Archives of Environmental Protection
Vol. 51 no. 3 pp. 54–73

PL ISSN 2083-4772
DOI:10.24425/aep.2025.156009

An innovative method of predicting the maximum flow in stormwater sewage systems using soft-sensors

Krzysztof Barbusiński[1], Bartosz Szeląg[2], Anita Białek[2]*, Marek Kalenik[3], Tomáš Bakalár[4]

[1]Department of Water and Wastewater Engineering, Silesian University of Technology, Gliwice, Poland
[2]Faculty of Environmental Engineering, Geomatics and Renewable Energy, Kielce University of Technology, Poland
[3]Institute of Environmental Engineering, Warsaw University of Life Sciences-SGGW, Poland
[4]Institute of Earth Resources, Faculty of Mining, Ecology, Process Control and Geotechnologies, Technical University of Košice, Slovak Republic

*Corresponding author's e-mail: anita_bialek@interia.eu

Keywords: Model uncertainty, soft sensors, maximum flow prediction, machine learning - ML, energy consumption optimization, MARS model

Abstract: Developing universal hydrological models for modeling urban catchments remains one of the major challenges in contemporary hydrology. This study aimed to create a model that integrates catchment characteristics, sewer network topology, sewer storage capacity, and rainfall data, along with a sensitivity analysis of input parameters. The goal was to evaluate the potential of advanced analytical methods, specifically, Multivariate Adaptive Regression Splines (MARS) and soft-sensor technology, to improve peak flow (Q_m) forecasting in stormwater systems. The results showed that combining MARS models with soft sensors yields high forecasting accuracy (R^2 = 0.96, RMSE = 0.038), even under variable rainfall conditions. However, the development of universally applicable model relationships proved challenging due to difficulties in parameterizing the model under changing rainfall scenarios. Additionally, the inclusion of a risk analysis method also enabled consideration of sewer network capacity and introduced a safety margin coefficient to assess system flexibility under future climate conditions. While the proposed approach does not lead to the creation of universal tools, it offers valuable insights for further research on adapting sewer systems to evolving hydrological conditions. The findings suggest promising directions for the development of cost-free, zero-emission soft sensors and models adaptable across diverse urban catchments.

Introduction

Peak flow (Q_m) is an important criterion used as the basis for selecting pipe diameters. To determine Q_m, Monte Carlo Method (MCM) models were commonly used. However, these models make it difficult to identify the factors such as catchment characteristics, topology, network layout, and storage on maximum flow in non-homogeneous catchments. This knowledge is crucial for decision-making regarding the modernization and reconstruction of sewer networks, especially in the context of climate change, urbanization, and sediment deposition in pipes.

Developing an MCM model is a complex task that requires collecting detailed data on land use, sewer network geometry (pipes, manholes), topology, and rainfall - runoff measurements. This involves the installation of stations to monitor flow rates and rainfall (Cristiano et al. 2019). The data collected enables calibration of MCM models (Guo et al. 2021, Fraga et al. 2016). However, these models tend to be over-parameterized, which complicates the identification of coefficients related to depression storage, sewer storage, and similar parameters. In urban catchments, it is particularly important to select an appropriate flow measurement method, choose suitable locations of the measurement point and define the measurement period. These factors influence the operational costs of measurement equipment, in particular, energy consumption, which affects measurement efficiency and accuracy.

Literature data (Bach et al. 2020, De Paola and Ranucci 2012) show that selecting an appropriate flow measurement method and optimizing sensor placement are key to effective stormwater system management (Tabuchi et al. 2020, Zhang et al. 2021). Techniques used to determine flow variation in sewer networks include pressure, acoustic, ultrasonic, electromagnetic, and optical methods (Shi et al. 2021, Luiso et al. 2017, Wong and Kerez 2016). These methods offer high measurement accuracy, provided that device placement complies with manufacturer requirements (Kumar et al. 2021, Wu et al. 2023, Moon et al. 2023). The choice of measurement technique should be based on a detailed analysis

Table 1. Advantages and disadvantages of flow measurement methods in stormwater sewage systems

Measurement Method	Advantages	Disadvantages	Flow Rate	Flow Velocity (v)	Location	Cost and Energy Reduction	Impact of Specific Conditions
Ultrasonic Sensors	- High accuracy (±1-2%) - Non-contact method, minimal interference - Real-time monitoring	- Sensitive to temperature changes and contamination - High purchase and installation costs	Effective across a wide range (0.1 to 10 m^3/s)	Velocities from 0.1 to 5 m/s	Small catchment areas	High initial costs, but long-term savings; low energy consumption	Sensitive to contamination, which may affect accuracy and maintenance costs
Electromagnetic Sensors	- Can measure over a wide range of flows - No moving parts, minimizing breakdowns - Rapid response to flow changes	- Requires calibration - Installation and maintenance costs	High accuracy at high flow rates (up to 25 m^3/s)	Velocities from 0.5 to 10 m/s	Small and large catchment areas	High initial costs, but long-term savings due to low maintenance	Requires adaptation to environmental conditions and sediment, which may affect efficiency
Hydraulic Measurements	- Simple installation, low cost - Low failure rate, easy to maintain - Effective in mountainous areas	- Low accuracy in low flow conditions - Sensitive to water level changes	Suitable for low and medium flows (up to 5 m^3/s)	Velocities from 0.1 to 3 m/s	Good for monitoring small catchments	Low cost, but limited energy-saving potential due to passive measurements	Sensitive to water level changes and sediment, which may affect accuracy and efficiency
Monitoring Cameras	- Allows analysis of water quality and flow - Can detect blockages and irregularities - Provides visual data	- High purchase and maintenance costs - Image quality affected by weather conditions	Effective in various flow conditions	Flow velocity estimation based on video analysis	Small catchment areas	Operational costs can be high, but provide valuable analytical data	Dependent on lighting conditions and contamination, which may limit effectiveness
Pressure Sensors	- Low purchase and installation costs - Enables measurement at various locations - Simple calibration	- Sensitive to contamination and sediment - Limited accuracy in low flow rates	Works well in low flow conditions	Velocities depend on the model and may be hard to determine	Small catchment areas	Low costs, but frequent maintenance and cleaning needed to avoid measurement errors	Frequent maintenance due to sediment can impact long-term costs
Doppler Sensors	- High accuracy in velocity measurements - Suitable for challenging conditions, including polluted waters - Real-time monitoring	- High purchase cost - Requires calibration and adaptation to specific local conditions	High accuracy across various flow rates	Velocities from 0.1 to 6 m/s	Small catchment areas	High initial costs, but can bring long-term savings	Requires adaptation to local conditions, which may increase installation and maintenance costs
Optical Sensors	- Ability to detect contamination and water properties - Rapid response to water state changes - Non-contact measurement method	- Sensitive to light changes and contamination - Purchase and maintenance costs may be high	Effective in diverse flow conditions	Velocities from 0.1 to 4 m/s	For small, polluted catchment areas	Initial costs may be moderate, but maintenance is needed to maintain accuracy	Variable lighting conditions may impact measurements; contamination may limit effectiveness
Acoustic Measurement	- Can measure in challenging conditions, e.g., polluted waters - Low system interference - Can detect blockages	- Highly sensitive to ambient noise - Requires calibration and regular maintenance	Effective for low and medium flows	Velocities from 0.1 to 5 m/s	Small and medium catchments	Low operational costs after initial investment	Impact on accuracy due to contamination and ambient noise

of the catchment's specific conditions, such as its size, land use variability, sewer network topology, and the directions of rainfall movement (Guo et.al. 2021, Rosenzweig et al. 2021). This analysis should be supplemented by an economic evaluation that includes installation and operational costs (Addison-Atkinson et al. 2022, Amiri et al. 2022). Ultrasonic, electromagnetic, and Doppler sensors generally have high initial operating costs. However, this is compensated for by the high measurement accuracy and the potential reduction in operating costs over time. Conversely, optical sensors and cameras may incur higher operating costs but provide valuable data for forecasting wastewater quality, which can ultimately reduce the costs associated with the installation and operation of measurement systems (Gong et al. 2022).

A detailed discussion of the advantages and disadvantages of the currently used flow measurement techniques, considering factors such as catchment size, wastewater quality, site specific conditions, cost, and energy efficiency, is given in Table 1.

Due to the challenges associated with developing mechanistic models and the high costs of direct flow measurements, machine learning (ML) approaches have gained increasing attention (Perdikaki et al., 2022, Morán-Valencia et al., 2023). ML techniques enable the prediction of peak flows based on input–output relationships (Szeląg et al., 2022b), however, many models lack mechanisms to quantify the influence of input variables. Applications to date include sewer flooding prediction (Szeląg et al., 2022a) and hydrograph forecasting (Yang and Chui, 2020, Palmitessa et al., 2022), typically using rainfall and land use characteristics as predictors (Duan et al., 2020). Analytical ML models, particularly those based on modified multivariate regression (e.g., incorporating threshold or smoothing functions), offer a means to evaluate input significance without requiring additional computations (Bhaskar et al., 2018), although their application to peak flow prediction remains limited (Li et al., 2022).

Risk analysis has become integral to assessing sewer system performance under uncertainties associated with climate change and catchment dynamics (Yao et al., 2023). Commonly evaluated metrics include maximum flow rates, flooding volumes, and manhole surcharge occurrences (Mondal et al., 2023). While recent studies have addressed the impacts of rainfall variability over extended periods (Beven, 2020, Napiorkowski et al., 2023), model calibration uncertainties are often overlooked. Accounting for increased runoff resulting from changes in climate and land use is critical for ensuring system resilience (Guan et al., 2015). To address these challenges, Willems et al. (2013) employed conceptual models such as CGM, SGT, and MCM to simulate rainfall impacts and inform system upgrade strategies, while Hauger et al. (2006) proposed a safety-factor-based risk approach to simplify catchment interactions. Adjustments to intensity–duration–frequency (IDF) curves have also been advocated to account for evolving rainfall extremes (Birgani et al., 2013).

Nevertheless, ML models for sewer flow prediction frequently remain oversimplified, often neglecting storm variability, storage processes, and network topology. In risk-based designs (Addison-Atkinson et al., 2022, Amiri et al., 2022), the omission of calibration uncertainties, such as storage capacity variations and roughness coefficients, can lead to underestimation of system vulnerabilities.

A methodology is presented for developing a model to forecast maximum flow based on rainfall data, land use, catchment storage, sewer network topology, and their spatial arrangement within the catchment using machine learning techniques. For this purpose, the MARS (Multivariate Adaptive Regression Spline) model was used. This model provides an analytical relationship that simplifies the identification interactions among the above-mentioned factors and enables straightforward evaluation of input-output relationships without the need for additional calculations. The resulting ML model

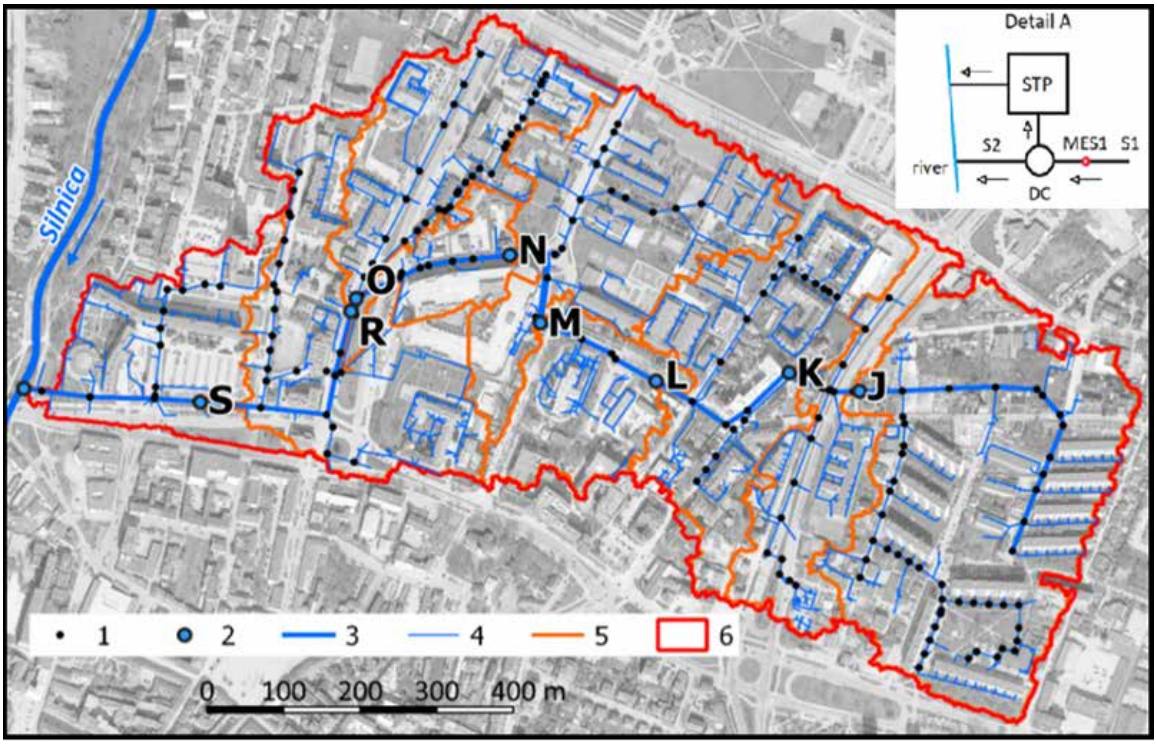

Figure 1. Diagram of analyzed urban catchment in Kielce

can serve as a cost-free soft sensor for predicting maximum flow in stormwater drainage networks. Additionally, a risk analysis was conducted, addressing the issue of catchment modernization in the context of maximum flow reduction under conditions of uncertainty in field and sewer storage.

Methods and data

Study area

The urban catchment area under consideration is located in Kielce, Poland - the capital of the Świętokrzyskie region. It is located in the southeastern part of the city (city center) and includes residential neighborhoods, public buildings, main and secondary streets (Fig.1).

The catchment area covers 63 ha, 40% of which is impervious. Spatial analyses per-formed using GIS tools showed that road density in the area is 108 m·ha-1 (Kiczko et al. 2018). The elevation difference between the highest (271.2 m above sea level) and lowest (260.0 m above sea level) points in the catchment is 11.2 m. The main sewer is 1,569 m long, with diameters ranging from 600 to 1,250 mm. Side sewer diameters range from 300 -to 1,000 mm, and sewer slopes vary from 0.04% to 3.90 %. The sewer system is a separation system. Stormwater from the catchment area flows through a collector to a diversion chamber (DC), from which, at a depth of less than 0.42 m, it flows to the stormwater treatment plant (STP) and then discharges into the Silnica River. During heavy rainfall, when the depth of the separation chamber (DC) exceeds the ordinate of the overflow threshold (OV), excess rainwater is discharged into the discharge canal (S2), which also drains into the Silnica River.

At the outflow from the catchment, an MES1 flow meter was installed 3.0 meters from the inlet of the collector (S1) to the diversion chamber (DC). This device records flow values at 1-minute intervals during heavy rainfall events. Analysis of the MES1 data collected between 2010 and 2020 showed that, during dry period, flows ranging from 1 to 9 $dm^3 \cdot s^{-1}$ were recorded, indicating the presence of infiltration. A rainfall station is located 2.5 km from the catchment boundary, where continuous rainfall measurements have been carried out since 2008, also with a resolution of 1 minute.

Delineation of sub-catchments and characteristics

In the analyzed catchment, 7 sub-catchments (J, K, L, N, M, R, and S) were separated (Fig. 1). This division was based, on one hand, on the alignment of the main collector and, on the other hand, on the need to account for variations in catchment characteristics, the sewer network, and its layout for the purpose of developing a maximum flow simulation model. The boundaries of the sub-catchments were established using spatial data, including land use analyses and sewer network layout (Szeląg et al.2013). These analyses were confirmed by Walek (2019), who separated sewer sub-basins, including side-sewer sub-basins, within the Silnica River catchment. He also estimated the variability of peak flows and their durations using a simplified approach based on the HEC - HMS model, which considered only land use. Based on these considerations, the study catchment was subdivided along the main sewer (Figure 1). The selection of model input characteristics was supported by a literature review (Szeląg et al., 2016), which examined the characteristics of the catchment and the sewer network parameters commonly used in models simulating stormwater system performance (e.g., maximum flow, flooding volume, manhole overflow). Table 2 presents detailed characteristics of land development and the sewer network for each of the separated sub-catchments.

Methodology

An algorithm was developed to created models for analyzing the influence of spatial catchment characteristics, sewer network topology and its layout (using the fractal dimension), and runoff-related parameters on peak flow in response to rainfall (Figure 2). The parameters considered include storage depth, the Manning roughness coefficient for impervious areas, a correction factor for mean slope and impervious area, runoff path width, and the Manning roughness coefficient for sewers.

Machine learning, specifically, the MARS model, was used to calculate the peak flow as an alternative to mechanistic models such as SWMM (Storm Water Management Model). The analytical form of the MARS model allows for the analysis

Table 2. Characteristics of the delineated sub-catchments

No.	F	Imp	Vk	Gk	R.t.	Vkp	Jkp	Impd	Gkd	Vrd·Vkd^{-1}	FD
	ha	-	m^3	$m \cdot ha^{-1}$	m	m^3	-	-	$m \cdot ha^{-1}$	-	-
min	12.66	0.35	157	0.0079	1.74	16.1	0.0036	0.40	0.0011	0.17	0.94
max	55.41	0.44	1240	0.0092	8.47	67.5	0.0102	0.55	0.0063	0.84	1.14

where: F - area of the catchment area; Imp - impervious of the catchment area; Vk - volume of the collector; Gk - length of the collector per impervious area in the catchment area; R.t. - difference in ordinates of thesewer; Vkp - volume of the sewer preceding the catchment closure section; dHp - difference in ordinates to the catchment closure section; Jkp - slope of the bottom of the sewer preceding the catchment closure section; Hst - depth of the manhole in the closure section; Imp - impervious of the catchment area below the closure section; Gkd - length of the collector per impervious area below the closing cross-section; Vrd - storage of the catchment area above the closing cross-section defined as Vrd = F - (Imp - d_{imp} + (1 - Imp) - d_{per}), Vkd - volume of the sewer network in the lower stream, FD - fractal dimension (calculations according to Section - Calculation of fractal dimensions).

of the influence and interaction between selected variables within their range of variability. The proposed model also supports risk analysis (Butler et al. 2014, Ursino et al. 2015), enabling the evaluation of sewer system performance under uncertainty (storage of catchment, sewerage network, rainfall due to climate change). This approach facilitates improvements in system operation by incorporating the probabilistic nature of SWMM model parameters.

Calculation of fractal dimensions

To estimate the fractal dimension, the box method was used. In this approach, the image (in this study, the route of the sewer network – Fig. A1), was placed on a grid of square boxes, each with side lengths of τ. The number of boxes $N(\tau)$ covering the image (i.e., the canal network) was then counted. This number depends on the size of the grid elements and decreases as τ is reduced in subsequent iterations. The box dimension was calculated by analyzing how $N(\tau)$ varies with τ. For complex sewer grids, the number of elements in successive iterations is not constant and the box dimension was defined as the boundary value where the length of the side of the box (τ) in the grid tends to zero. For the assumed number of boxes $N(\tau)$ with side length τ covering the image (the sewer network in the separated sub-catchment), the box dimension was calculated from the formula:

$$FD = lim_{\tau \to 0} \frac{log(N(\tau))}{log\left(\frac{1}{\tau}\right)} \quad (1)$$

Practically, the box dimension is determined by analyzing the relationship $\log(N(\tau) = f(\log(\tau\text{-}1))$, which is then approximated using a linear function. Based on the sewer network (Fig. A1) and the time distributions of rainfall events (Fig. A2), the fractal dimensions were calculated.

Mechanistic model

The mechanistic model of the catchment includes 92 sub-catchments, with areas ranging from 0.12 to 2.10 ha and imperviousness varying between 5 and 95%. The model includes 82 manholes and 72 pipes. The storage depths for impervious (D_{imp}) and pervious (D_{per}) areas were set to 2.50 mm and 6.00 mm, respectively. Manning's roughness coefficients were n_{imp} = 0.025 $m^{-1/3}\cdot s$ for impervious areas and n_{per} = 0.10 $m^{-1/3}\cdot s$ for pervious (n_{per}) areas, as determined during the model calibration step (Szeląg et al. 2016). The width of the runoff path (W) was calculated as $W = \theta \cdot A^{0.50}$, where θ is the flow path correction coefficient and A is the catchment area. The highest agreement between calculation and measurements was obtained for θ = 1.35. Model calibration using 6 rainfall-runoff events resulted in a Nash - Sutcliff efficiency coefficient ranging from 0.85 to 0.98, and a coefficient of determination between 0.85 and 0.99. The differences in hydrograph volumes and peak flows did not exceed 5% compared to measurements, confirming a high level of agreement between calculated and measured values (Szeląg et al. 2016).

Calculation of uncertainty by GLUE method (Generalized Likelihood Uncertainty Estimation)

In the present analyses, for the assumed uniform distributions of SWMM parameters (correction coefficient of impervious area, correction coefficient of flow path, Manning co-efficient of impervious and pervious area, Manning coefficient of sewer storage, storage depth of impervious and pervious area) (Tab. A1) and measured rainfall - runoff events (calibration: 24th July 2011 and 15th September; validation: 30th May 2010 and 30th July 2010), a posteriori distributions of SWMM parameters, 95% confidence interval, and likelihood function values were calculated. For the calibration data (30th May

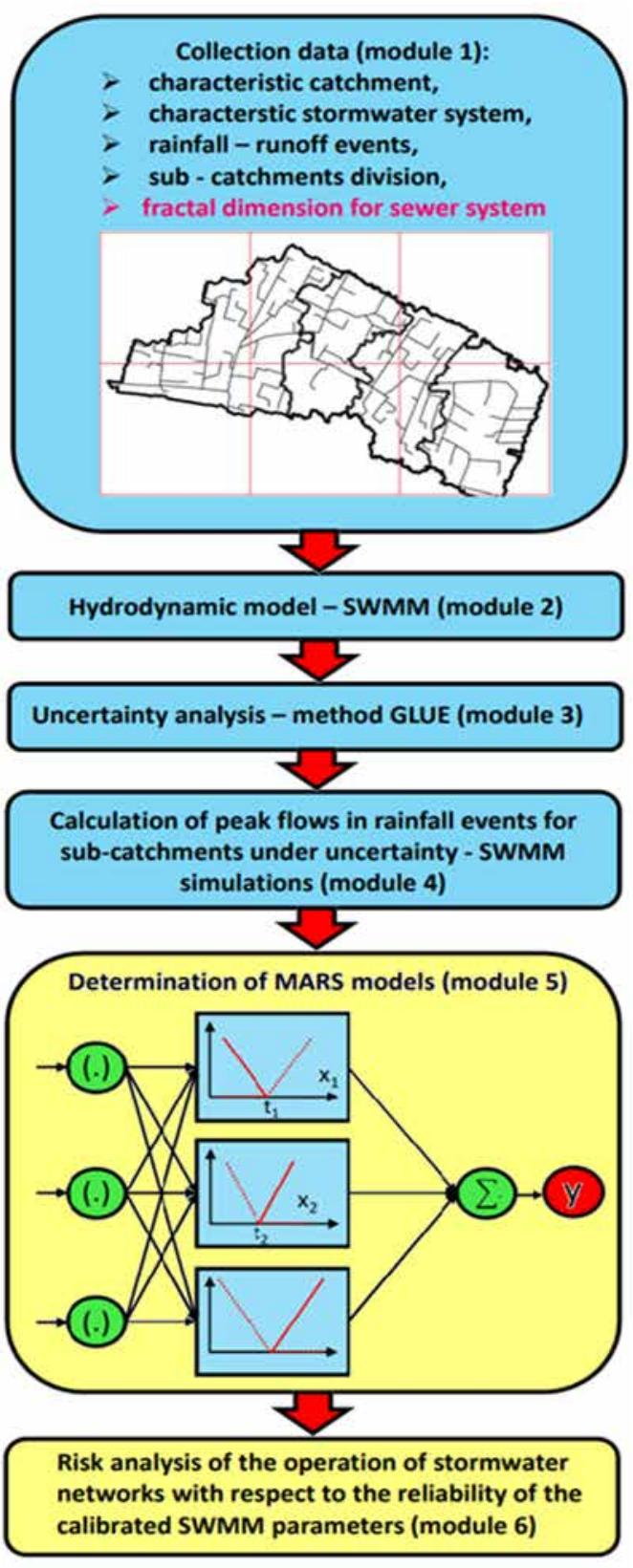

Figure 2. Scheme for the development of model to calculate peak flows in a sewer network and risk analysis

2010 and 8th July 2011), up to 96% of the measurement data were found to be within the calculated confidence intervals. For the validation set (15th September 2010), 90% of the observations were within the confidence interval, while for the event on 30th of July 2010, 70% of the observations covered the determined 95% confidence interval. A detailed discussion of the computational methodology used in the uncertainty analysis is provided in Szeląg et al. (2022 a).

Simulations of sewer network operation

Using the developed hydrodynamic model of the catchment, peak flow calculations were performed at the cross-sections closing sub-catchments J, L, N, R, and S under uncertainty associated with the calibrated SWMM parameters (the correction coefficient percentage area, storage depth of impervious and pervious areas, Manning coefficient for impervious and pervious areas, the correction coefficient for slope, and the Manning coefficient for sewer storage). Using the posteriori distributions of the SWMM parameters, peak flow simulations were performed for assumed rainfall data at the sub-catchment closure cross sections. The following rainfall events were assumed for the analyses: t_r = 30 min (P_t = 6.0 mm and 8.0 mm) and tr = 60 min (P_t = 6.0 mm) for temporal rainfall distribution types R2, R3, R4 (according to DWA A – 118, 2006). These data describe different rainfall loads on the sewer network (potential consequences of climate change) for the catchment in Kielce, and help to explain runoff variability and the occurrence of various flow conditions in the sewers (Szeląg et al. 2016). Rainfall with t_r = 60 min generates gravity flow conditions throughout the sewer network. In contrast, for tr = 30 min and Pt = 6.0 mm, storm overflow (OV) discharge occurs, and for P_t = 8.0 mm, pressure flows and hydraulic backflows are observed in the sewer system (Szeląg et al. 2016).

Development of MARS model

In the present analyses, the MARS method was used to describe the peak flow in the sewers closing the separated sub-catchments. The model can be expressed in general form as:

$$\boldsymbol{Q_m^* = \alpha_0 + \sum_{j=1}^{T} \cdot h(x_j^*, t_j^*)} \tag{2}$$

where: α_0, α_j are empirical coefficients estimated by recursive partitioning of the feature space (Friedman and Roosen 1995); T is the number of basis functions h(x_j*, t_j*), assuming that a linear basis function is described by the equation:

$$\boldsymbol{h(x_j^*, t_j^*) = \begin{cases} \alpha_j \cdot (x_j^* - t_j) \; for \; x_j^* > t_j^* \\ 0 \; for \; x_j^* \le t_j^* \end{cases}} \tag{3}$$

where: t_j* are threshold values for the T basis functions (Fig. A3); x_i*, Q*$_m$ denote the standardized independent variables (e.g., rainfall data, catchment characteristics, sewer network, fractal dimension, SWMM parameters) and the standardized maximum flow (Q_m).

In equation (3), the basis function takes the form of a linear relationship that incorporates threshold values (t_j). In this context, and drawing on the principles of linear regression and the Standard Regression Coefficient (SRC) sensitivity analysis method, the obtained values of α_j for individual independent variables, in their variation intervals, are interpreted as indicators of model sensitivity (Wang et al., 2019). For two threshold values ($t_1^* < t_2^*$) determined with respect to a single independent variable (x_j^*), the MARS function takes the form $Q^*_m = \alpha_j\text{-}x_j^* + ... + \alpha_0$ ($x_j^* > t_1^*$) i $y = \alpha_2\text{-}x_{j*} + ... + \alpha_0$ ($x_j^* > t_2^*$). In this case, the combined effect of α in the interval (t_1^*; t_2^*) is equal to $\alpha_1+\alpha_2$. To determine the number of nodes (T), threshold values (t_j^*), and coefficients α_j, the STATISTICA 10 software was used. This program automatically identifies statistically significant independent variables.

Risk analysis including reliability of SWMM parameters

To analyze the performance of the sewer network, the design peak flow was used, taking into account the allowable error in its identification (σ), which is described by the following equation (4):

$$\begin{aligned} \boldsymbol{Q_m^p} &= \boldsymbol{Q_m \cdot (1 \pm \sigma)} \\ \boldsymbol{Q_{m(L)}^p} &= \boldsymbol{Q_m \cdot (1 - \sigma)} \\ \boldsymbol{Q_{m(U)}^p} &= \boldsymbol{Q_m \cdot (1 + \sigma)} \\ \boldsymbol{Q_m^p} &= \boldsymbol{< Q_{m(L)}^p ; Q_{m(U)}^p >} \end{aligned} \tag{4}$$

where: σ represents the allowable uncertainty threshold of the maximum flow Q_m. Values of σ = 0.10, 0.15, and 0.20 were considered in the calculations, and can be interpreted as safety factors that influence the reserve capacity of the designed sewer system, e.g. due to climatic changes in rainfall.

The performance of the drainage system under uncertainty (e.g., variability in temporal rainfall distribution and SWMM model parameters) was assessed based on the probability of occurrence of the peak design flow:

$$\boldsymbol{hp_{Qm} = \frac{\sum_{i=1}^{k} Z_k}{N}} \tag{5}$$

where: Z_k represents the function describing the exceeding of the maximum flow Q^p_m:

$$\boldsymbol{Z_k = \begin{cases} Z_k = 1 \rightarrow Q_{m(U)}^p \ge Q_m(X_R, X_{CTCH}, x_1, x_2, \dots, x_p) \ge Q_{m(L)}^p \\ Z_k = 0 \rightarrow Q_m(X_R, X_{CTCH}, x_1, x_2, \dots, x_{p=P}) \ge Q_{m(U)}^p \\ Z_k = 0 \rightarrow Q_m(X_R, X_{CTCH}, x_1, x_2, \dots, x_{p=P}) \le Q_{m(L)}^p \end{cases}} \tag{6}$$

where: N – number of Monte Carlo (MC) samples; Q_m (x_1, x_2, x_3, ..., x_j) – value of the peak flow for the sub-catchment for the assumed characteristics of the catchment, the sewer network, and SWMM model parameters; X_R – vector of rainfall data combinations (e.g., rainfall depth, duration, and temporal distribution of the event), X_{CTCH} – vector of combinations of catchment and sewer network characteristics, p = 1, 2, 3, …, P.

The risk analysis proposed in the present study included the following computational steps:

- identification of a posteriori (GLUE) distributions of SWMM parameters (N = 5000 samples),
- simulation of the maximum flow in separated sub-catchments for rainfall events, taking into account uncertainties (Section A1),
- calculation of the probability of maximum design flow (pQm) from equation (5) in rainfall events with σ = 0.10, 0.15, and 0.20,

- determination of SWMM parameters - n_{imp}, d_{imp}, β, n_{sew} (neglecting the likelihood function) based on the a posteriori distributions:

$$f(x_p) = \{x_{p,1}, x_{p,2}, \dots, x_{p,s}\} \rightarrow x_{p,s} = \\ = F^{-1}\left(< Q^p_{m(L)}; Q^p_{m(U)} >, X_R, X_{CTCH}, x^*_{p-1}, \dots, x^*_P\right) \\ for\ p = 1, 2, 3, .., P \\ s = 1, 2, 3 \dots, N < 5000 \quad (7)$$

where: $F^{-1}\left(< Q^p_{m(L)}; Q^p_{m(U)} >, X_R, X_{CTCH}, x^*_{p-1}, \dots, x^*_P\right)$ inverse function $Q_m \pm \sigma \cdot Q_m$ to identify SWMM parameter distributions based on p_{Qm}; s - number of p-values of these SWMM parameters; f() – empirical distributions of p = 1, 2, 3, ..., P - SWMM parameters; x_p,s - s - p-values of this SWMM parameter assuming $Q^p_{m(U)} \geq Q^p_m \geq Q^p_{m(L)}$; x^*_{p-1}, …., x^*_p - p - 1, ..., P values of SWMM parameters obtained from uncertainty calculations (a posteriori distribution). The final result obtained was the vectors [x_p, ..., x^*_{p-1}, x^*_p].

Determination of 0.50 percentiles of SWMM parameters and preparation of the curves: $n_{imp} = f(Q_m \pm \sigma \cdot Q_m)$, $d_{imp} = f(Q_m \pm \sigma \cdot Q_m)$, $\beta = f(Q_m \pm \sigma \cdot Q_m)$, $n_{sew} = f(Q_m \pm \sigma \cdot Q_m)$ for σ = 0.10, 0.15, 0.20

In this study, the distributions of SWMM parameters (n_{imp}, d_{imp}, β, n_{sew}) were determined for two variants. In the first, the likelihood function was neglected, while in the second variant, it was included. For the determined SWMM parameters (x_p) that satisfy the condition described by relation (Section A1, equation 2A), the corresponding value of L(Q/θ) was obtained. On this basis, the 0.50 percentiles of the likelihood function were identified, and the corresponding vectors of SWMM parameters were determined, as follows: [n_{imp}, …., x_p^*], [d_{imp}, …., x_p^*], [β, …, x_p^*] i [n_{sew} …., x_p^*].

Results

Influence of uncertainty on the results of peak flow calculations

Based on the results of runoff simulations in the separated sub-catchments (J, L, N, R, S), considering the uncertainty in SWMM parameters for the assumed rainfall events (t_r = 30, 60 min), the variability of peak flow was assessed using percentiles (25%, 50%, 75%) (Fig. A5). Example simulation results of Q_m for the considered sub-catchments, with t_r = 30 min and P_t = 6.0 mm (rainfall type R2), are presented in Fig. 4. The highest value of Q_m = 0.75 $m^3 \cdot s^{-1}$ (percentile 0.50) was obtained for sub-catchment S, while the lowest was observed in the sub-catchment J, with Q_m = 0. 25 $m^3 \cdot s^{-1}$. The uncertainty in the SWMM parameters was found to have a significant impact on the variability of the peak flow (Q_m), as confirmed by the percentile values (25%, 50%, 75%) for the sub-catchments (Fig. 3).

For the sub-catchment S, the Q_m value varies between 0.48 and 1.82 $m^3 \cdot s^{-1}$, and for the sub-catchment J, it is in the range of 0.09 - 0.49 $m^3 \cdot s^{-1}$.

Development of MARS models for peak flow calculations

Fig. 4 shows the relationship between R, RMSE, and MAPE values for the obtained MARS models; Table A2 gives the results for the learning set (80%) and test set (20%).

The best fit between calculations and analysis was obtained for t_r = 30 min, P_t = 8.0 mm (type R3) obtaining R^2 = 0.96, MAE = 0.021, RMSE = 0.038, while the largest simulation errors Q_m^* occurred for rainfall type R2 (R^2 = 0.92, MAE = 0.034, RMSE = 0.058). Models developed for t_r = 60 min (rainfall types R2, R3, R4) exhibited worse predictive performance (R^2 = 0.90 - 0.92, MAE = 0.027 - 0.031, RMSE = 0. 050 - 0.059) compared to models determined for t_r = 30 min and P_t = 6.0 mm, P_t = 8.0 mm (type R2, R3, R4), which achieved R^2 = 0.92 - 0.96, MAE = 0.021 - 0.034, and RMSE = 0.038 - 0.053.

Identification of the impact of independent variables on maximum flow based on rainfall data

MARS model

Using the results of SWMM calculations in the sub-catchments (J, K, L, N, P, R) including associated uncertainties, MARS models were developed based on the assumed rainfall data (rainfall depth, duration, and temporal distribution of

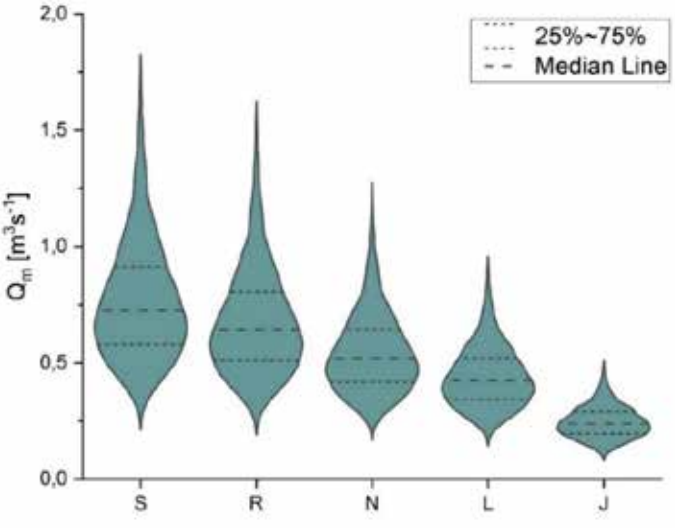

Figure 3. Influence of SWMM parameter uncertainty on peak flow (Qm) in sub-catchments (J, L, N, R, S) for rainfall type R2 and tr = 30 mm and Pt = 6.00 mm

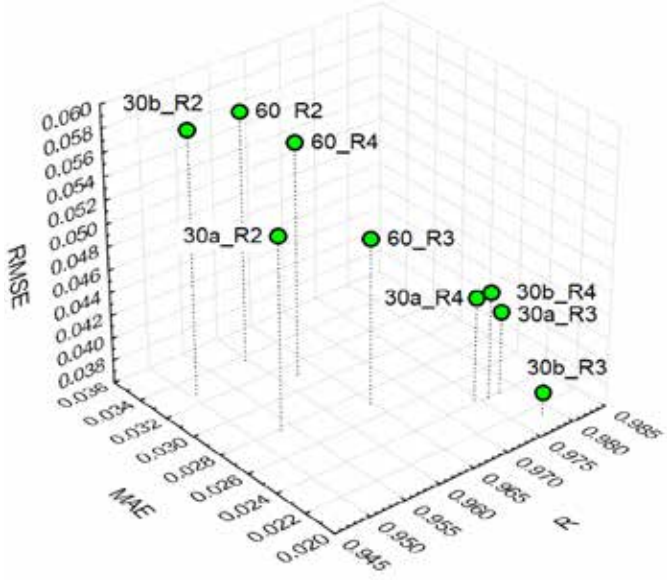

Figure 4. Relationship between RMSE, MAE, R for determined MARS models in relation to assumed rainfall data

rainfall in the event). The calculated threshold values (t) and empirical coefficients (αi) for MARS models are presented in Tables A3, A4 and A5 in Appendix A. An example model for rainfall 30a_R (t_r = 30 min, P_t = 8.0 mm, rainfall type R2) is described by the following equation:

$$Q_m^* = 0.146 + 0.198*\max(0; F_{imp}^*-0.00) - 0.196*\max(0; n_{sew}^*-0.451) + 0.276*\max(0; 0.451-n_{sew}^*) - 0.160*\max(0; d_{imp}^*-0.154) - 0.167*\max(0; 0.154-d_{imp}^*) - 0.01*\max(0; n_{imp}^*-0.329) + 0.173*\max(0; 0.329-n_{imp}^*) + 0.095*\max(0; \alpha-0.38) - 0.151*\max(0; 0.380-\alpha) + 0.07*\max(0; \beta-0.722) - 0.082*\max(0; 0.722-\beta) + 0.07*\max(0; FD-0.00) - 0.024*\max(0; 0.75-\gamma) + 0.071*\max(0; n_{sew}^*-0.225) + 0.068*\max(0; n_{imp}^*-0.863)$$

Across the full range of variation in F_{imp}* and FD* values, a uniform influence of these independent variables on Q_m* is observed. In contrast, for n_{sew}*, n_{imp}*, β*, α*, γ* within the range x_i^* = <0; 1>, their influence on Q_m* varies, as confirmed by the α_i coefficients values (Figure A5). For $n_{sew}^* \leq 0.225$, an increase in n_{sew}* led to an increase in Q_m* of 0.0276 per $0.1 \cdot n_{sew}^*$. For $0.451 > n_{sew}^* > 0.225$, an increase in Q_m* of 0.035 per $0.1 \cdot n_{sew}^*$ takes place. However, when $n_{sew} > 0.451$, a decrease in Q_m* of 0.0196 per $0.1 \cdot n_{sew}$ is observed. Table 3 presents the αi coefficients in the MARS models for the assumed rainfall conditions (type R2, R3, and R4 for scenarios 30a, 30b, 60).

Influence of rainfall data on sensitivity (catchment characteristics, drainage network)

F_{imp}* had the highest influence on Q_m* under the P_t = 6.0 mm, type R3 temporal rainfall distribution, and also for P_t = 8.0 mm type R2 rainfall. In contrast, Vkp exhibited the lowest influence (Table 3). An increase in from P_t = 6.0 mm to P_t = 8.0 mm for type R2 rainfall (e.g., the influence of climate change) led to a greater influence of F_{imp} on Q_m* (Fig. 5a). For rainfall types R3 and R4, the opposite trend was observed. For the same types (R3 and R4), an increase in P_t led to greater sensitivity of Q_m* to changes in FD (Fig. 5b). The highest influence of FD* on Q_m* was observed under type R3 rainfall conditions, while the lowest influence occurred under type R2, regardless of whether P_t was 6.0 mm or 8.0 mm.

Influence of rainfall data on sensitivity (SWMM parameters)

For rainfall types R2, R3, and R4, an increase in P_t from 6.0 mm to 8.0 mm resulted in an increase in the sensitivity of Q_m* by 2.0 - 27.7% for α, β, n_{imp} and by 12.5 - 80.0% for n_{sew}. In contrast, an opposite trend was observed for d_{imp}. Under t_r = 30 min and P_t = 6.0 mm (types R3 and R4), d_{imp} had the greatest influence on Q_m*, while n_{imp} had the least. For t_r = 30 min and P_t = 8.0 mm (types R2 and R4) n_{sew} showed the highest sensitivity of maximum flow to changes, with α_i values ranging from 0.18 to 0.26; under rainfall type R3, d_{imp} remained the dominant influencing factor. For t_r = 60 min and rainfall types R2, R3, and R4, d_{imp} had the highest sensitivity in the Q_m* model.

The impact of precipitation characteristics on estimated coefficients in the MARS model

Table 4 presents the results of correlation coefficient (R_{ij}) calculations between the MARS model coefficients and rainfall characteristics. Very high correlations (R_{ij} = 0.7- 0.9) were shown for Gk and P_t = 5 (R_{ij} = 0.91), α(FD) and $P_{t=10/Pt}$ (R_{ij} = 0.79) and α(α) and $P_{t=5}/P_{t=15}$ (R_{ij} = 0.72). High correlations (R_{ij} = 0.5 - 0.7) was found between $\alpha(F_{imp})$ and $P_{t=5,10}/P_{0.5tr}$, with the highest R value observed for $P_{t=10}/Pt$ (Table 4). Similarly, high correlations were obtained for α(β) and $P_{t=5}/P_{sr}$ (R_{ij} = 0.63), as well as for $\alpha(n_{imp})$ with both $P_{t=5}/P_{sr}$, and $P_{t=10}/P_{sr}$. Based on the above results, the empirical relationships were derived: α(FD) = $f(P_{t=10}/P_t)$ and α(Gk) = $f(P_{t=5})$. Both models yielded R^2 = 0.50, suggesting a nonlinear relationship between the variables (Fig A4).

Probability of maximum design flow – rainfall characteristics – catchment characteristics – SWMM model parameters

Based on sub-catchment flow simulations, the probability of exceeding the maximum design flow (p_{Qm}) was determined for the assumed rainfall data. The simulated maximum flows (Q_m) ranged from 25 to 2000 $L \cdot s^{-1}$, depending on catchment characteristics such as F_{imp} and FD (Figures 5a, 5b). On this basis, empirical cumulative distribution functions (CDFs)

Table 3. The absolute values of the coefficients in the MARS model depending on the distribution of rainfall (* - the table shows the coefficients α_i including the maximum range of variation x_i).

	F_{imp}*	n_{sew}*	d_{imp}*	β*	n_{imp}	α*	γ*	Gk*	FD*	Vkp*
30a_R2	0.20	0.20	0.16	0.09	0.10	0.09	0.02	0.00	0.07	0.00
30a_R3	0.20	0.08	0.20	0.11	0.07	0.11	0.02	0.03	0.17	0.00
30a_R4	0.26	0.10	0.20	0.11	0.07	0.11	0.02	0.00	0.11	0.00
30b_R2	0.43	0.26	0.10	0.10	0.11	0.10	0.02	0.09	0.00	0.01
30b_R3	0.16	0.09	0.12	0.11	0.07	0.11	0.02	0.04	0.18	0.00
30b_R4	0.19	0.18	0.14	0.14	0.07	0.14	0.02	0.03	0.18	0.00
60_R2	0.20	0.22	0.26	0.10	0.09	0.10	0.00	0.03	0.09	0.00
60_R3	0.24	0.01	0.35	0.01	0.06	0.01	0.00	0.00	0.01	0.00
60_R4	0.31	0.01	0.37	0.10	0.01	0.10	0.00	0.00	0.00	0.00

Table 4. Correlation coefficient (Rij) between estimated coefficients in MARS models and rainfall characteristics

	$\alpha(F_{imp})$	α(FD)	α(Gk)	$\alpha(n_{sew})$	$\alpha(d_{imp})$	α(β)	$\alpha(n_{imp})$	α(α)	α(γ)
Pt	0.37	0.48	0.88	0.42	0.00	0.05	0.42	0.63	0.16
q	0.37	0.48	0.88	0.42	0.00	0.05	0.42	0.63	0.16
$P_{t=5}$	0.45	0.59	0.91	0.38	0.08	0.10	0.17	0.63	0.20
FD_r	0.37	0.44	0.73	0.65	0.12	0.40	0.33	0.40	0.25
$P_{t=5}/P_{t=10}$	0.48	0.71	0.71	0.28	0.15	0.38	0.03	0.57	0.07
$P_{t=5}/P_{t=15}$	0.27	0.61	0.52	0.03	0.27	0.18	0.07	0.72	0.10
$P_{t=10}/P_{t=15}$	0.48	0.71	0.71	0.28	0.15	0.38	0.03	0.57	0.07
$P_{t=5}/P_t$	0.40	0.68	0.64	0.30	0.20	0.48	0.00	0.58	0.17
$P_{t=10}/P_t$	0.63	0.79	0.64	0.28	0.15	0.33	0.02	0.42	0.02
$P_{t=15}/P_t$	0.33	0.40	0.80	0.33	0.03	0.29	0.44	0.61	0.14
P_{sr}	0.45	0.71	0.60	0.15	0.08	0.55	0.18	0.52	0.20
$P_{t=5}/P_{sr}$	0.17	0.33	0.01	0.27	0.10	0.63	0.62	0.00	0.50
$P_{t=10}/P_{sr}$	0.03	0.09	0.28	0.40	0.13	0.48	0.62	0.27	0.53
$P_{t=15}/P_{sr}$	0.15	0.06	0.29	0.00	0.08	0.47	0.52	0.50	0.10
$P_{0.5tr}$	0.40	0.37	0.42	0.52	0.02	0.33	0.00	0.03	0.38
$P_{sr}/P_{0.5tr}$	0.43	0.59	0.55	0.18	0.08	0.35	0.28	0.35	0.02
$P_{t=5}/P_{0.5tr}$	0.58	0.69	0.41	0.10	0.10	0.30	0.20	0.18	0.05
$P_{t=10}/P_{0.5tr}$	0.59	0.49	0.50	0.46	0.07	0.19	0.03	0.02	0.14
$P_{t=15}/P_{0.5tr}$	0.27	0.34	0.52	0.49	0.02	0.41	0.02	0.23	0.32

where: FD_r - fractal dimension of the temporal distribution of rainfall in the event; $P_{t=5}$, $P_{t=10}$, $P_{t=15}$ - maximum 5-, 10-, 15-minute rainfall depth in the episode; P_t - rainfall depth in the episode; P_{sr} - average 5-minute rainfall depth in the episode, $P_{0.5\cdot tr}$ - rainfall depth for $t = 0.5 \cdot t_r$.

of SWMM parameters (n_{imp}, β, d_{imp}, n_{sew}), which condition the non-exceedance of Q_m values, were developed. The 50th percentile values were extracted from the CDFs, and curves of $d_{imp} = f(Q_m, F_{imp}, FD)$ were prepared for rainfall scenarios 30a, 30b, 60 of types R2, R3, and R4. Example curves for dimp, identified as the SWMM parameter with the greater influence on Q_m*, are shown in Figures 5c and 5d. The computed values of other SWMM parameters (n_{imp}, β and n_{sew}) are provided in Table A3.

A non-linear relationship between Q_m and the probability of exceeding the design flow (p_{Qm}) was observed (Figs 5a and 5 b). As Q_m increases, the value of p_{Qm} initially rises to a peak and then decreases. A decrease in rainfall depth (P_t) leads to a decrease in p_{Qm}, while an inverse relationship was found for F_{imp}. The highest p_{Qm} values (σ=20%) were obtained for rainfall type R3 ($p_{Qm} = 0.70$), and the lowest for R2 ($p_{Qm} = 0.51$) (Figure 5b). Additionally, increasing the allowable uncertainty threshold in Q_m leads to a reduction of p_{Qm}. For rainfall types R3 and R4, the value of σ has a negligible influence on the maximum flow. However, for rainfall type R2, it leads to a decrease in $Q_m = f(\max\{p_{Qm}\})$ by 4% (from 1270 $L \cdot s^{-1}$ to 1220 $L \cdot s^{-1}$).

The figures (Fig. 5a and Fig. 5b) confirmed the strong interaction between catchment characteristics (F_{imp}, FD), rainfall data, and the acceptable uncertainty threshold (σ). A change in rainfall distribution, from R3 to R2 (F_{imp} = 24.26 ha), in the context of calculations for verifying sewer capacity, performance evaluation, or climate-smart sewer design, leads to a 33.3% increase in Q_m (Figure 5b). Meanwhile, an increase in P_t = 6.0 mm to P_t = 8.0 mm (under R2 temporal rainfall distribution) results in a 68% increase in Q_m (Figure 5a). The calculation results for the remaining variants are presented in Figure A8 in Appendix A.

An increase in catchment imperviousness (from F_{imp} = 16.26 ha to F_{imp} = 24.26 ha), for rainfall duration t_r = 30 min

and P_t = 6.0 mm (type R2), led to a 42% increase in Q_m. It was found that as σ increases, the range of variability of Q_m also increases for the assumed values of p_{Qm} = const, across rainfall types R2, R3, and R4 (Figure 5b). In practice, an increase in σ indicates a greater reserve of sewer capacity at the design stage and an increase in the reliability of system operation during heavy rainfall.

In addition to catchment characteristics, the significant influence of SWMM model parameters on Q_m simulation results was confirmed (Figures 5c and 5d). It was found that a reduction in the Manning roughness coefficient of impervious areas (n_{imp}) and the sewer roughness coefficient (n_{sew}) leads to an increase in Q_m (0.5th percentile), while the impervious area correction factor (β) shows an inverse relationship. For a rainfall duration of t_r = 30 min and P_t = 8.0 mm, it was shown that despite an increase from F_{imp} = 11.17 ha (FD = 1.08) to F_{imp} = 15.26 ha (FD=1.14), as a result of catchment urbanization, it is possible to maintain a constant flow of Q_m = 1000 $L \cdot s^{-1}$ by increasing the Manning roughness coefficient from n_{imp} = 0.0157 $m^{-1/3} \cdot s$ to 0.018 $m^{-1/3} \cdot s$ (Figure 5d).

The inclusion of the likelihood function was shown to have a significant impact on the calculated values of d_{imp}, n_{imp}, n_{sew}, and β, as confirmed by the resulting curves (Fig. 5c, d). The introduction of the likelihood function during the identification of SWMM parameters influencing Q_m, under a defined uncertainty threshold (σ), leads to an increase in their values by approximately 2 – 15% compared to the variant where the likelihood function is omitted.

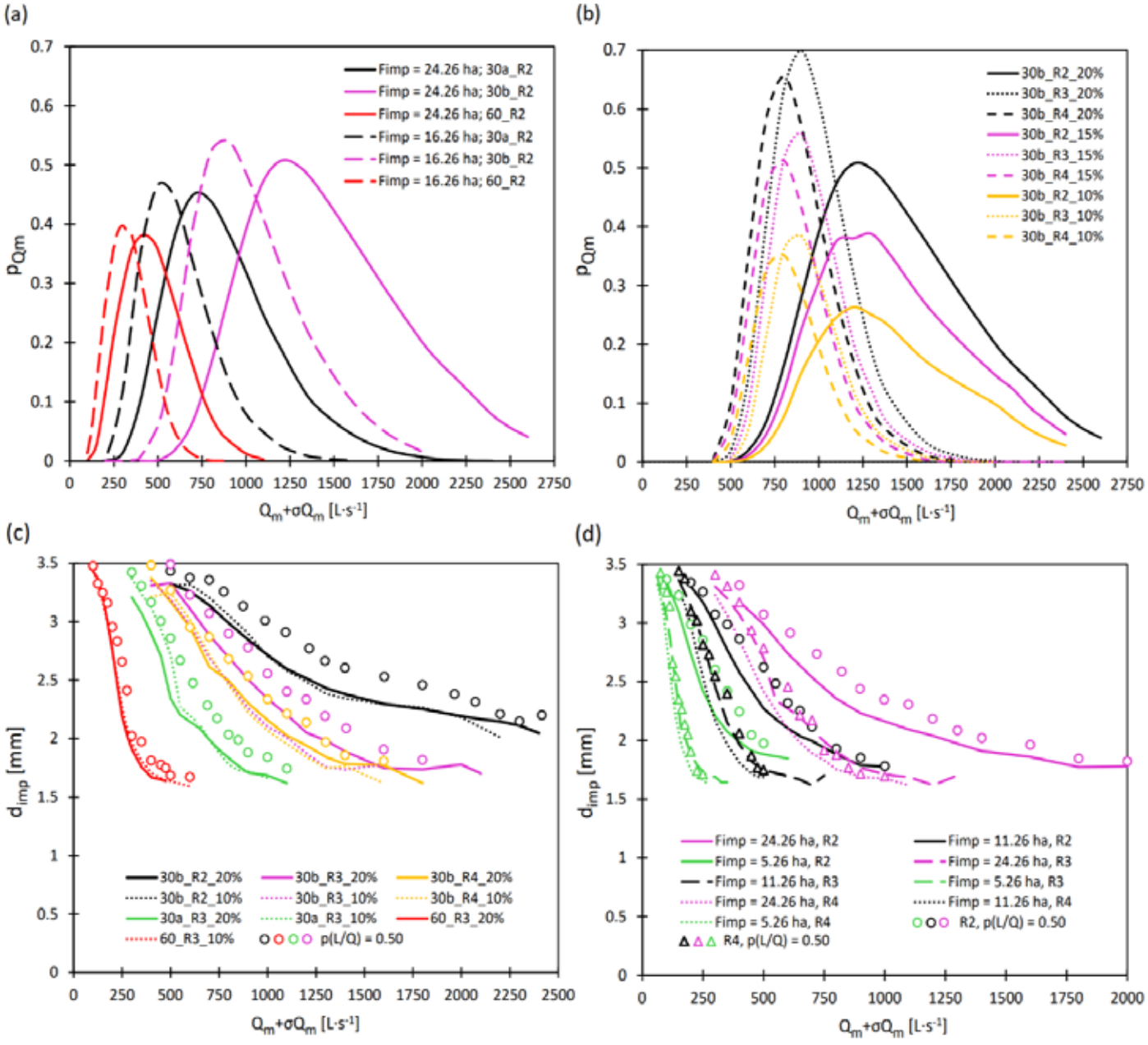

Figure 5. (a) Influence of maximum design flow (Q_m), catchment characteristics (F_{imp}), rainfall data on p_{Qm} probability; (b) Influence of maximum design flow (Q_m), rainfall distribution (type R2, R3, R4), acceptable uncertainty threshold (σ) on p_{Qm} probability; (c) Influence of maximum design flow (Q_m), rainfall data, permissible uncertainty threshold (σ) on d_{imp} with neglect and consideration of likelihood of SWMM parameters; (d) Influence of maximum design flow (Q_m), catchment characteristics (F_{imp}), temporal rainfall distribution on d_{imp} with neglect and consideration of reliability of SWMM parameters.

Discussions

Model for simulating maximum flow

The MARS method was used to simulate processes of 5 sub-catchments and identify the condition of conduits, flows, slopes, etc. To date, this method has not been applied to the calculation of maximum flows in sewer lines (Table 5), even though it has been used to simulate flows in river catchments (Fig. 1). The ML computational models presented in this study differ from those developed by other authors (She and You, 2019; Yang and Chui, 2020; Kratzert et al., 2019) in that they are expressed as analytical equations. This format allows the influence of independent variables on maximum flow simulation results to be assessed directly, without the need for additional calculations. In the present study, non-linear relationships between rainfall data and the estimated

Table 5. Review of a mechanistic machine learning model incorporating input data for predicting maxi-mum flow

Studies	Variable	Methods	Input				
			RN	SW	CT	SP	UN
Vojnovic et al. (2003)	Q(t)	ML (MLP)	1	0	0	0	0
Zhou et al. (2014)	Q	ML (SLRM)	1	1	1	1	0
Yao et al. (2016)	Qm, V	ML (MLR)	1	0	1	1	0
Gargano et al. (2015)	Q(t)	ML (SVM)	1	0	0	0	0
Yao et al. (2017)	Qm	MC (MLR)	1	0	1	1	0
Liu et al. (2017)	Qm	MC (MQR)	1	1	1	0	0
Jato-Espino et al. (2017)	Qm	ML (MQR)	0	0	0	0	1
Jato-Espino et al. (2018)	Qm	ML (MQR)	1	1	1	0	0
Kratzert et al. (2019)	Qm	ML (LSTM)	1	0	1	0	0
She and You (2019)	Q(t)	MC (RBF-NARX)	1	0	0	0	0
Yang and Chui (2020)	Q(t)	ML (XGboost)	1	0	0	0	0
Bell et al. (2020)	Qm, V	ML (logit)	0	0	1	0	0
Fatone et al. (2021)	Qm	ML (logit)	1	0	0	0	1
Palmitessa et al. (2022)	Q(t), h, Vov	ML (DNN)	1	0	0	0	0
Meierdiercks et al. (2010)	Q(t)	MC (SWMM)	1	1	1	1	1
Mejia et al. (2010)	Q(t)	MC(GUIH)	1	1	1	1	1
Cantone and Schmidt (2011)	Q(t)	MC (IUHM)	1	1	1	1	1
Zhang et al. (2014)	Q(t)	MC (GSSH)	1	1	1	1	1
Chadalawada et al. (2017)	Q(t)	MC (IUH)	1	1	1	1	1
Ichiba et al. (2018)	Q(t)	SWMM	1	1	1	1	1
Qui et al. (2020)	Q(t)	SWMM	1	1	1	1	1
Wolfs and Willems (2016)	Q(t)	MC+ML	1	1	1	1	1
this study	**Qm**	**MC (MARS)**	**1**	**1**	**1**	**1**	**1**

where: RN – rainfall; SW – sewer characteristics; CT – catchment characteristics; SP – spatial; UN – uncertainty (calibration parameter describe of the storage catchment, capacity of sewer); 1 – functional; 2 – unfunctional; SVM – support vector machines; DNN – deep neural network; XGboost – Extreme Gradient Boosting; MLR – multinomial linear regression; MQR – mulitnomial polynomial regression; RBF – radial basis function; LSTM – long short term memory; MLP – multilayer perceptron; MC – mechanistic model; GUIH – Geomorphological Unit Hydrogram Instantaneous Hydrogram; IUHM – Illinous Unit Hydrogram Model; GSSH - ; IUH – Instantaneous Unit Hydrogram.

coefficients in the MARS model were identified, as confirmed by the obtained correlation coefficients. However, it was not possible to establish generalized relationships for determining these coefficients, which means the developed model is not universal.

Similar case-study-based approaches were proposed by Fatone et al. (2021), who considered SWMM parameters (α, n_{imp}, $d_{imp,}$ n_{sew}) and rainfall data but neglected catchment and sewer network storage and topology. Yao et al. (2016) considered catchment imperviousness and, in a simplified manner, the layout of the sewer network by classifying 6 sub-catchment types. Bell (2020) included only impervious areas in his model. Among the reviewed studies, only Jato-Espino et al. (2018) developed a model for maximum flow prediction that simultaneously considers catchment imperviousness and sewer storage as a function of rainfall frequency. This model shows potential for application to other catchments, but it has not been widely studied.

Impact of catchment characteristics and sewer sewage system topology on maximum flow

It was found that the impervious area of the catchment contributes significantly to an increase in maximum flow, while other catchment characteristics were not statistically significant. This finding is consistent with the analyses of Yao et al. (2016) and Jato-Espino et al. (2018) for small urban catchments. Bell et al. (2020), based on a review of 55 publications, developed a logistic regression model to estimate the probability of exceeding the maximum flow based on the area of green infrastructure (GI) facilities in the catchment and rainfall data. However, the spatial layout of these features was not considered due to limited available data. Among drainage network characteristics, both the fractal dimension and sewer density were found to influence maximum flow, with observed correlations to rainfall depth. Yao et al. (2016) showed the influence of land use spatial layout and sewer system topology using 5 sub-catchments in Beijing. However, their study did not attempt to parameterize these spatial features using indicators such as fractal dimension or semivariogram parameters.

To date, fractal dimension has been determined through mechanistic model analyses, where its influence on sewer system performance has been confirmed (Gires et al. 2017). Ogden et al. (2011), using the GSSH model, showed that increasing sewer density leads to higher peak flows. This effect was attributed to changes in the routing of the sewer network in the catchment, but the spatial configuration of the drainage system and its effect on runoff were ignored. These findings are consistent with the results of Gironas et al. (2010), who used the advanced GIUH model to simulate sewer network outflow hydrogram. Gires et al. (2017) further demonstrated the potential of an integrated modeling tool combining MCM and fractal geometry to simulate sewer systems, impervious area, and building rooftops across 3 different networks.

Despite these advancements, no prior study has fully established a quantifiable relationship between catchment characteristics and maximum flow. The relationship identified in this study, linking fractal dimension, catchment properties, and sewer network characteristics, enhances our understanding of the interactions among these factors and their implications for urban catchment hydrology.

Impact of Rainfall Characteristics on the Relationship: SWMM parameters – maximum flow (Sensitivity Analysis)

The complex influence of rainfall load on the coefficients in the MARS model has been established. An increase in the values of SWMM parameters (d_{imp}, n_{sew}, β, etc.), depending on the range of their variability, can either increase or decrease Q_m. These results showed the influence of the spatial layout of land use and sewer network configuration on the outflow hydrogram, as different layouts generate different hydraulic conditions in the sewer network, particularly in urban catchments (Fig. 1). In many previous studies (Szeląg et al. 2016, Barco et al. 2008), the impact of calibrated parameters on sewer performance (hydrogram, flooding volumes, etc.) was analyzed using simplified approaches, which often involved a single rainfall event, limited combinations of calibrated parameters, and the neglect of temporal rainfall distribution, leading to overly generalized conclusions (Szeląg et al. 2022 b). Fatone et al. (2021) found that increases in d_{imp}, n_{imp}, n_{sew} could lead to a decrease in maximum flow, while an opposite relationship was observed for α. The difference in methodological approaches, classification models (logit) versus regression models (MARS), enabled the identification of consistent relationships across both modeling frameworks.

The results of the calculations showed that for rainfall with t_r = 30 min (type R2), the Manning roughness coefficient of sewer (n_{sew}) has the greatest influence on the maximum flow, and for rainfall with t_r = 60 min, the storage depth of impervious areas (d_{imp}) was the most influential. The influence of the temporal distribution of rainfall and catchment characteristics on the results of sensitivity analysis was confirmed by Cristiano et al. (2018), who performed analyses for 4 urban catchments (2.5 - 5.3 km^2). Similar results were obtained by Fraga et al. (2016) for a small road catchment (0.005 km^2), who found an influence of rainfall data on sensitivity coefficients and also showed the key influence of n_{imp} on the hydrogram. Fatone et al. (2021), using a logistic regression method (rainfall types R2, R3, and R4), confirmed the highest impact of n_{sew} on maximum flow (Tables 3 and 4). They found that an increase in rainfall intensity led to a reduction in sensitivity coefficients, with highest values obtained for type R1 rainfall distribution and the lowest for type R2. Krebs et al. (2014) (Imp = 86%) confirmed the key influence of sewer roughness coefficient on the catchment outflow hydrogram. On the other hand, Barco et al. (2008) (Imp = 65%), Krebs et al. (2014) (Imp = 19, 86%) found that maximum flow and runoff volume were most influenced by depression storage and catchment imperviousness; their calculations did not include rainfall data.

Risk analysis

Most current risk analysis methods (Fu and Butler 2014, Ursino et al. 2015) focus on the impact of rainfall uncertainty on the performance of the sewer networks, particularly regarding maximum flow, flood volume, and manhole overflows. , etc.). In contrast, the proposed risk analysis method considers not only rainfall uncertainty but also the characteristics of the catchment and sewer network, as well as uncertainty in depression storage and sewer capacity, as represented by SWMM parameters. The model includes a so-called acceptable uncertainty threshold for maximum flow calculations, which

can be interpreted as a reserve of canal capacity. This reserve accounts for climate change and varying catchment conditions, which is important given potential changes in catchment storage and the hydrological cycle. A similar approach was proposed by Sharifi et al. (2024), but their model was limited to analyzing the impact of climate change on sewer network design.

Soft-sensors for predicting sewer sewage system performance – field studies

Literature data (Table 6) confirm that an alternative to classical measurement methods is the used of so-called soft sensors, which operate based on ML models. This approach is gaining increasing attention, as evidenced by the growing number of studies published in this area (Guo et al. 2023). For the development of such models, it is essential to collect an sufficiently large datasets, which requires continuous monitoring of rainfall and sewer network parameters such as flow, flooding, and water depth. ML models have been successfully applied to predict catchment flow rates, overflows, water levels, flood volumes and water depths (Li et al. 2022, Guo et al. 2023). However, most of these models are case-specific and typically developed for individual catchments. As a result, their applicability is limited and dedicated models must be developed for each catchment restricting their universality. Consequently, continuous data collection is required for every application, leading to operational costs associated with maintaining high-resolution and high-quality measurement data (Kwon and Kim 2021, Al Mehedi 2023). Additionally, issues such as improper calibration or placement of flow meters can negatively affect the performance of mechanistic models and reduce the accuracy of predictions, limitations that are detailed in Table 7.

The aforementioned disadvantages - technical limitations, data quality and integration issues, and errors resulting from simplified model assumptions, have been addressed in MC models that incorporate rainfall data, land use, sewer network

Table 6. Review publications on the role of ML models in forecasting and cost reduction of measurement systems

No.	Authors and Year	Duration of Study	Measurement Devices and Measurement Resolution	Input to ML model	Measurement Issues
1	J. Liu, HS Cho (2022)	Typhoon Committee conducted urban flood monitoring projects for several years (UFRM and OSUFFIM)	LSTM Sensors (Long Short-Term Memory); automatic image monitoring based on IoT, weather radar, satellite monitoring - weather radar and satellite images providing good spatial and temporal resolution.	Rainfall data – in tropical cyclone (TC) regions, flood events in densely populated high-risk areas in the Asia-Pacific region.	Issues in real-time urban flood forecasting: low accuracy and time-consuming computations, inconsistent access to updated technology, and geographical variability in data collection leading to gaps in measurements.
2	SH Kwon, JH Kim (2021)	16 studies, including five operational UDS studies; research from the last ten years	Flow sensors, UDS system; long-term memory, LSTM - high resolution	Rainfall data, real-time data used – heavy rainfall	Discontinuous data; studies require large sample sizes (i.e., big data); challenges with generalizing ML models, data diversity, especially in the context of overflow prediction.
3	MA Al Mehedi (2023)	Five years, focusing on the rain garden in Villanova, Pennsylvania	IoT smart sensors, humidity sensors at various depths, water depth sensors, and devices monitoring air and sea temperature - high resolution due to continuous data from sensors allowed in the LSTM model.	Flow data; the LSTM model was trained on operational units such as air temperature, outlets at different depths and waters. The goal was to predict water recession rates.	0.51 ha, small catchment area with urban infrastructure for green stormwater retention.
4	G Fu, S Sun (2022)	No data	LSTM sensors; various sensors used for monitoring water, leak detection, sewer defect detection, and forecasting sewer system conditions - high, enabling continuous monitoring and forecasting based on input data.	Data includes forecasts for water demand, leak detection, forecasting sewer system conditions – synthetic data.	Lack of data synchronization; data issues; problems with data values, privacy, and trust in algorithms.
5	A Sharifi (2024)	No data	IoT, smart sensors, a series of sensors for real-time data collection - high, enabling real-time data collection and analysis.	Hydrological parameters (flow, temperature, and precipitation) of urban drainage systems.	Issues with location and inaccurate definitions of graphic concepts.

layout, and surface runoff conditions (Tabuchi 2020, Moon 2023). Guo et al. (2023) showed the predictive generalization capacity of such models to forecast sewage flooding for urban catchments, allowing their use as soft sensors. Since machine learning methods determine input - output relationships based on statistical correlations, the development of a universal model is feasible. This can be achieved by collecting data from various catchments with diverse physical and geographical characteristics, or by using mechanistic models as simulators to generate synthetic datasets for training machine learning models applicable across multiple catchments. This approach aligns with ongoing efforts to develop a low-cost soft sensors capable of predicting sewer network flow rates and identifying the volume and extent of sewage flooding.

Table 7. MC models for predicting flow intensity in sewer network catchments of different sizes

No.	Authors and Year	Catchment Size	Measurement Devices	Measurement Resolution	Measurement Data	Measurement Issues
1	Tabuchi (2020)	Paris, France; 1800 km²; 500 km of main sewage channels, modeled by 3113 nodes (23,000 computational nodes)	Flow control of the Seine in Paris, and the 5-year and 10-year wet and dry flow rates; MAGES - real-time control system in the Paris Metropolitan Region - 150 management sites (basins, valves, and pumps)	Simulation (approx. 3 minutes within 24 hours); 5-minute cycle	Real-time rainfall data; Flow data; Historical and operational data from the sanitation system, including sewer network performance, maintenance records, and real-time environmental data.	Processing rainfall data and configuring real-time data exchange platforms between SIAAP (Interdepartmental Syndicate for Sanitation in the Paris Region) and each partner: limitations related to sewage pollution and the impact of discharges on the natural environment.
2	Tao (2020)	Kentucky, USA; 385 square miles catchment, 11 Ohio River catchments	The RTC system has ~30 control facilities, including retention basins, storage, diversion controls, and pump stations.	RTC system updated based on five-minute data sampling intervals	Sewage flow data; Rainfall forecasts; Water levels; Flows; Operational data: gate positions, pump rates, treatment capacities	Data quality inconsistencies and challenges in integrating diverse data sources; Infrastructure limitations, as well as inadequate spatial and temporal resolution of data.
3	Moon (2023)	Seoul, South Korea; 42.50 km²	Weather system (https://www.kma.go.kr), rain gauge sensors; water level monitoring systems	Rainfall data collected at 1-minute resolution for 21 years (2001–2021).	Rainfall data, water levels; Additional synthetic data generated by a rainfall-runoff model	Errors in R-R simulation results (SWMM), uncertainty in catchment data for estimating parameters is relatively high.
4	Wu (2023)	Wuhan, China; 8494 km²	Pump performance; technologies for monitoring rainfall and floodwater levels	High spatial and temporal resolution	Rainfall data; Overflow data; Pump performance data	Found a nonlinear response relationship between pump performance and the area covered by LID and UFS.
5	M. Zawilski (2013)	Łódź, Poland; 335 ha catchment, sealing degree - 40%; 60 subcatchments (1.79 - 17.16 ha)	Flow sensors, rain gauges, standalone monitoring stations for Total Suspended Solids (TSS)	Continuous quantitative and qualitative monitoring from 1989–1992, pluviometric data from 2010–2012	Precipitation data, flows in the sewage system, suspended solids data	Technical limitations: Discontinuous data, Measurement accuracy issues, Calibration difficulties.
6	Nowogoński (2019)	Poland, 288 ha catchment, 72 subcatchments of 4 ha each	Simulation, rain gauge sensors (so-called pluviometers), flow sensors in channels, and water level sensors	Temporal and spatial resolution	Precipitation data, hydrological data: flow and water level measurements, spatial data - velocity, assuming uniform runoff from rainfall	Errors arise from assuming too few subcatchments, which underestimates maximum runoff and stabilizes it at a lower level.
7	M. Wawrzyniak (2023)	Szczecin, Poland; 300.55 km²	IMGW-PIB meteorological station devices in Szczecin, flow meters	Minute data for 30 years from 1986–2015	Rainfall data; Flows; Real-time ML algorithm training	Low effectiveness of the Błaszczyk formula: Major errors, poor fit to data. Better fit for R&D models (e.g., KOSTRA atlas, PANDa, PMAXTP).

Limitations and future directions

The presented model is conceptual in nature and was developed using data from a single urban catchment with an area of 63 ha. As such, it should be regarded as a case study aimed at demonstrating the potential of machine learning-based soft sensors for predicting peak flows in stormwater systems without the need for continuous flow monitoring. Although the model was calibrated for a specific location, it is designed for broader applicability, both within Poland and internationally, pending further validation with data from catchments exhibiting diverse land use patterns and sewer system configurations.

A principal limitation of the study is the reliance on precipitation data from a single pluviometer located outside the study catchment. While this approach is generally acceptable for small catchments (Del Giudice et al., 2013; Bell et al., 2020), the use of denser rain gauge networks or radar-based rainfall estimates is recommended for larger areas to more accurately capture spatial variability in rainfall.

Hydrodynamic calculations and simulations based on the studies of Barco et al. (2008) and Catone et al. (2011) demonstrated the influence of infiltration processes on surface runoff in large urban catchments. In contrast, the present study focused on a small catchment where the contribution of green areas to runoff generation was negligible. Sobol sensitivity analysis revealed a marginal impact of infiltration-related parameters, particularly d_{imp} and β (Kiczko et al., 2018), which justified their exclusion from the modeling process. These findings are consistent with those of Fatone et al. (2021), who also reported the influence of these parameters as statistically insignificant.

Additionally, uniform parameterization was applied across the catchment, which is a common practice when modeling small urban areas in Poland (Wałęga et al., 2013, Banasik et al., 2014). These studies, focusing on rainfall-runoff modeling using SCS-CN method, demonstrated that spatial variability had a limited effect on peak flow estimation in small catchments, such as the Służewiecki Stream catchment in Warsaw. Nonetheless, future research should account for spatial heterogeneity to further enhance model accuracy.

Calibration events were carefully selected to exclude sewer overflows and surface flooding, thereby ensuring mass balance consistency, as recommended in the literature (Jato-Espino et al., 2017). Consequently, the current model does not account for hydraulic overload scenarios, an important limitation that should be addressed in future work to improve model robustness.

Future research should explore the application of dimensionality reduction techniques and the development of hybrid models that integrate data-driven approaches with conceptual hydrological modeling. Despite its limitations, the proposed soft sensor offers an energy-efficient alternative to traditional flow monitoring devices and shows significant potential for integration into real-time stormwater management systems.

Conclusions

One of the currently major challenges in hydrology is the development of universal models that can be applied to model different urban catchments. Mechanistic models often fall short in this regard due to their limited flexibility.

As part of the research, an attempt was made to develop a universal model that incorporates key factors such as catchment characteristics, sewer network topology, sewer storage, and rainfall data. This process involved both model construction and sensitivity analysis. However, it was not possible to parameterize the estimated coefficients of the MARS model with rainfall data characteristics. As a result, the proposed models lack universality and cannot be directly transferred to other urban catchments. Despite this limitation, the obtained results revealed complex relationships between depression storage, sewer network, and peak flow, which were captured through the analytical structure of the MARS model. Notably, although empirical relationships for unambiguous coefficient identification could not be established, a very strong correlation was observed between the estimated coefficients and rainfall data. This finding holds significant cognitive value in understanding the behavior of the studied catchment.

Also noteworthy is the developed risk analysis method, which accounts for both the reliability of calibrated parameters and the capacity of sewers. This analysis was conducted based on the existing system conditions and incorporated a safety margin coefficient that defines the potential excess capacity. This approach enabled the estimation of the additional volume of wastewater that the network could convey under extreme conditions. The introduction of this coefficient provides a valuable metric for assessing and enhancing the adaptability of existing sewer networks to future climate change impacts.

Considering the utilitarian nature of the approach and the potential application of the obtained results to the analyzed catchment, further research is warranted. The goal of future work will be to develop universal relationships that can be applied to other catchments. Ultimately, objective is to create a tool that will enable efficient management of urban catchments using modern hydrological solutions, including zero-emission soft sensors that require no physical installation and generate zero energy consumption. This tool will be designed to adapt to evolving hydrological and climatic conditions, enabling effective monitoring of phenomena such as heavy rainfall or droughts. In turn, it will support sustainable water resource management, increase the resilience of urban infrastructure to climate change impacts, and help mitigate the risk of damage caused by extreme weather events.

References

Addison-Atkinson, W., Chen, A.S., Memon, F.A. & Chang, T.J., 2022. Modelling urban sewer flooding and quantitative microbial risk assessment: a critical review. *J. Flood Risk Mangement*. 15, e12844.

Amiri, H., Azadi, S., Javadpour, S., Naghavi, A.A. & Boczkaj, G., 2022. Selecting wells for an optimal design of groundwater monitoring network based on monitoring priority map: a Kish Island case study. *Water Resour. Ind*. 27, 100172 DOI:10.1016/j.wri.2022.100172.

Bach, P. M., Kuller, M., McCarthy, D. T. & Deletic, A. (2020). A spatial planning-support system for generating decentralised urban stormwater management schemes. *Science of The Total Environment*, 138282. DOI:10.1016/j.scitotenv.2020.138282.

Banasik, K., Krajewski, A., Sikorska-Senoner, A. E. & Hejduk, L. (2014). Curve number estimation for a small urban catchment

from recorded rainfall-runoff events. *Archives of Environmental Protection*, 40, 3, pp. 21–30. DOI:10.2478/aep-2014-0032.

Barco, J., Wong, K.M. & Stenstrom, M.K. (2008). Automatic calibration of the U.S. EPA SWMM model for a large urban catchment. *J. Hydraul. Eng*. 134, pp. 466–474. DOI:10.1061/(ASCE)0733-9429(2008)134:4(466).

Bell, C. D., Wolfand, J. M., Panos, C. L., Bhaskar, A. S., Gilliom, R. L., Hogue, T. S. & Jefferson, A. J. (2020). Stormwater control impacts on runoff volume and peak flow: A meta-analysis of watershed modeling studies. *Hydrological Processes*. DOI:10.1002/hyp.13784.

Beven, K. (2020). Deep learning, hydrological processes and the uniqueness of place. *Hydrological Processes*, 34, 16, pp. 3608–3613. DOI:10.1002/hyp.13805.

Bhaskar, A. S., Hogan, D. M., Nimmo, J. R. & Perkins, K. S. (2018). Groundwater recharge amidst focused stormwater infiltration. *Hydrological Processes,* 32, 13, pp. 2058–2068. DOI:10.1002/hyp.13137.

Birgani, Y. T., Yazdandoost, F. & Malaekepour, S. M. (2013). Urban Drainage Systems Risk Assessment under Climate Change Conditions.

Butler, D., Farmani, R., Fu, G., Ward, S., Diao, K. & Astaraie-Imani, M. (2014). A new approach to urban water management: Safe and sure. *Procedia Engineering*, 89, pp. 347-354.

Cantone, J. & Schmidt, A. (2011). Improved understanding and prediction of the hydrologic response of highly urbanized catchments through development of the Illinois Urban Hydrologic Model. *Water Resources Research*, 47, 8. DOI:10.1029/2010wr009330.

Chadalawada, J., Havlicek, V. & Babovic, V. A (2017). Genetic Programming Approach to System Identification of Rainfall-Runoff Models. *Water Resour Manage* 31, pp. 3975–3992. DOI:10.1007/s11269-017-1719-1.

Cristiano, E., ten Veldhuis, M.C., Gaitan, S., Rodriguez, S.O. & van de Giesen, N. (2018). Critical scales to explain urban hydrological response. *Hydrol. Earth Syst. Sci. Discuss.* DOI:10.5194/hess-2017-715.

Cristiano, E., ten Veldhuis, M.C., Wright, D.B., Smith, J.A. & van de Giesen, N. (2019). The influence of rainfall and catchment critical scales on urban hydrological response sensitivity. *Water Resour. Res*. 55, pp. 3375–3390. DOI:10.1029/2018WR024143.

De Paola, F. & Ranucci, A. (2012). Analysis of spatial variability for stormwater capture tank assessment. *Irrig. Drain*. 61, pp. 682–690. DOI:10.1002/ird.

Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P. & Rieckermann, J. (2013). Improving uncertainty estimation in urban hydrological modeling by statistically describing bias. *Hydrology and Earth System Sciences*, 17, 10, pp. 4209–4225. DOI:10.5194/hess-17-4209-2013.

Duan, H., Akker, B. van den, Thwaites, B. J., Peng, L., Herman, C., Pan, Y. & Ye, L. (2020). Mitigating nitrous oxide emissions at a full-scale wastewater treatment plant. *Water Research,* 116196. DOI:10.1016/j.watres.2020.116196

DWA-A 118E, 2006. Hydraulic Dimensioning and Verification of Drain and Sewer Systems. DWA German Association for Water, Wastewater and Waste, Hennef, Germany.

Fatone, F., Szeląg, B., Kiczko, A., Majerek, D., Majewska, M., Drewnowski, J. & Łagód, G. (2021). Advanced sensitivity analysis of the impact of the temporal distribution and intensity of rainfall on hydrograph parameters in urban catchments. *Hydrol. Earth Syst. Sci*. 25, pp. 5493–5516. DOI:10.5194/hess-25-5493-2021.

Fraga, I., Cea, L., Puertas, J., Su´arez, J., Jim´enez, V. & J´acome, A. (2016). Global sensitivity and GLUE-based uncertainty analysis of a 2D–1D dual urban drainage model. *J. Hydrol. Eng*. 21, 04016004. DOI:10.1061/(ASCE)HE.1943-5584.0001335.

Friedman, J. H. & Roosen, C. B. (1995). An introduction to multivariate adaptive regression splines. *Statistical Methods in Medical Research*, 4, 3, pp. 197–217. DOI:10.1177/096228029500400303.

Fu, G. & Butler, D. (2014). Copula-based frequency analysis of overflow and flooding in urban drainage systems. *J. Hydrol*. 510, pp. 49–58. DOI:10.1016/j.jhydrol.2013.12.006.

Fu, G., Jin, Y., Sun, S., Yuan, Z. & Butler, D. (2022). The role of deep learning in urban water management: A critical review. *Water Research* 223, 1 118973, DOI:10.1016/j.watres.2022.118973.

Gargano, R., Di Palma, F., de Marinis, G., Granata, F. & Greco, R. (2015). A stochastic approach for the water demand of residential end users. *Urban Water Journal*, 13, 6, pp. 569–582. DOI:10.1080/1573062x.2015.1011.

Gires, A., Tchiguirinskaia, I., Schertzer, D., Ochoa-Rodriguez, S., Willems, P., Ichiba, A. & ten Veldhuis, M.-C. (2017). Fractal analysis of urban catchments and their representation in semi-distributed models: imperviousness and sewer system. *Hydrology and Earth System Sciences*, 21, 5, pp./ 2361–2375. DOI:10.5194/hess-21-2361-2017.

Gironás, J., Roesner, L. A., Rossman, L. A. & Davis, J. (2010). A new applications manual for the Storm Water Management Model (SWMM*), Environ. Modell. Softw*., 25, pp. 813–814. DOI:10.1016/j.envsoft.2009.11.009, 2010.

Gong, Y., Zhang, G., Hao, Y. & Nie, L., 2022. Enrichment Evaluation of Heavy Metals from Stormwater Runoff to Soil and Shrubs in Bioretention Facilities. Water, 14(4), 638. DOI:10.3390/w14040638

Guan, M., Sillanpää, N. & Koivusalo, H. (2015). Storm runoff response to rainfall pattern, magnitude and urbanization in a developing urban catchment. *Hydrological Processes*, 30, 4, 543-557. DOI:10.1002/hyp.10624.

Guo, K., Guan, M. & Yu, D. (2021). Urban surface water flood modelling – a comprehensive review of current models and future challenges. *Hydrol. Earth Syst. Sci*. 25, pp. 2843–2860. DOI:10.5194/hess-25-2843-2021.

Guo, K., Guan, M., Haochen, Y. & Xilin, X. (2023). A spatially distributed hydrodynamic model framework for urban flood hydrological and hydraulic processes involving drainage flow quantification. *Journal of Hydrology*. 625. DOI:10.1016/j.jhydrol.2023.130135.

Hauger, M. B., Mouchel, J. M. & Mikkelsen, P. S. (2006). Indicators of hazard, vulnerability and risk in urban drainage. *Water science and technology*, 54, 6-7, pp. 441-450. DOI:10.2166/wst.2006.622.

Ichiba, A., Gires, A., Tchiguirinskaia, I., Schertzer, D., Bompard, P. & Ten Veldhuis, M.-C. (2018). Scale effect challenges in urban hydrology highlighted with a distributed hydrological model. *Hydrology and Earth System Sciences*, 22, 1, pp. 331–350. DOI:10.5194/hess-22-331-2018.

Jato-Espino, D., Sillanpää, N., Andrés-Doménech, I. & Rodriguez-Hernandez, J. (2017). Flood Risk Assessment in Urban Catchments Using Multiple Regression Analysis. *Journal of Water Resources Planning and Management*, 144, 2. DOI:10.1061/(ASCE)WR.1943-5452.0000874.

Jato-Espino, D., Sillanp¨a¨a, N., Andr´es-Dom´enech, I. & Rodriguez-Hernandez, J. (2018). Flood risk assessment in urban catchments using multiple regression analysis. *J. Water Resour. Planning Manag*. 144, 04017085. DOI:10.1061/(ASCE)WR.1943-5452.0000874.

Kiczko, A., Szeląg, B., Kozioł, A. P., Krukowski, M., Kubrak, E., Kubrak, J. & Romanowicz, R. J. (2018). Optimal capacity of a stormwater reservoir for flood peak reduction. *Journal of Hydrologic Engineering*, 23, 4. DOI:10.1061/(ASCE)HE.1943-5584.0001636

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S. & Nearing, G. S. (2019). Towards Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning. *Water Resources Research*. DOI:10.1029/2019wr026065

Krebs, G., Kokkonen, T., Valtanen, M., Setälä, H. & Koivusalo, H. (2014). Spatial resolution considerations for urban hydrological modelling, *J. Hydrol*., 512, pp. 482–497. DOI:10.1016/j.jhydrol.2014.03.013.

Kumar, S., Agarwal, A., Ganapathy, A., Villuri, V.GK. Pasupuleti, S., Kumar, D., Kaushal, D.R., Gosain, A.K. & Sivakuma, B. (2022). . Impact of climate change on stormwater drainage in urban areas. *Stoch Environ Res Risk Assess* 36, pp. 77–96. DOI:10.1007/s00477-021-02105-x.

Kwon, S. H. & Kim, J. H. (2021). Machine Learning and Urban Drainage Systems: State-of-the-Art Review. *Water,* 13, 24, 3545. DOI:10.3390/w13243545.

Li, X. & Willems, P. (2020). A hybrid model for fast and probabilistic urban pluvial flood prediction. *Water Resour. Res*. 56, 6. DOI:10.1029/2019WR025128.

Li, X., Erpicum, S., Mignot, E., Archambeau, P., Pirotton, M. & Dewals, B. (2022). Laboratory modelling of urban flooding. *Sci. Data*, 9, 1. DOI:10.1038/s41597-022-01282-w.

Liu, C. Y. & Chui, T. F. M. (2017). Factors influencing stormwater mitigation in permeable pavement. *Water*, 9, 12, 988. DOI:10.3390/w9120988.

Liu ,J., Cho, H-S., Osman, S., Jeong, H-G. & Lee, K. (2022). Review of the status of urban flood monitoring and forecasting in TC region, Tropical Cyclone Research and Review 11, 2, pp.103-119.. DOI:10.1016/j.tcrr.2022.07.001.

Luiso, J. E. & Giribet, J. I. (2017). Optical flow sensor. 2017 XVII Workshop on Information Processing and Control (RPIC). DOI:10.23919/rpic.2017.8211652.

Meierdiercks, K.L., Smith, J.A., Baeck, M.L. & Miller, A.J. (2010). Analyses of urban drainage network structure and its impact on hydrologic response. *JAWRA Journal of the American Water Resources Association* 46, pp. 932–943. DOI: 10.1111/j.1752-1688.2010.00465.x.

Mejía, A. I. & Moglen, G. E. (2010). Impact of the spatial distribution of imperviousness on the hydrologic response of an urbanizing basin. *Hydrological Processes*, 24, 23, pp. 3359–3373. DOI:10.1002/hyp.7755.

Al Mehedi, M. A., Amur, A., Metcalf, J., McGauley, M., Smith, V. & Wadzuk, B. (2023). Predicting the performance of green stormwater infrastructure using multivariate long short-term memory (LSTM) neural network. *Journal of Hydrology*, 625, 130076. DOI:10.1016/j.jhydrol.2023.130076.

Mondal, K., Bandyopadhyay, S. & Karmakar, S. (2023). Framework for global sensitivity analysis in a complex 1D-2D coupled hydrodynamic model: Highlighting its importance on flood management over large data-scarce regions. *Journal of Environmental Management,* 332, 117312. DOI:10.1016/j.jenvman.2023.117312

Moon, H., Yoon, S. & Moon, M., (2023). Urban flood forecasting using a hybrid modeling approach based on a deep learning technique. *Journal of Hydroinformatics,*25, 2, pp. 593–610. DOI:10.2166/hydro.2023.203.

Mor´an-Valencia, M., Flegl, M. & Güemes-Castorena, D. (2023). A state-level analysis of the water system management efficiency in Mexico: two-stage DEA approach. *Water Resour. Ind*. 29, 100200 DOI:10.1016/j.wri.2022.100200.

Napiorkowski, J.J., Piotrowski, A.P., Karamuz, E. & Senbeta, T.B. (2023). Calibration of conceptual rainfall-runoff models by selected differential evolution and particle swarm optimization variants. *Acta Geophys*. 71, pp. 2325–2338. DOI:10.1007/s11600-022-00988-0.

Nowogoński, I. (2019). Emporal and spatial variability of rainfall in modelling of stormwater outflows. *CEER*; 29, 4, pp. 267-278. DOI: 10.2478/ceer-2019-0060.

Ogden, F.L., Pradhan, N.R., Downer, C.W. & Zahner, J.A. (2011). Relative importance of impervious area, drainage density, width function, and subsurface storm drainage on flood runoff from an urbanized catchment. *Water Resour. Res*. 47, W12503. DOI:10.1029/2011WR010550.

Palmitessa, R., Grum, M., Engsig-Karup, A.P. & Löwe, R. (2022). Accelerating hydrodynamic simulations of urban drainage systems with physics-guided machine learning. *Water Res*. 223, 118972 DOI:10.1016/j.watres.2022.118972.

Perdikaki, M., Makropoulos, C. & Kallioras, A. (2022). Participatory groundwater modelling for managed aquifer recharge as a tool for water resources management of a coastal aquifer in Greece. *Hydrgeol. J*. 30, 1, pp. 37–58. DOI: 10.1007/s10040-021-02427-8

Qiu, Y., da Silva Rocha Paz, I., Chen, F., Versini, P. A., Schertzer, D. & Tchiguirinskaia, I. (2020). Space variability of hydrological responses of Nature-Based Solutions and the resulting uncertainty. *Hydrology and Earth System Sciences Discussions*, 1-34. DOI:10.5194/hess-2020-468

Rosenzweig, B.R., Cantis, H., Kim, Y., Cohn, A., Grove, K., Brock, J., Yesuf, J., Mistry, P., Welty, C., McPhearson, T., Sauer, J. & Chang, H. (2021). The value of urban flood modeling. *Earth's Future* 9, e2020EF001739. DOI:10.1029/2020EF001739.

Sharifi, A., Beris, A.T., Javidi, A.S., Nouri, M., Lonbar, A.G. & Ahmadi, M. (2024). Application of artificial intelligence in digital twin models for stormwater infrastructure systems in smart cities. *Advanced Engineering Informatics*, 61, 102485. DOI:10.1016/j.aei.2024.102485.

Sharma, S., Lee, B. S., Nicholas, R. E., & Keller, K. (2021). A safety factor approach to designing urban infrastructure for dynamic conditions. *Earth's Future*, 9, 12, e2021EF002118. DOI:10.1029/2021EF002118.

She, L. & You, X.Y. (2019). A dynamic flow forecast model for urban drainage using the coupled artificial neural network. *Water Resour. Manag*. 3, pp. 3143–3153. DOI:10.1007/s11269-019-02294-9.

Shi, B., Catsamas, S., Kolotelo, P., Wang, M., Lintern, A., Jovanovic, D. & McCarthy, D. T. (2021). A Low-Cost Water Depth and Electrical Conductivity Sensor for Detecting Inputs into Urban Stormwater Networks. *Sensors,* 21, 9, 3056. DOI:10.3390/s21093056.

Szeląg, B., (2013). Influence of the Hydrogramme Shape on the Capacity and Selection of Drains of a Small Retention Reservoir. University of Technology, Kielce. PhD thesis.

Szeląg, B., Kiczko, A. & Dąbek, L. (2016). Sensitivity and uncertainty analysis of hydrodynamic model (SWMM) for storm water runoff forecating in an urban basin – A case study, *Ochr. Sr.*, 38, pp. 15–22.

Szeląg, B., Majerek, D., Kiczko, A., Łagód, G., Fatone, F. & McGarity, A. (2022a). Analysis of sewer network performance in context of modernization: modeling, sensitivity, uncertainty analysis. *Journal of Water Resources Planning and Management*. 148: pp[. 1– 12. DOI:10.1061/(ASCE)WR.1943–5452.0001610.

Szeląg, B., Suligowski, R., De Paola, F., Siwicki, P., Majerek, D. & Łagód, G. (2022b). Influence of urban catchment characteristics and rainfall origins on the phenomenon of stormwater flooding: case study. *Environ. Model. Softw*. 150, 105335 DOI:10.1016/j.envsoft.2022.105335.

Tabuchi, J.-P., Blanchet, B. & Rocher, V. (2020). Integrated Smart Water Management of the sanitation system of the Greater Paris region. *Water International*, 45, 6, pp. 574–603. DOI:10.1080/02508060.2020.1830584.

Tao, D. Q., Pleau, M., Akridge, A., Fradet, O., Grondin, F., Laughlin, S. & Shoemaker, L. (2020). Analytics and Optimization Reduce Sewage Overflows to Protect Community Waterways in Kentucky. *INFORMS Journal on Applied Analytics*, 50, 1, pp. 7–20. DOI:10.1287/inte.2019.1022.

Ursino, N. (2015). Risk analysis of sustainable urban drainage and irrigation. *Advances in Water Resources,* 83, pp. 277–284. DOI:10.1016/j.advwatres.2015.06.011.

Vojinovic, Z., Kecman, V. & Babovic, V. (2003). Hybrid Approach for Modeling Wet Weather Response in Wastewater Systems. *Journal of Water Resources Planning and Management,* 129, 6. DOI:10.1061/(ASCE)0733-9496(2003)129:6(511).

Wałek, G. (2019). The influence of roads on the formation of surface runoff in the urbanized area on the example of the Silnica River catchment area in Kielce. Jan Kochanowski University Press, Kielce. (in Polish).

Wałęga, A., Cupak, A., Michalec, B. & Wachulec, K. (2013). The influence of physical and geographical catchment parameters and precipitation characteristics on the runoff time of concentration. *Journal of Civil Engineering, Environment and Architecture*, 30, 60, pp. 143–160. DOI:10.7862/rb.2013.44.

Wang X., Kvaal K. & Ratnaweera H. (2019). Explicit and interpretable nonlinear soft sensor models for influent surveillance at a full-scale wastewater treatment plant. *J. Process Control*, 77, pp. 1-6. 10.1016/j.jprocont.2019.03.005 .

Wawrzyniak, M. & Wdowikowski, M. (2023). Heavy rainfall modeling in an urban catchment area on the example of Szczecin, Gaz, Woda i Technika Sanitarna. pp.22-30. DOI:10.15199/17.2023.7.4.

Willems, P. (2013). Revision of urban drainage design rules after assessment of climate change impacts on precipitation extremes at Uccle, Belgium. *Journal of Hydrology*, 496, pp. 166–177. DOI:10.1016/j.jhydrol.2013.05.037.

Wolfs, V. & Willems, P. (2016). Modular Conceptual Modelling Approach and Software for Sewer Hydraulic Computations. *Water Resources Management*, 31, 1, pp. 283–298. DOI:10.1007/s11269-016-1524-2.

Wong B. P. & Kerkez B. (2016). Real-time environmental sensor data: an application to water quality using web services. *Environmental Modelling and Software* 84, pp. 505–517. DOI:10.1016/j.envsoft.2016.07.020.

Wu, Y., She, D., Xia, J., Song, J., Xiao, T. & Zhou, Y. (2023). The quantitative assessment of impact of pumping capacity and LID on urban flood susceptibility based on machine learning. *Journal of Hydrology*, 617, 129116. DOI:10.1016/j.jhydrol.2023.129116.

Yang, Y. & Chui, T. F. M. (2020). Modelling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods, *Hydrol. And Earth Syst. Sci.*, 25,11, pp. 5839-5858. DOI:10.5194/hess-2020-460, in review, .

Yao, L., Wei, W. & Chen, L. (2016). How does imperviousness impact the urban rainfall-runoff process under various storm cases? Ecological Indicators, 60, pp. 893–905. DOI:10.1016/j.ecolind.2015.08.041 10.1016/j.ecolind.2015.08.041.

Yao, Y., Li, J., Jiang, Y. & Huang, G. (2023). Evaluating the response and adaptation of urban stormwater systems to changed rainfall with the CMIP6 projections. *Journal of Environmental Management,* 347, 1, 119135. DOI:10.1016/j.jenvman.2023.119135.

Zawilski, M. & Sakson, G. (2013). Assessment of total suspended solid emission discharged via storm sewerage system from urban areas. OCHRONA SRODOWISKA, 35, 2, pp. 33-40.

Zhang, K., Liu, Y., Deletic, A., McCarthy, D. T., Hatt, B. E., Payne, E. G. Chandrasena, G., Li, Y., Pham, T., Jamali, B., Daly, E., Fletcher, T.D. & Lintern, A. (2021). The impact of stormwater biofilter design and operational variables on nutrient removal-a statistical modelling approach. *Water Research*, 188, 116486.

Zhou, Q. (2014). A Review of Sustainable Urban Drainage Systems Considering the Climate Change and Urbanization Impacts. *Water*, 6, 4, pp. 976–992. DOI:10.3390/w6040976.

Innowacyjna metoda prognozowania maksymalnego przepływu w systemach kanalizacji deszczowej z wykorzystaniem soft-sensorów.

Streszczenie. Tworzenie uniwersalnych modeli hydrologicznych do modelowania zlewni miejskich jest jednym z największych wyzwań współczesnej hydrologii. Niniejsze badanie podejmuje próbę opracowania modelu uwzględniającego różnorodne cechy zlewni, topologię sieci kanalizacyjnej, retencję w kanałach oraz dane opadowe, a także przeprowadzenia analizy wrażliwości modelu na zmiany parametrów wejściowych. Celem było zbadanie, w jakim stopniu zastosowanie zaawansowanych metod analitycznych, takich jak modele MARS (Multivariate Adaptive Regression Splines) oraz technologia soft-sensorów, może poprawić prognozowanie przepływów szczytowych (Qm) w systemach kanalizacji deszczowej. W badaniu wykorzystano zaawansowane metody analityczne, w tym modele MARS, do prognozowania przepływów szczytowych w systemach kanalizacji deszczowej. W proces modelowania włączono technologię soft-sensorów, uwzględniającą różnorodne cechy zlewni, topologię sieci kanalizacyjnej, retencję w kanałach oraz zmienne dane opadowe. Dodatkowo przeprowadzono analizę wrażliwości w celu oceny reakcji modelu na zmiany parametrów wejściowych. Wyniki pokazały, że połączenie modeli MARS z soft-sensorami pozwala na uzyskanie wysokiej dokładności prognoz (R^2 = 0,96, RMSE = 0,038), pomimo

zmienności warunków opadowych. Jednakże, nie udało się opracować uniwersalnych zależności modelowych z powodu trudności w parametryzacji modelu w kontekście zmiennych danych opadowych. Chociaż proponowane podejście nie prowadzi do stworzenia uniwersalnych narzędzi, dostarcza cennych wskazówek do dalszych badań nad adaptacją systemów kanalizacyjnych do zmieniających się warunków hydrologicznych. Zastosowana metoda analizy ryzyka umożliwiła uwzględnienie przepustowości sieci kanalizacyjnej oraz wprowadzenie współczynnika marginesu bezpieczeństwa, który ocenia elastyczność systemu w kontekście przyszłych zmian klimatycznych. Badanie wskazuje kierunki przyszłych prac nad modelami, które mogłyby być stosowane w różnych zlewniach miejskich, szczególnie poprzez rozwój bezkosztowych, zeroemisyjnych soft-sensorów.

Section A1. GLUE (Generalized Likelihood Uncertainty)

The uncertainty analysis involves the following steps:

- Selection of SWMM model parameter ranges (determination of the prior distribution) – Table A1,
- Simulation of parameters (N=5000N = 5000 samples) using the Monte Carlo method,
- Simulation of runoff hydrographs from the catchment for rainfall-runoff events, considering uncertainty,

The transformation of the a priori distribution P(θ) to the a posteriori P(θ⁄Q) was performed by the likelihood function L(θ⁄Q) using the equations (1A) (2A):

$$P(Q/\theta) = \frac{L(Q/\theta)P(\theta)}{\int L(Q/\theta)P(\theta)} \tag{1A}$$

where: P(θ) represents the prior coefficient distribution (see Table A1), and L(θ⁄Q) is the likelihood function used to assign weights to the Monte Carlo sample based on model fit to observed basin flows Q and P(θ⁄Q), yielding the posterior distribution of model coefficient θ. The likelihood function used is from Romanowicz and Beven (2006):

$$L(Q/\theta) = exp\left(\frac{\sum_{i=1}^{N}\left(Q_i - \widehat{Q}_i\right)^2}{\kappa \cdot \sigma^2}\right) \tag{2A}$$

where: Q_i and $\widehat{Q}_i$ denote the i -values of these measurements and simulated flows; $\kappa$is a scaling factor for the variance σ^2 of model residual, used to adjust the width of the confidence intervals. The value of κ (the factor used to control the variance of the a posteriori distribution) was determined ensuring, that 95% of measured flow points is enclosed by 95% confidence intervals of the model output (Kiczko et al. 2018).

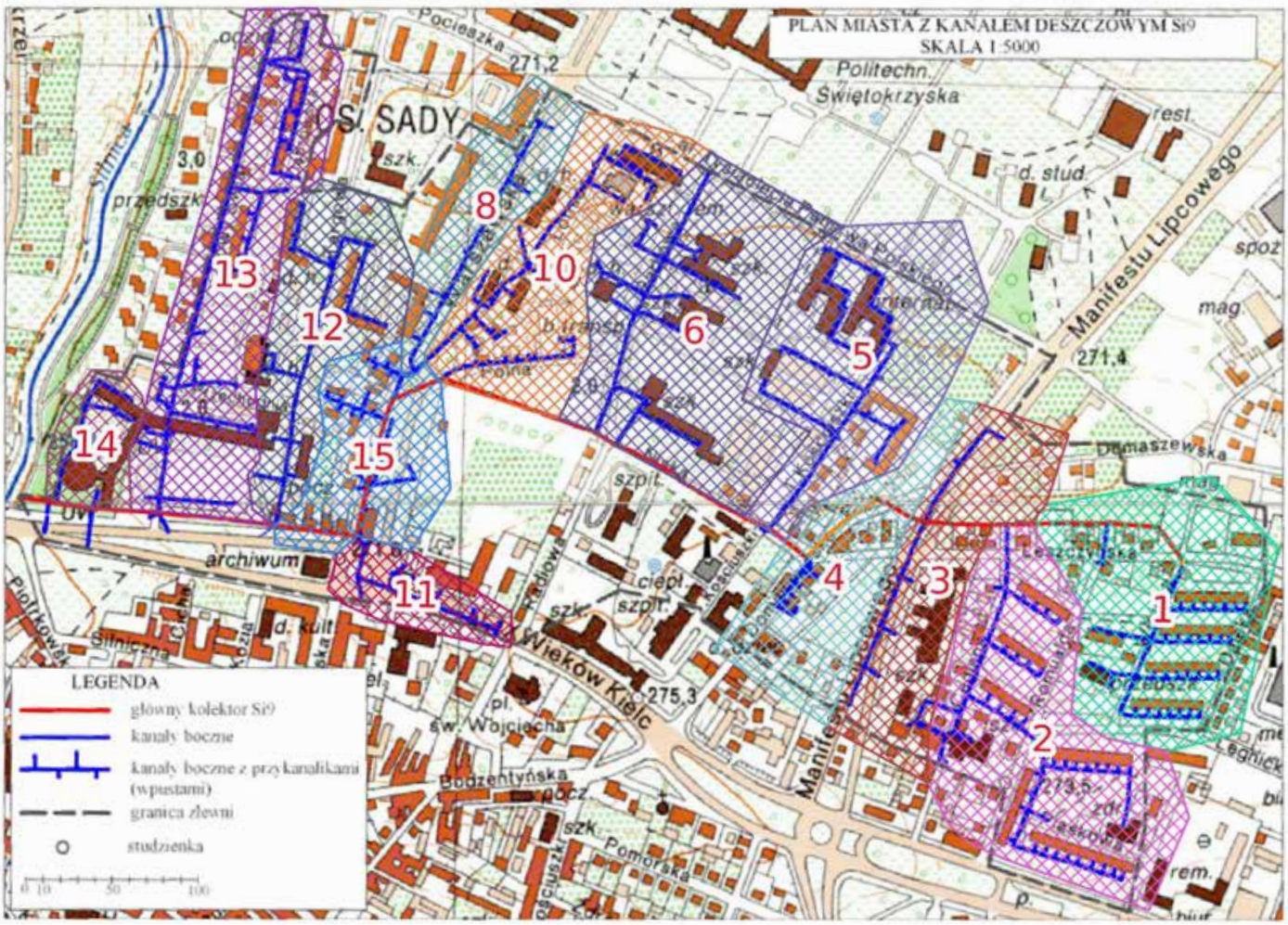

Figure A1. Map of the situational stormwater sewage network

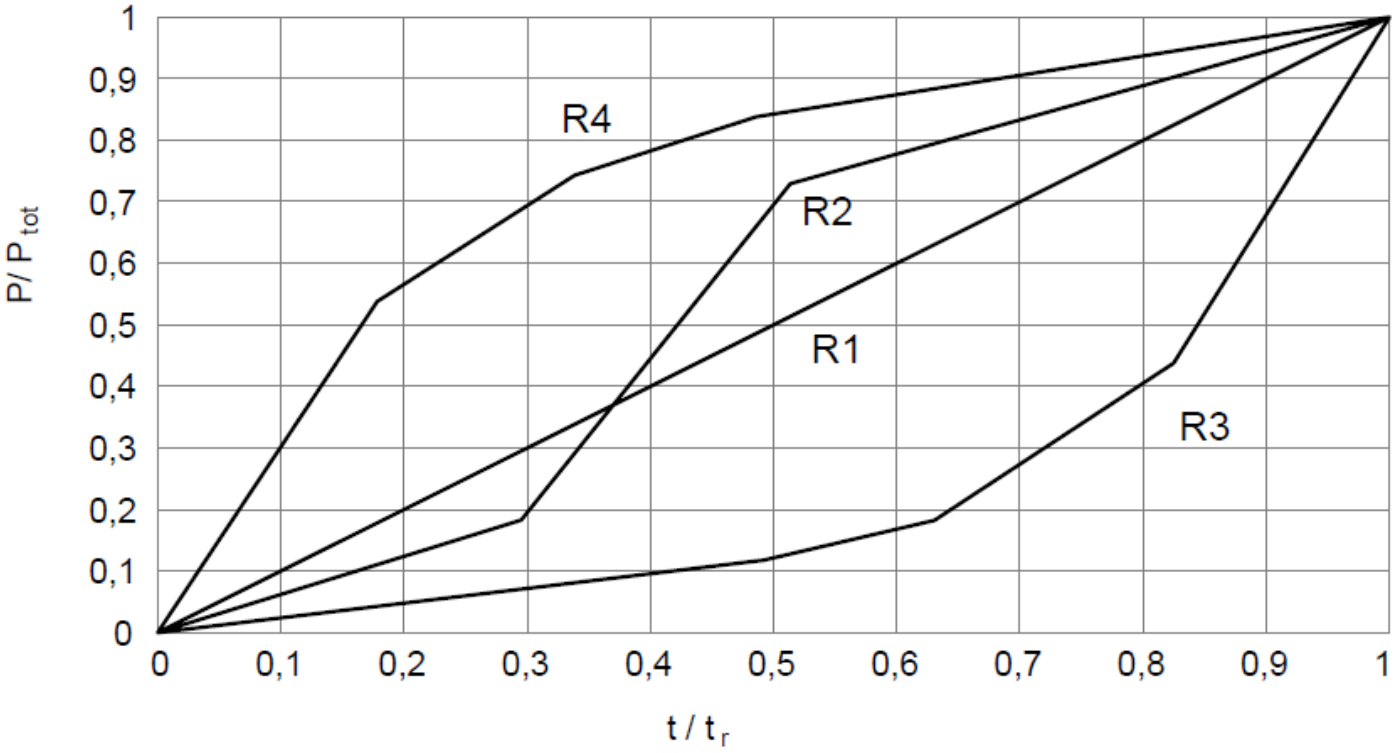

Figure A2. Temporal rainfall distributions (ξ = R1, R2, R3, R4)

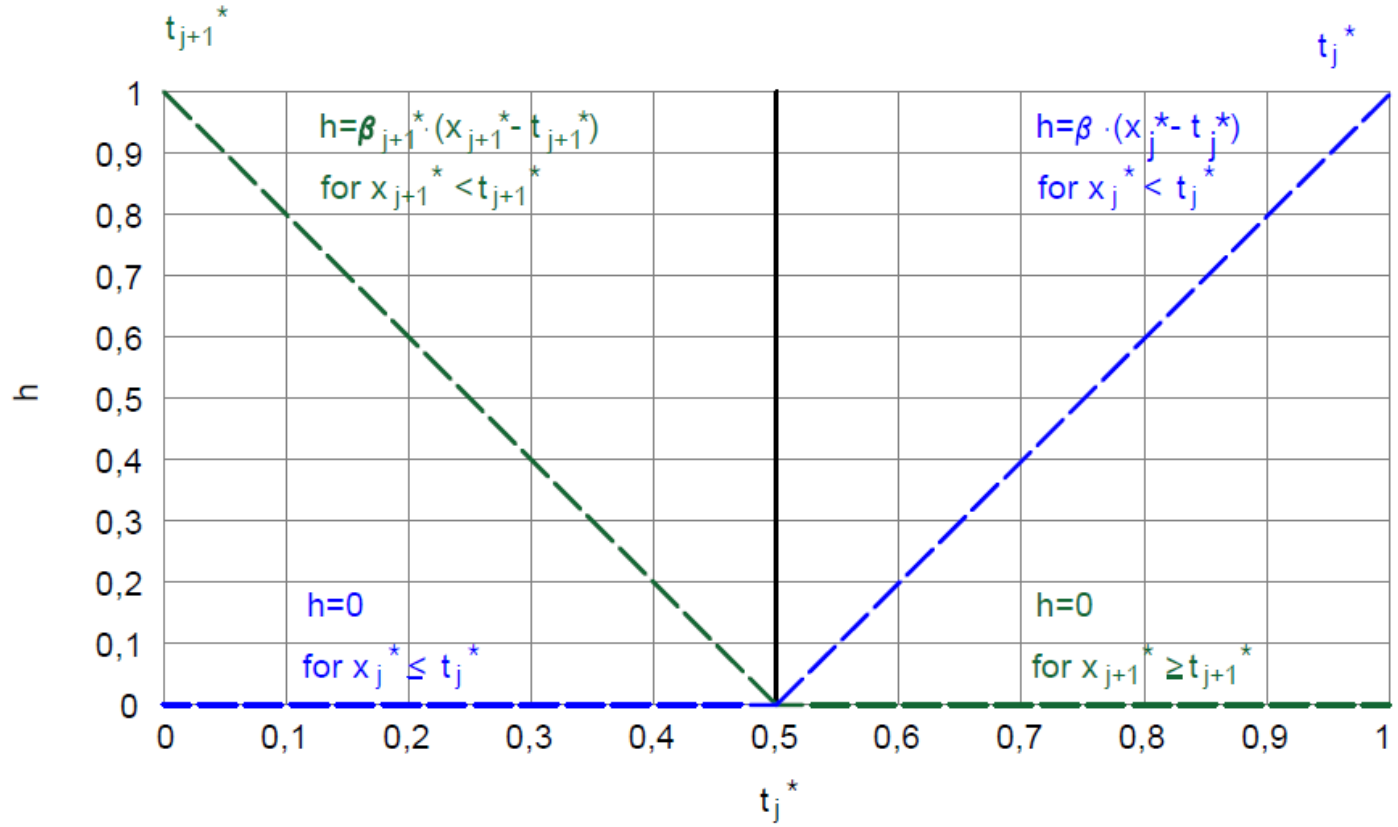

Figure A3. The base function in the MARS model