) or \( \Pi_1 \) is negative, an area is flooded, but when both \( \pi_1 \) and \( \Pi_1 \) are positive, an area is not flooded. Thus, the binary result of flooded (e.g., 1) and not flooded (e.g., 0) is found by applying a multidimensional Heaviside function, \( \Theta(\cdot, \cdot) \), i.e.,

\[
1 - \Theta(\Pi_1, \pi_1) = 1 - \Theta[f(\pi_2, \pi_3, \pi_4), \ln(\Pi_2)],
\]

for which the Heaviside step function is only 1 if both arguments are positive, the function is right continuous, i.e., \( \Theta(0, 0) = 1 \), and the function \( f(\pi_2, \pi_3, \pi_4) \) is unknown. Here, we consider that an ML algorithm may subsume the overall function \( \Theta(f(\pi_2, \pi_3, \pi_4), \ln(\Pi_2)) \). Specifically, we consider the logistic regression for the likelihood (probability of flooding), i.e.,

\[
1 - \Theta(\Pi_1, \pi_1) \approx \Theta[P(\ln[\pi_2], \ln[\pi_3], \ln[\pi_4], \ln[\Pi_2]) - \epsilon],
\]

where \( P(\ln[\pi_2], \ln[\pi_3], \ln[\pi_4], \ln[\Pi_2]) \) is the logistic regression probability, the natural log, \( \ln[\cdot] \), is applied to all terms because we assume a multiplicative relationship (i.e., addition in log space results in multiplication of the variables), and the discrimination threshold, \( \epsilon \), allows for the step function, \( \Theta(\cdot) \), to filter the logistic regression probability into flooded (i.e., 1) or not flooded areas (i.e., 0).
Table 2. Input data sources and resolution

| Data Type                  | Resolution (m) | Source                                      | Base URL                                      |
|----------------------------|----------------|---------------------------------------------|-----------------------------------------------|
| Digital Elevation Model    | 1-10           | USGS Elevation Products (3DEP)              | https://apps.nationalmap.gov                   |
| Land Use                   | 30             | NLCD<sup>a</sup>                            | https://www.mrlc.gov                          |
| Impervious Surface         | 30             | NLCD Impervious Surface                     | https://www.mrlc.gov                          |
| Roads                      | 30             | NLCD Impervious Descriptor                  | https://www.mrlc.gov                          |
| Soil Type                  | 30             | NRCS SSURGO<sup>b</sup>                     | https://www.nrcs.usda.gov                     |
| Hydrography                | n/a            | National Hydrography Dataset (NHD)          | https://apps.nationalmap.gov                   |
| Watershed Boundaries       | n/a            | USGS Hydrologic Units (HU)                 | https://apps.nationalmap.gov                   |
| Flood Maps                 | 1-5            | FEMA Risk MAP Program                      |                                               |
| Climatology                | 10<sup>3</sup> | Daymet                                      | https://daymet.ornl.gov                       |

<sup>a</sup> Natural Land Cover Database

<sup>b</sup> Natural Resources Conservation Service Soil Survey Geographic Database

3 Data and Machine Learning Methodology

3.1 Dimensionless Input Feature Estimation

The input feature π and Π groups are derived from the variables of Q, A, n, Rh, zd, So, q1, r, ks, and S (Table 1). For each HUC-12 watershed, these variables are derived by transforming the raw data (Table 2) with GRASS GIS (Jasiewicz & Metz, 2011). Before processing the variables with GRASS GIS, the raw DEM was hydro-enforced with the stream and river hydrography centerlines.

3.1.1 Landscape properties at each point

Landscape properties were calculated on a grid at the resolution of the raw DEM data, including Manning’s n, saturated hydraulic conductivity, slope, contributing area, conveyance velocity, and recharge rate. The DEM resolution was either 1-m, 3-m, or 10-m. Manning’s roughness coefficient, n, was based on land use data as described in the Hydrologic Engineering Center’s River Analysis System (HEC-RAS) (USACE Hydrologic Engineering Center, 2024). Saturated hydraulic conductivity, ks, was extracted from the Soil Survey Geographic Database (SSURGO) database as the most restrictive soil component (i.e., lowest ks value) in the top 5 meters. Slope and contributing area were estimated from the DEM.

The conveyance velocity, Kv, was calculated in two steps: 1) initially, we estimated Kv based on an assumed sheet flow and Manning’s n, and 2) we then replaced these initial point values with an average of the upstream conveyance velocities (tributary to each point), scaled by the factor \( \frac{a_{c}^{1/2}S_{1/2}}{a_{c}^{1/2}S_{1/2}} \), where the overline (of \( a_{c}^{1/2}S_{1/2} \)) indicates the upstream average of all contributing points (Rinaldo et al., 1991; D’Odorico & Rigon, 2003; Maidment et al., 1996). For the initial calculation of the point conveyance velocity, we assumed a hydraulic radius corresponding to a sheet flow condition with a depth of 5 cm—a value that should be linked to the rainfall intensity in future work.

At each point, the long-term recharge rate, r, was approximated as \( r = R(1 - \frac{ET}{R}) \), where R represents annual rainfall on a daily average basis, and ET/R is the evapotranspiration-to-rainfall ratio. This ratio ET/R was retrieved from the Budkyo curve equation \( ET/R = \{D_I[1 - \exp(-D_I) \tanh(1/D_I)]\}^{0.5} \), where \( D_I = ET_{\text{max}}/R \) was the dryness index based on the annual potential evapotranspiration \( ET_{\text{max}} \) (Rodríguez-Iturbe & Porporato, 2004). These annual values of \( ET_{\text{max}} \) and R, both on a daily average basis, were obtained from the Daymet dataset at 1-km resolution for 2022.
3.1.2 Peak flow model

To determine the maximum flood extents, we replaced the time varying flow, $Q(t)$, of $\pi_2$ with a peak flow $Q_{\text{max}}$ taken from a flow hydrograph specific to each storm event (Fig. 1). This flow hydrograph captured the spatial runoff variability based on a semi-distributed rainfall-runoff model, where the unit-area runoff, $Q$, varied over three areas:

\begin{align}
Q_{\beta} &= R F_t(\bar{S}, \bar{R}) + \bar{R}(1 - F_t(\bar{S}, \bar{R}))(1 - \bar{c}) \\
Q_{(1-\beta)F_t} &= \bar{R} \\
Q_{(1-\beta)(1-F_t)} &= 0
\end{align}

where $\bar{R}$ is unit area rainfall, and over a fraction of the watershed $\beta$, runoff occurs prior to saturation, while over the fraction of area, $(1 - \beta)$, runoff only occurs over the saturated fraction of area, $F_t$ (Bartlett et al., 2016a, 2016b, 2017). Note that the overall unit area runoff, $\beta Q_{\beta} + (1 - \beta)F_t Q_{(1-\beta)F_t}$, is equal to the general runoff equation (10) of Bartlett et al. (2016b). The fraction of saturated area, $F_t$, is calculated based on a hydrology model, e.g., the variable infiltration capacity (VIC) model, TOPModel, and the NRCS-CN approach (Bartlett et al., 2016a, 2016b, 2017).

For this study, the fraction of area $\beta$ reasonably was represented by streets and other impervious areas. In these areas, it was assumed that the antecedent moisture deficit, $\bar{c}$, was negligible, i.e., $\bar{c} \approx 0$. Thus, within the fraction of area $\beta + (1 - \beta)F_t$, the runoff was simply $\bar{R}$. With these assumptions, the peak flow was calculated as,

$$Q_{\text{max}} = \max \left[ a_{cr} \int_0^t f_{(\bar{c}+(1-\beta)F_t)}(t-\tau) \bar{R}(\tau) d\tau \right],$$

where $f_{(\bar{c}+(1-\beta)F_t)}(\cdot)$ is the geomorphological instantaneous unit hydrograph (GIUH) for the fraction of watershed area, $\beta + (1 - \beta)F_t$, where the unit area runoff is the unit area rainfall, $\bar{R}$, $a_{cr}$ is the runoff producing area, i.e., $\beta + (1 - \beta)F_t$ multiplied by the watershed area, and $\tau$ is the time from the beginning of the storm. For Eq. (21), the GIUH, $f_{(\bar{c}+(1-\beta)F_t)}(\cdot)$, was based on the travel times of the watershed points within the fraction of area $\beta + (1 - \beta)F_t$. Travel times were calculated using average water velocities in the hillslope and channel areas (Rigon et al., 2011). In each area, the respective averages were over the point values of the conveyance velocities calculated in Section 3.1.1.

While the fraction of area $\beta$ was based on streets and other impervious areas (based on impervious descriptor map of the NLCD), the saturated fraction of area $F_t$ was mapped to the watershed points based on $\Pi_2$ because it is akin to the topographic wetness index (TWI). From the values of $\Pi_2$ over the fraction of area $1 - \beta$, we calculated the empirical quantile function, $P_{\Pi_2}^{-1}(\cdot)$. In turn, we calculated the quantile of $\Pi_2$ that corresponded to the saturated fraction of the watershed, i.e., $P_{\Pi_2}^{-1}(1 - F_t)$, and spatially mapped the saturated area as those points (over the fraction of area $1 - \beta$) where $\Pi_2$ was greater than this quantile value, i.e.,

$$\Pi_2 \geq P_{\Pi_2}^{-1}(1 - F_t),$$

where $F_t$ was calculated with Eq. (24) of Bartlett et al. (2016a) for the extended NRCS-CN method.

For any unit-area precipitation hyetograph, the peak flow was defined through Eq. (21). In turn, the maximum flow roughly was inferred as a function of the contributing area (Nardi et al., 2006; Rigon et al., 2011), i.e.,
The relationship between peak flow, $Q_{\text{max}}$, and the runoff producing contributing area, $a_{c_r}$, is based on a series of $Q_{\text{max}}$ values calculated with Eq. (21). Each $Q_{\text{max}}$ is derived from 1) a GIUH specific to the runoff producing area, $a_{c_r}$, and 2) a rainfall hyetograph based on the 100-year rainfall event (Table S1) and the 2nd quartile, 50th decile event temporal distribution from NOAA Atlas 14.

$$Q_{\text{max}} = c_1 (a_{c_r})^{c_2}$$

(23)

where the contributing area $a_{c_r}$ was derived only accounting for the area producing runoff, i.e., the fraction of watershed area $\beta + (1 - \beta)F_{t}$, and $c_1$ and $c_2$ were found by fitting the equation to multiple (watershed specific) data points derived from the area and flow relationship of Eq. 21 (Fig. 4).

Consequently, for each HUC 12 watershed, we divided the watershed into a series of nested watersheds down to an area of approximately 10 km$^2$, and calculated a series of data points that related the contributing area to the maximum flow. Subsequently, we found $c_1$ and $c_2$ from the best-fitting line and then mapped the contributing area to the peak flow based on Eq. (23) (Rigon et al., 2011). Generally, a log-linear relationship adequately described how the peak flow scaled with the contributing area for each HUC 12 watershed (Fig. 4). Precipitation inputs to the watersheds were estimated to be consistent with the FEMA HEC-RAS approach (Table S2). The precipitation hyetographs
for the 10-, 100-, and 1000-yr. events were estimated from the NOAA Atlas 14 24-hour event for both the total rainfall amount and the temporal distribution, which was the second quartile, 50th decile event. This precipitation hyetograph was then input into the $Q_{\text{max}}$ flow model of Eq. (21) with the saturated fraction of area based on the CNs of Table S2 and the extended NRCS-CN method of Bartlett et al. (2016a).

### 3.1.3 Channel properties

The channel network was delineated based on a specified flow accumulation threshold, culminating in the assignment of a flow capacity and hydraulic parameters at each watershed point. Channel delineation thresholds for local and non-local scales were 0.01 mi² and 1 mi², respectively. For the respective channel networks, each watershed point was related to a channel flow capacity when the channel stage equaled the HAND of the point. This capacity was based on $n$ and $S_o$ from the nearest downslope channel centerline and $A_c$ and $R_h$ respectively calculated as the product of the reach averaged hydraulic values (at the HAND depth) and an adjustment factor based on the channel geometry specific to the point (Appendix A). Note that the reach averaged hydraulic values were calculated based on the overall, respective nonlocal and local channel networks (Zheng, Maidment, et al., 2018; Garousi-Nejad et al., 2019). For channel networks extracted from the DEM, the inclusion of bathymetry would alter the extracted flow paths and provide a more accurate representation of the channel centerlines.

### 3.1.4 Indices at different scales

The dimensionless indices $\pi_2$, $\pi_3$, $\pi_4$, and $\Pi_2$ were calculated at different scales: 1) at each point, i.e. $\pi_2(p)$ and $\pi_4(p)$; 2) in relation to localized flow paths (e.g. streets), i.e. $\pi_2(l)$, $\pi_3(l)$, $\pi_4(l)$, and $\Pi_2(l)$; and 3) in relation to major streams and rivers that can create non-localized flooding, i.e. $\pi_2(nl)$, $\pi_3(nl)$, $\pi_4(nl)$, and $\Pi_2(nl)$. Note that $\pi_3$ and $\Pi_2$ do not exist at a point because the height above drainage is nonexistent without reference to a stream. An example of indices calculated at the 3 different scales is shown on Fig. 5.

The dimensionless indices were generally not collinear (Figure S2). Of the 45 pairwise correlations, only 2 were greater than 0.6. From the channel hydraulics formulation, the ratio of gravitational to frictional force at the point scale ($\pi_4(p)$) and the channel expansion ratio at the non-local scale ($\pi_3(nl)$) were both correlated with the hydrologic index TWI at the local scale ($\Pi_2(l)$) (0.66 and 0.73, respectively). The low correlations of the indices made them suitable for an ML application (Murphy, 2022).

### 3.2 Machine Learning Methodology

#### 3.2.1 Pluvial flooding ground-truth (label) data

In this study, we benchmarked ML training and validation using FEMA’s available rainfall/runoff inputs and flood hazard output data from 2D HEC-RAS models, which were run with a spatially uniform but time-varying runoff hyetograph. These pluvial flood risk data represent a relatively recent addition to FEMA’s inventory, which traditionally consists of approximately 1.2 million miles of fluvial stream studies (Government Accountability Office, 2021). Since 2019, FEMA has been expanding its flood risk data by incorporating HEC-RAS 2D models, where runoff is applied over a 2D watershed area to enhance flood hazard assessments (Council, n.d.).

#### 3.2.2 Model training and performance

We hypothesized that using dimensionless input features would improve the generalization of the ML model. To test this hypothesis, watersheds from 2 distinct regions
Figure 5. For HUC 071200040403, the dimensionless indices at the different scales of 1) the nonlocal, major streams and rivers with a contributing area greater than 1 sq. mile, i.e., $\pi_2(nl)$, $\pi_3(nl)$, $\pi_4(nl)$, and $\Pi_2(nl)$, 2) the localized flow paths (e.g., streets) with a contributing area greater than 0.01 sq. mile, i.e., $\pi_2(l)$, $\pi_3(l)$, $\pi_4(l)$, and $\Pi_2(l)$, and 3) at each point, i.e., $\pi_2(p)$ and $\pi_4(p)$. The $\pi_2$ terms are based on $Q_{max}$ derived from the 100-year, 24-hour rainfall event of NOAA Atlas 14 (Table S1), the second quartile, 50th decile temporal distribution of the NOAA Atlas 14, and the extended NRCS-CN method of Bartlett et al. (2016a) (see Table S1).
were selected for model training and testing – 4 watersheds in the Chicago area located
in the Des Plaines and Skokie river watersheds (HUC 12 codes starting with 0712) and
4 watersheds in New Jersey in the Hackensack and Passaic river watersheds (with HUC
12 codes starting with 0203) (Figure S1, Table S1). Compared to the Chicago watersheds,
the New Jersey watersheds have a 34-percent decrease in medium- and high-density de-
veloped space, an approximately 80-percent increase in topographic slope (on a water-
shed average basis), and an increase in the density of streams (Table S1).

Since there was a spatial dependency of the data points, the train-test splits were
performed so that entire watersheds were either included or held out from training. For
the ML model training, We considered two combinations of these watersheds. In Case
1, all the Chicago watersheds were used for training, and the model was tested separately
on each of the New Jersey watersheds (50%-50% train-test split). In Case 2, one water-
shed from Chicago and one from New Jersey were held out for testing, while the remain-
ing six watersheds were used for training (75%-25% train-test split). In each case, a
logistic regression model was built for each of the the 10-, 100-, and 1000-year events (Cox,
1958), and each model captured different rainfall ranges: 3.5–6 inches for the 10-year event,
6–8.5 inches for the 100-year event, and 8.5–13.5 inches for the 1000-year event (see Ta-
ble S2). To benchmark the performance of dimensionless features, logistic regression mod-
els also were developed using the dimensional features. Both the dimensional and dimen-
sionless features were standardized, i.e., centered around a mean value and scaled to a
unit variance.

Logistic regression was selected for the ML model to maintain tractability, inter-
pretability, and to isolate the benefit of the dimensionless features from the selection of
the ML model. By using this simpler ML model, we avoided confounding the effects of
the dimensionless features with those of a more sophisticated ML model. The logistic
regression models were trained by adjusting model weights to maximize the performance
of the algorithm in matching the ground truth data (HEC-RAS-based flood maps). First,
the weights were selected to minimize the log-loss with an L1 regularization penalty (Murphy,
2022). Secondly, a discrimination threshold used to classify the points as flooded (class
1) or not flooded (class 0) was estimated to best match to the ground truth flood maps.
Finally, this optimal threshold was applied to classify the points within the test water-
sheds.

Performance was assessed using two metrics: the receiving operator characteris-
tic (ROC) curve and the \( F_\beta \)-score. For the ROC curve, performance was summarized
as the area under the curve (AUC) (Murphy, 2022). Generally, an AUC score of 0.8 to
0.9 is considered excellent discrimination, while an AUC score greater than 0.9 is con-
sidered outstanding (Hosmer Jr et al., 2013). An AUC score of 0.5 provides no discrim-
ination between flooded and non-flooded areas and is no better than a random guess.
For the \( F_\beta \)-score, the recall was weighted as twice as important as the precision (i.e., \( \beta = \)
2, see Appendix B). The choice of \( \beta = 2 \) prioritizes the minimization of false negatives.
False negatives are particularly costly in flood response as they may direct resources away
from flood-prone areas. The \( F_2 \) score ranges between 0 and 1, with higher values indi-
cating better performance. The \( F_2 \)-score depends on the choice of discrimination thresh-
old and is sensitive to class imbalance and may vary as event size decreases (i.e., there
is less flooded area). Differently, the AUC is independent of the discriminationthresh-
old and more robust to class imbalance; however, the AUC can give misleadingly high
scores for imbalanced datasets by giving too much weight to the negative class. For such
highly imbalanced datasets, the \( F_2 \) score is better for assessing the detection of the rare
positive class (e.g., floods) because the \( F_2 \) score emphasizes recall.
4 Results

4.1 Model performance with dimensionless features

The dimensionless feature ML model performed well according to the AUC scores (Table S3, Figure S3). Across all events (e.g., 10-, 100-, and 1000-year), the average AUC scores for dimensionless and dimensional features were, respectively, 0.89 and 0.84 for Case I and 0.89 and 0.87 for Case II, respectively. AUC scores ranged from 0.75 to 0.96. The model consistently performed better with dimensionless features compared to with dimensional features. This improvement in model performance was most pronounced for the 10-year storm event, with the benefit decreasing for higher return periods as flooding became more widespread. Across storm event return periods, the dimensionless feature model AUC values decreased from 0.9-0.95 (10-yr.) to 0.83-0.90 (100-yr.) and 0.83-0.92 (1000-yr.).

The $F_2$ score is preferred over the AUC for highly imbalanced datasets when detecting the rare positive class (i.e., flooded areas) because it weights the recall more than the precision, i.e., the $F_2$ prioritizes the correct classification of flooded areas. Across all events, the dimensionless feature $F_2$ scores ranged from 0.48 to 0.85, while the dimensional feature $F_2$ score ranged from 0.37 to 0.85. The average $F_2$ scores for the dimensionless and dimensional features were 0.65 and 0.62 for Case I and 0.62 and 0.55 for Case II, respectively (Table 3). Accordingly, the $F_2$ scores indicated better model performance with dimensionless features with an average $F_2$ increase of 9% in Case I and 15% in Case II. Unlike the AUC, the $F_2$ scores indicated better performance for the larger return period events (e.g., 100-year and 1000-year). However, it is important to note that $F_2$ is sensitive to class imbalance, and differences in $F_2$ between events may result from the decreasing class imbalance (i.e., more flooded area) as the event return period increases.

$F_2$ score improvements with the dimensionless features varied across watershed and storm rainfall depth. $F_2$ increased more for the 10-yr. storm than the 100- and 1000-yr. storms on average (20.2%, 5.3%, and 7.3% average increases, respectively). Notably, for watershed 020301030804, the $F_2$ increased by 21.4% for the 100-year storm and 37.5% for the 1000-yr. storms. Out of the 18 dimensionless and dimensional $F_2$ score comparisons (Table 3), half showed no more than a 5% difference between the dimensionless and dimensional $F_2$ scores; however, in the other half, the dimensionless features provide an average $F_2$ uplift of 23%. This average uplift of 23% showcases how the dimensionless features improve the ML model generalization across watersheds.

In Case I where the model was trained in Chicago and tested in New Jersey, the dimensionless features provided an average $F_2$ uplift of 16% for the 10-year storm, 2.8% for the 100-year storm, and 4.4% for the 1000-year storm (Table 3). For Case II, in which training data came from both regions, the dimensionless features provided an even greater uplift in the $F_2$ score of 27.7% for the 10-year storm, 10.1% for the 100-year storm, and 7% for the 1000-year storm (Table 3). In both cases, the dimensionless features improved ML performance most for the 10-yr. storm and for the watersheds that performed worst with dimensional features. For example, $F_2$ for watershed 020301030804 increased from 0.42 to 0.51 for the 100-yr. storm (and 0.44 to 0.55 for the 1000-yr. storm). Therefore, dimensionless features provide an advantage compared to dimensional features with respect to generalization, especially for watersheds with poor performance.

For both Case I and Case II, the optimal discrimination thresholds (Table 3) were relatively similar between the dimensionless and dimensional cases and varied more across storm rainfall depths, as expected. Compared to dimensional features, the dimensionless features tended to increase the true negative (TN) and decrease the true positive (TP) rates. As an example, for the 100-yr. event, the TN rate increased from 0.82 to 0.90 with dimensionless features (Figures 4 and S5). The TP rate decreased from 0.72 to 0.65. Therefore, the dimensionless features improve the ability of the model to correctly pre-
Table 3. $F_2$ scores across 10-yr., 100-yr., and 1000-yr. return period events for Case I and Case II. In Case I, the model was trained on Chicago HUC 12 watersheds and tested on New Jersey HUC 12 watersheds. In Case II, the model was trained on 3 watersheds from each region (6 total) and tested on the remaining 2 watersheds.

|                      | 10-yr. |   | 100-yr. |   | 1000-yr. |   |
|----------------------|--------|---|---------|---|----------|---|
|                      | Dim.-less | Dim. | Dim.-less | Dim. | Dim.-less | Dim. |
| Discrimination Threshold | 0.82  | 0.79  | 0.57  | 0.60  | 0.52  | 0.56  |
| Case I $F_2$          |        |      |        |      |        |      |
| 020301030606          | 0.70  | 0.61  | 0.80  | 0.83  | 0.85  | 0.85  |
| 020301030503          | 0.58  | 0.49  | 0.69  | 0.7  | 0.72  | 0.74  |
| 020301030703          | 0.61  | 0.45  | 0.57  | 0.6  | 0.61  | 0.64  |
| 020301030804          | 0.66  | 0.68  | 0.51  | 0.42  | 0.55  | 0.44  |
| Case II $F_2$         |        |      |        |      |        |      |
| 071200030101          | 0.49  | 0.37  | 0.61  | 0.52  | 0.68  | 0.62  |
| 020301030503          | 0.48  | 0.39  | 0.70  | 0.68  | 0.74  | 0.71  |

dict areas without flooding, while decreasing the ability of the model to correctly predict areas with flooding.

Fig. 6 (Case I) and 7 (Case II) show the ML prediction of flooding compared to the labels derived from 2D HEC-RAS simulations. The ML model captures the spatial pattern of flooding better for the 10-yr. storm compared to the 100- and 1000-yr. storms. The ML model tended to underpredict flood extent (i.e., predict no flooding where the HEC-RAS model predicted flooding). Despite the underpredictions, the ML model generally captured the networked pattern of flash flooding from the HEC-RAS 2D models.

4.2 Sensitivity to non-local scale threshold

To study the sensitivity of model performance to the non-local scale threshold, we decreased the contributing area threshold for the non-local flow paths from 1 sq. mile to 0.2 sq. mile and evaluated the absolute change in the AUC (Table 4). Generally, performance decreased, with the most significant performance decreases occurring for the 10-year event (Table 4). The 10-year event has the most dramatic performance decrease because most of the flooding for a 10-year event is centered around the major rivers and streams, and a lower non-local threshold (of 0.2 sq. miles) does not define major rivers and streams well. The magnitude of the change in some cases demonstrates the importance of optimizing the thresholds for flow path delineation in relation to the performance of the ML model.

5 Discussion

Recent studies have demonstrated that dimensionless features can improve ML generalization (Gunaratnam et al., 2003; Oppenheimer et al., 2023) and computational efficiency (Hu et al., 2019). Dimensionless features improve generalization by redefining the training space such that a larger number of test cases fall within the training set (Oppenheimer et al., 2023). This reduces extrapolation for new cases. In a similar vein, Hu et al. (2019) used proper orthogonal decomposition (POD) and singular value decomposition (SVD) to define a reduced-order space in which a long short-term memory (LSTM) model was trained to predict flood depth. They noted improved computational efficiency as the LSTM was trained with a smaller number of features. In contrast, we use the Buckingham II Theorem to define dimensionless input features in a reduced-order space and show that
Figure 6. For the Case I test HUC 12 watershed of 020301030503, the ML predictions versus the HEC-RAS generated flood maps (labels) for different return period events, where the discrimination threshold (probability above which a point is flooded) for the 10-year, 1000-year, and 10000-year events are 0.82, 0.57, and 0.53, respectively. The difference is calculated as prediction minus label.
Figure 7. Case II test HUC 12 watershed of 071200030101, the ML predictions versus the HEC-RAS generated flood maps (labels) for different return period events, where the discrimination threshold (probability above which a point is flooded) for the 10-year, 100-year, and 1000-year events are 0.84, 0.58, and 0.53, respectively. The difference is calculated as prediction minus label.
Table 4. Sensitivity of AUC to the local scale threshold. Absolute change in AUC after reducing the non-local contributing area threshold from 1 mi$^2$ to 0.2 mi$^2$.

| Case I          | Test AUC change | 10-yr. | 100-yr. | 1000-yr. |
|-----------------|-----------------|--------|---------|----------|
| 020301030606    | -0.01           | -0.01  | -0.01   |
| 020301030503    | -0.05           | -0.02  | -0.02   |
| 020301030703    | -0.08           | -0.01  | -0.01   |
| 020301030804    | -0.08           | 0      | 0       |

Case II

| 071200030101 | -0.14 | -0.05 | -0.05 |
| 020301030503 | -0.05 | -0.02 | -0.03 |

dimensionless features perform better than traditional dimensional features with respect to model generalization between regions. Our model is also more computationally efficient as 9 input layers (e.g., for each of the local and non-local features) are combined into 4 dimensionless features prior to model training. This is the first study, to our knowledge, to apply Buckingham II for ML in hydrology and we anticipate wide applicability in flood and streamflow forecasting.

The dimensionless features performed comparable to previous ML flood modeling efforts (e.g., Manfreda, Nardi, et al. (2014); Samela et al. (2016)), even when the model was trained and tested in different regions (e.g., Chicago and New Jersey). Therefore, the use of dimensionless features in ML flood modeling has potential to improve model generalization across terrains. Other approaches have been proposed to address model generalization, such as including multi-scale contextual terrain information (Cache et al., 2024), reduced-order feature selection (Hu et al., 2019; Pakdehi et al., 2023), transfer learning (Seleem et al., 2022), and the application of convolutional neural networks (CNNs) (Guo et al., 2022). Here, we used logistic regression to isolate and maintain focus on performance of the dimensionless features. However, we anticipate that multi-layer neural networks would improve model performance.

The dimensionless indices when coupled to the ML model allow for a rapid prediction of pluvial (flash flood) extents. The dimensionless features are largely based on static terrain data (e.g., DEM, land cover, stream network) and, therefore, only need to be processed once. The processing of the dimensionless data takes about 2-3 hours for DEM resolutions of 1- to 3-meters. Once the derivative data are created, they can be combined with runoff hyetographs in seconds and new ML results produced within minutes. Thus, runoff hyetographs from the latest weather forecasts (e.g., from the High-Resolution Rapid Refresh (HRRR) model, (Dowell et al., 2022)) could be rapidly transformed into maps that inform flash flood warnings. Such rapid, high-resolution mapping would be an improvement over current flash flood warnings that generally are given on a regional basis without specificity. This lack of specificity makes it difficult to truly understand the best actions to preserve life and maintain safety during a flash flood warning.

The definition of the dimensionless $\pi$ and $\Pi$ terms is not unique — multiple valid formulations exist under Buckingham II theorem. Future work could explore and quantify the performance of different formulations of the dimensionless indices. Identifying further performance improvements could involve 1) exploring optimal thresholds for delineating flow paths at different scales to better capture flood processes (Sangireddy et al., 2016) and 2) resolving reach-averaged hydraulics at a subwatershed level within each HUC 12 watershed. Currently, reach-average hydraulics are calculated across entire HUC 12 watershed. Additionally, advanced image processing techniques could be integrated into the model, either as a pre-processing step or as part of the machine learning pro-
cess itself. A fully convolutional neural network (e.g., the U-Net ML model for image segmentation Ronneberger et al. (2015)) could further improve the flash flood prediction by capturing complex spatial patterns, offering a more flexible alternative to the current logistic regression framework.

The runoff estimation technique used in this study could improve regression analyses commonly used to estimate stream flow. Specifically, the geomorphological instantaneous unit hydrograph (GIUH) when convolved with a runoff hyetograph provides a watershed wide signature of the flow versus the contributing area (e.g., Fig. 4). The GIUH approach relies on data that are readily available at continental and global scales (e.g., DEM and land cover), suggesting that integration of GIUH attributes into streamflow estimation could be widely beneficial (Capesius & Stephens, 2009). The performance of the ML model, which leverages the GIUH, shows that the GIUH has promise in estimating flow over regional areas.

Lastly, a number of assumptions were employed here that warrant further study. First, a small set of training and testing watersheds was used here to test the hypothesis that the dimensionless features improved model generalization. Secondly, the model was trained and tested on HEC-RAS model output, not observed flood events. Lastly, models were generalized only across regions and not across storm depths. Future testing and development of dimensionless features for flood modeling should expand to a larger number of watershed contexts, test against observed flood extents, and address generalization across events.

6 Concluding Remarks

Dimensionless features can improve ML model performance, especially when generalizing to never-before-seen conditions. By reformulating a physical problem into a dimensionless context with Buckingham II theorem, we may capture the similarity of a process across different environments, conditions, and scales. With respect to flood mapping with ML, this approach can be expanded to other regions and catchments, a wider range of storm events, and other flooding mechanisms such as riverine, coastal, and compound flooding. We also anticipate that this approach will find general utility in assessing different geohazard risks.

The concepts described here can be applied to the broader area of reproducibility in ML models, which includes feature engineering, generalizability, interpretable ML, and efficiency and scalability. The physical processes described by the input features often involve relationships between variables that are maintained across different scales and conditions (e.g., climate, fluid dynamics, financial systems). Derivation of these relationships from fundamental physical principles (via the Buckingham II theorem) is an innovative method to improve model transferability. Furthermore, this approach can be considered more interpretable, as the input features are meaningfully constrained by the underlying physics, represent known controls on flooding extent, and are dimensionally independent. Lastly, models with a minimal number of features that generalize better will improve efficiency and scalability. Models with reduced computational demand while maintaining accuracy and generalization can be deployed in real-time scenarios. The methods described in the article not only advance flood inundation mapping but also offer insights into general ML practices. The proposed framework provides a path forward for developing more robust, scalable, and reproducible ML models that can be applied to diverse and dynamic environments.

Appendix A Hydraulic Radius Adjustment factor

The hydraulic radius is the cross sectional area divided by the wetted perimeter. Accordingly, the factor adjusting the hydraulic radius, \( a_{R_h} \), is the factor for adjusting
the cross sectional area, \( \alpha_A = \frac{2y_d}{w} \), multiplied by a factor for adjusting the wetted perimeter. Here, the wetted perimeter factor is a weighted combination of the factors for a triangular cross section, \( \alpha_T \), and a parabolic cross section, \( \alpha_P \), i.e.,

\[
\alpha_R = \begin{cases} 
\alpha_T \alpha_A & A < A_T \\
\alpha_T \alpha_A \left( \frac{A - A_P}{A_T - A_T} \right) + \alpha_P \alpha_A \left( 1 - \frac{A - A_T}{A_P - A_T} \right) & A_T \leq A < A_P \\
\alpha_P \alpha_A & A_P \leq A
\end{cases}
\]

where the overall factor varies linearly depending on how the cross sectional area, \( A \), compares to an assumed triangular cross section, \( A_T = y_d z_d \), and parabolic cross section, \( A_P = \frac{2}{3} y_d z_d \)–both calculated based on the distance from drainage, \( y_d \), and height above drainage, \( z_d \). The wetted perimeter factors for the triangular cross section, \( \alpha_T \), and parabolic cross section, \( \alpha_P \), respectively are (Chow, 1959)

\[
\alpha_T = \frac{2z_d \sqrt{\frac{w}{z_d} + 1}}{2z_d \sqrt{\frac{2y_d}{z_d} + 1}} \quad \text{(A1)}
\]

\[
\alpha_P = \frac{w + \frac{8y_d^2}{3z_d^2}}{2y_d + \frac{8y_d^2}{3z_d^2}} \quad \text{(A2)}
\]

where each factor is the wetted perimeter based on the reach averaged width, \( w \), divided by the wetted perimeter based on the channel width (estimated as \( 2 \cdot y_d \)).

**Appendix B  \( F_\beta \)-score**

Binary classification produces predictions of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). A successful statistical inference model maximizes TP and TN relative to TP and TN. These four categories (i.e., TP, TN, FP, and FN) are preferred over raw accuracy for imbalanced datasets such as flood datasets where, in most cases, dry areas exceed flooded areas in extent. The raw accuracy is calculated as

\[
\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{All data points}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad \text{(B1)}
\]

Assessing model performance on the raw accuracy of Equation B1 is inappropriate for imbalanced data because it allows for models to perform poorly on the minority class. For example, if a geospatial dataset is 5% flooded by extent, one can train a model with 95% accuracy by predicting all pixels as dry.

The \( F_\beta \)-score is an improvement on raw accuracy for imbalanced data. Once the model is trained, it is evaluated by calculating the \( F_\beta \)-score on the test set, which weights recall as \( \beta > 0 \) more significant than precision, i.e., (Murphy, 2022)

\[
F_\beta = \frac{1}{\frac{1}{1+\beta^2} P + \frac{1}{1+\beta^2} R}, \quad \text{(B2)}
\]

where the precision, \( P \), is written as

\[
P = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad \text{(B3)}
\]
and the recall, $R$, is written as

$$R = \frac{TP}{TP + FN}. \quad (B4)$$

When the precision and recall are equally weighted, i.e., $\beta = 1$, then we retrieve the $F_1$-score that is the harmonic mean of the precision, $P$, and recall, $R$, i.e.,

$$F_1 = \frac{2PR}{P + R}. \quad (B5)$$

When the recall is weighted twice as important as the precision, i.e., $\beta = 2$, then we retrieve the $F_2$-score,

$$F_2 = \frac{PR}{0.8P + 0.2R}. \quad (B6)$$

Equations B3 and B4 both have TP in their numerators and denominators which suggests that the $F_\beta$ score prioritizes the positive class over the negative one. This is appropriate when the imbalance in the data is in the negative class’ favor; this is when the positive class is in the minority. More intuitively, the recall, $R$, also may be thought of as a ‘hit rate’, i.e., the fraction of predictions that capture flooding accurately in comparison to the assumed ground truth data, while 1 minus the precision, $1-P$, may be thought of as a ‘false alarm rate’, i.e., the fraction of predictions that falsely predicted flooding in comparison to the assumed ground truth data.

**Acknowledgments**

Funding for this work was provided by Stantec and The Water Research Foundation (WRF) project 5084, *Holistic and Innovative Approaches for Flood Mitigation Planning and Modeling under Extreme Wet Weather Events and Climate Impacts*. We thank Harry Zhang of WRF for his support and encouragement.

**Data Availability Statement.** The data supporting this manuscript is available at the locations shown in Table 2. The data processing was performed with GRASS GIS (GRASS Development Team, 2017), while the machine learning was performed using Scikit-learn (Pedregosa et al., 2011).

**References**

Ashley, S. T., & Ashley, W. S. (2008). Flood fatalities in the united states. *Journal of Applied Meteorology and Climatology, 47*(3), 805–818.

Bartlett, M. S., Parolari, A. J., McDonnell, J., & Porporato, A. (2017). Reply to comment by Fred L. Ogden et al. on “Beyond the SCS-CN method: A theoretical framework for spatially lumped rainfall-runoff response.”. *Water Resources Research, 53*(7), 6351-6354.

Bartlett, M. S., Parolari, A. J., McDonnell, J. J., & Porporato, A. (2016a). Beyond the SCS-CN method: A theoretical framework for spatially lumped rainfall-runoff response. *Water Resources Research, 52*(6), 4608–4627.

Bartlett, M. S., Parolari, A. J., McDonnell, J. J., & Porporato, A. (2016b). Framework for event-based semidistributed modeling that unifies the SCS-CN method, VIC, PDM, and TOPMODEL. *Water Resources Research, 52*(9), 7036–7052.

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. *Hydrology and Earth System Sciences Discussions, 2022*, 1–50.
Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of hydrology, 320*(1), 18–36.

Beven, K. (2012). *Rainfall-runoff modelling: The primer*. Wiley.

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. *Journal of hydrology, 249*(1), 11–29.

Beven, K., & Kirkby, M. (1979). A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d’appel variable de l’hydrologie du bassin versant. *Hydrological Sciences Journal, 24*(1), 43–69.

Bui, D. T., Fanahi, M., Shahabi, H., Singh, V. P., Shirzadi, A., Chapi, K., . . . others (2018). Novel hybrid evolutionary algorithms for spatial prediction of floods. *Scientific reports, 8*(1), 1–14.

Cache, T., Gomez, M. S., Beucler, T., Blagojevic, J., Leitao, J. P., & Peleg, N. (2024). Enhancing generalizability of data-driven urban flood models by incorporating contextual information. *Hydrology and Earth System Sciences Discussions, 2024*, 1–23.

Capesius, J. P., & Stephens, V. C. (2009). *Regional regression equations for estimation of natural streamflow statistics in colorado*. US Department of the Interior, US Geological Survey.

Chow, V. T. (1959). Open-channel hydraulics. *McGraw-Hill civil engineering series*.

Collins, E. L., Sanchez, G. M., Terando, A., Stillwell, C. C., Mitasova, H., Sebastian, A., & Meentemeyer, R. K. (2022). Predicting flood damage probability across the conterminous united states. *Environmental Research Letters, 17*(3), 034006.

Cornwall, W. (2021). *Europe’s deadly floods leave scientists stunned*. American Association for the Advancement of Science.

Council, T. M. A. (n.d.).

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 20*(2), 215–232.

D’Odorico, P., & Rigon, R. (2003). Hillslope and channel contributions to the hydrologic response. *Water resources research, 39*(5).

Dowell, D. C., Alexander, C. R., James, E. P., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., . . . others (2022). The high-resolution rapid refresh (hrrr): An hourly updating convection-allowing forecast model, part i: Motivation and system description. *Weather and Forecasting, 37*(8), 1371–1395.

Ducharme, A. (2009). Reducing scale dependence in topmodel using a dimensionless topographic index. *Hydrology and Earth System Sciences, 13*(12), 2399–2412.

Ercan, A., Kavvas, M. L., & Haltas, I. (2014). Scaling and self-similarity in one-dimensional unsteady open channel flow. *Hydrological Processes, 28*(5), 2721–2737.

Garonsi-Nejad, I., Tarbonton, D. G., Aboutalebi, M., & Torres-Ruiz, A. F. (2019). Terrain analysis enhancements to the height above nearest drainage flood inundation mapping method. *Water Resources Research, 55*(10), 7983–8009.

Giovannettone, J., Copenhaver, T., Burns, M., & Choquette, S. (2018). A statistical approach to mapping flood susceptibility in the lower connecticut river valley region. *Water Resources Research, 54*(10), 7603–7618.

Government Accountability Office. (2021). Better planning and analysis needed to address current and future flood hazards. (GAO-22-104079).

GRASS Development Team. (2017). Geographic resources analysis support system (grass gis) software, version 7.2 [Computer software manual]. Retrieved from http://grass.osgeo.org

Gumaratnam, D. J., Degroff, T., & Gero, J. S. (2003). Improving neural network models of physical systems through dimensional analysis. *Applied Soft Computing, 2*(4), 283–296.
Guo, Z., Moosavi, V., & Leitão, J. P. (2022). Data-driven rapid flood prediction mapping with catchment generalizability. *Journal of Hydrology, 609*, 127726.

Hocini, N., Payrastre, O., Bourgin, F., Gaume, E., Davy, P., Lague, D., ... Pons, F. (2021). Performance of automated methods for flash flood inundation mapping: a comparison of a digital terrain model (dtm) filling and two hydrodynamic methods. *Hydrology and Earth System Sciences, 25*(6), 2979–2995.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Hosseiny, H., Nazari, F., Smith, V., & Nataraj, C. (2020). A framework for modeling flood depth using a hybrid of hydraulics and machine learning. *Scientific Reports, 10*(1), 1–14.

Hu, R., Fang, F., Pain, C., & Navon, I. (2019). Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method. *Journal of Hydrology, 575*, 911–920.

Ivanov, V. Y., Xu, D., Dwelle, M. C., Sargsyan, K., Wright, D. B., Katopodes, N., ... others (2021). Breaking down the computational barriers to real-time urban flood forecasting. *Geophysical Research Letters, 48*(20), e2021GL093585.

Jakeman, A., & Hornberger, G. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research, 29*(8), 2637–2649.

Jasiewicz, J., & Metz, M. (2011). A new grass gis toolkit for hortonian analysis of drainage networks. *Computers & Geosciences, 37*(8), 1162–1173.

Johnson, J. M., Munasinghe, D., Eyelade, D., & Cohen, S. (2019). An integrated evaluation of the national water model (nwm)–height above nearest drainage (hand) flood mapping methodology. *Natural Hazards and Earth System Sciences, 19*(11), 2405–2420.

Maidment, D., Olivera, F., Calver, A., Eatherall, A., & Fraczek, W. (1996). Unit hydrograph derived from a spatially distributed velocity field. *Hydrological processes, 10*(6), 831–844.

Manfreda, S., Nardi, F., Samela, C., Grimaldi, S., Taramasso, A. C., Roth, G., & Sole, A. (2014). Investigation on the use of geomorphic approaches for the delineation of flood prone areas. *Journal of hydrology, 517*, 863–876.

Manfreda, S., Samela, C., Sole, A., & Fiorentino, M. (2014). Flood-prone areas assessment using linear binary classifiers based on morphological indices. In *Vulnerability, uncertainty, and risk: Quantification, mitigation, and management* (pp. 2002–2011).

Meles, M. B., Younger, S. E., Jackson, C. R., Du, E., & Drover, D. (2020). Wetness index based on landscape position and topography (wilt): Modifying twi to reflect landscape position. *Journal of environmental management, 255*, 109863.

Moussa, R., & Bocquillon, C. (2000). Approximation zones of the saint-venant equations flood routing with overbank flow. *Hydrology and Earth System Sciences, 4*(2), 251–260.

Murphy, K. P. (2022). *Probabilistic machine learning: an introduction*. MIT press.

Nachappa, T. G., Pirahelic, S. T., Gholamnia, K., Ghorbanyadeh, O., Rahmati, O., & Blaschke, T. (2020). Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using dempster shafer theory. *Journal of Hydrology, 590*, 125275.

Nardi, F., Vivoni, E. R., & Grimaldi, S. (2006). Investigating a floodplain scaling relation using a hydrogeomorphic delineation method. *Water Resources Research, 42*(9).

Oppenheimer, M. W., Doman, D. B., & Merrick, J. D. (2023). Multi-scale physics-informed machine learning using the buckingham pi theorem. *Journal of Computational Physics, 474*, 111810.

Pakdehi, M., Ahmadisharaf, E., Nazari, B., & Cho, E. (2023). Transferability of machine learning-based modeling frameworks across flood events for hindcasting.
maximum river flood depths in coastal watersheds. *Natural Hazards and Earth System Sciences Discussions*, 2023, 1–57.

Pedregosa, F., Varoquaux, G., Granfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Porporato, A. (2022). Hydrology without dimensions. *Hydrology and Earth System Sciences*, 26(2), 355–374.

Rahmati, O., Darabi, H., Panahi, M., Kalantari, Z., Naghibi, S. A., Ferreira, C. S. S., . . . others (2020). Development of novel hybridized models for urban flood susceptibility mapping. *Scientific reports, 10*(1), 1–19.

Rigon, R., D’Odorico, P., & Bertoldi, G. (2011). The geomorphic structure of the runoff peak. *Hydrology and Earth System Sciences, 15*(6), 1853–1863.

Rinaldo, A., Marani, A., & Rigon, R. (1991). Geomorphological dispersion. *Water Resources Research, 27*(4), 513–525.

Rodríguez-Iturbe, I., & Porporato, A. (2004). *Ecohydrology of water-controlled ecosystems: soil moisture and plant dynamics*. Cambridge University Press.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–miccai 2015: 18th international conference, munich, germany, october 5-9, 2015, proceedings, part iii* (pp. 234–241).

Rudolph, S., et al. (1998). On the context of dimensional analysis in artificial intelligence. In *International workshop on similarity methods*.

Samela, C., Manfreda, S., Paola, F. D., Giugni, M., Sole, A., & Fiorentino, M. (2016). Dem-based approaches for the delineation of flood-prone areas in an ungauged basin in africa. *Journal of Hydrologic Engineering, 21*(2), 06015010.

Sangireddy, H., Stark, C. P., Kladyzyk, A., & Passalacqua, P. (2016). Geonet: An open source software for the automatic and objective extraction of channel heads, channel network, and channel morphology from high resolution topography data. *Environmental Modelling & Software, 83*, 58–73.

Scriven, B. W. G., McGrath, H., & Stefanakis, E. (2021). Gis derived synthetic rating curves and hand model to support on-the-fly flood mapping. *Natural Hazards, 109*(2), 1629–1653.

Seleem, O., Ayzel, G., Bronstert, A., & Heistermann, M. (2022). Transferability of data-driven models to predict urban pluvial flood water depth in berlin, germany. *Natural Hazards and Earth System Sciences Discussions, 2022*, 1–23.

Tavares da Costa, R., Zanardo, S., Bagli, S., Hilberts, A. G., Manfreda, S., Samela, C., & Castellarin, A. (2020). Predictive modeling of envelope flood extents using geomorphic and climatic-hydrologic catchment characteristics. *Water Resources Research, 56*(9), e2019WR026453.

USACE Hydrologic Engineering Center. (2024). *Hec-ras 2d user’s manual*. The United States Government, US Army Corps of Engineers, Hydrologic Engineering Center. Retrieved from https://www.hec.usace.army.mil/confluence/rasdocs/r2dum/6.2/developing-a-terrain-model-and-geospatial-layers/creating-land-cover-mannings-n-values-and-impervious-layers

Wagenaar, D., Lüdtke, S., Schröter, K., Bouwer, L. M., & Kreibich, H. (2018). Regional and temporal transferability of multivariable flood damage models. *Water Resources Research, 54*(5), 3688–3703.

Wasko, C., Nathan, R., Stein, L., & O’Shea, D. (2021). Evidence of shorter more extreme rainfalls and increased flood variability under climate change. *Journal of Hydrology, 603*, 126994.

Yang, Z., Bai, F., & Xiang, K. (2019). A lattice boltzmann model for the open channel flows described by the saint-venant equations. *Royal Society open science, 6*(11), 190439.
Zheng, X., Maidment, D. R., Tarboton, D. G., Liu, Y. Y., & Passalacqua, P. (2018). Geoflood: Large-scale flood inundation mapping based on high-resolution terrain analysis. Water Resources Research, 54(12), 10–013.
Zheng, X., Tarboton, D. G., Maidment, D. R., Liu, Y. Y., & Passalacqua, P. (2018). River channel geometry and rating curve estimation using height above the nearest drainage. JAWRA Journal of the American Water Resources Association, 54(4), 785–806.