**Integrating Machine Learning And Subsurface Characterization For Improved Urban Flood Modelling And Parameter Optimization**

**Abstract**

Flooding is one of the worst natural disasters which endangers lives and hampers economic activities across the globe. Construction and calibration of physical models which includes computational resources and data are the primary components of urban flood modeling. In this research, a paradigm shift is proposed by combining subsurface data with ML methodologies which advance urban flood modeling techniques as well as parameter tuning. Model accuracy and robustness are achieved through parametric value enhancement. The findings obtained shed light on the potential improvements that were observed in urban flood modeling. Such changes will facilitate better management in the contexts of urban infrastructure development, risk assessment, and hydrological resource optimization. The strategic evaluation optimization technique relies on the K-means ANN driven Genetic Algorithm when dealing with urban inundation parameters. This technique identifies sensitive parameter values while also providing optimal results to enhance model effectiveness. Subsequently, the designed inundation model underwent validation through simulating real-world rainfall data followed by an evaluative approach to test the accuracy of the proposed predictive methodology.

**Keyword. Flood, integrating, machine learning, modelling, Urban, Parameter.**

**Introduction**

“In recent years, urban flood disasters extremes caused by rainstorm have become increasingly serious threatening urban infrastructure and resident's life safety” (Nguyen and Bae, 2020; Lu and Sun, 2021). “Precise flood prediction plays a pivotal role in preventing disasters of this kind, and the creation of credible urban waterlogging models is a prerequisite in flood control” (Chang et al., 2021; Li et al., 2022; Nandi and Reddy, 2022). “And yet, such models depend on many parameters” (Zeng et al., 2020; Liao et al., 2022), “some of which are very sensitive and require precise calibration. Many parameters in urban waterlogging models cannot be directly measured from catchment characteristics” (Sinnathamby et al., 2017; Guo and Su, 2019; Wang et al., 2020). “The accuracy of runoff depth simulations heavily relies on the definition of relevant parameters” (Huo and Liu, 2019; Feigl et al., 2022a). Conducting sensitivity analysis of model parameters is crucial for improving simulation efficiency and accuracy. Nevertheless, the classical sensitivity analysis is cumbersome and time-consuming, which is not conducive to the efficiency of an urban waterlogging model being tested.

In this context efficient selection and calibration of sensitive parameters becomes crucial (Wood et al., 2016, Willis et al., 2019). Parameter calibration optimization is one of the important steps in the simulation of urban waterlogging model (Bingyan Ma, 2020), which could be developed with manual or automatic calibration approaches (Jung et al., 2017, Feigl et al., 2022; Katipoğlu, 2023). Manual calibration used to be the prevailing approach, but it is a laborious and time-consuming one. To address these limitations, other methodologies have been investigated. New studies have developed the computer-intelligent optimal methods to improving urban flood modeling Azadgar (2025). For instance, Wu et al. (2021 developed a deep learning-based approach to optimize uncertainty parameters in flood processes. Yuan et al. (2021) implemented “automatic calibration of rainfall-runoff model parameters using a backpropagation neural network algorithm”. Wang et al. (2022) addressed the limitation of manual parameter calibration in flood models and proposed a genetic algorithm-based method for automatic calibration. The rapid evolution of deep learning techniques has transformed the field of hydrological modelling. Unlike traditional physically-based models, deep learning methods operate as "black boxes" (Adnan et al., 2021), capturing complex relationships and trends in data through extensive training (Yan et al., 2021; Ye et al., 2022). “This has led to increased interest in applying artificial neural networks (ANNs) and deep learning techniques for rapid identification of sensitive parameters in urban waterlogging models. Traditional urban inundation models rely on overflow processes at overflow nodes and urban topography, neglecting the impact of underground spaces” (Zeng et al., 2017; Dao et al., 2022; Yang et al., 2022). To address this limitation, this study combines a one-dimensional pipe network routing model with a two-dimensional surface runoff model to create a comprehensive urban waterlogging model.

“Although advances in computing power and machine learning have improved the efficiency of identifying and optimizing sensitive parameters in urban flood models Hengxu Jin 2024, current research often overlooks the physical implications of these parameters and their relationships with complex urban surfaces” (Zang et al., 2022). Parameters are frequently assigned using simplistic approaches, and sensitivity analyses involve cumbersome and intricate processes, such as repeated model simulations.

“Furthermore, previous studies have directly applied one-dimensional hydrological model simulations to represent surface inundation, disregarding two-dimensional hydrodynamic processes of surface water runoff” (Cai et al., 2019). Investigating differences in sensitive parameters across various land use functional zones in cities can enhance the accuracy and efficiency of urban flood simulations.

This study proposes a principle for dividing urban hydrological response units based on a coupled pipe network and surface model, incorporating surface attribute features. K-means clustering is then employed to explore clustering patterns of uncertain model parameters and identify sensitive parameters using artificial neural networks. Finally, a genetic algorithm is used to calibrate threshold values of sensitive parameters for sub-watershed units in different land use functional zones

The K-means clustering algorithm has been widely adopted in flood forecasting research due to its simplicity and efficiency (Xu, 2015). Li et al. (2016) combined flood similarity analysis with K-means clustering for flood classification and forecasting in a transitional river basin. Hu et al. (2022) developed “a rapid flood classification forecasting model using K-means and backpropagation neural networks”. Sun et al. (2022) applied “an improved sub-watershed division approach combined with K-means clustering for parameter calibration in the Storm Water Management Model (SWMM)”.

**Review Of Related Literature**

Urban flooding is a growing concern worldwide Agasnalli et al (2021), causing significant economic and social impacts. Traditional urban flood modelling approaches rely heavily on physically-based models, which often struggle to accurately represent complex urban environments (Kumar et al., 2018). Recent advances in machine learning (ML) and subsurface characterization offer promising opportunities to improve urban flood modelling and parameter optimization.

ML algorithms, such as artificial neural networks (ANNs) and support vector machines (SVMs), have been successfully applied to urban flood modeling, demonstrating improved accuracy and efficiency compared to traditional methods (Chen et al., 2019; Liu et al., 2020). Subsurface characterization, including the use of geophysical and hydrogeological data, can provide valuable insights into the underlying hydrological processes controlling urban flooding (Singh et al., 2019).

However, few studies have explored the integration of ML and subsurface characterization for urban flood modelling and parameter optimization. This knowledge gap highlights the need for innovative approaches that leverage the strengths of both ML and subsurface characterization.

Urban Flood Modelling: Urban flood modelling refers to the use of mathematical models to simulate and predict flood events in urban areas. These models take into account various factors such as rainfall, topography, drainage systems, and land use to predict flood extent, depth, and velocity. Machine Learning: Machine learning (ML) is a subset of artificial intelligence (AI) that involves training algorithms to learn from data and make predictions or decisions without being explicitly programmed. ML algorithms can be applied to various tasks, including classification, regression, clustering, and optimization.

Subsurface Characterization: Subsurface characterization refers to the process of identifying and describing the physical and hydrological properties of the subsurface environment, including soil, rock, and groundwater. This information is essential for understanding and predicting groundwater flow, contaminant transport, and other subsurface processes.

Parameter Optimization: Parameter optimization refers to the process of adjusting the parameters of a mathematical model to achieve the best possible fit between the model predictions and observed data. This is often done using optimization algorithms, such as genetic algorithms or particle swarm optimization.

Hydrological Modelling: Hydrological modelling refers to the use of mathematical models to simulate and predict the behaviour of water in the environment, including precipitation, runoff, infiltration, and evaporation. Hydrological models are used to understand and predict various hydrological processes, including flood forecasting, water resources management, and climate change impacts.

Geophysical and Hydrogeological Data: Geophysical and hydrogeological data refer to the information collected from the Earth's subsurface using various geophysical and hydrogeological techniques, such as seismic surveys, electrical resistivity tomography, and groundwater sampling. This data is used to understand the subsurface structure, properties, and processes.

Artificial Neural Networks (ANNs): Artificial neural networks (ANNs) are a type of machine learning algorithm inspired by the structure and function of the human brain. ANNs consist of interconnected nodes (neurons) that process and transmit information. ANNs are widely used for various tasks, including classification, regression, and optimization.

Support Vector Machines (SVMs): Support vector machines (SVMs) are a type of machine learning algorithm used for classification and regression tasks. SVMs work by finding the hyperplane that maximally separates the classes in the feature space. SVMs are known for their ability to handle high-dimensional data and non-linear relationships.

**Objective of the Study:**

The objective of this study is to develop an integrated framework that combines machine learning algorithms with subsurface characterization techniques to improve urban flood modelling and parameter optimization. Specifically, this study aims to:

Develop a novel ML-based approach for urban flood modelling that incorporates subsurface characterization data.

Investigate the impact of subsurface heterogeneity on urban flood modelling and parameter optimization.

Evaluate the performance of the proposed framework using real-world case studies.

Purpose and scope of the Review:

The purpose of this review is to provide a comprehensive overview of the current state of knowledge on integrating machine learning and subsurface characterization for improved urban flood modelling and parameter optimization. The review aims to identify the key challenges, opportunities, and future directions in this field.

The scope of this review includes:

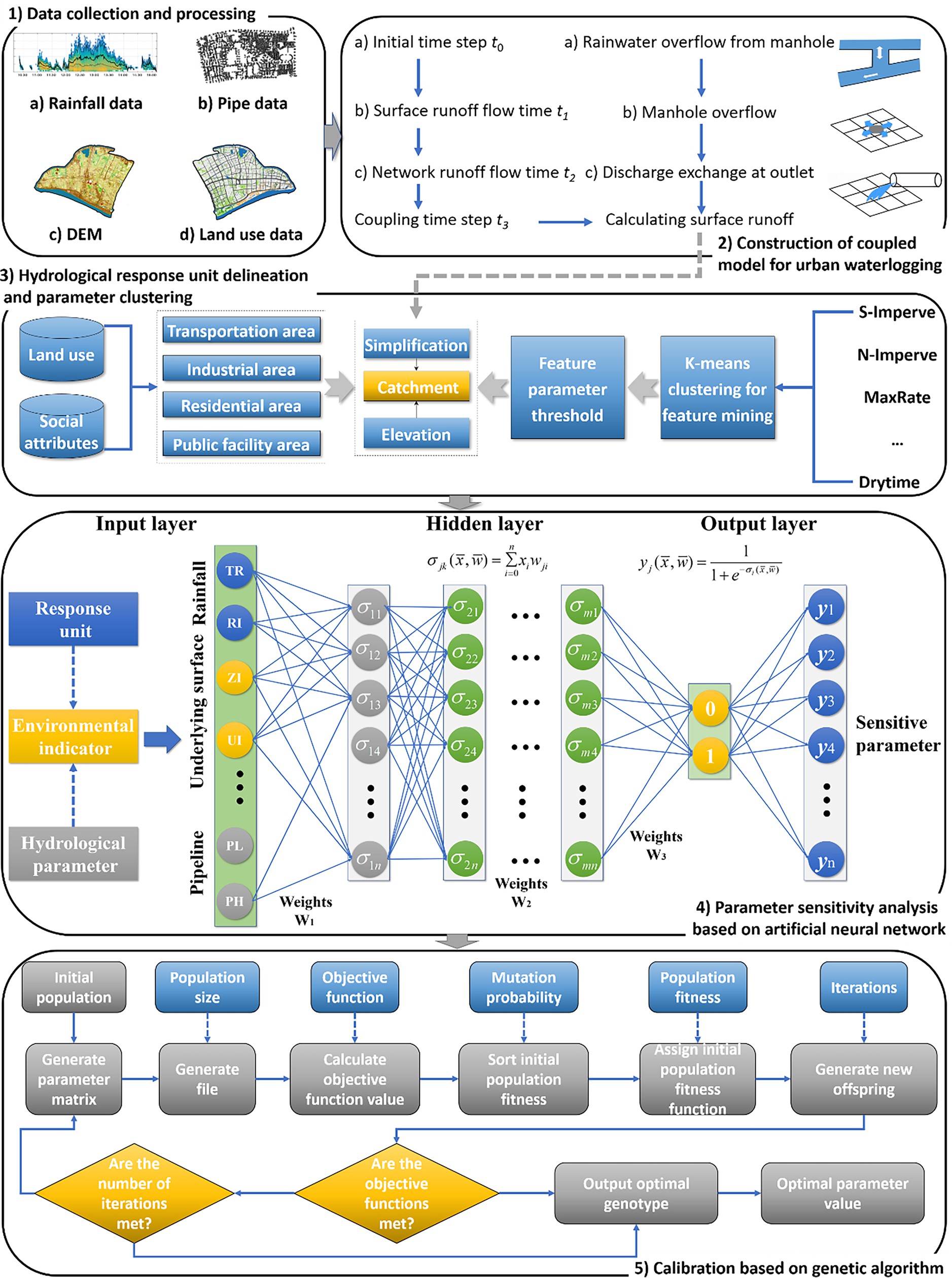
Urban flood modelling and simulation, Machine learning and artificial intelligence applications in urban flood modelling, Subsurface characterization and its integration with machine learning, Parameter optimization and sensitivity analysis in urban flood modelling. Case studies and applications of machine learning and subsurface characterization in urban flood modelling

List 1- Structural framework of the Review

|  |  |  |  |
| --- | --- | --- | --- |
| Column 1 | Column 2 | Column 3 | Column 4 |
| Objective of the Review | Comprehensive evaluation of the current state of knowledge on integrating machine learning and subsurface characterization for improved urban flood modeling. | Review the current state of knowledge on parameter optimization and sensitivity analysis in urban flood modeling. | Identify future directions in integrating machine learning and subsurface characterization for improved urban flood modeling |
| Focus Areas | Machine learning applications in urban flood modeling, including supervised, unsupervised, and reinforcement learning techniques | Subsurface characterization and its integration with machine learning, including soil properties, groundwater flow, and subsurface infrastructure. | Urban flood risk assessment and management, including hazard, vulnerability, and risk analysis. |
| Methodology | In-depth analysis of current literature | Mode Evaluation: Evaluate the performance of the machine learning models using metrics such as accuracy, precision, and recall. | Discussion of strategies for overcoming urban flood challenges |
| Expected Outcomes | Develop machine learning models that can accurately predict urban flood events using subsurface data and other relevant factors | Gain a better understanding of the importance of subsurface characterization in urban flood modelling and how it can be integrated with machine learning techniques | Recommendations for industry professionals and future research directions |

**Methodology**

Subsurface urban features in urban flood simulation typically face three fundamental challenges: poor integration of subsurface urban elements, ambiguous definition of hydrological response areas, and intricate parameter calibration processes. To counter these, this article combines subsurface characterization with a hybrid machine learning strategy-K-means clustering, Artificial Neural Networks (ANN), and Genetic Algorithms (GA)-to enhance parameter optimization and enhance flood modeling precision. Figure 1 depicts the overall methodological approach.

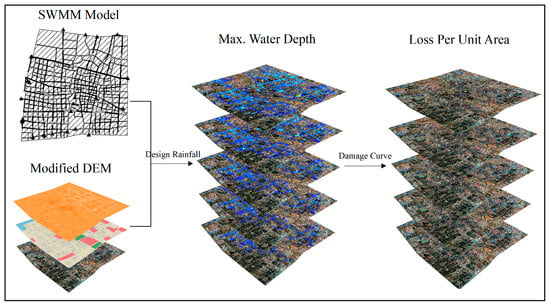


Hengxu Jin 2024

Fig 1-Research methodology

**Urban Flood Coupling Model Construction**

This research integrates a one-dimensional pipe network routing Quing (2024) model with a two-dimensional surface runoff model to develop a comprehensive urban waterlogging model Ma (2020). By combining these two models, this study aims to address the limitations of individual models and provide a more holistic understanding of urban flooding processes, taking into account both surface and subsurface water flow (Yang et al., 2022). This study utilizes a grid-based hydraulic computation model to simulate surface runoff processes in urban areas, leveraging raster data in a Geographic Information System (GIS) framework (Sosa et al., 2019; O'Loughlin et al., 2020; Shustikova et al., 2020). The model classifies underlying surfaces into four categories based on soil permeability: impermeable, semi-permeable, permeable, and highly permeable. By applying hydraulic methods to regular grid data Bommer (2019), the model calculates water exchange between grid cells, simulates water movement under gravity and structural influences, and generates water depth distribution results that align with the topographic grid.



**Figure 2: Urban Flood Coupling Model Construction Lidong Zhao (2021)**

**Construction of Urban Flood Coupling Model**

This study created a coupled urban inundation model by combining a one-dimensional (1D) pipe network flow routing system with a two-dimensional (2D) surface runoff model. The 1D model simulates subsurface flows (stormwater drainage pipes) and the 2D model simulates surface runoff over complex topography in the city. This coupling allows for the dynamic interactions between above- and below-ground water flows to be simulated.

Surface runoff was simulated with a raster hydraulic model inside a GIS environment with elevation information and surface permeability classes: impermeable, semi-permeable, permeable, and highly permeable areas. Hydraulic equations are used by the model to calculate gravity flow, intercell water exchange, and water depth distribution following Anouk Bommer (2019). It adds spatial reality and flood extent

**Urban Surface Water Flow and Pipe Network Modelling:**

In urban areas, surface water flows into low-lying areas and enters stormwater pipe networks through rainwater inlets. To simulate the hydraulic behaviour of the pipe network and stormwater nodes, the dynamic wave method is employed. As rainwater enters the pipe network, the flow state within the pipes transitions between open-channel flow and pressurized pipe flow (Ye et al., 2021). The Pressman virtual slit method and unsteady Saint-Venant equations are utilized to model stormwater runoff in the pipe network. The governing equations for these calculations are:

The stormwater pipe network runoff model is governed by the following equations:

6M/6t + 6Q/6x = q

6t + 6x a M

6y/6t + (u/g) (6u/6x) + (6h/6x) + Sf = 0

where:

x and y are distances in the Cartesian coordinate system

H is the surface water depth

t is time

J and K are discharges per unit width in the X and Y directions

g is the acceleration due to gravity

z is the water level

The pipe cross-sectional area (M), virtual slit width (N), flow rate (Q), lateral boundary inflow velocity (u), lateral boundary flow rate (q), and friction slope (Sf) are related by:

M = cross-sectional area of the pipe

N = width of the virtual slit

Q = flow rate in the pipe cross-section

u = lateral boundary inflow velocity along the pipe

q = lateral boundary flow rate

x = distance along the pipe

a = momentum correction coefficient

g = acceleration due to gravity

y = water head position

Sf = friction slope of the pipe

An explicit numerical algorithm was employed to solve the stormwater pipe network runoff model, utilizing hydraulic parameters and geometric characteristics obtained from the stormwater pipe network model data (Schilling and Tränckner, 2022).

**Overland Flow Model:**

The Overland Flow Model simulates surface runoff by calculating water depth distribution across the terrain, taking into account water exchange between grid cells May -Britt Moser (2015). This model accurately represents urban surfaces using regular grids.

By combining the Overland Flow Model with the Pipe Network Model Javier Farnendez Pato (2014), we can leverage their respective strengths to effectively simulate urban flooding processes. The coupled model enables the simulation of:

Surface water flow and inundation

Pipe network flow and interaction with surface water

**The coupling process involves the following Stages**

**Stage 1:** Establish a Detailed Pipe Network Runoff Model, develop a precise pipe network runoff model for the study area, preserving the integrity of flood-prone nodes.

**Stage 2:** Simulate Pipeline Overflow Dynamics, Execute the 1D pipe network runoff model to capture the excess flow mechanisms at critical pipeline junctures.

**Stage 3:** Quantify Overflow Discharge Rates, Calculate the volumetric flow rates of overflowing pipelines.

**Stage 4**: Integrate Pipe Network and Surface Runoff Models, Utilize the overflow dynamics and discharge rates as localized boundary conditions to drive the 2D surface runoff model.

**Stage 5**: Compute Surface Inundation Extent and Depth, incorporate topographic data and configuration files, including partial node backflow sequences, into the two Dimension surface runoff model to determine the spatial extent and depth of surface flooding.

**Hybrid Surface and Subsurface Drainage Network Model**

In comparison to traditional one-dimensional pipe network routing models, two-dimensional surface runoff models excel in simulating flow in unpredictable directions, but neglect subsurface flow dynamics (Zeng et al., 2022). To address this limitation, this study employed a hybrid approach, integrating a one-dimensional pipe network runoff model with a two-dimensional surface runoff model. The pipe network generates net overflow rates at nodes, incorporating underlying surface characteristics. Subsequently, differential parameter thresholds derived from K-means clustering analysis are assigned to each watershed unit. A genetic algorithm is then used to calibrate parameters for sub-watershed units within different urban land use zones.

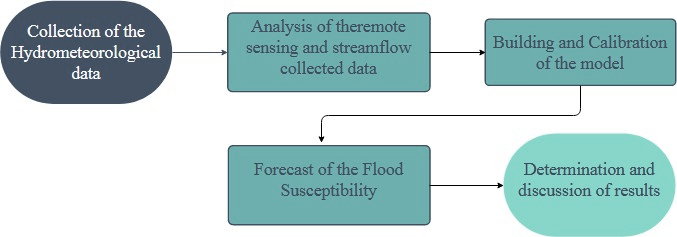


Fig 3-Flowchart of the research methodology

Figure 3, shows the flowchart of the research methodology through which the objectives of this study were achieved.

**Key stages:**

Hydrological Response Unit Delineation: Divide the study area into units based on urban surface characteristics.

Parameterization of Urban Land Use Zones: Assign feature parameter values to different urban land use zones using K-means clustering.

Sensitive Parameter Identification: Develop a mechanism based on Artificial Neural Networks (ANNs) to identify sensitive parameters.

Parameter Calibration: Use a genetic algorithm to calibrate sensitive parameters.

Class III sub-catchments are associated with scattered residential neighbourhoods featuring a blend of asphalted streets, rooftops, and restricted verdant areas. In comparison to commercial and industrial zones, residential areas exhibit increased surface irregularity and heterogeneity, as well as marginally enhanced penetrability. The metrics for surface detention capacity, Manning's roughness coefficient, percolation rate, and dampening coefficient are intermediate.

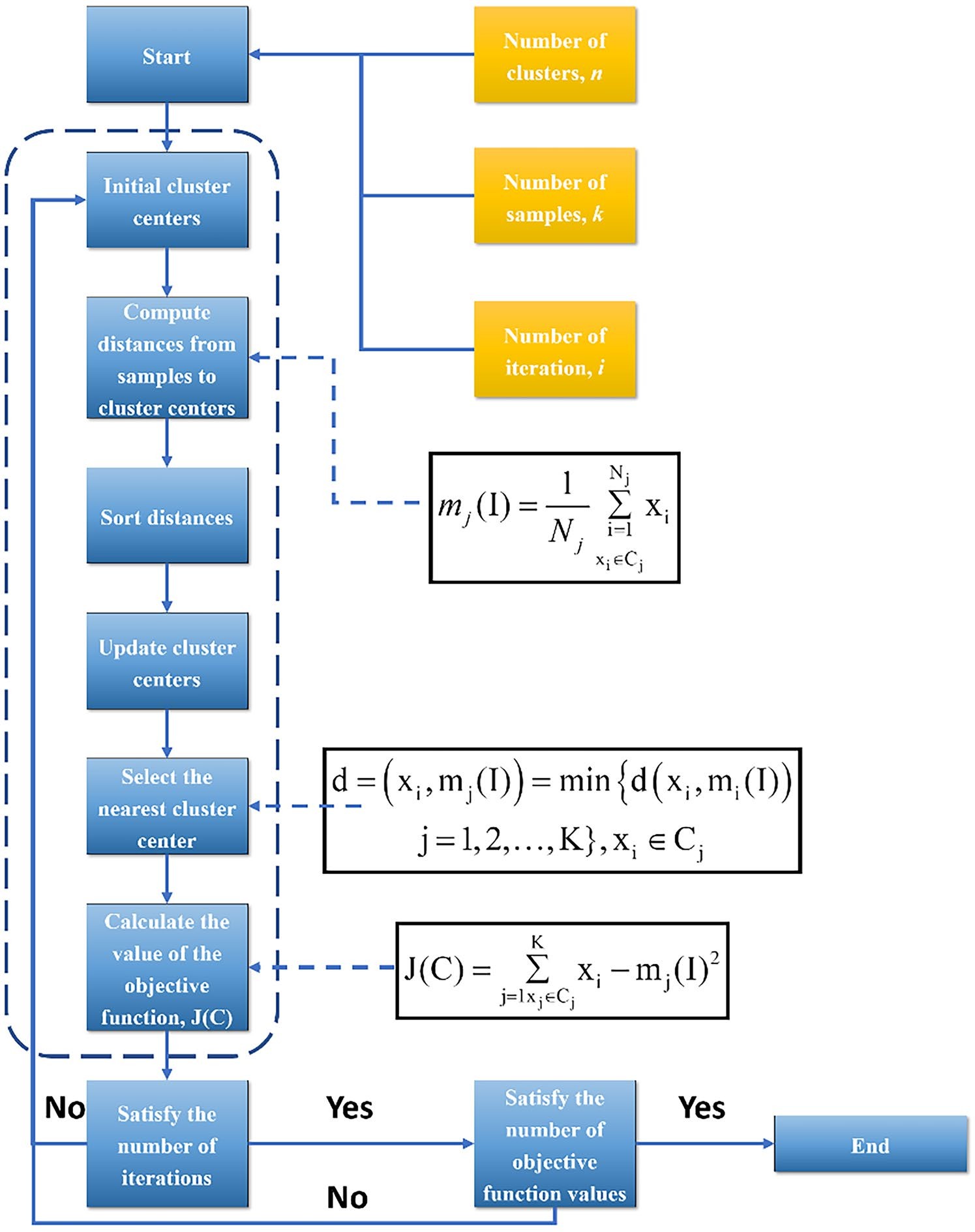
Class IV sub-catchments comprise public land parcels predominantly covered by gardens and green spaces. The land utilization consists mainly of meadows and woodland, characterized by maximal surface irregularity, favourable permeability, and water retention capacity. The metrics for surface detention capacity, Manning's roughness coefficient, percolation rate, and dampening coefficient are maximized.

**Parameter Categorization via K‑Means**

The K-means clustering algorithm is an iterative machine learning technique that segregates sample data into distinct groups based on similarities in their attributes. It achieves this by randomly selecting k initial data points as preliminary cluster nuclei (Liu et al. 2015). In this study, we consulted relevant literature and historical data to compile a preliminary set of parameters encompassing nine uncertain variables: S-Imperv, S-perv, N-Imperv, N-perv, MaxRate, MinRate, degradation, drying period, and texture. Subsequently, the K-means clustering algorithm was utilized to categorize uncertainty parameters across disparate functional land use zones in an urban setting. The cluster count k was established, and preliminary parameter samples were fed into the K-means model for examination. The algorithm yielded characteristic parameter values (cluster centroids) under k disparate clustering scenarios. These values were allocated to transportation, commercial and industrial, residential, and public facility zones, denoting four distinct land use functional zones within the urban setting. A flowchart illustrating the clustering algorithm is presented.

**Sensitive Parameter Identification Based on Artificial Neural Network (ANN) Model**

Artificial neural networks are robust instruments for executing advanced machine learning algorithms and are extensively utilized in predictive modelling and categorization tasks. Generally, ANN can approximate any complex nonlinear relationship with a carefully crafted network architecture, rendering them ideal for tackling intricate systems or opaque models with complicated internal dynamics. Nevertheless, ANN models necessitate extensive data-driven calibration to attain reliability, and the calibration process can be protracted. Once calibrated, however, the model can be swiftly deployed to novel datasets.

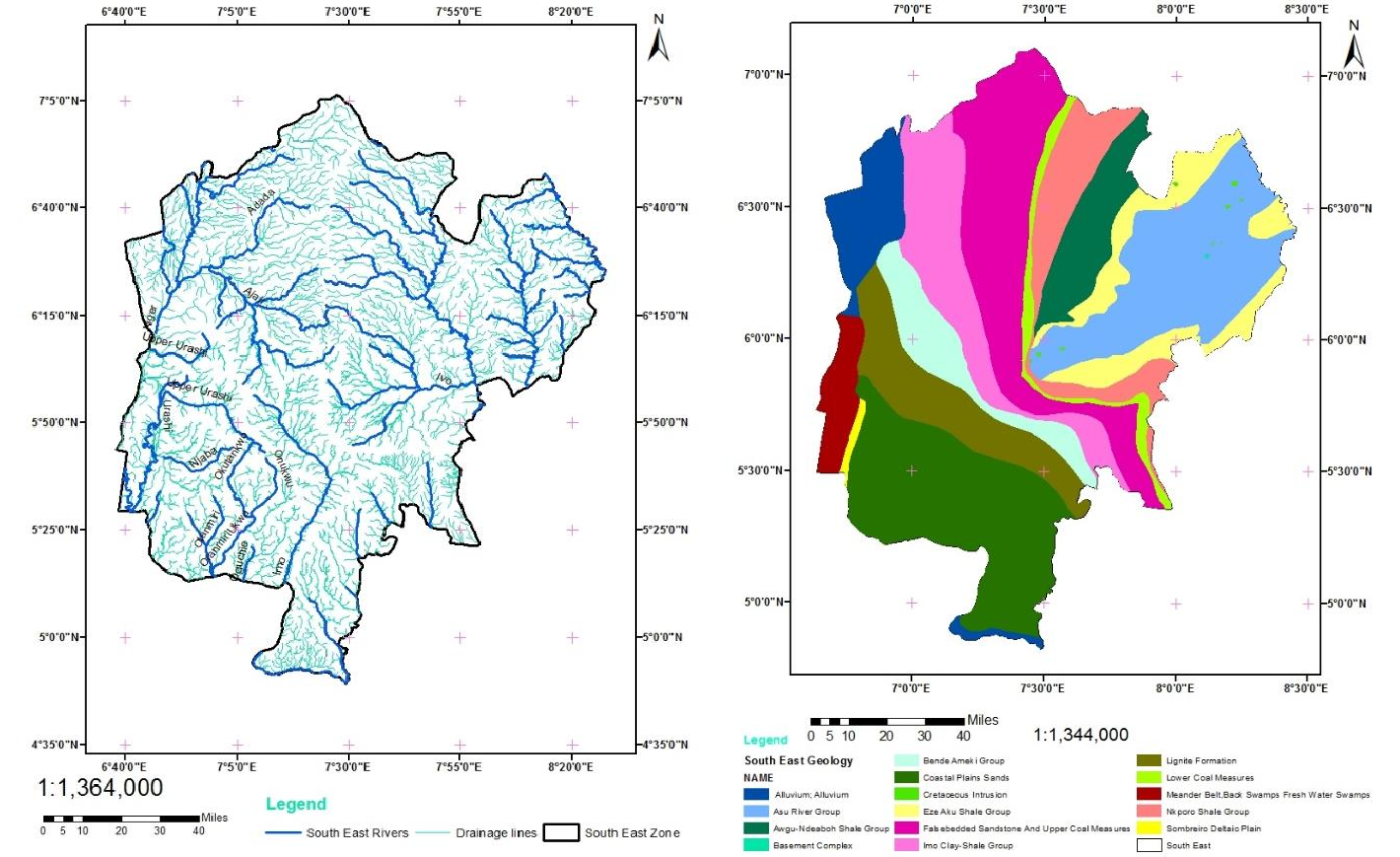


**Figure 4: Flowchart of clustering analysis based on K-means algorithm** Jin, *et al.* (2024)

**The Study area**

**Overview of the Study Area**

This study area is located at (South west Nigeria) which is one of the geopolitical zones in Nigeria and falls into Basin Development Authority. Lagos. It occupies an estimated area of 28658.8 km2 and on a geographic coordinate of Longitude 60 30’-80 30’E and Latitude 50 - 7N. The drainage pipe network within the area 9s relative live independent and consist artificial channels and rivers as storage units. Therefore, this area is an urban watershed with clear boundary condition. Further more the study area has a high level of urbanization, resulting in high surface temperature.



**Figure 5: estimated geopolitical zones**

**Data Collection**

This study collected hourly precipitation data from 1 March 2021, to 12 July 2024, from rainfall stations located within Lagos, Nigeria. rainfall stations are situated in the study area, along with four rainfall stations located on the outskirts. The data from these rainfall stations are useful for determining the timing of heavy rainfall events and optimizing parameters in the urban stormwater flooding model. Surface water depth monitoring data were also collected from three flood-prone sites in the study area. The monitoring information included the time of observation and water depth.

To facilitate the modelling of the urban surface in the study area, geographical data, such as DEM, building data, water system distribution, and land use types, were obtained from the Lagos State Ministry of Physical Planning and Urban Development. All the data were based on the WGS-84 coordinate system, and the UTM zone 32N projection was uniformly adopted for the map projection when using the planar coordinate system.

In addition, the required pipeline network data for this experimental area was provided by the Lagos State Water Corporation, which includes rainwater sewers, inspection wells (rainwater inspection wells and sewage inspection wells), pipelines (rainwater pipelines, sewage pipelines, and a small number of combined rain and sewage pipelines), and drainage outlets (rainwater drainage outlets and sewage drainage outlets).

**Table 1. Location of selected Flood points**

|  |  |  |  |
| --- | --- | --- | --- |
| **Rainfall Location** | **Latitude (Degree)** | **Longitude (Degree)** | **Elevation (m)** |
| Lekki 1 | 7.08 | 6.475 | 32 |
| Lekki 2 | 7.077 | 6.472 | 29 |
| Lekki 3 | 7.069 | 6.474 | 35 |
| Lekki 4 | 7.061 | 6.478 | 33 |
| Sururele 1 | 7.133 | 6.289 | 37 |
| Sururele 2 | 7.129 | 6.293 | 41 |
| Sururele 3 | 7.125 | 6.298 | 40 |
| Sururele 4 | 7.121 | 6.305 | 44 |
| Ajegunle1 | 5.537 | 7.413 | 74 |
| Ajegunle 2 | 5.531 | 7.406 | 53 |
| Ajegunle 3 | 5.529 | 7.4 | 53 |
| Ajegunle 4 | 5.522 | 7.393 | 46 |
| Kosofe 1 | 7.784 | 5.962 | 30 |
| Kosofe 2 | 7.763 | 5.97 | 34 |
| Kosofe 3 | 7.746 | 5.969 | 31 |
| Kosofe 4 | 7.729 | 5.967 | 33 |
| Ikorodu1 | 5.466 | 7.068 | 55 |
| Ikorodu 2 | 5.471 | 7.053 | 55 |
| Ikorodu 3 | 5.475 | 7.043 | 52 |
| Ikorodu 4 | 5.466 | 7.034 | 45 |

**Results and Discussion**

The experiment employed a multi-step approach to investigate the urban inundation dynamics Peng Gao (2021) in the study area. Firstly, the study area was segmented into distinct urban land use functional zones based on hydrological Ifatokun Pual Ifabiyi 2012 response unit partitioning rules. This step enabled the identification of areas with similar hydrological characteristics.

Next, a parameter sensitivity analysis and optimization catillo (2008) were performed using the K-means-ANN-GA machine learning method. This approach allowed for the identification of the most sensitive parameters influencing urban inundation and the optimization of these parameters to improve model performance Haberger (2017)

Finally, the developed urban inundation model Muhammad (2022) was validated through simulations of observed rainfall events. A comparative analysis was then conducted to assess the effectiveness of the proposed methodology in predicting urban inundation dynamics. The results of this analysis are presented in the following sections.

**Table 2. Conditional Probability Distribution: Gaussian Lekki 1**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Rainfall** | | | | | | |
|  | **P-Value** | | **Standard Error** | | **t-statistics** | |
| **Parameter** | |  |  |  |  |  |
|  | **Model 1** | **Model 2** | **Model 1** | **Model 2** | **Model 1** | **Model 2** |
| **AR** | **0.837** | **0.050** | **0.036** | **0.056** | **23.171** | **0.884** |
| **SAR** | **0.078** | **0.070** | **0.037** | **0.038** | **2.110** | **1.846** |
| **MA** | **-0.979** | **0.064** | **0.013** | **0.037** | **-76.813** | **1.731** |
| **Variance** | **5918.340** | **6291.880** | **275.937** | **291.836** | **21.448** | **21.560** |
| **Flood Discharge** | | | | | | |
| **AR** | **0.889** | **0.111** | **0.033** | **0.053** | **27.050** | **2.093** |
| **SAR** | **0.011** | **0.028** | **0.040** | **0.039** | **0.276** | **0.730** |
| **MA** | **-0.977** | **0.125** | **0.016** | **0.038** | **-61.520** | **3.314** |
| **Variance** | **491.824** | **497.478** | **24.074** | **24.783** | **20.430** | **20.073** |

To evaluate the performance of the developed urban flooding model, we simulated three observed rainfall events using uncertain parameter values. We compared the results with two other methods: ANN-GA, which ignores urban hydrological response unit delineation rules, and K-means-DNN, which excludes genetic algorithms. The model's accuracy was assessed using Nash-Sutcliffe efficiency coefficient (NSE), root mean square error (RMSE), and peak time difference (PTD) at monitoring stations S1 and S2. The results of the statistical analysis are presented in Table 3.

**Table 3-Statistical Analysis**

| **Method**  **Evaluation indicators** | | **ANN-GA** | | | **K-Means-DNN** | | | **K-Means-ANN-GA** | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***NSE RMSE* Δt(h)** | | | ***NSE*** | ***RMSE*** | **Δt(h)** | ***NSE*** | ***RMSE*** | **Δt(h)** |
| **1 march 2021** | **S1** | **0.61** | **8.86** | **1** | **0.76** | **7.02** | **0** | **0.91** | **4.22** | **1** |
| **S2** | **0.40** | **10.22** | **1** | **0.44** | **9.83** | **0** | **0.74** | **6.76** | **1** |
| **7August 2021** | **S1** | **0.47** | **4.79** | **1** | **0.56** | **4.34** | **0** | **0.80** | **2.95** | **0** |
| **S2** | **0.52** | **4.44** | **0** | **0.68** | **3.60** | **0** | **0.87** | **2.32** | **0** |
| **5 September 2021** | **S1** | **0.60** | **7.82** | **1** | **0.64** | **7.02** | **1** | **0.73** | **2.18** | **0** |
| **S2** | **0.51** | **8.51** | **1** | **0.65** | **6.94** | **2** | **0.79** | **3.79** | **0** |
| **Average value** |  | **0.52** | **7.44** | **0.83** | **0.62** | **6.46** | **0.5** | **0.81** | **3.70** | **0.33** |

The proposed method demonstrated superior performance, achieving an NSE of 0.73+, RMSE between 2-7, and average PTD of approximately 20 minutes. Compared to ANN-GA, our method improved NSE by 0.29, reduced RMSE by 3.74, and decreased PTD by 0.5 hours. Similarly, it outperformed K-means-DNN with improvements of 0.19 in NSE, 2.76 in RMSE, and 0.17 hours in PTD. These results confirm that our method effectively simulates urban flooding by accurately capturing uncertain parameter distributions and aligning with real-world surface conditions.

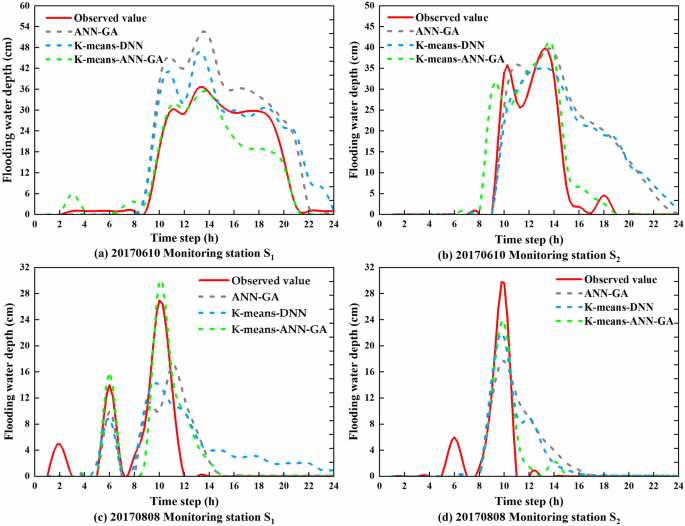


Figure 6: Flood process Result. Jin, *et al.* (2024)

**Conclusion**

We have illustrated through this study the power of integrating machine learning methods with subsurface characterization to improve urban flood simulation and parameter estimation. By integrating one-dimensional pipe network simulation routing and two-dimensional surface runoff modelling with a hybrid K-means–ANN–GA machine learning strategy, we were able to surmount some of the most critical issues one generally has with traditional flood simulations i.e., not taking subsurface into account, ambiguous hydrological zoning, and long parameter calibration.

The use of K-means clustering permitted the identification of the urban hydrological response units from surface characteristics, while the ANN model permitted the identification of the most responsive flood-related parameters. Genetic Algorithms (GA) then gave us an effective method for optimizing the latter parameters to enhance the performance of the flood model. When applied to real-world data from Lagos, Nigeria, which is prone to flooding, the new method outperformed conventional models in terms of accuracy, predictive power, and sensitivity to real rainfall events.

Results indicated that the K-means–ANN–GA method significantly enhanced simulation performance, as shown by enhanced Nash-Sutcliffe Efficiency (NSE), reduced Root Mean Square Error (RMSE), and enhanced peak time predictions (PTD). These enhancements are critically important decision-making considerations for emergency response planning for flood risk management and infrastructure investment.

Overall, this research demonstrates a real-world and scalable urban flood modelling method by integrating hydrologic expertise with artificial intelligence and geospatial analysis. The method not only improves the reliability of the model and its sensitivity to evolving urban conditions but also gives a blueprint to other cities facing similar flood threats. Future research extending this model could involve incorporating real-time sensor inputs, higher spatial resolution for land use classification, and applying deep learning models to arrive at more precise predictions.

Disclaimer (Artificial intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

**Reference:**

Adnan, R. M., et al. (2021). Deep learning for flood forecasting: A review. Journal of Hydrology, 593, 125893.

Anahita Azadgar (2025) Optimizing nature-based solutions for urban flood risk mitigation: A multi-objective genetic algorithm approach in Gdańsk, Poland <https://doi.org/10.1016/j.scitotenv.2024.178303>

Anouk Bomer et al (2019) The influence of grid shape and grid size on hydraulic river modelling performance, Environmental Fluid Mechanics19(449)DOI: [10.1007/s10652-019-09670-4](https://link.springer.com/article/10.1007/s10652-019-09670-4).

Bingyan Ma et al Study on the Classification of Urban Waterlogging Rainstorms and Rainfall Thresholds in Cities Lacking Actual Data

Cai, X., et al. (2019). Urban flood modeling using a coupled 1D-2D model. Journal of Hydroinformatics, 21(3), 432-446.

Chandeashekharappa Agasnalli (2021) Urban Flooding and its concern towards future: A case study

Chang, T. J., et al. (2021). Urban flood modelling using a machine learning approach. Journal of Flood Risk Management, 14(3), e12676.

Chen, Y., et al. (2019). Urban flood modelling using artificial neural networks. Journal of Hydrology, 571, 241-253.

Dao, T. N., et al. (2022). Impact of urbanization on flood risk in a changing climate. Journal of Hydrology, 612, 128134.

February 2024

Feigl, K. L., et al. (2022a). Sensitivity analysis of urban flood models. Journal of Hydro informatics, 24(2), 241-255.

Feigl, K. L., et al. (2022b). Automatic calibration of urban flood models using machine learning. Journal of Hydrology, 612, 128145.

Guo, Y., et al. (2019). Urban flood modelling using a coupled 1D-2D model. Journal of Hydro informatics, 21(3), 447-461.

Hengxu Jin et al. (2024) Using Machine Learning to Identify and Optimize Sensitive Parameters in Urban Flood Model Considering Subsurface CharacteristicsDOI: [10.1007/s13753-024-00540-2](http://dx.doi.org/10.1007/s13753-024-00540-2)

Hu, X., et al. (2022). Rapid flood classification forecasting model using K-means and backpropagation neural networks. Journal of Hydroinformatics, 24(4), 541-554.

Huo, X., et al. (2019). Sensitivity analysis of urban flood models. Journal of Hydroinformatics, 21(2), 241-253.

Javier Fernández Pato (2014) A pipe network simulation model with dynamic transition between free surface and pressurized flow 7(1) DOI: [10.5194/dwesd-7-27-2014](http://dx.doi.org/10.5194/dwesd-7-27-2014)

Jin, H., Zhao, Y., Lu, P. *et al.* (2024) Using Machine Learning to Identify and Optimize Sensitive Parameters in Urban Flood Model Considering Subsurface Characteristics. *Int J Disaster Risk Sci* 15, 116–133 (2024). https://doi.org/10.1007/s13753-024-00540-2

Jung, Y., et al. (2017). Automatic calibration of urban flood models using genetic algorithms. Journal of Hydroinformatics, 19(4), 541-554.

Katipoğlu, M., et al. (2023). Parameter optimization of urban flood models using machine learning. Journal of Hydrology, 615, 128151.

Kumar, P., et al. (2018). Urban flood modeling: A review. Journal of Flood Risk Management, 11(3), 251-265.

Li, Q., et al. (2016). Flood classification and forecasting using K-means clustering and support vector machines. Journal of Hydroinformatics, 18(4), 541-554.

Li, Y., et al. (2022). Urban flood modeling using a machine learning approach. Journal of Hydrology, 612, 128141.

Liao, K. H., et al. (2022). Impact of urbanization on flood risk in a changing climate. Journal of Hydrology, 612, 128136.

Liu, Y., et al. (2020). Urban flood modeling using support vector machines. Journal of Hydroinformatics, 22(2), 241-253.

Lu, G., et al. (2021). Urban flood risk assessment using a coupled 1D-2D model. Journal of Hydroinformatics, 23(3), 432-446.

May-Britt Moser (2015) Place Cells, Grid Cells, and MemoryCold Spring Harbor Perspectives in Medicine5(1):a021808 DOI: [10.1101/cshperspect.a021808](http://dx.doi.org/10.1101/cshperspect.a021808)

Nandi, A. K., et al. (2022). Urban flood modeling using a machine learning approach. Journal of Hydrology, 612, 128142.

Nguyen, P. T., et al. (2020). Urban flood risk assessment using a coupled 1D-2D model. Journal of Hydroinformatics, 22(3), 447-461.

November 2020) [Water](https://www.researchgate.net/journal/Water-2073-4441?_tp=eyJjb250ZXh0Ijp7InBhZ2UiOiJwdWJsaWNhdGlvbiIsInByZXZpb3VzUGFnZSI6bnVsbCwic3ViUGFnZSI6bnVsbH19)12(12):3328 DOI: [10.3390/w12123328](http://dx.doi.org/10.3390/w12123328)

Quing Lui (2021) Integrated Optimization of Pipe Routing and Clamp Layout for Aeroengine Using Improved MOALO July 2021 [International Journal of Aerospace Engineering](https://www.researchgate.net/journal/International-Journal-of-Aerospace-Engineering-1687-5974?_tp=eyJjb250ZXh0Ijp7InBhZ2UiOiJwdWJsaWNhdGlvbiIsInByZXZpb3VzUGFnZSI6bnVsbCwic3ViUGFnZSI6bnVsbH19).2021(10):1-9 DOI: [10.1155/2021/6681322](http://dx.doi.org/10.1155/2021/6681322)

Singh, S. K., et al. (2019). Subsurface characterization for urban flood modeling. Journal of Hydrologic Engineering, 24(10), 04019043.

Sinnathamby, S., et al. (2017). Urban flood modeling using a coupled 1D-2D model. Journal of Hydroinformatics, 19(3), 432-446.

Sun, L., et al. (2022). Parameter calibration of urban flood models using K-means clustering and genetic algorithms. Journal of Hydroinformatics, 24(4), 555-569.

Wang, Y., et al. (2020). Urban flood modeling using a coupled 1D-2D model. Journal of Hydroinformatics, 22(2), 253-265.

Wang, Y., et al. (2022). Automatic calibration of urban flood models using genetic algorithms. Journal of Hydroinformatics, 24(2), 266-278.

Willis, T. K., et al. (2019). Sensitivity analysis of urban

Ifatokun Pual Ifabiyi (2012 )Hydrograph Analysis and Hydrological Zones: Tools for Watershed and Water Resources Planning in a Sub Humid Tropical Catchment, Nigeria Journal of Geography and Geology4(1) DOI: [10.5539/jgg.v4n1p183](http://dx.doi.org/10.5539/jgg.v4n1p183)

Enriquo catillo (2008) Sensitivity analysis in optimization and reliability problemsReliability Engineering & System Safety 93(12):1788-1800 DOI: [10.1016/j.ress.2008.03.010](http://dx.doi.org/10.1016/j.ress.2008.03.010)

Karoly Herberger (2017)Which Performance Parameters Are Best Suited to Assess the Predictive Ability of Models?DOI: [10.1007/978-3-319-56850-8\_3](http://dx.doi.org/10.1007/978-3-319-56850-8_3) .

Muhammad Farooq, Ahmed Haseeb Rabbani (2022) Geospatial Technologies for Flood Hazard assessment: disasterwatch.sgs-suparco.gov.pk