https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

Structured exploration of machine learning model complexity for spatio-temporal forecasting of urban flooding

Candace Agonafir[1,2], Tian Zheng[1,2]

[1]Learning the Earth with AI and Physics (LEAP) and Department of Statistics, Columbia University, New York, New York 10027

[2]National Science Foundation (NSF) Science and Technology Center, United States

Correspondence to: Candace Agonafir, candace.a@columbia.edu

Abstract

Urban flooding may lead to significant socio-economic impacts and loss of life. To afford preventative actions, researchers have implemented various modeling techniques to gain insight into urban flood occurrences. Using New York City (NYC) as the study area, data-driven techniques, specifically statistical and neural network models with increasing spatio-temporal complexity, are formulated and tested, assessing the potential relative contribution of different modeling constructs. Zones, based on flood characteristics, are first delineated using the unsupervised machine learning technique of spectral clustering. Then, the models are applied to each cluster, with comprehensive performance evaluation, as to understand which algorithmic, structural aspects contribute to the reduction of prediction errors. A chief discovery of this study is the emergence of the Graph Wavenet (GWN) as the most effective model due to its proficiency in capturing spatio-temporal aspects and implementing dynamic graph creation. Furthermore, it is seen that the enhancement of specific temporal and spatial components within a modeling technique proves beneficial, and a novel adoption of graph-based architectures is additive. Offering a unique exploration of spatio-temporal aspects, emphasizing the benefits of component enhancement and the adoption of graph-based architectures, this paper identifies modification techniques, which would allow for insights to prevail in urban flood modeling despite being confronted with limited data availability.

1. Introduction

Urban flooding, a natural disaster with dynamic ramifications, requires circumspect consideration. At the onset of rainfall, as runoff traverses through an urban setting, numerous obstacles, including an inability to permeate structures, sidewalks, and streets, and restricted entry into drainage systems due to debris blockages, are experienced. Compounding the issue, additional rainfall in a brief timeframe overwhelms even the unobstructed drains, preventing water admission and exacerbating overflow departure attempts. Consequently, runoff either persists on the streets or seeps through gaps in homes and buildings. In the resulting circumstance of urban flooding, the community is harmed. Most egregiously, deaths may occur,

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

as without adequate warning, individuals are not able to implement precautionary measures, such as avoiding outdoor travel or relocating to higher elevated areas when indoors. Veritably, this tragic outcome was demonstrated during the post-tropical depression Ida event, where many of 44 deaths within the New York City (NYC) metropolitan region were caused either by exposure to outdoor hazards, such as vehicular drownings or being swept away by the waters, or indoor below-ground dangers, such as drownings in flooded basements (Falconer, 2021; Plumer, 2021). Furthermore, in addition to fatalities, urban flooding exacts economic strain, as there may be destruction to the infrastructure, interruption to transportation services, and structural damage to the buildings and vehicles. Indeed, in the event of a large flooding disaster, direct costs may be incurred at the extent of billions of dollars; additionally, when examining smaller, frequent floods, long-term costs collect over the years by the chronic strains to the structural, plumbing and electrical systems (Agonafir et al., 2023). Hence, prioritizing the analysis of the diverse elements and influences on water behavior in an urban environment proves essential.

Now, as accurate flood forecasting models, allowing for the implementation of disaster deterrent measures, would offer significant health and financial relief, there is continuous progression in model development. Traditionally, hydrodynamic models have been widely used for flood prediction and risk assessment; however, the employment of these physics-based models is limited in certain metropolitans (Agonafir et al., 2023). Specifically, hydrodynamic models rely on extensive calculations to determine water flow, requiring detailed drainage network plans; as thus, in certain urban cities, such as NYC, where drainage details are unobtainable to researchers (Al-Suhili et al., 2019), the implementation of physics-based models becomes infeasible. Therefore, there has been a turn towards data-driven techniques to provide insight into water behavior when existing physical information is limited. With the provision of influencing variables, via statistical calculations and artificial intelligence (AI) capabilities, the models possess the ability to assess an occurrence and then create forecasts or ascertain vulnerabilities. Hence, the objective of understanding urban flooding is met without the need to simulate the exact water path. Ultimately, due to the convenience of use, data-driven techniques, particularly AI methods, have risen drastically in flood literature (Mosavi et al., 2018). Accordingly, an in-depth study into the efficiency of emerging AI techniques, within the field of urban flooding, affords complementation to the trend.

The mission of this study is formulated in consideration of the dire human and economic devastations of urban flooding, the modeling limitations due to data availability, and the recent advances in data-driven models to remedy the issues. Accurately assessing the intricacies of urban flood occurrence in NYC, by the employment of physical and crowdsourced data, this research provides a unique analysis of added components by presenting a cascade of statistical and neural network models, each with ascending complexities. In the exploration, a preliminary step involves the delineation of zones based on urban flood characteristics, using the unsupervised machine learning technique, spectral clustering. Then, a particular set of models, the Poisson Generalized Linear Regression (GLM), Feed Forward Neural Network (FFN), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Graph Convolutional Network (GCN) and the Graph Wavenet (GWN), is selected as to assess the benefits of auxiliary, advanced spatio-temporal aspects, dynamic graph creation and convolutional node-messaging

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

capabilities. The conduction of the experiment follows such that every model, receiving identical time-series input data and undergoing training (8 years) and testing (2 years) using the same set of dates, is tasked with the production of daily predicted street flooding counts for the testing period. By goodness-of-fit determinations, model comparisons are performed and dissected to discover the impacts of additional complexities. Therefore, the comprehensive examination of diverse models imparts invaluable modeling guidance towards urban flood research in the data-driven era.

Regarding existing urban flood literature, there is limited utilization of graph-structured artificial neural network (ANN) models. Moreover, of the few existing studies, an inclusion of pluvial, urban flooding is notably absent. For instance, in Farahmand et al., a spatial-temporal graph-based model (ASTGCN) for nowcasting in Harris County, Texas is developed (Farahmand et al., 2023). However, the examination is conducted on the singular flooding incident of Hurricane Harvey and its direct landfall onto the county, thereby effectively assessing the model's accuracy only in regards to coastal flooding. In contrast, the research of this paper delves into urban flooding over a 10-year duration, encompassing both large-scale events and persistent, frequent flash flood and pluvial occurrences. Furthermore, the primary objective of Farahmand et al is to highlight the proficiencies of the new model, rather than to conduct an in-depth exploration of the specific advantages of its added elements. In contrast, this research traverses a range of models, inspecting the benefits of each advancement. In another existing study, Wang et al, a graph-structed model is also created to benefit urban flood insights (Z. Wang et al., 2023). Nonetheless, the paper does not serve as a comparative analysis of varying models; also, the model developed identifies flood susceptibility, as opposed to producing forecasts. Finally, it is worth noting that Santos et al develops a graph-based, deep learning model for flood prediction (Oliveira Santos et al., 2023). Yet, the forecasts are for riverine flooding instead of urban flooding. As urban flooding involves multiple factors distinct from riverine flood variables (i.e., lack of infiltration and complex drainage networks), a model analysis specifically tailored towards urban concerns has more utility to metropolitan stakeholders. Therefore, this paper is the first of its kind to pioneer the adaptation of graph-based neural networks for urban, pluvial flooding, while conducting an expansive exploration of the spatio-temporal aspects within the domain.

This paper follows a structured sequence. In Section 2, the study initiates by discussing the study area of NYC and the urban characteristics which make it ideal for experimentation. Also, Section 2 demonstrates the data collection and pre-processing steps. Next, Section 3 delves, with extensive detail, into the methodology employed for each machine learning and statistical technique, laying a solid foundation for the thorough analysis presented in Section 4. Then, in Section 4, a detailed presentation and discussion of results unfold, exploring risk zones and model performances across diverse flood attributes and model features. Finally, in Section 5, conclusions are drawn, as the findings from the study are synthesized, offering a comprehensive overview of the outcomes and their implications for urban flood research.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

2. Data pre-processing

2.1 Study area

Situated along the northeastern coast of the United States, NYC emerges as the metropolitan landscape—distinctly impervious and densely populated. Lacking efficient infiltration with approximately 72% impervious cover, and encompassing a mere 800 square kilometers, while boasting roughly 8 million residents (City of New York, 2022; U.S. Census Bureau QuickFacts, 2012.), NYC embodies urbanization. Moreover, crucial details, such as the locations and widths of stormwater inlet drains and digitized maps of the sewer network are notably absent from public records. This lack of drainage data poses a challenge to hydro-hydraulic flood modeling and emphasizes the need for alternative methods to mitigate flooding issues within the city. Moreover, NYC holds economical influence, contributing approximately $1.8 trillion annually to the U.S. gross domestic product (Bureau of Economic Analysis, 2021). Therefore, the intricate urban fabric, intertwined with the economic significance and challenges of minimally accessible data, poses NYC as the ideal study area for the application of data-driven techniques in urban flooding.

NYC has 59 localized, politically based districts, called Community Boards (DISs). DISs manage zoning and land-use policies and address general municipal concerns (City of New York, 2023a). By borough, the breakdown of DISs is the following: 12 in Manhattan, 12 in the Bronx, 18 in Brooklyn, 14 in Queens, three in Staten Island. Due to the quantity and extent of the districts, they serve as an ideal starting point for variable aggregation and clustering analysis. In this study, shapefiles were downloaded from NYC Open Data (City of New York, 2023b), and the processing of data was conducted by ArcGIS and Python. In the proceeding sections, the methods of aggregation are further detailed.

2.2 Dynamic Variables

2.2.1 NYC 311 Platform

NYC311, a crowdsourcing platform, provides gainful insights into sewer related conditions in NYC. Observations of city issues are reported by residents of the city, where the date, time, and longitude and latitude coordinates of the incident are listed. Hence, the detailed temporal and locational information affords researchers the opportunity to employ data-driven techniques for analyses. For this study, NYC311 street flooding (SF), sewer backups (SB), and catch basin blockages (CB) reports, ranging from January 1, 2010 through December 31, 2019, were downloaded from the NYC Open Data website: http://data.cityofnewyork.us (Dates after 2019 were excluded, as the COVID pandemic is assumed to have an impact on reporting behavior, particularly in NYC, where residents and visitors relocated and returned at various intervals). For each report type, daily counts were aggregated to the DIS level. SF complaints, witnessed incidents of street flooding, served as the response variable. SB and CB reports were chosen for inclusion, as they are known influences towards SF (Agonafir et al., 2021; City of New York, 2022a). Specifically, SB indicate an internal issue within the drainage network, where it is overtaxed (Schmitt et al., 2004).); moreover, Agonafir et al has shown that SB reports are a significant predictor towards SF reports in almost half of NYC zip codes (Agonafir et al., 2021).

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

Thus, the addition of SB may strengthen a model's ability to make predictions. Also, in the Agonafir et al study, CB was found to have consequence on SF in roughly half the NYC zip codes (Agonafir et al., 2021). When a catch basin is blocked or clogged, runoff is not efficiently extracted into the stormwater drains, thereby allowing for ponding. In brief, from the NYC311 platform, the dynamic (values varying with time) infrastructural predictors, SB and CB, and the predictand, SF, were obtained. SF, SB, and CB counts were totaled per day per DIS by using the timestamps and longitude and latitude coordinates.

2.2.2 Radar and Gauge Data

Precipitation, rain and snow, drive urban flooding occurrence. While it is apparent that rainfall is the primary contributor (Agonafir et al., 2023; Qin et al., 2013; Schmitt et al., 2004; Sharif et al., 2006; Valeo & Ho, 2004), snowmelt also has influence, as when large amounts of snow liquifies, streets may be flooded (Semádeni-Davies & Bengtsson, 1998; Valeo & Ho, 2004). Concerning rainfall, there are intense rainstorms (large amounts of rain in a brief time interval), which contribute to flash floods, and there are prolonged rainy days, where the rainfall may not be intense, yet there is sufficient amount of water over a longer duration. In both cases, the capacity of the stormwater drains may become exceeded (Agonafir et al., 2023). Therefore, this study used three predictor variables representing precipitation: Max Hourly Rainfall (MR), Total Daily Rainfall (TR), and Snowfall (SN). Now, there are also varied methods of rainfall collection: in-situ (gauge) and remote sensing (radar and satellite). Some benefits of in-situ measurements include not being encumbered with cloud top reflectance, thermal radiance, retrieval algorithm and overpass frequency issues (AghaKouchak et al., 2009); whereas radar data is advantageous in terms of spatial distribution (Thorndahl et al., 2017), capturing precipitation amounts at more locations within an area. As both techniques are considered standard measurement methods, this study employed radar data for the MR and gauge data for the TR and SN variables. Each of the precipitation variables were determined at the DIS level.

The radar rain data was taken from the National Center for Atmospheric Research (NCAR)/Earth Observing Laboratory (EOL) website, and the gauge rain and snow data were retrieved from NOAA's Climate Data Online. For the radar, the resolution is 4 km by 4 km, and the gridded data is Stage IV, benefiting from manual quality control (EOL, 2022). Hourly totals were gathered for the dates ranging from January 1, 2010 through December 31, 2019. Then, the maximum hourly value (MR) was taken for each radar point for each day. Ultimately, the MR values were assigned to each DIS based on the radar point's proximity to its centroid. Now, regarding the gauge data, the Global Historical Climatology Network (GHCN) by NOAA's National Centers for Environmental Information (NCEI) provides daily climatology details from land surface stations globally (National Centers for Environmental Information, 2023). With respect to the daily rain totals (TR) and snow totals (SN), data was collected from the GHCN station, NY CITY CENTRAL PARK, NY US, for the dates ranging from January 1, 2010 through December 31, 2019. These are direct measurements provided by the station. The TR and SN 24-hour amounts were assigned to each DIS. Therefore, by a combination of radar and gauge determinations, precipitation, representing rainfall intensity, total rainfall, and total snowfall, were assigned to each DIS at the daily level as predictors for the models.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

2.3 Static Features

There are multiple factors driving the occurrence of urban flooding. First, there are topographical variables, such as slope (SLP) and elevation (ELV). Regarding slope, the greater the incline of a surface, the greater the velocity and discharge of water; hence, at the bottom of the slope, the water will pond quickly (Bruwier et al., 2020). Concerning elevation, studies have also shown lower elevated areas to be at a higher risk of flooding. For one, lower elevated areas are more vulnerable to storm surges from coasts and rivers, and secondly, as mentioned prior, lower elevated areas may be located at the edge of a sloped surface (Ouma & Tateishi, 2014; Woodruff et al., 2013). In addition to the topographical, there are urban features, specifically the quantity of buildings (BLD) and the extent of building footprint (FTP), which affect flooding. Buildings are an impervious surface, such that water is unable to infiltrate through the ground. Moreover, multiple studies have found buildings to have a dominating influence on urban flooding, compared to other common flood factors (Agonafir et al., 2022; Bruwier et al., 2020; J. Lin et al., 2021). Another variable included represents percent impervious (IMP); it depicts the percentage of all impervious surfaces, such as buildings, sidewalks, and streets, within a neighborhood. Next, a variable representing the area (SIZ) of the DIS was included in the study. The size of a region does not increase flooding occurrence, yet, larger areas have more opportunity for flooding occurrences, leading to higher flood counts; thus, the machine learning model will benefit from the information. Lastly, concerning locational, latitude (LAT) and longitude (LNG) coordinates, the variables allow for a directionality indication of flood occurrence. For instance, Agonafir et al found that street flooding in NYC had a southern and eastern locationality of increased flood incidents (Agonafir et al., 2022). While the exact cause of flooding is not given by the location, the variable allows for a geographical pattern to be learned within machine learning models. Overall, physical features including slope, elevation, building extent, area and geographical coordinates are useful variables in understanding urban flooding via modeling.

For the physical features, shapefiles were downloaded from NYC Open Data and processed via ArcGIS analysis tools. For the SLP and ELV variables, a shapefile of elevation points was downloaded, and the mean elevation in meters (m) and mean degree of slope were determined per DIS. Also, a shapefile, providing the number of buildings, was retrieved, and the total number of buildings per square kilometer (km^2) area of each DIS was tallied to represent the BLD variable. Regarding the FTP variable, a shapefile of building footprints was used, and via ArcGIS, for each building footprint, the area in km^2 was calculated. In each DIS, the sum of the footprint areas was determined. For the SIZ, LAT, and LNG variables, with the DIS shapefile, the values were determined via geometry processing tools of each DIS polygon, where SIZ was calculated in km^2, and the LAT and LNG represent the centroid points of each polygon. Therefore, by geoprocessing, the static features for the analysis were collected.

3. Methodology

With the overarching goal being the achievement of a model, which produces profound insights on urban flooding despite limited data, this study investigates the value of added components, ascending in complexity. The methodology is outlined as a flowchart in Fig. 1. The

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

preliminary step for this analysis was the delineation of meaningful zones. Borders of zip codes, DISs, or boroughs are not based on topographical or urban flood characteristics. Therefore, while the DIS serves as a sufficient starting point for data aggregation, a further outlining of risk zones based on flood factors, as opposed to political or postal bordering, has more utility in modeling endeavors. Accordingly, a spectral clustering technique was applied to identify areas of similar flood vulnerability characteristics. Specifically, the features used as inputs to the clustering algorithm were SLP, ELV, BLD, FTP and IMP. Then, six clusters (zones) based on likeness were created (Fig. 2), and the dynamic and static variables were aggregated from the DIS to the cluster level. Thus, with zones of NYC regions aligned on related flood factors, predictive modeling was able to be performed.

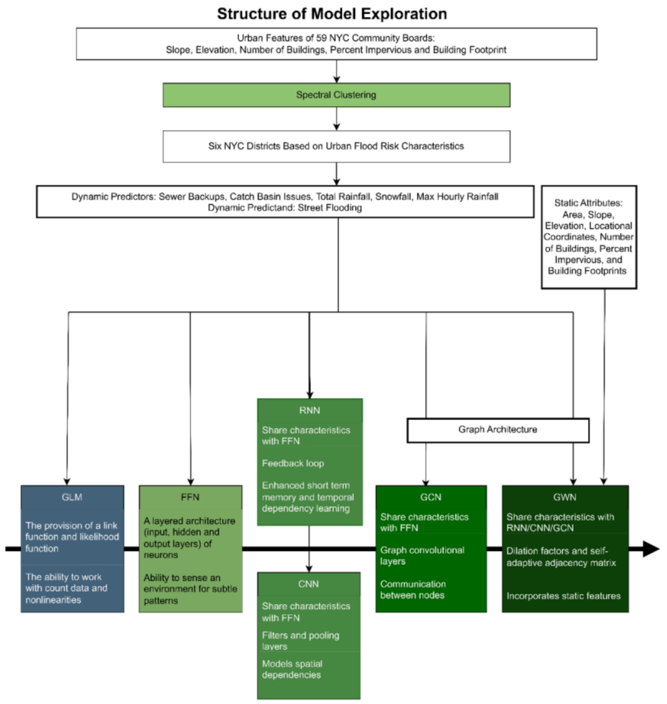

Fig. 1 A flow chart outlining key processes.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

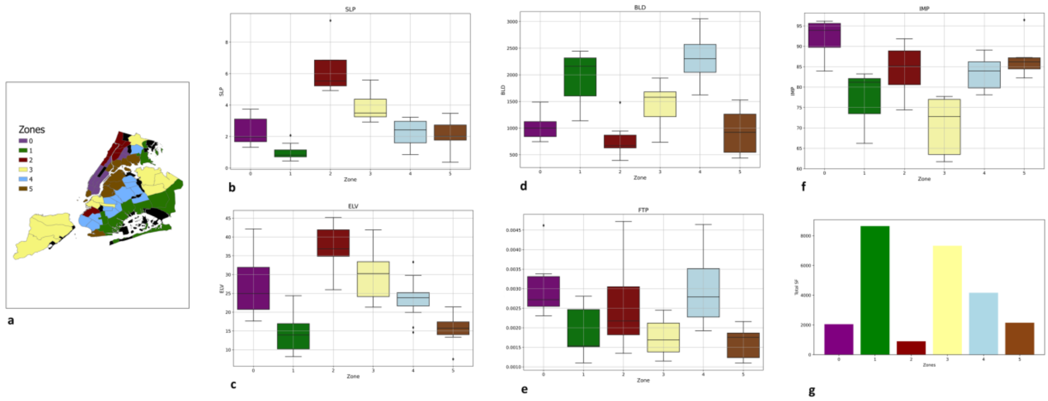

Fig. 2 The six zones delineated by urban flood risk characteristics. Fig. 2a depicts the mapping of the zonal spreads per DIS. Fig. 2b-f are box plots depicting the range of flood characteristics of DISs within each zone. For each zone, the extent of the mean slope, mean elevation, number of buildings per DIS area, the sum of total building footprint area per DIS area, and mean percent of impervious cover of each DIS are shown in Fig. 2b, 2c, 2d, 2e, and 2f, respectively. Fig. 2g are the bar plots illustrating the sum of the total SF reports in each zone over a 10-year period.

The first set of models were created to predict street flooding occurrence (SF) based on daily values of climatic (MR, TR, SN) and infrastructural influences (SB and CB) over a 10-year time span. The research begins with a classical model - the GLM. Features of GLMs include the provision of a link function and likelihood function; in addition, GLMs possess the ability to work with count data and nonlinearities (Hardin & Hilbe, 2007). Next, the research moves to a more complex, yet classical AI-based model – the FFN. FFNs introduce a layered architecture (input, hidden and output layers) of neurons, which gives the models the ability to sense an environment for subtle patterns (Fine, 2006; Setiono, 2001). Also, FFNs have the benefit for adaptability, where elements may be added, allowing for varying capabilities. Particularly, the architecture of FFNs provides the basis for the RNN and CNN. For instance, with the RNN, there is the feed forward mechanism as found in the FFN; however, the RNN is also equipped with a feedback loop, thereby enhancing short term memory and temporal dependency learning (Fei & Lu, 2018; Schmidt, 2019). Regarding the CNN, the feed forward process comprises of convolutional and pooling layers, where spatial (patterns across neighboring time steps) aptitude is achieved via filtering kernels (Albawi et al., 2017; Durairaj & Mohan, 2022; Koprinska et al., 2018; M. Sun et al., 2017). When working with time-series data, coordinates of observation values are created, and the spatial capability refers to the comprehension these numerical relationships in time, which is processed in parallel, as opposed to in sequence as with the RNN (K. Wang et al., 2019). Hence, the RNN and CNN are both capturing temporal dependencies; however, the mechanism of temporal incorporation differs. This study investigates the RNN and CNN to explore the effects of each added component towards the forecasting of SF within this specific data set. Overall, the beginning models, varying in architecture, temporal capabilities, utilize the time-series predictors to make predictions on the response variable representing street flooding in the urban setting.

Graph-based neural networks consider locational aspects. Since there is a geographical component to urban flooding, as localized areas of susceptibility are to be discovered, this model

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

study further explores the benefits of the graph neural network architecture. For the GCN, the predictors remain as MR, TR, SN, SB, and CB, with the response variable as SF. As there are only 6 nodes in this study, all within a reasonably close geographical proximity, the model was created so that each cluster is connected to all other clusters. Each cluster represents a node, and via edges, external information is communicated among the nodes, thereby bridging potential gaps within each node's incomplete internal data (Jiang et al., 2023; Piao et al., 2022; Scarselli et al., 2009; Y. Wang et al., 2023). Ensuingly, this study explores the GCN and the GWN. For the GCN, the inputs are as described above. A tensor creates a fully-connected graph, where features from neighboring nodes are aggregated to assess how conditions in one zone may have affect another zone. The process is achieved via two graph convolutional layers. Lastly, the graph neural network architecture is transfigured to the GWN. Here, the GWN had been adapted from Sun et al (A. Y. Sun et al., 2021). Considered the most complex model of this study, the GWN incorporates all the aspects of the previous neural network models - locational, spatial, and recurrent elements - while also introducing novel features of its own, such as gated layers and the self-adaptive adjacency (SAA) matrix. Moreover, due to the SAA, the GWN is able to incorporate static features. For this study, the following static characteristics were fed into the GWN: SIZ, SLP, ELV, LAT, LNG, BLD, FTP, and IMP.

For every model in the exploration, the coefficient of determination (R^2) was used as a validation measure. The dataset of variables was partitioned into training and testing sets. The training data ranged from January 1, 2010 through January 1, 2018; the testing data ranged from January 2, 2018 through December 31, 2019. Each model made predictions spanning the testing date range; the predictions were compared to the observed, and the R^2 was ultimately computed. Hence by evaluating the goodness of fit, model comparisons were conducted.

3.1 Spectral Clustering

Spectral clustering, an unsupervised machine learning technique, partitions groups based on similarities. For this study, the SpectralClustering tool from the sklearn module is utilized in Python (Scikit-learn, 2023). The data points of each DIS are SLP, ELV, BLD, FTP and IMP. Each feature was transformed independently via Standard Scaler, processing for a mean of zero and standard deviation of one, as to prevent disproportional influence on the algorithm's computation. Here, the features were represented as $x_1, \ldots, x_n$. For each vertex (DIS), edges were constructed from x_i to its k-nearest neighbor, x_{ij}. The Euclidean distance [$t(x_i, x_j)$] between each unique pair of x_i and x_j was calculated (Scikit-learn, 2023). Then, a measure of similarity (s_{ij}) was determined as follows (Scikit-learn, 2023):

$$s_{ij} = e^{-10 \times t(x_i, x_j)^2} \ldots (1)$$

Edges were created between each pair, and similarity is used as the edge weight. The purpose of similarity weights is for the edges between a pair of points in the same group to have greater weights than the edges between a pair of points that lie in separate groups (von Luxburg, 2007). An unnormalized Laplacian graph was formed, with the matrix defined by Luxburg as follows (von Luxburg et al., 2008):

$$L = D - S \ldots (2)$$

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

S was the similarity matrix: S: $= (s_{ij})_{i,j=1,\ldots,n}$ and D was the diagonal matrix with passes: d_{ij}: $= s_{ij} \sum_{j=1}^{n} s_{ij}$. With the computation of L, eigenvalue decomposition was initialized. The solver used was ARPACK, which computes k eigenvectors of L: $v_1, \ldots, v_k$ (von Luxburg, 2007). Here, k is six (the number of desired flood zones). Let V be the matrix, where the eigenvectors were columns, and q_i represented the vector in the i-th row of V, then via the k-means algorithm, the points, $(q_i)_{i=1,\ldots,n}$ were grouped into clusters (von Luxburg, 2007). Hence, with the preceding machine learning technique, each of 59 DISs were grouped into six zones (labeled 0, 1, 2, 3, 4, and 5), based on the physical and urban traits.

There are physical characteristics which influence a region's susceptibility to SF complaints. The primary intent of the spectral clustering application is informed delineation. Nevertheless, a further advantage is the depiction of regions sharing similar extents of known, physical, flood factors, as detected in the Random Forest model by Agonafir et al (Agonafir et al., 2022) and discussed in previous urban flood literature. The Standard Scaler function was employed so that each attribute holds comparable influence. Hence, for this study, the purpose of spectral clustering model is not to serve as a discovery or predictive model, as it is preparing the data for the statistical model and supervised machine learning explorations.

By spectral clustering, six zones were designated based on the flood factors. The quantity of six was chosen, as to provide a higher degree of localization compared to county or borough levels. To illustrate the prevalence of each trait within a zone, box plots were created (Fig. 2b-f). Each box plot is comprised of the values of the DISs within the specified zone. The plots illustrate the range of SLP, ELV, BLD, FTP and IMP. Also, the total SF complaints, over the 10-year timespan, for each zone is depicted in Fig. 2g. Since each DIS now belongs to a zone, the data must also be aggregated to the zonal level. For the dynamic variables, SF, SB, and CB, the totals of each DIS with a zone were taken, and for MR, the mean of the DIS values within a zone were calculated. For TR and SN, the measurements were previously taken from a single source; hence, no aggregation was needed. For the static attributes [only used as an input to the GWN], SLP, ELV, LAT, LNG, BLD, FTP, and IMP, mean values for each DIS within a zone were determined, and for SIZ, the sum of the areas of each DIS in a zone was calculated. Therefore, the dynamic predictors and response variables and the static characteristics for each zone were prepared for the performance of predictive modeling.

3.2 Poisson Generalized Linear Regression Model

The GLM used in this study is the Poisson GLM (hereafter referred to as GLM). Here, the GLM uses a log link function and a Poisson distribution of the exponential family. For each zone, the target variable and the explanatory variables were expressed at i-th observations as follows:

$$SF_i \sim Poisson(\lambda_i) \ldots (3)$$

where,

$$\lambda_i = e^{\beta_i^0 + \beta_i^1 SB_i + \beta_i^2 CB_i + \beta_i^3 MR_i + \beta_i^4 TR_i + \beta_i^5 SN_i} \ldots (4)$$

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

The β^k coefficients represent the strength of change in the log-relative rate of the SF for a one-unit change in the associated predictor variable, and β^0 is the intercept, which is the baseline rate when the predictors are zero.

3.3 Feed Forward Neural Network

The FFN, applied to each zone independently, is composed of three layers: input, hidden, and output. Fig. 3 depicts the architecture. The input layer has the two infrastructural and the three climatic predictors; each predictor is normalized using the Python sklearn's Standard Scaler function. For this model, at each time step, there are 32 neurons within the hidden layer, and each predictor feeds forward to all neurons. For each connection, via the Adam optimizer gradient descent method, weights, ω, are initialized; moreover, for all neurons, a distinctive bias, b, is computed by random initialization. Then, for every neuron, i, there is a weighted sum calculation, as follows:

$$S_i = SB\omega_{SBi} + CB\omega_{CBi} + MR\omega_{MRi} + TR\omega_{TRi} + SN\omega_{SNi} + b_i \dots (5)$$

The sum then enters into the ReLU activation function, σ^R. Activation functions aid in understanding nonlinear relationships. The ReLU was chosen, as it is known for its accuracy and is widely used in deep learning modeling (Dubey & Jain, 2019). The next step in the FFN is the forward movement of information from the last hidden layer to the output layer. Similar to the last weighted sum calculation, the weighted sum at the output neuron is computed:

$$S_{output} = \left(\sum \omega_i \sigma_i^R\right) + b \dots (6)$$

From the output layer, the data enters a linear function and produces the predicted counts of SF. The model was constructed via the Pytorch NN module (PyTorch, 2023c). The model was run with a learning rate of 0.001, batch size of 32 and 100 epochs.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

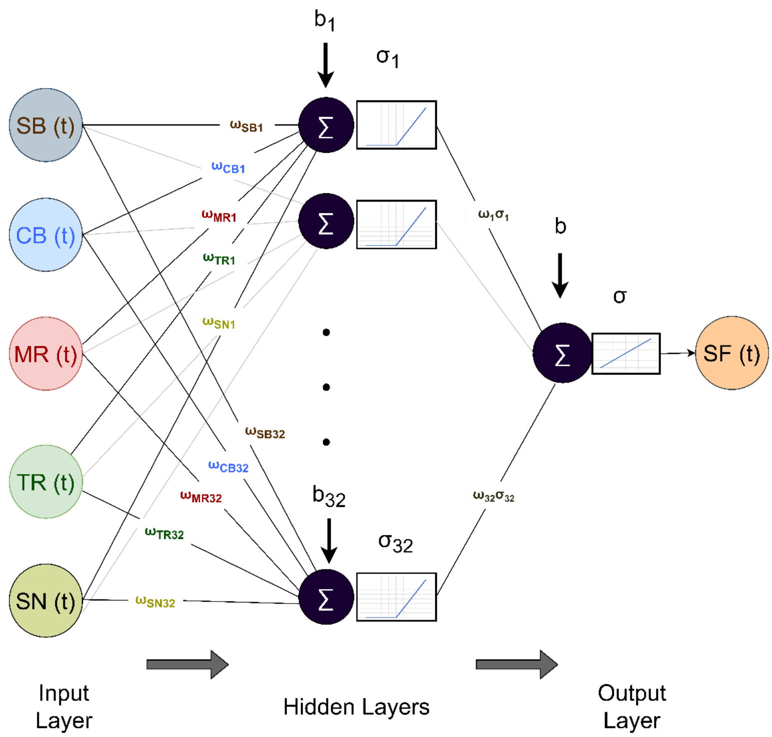

Fig. 3 The FFN architecture.

3.4 Recurrent Neural Network

The RNN architecture, built upon a network of neurons, is similar to the FFN structure. However, there is a difference within the hidden layer composition, where, as opposed to possessing only feed forward (FF) layers, the RNN includes a preceding Long Short-Term Memory (LSTM) layer (Fig. 4). First, a concatenated input vector of the predictors, X at each time step enters each neuron. X is then concatenated with the hidden state vector, H, at the previous time step. Then, via gradient descent optimization, unique (per neuron) input weight and bias are calculated. The input gate, i, controls the extent of input information entering the cell state (Schmidt, 2019; Tsantekidis et al., 2022), and the computations for each neuron, n, of the 16 neurons of the RNN layer are as follows:

$$i_{n,t} = \omega_{i,n}\left[H_{t-1,n}, X_t\right] + b_{i,n} \ldots (7)$$

After the initial computation, the input enteres the sigmoid activation, $\boldsymbol{\sigma^S}$ function:

$$I_{n,t} = \sigma_n^S\left(i_{n,t}\right) \ldots (8)$$

Now, the forget gate, f, also receives the input vector of predictors; yet, it's function is to filter out irrelevant information from the previous cell state (Schmidt, 2019) .The calculations for f are as follows:

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

$$f_{n,t} = \omega_{f,n}\left[H_{t-1,n}, X_t\right] + b_{f,n} \ldots (9)$$

Then, the sigmoid function is applied:

$$F_{n,t} = \sigma_n^S\left(f_{n,t}\right) \ldots (10)$$

With the productions of the input gate and the forget gate, the cell state, C, is computed. In the cell state computations, the tanh activation function, σ^T are applied:

$$C_{n,t} = F_{n,t}C_{n,t-1} + I_{n,t}\sigma_n^T\omega_{c,n}\left[H_{t-1,n}, X_t\right] + b_{c,n} \ldots (11)$$

Additionally, the input vector of predictors also pass through the output gate. The role of the output gate is to control the flow from the cell state to the hidden state and ultimately producing the output of the LSTM neuron (Chung et al., 2014). The output gate calculations are shown here:

$$o_{n,t} = \omega_{o,n}\left[H_{t-1,n}, X_t\right] + b_{o,n} \ldots (12)$$

$$O_{n,t} = \sigma_n^S\left(O_{n,t}\right) \ldots (13)$$

Finally, with the product of the output gate, the hidden state, the information that passes to the next layer (the FF layer), is computed. The figuration was as follows:

$$H_{n,t} = O_{n,t}\sigma_n^T C_{n,t} \ldots (14)$$

The process continues through the FF layer to predict SF counts. Here, the FF had 16 neurons. The RNN model underwent 1000 epochs, with a batch size of 32, a learning rate of 0.007, and a sequence length of 6 days. The model was constructed via the Pytorch LSTM module (PyTorch, 2023b).

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

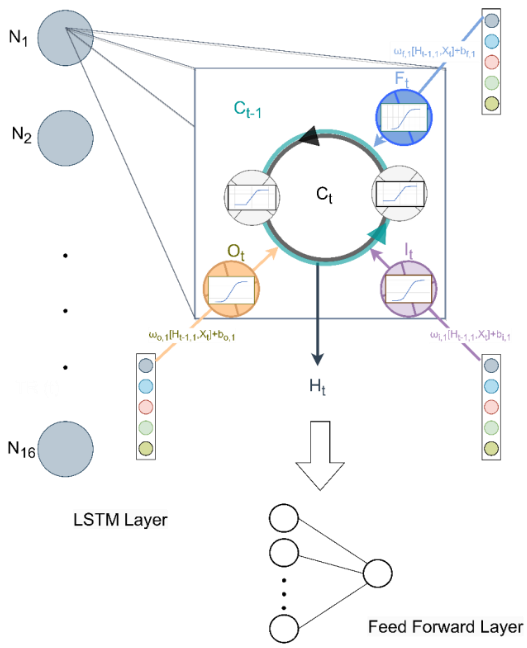

Fig. 4 The hidden layer structure of the RNN.

3.4 Convolutional Neural Network

The CNN is also comprised of an input, hidden and output layer. The input layer and the output product are the same as the FFN; hence, the main difference from the FFN is the composition of the hidden layers. Particularly, the CNN has two convolutional layers, a global average pooling layer and a FF layer. For the first convolutional layer, there are 16 kernels (or filters), f, of size one (Each kernel is applied independently to each predictor at each time step). Each filter produces an output, c, by generating a unique bias and weight at each time step, t, and for each predictor. Therefore, in the CNN, the convolutional layers involve the weighted sum of predictors over time for each filter; this differs from the FNN, as in the FNN, the predictions are based on the weighted sum of predictors without considering the time dimension. Fig. 5 illustrates the hidden layers of the CNN. The overview of computations are depicted here:

$$c_{f,t} = \omega_{f,t,SB}SB + \omega_{f,t,CB}CB + \omega_{f,t,MR}MR + \omega_{f,t,TR}TR + \omega_{f,t,SN}SN + b_{f,t} \ldots (15)$$

The output of the first convolutional layer enters the second convolutional layer, c^*, via the ReLU activation function and computes the following:

$$C_{f,t} = \sigma(c_{f,t}) \ldots (16)$$

The second convolutional layer has 32 kernels, f^*. For each filter output, c^*, at a time step, the calculation is shown:

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

$$c^*_{f^*,t} = \sum_{f=1}^{16} \omega_{f^*,t,C_{f,t}} C_{f,t} + b_{f^*,t} \ldots (17)$$

The output of each kernel in the second convolutional also enters the ReLU function to gain enhanced pattern recognition:

$$C^*_{f^*,t} = \sigma\left(c^*_{f^*,t}\right) \ldots (18)$$

After exiting the second convolutional process, there is input into the global average pooling layer. For each of the 32 channels, p, an average, g, is taken across all time steps (3,652 days):

$$g_p = \frac{1}{3652} \sum_{t=1}^{3652} C^*_{f^*,t} \quad for\ f^* = p\ \ldots (19)$$

Now, the pooling output units are then processed by a feed forward layer of 32 neurons, producing predicted SF counts. The model was run with a learning rate of 0.001, batch size of 32 and 100 epochs. The model was constructed via the Pytorch CONV1D module (PyTorch, 2023a).

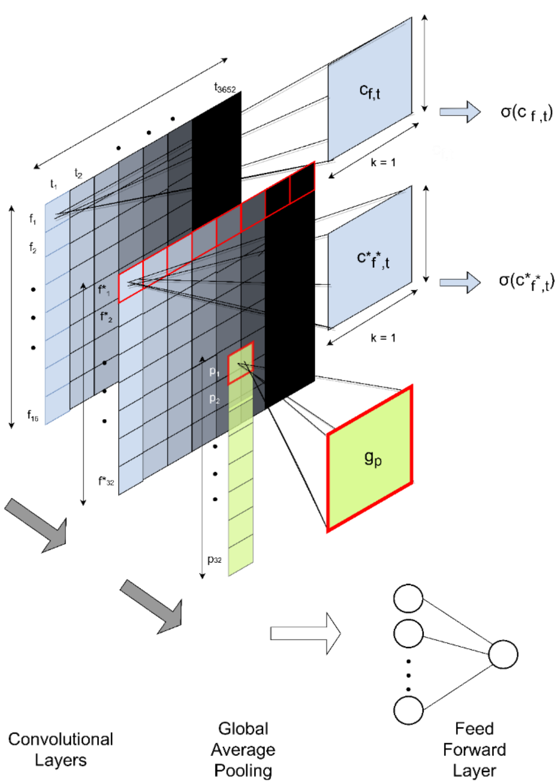

Fig. 5 The hidden layer structure of the CNN.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

3.5 Graph Convolutional Neural Network

The main contribution of graph-based methods is the sharing of information via neighbors. The input predictors are the same as the previous models; however, as opposed to assessing each node's predictors singularly and running separate models for each, the input vector includes all the nodes and their respective predictor values at each time step. Within the model, individual assessments occur, and SF counts are produced for each node. At the hidden layer, the GCN begins the process by the creation of edges, the advanced communication channels between nodes of locational proximity. In this study, each cluster represents a node, and the edges are constructed such that each node is fully connected to every other node. Fig 6 illustrates the edges connecting the clusters (nodes) of this study. An input vector of the dynamic predictors, X at each time step is entered into each node, c. Additionally, X is also messaged into neighboring nodes, N (those nodes connected to the node by edges), via aggregation by edge weights. The messaging, m, calculation, for a node is as follows:

$$m_c = \sum_{i \in N_c} \omega_{c,i} X_i \ldots (20)$$

After messages are aggregated, feature representation is updated at the next time step:

$$X_{c,t+1} = f\left(X_{c,t} m_c\right) \ldots (21)$$

The aggregated updated node representations occur within the graph convolutional layers. The model is constructed via the Pytorch Geometric GCN module (PyTorch Geometric, 2023). There are two convolutional layers, of which the model acquires knowledge of the effect each neighbor's feature may have on another. The ReLU activation function was employed. The learning rate was 0.001, and the GCN underwent 100 epochs.

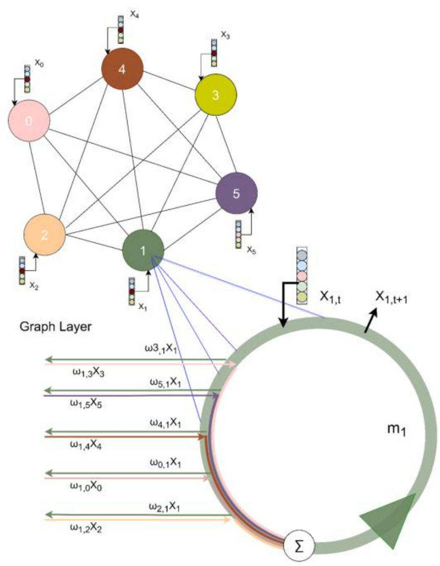

Fig. 6 The graph structure of the GCN.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

3.6 Graph Wavenet

The GWN is graph-based, with sophisticated recurrent and convolutional aspects. The key advances from the GCN, RNN, and CNN are the additions of the adjacency matrix, A, dilation factors, l, and skip connections. The input set of dynamic predictors and the output of predicted SF counts are similar to the other graph-based models. Although, for the GWN, the arrangement of the input differs. The input is a vector, D, which includes the nodes and the dynamic predictors per node for each sequence, q. Here, the sequence is 6; thus, the vector includes information from t, t-1, and so forth until t-5. D then enters the A, along with the static attributes (SIZ, SLP, ELV, LAT, LNG, and BLD), T. With the data, A makes informed decisions towards graph construction and the determination of node neighbors, via edges. Moreover, for each unique filter, A creates a unique transformed input vector for a particular node. Now, for each node, n, there is a transformed input vector, described as follows:

$$X_{n,f} = f(D, T, q) \ldots (22)$$

The transformed vectors for each node then enter convolutions. Recall from the CNN section, that the convolution operation, c, is a function of the predictors, weights and biases, where the weights and biases are unique for each time step and filter. It is similar here; however, for the GWN, it is also a function of the dilation factor. For the GWN, the output of the convolution operation will be referred to g:

$$g_{n,f} = f\left(X_{n,f}, \omega_f, b_f, l\right) \ldots (23)$$

Next, g is split, where one enters into the tanh activation function, and the other is processed by the sigmoid activation function. While information passes through the sigmoid and tanh activation functions in the RNN, this occurs at sequential steps. Whereas, for the GWN, the transformations by the sigmoid and tanh activation functions are brought together in element-wise multiplication. The determinations of the output tensors of the sigmoid activation function, G^S, and the output of the tanh activation function, G^T, are shown here:

$$G_{n,f}^S = \sigma_f^S\left(g_{n,f}\right) \ldots (24)$$

$$G_{n,f}^T = \sigma_f^T\left(g_{n,f}\right) \ldots (25)$$

An element-wise multiplication, M is then performed on both tensors.

$$M_{n,f} = G_n^S \odot G_n^T \ldots (26)$$

The input tensor, M, then passes through a 1X1 convolution, where a point-wise convolution operation takes place, reducing the hidden dimensions from 11 to one. The output, P, is described here:

$$P_{n,f} = f\left(M_{n,f}, \omega_{n,f}, b_{n,f}\right) \ldots (27)$$

Now, during this process, the original input tensor X is also preserved. It is added, elementwise, to P, as to produce a residual connection, **r**. In this way, the output not only learns from the

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

transformed input, via convolutions and gating mechanisms, but it also learns from its original input.

$$r_{n,f} = P_{n,f} \oplus X_{n,f} \text{ ... } (28)$$

Meanwhile, there is another part of the output of the 1x1 convolution with a different utility. This output, known as a skip connection, will be designated as s:

$$s_{n,f} = f\left(M_{n,f}, \omega_{n,f}, b_{n,f}\right) \text{ ... } (29)$$

At this point, the calculations have been shown at the filter level. As explained in the CNN section, the calculations of each filter are aggregated to the layer level. Hence, describing the calculations at the layer level, the residual connection exiting the layer, y, will be denoted as R, and the skip connection exiting y will be denoted as S. For this model, R exits the filter and is utilized in subsequent skip connections, and S exits the filter, and residual connections from previous layers are incorporated to form a feedforward output, K. Let z represent the quantity of layer skips, then the calculation of K is as follows:

$$K_y = s_{n,f,y} \oplus \sum_{z=0}^{Z} R_{n,f,y-z} \text{ ... } (30)$$

The final output then undergoes a ReLU activation function, and the model prepares predictions for SF counts on the testing dates. The process is shown in Fig. 7. For this study, the hyperparameters for the GWN included a 0.01 learning rate, batch size of 15, and 50 epochs.

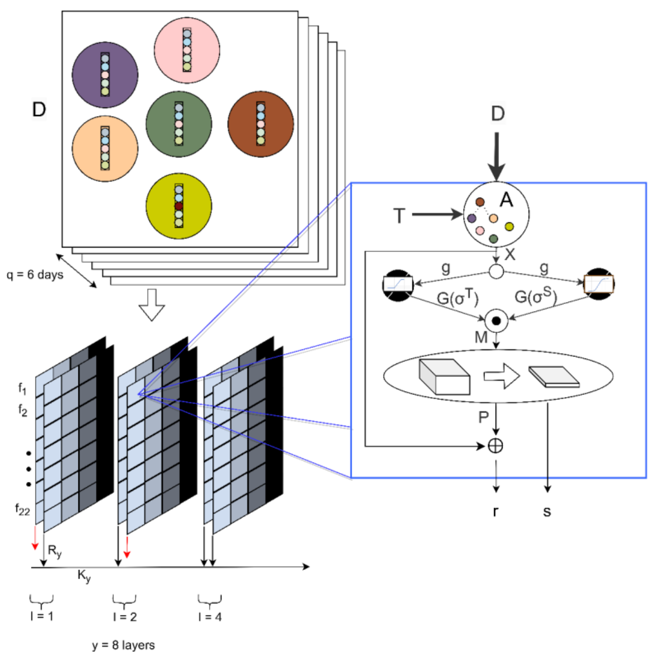

Fig. 7 A diagram depicting GWN configuration.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

4 Results and discussion

4.1 Risk Zones

Spectral clustering, an unsupervised machine learning technique, created six zones based on the flood attributes of each DIS. Fig 2 shows the predominance of each of the five characteristics within a zone and plots total SF occurrence in each zone. By a visual analysis of the plots, elements of risk are conveyed. For instance, zone 2 is shown to have the highest mean incline (Fig. 2b), allowing for water to flow, as opposed to ponding. Moreover, zone 2 has the greatest elevation (Fig. 2c) and the least number of buildings per unit area (Fig. 2d). A higher elevation and lower quantity of buildings are known to reduce urban flood susceptibility. Subsequently, the physical qualities of zone 2 may serve as a plausible explanation for the zone having the lowest total SF complaints (Fig. 2g). Similarly, zonal characteristics may also explain SF occurrence in zone 1. When viewing Figs. 2b-d and Fig 2g, it is shown that zone 1 has the opposite extent, with the flattest surface (low slope), lowest elevation, and the second highest number of buildings per square unit; notwithstanding, zone 1 has the most SF reports. Thus, the extensive flood risk characteristics of zone 1 may be the antecedent for its high complaints. Concerning the remaining zones, the prevalence of a combination of flood attributes is not as strongly skewed. For example, zone 0, which has a low SF total, ranges mediumly in SLP, ELV, and BLD; although, it prevails on the higher end for FTP, and it has the greatest IMP. Regarding zone 3, there are no extremes in flood characteristics; yet, zone 3 has the second highest total SF. With zone 4, it has the highest BLD and FTP; yet the values of the other flood attributes extend moderately. Lastly, concerning the box plot of zone 5, it does not depict many extremes in flood characteristics, except a relatively low FTP and high IMP. Regarding total SF, zone 5 retains low SF reports (slightly higher than that of zone 0). Overall, via visual inspection, it appears that some flood characteristics, particularly BLD, SLP, and ELV, maintain stronger effects on total SF reports. This was also found in (Agonafir et al., 2022), where the random forest algorithm detected the relative importance of BLD, SLP, and ELV to be greater than IMP and FTP. For the zones with moderate flood characteristics, modeling forecasting techniques, specifically machine learning methods, have utility, as the algorithms possess the ability to detect intricacies within a learned environment.

4.2 Model Results

4.2.1 Validation Results of Models

The evaluation of the models with varying complexities is conducted via the assessment of R^2 values. Each model utilized a training set between 2010 to 2018 to produce daily predictions for the testing range from 2018 through 2019, and goodness-of-fit was determined. Each model, aside from the GLM, underwent 50 runs, and the average of the mean and median R^2 values of all nodes were tabulated in Table 1. The model results are mapped, at the zonal level, in Fig. 8. Ordered from least to greatest, the results, depicting the mean R^2, for the models are the following: GLM (0.26), FFN (0.35), RNN (0.36), CNN (0.36), GCN (0.43), and GWN (0.51). Thus, the GWN is the model with the best performance. Moreover, the delta between the model with the lowest R^2 (GLM) and the highest (GWN) is found to be significant at a value of 0.25.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

Furthermore, when only comparing ANNs, there is a 0.16 delta between the GWN's R^2 (highest) and FNN's R^2 (lowest). Lastly, inspecting the difference in R^2 between the simplest introduction of the feed forward (FFN) to the statistical model (GLM), there is a 0.09 delta, and between the simplest graph neural network (GCN) to the GLM, there exists a 0.08 delta. Contiguously, the results bring forth apparent connections. Firstly, there is notable performance improvement from the feed forward models (FFN, RNN, and CNN) to the GWN, while minimal (in some zones) to no improvement exists from the FFN to the RNN or to the CNN. This indicates that LSTM and convolutional layers benefit from the extensive structural detail, such as gating mechanisms, dilations, and skip connections, of the GWN. Secondly, as illustrated by the increase in R^2 achieved by the FFN when comparing to the GLM, a simple machine learning model outperforms the generalized linear regression model. Finally, there is an overall improvement from the introduction of the graph architecture to the feed forward process, as highlighted by the mean performance of the GCN being greater [by 0.07] than the RNN, and CNN. This suggests that the graph-based structure assists in environmental learning for certain datasets. Principally, due to the substantial differences in model performance, this exploration proves that careful consideration must be taken when choosing an appropriate technique for forecasting endeavors.

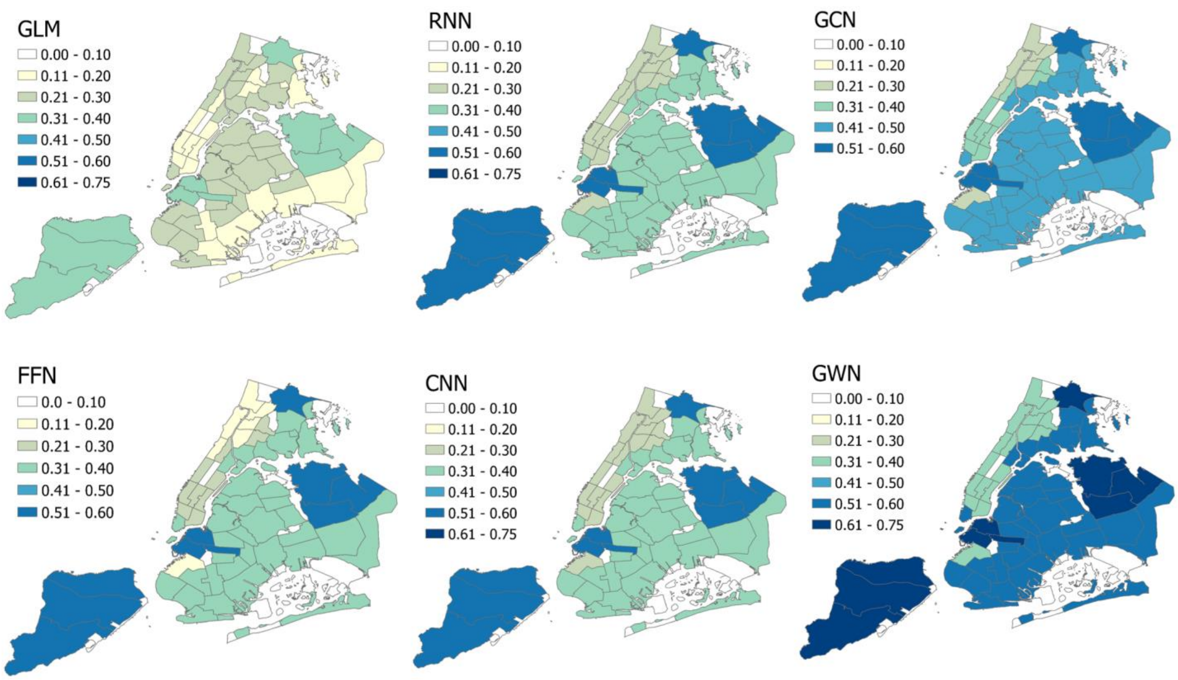

Fig. 8 Maps of the mean R^2 values for each of the models in the exploration. The R^2 values of the GLM, FFN, RNN, CNN, GCN and GWN are shown in Fig. 8a, 8b, 8c, 8d, 8e, and 8f, respectively.

Table 1. A listing of the R^2 values for each model at each zone and the mean and median determinations.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

Zones	0	1	2	3	4	5	Median	Mean	
GLM	0.20	0.20	0.24	0.38	0.25	0.27	0.25	0.26	
FFN	0.26	0.39	0.17	0.57	0.39	0.34	0.35	0.35	
RNN	0.25	0.39	0.23	0.56	0.40	0.33	0.36	0.36	
CNN	0.26	0.39	0.22	0.57	0.40	0.34	0.36	0.36	
GCN	0.37	0.45	0.28	0.58	0.44	0.44	0.43	0.43	
GWN	0.35	0.59	0.31	0.72	0.55	0.52	0.54	0.51	
Median	0.26	0.39	0.24	0.57	0.40	0.34			
Mean	0.28	0.40	0.24	0.56	0.41	0.37			

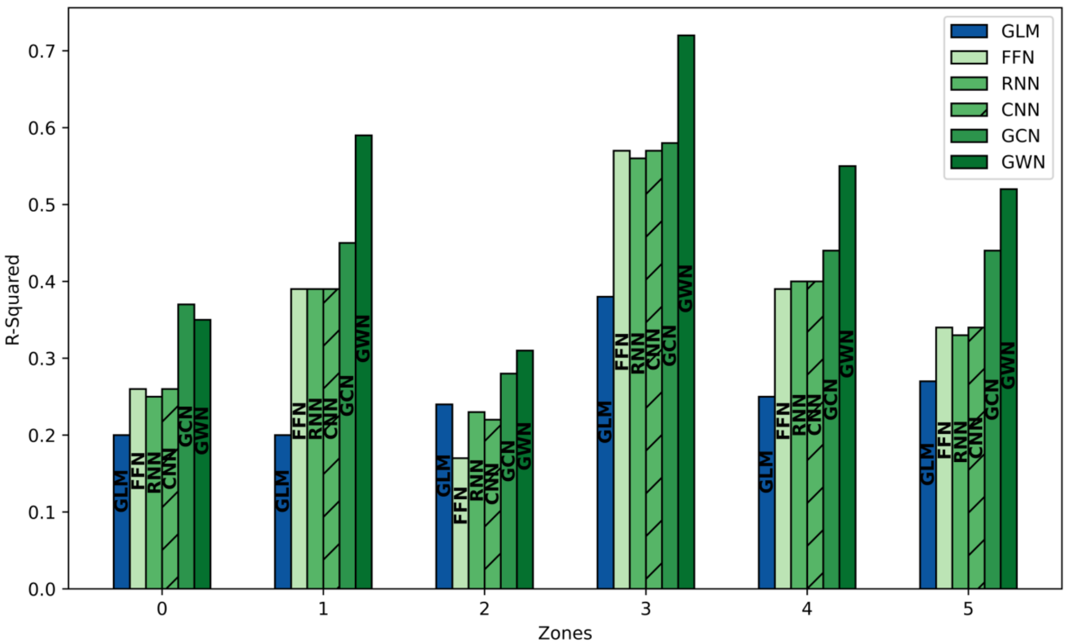

Fig. 9 The bar plots of the mean R^2 of each model at each zone.

4.2.2 Zonal Analysis

As spectral delineation has created each zone with varying physical traits, analyzing the results at the zonal level provides additional insights into model strengths. First, it is essential to inspect the relationship between R^2 (Fig. 9) and total SF per zone (Fig. 2g). There appears to be a general trend, where R^2 values are greater in zones with a larger quantity of total complaints. For instance, zones 0 and 2 have lower total SF reports, and the R^2 values for these zones are also the lowest. Moreover, zone 3 has the second highest quantity of complaints, and it has the best performance of all models. Thus, more available response data appears to be a benefit to the models. Moreover, the graph-based models significantly boost the quality of predictions, especially in the cases of zones 0 and 5. This may be attributed to the message-passage capability of separate nodes (zones), such that a zone is not only learning its specific environment, but also

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

gaining a sense of the surroundings. Nonetheless, while there appears to be a positive link between quantity of complaints and validation values, there are other factors influencing model performance. For instance, zone 1 has the highest number of SF reports; yet, the mean R^2 values are lower (0.40) than those of zone 4 (0.41), which has less than half the complaints. In addition, zone 5 outperforms zone 0 in all models [by a difference of 0.09 mean R^2], despite having a similar quantity of SF reports. These observations indicate the presence of other factors influencing zonal differences in modeling prediction aptitude. For example, concerning zone 1, additional variables may prove useful as model inputs. Specifically, zone 1 is comprised of regions mostly along the waterbodies of the Long Island Sound, Lower New York Bay, Jamaica Bay, and the Atlantic Ocean; furthermore, this sector also has the lowest elevation. Thus, the zone's vulnerability to sea level rise could be heightened due to a combination of low elevation and proximity to water bodies. Accordingly, an additional variable expressing sea level rise may benefit modeling endeavors for the zone. Regarding zone 0, the lower performance within the cluster compared to zone 5 may be attributed to potential bias within the crowdsourced platform. Agonafir et al found that commuters who drive are more likely to report SF (Agonafir et al., 2022). Given that zone 0 encompasses various Manhattan neighborhoods, and Manhattan is the borough with the highest influx of commuters (City of New York, 2019), employing both public transportation and vehicles, the crowdsourced response data may exhibit subtle inconsistencies, posing challenges in detecting flooding patterns. Hence, the model performance in a zone may be affected by the amount of SF complaints, an insufficient set of variables or bias in the crowdsourced platform.

Next, the zonal analysis will gear towards the examination of individual model performance. An evident observation of the model results is the GWN exhibiting the highest R^2 [oftentimes, at a great margin] across all the zones, except for one. Now, as the sole deviation from this trend of GWN dominance occurs at zone 0, the continued examination at the zonal level will begin at this curious exception. In the case of zone 0, the GCN outperforms the GWN by a marginal delta of 0.02. The difference is slight, as thus, considering the low volume of response data in the sector, the variation likely does not hold significance in model comparison. Nevertheless, within zone 0, there exist remarkable takeaways. The graph neural networks (GCN and GWN), when compared to the FFN, RNN, and CNN, demonstrate stronger prediction accuracies. As the differences range from 0.09 to 0.12 in R^2 values at zone 0, the results pronounce the benefits of the graph structure. The graph neural networks add value by not only including locational elements, but by also allowing various areas to be connected and communicate with each other, and for a location, such as zone 0, with limited data, feeding a model with added information assists in the learning of environmental patterns. This merit of the graph-based architecture is additionally seen when inspecting the results of zone 1, 2, 4, and 5. Specifically, zone 5, also with low response data (the total SF complaints of zone 5 are less than a third of zone 3), obtains GCN and GWN R^2 values, ranging from deltas of 0.10 to 0.19 greater than those of the non-graph-based ANN models. The region where there is a balance in model performance between the GCN and the FFN, RNN, and CNN is zone 3. Since zone 3 has a high number of SF complaints, the non-graph-based ANNs are not as encumbered by low data volume; hence, their performance is competitive with the GCN. Nonetheless, for all zones, the GWN transcends, and, due to key elements in the graph-

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

based neural network structure, such as neighboring nodal data gains, the GCN also attains strong prediction results.

Also, at a zonal inspection, it is observed that the ANNs overwhelmingly outperform the GLM. The difference in R^2 values from GLM to an ANN range from 0.05 to 0.17 for zone 0, 0.19 to 0.39 for zone 1, 0.18 to 0.34 for zone 3, 0.14 to 0.30 for zone 4, 0.06 to 0.25 for zone 5. Hence, it is seen that the employment of even a simple neural network may have substantial benefits to data-driven urban flood modeling. The only exception to the observation lies in zone 2. In zone 2, the GLM achieves a higher R^2 than the FFN, RNN, and CNN by a difference of 0.07, 0.01, and 0.02, respectively. Zone 2 has the lowest SF reports, at roughly half of the response data as that of zone 0 (the zone with the second fewest) and less than one-eighth the SF complaints as zone 1 (the zone with the highest SF complaints). Therefore, an inference is that the GLM is not necessarily outperforming the other ANNs; however, due to very limited response data, the other ANNs are not performing at their fullest aptitudes.

Finally, while the zonal analysis has illustrated the advantages of ANN modeling for urban flooding and the benefits of applying a graph-based structure, the examination principally underscores the aptitude of the GWN. In Fig. 9, the comparative analysis demonstrates that under conditions where the R^2 values of the FFN, CNN, or RNN edge closely to those of the GCN, the GWN maintains a dominant performance. Conversely, when the validation determinations of the GCN ascend the other ANNs, the GWN upholds its position. The observed outcomes stem from the comprehensive nature of the GWN, incorporating key elements from preceding ANN models. These advanced features include the integration of convolutional and LSTM layers and the incorporation of the self-adaptive adjacency matrix within the graph architecture. Moreover, the structure within each component of the GWN are multiplexed, as opposed to simple additions. In a latter section, 4.3 Graph Wavenet Deconstruction, the ancillary complexities and their influence on model results will be explored in more detail.

4.2.3 Feature Importance

To illustrate how flood-related factors affect a model's performance, a feature importance analysis is conducted. The infrastructural dynamic variables (SB and CB) are removed from each model, and the results are compared against the original R^2 values of each model (when all the variables are present). Likewise, the climatic dynamic variables (MR, TR, and SN) are also removed and compared against the original R^2 values. This approach, which assumes a linear relationship between input variables and the target variable, provides insights into the contributions of the variables in isolation. The difference in R^2 values for each set of variable removal from the original R^2 is divided by the original R^2 to see the extent of effect. The results are shown in Fig. 10. While acknowledging the simplicity of this method and its reliance on linear relationships, it offers an interpretable way to rank the importance of dynamic predictors within the context of the models.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

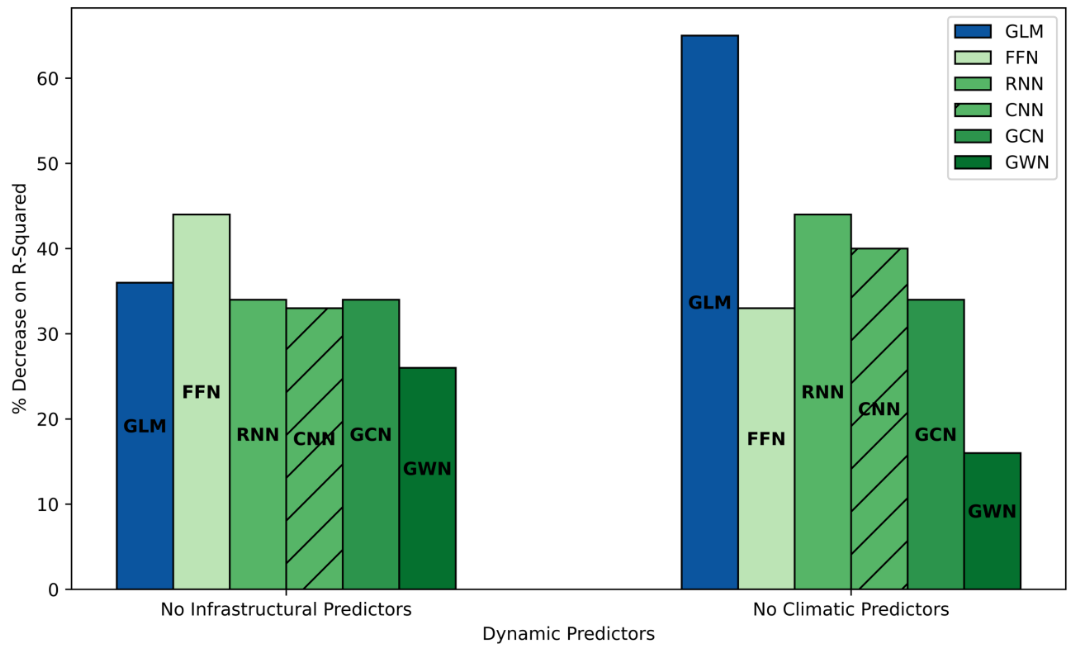

Fig 10 Plots of R^2 decrease with variable exclusion.

There are a few notable observations of the feature importance results. First, the most perceptible effect is that of the climatic variables on the GLM. The GLM experiences significant decrease when the precipitation predictors are absent. As precipitation is the fundamental cause of urban flooding, without its representation, the simplistic calculations of the GLM do not suffice for strong prediction capabilities. Next, it is seen that the performance of the FFN is more reliant on the infrastructural variables than that of the CNN and RNN. An implication of this finding may be that the CNN and RNN models have more adept utilization of the seasonality and temporal nature of precipitation occurrence due to their enhanced spatial and sequential pattern recognition time-series data. Lastly, an observation is that the GWN appears to be less dependent on either set of variables. Of all the models, the R^2 of the GWN decreases by the least relative extent for both exclusions. This strengthens the assertion that the GWN is more robust, as it possesses an improved capability of acquiring the environment, despite being given a lean set of variables. Nonetheless, there are differences in performance decline, as the GWN performs better with the set of infrastructural-only variables. This observation may be attributed to its graph-based structure. Particularly, the graph-based models are aided by locational information, and the infrastructural predictors allow the models to sense the presence of chaos. If there are sewer backups and catch basin issues reported throughout the city, the models are alerted towards a more probable occurrence of street flooding. On an ending note, the feature importance plots identify model aptitude within a particular set of variables, further dissecting key model strengths and limitations.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

4.3 Graph Wavenet Deconstruction

The GWN, extending beyond a basic encompassment of the characteristics of the ANN models in this study, elevates each fusion with intricate compositions and pathways. For instance, in its graph-based structure, although the foundational nodes and messaging through vertices (edges) are shared by the GCN, the GWN transcends with the incorporation of the advanced adjacency matrix. The matrix, skillfully integrating static attributes and facilitating early temporal review through sequenced data during the node-edge creation process, provides tailored inputs to each filter. Furthermore, in regards to the recurrent and convolutional aspects, the GWN builds upon the passage of information via dilation factors and residual and skip connections, exceeding the simplified structures of the RNN and CNN. Specifically, the RNN features a feedback mechanism via the LSTM layer, which learns information via a lookback period of six days. While the GWN also includes a lookback of six days, the temporal learning onsets at the adjacency matrix; additionally, the inputs proceed through gated activation functions and merge by element wise multiplication, as opposed to sequential summation processes. This careful procedure allows for improved biases and weights; moreover, it reduces the occurrences of vanishing gradients or exploding activations, which are known risks of RNN modeling (H. Lin et al., 2022; Rangapuram et al., 2018). Now, transitioning to the comparison of the convolutional structure between the GWN and the CNN, the layers of the GWN exhibit a higher level of knowledge transfer. First, the dilation factor of the GWN allows for different convolutional layers to capture varying time ranges (Rathore et al., 2021); therefore, each layer brings a distinctive evaluation of pattern recognition, allowing for a holistic perception of the temporal environment. Second, the residual and skip connections enable a direct extraction from layers with pertinent information. To prevent distortion of the vital information as it passes sequentially layer to layer until reaching the current layer, the skip connections allow the current layer to retrieve the information a previous (not immediately preceding) layer before it undergoes subsequent convolutions. As thus, the sophistication of the GWN convolutions allow for a more evolved and exclusive learning progression. Due to the complexity of the graph, recurrent and convolutional fundaments, the GWN not only incorporates, but ascends.

To highlight the improvements facilitated by the added components of the GWN, an additional model run is performed. In this GWN simulation (hereinafter referred to as GCR), the adjacency matrix is excluded, while preserving spatial and temporal aspects. The mean and median R^2 values are determined to be 0.46 and 0.47, respectively (Table A1). When contrasting the GCR with the GWN, which achieves mean and median values of 0.51 and 0.54, respectively, the impact of excluding the adjacency matrix becomes evident. The GCR experiences a decrease of 0.05 in the mean and 0.07 in the median. Additionally, at the zonal level, every cluster demonstrates a reduction in performance. Furthermore, this isolation of the GWN from the adjacency matrix emphasizes the vital role played by the intricate convolutional and recurrent features of the GWN. Notably, the primary distinction between the GCN and GCR lies in the temporal and convolutional aspects. The results, wherein the GCR outperforms the GCN with a mean and median R^2 difference of 0.03 and 0.04, respectively, emphasize the strengths of temporal and spatial learning introduced by the GWN. Therefore, as theorized, the success of the GWN may be attributed to the multiple traits within the algorithm.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

4.4 Limitations and Future Considerations

4.4.1 The FFN, RNN and CNN

The added complexities by the RNN and CNN yield minimal model performance. The FFN obtains a mean R^2 value of 0.35, and the RNN and CNN each obtain mean R^2 values of 0.36. Only in zone 2, there exists a substantive improvement by the RNN and CNN, where the R^2 values are 0.06 and 0.05 greater than the FFN, respectively. A plausible explanation for the limited benefits is the simplicity of the model layers, of which lack the detail and mechanisms needed to produce discernable results. This assertion gains credibility when considering the performance of the GWN, where an interleaved system of data flow, employing skip connections and advanced gating mechanisms, achieves superior prediction accuracy. Another contributing factor may be the absence of significant temporal dependency within the dataset. Given the nature of urban flooding, often ensuing in the form of a flash flood, where the onset and finality of the disaster occurs within a brief timespan of 6 hours (NWS, 2022), flooding on one day, oftentimes, does not exert influence on the following day. Simplified models designed to capture spatial and temporal dependencies may overlook these subtle patterns. Lastly, many of the variables, including the response variable, are retrieved from the NYC 311 dataset, and crowdsourced data is not as accurate in illustrating the environment as physical measurements. Thus, a simplified feedback loop or spatial assessment may not suffice. The limitations of the crowdsourced platform are further discussed in the next section. In summary, while the neural network architecture attains noticeable improvements in model performance, the basic LSTM or convolutional layers are not as advantageous; this may be attributed to the temporal nature of urban flooding and the limitations of crowdsourced data.

4.4.2 The Crowdsourced Platform

Crowdsourcing has been applied in previous urban flood modeling initiatives, particularly in cities like NYC, where flood data is scarce. The incorporation of residential reports provides insights into flooding occurrences, of which, otherwise, would not be obtained. However, while this method is valuable, the leveraging of eyewitness accounts is not as exact as physical measurements. Also, another issue of crowdsourced data is the potential for bias – a greater inclination of certain types of people to report issues. For example, it has been indicated that certain socio-demographical attributes may be factors in SF complaints, thereby possibly indicating that a particular set of residents are more likely to utilize the platform (Agonafir et al., 2022). Hence, there are bias concerns. Nonetheless, there exists strength in the reports, as they are taken by individuals observing an event. For instance, the bias attributed to socio demographics is a *potential* consideration, since it may also be inferred that specific neighborhoods prone to flood occurrences may be comprised of a certain set of socio-demographics. Moreover, the validity of the crowdsourced data's depiction of flooding occurrence is reinforced by the climatic, topographical, and infrastructural predictors holding substantial significance in the crowdsourced response variable, SF complaints (Agonafir et al., 2021, 2022). Furthermore, the 311 NYC street flooding reports, at the very least, capture the concerns of residents inclined to report, potentially identifying those most at personal risk. Indeed, the attribute of commuters who drive holds a sizable relative importance in Agonafir et

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

al. (Agonafir et al., 2022), and as the leading cause of death from flooding is vehicular, accurate predictions in these regions may prove lifesaving. In essence, in metropolitans with limited data, crowdsourcing, despite some drawbacks, enables the continuity of predictive modeling, sustaining efforts that would otherwise cease.

Moreover, in metropolitans, particularly in the specific study area of NYC, flood sensors [measuring water levels] are being installed, enabling proximate applicability of this explorative analysis. The findings presented here accentuate the models best suited for the local landscape, with the GWN delivering promising results. Despite the constraints of the crowdsourced platform, the GWN attains an R^2 of 0.72 for zone 3, demonstrating its potential. Anticipating even greater predictive accuracy with actual measurements from sensors, this study outlines techniques applicable to the urban city. Once physical data becomes accessible, this model exploration provides policymakers and stakeholders with an outline of the strengths and weaknesses of models, ascending in complexity, while also pinpointing the overall, most effective model for forecasting floods during a predicted rain event.

5. Conclusions

By a diverse, novel pooling of machine learning techniques, this study advances our understanding of urban flooding, offering detailed insights into risk zones, comparing the performance of various models, and emphasizing the effectiveness of graph-based neural networks, particularly the Graph Wavenet. Listed below are the key appreciations of the research:

- Spectral clustering has utilization in risk zone identification and border delineation. The analysis of these zones reveals relationships between specific physical characteristics (such as slope, elevation, and building density) and the occurrence of street flooding. Notably, zones with higher elevation and lower building density exhibited lower susceptibility to flooding, emphasizing the importance of urban characteristics in flood risk assessment.
- Machine learning models demonstrate superior performance to the GLM. Unlike the GLM, which assumes linearity in the parameters, machine learning models offer greater flexibility by adapting to complex, nonlinear patterns present in the data.
- By a systematic evaluation of the performance of varying flood prediction models, ranging from traditional statistical models to advanced neural networks, the GWN emerges as the most suitable model for urban flood forecasting in NYC, outperforming other models, including the GCN, CNN and RNN. Hence, there is significance in incorporating advanced spatio-temporal aspects and dynamic graph creation for accurate flood forecasting.
- Via zonal analysis, graph-based structures are shown to be particularly beneficial in areas with sparse data, where traditional models may struggle.

The discoveries of this research accord practical value onto urban stakeholders, especially in cities such as NYC, where water measurement sensors are currently being deployed. This synergy between advanced modeling techniques, particularly spatio-temporal graph neural

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

networks, and emerging sensor technologies ensures informed decision-making, enabling urban planners and emergency responders to safeguard communities, reduce economic losses, and enhance overall resilience to the challenges posed by urban flooding.

6. Appendix A

Table A.1

Zones	0	1	2	3	4	5	Median	Mean
GCR	0.29	0.56	0.28	0.68	0.52	0.42	0.47	0.46

7. Code Availability

The processed data and the codes used in this study are available from the corresponding authors upon reasonable request.

8. Data Availability

Physical, topographical, and crowdsourced data from this study are publicly available at NYC Open Data and NYC311. Radar and gauge data are available from EOL/UCAR and NOAA databases, respectively.

9. Author Contribution

Candace Agonafir: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft and Preparation. **Tian Zheng**: Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Resources, Supervision.

10. Competing Interests

The authors declare that they have no conflict of interest.

11. Acknowledgements

The authors would like to acknowledge funding from NSF Grant 2019625. We also thank Dr. Upmanu Lall for providing valuable feedback on this work.

12. References

AghaKouchak, A., Nasrollahi, N., & Habib, E. (2009). Accounting for uncertainties of the TRMM satellite estimates. *Remote Sensing*, *1*(3), 606–619. https://doi.org/10.3390/rs1030606

Agonafir, C., Lakhankar, T., Khanbilvardi, R., Krakauer, N., Radell, D., & Devineni, N. (2022). A Machine Learning Approach to Evaluate the Spatial Variability of New York City's 311 Street Flooding Complaints. *Computers, Environment and Urban Systems*.

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

Agonafir, C., Lakhankar, T., Khanbilvardi, R., Krakauer, N., Radell, D., & Devineni, N. (2023). A review of recent advances in urban flood research. In *Water Security* (Vol. 19). Elsevier B.V. https://doi.org/10.1016/j.wasec.2023.100141

Agonafir, C., Ramirez Pabon, A., Lakhankar, T., Khanbilvardi, R., & Devineni, N. (2021). Understanding New York City Street Flooding through 311 Complaints. *Journal of Hydrology*, *605*(March 2021), 127300. https://doi.org/10.1016/j.jhydrol.2021.127300

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*, 1–6. https://doi.org/10.1109/ICEngTechnol.2017.8308186

Al-Suhili, R., Cullen, C., & Khanbilvardi, R. (2019). An urban flash flood alert tool for megacities-Application for Manhattan, New York City, USA. *Hydrology*. https://doi.org/10.3390/HYDROLOGY6020056

Bruwier, M., Maravat, C., Mustafa, A., Teller, J., Pirotton, M., Erpicum, S., Archambeau, P., & Dewals, B. (2020). Influence of urban forms on surface flow in urban pluvial flooding. *Journal of Hydrology*, *582*(December 2019). https://doi.org/10.1016/j.jhydrol.2019.124493

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. http://arxiv.org/abs/1412.3555

City of New York. (2019, September). *The Ins and Outs of NYC Commuting*. https://www.nyc.gov/assets/planning/download/pdf/planning-level/housing-economy/nyc-ins-and-out-of-commuting.pdf

City of New York. (2022a). *Flood Prevention*. https://www1.nyc.gov/site/dep/environment/flood-prevention.page

City of New York. (2022b). *Impact of NYW Bonds*. https://www1.nyc.gov/site/nyw/investing-in-nyw-bonds/the-impact-of-investing.page

City of New York. (2023a). *About Commmunity Boards*. https://www.nyc.gov/site/cau/community-boards/about-commmunity-boards.page

City of New York. (2023b). *NYC Open Data*. https://data.cityofnewyork.us/City-Government/Community-Districts/yfnk-k7r4

Dubey, A. K., & Jain, V. (2019). Comparative Study of Convolution Neural Network's Relu and Leaky-Relu Activation Functions. In S. Mishra, Y. R. Sood, & A. Tomar (Eds.), *Applications of Computing, Automation and Wireless Systems in Electrical Engineering* (pp. 873–880). Springer Singapore.

Durairaj, D. M., & Mohan, B. H. K. (2022). A convolutional neural network based approach to financial time series prediction. *Neural Computing and Applications*, *34*(16), 13319–13337. https://doi.org/10.1007/s00521-022-07143-2

EOL. (2022). *NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data*. https://data.eol.ucar.edu/cgi-bin/codiac/fgr_form/id=21.093

Falconer, R. (2021). Hurricane Ida death toll rises past 60. *Axios*. https://www.axios.com/2021/09/06/hurricane-ida-death-toll-rises-power-outage

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

Farahmand, H., Xu, Y., & Mostafavi, A. (2023). A spatial–temporal graph deep learning model for urban flood nowcasting leveraging heterogeneous community features. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-32548-x

Fei, J., & Lu, C. (2018). Adaptive Sliding Mode Control of Dynamic Systems Using Double Loop Recurrent Neural Network Structure. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(4), 1275–1286. https://doi.org/10.1109/TNNLS.2017.2672998

Fine, T. L. (2006). *Feedforward Neural Network Methodology*. Springer New York. https://books.google.com/books?id=s-PlBwAAQBAJ

Hardin, J. W., & Hilbe, J. M. (2007). *Generalized Linear Models and Extensions* (Second). Stata Press.

Jiang, B., Wang, B., & Luo, B. (2023). Sparse norm regularized attribute selection for graph neural networks. *Pattern Recognition*, *137*, 109265. https://doi.org/https://doi.org/10.1016/j.patcog.2022.109265

Koprinska, I., Wu, D., & Wang, Z. (2018). Convolutional Neural Networks for Energy Time Series Forecasting. *2018 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN.2018.8489399

Lin, H., Gharehbaghi, A., Zhang, Q., Band, S. S., Pai, H. T., Chau, K.-W., & Mosavi, A. (2022). Time series-based groundwater level forecasting using gated recurrent unit deep neural networks. *Engineering Applications of Computational Fluid Mechanics*, *16*(1), 1655–1672. https://doi.org/10.1080/19942060.2022.2104928

Lin, J., He, X., Lu, S., Liu, D., & He, P. (2021). Investigating the influence of three-dimensional building configuration on urban pluvial flooding using random forest algorithm. *Environmental Research*, *196*, 110438. https://doi.org/10.1016/J.ENVRES.2020.110438

Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, *10*(11). https://doi.org/10.3390/w10111536

National Centers for Environmental Information. (2023). *Global Historical Climatology Network daily (GHCNd)*. https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily

NWS. (2022). *Flash Flooding Definition*. https://www.weather.gov/phi/FlashFloodingDefinition

Oliveira Santos, V., Costa Rocha, P. A., Scott, J., Thé, J. V. G., & Gharabaghi, B. (2023). A New Graph-Based Deep Learning Model to Predict Flooding with Validation on a Case Study on the Humber River. *Water (Switzerland)*, *15*(10). https://doi.org/10.3390/w15101827

Ouma, Y. O., & Tateishi, R. (2014). Urban Flood Vulnerability and Risk Mapping Using Integrated Multi-Parametric AHP and GIS: Methodological Overview and Case Study Assessment. In *Water* (Vol. 6, Issue 6). https://doi.org/10.3390/w6061515

Piao, Y., Lee, S., Lee, D., & Kim, S. (2022). *Sparse Structure Learning via Graph Neural Networks for Inductive Document Classification*. https://github.com/qkrdmsghk/TextSSL

Plumer, B. (2021). Flooding From Ida Kills Dozens of People in Four States. *The New York Times*. https://www.nytimes.com/live/2021/09/02/nyregion/nyc-storm

PyTorch. (2023a). *CONV1D*. https://pytorch.org/docs/stable/generated/torch.nn.Conv1d.html

PyTorch. (2023b). *LSTM*. https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html

PyTorch. (2023c). *Module*. https://pytorch.org/docs/stable/generated/torch.nn.Module.html

PyTorch Geometric. (2023). *Torch_gemoteric.nn*. https://pytorch-geometric.readthedocs.io/en/latest/modules/nn.html

Qin, H. peng, Li, Z. xi, & Fu, G. (2013). The effects of low impact development on urban flooding under different rainfall characteristics. *Journal of Environmental Management*, *129*, 577–585. https://doi.org/10.1016/J.JENVMAN.2013.08.026

Rangapuram, S. S., Seeger, M. W., Gasthaus, J., Stella, L., Wang, Y., & Januschowski, T. (2018). Deep state space models for time series forecasting. *Advances in Neural Information Processing Systems*, *31*.

Rathore, N., Rathore, P., Basak, A., Nistala, S. H., & Runkana, V. (2021). Multi Scale Graph Wavenet for Wind Speed Forecasting. *2021 IEEE International Conference on Big Data (Big Data)*, 4047–4053. https://doi.org/10.1109/BigData52589.2021.9671624

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, *20*(1), 61–80. https://doi.org/10.1109/TNN.2008.2005605

Schmidt, R. M. (2019). *Recurrent Neural Networks (RNNs): A gentle Introduction and Overview*. http://arxiv.org/abs/1912.05911

Schmitt, T. G., Thomas, M., & Ettrich, N. (2004). Analysis and modeling of flooding in urban drainage systems. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2004.08.012

Scikit-learn. (2023). *Sklearn.cluster.SpectralClustering*. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.SpectralClustering.html

Semádeni-Davies, A., & Bengtsson, L. (1998). Sensibilité de la fonte des neiges au rayonnement en milieu urbain. *Hydrological Sciences Journal*, *43*(1), 67–89. https://doi.org/10.1080/02626669809492103

Setiono, R. (2001). Feedforward Neural Network Construction Using Cross Validation. *Neural Computation*, *13*(12), 2865–2877. https://doi.org/10.1162/089976601317098565

Sharif, H. O., Yates, D., Roberts, R., & Mueller, C. (2006). The use of an automated nowcasting system to forecast flash floods in an urban watershed. *Journal of Hydrometeorology*. https://doi.org/10.1175/JHM482.1

Sun, A. Y., Jiang, P., Mudunuru, M. K., & Chen, X. (2021). Explore Spatio-Temporal Learning of Large Sample Hydrology Using Graph Neural Networks. *Water Resources Research*, *57*(12). https://doi.org/10.1029/2021WR030394

Sun, M., Song, Z., Jiang, X., Pan, J., & Pang, Y. (2017). Learning Pooling for Convolutional Neural Network. *Neurocomputing*, *224*, 96–104. https://doi.org/https://doi.org/10.1016/j.neucom.2016.10.049

https://doi.org/10.5194/egusphere-2024-551
Preprint. Discussion started: 15 March 2024

Thorndahl, S., Einfalt, T., Willems, P., Ellerbæk Nielsen, J., Ten Veldhuis, M. C., Arnbjerg-Nielsen, K., Rasmussen, M. R., & Molnar, P. (2017). Weather radar rainfall data in urban hydrology. *Hydrology and Earth System Sciences*, *21*(3), 1359–1380. https://doi.org/10.5194/hess-21-1359-2017

Tsantekidis, A., Passalis, N., & Tefas, A. (2022). Chapter 5 - Recurrent neural networks. In A. Iosifidis & A. Tefas (Eds.), *Deep Learning for Robot Perception and Cognition* (pp. 101–115). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-32-385787-1.00010-5

U.S. Census Bureau QuickFacts: New York city, New York; Bronx County (Bronx Borough), New York; Kings County (Brooklyn Borough), New York; New York County (Manhattan Borough), New York; Queens County (Queens Borough), New York; Richmond County (Staten Isl. (n.d.). https://www.census.gov/quickfacts/fact/table/newyorkcitynewyork,bronxcountybronxboroughnewyork,kingscountybrooklynboroughnewyork,newyorkcountymanhattanboroughnewyork,queenscountyqueensboroughnewyork,richmondcountystatenislandboroughnewyork/PST045219

Valeo, C., & Ho, C. L. I. (2004). Modelling urban snowmelt runoff. *Journal of Hydrology*, *299*(3–4), 237–251. https://doi.org/10.1016/J.JHYDROL.2004.08.007

von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, *17*(4), 395–416. https://doi.org/10.1007/s11222-007-9033-z

von Luxburg, U., Belkin, M., & Bousquet, O. (2008). Consistency of Spectral Clustering. *The Annals of Statistics*, *36*(2), 555–586. http://www.jstor.org/stable/25464638

Wang, K., Li, K., Zhou, L., Hu, Y., Cheng, Z., Liu, J., & Chen, C. (2019). Multiple convolutional neural networks for multivariate time series prediction. *Neurocomputing*, *360*, 107–119. https://doi.org/https://doi.org/10.1016/j.neucom.2019.05.023

Wang, Y., Bu, F., Lv, X., Hou, Z., Bu, L., Meng, F., & Wang, Z. (2023). Attention-based message passing and dynamic graph convolution for spatiotemporal data imputation. *Scientific Reports*, *13*(1), 6887. https://doi.org/10.1038/s41598-023-34077-z

Wang, Z., Lyu, H., & Zhang, C. (2023). Pluvial flood susceptibility mapping for data-scarce urban areas using graph attention network and basic flood conditioning factors. *Geocarto International*, *38*(1). https://doi.org/10.1080/10106049.2023.2275692

Woodruff, J. D., Irish, J. L., & Camargo, S. J. (2013). Coastal flooding by tropical cyclones and sea-level rise. *Nature*, *504*(7478), 44–52. https://doi.org/10.1038/nature12855